A system for a computer-based content analysis of interview data

Bierschenk, Bernhard; Bierschenk, Inger

1976

Published by the University of Lund, the Malmö School of Education, and the Society of Psychology and Pedagogy:
STUDIA PSYCHOLOGICA ET PAEDAGOGICA

SERIES ALTERA

1. Edlund, Sven: Diskussionen om begåvningsurvalet under reformations- och stormaktstiden. I. 1947.
2. Agrell, Jan: Nyinrättade professurer inom filosofiska fakulteten i Lund under 1800-talets första hälft. 1949.
3. Smith, Gudmund: Psychological Studies in Twin Differences. 1949.
4. Husén, Torsten: Om innebörden av psykologiska mätningar. Några bidrag till psykometrikens metodlära. 1949.
5. Smith, Gudmund: Interpretations of Behaviour Sequences. 1952.
6. Johannesson, Ingvar: Studier av sociala relationer mellan barn i folkskoleklasser. 1954.
7. Kragh, Ulf: The Actual-Genetic Model of Perception-Personality. 1955.
8. Bjerstedt, Åke: Interpretations of Sociometric Choice Status. 1956.
9. Werdelin, Ingvar: The Mathematical Ability. Experimental and Factorial Studies. 1958.
10. Werdelin, Ingvar: Geometrical Ability and the Space Factors in Boys and Girls. 1961.
11. Borg, Gunnar A. V.: Physical Performance and Perceived Exertion. 1962.
12. Lindell, Ebbe: Den svenska rättskrivningsmetodiken. 1964.
13. Bokander, Ingvar: Stereoscopically Induced Perceptual Conflict between Facial Photographs. A Discussion and Summary of Results. 1968.
14. Bjerstedt, Åke: Twelve Years of Educational and Psychological Research in Sweden: A Bibliography. 1968.
15. Larsson, Bernt: Bayes Strategies and Human Information Seeking. 1968.
16. Bjerstedt, Åke: Educational Technology. Instructional Programming and Didakometry. 1972.
17. Löfgren, Horst: The Measurement of Language Proficiency. 1972.
18. Bierschenk, Bernhard: Självkonfrontation via intern television i lärarutbildningen. 1972.
19. Gestrelius, Kurt: Job Analysis and Determination of Training Needs. Examples of Methods Applied to Teacher Trainers. 1972.
20. Sjöbäck, Hans: The Psychoanalytic Theory of Defensive Processes. 1973.
21. Larsson, Inger: Individualized Mathematics Teaching. 1973.
22. Fredriksson, Lennart: The Use of Self-Observation and Questionnaires in Job Analysis for the Planning of Training. 1974.
23. Idman, Pekka: Equality and Democracy: Studies of Teacher Training. 1974.
24. Bjerstedt, Åke: Social Development and Training in School. 1974.
25. Alwall, Ellen: The Religious Trend in Secular English—Irish Schoolbooks 1850—1861 and 1871—1880. 1974.
26. Holmberg, Ingrid: Effects of Some Trials to Improve Mathematics Teaching. 1975.
27. Lindsten, Carin: Hembygdskunskap i årskurs 3. Att inhämta, bearbeta och redovisa kunskaper. 1975.
28. Bjerstedt, Åke: Pedagogisk forskning i Sverige: En sexårsbibliografi. 1975.
29. Nordén, Kerstin: Psychological Studies of Deaf Adolescents. 1975.
30. Westerlundh, Bert: Aggression, Anxiety and Defence. 1976.
31. Jernryd, Elisabeth: "Optimal auktoritets- och propagandaresistens": En sambands- och utvecklingsstudie. 1976.
32. Bierschenk, Bernhard & Bierschenk, Inger: A System for a Computer-Based Content Analysis of Interview Data. 1976.

Liber

ISBN 91-40-04057-7

Bernhard Bierschenk & Inger Bierschenk: A System for a Computer-Based Content Analysis of Interview Data

BERNHARD BIERSCHENK
INGER BIERSCHENK

A SYSTEM FOR A COMPUTER-BASED CONTENT ANALYSIS OF INTERVIEW DATA

CWK GLEERUP

BERNHARD BIERSCHENK
INGER BIERSCHENK

A System for a Computer-Based Content Analysis of Interview Data

STUDIA PSYCHOLOGICA ET PAEDAGOGICA
SERIES ALTERA XXXII

BERNHARD BIERSCHENK
INGER BIERSCHENK

A System for a Computer-Based Content Analysis of Interview Data

LUND/CWK GLEERUP

CWK Gleerup is the imprint for the scientific and scholarly publications
of LiberLäromedel Lund.

Berlingska Boktryckeriet, Lund 1976
ISBN 91-40-04275-8

Contents

Preface 7

1. *Development of a computer-based content analysis* 9
   - 1.1 Directions for written and spoken text 10
   - 1.2 Adjustment of models to verbal data 12
   - 1.3 Selection of text 12
   - 1.4 Segmentation of text 12
   - 1.5 Unit of analysis 13
   - 1.6 Development of rules for formalization of text 13
   - 1.7 Assignment of codes to text 14
   - 1.8 Supplementation of text 14
   - 1.9 Choice of test criterion 15
   - 1.10 Control of coder agreement 15
   - 1.11 Construction of dictionaries 16
   - 1.12 Quantification of concepts 16
   - 1.13 Design of search logics 16
   - 1.14 Formulation of search questions 17
   - 1.15 Statement of hypotheses 17
   - 1.16 Data processing 17
   - 1.17 Data analysis 18
   - 1.18 Inference 19
   - 1.19 Construction of theories and models 19

2. *A psycholinguistic model for the analysis of research processes* 22
   - 2.1 Sensory coding of information 27
   - 2.2 Indexing information 30
   - 2.3 Operationalization of information 31
   - 2.4 Organization of information 33
   - 2.5 Functionalization of information 35

3. *An empirical analysis of concepts in context* 38
   - 3.1 Identification of concepts 38

3.2 Generation of concepts 39
3.3 Assignment of codes to conceptualizations 39
3.4 Assignment of codes to relations bewteen conceptualizations 51
3.5 Control of coder agreement 62
3.6 Computer input of text 64
3.7 Control of punch cards 66

4. *Construction of dictionaries and quantification of concepts* 70
4.1 Truncation of adjectives and verbs 70
4.2 Truncation of nouns 71
4.3 Scaling of properties, actions and states 73
4.4 Procedure for scaling adjectives and verbs 76
4.5 Processing and description of data 78
4.6 Analysis of data 79

5. *Design of search logics* 90
5.1 Statement of search questions 90
5.2 Formulation of hypotheses 91

6. *Data processing* 94

7. *Data analysis and inference* 96

8. *Construction of theories and models: A recapitulation* 98

9. *References* 101

Preface

This book presents no final results, but explorations and experimentation. We have tried to develop a method and construct a system for a computer-based content analysis of interview texts.

We have called our approach an ANAlysis of CONcepts by DAta processing (hence the acronym, ANACONDA). After having tested and presented our first attempts at this analysis, it has become important for us to anchor this approach in its true context with regard to subject and method. ANACONDA has passed through several stages (and will hopefully pass through more) and cannot yet be considered "ready". This presentation aims at showing on what points it can be compared to other similar experiments. Consequently we discuss a few models and theories from the fields of both linguistics and psychology.

The problem of gaining access to computer-based content analysis and a technique for using it has featured largely in a research project on search and steering strategies in educational and psychological research planning. This project has been financed by the Swedish Board of Education. The work within this project was initiated with an interview study involving forty randomly selected researchers working in departments of educational and psychological research in Sweden.

The research has been directed by the first author, while the second writer has borne the main responsibility for the linguistic part of the work.

We wish to thank Professor Åke Bjerstedt for valuable points of view, ideas and suggestions for concrete improvements during the different phases of the work.

We have also gratefully received valuable suggestions and advice from Docent Bernt Larsson and Lektor Hans Arte.

We would like to take this opportunity of specially thanking Fil. Kand. Leif Robertsson of the Computing Centre of Lund University for his contributions to the development of computer programmes and for carrying out the individual analysis programmes.

Our gratitude also goes to all the researchers who contributed to this study by kindly participating as interviewees and as members of an assessment panel.

Miss Berit Lundin and Mrs. Anna-Lisa Gustavsson has typed our manuscripts with great care and attention and Mrs. Ingegärd Johansson has done excellent work in punching the extensive test material.

Last but not least we would like to convey our thanks to Gillian Nilsson, B.A., for her discerning translation.

1. Development of a computer-based content analysis

The mediation of information by means of symbols is a typical human action. It takes place, for example, when we read newspapers or books. On the basis of the information mediated, conclusions are drawn about various events or other people, i.e. the interpretations are made the foundation of a conception or a frame of reference. When this process is based partly on extracted cues and partly on a person's indefinable intellectual ability, the procedure can be called "an impressionistic content analysis". This type of analysis is based on intuition, insight and impressions, which means that the interpretation is based on subjective analysis results. Content analyses based on frequency distributions differ from impressionistic analysis and interpretations of written or spoken text. This type of analysis is objective insofar as it requires an explicit analysis procedure and a formalized analysis. Objectification means that a person transfers certain typical human functions to objects, i.e. tools, and that machines are developed that can carry out functions that were originally subjective. In this respect the development of a Computer-based Content Analysis (CCA) is an attempt to objectify the method of content analysis.

In behavioural science research, many different content analysis techniques are used and have been used. Consequently behavioural scientists are well-acquainted with both the theoretical and the methodological, technical and practical problems involved in the use of classical content analysis techniques. A scientifically conducted content analysis implies that the researcher, regardless of a particular result, must be able to account for the reliability and validity of the method chosen.

A thorough, reliable and valid analysis of text is extremely time-consuming, however, and extensive text analyses require the development of mechanized or automated routines. Now that computers can be used for memorizing and logical selection, we hope to develop a CCA method with a greater degree of objectivity and flexibility than the classical content analysis techniques have had. We do not aim, however, at fulfilling the demand made by Waterman & Newell (1971, p. 287) that

"one should aim at full automatization and not at some optimal man-machine symbiotic system, even though the latter is a desired goal".

In the development of a CCA method research results from different scientific fields such as mathematical and computational linguistics, cognitive psychology and artificial intelligence, and the computer sciences have been used. The following central postulations form the basis for the development of a CCA method suited to varying purposes within the field of behavioural science:

1. an organization in the basic material, the structure of which can be revealed by means of a content analysis method
2. a theory that directs the researcher in his order-creating activities
3. algorithms that steer order-creating activities
4. a basic element that can be isolated and selected and an analysis unit that can be counted and measured
5. a set of logical operations by means of which problems can be formalized and hypotheses tested
6. statistical methods that are congruent with the theory on which the chosen analysis method is based.

A great deal of work is required to analyze the content of a complex material that also has a low degree of structurization. The work involves the demarcation of suitable analysis units, the development of a category system and coding of this information. The demands made on the degree of structurization of the basic material increase with increasing amounts of data. But the search for information and greater precision in the hypotheses that are formulated also add to the demands made on content analysis techniques, i.e. the retrieving capacity of the technique, which can only be compensated to a limited extent by the researcher's patience, hard work and ability to remember.

A CCA method is almost unavoidable in the cases where researchers wish to carry out sophisticated analyses and where they want to try out different theories and models on the same basic material without at the same time needing to reconstruct complex category systems manually and recode large amounts of data.

The development of the CCA method and the construction of the system described below have taken place in accordance with the steps shown in Figure 1. A brief account is given of the implications of each of them.

1.1 Directions for written and spoken text

The directions given for writing down the interview material have not included phonological transcription rules. Our aim has not been to make a study of different components of the spoken language. Moreover, specially schooled staff would have been required for the transcription. Nevertheless the importance of an authentic recording of the audio-tape material has been empha-

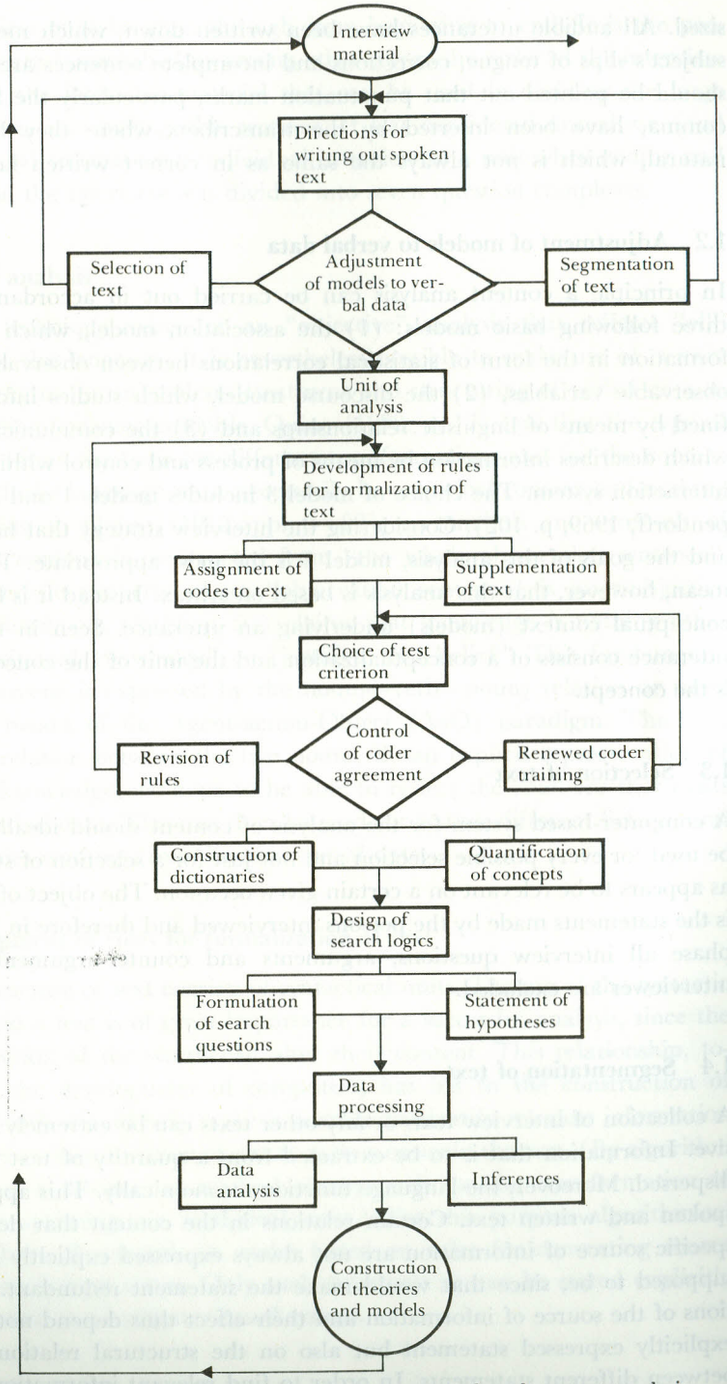


Figure 1. Flow-chart for designing a computer-based content analysis

sized. All audible utterances have been written down, which means that the subject's slips of tongue, corrections and incomplete sentences are included. It should be pointed out that punctuation marks, particularly the full-stop and comma, have been inserted by the transcribers where they have seemed natural, which is not always the same as in correct written language.

1.2 Adjustment of models to verbal data

In principle a content analysis can be carried out in accordance with the three following basic models: (1) the association model, which presents information in the form of statistical correlations between observable and non-observable variables, (2) the discourse model, which studies information defined by means of linguistic relationships and (3) the communication model, which describes information by means of process and control within a dynamic interaction system. The choice of model 3 includes models 1 and 2 (see Krippendorff, 1969, p. 102). Considering the interview strategy that has been used and the goals of this analysis, model 2 is the most appropriate. This does not mean, however, that this analysis is based on syntax. Instead it is based on the conceptual context (models) underlying an utterance. Seen in this way, an utterance consists of a conceptualization and the unit of the conceptualization is the concept.

1.3 Selection of text

A computer-based system for the analysis of content should ideally be able to be used for every possible selection and not just for a selection of such material as appears to be relevant on a certain given occasion. The object of the analysis is the statements made by the persons interviewed and therefore in the selection phase all interview questions, arguments and counter-arguments from the interviewer are excluded.

1.4 Segmentation of text

A collection of interview texts or any other texts can be extremely comprehensive. Information that is to be extracted from a quantity of text can be very dispersed. Moreover, the language functions economically. This applies both to spoken and written text. Certain relations in the content that derive from a specific source of information are not always expressed explicitly and are not supposed to be, since that would make the statement redundant. The intentions of the source of information and their effect thus depend not only on an explicitly expressed statement but also on the structural relations that exist between different statements. In order to find relevant information each individual interview must be gone through from beginning to end. It is the permanent structure (established by among other things the order of the interview questions) of the interview material that can be used in dividing the large amount of texts into manageable sections. It was considered unrealistic (and proved to be so) to treat each individual interview as a unit when coding and for this reason the interview was divided into seven question complexes.

1.5 Unit of analysis

Even if no technique exists for an "objective" analysis that reflects "all" dimensions of the language, it is nevertheless possible to make use of certain general paradigms for an analysis, treatment and structuring of verbal data so that information becomes available. Quine (1972, p. 17) says that the type of content that forms the basis for different transformations and for the content of the individual's language must necessarily be empirical content and nothing else. According to Quine (1972, pp. 9—22), an analysis and synthesis of empirical phenomena is constructed from "the whole observation sentence". The characteristic property of an "observation sentence" is intersubjective agreement. Rozeboom (1972, p. 97) claims that knowledge is nothing other than "propositional knowledge" or "justified true belief". This fundamental form of statement is expressed by the $noun_1$—verb—$noun_2$ relation or more formally by means of the Agent-action-Object (AaO) paradigm. The verb denotes the relation between the two nouns. By an explicit representation of this type of knowledge, we hope to be able to reflect the *evidence* that exists in a set of verbal data. This can then be used for testing different behavioural science theories working from the same set of data.

1.6 Development of rules for formalization of text

The basic structure of text consists of syntactical units. The use of the syntactic information in a text is of great importance for a successful analysis, since the syntactic position of the words can alter their content. This relationship, together with the development of computers, has led to the construction of algorithms which should help make it possible to identify relevant information as opposed to an identification of words as they occur in the text. (By algorithm is meant here a mechanical method of approach for the transformation of utterances to unambiguous analytical units.) For this purpose algorithmical codes have been developed, i.e. codes based on rules for converting source material to equivalent terms. Only such structures as can be stated explicitly can be delegated to a computer-based system.

1.7 Assignment of codes to text

An exact description of a text requires that a basic element can be isolated and selected. The basic elements must be alike (approximately identical), particularly when they are to form the basis for a measurement of equivalent properties. The basic element should be unequivocal, i.e. no element should contain more than one variable (assume no more than one value). It should be exchangeable and it should be more coherent internally (within the unit of analysis) than externally (between different units of analysis). The basic element should be more manifest than latent.

The maximal unit is a sentence, which can be divided into clauses of different degrees. A clause is complete as soon as it contains the two main constituents, subject and verb (phrase). These are coded. In addition to the individual parts of the clause, the statement's tense, mode etc. have been assigned codes. Furthermore, there are a number of codes for overall structures that the coding of separate units cannot give. This analysis works with the sequence of clauses. Each desired facet cannot be stated in advance, nor be extracted from a text material. For this reason, we have, in addition to the clause codes discussed, also devised codes for the main theme, so that the fundamental information, which cannot be retrieved or mediated by means of clause codes, does not get lost.

1.8 Supplementation of text

Some sentences can be fragments that cannot be supplemented into independent conceptualizations according to the AaO paradigm. In these cases in which the coder does not understand an utterance, it is to be deleted. The utterance must be completely comprehensible, which means that different types of relation words (e.g. pronouns and adverbs) must be supplemented to their right meaning in the context. Supplements are placed in parenthesis, so that the analysis does not lose track of what the person interviewed in fact says. When choosing the words to be used in the supplements, those already used by the person interviewed are taken first, if the context does not make this impossible.

During the recording, even text concerning practical or technical details related to the interview procedure is taken up on the tape. These parts of the text are deleted when the material is segmented.

When defining a sentence, it cannot always be assumed that each sentence in the text has been concluded with a full stop (see Chap. 1.2 above). A unit between two full stops can consist of several sentences, either separated by means of pauses that are marked in the transcription by a line of dots, or fragments that can be supplemented and made into complete sentences. An-

other way of marking the beginning and end is by linking with "and" or other conjunctions, which in this analysis are taken as being the first unit in a sentence. (This does not apply to a conjunction that links two objects in the same clause.) In the cases in which obvious corrections are made by the person interviewed, the utterance that is immediately corrected is not coded.

1.9 Choice of test criterion

Starting from a coded material in accordance with the discourse model, it becomes possible to represent events or ideas within the source of information (the researcher). The use of the model presupposes that independent coders can assign codes to the text with a satisfactory degree of agreement.

Two methods of assessment were applied. The first method (Osgood et al. 1956, p. 57) states the proportional agreement. The (1) separating of relevant from non-relevant text material, (2) segmenting of text into meaningful units and (3) identification of syntactical relationships were estimated according to this method. Osgood's technique was applied primarily for the purpose of making it possible to compare our results with those presented by Osgood. The second method is based on the binomial division hypothesis, i.e. the binomial test.

If we have estimated the intercoder agreement, irrespective of which method of assessment has been used, it is usually very difficult to judge whether the calculated index value can be considered satisfactory. It can be very difficult to determine a reasonable level of agreement, since there is no simple solution to this problem. Moreover, it is only possible to decide what can be considered a satisfactorily "reliable" coding within the frame of a given problem.

1.10 Control of coder agreement

The computer-based processing of text according to ANACONDA implies pattern recognition on the basis of manually inserted clause codes. The placement of a basic element in one and the same category by two independent coders can best be considered as parallel "tests". At the same time this assumes that both coders have at least equivalent frames of reference. An examination of the precision of the assignment done by the coders is one of the prerequisites if we are to be able to demonstrate the objectivity in content analytical processing of verbal material. The "reliability" of the assignments is above all a problem of communication, i.e. the precision of the coding is dependent on the communicability of the criteria stated in ANACONDA. To summarize, it can be said that the reliability of the coding is a function of (1) the unequivocality of the information units, (2) the unequivocality of manual and category functions and (3) the coders' special frame of reference, e.g. knowl-

edge of linguistics and knowledge of the subject. The coders are the measuring instrument in the analysis. In addition the unequivocality of the information contained in the basic elements influences the reliability considerably. But since it is very difficult, if not impossible, to get the entire process under control, the possibility of increasing the reliability is usually restricted to manipulation of the coders and/or manual. For this reason it is more justifiable to use the term "intercoder agreement", at least as long as the assignment of codes cannot be done mechanically.

1.11 Construction of dictionaries

Dictionaries for content analytical processing function as links between the natural language and a more formal, theory-oriented language. The analysis technique that has been developed for the interview material requires at least three different registers: (1) Independent concepts (subject and object terms), (2) Dependent concepts (attributes), (3) Action or copula (verbs). Using the computer, lists are produced of these parts of speech. Files are then compiled on the basis of these lists. By means of a KWIC programme, the dictionaries can be adapted very closely to the verbal behaviour of the interviewees.

1.12 Quantification of concepts

In constructing dictionaries 2 and 3, some of Osgood's semantic differentials were used. Each term is defined with regard to (1) evaluation, (2) activity and (3) potency. The assessment is made according to seven-point and bipolar scales with the respective pairs of adjectives (1) negative/positive, (2) passive/active and (3) weak/strong. The advantages of this scaling technique are that it is simple to use and that we can study three independent dimensions.

By means of the evaluation dimension, the extent to which the researcher assesses different aspects as good or bad can be studied. The activity dimension measures the extent to which the researcher considers that a particular aspect has influenced the development of project outlines or behaviour during the initial phase of the research process. The potency dimension measures the researcher's sensitivity or responsiveness. Dimensions two and three together express dynamics.

1.13 Design of search logics

Since the basic material displays a high degree of structurization, the analysis units can easily be re-defined and new information quickly extracted. It must be possible to predict the statements or the latent structure that exist in a text.

A CCA method assumes predictable statements and structural relationships between the statements. There are two types of relation, namely relations within concepts and relations between concepts. While the latter must always contain an object element, the former need not. Boolean algebra is used in linking them.

1.14 Formulation of search questions

Before a CCA method can be realized, the researcher must state his theoretical standpoint, i.e. define his concepts. It is necessary to establish in advance which aspects of the material are to be focussed on. The questions that have guided the planning of our investigation and the analysis of interview material are:

1. Which intentions or fundamental attitudes influence the selection of problems?
2. What ideas guide the researcher, i.e. which facts and values are important for research planning?
   2.1 Which results are anticipated, i.e. which hypotheses are stated and in which way are these to be tested (theoretically, empirically)?
3. What plans does the researcher develop, i.e. which methods are of importance for steering and controlling a systematic search for knowledge?
   3.1 What investigation designs are drawn up?
4. What strategies does the researcher develop, i.e. which skills and which aids are coordinated?
   4.1 What behavioural patterns does the researcher develop for the purpose of attaining his/her scientific goals?

1.15 Statement of hypotheses

Syntax implies sequence or a relation between the different parts of an utterance. There are fixed and mobile positions in this structure. If these positions are utilized in an analysis of text, hypotheses should be formulated for the purpose of proving whether or not the stipulated syntactical or psychological relations are meaningful. Thereby each concept category can also be related to each of the others by means of the conditions stated by the hypotheses. The purpose of our analysis is primarily to establish (1) which actions (with or without explicitly stated objects) are carried out by researchers and (2) which modifiers are used in the process of specifying a proposition.

1.16 Data processing

If we are to observe each individual researcher from the point of view of different manifest variables, this presupposes that we can define observable

fundamental elements. If such elements exist, a complicated phenomenon can be described or represented as regular compositions, i.e. a profile reflecting manifest values ("score profile").

But what we are more interested in is the dimensionality of a phenomenon, i.e. a profile that reflects latent values ("universe scores"). In an analysis of the relations between different concepts, it is always the relations that evade direct observations that attract the interest of the researcher. They are namely particularly important when two or more variables are to be interpreted simultaneously, since differences between the manifest values for a particular variable can be a result that reflects nothing other than inadequate observations (see Cronbach, Gleser, Nanda & Rajaratnam, 1972, p. 314).

Since dimensionality is a central concept in every form of scientific analysis, the questions that are to direct the continuing research work are formulated around such concepts as evade direct observations. A scientific analysis and description of a phenomenon can thus take place on two different levels, namely one manifest and one latent level. When the researcher demonstrates which aspects are to be mapped, he often constructs models and data matrices, in which the lines usually represent the measuring objects of the investigation, i.e. everything that can be measured and calculated, while the columns represent attributes or descriptors that refer to the measuring objects of the investigation. When using psychological tests, assessment scales or questionnaires with fixed alternative answers, one gets test values that can be used directly for setting up data matrices. Such values are not obtained immediately, however, when the basic material consists of verbal data. Thus, it will be necessary to discuss both theoretical and psychometrical problems in connection with the development of a CCA method by means of which interview data can be transformed into numerical values.

1.17 Data analysis

In manual content analyses only association models are normally used. They regard the information in a text as a result of simple frequency calculations that form the foundation for statistical correlations between manifest and latent variables. Naturally this type of processing only permits rough estimations of the latent structure of the text and the result of the analysis can hardly be considered an adequate base for valid interpretations referring to the entire association structure.

The obvious limitations that are a consequence of an interpretation of paired correlations have led to the use of some more flexible analysis models. In connection with the development of computer-based information and documentation systems, linear regression equations have been used to specify the relations between the input and output of the systems (see Salton, 1971,

p. 456). But again this methodological improvement does not make it possible to take into consideration the relations and interactions that exist within and between concepts. Only a coding of the structural relations of the linguistic elements and a multivariate analysis can possibly lead to an adequate representation of the complex structures that can be assumed to exist behind verbal utterances.

1.18 Inference

A verbal utterance is organized by the speaker in accordance with implicit models and a system of rules that applies to a particular language. Using the model on which ANACONDA is based, we should in principle at least be able to predict unequivocal concept and conceptual relations for a specific clause at a given point in time. Our aim is to be able to code conceptual information and address such information. For this reason complete syntactical analyses will be superfluous. ANACONDA presupposes access to syntactical information as a *pointer* to conceptual information. If we know that we need a certain type of conceptual information we should be able to seek this information by predicting in which syntactic form and in which place it probably exists. If we find unexpected information, however, its content will be analyzed. The result of such an analysis determines whether we need to change our rules for connecting syntactical codes or our addressing routines.

The cognitive structure of a specified individual can be defined by the perceived relations that exist between the properties that characterize an object. If these relations can be quantified by means of the values that represent co-variations of these properties, it will also be possible for us to determine the weights that each property should be given in a prediction of an object's attributes. If we want to make explicit which theories or models are guiding a researcher's approach to his work, we should study (1) which implicit models form the basis for the selection of information, (2) which structures the implicit models have, i.e. which attributes specify a particular model and which relations exist between the models and (3) which inferences researchers make on the basis of implicit models.

1.19 Construction of theories and models

Language is an expression of process (actions, events, conditions and relationships and associated persons, objects and abstractions). This process takes place within a structure: the clause. The process itself is represented by the verb. Participators in the process are e.g. persons and objects. They take the role of agent and goal. This role-playing in relation to the verb is called transitivity. This means that we cannot extract information from a text if we

only work with individual words. When people utter a thought, this takes place as economically as the situation permits. In a dialogue between e.g. researchers with a common frame of reference, the researchers find it easy to communicate, since the verbal representation they use produces the same or similar conceptualization in both parties involved (cf. Miller, 1967, p. 67).

By conceptualization is meant the individual's use of certain rules for relating concepts. Conceptualizations may be simple or complex. In this way an utterance in an interview situation can be rich in simultaneously underlying conceptualization and make it difficult to represent these in a sentence. Consequently a sentence in a text can contain many completely expressed ideas and idea relations. The condensed information that is a result of the inherent economy in clause-linking can thus only be obtained in a content analysis if supplementation is used. But to carry out our analysis, we need a starting point from which we can build up the structure in an utterance. In this analysis we begin with the action or the verb. An action can be said to be something that an "agent" can achieve in relation to an "object". Agent is used in the sense "action centre" and object consists of the means or the goal of an action. In principle only two cases exist, namely (1) agent and object coincide and (2) agent and object consist of two separate units. Different attributes which qualify and describe agent and object are arranged around these units while attributes that characterize actions will be grouped around the verb.

The problem in an empirical analysis of a text is choosing suitable or strategic parts. Moreover, this cannot take place independently of a relatively explicitly described model, i.e. a theory. The basic problem that must be solved in connection with the development of a computer-based content analysis is how the information that exists in a text is to be structured so that it can be recovered in many different ways. In classical content analyses researchers from different disciplines have developed almost as many techniques as there are users of the content analysis method. At the same time this means that in a stricter sense different content analysis results cannot be compared. Not until now are there signs that these techniques can become uniform through the development of algorithms, which make the recovery more objective, more flexible and more general. In connection with manual analyses words as "basic element" are sometimes used, but most often several words form a basic element. Since the interpretation of words and groups of words may produce very different results, manual content analyses in a strict scientific sense lack an objective and uniform theoretical foundation. Moreover, the chances of re-analyzing verbal material on the basis of re-defined basic elements is judged to be non-existent.

A computer-based processing of text assumes that algorithms can be designed and that computer programmes can be written that

1. accept the structure that characterizes natural language
2. systematically identify linguistic signs and strings that occur singly or together with other signs or strings in a text
3. sort out linguistic elements
4. reorganize linguistic elements in accordance with a certain given syntactical position
5. carry out logical selections
6. calculate frequencies and print out distributions, e.g. in the form of data matrices.

For a computer-based retrieval of information that is relevant to an investigation it is necessary that the statements have been stored in their original form. This means that complex concepts or compound concepts with a complex content are analyzed, i.e. divided into their linguistic elements and that the structure (the original form) is preserved.

The most important factor in a computer-based analysis, however, is that the use of computers assumes that algorithms can be constructed and theories formulated. In this way the researcher is forced to make explicit analytical methods of approach that previously were understood more or less intuitively.

2. A psycholinguistic model for the analysis of research processes

Human beings have a number of different symbolic behaviours. Verbal behaviours are those used primarily in structuring and organizing intra-personal and inter-personal experiences, even though there is no complete agreement between the symbolic representation and what is to be represented. The development of ideas and the formulation of problems are "behaviours" that are intimately associated with a person's specific ability to express himself verbally. For this reason the content of the language is our primary source concerning the researcher's problem-perception and problem-formulation during the initial phase of the research process. The analytical problem in the use of spoken or written text is, as with all forms of raw data, that we must infer specific events, behaviours or properties that are connected with the object being measured.

Psychological research, and in particular its psycholinguistic branch, has long been trying to map the psychological processes that underlie linguistic sentences. The specific human ability to collect data and transform them into information that is then transferred into symbols will be discussed from the point of view of the psycholinguistic process model given i Figure 2. It is based on assumptions of a general theory of systems. The central basic concepts of the model are choice of information, steering and control. It is an open system (see Bertalanffy, 1968), of which cybernetic models are a special case. Regarding human beings in the light of a theory of systems, we have good reason to assume that the essential results in the research will even in future be produced by individual persons. This view may appear far too reminiscent of the psychology of cognition, but research processes consist fundamentally of the collection of data and its transformation to information, which is then provided with symbols and models. Data have a physical existence in the sense that data can be classified, counted or measured, while the term information refers to the transformation of data.

The model in Figure 2 contains six different geometric forms. They have the following meaning:

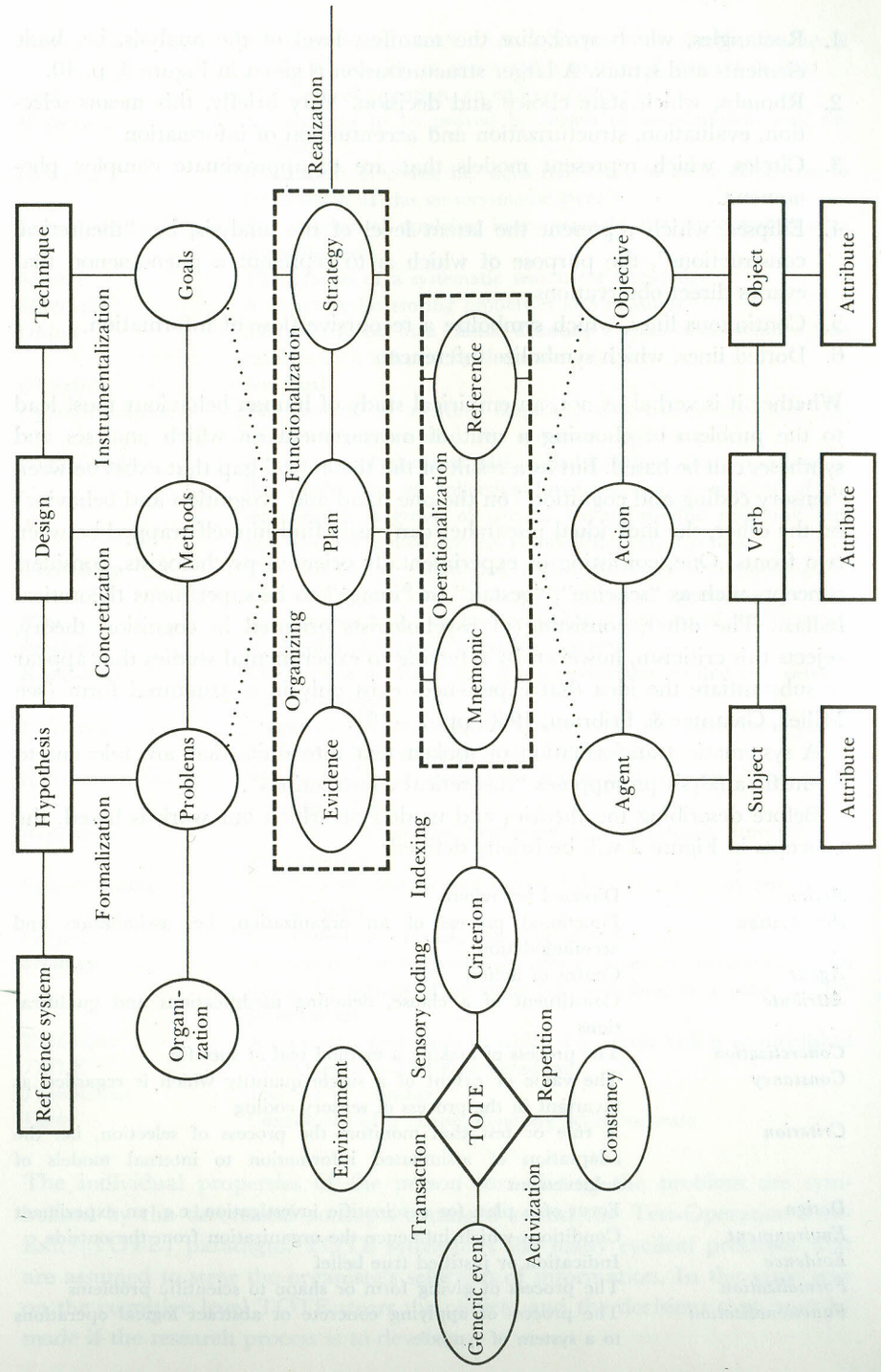


Figure 2. A psycholinguistic process model

1. Rectangles, which symbolize the manifest level of the analysis, i.e. basic elements and syntax. A larger structurization is given in Figure 3, p. 40.
2. Rhombs, which state choice and decision. Very briefly, this means selection, evaluation, structurization and accentuation of information.
3. Circles, which represent models that are to approximate complex phenomena.
4. Ellipses, which represent the latent level of the analysis, i.e. "theoretical constructions", the purpose of which is to represent a phenomenon that evades direct observations.
5. Continuous lines, which symbolize a recoursive flow of information.
6. Dotted lines, which symbolize inference.

Whether it is verbal or not, an empirical study of human behaviour must lead to the problem of choosing a unit of measurement on which analyses and syntheses can be based. But as a result of the theoretical gap that exists between "sensory coding and cognition" on the one hand and "cognition and behavior" on the other, the individual researcher can easily find himself trapped between two fronts. One, consisting of experimentally oriented psychologists, considers concepts such as "scheme", "gestalt" or "image" to be superfluous theoretical ballast. The other, consisting of psychologists oriented in cognition theory, rejects this criticism, however, by referring to experimental studies that appear to substantiate the idea that experiences exist only in a structured form (see Miller, Galanter & Pribram, 1960, pp. 2—13).

A systematic transformation of spoken text into units that are relevant to scientific analysis presupposes "theoretical constructions".

Before describing the theories and models on which our work is based, the concepts in Figure 2 will be briefly defined:

Action	Directed behaviour
Activization	Functional process of an organization, i.e. assimilation and accommodation
Agent	Centre of action
Attribute	Constituent of a clause, denoting modifications and qualifications
Concretization	The process of making a method real or specific
Constancy	The value or extent of a single quantity which is regarded as invariant in the process of sensory coding
Criterion	A rule or test that monitors the process of selection, i.e. the adaptation of assimilated information to internal models of achievement
Design	Form of a plan for a scientific investigation, e.g. an experiment
Environment	Conditions which influence the organization from the outside
Evidence	Indication or justified true belief
Formalization	The process of giving form or shape to scientific problems
Functionalization	The process of applying concrete or abstract logical operations to a system of symbols

Generic event	An occurrence, incident or experience of significance at a particular position in space and a particular moment in time
Goal	The focus of attention, an end or objective
Hypothesis	A statement to be proved by means of an empirical test, i.e. evidence
Indexing	Assignment of codes, i.e. signs that serve to guide references to information at the sensory-motor level
Instrumentalization	The process of applying instruments in the performance of a scientific task
Method	The process of a systematic search for knowledge
Mnemonic	A structure for storing properties and relations
Object	The constituent of a clause, denoting a noun or substantive that receives or is affected by the action of a verb
Objective	See Goal
Operationalization	The process of relating empirical meaning to structure, i.e. mnemonic are included in a system of experiences
Organization	A structure of elements with varied functions that contribute to the whole and to collective functions, e.g. a number of researchers or a group of persons having specific responsibilities and who are united for the accomplishment of a task
Organizing	The process of eliminating and successively accumulating information through the application of logical operations, i.e. rules
Plan	A sequential order or hierarchical arrangement of TOTE units
Problem	A complex of ideas or cognitive elements
Realization	The process of materializing a phenomenon, so that its existence may be verified
Reference	Properties of an object of observation that have been linked to the structure of that object
Reference system	The structure of natural or artificial facts and values in a specified context
Repetition	The act or process of producing an event again, especially for memorization
Sensory coding	Selection of internal and external data by means of representative sampling
Subject	Constituent of a clause, denoting an action centre
Strategy	A sequential or hierarchical order or a set of instructions which steers the actions of an organism according to a plan, i.e. the performance
Technique	A systematic procedure by which a scientific task is accomplished
TOTE	Test-Operation-Test-Exit paradigm
Transaction	The act of transfer
Verb	Constituent of a clause, denoting action or state

The individual properties of the person formulating the problem are symbolized by the theoretical concepts arranged round the Test-Operation-Test-Exit (TOTE) paradigm. TOTE symbolizes the many cyclical processes that are assumed to steer the organism's selection of information. In the same way on the cognitive level TOTE steers the choices and the decisions that must be made if the research process is to develop.

Problem formulation and problem solving are fundamental human operations. But since science cannot be regarded as a determinable object or a determinable set of problems, but must be considered as a way of attacking problems (methods and goals), everything can be changed into an object for scientific examination. Thus in the model it is the method that defines the content in the Problem-Method-Goal paradigm. On the latent level it is (1) generally available knowledge (Evidence or "justified true belief") and (2) anticipated problem solutions (plans and strategies) that form indispensable elements in the research process. Moreover scientific problems, whether they be original problems or routine problems, do not arise from a vacuum, but are based on existing knowledge, they are constructed on the basis of empirical generalizations and have their origin in theories and methods. Thus the choice of problems is decided by means of existing knowledge or the gaps in our knowledge, by our goals and by our methodological facilities. The purpose of all scientific activity is to show either that a scientific problem can be wholly or partly solved or that a problem cannot be solved by the methods that are at that time available. Research organizations (institutes and laboratories) form the outer framework within which research is carried out or is to be carried out. (For a more detailed discussion, see B. Bierschenk, 1974).

Every content analysis assumes that the researcher can define his objects of measurement, i.e. that which is to be measured and calculated. Starting from the Agent-Action-Objective paradigm, the unit of the analysis is defined in the model. In the same way as the choice of method decides the extent to which a problem is considered scientific or not, "action" defines the import of "agent" and "objective". The AaO paradigm demarcates the components that form a natural context, i.e. *"the whole observation sentence"*. While the agent and object (noun) are specified by means of the attributes linked to them, the verb states the relation between the nouns, i.e. actions, events or state. The order between these basic elements is stated by means of syntax. By using a dictionary and system of rules (directions for logical operations), we hope to be able to construct theories and models that can be used to describe and predict the initial phase in a research process.

Any attempt to explain complex human behaviours without a theoretical foundation is doomed to fail. Scientific analysis implies namely an attempt to arrange empirical facts in agreement with a theory or model. Despite the fact that there is an abundance of psychological theories, they all seem to be based on only three different basic paradigms, namely (1) the reflex arc paradigm, (2) the genetic paradigm and (3) the cybernetic paradigm. All three refer to a biological basic element. The first assumes associations as the basic elements in a theory of behaviour, in the second the components are an a priori determined "structure" or gestalt and in the third they are feed-

back and control of information. While the association theoreticians, especially within the S-R tradition, emphasize the importance of learning, this is of lesser importance within the framework of the gestalt theories. The process theoreticians assume that experience (environment) and a set of rules for logical operations are necessary to explain the behaviour of the individual. In this case the assimilation and accommodation assumptions play a large part. It is this last basic paradigm that will guide our work. The most influential sources have been Miller, Galanter & Pribram's information psychological model, TOTE, which is presented in the book "Plans and the structure of behavior" and which was first published in 1960. The implications of many of the assumptions made in this book had been explained as early as 1952 on a deeper psychological level in Piaget's (1963) work, "The origins of intelligence in children". The importance of the "cybernetic hypothesis" for an analysis of human behaviour has been further emphasized in Monod's work, "Chance and necessity" which was published in Swedish in 1972 and by Watson's description of the DNA structure in "The double helix" (1968).

It is primarily the coding concept that has in more recent years been used with a diversified (information theoretical, neurophysiological and psychological) import for the purpose of explaining how people become aware of themselves.

2.1 Sensory coding of information

It has been generally observed that people perceive selectively. It is now an accepted fact that our senses do not function as automatic transmitters of information but as a perception system or a selection mechanism. Monod (1972, p. 49) says that neurophysiology and the advances made in experimental psychology are starting to show that

> "the central nervous system cannot and certainly should not pass on to the consciousness any information that is not codified, reshaped and set in predetermined norms: in brief, assimilated and not simply reproduced information".

Sensory coding and the remembering of information has been described as a "content addressing" or "self-addressing mechanism" (see Uttal, 1973, pp. 1—2). The psychological implication of these results is that the fundamental principle that steers all human behaviour is the selection of information and that these selection processes are interactive. It was among other things such observations that led Wiener (1948) to formulate the "cybernetic hypothesis". This is based on the assumption of a steering and control of information as a prerequisite for systems that build up themselves.

Pribram (1972, pp. 449—480) argues that knowledge should be seen as "codified information consensually validated". According to Pribram (1972,

p. 463), coding is the key to knowledge and he names the underlying brain mechanism "hologram". This is defined as a mechanism that takes small but adequate random samples of relevant elements in order to create organized wholes again.

Nowadays the nervous system's most elementary signal system is thought to be of the presence-absence type. The nervous system then uses an inhibition mechanism to create sequences or signal groups that are coded. Expressed in Pribram's terms, the hologram mechanism would lead to at least four different products, namely (1) "Images-of-Event", which correspond to "Environment" in Figure 2, (2) "Images-of-Action", which correspond to "Generic event", (3) "Images-of-Achievement", which correspond to "Criterion" and (4) "Monitor-Images" which correspond to "Constancy". From these a limited number of variables are extracted that are coded, but no longer in an elementary form as presence or absence but as indicators, which state the relations between them. This indexing results in what is called a Mnemonic in Figure 2, which means a basic structure for storing properties and relations. Thus a Mnemonic can be seen as a holder for attributes. Piaget (1963, p. 119) is of the opinion that this transformation takes place through processes of assimilation and adaptation. Accommodation or adaptation of information to internal schemes comes according to Piaget (1963, p. 175) from simple differentiations of internal models. He borrows Poincaré's idea of a constitutive or intrinsic logic (similar to the structure of mathematical groups) in the actions of the organism. The structural differentiation that is a consequence of a differentiation and generalization gradually transforms the assimilation into perception of objects. Generalized selection processes that are a result of long-term repetition (cf. Constancy on the sensory level in Figure 2), are considered to lead to directed activity and should be able to replace concepts such as intention and will (see Piaget, 1963, p. 135; Miller et al., 1960, p. 27; Monod, 1972, pp. 29—30). This implies that a selective code system exists, which is based on attention, i.e. the organism is equipped with a network that accepts information and then decides what is to reach our conscious attention.

According to Piaget (1963, p. 148), consciousness arises from "dis-adaptation" and develops from the periphery towards the centre. Experiments with "the distorted room" have shown what such "centrifugal" activities, which function as powerful modulators of mnenonics, can achieve. These experiments have shown that our senses function as filters. This fact has become known in a socio-psychological context as the "Honi phenomenon". In the experiment with "the distorted room" a perspectivist displacement is used of a room that is shaped like an apparent parallelogram, but with one wall shorter and from the position of the subject "further away" than the other one. The subject of the experiment is asked to look through a hole in the wall and observe how two people (a child and an adult) walk towards each other and

change places. The observer gets the impression that the child becomes larger and larger, while the adult becomes smaller. When they have reached the opposite corner, the child is very tall and the adult very short (see Wilson, 1974, pp. 256—257).

Thus where we store information should depend on the way in which we perceive an object at the time when this information is stored in a memory that is under conscious control. At the same time this would mean that errors in the coding of incoming information would on a later occasion be reflected as errors of memory.

The concept information is admittedly not equally self-evident in scientific contexts as mass or energy, but in recent years it has become increasingly important. While the TOTE paradigm is based on "feedback of information" in the form of a control of instructions, the reflex arc paradigm is based partly on relatively discrete operations, partly on a special form of "information feedback", namely reinforcement of a behaviour. This means that the TOTE paradigm can be used for the purpose of comparing and testing, while the reflex arc paradigm assumes some drive reduction. On the basis of this paradigm the association theoreticians postulate as a basic component a conditioned association between stimulus and response, i.e. the theory is constructed of associations. In explaining such complex phenomena as the acquisition of language the association theoreticians say that this takes place by means of the principles of association and in this it is assumed that language is a set of associations. This idea emerges most clearly in Skinner (1957).

In the field of psychology it is primarily Lewin who has introduced conceptions such as "intention" and "valency" to counteract the postulation of the S-R theory that a behaviour must always be reinforced if it is to be successfully established or maintained. But within the frame of the TOTE paradigm evaluation forms one kind of empirical experience (Miller et al., 1960, pp. 62—66), which helps to shape a person's reference or frame of reference.

The difficulty in keeping apart knowledge or facts and values seems to originate in the problem of preserving the distinction between both categories, even though every meaningful behaviour combines them (Monod, 1972, pp. 160—161). But since there is in every behaviour an intimate link between intentions, means and goal, it becomes necessary to say something of how this interrelation arises. Piaget (1963, pp. 148—149) thinks that the problem of differentiating between facts and value judgements arises through multiple and generalized combinations of "schemes" (internal representation of information). These relations lead to goal-means hierarchies, which are influenced by conscious, directed activities or "intentions". Thereby intentions act as an extension of the whole schedule complex and the relations that exist between subgroups. This process leads to a "distinction", i.e. to mnemonics that represent "reality" and mnemonics that represent the "ideal". In Piaget's opinion the

evidence structures, plans and strategies are built up and changed.

If the TOTE paradigm and the biological and psychological processes are generalized into TOTE units of a greater complexity, it is no longer a question of transmission, transformation and representation of information. Instead it is a question of the order in which the instructions are carried out and how evidence structures, plans and strategies are built up and changed.

The building block in a theory of information psychology is information processing. In explaining the acquisition of language the process theoreticians assume a cognitive mechanism, i.e. continuous processes of differentiation and integration, and rules for inference, i.e. abstraction of implicit models as a result of the individual's observations. Thus it is assumed that there are regularities in complex phenomena that can be observed or predicted. A basic quality that is typical of all self-organizing systems is that TOTE organizations form hierarchical structures or plans. By studying the structure in a plan and analyzing its functions, we can investigate problems that are characterized by "organized complexity". On the macro-level holograms and mnemonics lead to evidence, which form the basis of the researcher's plans, strategies and goals. TOTE units interact with each other and the structure in these TOTE units is determined by their organization on the micro-level.

2.2 Indexing information

Thus from the perspective of information psychology a human being is a system that processes information and human behaviour is considered to be the result of this processing. The new interest in the "cybernetic hypothesis" is reflected in the psychological experiments that have been conducted for the purpose of studying the use made by human beings of internal mediators (mnemonics) and mnemo-techniques for memorizing lists of verbal material. Research on *pattern recognition* has attracted a great deal of attention. Experimental studies (see Hunt, 1973, pp. 343—371) show e.g. that an individual can perceive the structural and operational properties of an object. The structural properties of a given object here form the person's mnemonic of the object, while the operational properties, i.e. the individual's reference, form the basis of the conception that is formed.

Gibson's (1972, p. 215) theory on visual perception assumes

"the existence of stable unbounded and permanent stimulus-information in the ambient optic array. And it supposes that the visual system can explore and detect this information".

This is a new theory in the sense that it is "based on information and not sensation" and that the theory assumes an active extraction of information

from the environment, together with an active construction of models of the environment. According to Gibson (1972, p. 217) the theory differentiates between "stimulation by light" and "information in light". The relation between optic stimulation and optic information appears to be the following. The stimulation of the photo-receptors by means of light is a prerequisite for visual perception. The activity in the visual system depends on the surrounding light. There is no vision in the dark. But another prerequisite (condition) for visual perception is an area of surrounding light. This must be structured and differentiated. If the surrounding light is homogeneous, on the other hand, no perception can take place, even though the sensation by means of light continues.

Gibson (1972, p. 223) presents the following result: The *contour* or basic feature in an area is "invariant" compared to most changes in the lighting; the *structure* or the nature (area) of the object is "reliable invariant" compared to the changes of the observation point; the *qualities* of the contour (closed, open) are always "invariant"; the *shape* of a closed contour (of an area) is independent of light but "highly variant" compared to changes in the observation point.

This theory is well suited for an explanation of perception processes from the point of view of system theory. It supposes an abstraction of implicit models and it is built directly on the sensory coding process described, i.e. the hologram assumption. Gibson (1972, p. 227) writes namely:

"The eye is a biological device for sampling the information available in an ambient optic array."

Using Gibson's theory one no longer asks how the individual can "know", but asks instead in what sense an object is real and this can be indicated by suitable measuring instruments. Thus the theory makes a radical break with traditional perception theories, which assume that there is always an objective contribution in the form of sensations and a subjective one in the form of intrinsic (original) ideas or gestalts. In other words it is no longer supposed that there is any biologically anchored behavioural system or a biological disposition for the discovery of objects, e.g. language, as in Chomsky (1957). The fundamental factor in Chomsky's model is the assumption that there is a predetermined "grammar" and that the individual simply has to be able to discover it.

2.3 Operationalization of information

Mnemonics form codes to which references are linked. According to Gibson's theory (1972) invariances are extracted and this entire active process of ex-

ploration that the theory presupposes ought to be able to explain the phenomenon called reality. By means of the operationalization process mnemonics are assigned meanings, i.e. arranged in a system of experiences. To operationalize mnemonics, the individual must be able to formulate hypotheses or learn rules (see Rozeboom, 1972, p. 66), through which mnemonics are given reference status. More specifically a reference can be said to be the result of processes that *represent, relate* and *refer*. This is clearly different from what is called Evidence in Figure 2. Evidence refers to the acquisition of knowledge, if knowledge is defined as "justified true belief". (Boulding's, 1956, "image".) In this sense, therefore, a reference system includes more than a knowledge structure does. While the former can also include false beliefs, the latter is limited to true beliefs.

The consequences of disturbing the interpretation mechanism have been described very graphically by Luria (1969, pp. 33—58). The *mnemonist* could leave and recall "images" to an unlimited extent. But although this individual was

"exceptionally skilled at breaking down material into meaningful images, which he could carefully select, he proved to be quite inept at logical organization".

This lack of the ability to shape logical relations, i.e. interpret "image", means that abstractions and an intellectual behaviour are impossible.

This thinking in references appears to be typical of children. In connection with a presentation of empirical methods for a study of semantics, Miller (1967, pp. 51—73) describes a cluster analysis of children's (aged 8.5, 12.0 and 16.0 years) judgement of words belonging to different syntactical classes, which shows that when children are to judge the similarity between words, they assign them to a particular category depending on whether they are used together with the same word or not, e.g. the verb "eat" with the noun "apple". This is quite contrary to the groupings based on parts of speech that are so essential to adults. Miller (1967, pp. 59—60) writes:

"The thematic combination of words from different parts of speech, which is generally called a 'syntagmatic' response, can be seen to decline progressively with age and the putting together of words in the same syntactic category generally called a 'paradigmatic' response, increases during the same period."

Thus studies in child psychology show that a child begins to assign roles to the persons and objects that exist in the child's environment. In this the relations are identified by means of the role of those participating in this interaction. At this level of development, therefore, the structuring must take place by means of referents or so-called key words, although there does not yet appear to be any system of rules or logical organization in the way these are

used. Instead the references are linked by means of connected roles.

We needed a model that was suited not only for an adequate symbolic representation of information, i.e. concepts (mnemonics and reference) but also for an adequate description of the relations between concepts. The requirements we laid down led among other things to the development of ANACONDA, which was strongly influenced by the hypotheses of Schank (1972, pp. 552—631) and Abelson (1973, pp. 287—339).

2.4 Organization of information

As the discussion has shown, the processing of information in different phases leads to elimination and successive cumulation and to an empirical and logical operationalization. A human being's intellectual ability to organize and reorganize symbols leads to plans. While the classical scientific method was developed to study one-way causality, i.e. cause and effect between two or a few variables, the new scientific order today concerns "the world as organization". From a psychological point of view, the new basic view means that research no longer concerns a study of a "stimulus" as an independent variable and a "response" as a dependent variable, i.e. a study of "unorganized complexity" or statistical phenomena as the result of random events. Instead the interest is focussed on the development of methods for a study of "organized complexity" (see Bertalanffy, 1968, p. 234). Bertalanffy writes (p. 40) that

"we must look for principles and laws concerning 'organization', 'wholeness,' 'order of parts and processes', 'multivariate interaction' . . . to be elaborated by a 'general system theory' ".

A system is defined as "complex of interacting elements" and within the frame of a system theoretical model a "dynamic interaction between many variables" is assumed (cf. Bertalanffy, 1968, p. 30). To be able to develop a plan, it is necessary that "logical operations" can be utilized. Thus plans lay the foundation for a sequential or hierarchical arrangement of actions. While single actions are defined by means of a space and time coordinate, a plan is defined as a time-continuum along which different goal-directed activities are related to each other. When there are clearly defined criteria for a desired result, this is used to create conditions for a goal-directed behaviour. Bertalanffy (1968, p. 50) writes:

"Even under constant external conditions and in the absence of external stimuli, the organism is not a passive but a basically active system. This applies in particular to the function of nervous system and to behavior. It appears that internal activity rather than reaction to stimuli is fundamental."

The central importance of the directed activity can be seen, among other things, by the function of the verb in determining the nature of the AaO paradigm. The same basic view emerges from Abelson's (1973, p. 282) dicussion of the importance of the verb for the designing of plans. The basic unit in his model is "generic event". In his system this is represented by a verb category that is squeezed in between two nouns. Schank's (1973, pp. 187—247) hypothesis is that experiences are represented by relations between nouns. This means that a relation must encompass a process expressing goal, actions (event) and result. Finally Miller et al. (1960, p. 56) consider that a human being's verbal ability is in all probability very intimately related to his planned activity and since a person's plans are often of a verbal nature, they can be communicated. But despite the important function of the verb in a clause, i.e. on the manifest level in our model, the verb is not represented on the latent level. The same assumption appears to form the basis of the design of semantic networks (see Simmons, 1973, p. 71). A semantic network can be said to consist of coded properties and relations. The network consists of words that are part of natural language and of phrases that form "nodes". These in their turn are linked to other phrases by means of special groups of nodes, which are called semantic relations (see Simmons, 1973, p. 63).

Wearing (1972, pp. 77—86) conducted an experiment in order to study the way in which a sentence is processed and stored in the memory (in contrast to perceptual segmentation). The experiment shows significant differences between different parts of a sentence when it comes to remembering complex sentences. They were remembered most effectively with "objects" as "cues". "Subject", "adverb" and "verb" followed in that order. In the discussion, Wearing suggests several explanations for this differential influence. The elimination that arises in the memory works, in Wearing's opinion, directly on the term itself and not on the associative linkages of the term. The fact that the verb is the weakest code is explained by verbs having more common properties than nouns have. Moreover verbs have fewer unique properties compared to nouns, and consequently the meaning of one verb can easily be confused with the meaning of another. The possibility is discussed of the nouns being retained as distinct units, while the verb in a sentence is broken down into its component parts, which are then linked to the nouns. This explanation implies that there is a *semantic message* and a *transmission code*. This interpretation is supported by Piaget's (1968, p. 2) experiment. He writes concerning his results on the operational development of thought:

"... if we thus admit the existence of a progressive structuring of reality by means of operations gradually constructed one after another or on the basis of one another, then the most likely hypothesis is that the memory code itself depends on the subject's operations and that therefore this code is modified during development, and depends at any given moment on the subject's operational level."

Wearing's (1972, p. 84) hypothesis is that the subject and object in a clause are stored as distinct units in the memory. The other elements in the sentence (adjective, verb, etc.) lessen in importance, i.e. they are stored as abstract attributes to the subject/object. Thus the verb would be stored as an abstract relation between nouns. The consequence of this line of argument should be that the meaning of the verb is preserved in the same way as that of nouns, but that the exact structure of the verb is not preserved. Miller (1967, p. 59) says that the verb stands for "complex functions into which particular nouns can be substituted as arguments", but the classification of these functions is much more difficult than the classification of the arguments of the functions.

Reid (1974, p. 326) comes to the conclusion that verb, adverb and adjective on the latent level are only represented indirectly. He writes:

"... adjectives are syntagmatically related to nouns in surface structure and lexical memory, but in the image they are realized as features or qualities of one of the participants."

In the psychological model on which ANACONDA is based, the operationalizing and interpreting functions are summarized under the term "concept". It is assumed that every utterance is based on concepts that form the basis for key words in a clause. ANACONDA is based on only two types of concept and on only two role functions. The verb has admittedly important functions, both to pull together the key-words in a clause and to function selectively, but on the latent level properties, actions and states exist that are not independent of nouns. If we can in addition assume that the content of an utterance implies a choice, then a unit within the framework of the AaO paradigm only is non-redundant in the extent to which it is in contrast to other units, which could have appeared in a particular context (see Reid, 1974, p. 327).

2.5 Functionalization of information

Plans have been defined as a sequential or hierarchical arrangement of TOTE units, on which the behaviours of the organism are based. But to carry out these plans, a strategy must be developed, which means sequentially or hierarchically arranged instructions. In this way plans are functionalized, which means that properties and relations are placed in order. A functionalization of plans can lead to both concrete and abstract actions. They are usually not tied to separate discrete or fixed parts of a plan, but refer to more or less complex plans. Thus a functionalization means that an action can be carried out both concretely and in the mind. In the latter case the actions do not consist of a simple single-valued function but are complex, since they become reversible (see Piaget, 1970), and they have a complex relation to each other.

Abelson's (1968, pp. 112—139) hypothesis is that cognitive structures consist of "cognitive elements" and that ordered pairs of elements (nouns) in a clause should be linked to each other through perceived actions (relations). Further it was assumed that it should be possible to classify each relation as either positive, negative, ambivalent or empty. Quantification could take place along the dimension "value-centrality". This is regarded as the "strength" in e.g. a positive relation between "ego" and the element (object) concerned in the problem. Since Abelson places most emphasis on being able to state the direction of an action rather than a state which is the result of an action, and also wishes to study attitude structures, his system has more direct implications for our work than Schank's.

We have now presented a theoretical frame of reference that states the boundaries for the interpretation of the content of a clause. Something must still be said, however, about the *construction of a dictionary.* As far as we know, no objective method has yet been suggested by which we can extract content from a text directly. While linguists have mainly concentrated on an analysis of the structure and the elements in a clause, psychologists have studied "semantic distance" (see Miller, 1967, p. 51).

The purpose of developing a method for a computer-based content analysis is to build up a system that makes use of the advances that have been made within both scientific branches. As has been shown, this methodological development is based on the process paradigm and the assumptions underlying a theory of systems, which means that we are not interested in what words mean when isolated from their text. Instead we want to analyze how they function within the framework of a clause. By relinquishing both the association and the gestalt paradigm we wish to show that we consider a content of a clause to be neither a result of word associations nor the result of inborn grammar. Instead content should be regarded as interactions between words. We consider that a content of a clause is dependent on its context and on the experiences of the speaker and listener. Thus it is the communicative functions of the language that are important in the development of ANACONDA. In the development of a computer-based content analysis of text, therefore, we focus our attention on the design of the functional properties of the system.

A method for a computer-based content analysis differs fundamentally from a method for automatic text comprehension. The latter is theoretically anchored in linguistic "competence models" and puts the question: Can other (non-human) biological or non-biological systems acquire a natural language? The former, on the other hand, is theoretically anchored in a communication model ("linguistic performance") and puts the question: What must a system be able to do to prove that it has a language? Premack's (1969, 1971) research results on language show that these are two fundamentally different starting-points that can lead to quite divergent results.

For every empirical attempt to determine content in a text, we must select the units that are to be included in a dictionary. Thus a selection of strategic units is needed, carrying the linguistic information that makes meaningful logical operations possible.

3. An empirical analysis of concepts in context

In the behavioural sciences text analyses are usually based either on an examination of the text, looking for specific key concepts, or on the counting of how often a specific concept occurs in the text. In the latter case the method of analysis is based on the assumption that several independent coders can distinguish the analysis units describing a concept or a system of concepts. It is usually assumed that coders can remember all or the great majority of the categories. This content analysis technique, which is so familiar to behavioural scientists causes, however, a large number of theoretical and methodological problems (for a detailed discussion, see B. Bierschenk, 1972).

The application of a category system in the coding of text assumes that the latent structure of the text is reflected in the concepts and in the structure represented by the category system. A content analysis based on dichotomous decisions about or frequency distributions of concepts can, however, prove to be insensitive to the interviewee's own terminology and way of structuring text.

3.1 Identification of concepts

An analysis of language must take place on two levels. One is the manifest level which has been stated in Figure 2 by means of the AaO paradigm. The other is the latent level that has been indicated in Figure 2 by a dotted line. The manifest level forms the base while the other level states the concepts and concept relations that are assumed to lie below the speaker's (here the researcher's) construction of clauses and sentences. The theory for the representation of text that we have found best suited to our analysis is Schank's (1973) "Conceptual dependency theory". According to this theory, there are only three elements, "a nominal", "an action" and "a modifier". These are either independent or dependent.

Nominals, i.e. nouns, are independent concepts that do not need any addition to be understood. All others are dependent, i.e. they must be related to other concepts in order to have a complete meaning. By modifier is meant either adverb or adjective. The verb is regarded as "independent", but its meaning is specified by the noun(s) to which it is related. Schank calls the

nominal PP ("picture producer"), the verb ACT ("action"), the adjective PA ("picture aider") and the adverb AA ("action aider"). These abbreviations will be used from now on when appropriate in the comparison between this theory and our analysis. The properties of the concepts apply to the conceptual level and not to the sentential level. It is after all possible to have a sentence without a verb or an adjective without its noun and yet have an utterance that is quite comprehensible and still a sentence in a communicative sense. This does not apply, however, on the conceptual level. Since these are rules of dependency between concepts, an independency must also exist.

3.2 Generation of concepts

A method for content analysis that is suited to an approximation of the interviewee's conceptual structure (implicit models of the research process) cannot be satisfied with a traditional dictionary as a base. Such a method must be able to take into consideration context and syntactical order. The experiment carried out by Oller & Sales (1969, p. 209—232) shows that a given syntactical order limits the possible interpretations of the elements in the analysis.

Starting from the hypothesis that the interviewees in our investigation make use of syntax and a dictionary in order to formalize their thoughts and express their ideas about the initial phase of the research process, we intend to examine the interview material on the basis of the flow-chart presented in Figure 3. The assumptions on which this schedule has been designed have already been discussed in detail. To sum up, however, it can be said that we assume that the syntactical order between independent and dependent concepts is determined by conceptual rules.

The flow-chart in Figure 3 shows how we intend (starting from elements carrying linguistic information) to identify concepts in a given context. This presupposes a system of rules stating how different elements are to be linked to each other.

3.3 Assignment of codes to conceptualizations

A conceptualization expresses an event and thus requires a verb and at least two nouns. The way in which a clause is interpreted depends on the conceptual rules. Formally defined dependency relations exist between given categories of concepts. These dependencies form the structure on the conceptual level.

Schank (1973, pp. 194—195) has developed a so-called C-diagram ("conceptual dependency network") to express symbolically dependencies between concepts.

The purpose of the following sections is to compare certain parts of Schank's (1973) dependency theory with the ANACONDA system. Therefore the

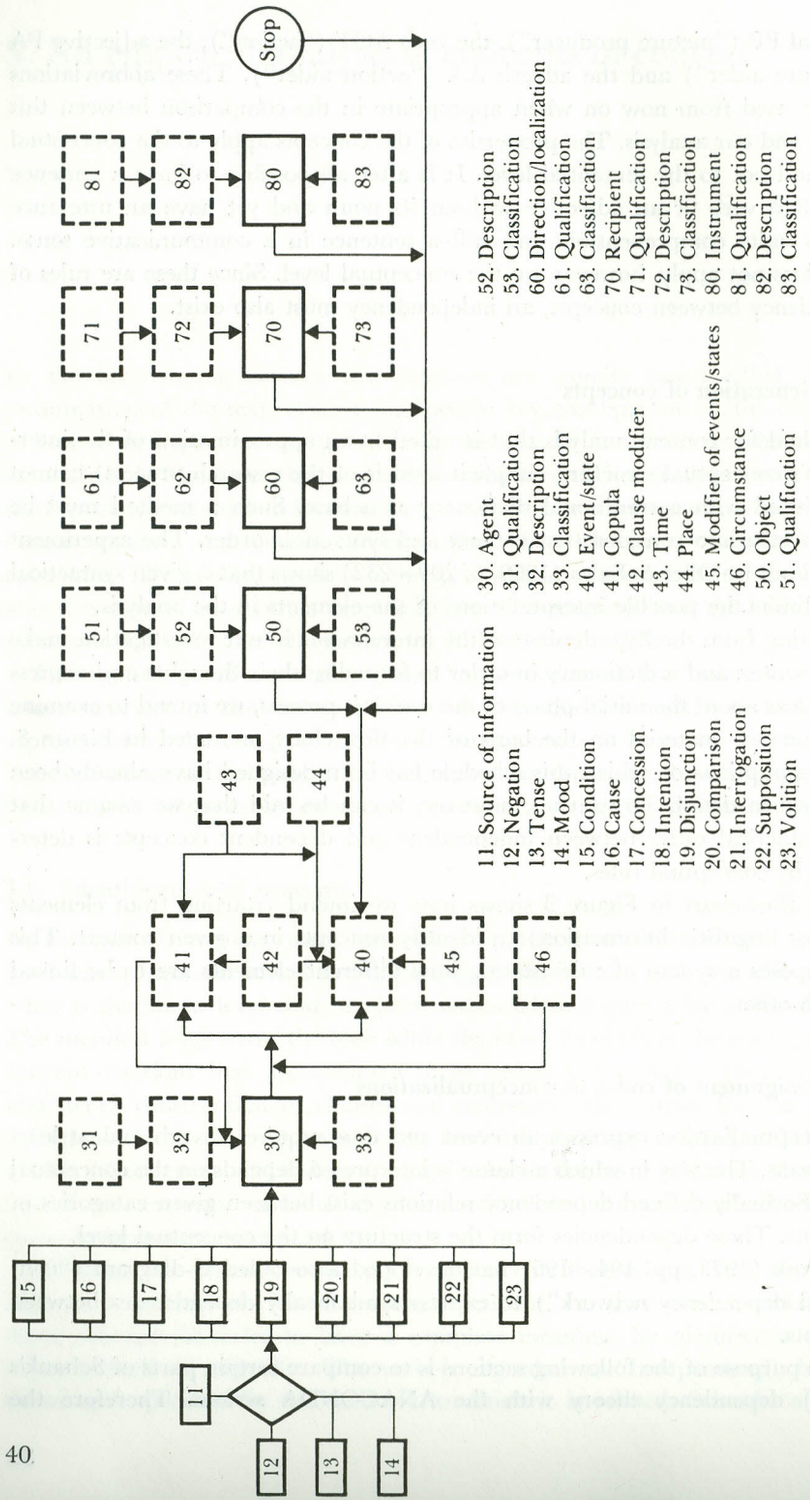


Figure 3. Flow-chart for an analysis of text and the formation of concepts

account follows Schank's presentation. For this reason we have not thought it necessary to burden the text too much with references and quotations. The way in which we have interpreted Schank for our purposes will be clear if the reader goes directly to this source.

For the systematic use of conceptual rules for feeding text into computers, it is necessary to find a way to represent concepts and relations between concepts by means of a code system. It is essential to point out that semantic rules or interpretations are not primarily to be expressed by coding at this level. An interpretation of the text is unavoidable, however; it would not be possible to segment the concepts otherwise. In our opinion syntax and semantics are each a prerequisite for the other and we have utilized syntax in order to be able to use this structure to limit the concepts within a clause or a sentence.

As was mentioned earlier, we want to build up our sentences mainly in accordance with the AaO paradigm. We shall attempt to show here how the relations between these labels can be coded symbolically. In addition we have tried to use the coding system to state the dependency structure within the concept complex. The way in which we link up with Schank's theory is shown in Box 1.

Box 1. Comparison between C-rules and ANACONDA: Concept coding

C-rule	ANACONDA	Symbols	Content of symbols
1. PP $\Longleftrightarrow$ ACT	30 + 40	$\Longleftrightarrow$	"mutual dependency"
2. PP $\Longleftrightarrow$ PA	30 + 41 + 32		
3. PP $\Longleftrightarrow$ PP	30 + 41 + 30	$\Longleftrightarrow$	"attributive conceptualization (set membership)"
4. PP $\uparrow$ PA	32 + 30	$\uparrow$	"Conceptual attributes predicated"
5. PP $\Uparrow$ PP	30 + 33 31 + 30	$\Uparrow$	"Concept that is attributively differentiated"
6. ACT $\overset{o}{\leftarrow}$ PP	40 + 50	$\overset{o}{\leftarrow}$	"Objective dependency"

For the meaning of the figures, see Figure 3, p. 40

Our code system is built up in such a way that each concept is specified by means of a two-figure code. AaO is expressed as 30 + 40 + 50. Code 30 denotes an agent function. (We do not state whether the agent is a person or an abstract concept. A later categorization within the respective codes takes care

of such statements.) A concept complex consists of main words and qualifiers of various kinds. The main concept has a code number ending with 0, while dependent concepts have a final number other than 0.

The relationship between PP and PA expresses dependency, in which PA is a dependent concept of PP. We express this kind of dependency within the PP structure by means of the combination of figures (30 + 32). Rule 2 states that the conceptualization that is formed requires a copula. We differentiate copula-verbs from other verbs and get the code 41. This also makes it easy for us to distinguish in the material all sentences expressing evaluations and classifications in the form of predicative statements.

Rule 3 functions in the same way, but here the complement is not an adjective (attribute) but a noun, as independent as the first.

Instead of a copula construction, the relations between PP and PA can be expressed as shown in Rule 4. The adjective stands in front of its substantive (e.g. the tall man). Schank's arrow under PP means that this concept complex does not form a complete syntactical sentence, since ACT is lacking, i.e. a concept that is placed horizontally in relation to PP. The same applies to the fifth rule, which shows the dependency between substantives in a concept complex. According to Schank, there are three kinds of dependency expressed by this rule, *containment, location* and *possession*. We call this pre or post qualification, which is not an adjective attribute. We would code *the man in New York*, or *the peas in the tin* as 30 + 33. *John's dog* is a state of possession expressed in a genitive form, which in our case is coded 31 + 30.

Rule 6 states dependency between the verb and its object. Schank's symbol says that there is a dependency, insofar as a verb can require an object in a complete conceptualization, but that the object is not otherwise a dependency concept. It is only on the horizontal line that it is in some cases necessary. This relationship refers to the question of transitive and intransitive verbs. We code the object with an independency code, since a PP in this position as a concept has the same structure as a PP as an agent. In one sentence there can namely be concepts that are referred to the object in the form of qualifications of various kinds (see Box 2). If we had assigned the object a code that made it belong to the verb (with 4 as the first number), we would have no symmetry between agent-complex and the object-complex. Instead we differentiate between the kinds of object. The object coded with 50 is the one related most closely to the verb, corresponding to the one that is traditionally called the direct object. (There is a second object and this is presented further on.)

Before introducing new comparisons, we would like to give an example of our coding of a basic sentence, *John hit his little dog*, which Schank presented initially, in order to show how his theory can be used in practice in input.

Box 2. Comparison between C-diagram and ANACONDA: A basic sentence

C-diagram		ANACONDA	
John ⟺ hit ←(o) dog		John	30
		hit(s)[1]	40
↗ ⇑	POSS-BY	his (John's)	51
little John		little	52
		dog	50

[1] Tense is not marked in the C-diagram

Box 2 shows the way in which the object is treated. It has the same basic structure as the agent from the point of view of their composition of independent and dependent concepts. Therefore we have the two-figure code system in order to emphasize and keep apart PP syntactically. Thereby we can treat the concepts separately in a flexible way. It was very easy, for example, to extract all adjective attributes prior to the scaling (see Chap. 4) that we have carried out, since they are specified by means of the second figure in the code.

The relationship between the two PP *John* and *dog* must be represented unambiguously. Since we do not work with automatic recognition of items, but prepare the mechanical computer processing manually, all pronouns are specified by a supplement in parenthesis. (See also p. 50.)

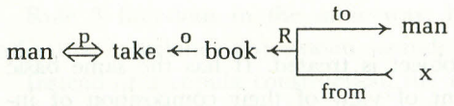
The information that is left out but that can be predicted by the dependency structure between concepts is in this context important for our analysis method. The pronoun's reference is the first thing we must take into consideration in order to be able to work with the concepts. Without reference to the concepts a large amount of material would be lost. However, the pronoun's reference is not a main issue in this context. The supplementation of a concept complex (e.g. within the agent structure) or the necessary parts in the syntactic paradigm is of major interest in a theory of concept formations.

The central theme in Schank's argument is the importance of the verb. ACT means an event or a process expressing movement or condition. The direction of a verb of motion is usually denoted by language researchers as transitivity. In these cases conceptualization means that one knows the goal of this direction and that one knows that there should be a goal. Schank describes his theory about TRANS by means of a number of sentences with verbs containing underlying but necessary cases.

Ex. *The man took a book* is analyzed conceptually:

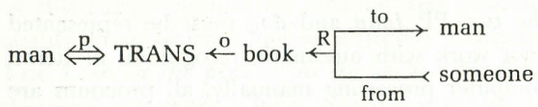
(*p* indicates past tense, *o* states that the verb requires an object.)

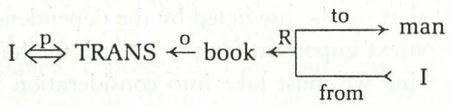
This sentence is not completely represented since the verb *take* in addition to an object must have concepts for "from whom or what" together with a recipient. The network looks like this:



(*R* stands for recipient. The same applies to the verb *give*, but the *x* is then known, e.g. *from I.*)

The conceptualizations underlying the sentences *The man took a book* and *I gave the man a book* are represented in this way:





Give is defined as TRANS when the agent and source ("originator") are identical, *take* is TRANS when agent and recipient are identical. Schank (1973, p. 198) explains in more detail:

"This conceptual rule states that certain ACT's require a two-part recipient in a dependency similar to that of objective dependency. The similarity lies in the fact that this type of dependency is demanded by certain members of the category ACT. If it is present at all, it is because it was required. /. . ./ . . . a conceptualization is not complete until all the conceptual cases required by the ACT have been explicated."

The four conceptual cases that Schank works with are OBJECTIVE, RECIPIENT, DIRECTIVE and INSTRUMENTAL. These cases are represented in the C-diagram as shown in Box 3. Rule 6 from Box 1 is repeated here. As a comparison, the way in which the ANACONDA system would code these case relations is also presented.

Box 3. Comparison between C-diagram and ANACONDA: Conceptual cases

C-diagram	ANACONDA
6. ACT $\xleftarrow{o}$ PP	40 + 50
7. ACT $\xleftarrow{R}$ PP, PP	30 + 40 + 50 + 70
8. ACT $\xleftarrow{I}$	40 + 80/45
9. ACT $\xleftarrow{D}$ PP, PP	30 + 40 + 60 30 + 40 + 50 + 60

For the import of the figures, see Figure 3, p. 40

Rule 6 has already been explained. Rule 7 states that an action that has a PP as recipient must also have a PP as initiator or agent, in addition to the PP that is connected with the verb, i.e. such an action requires two objects. Our second object has the code 70 and thus has the same structure as other independent concepts in the paradigm. In the coding rules we have called this indirect object (in accordance with the traditional way of analyzing clauses) or goal, corresponding to "recipient".

Rule 8 shows that the case that is called instrumental can be considered through the vertical arrow as being dependent on the action. In this interpretation of ours it corresponds most closely to an adverbial of manner (code 45), which modifies the verb. If on the other hand this concept consists of an independent concept (noun) it is included in our paradigm as a main code (see Fig. 3). For example:

Mary shouted *furiously* (code 45)
John killed his wife *with a big hammer* (code 80)

We think this differentiation is practical, above all since we must be able to separate the dependency concept in the complex (the with-phrase), i.e. the attribute *big*, which is assigned code 82. If a with-phrase contains an abstract noun, which can easily be transformed to an adverb without changing the meaning, we have considered regarding the concept as an adverb, as in

John looked at Mary *with anger/angrily.*

Rule 9 is explained by Schank as follows (Schank, 1973, p. 202):

"The DIRECTIVE case indicates that PP's may serve as direction indicators of a directional action. /. . ./ The directive case is extremely similar to the recipient in form and

is almost in complementary distribution with it. That is, the two never appear together and would seem to be different forms of the same phenomenon. The most common ACT that takes directive case is 'go'."

In our tests this case has proved to be difficult to differentiate with regard to "direction where" and "goal" (or "recipient"). This has been solved by using the code for "recipient", if the object of the action changes "owner". If the agent or object changes position, the code for "direction" is used:

I sent the report *to the institute* (code 70)
I went *to New York* (code 60)
I put the letter *into the box* (code 60).

We also have a code (44) that does not express transitivity and that is an indication of place, e.g.

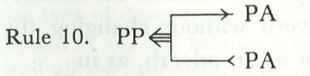
I live *in Stockholm* (code 44).
I saw him *in the street* (code 44).

It should be pointed out that code 44 does *not* express "location" as defined in Rule 5 (Box 1). Identification of "place" is further exemplified in connection with Box 6 and Figure 7.

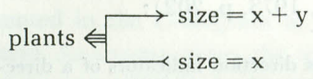
The difference between objective and instrumental can be difficult to clarify. Schank (1973, pp. 199—200) gives an example:

John grew the plants with fertilizer.

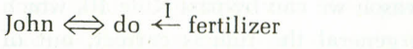
The concept fertilizer is the syntactical instrument of *grew*. But it is Schank's opinion that on the conceptual level the verb *grow* cannot be an action that a person can perform towards anything. It is the plants that become bigger as a result of what John did. Thus it is a question of *a change of state*, which must be expressed in a new rule:



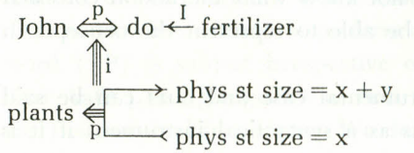
which, represented in the sentence above, becomes:



John's action is represented by means of a so-called dummy verb, *do*:



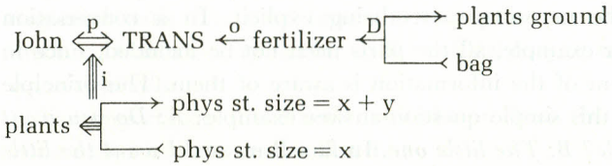
Hereby we have got two conceptualizations, which must be related to each other in some way, namely by means of a causal link, which is stated by ⇑ between the two clauses (*i* states that the causation was intentional):



Really *fertilizer* is not an instrument but an object, since what happened was probably the following:

"John took his fertilizer bag over to the plants and added the fertilizer to the ground where the plants were. This enabled the plants to grow."

On the conceptual level this is another kind of TRANS:



The conclusion is that what looked like being a syntactical instrument, i.e. an instrument on the syntactical level, is on the conceptual level an object. Schank says that this always happens with a syntactical instrument, since a single PP cannot be a conceptual instrument but only the object of an action. This is the explanation of Rule 8 above.

In this way Schank continues to investigate underlying structures in sentences. The representation of the sentence *John ate ice cream with a spoon* has an even larger network on the conceptual level, since the verb *eat* involves a series of actions that each have necessary instruments. Schank's idea is that each action requires an instrumental case, but that it is not necessary to state these with verbs like eat, where every listener knows which instruments are required and which are possible. One does not think about them actively. In coding natural language, in the way a listener does when he understands what

a speaker says, all the series of actions are implicit in such a verb. The goal for us is to state which concepts in the clauses indicate action. Underlying actions and instruments are irrelevant. For this reason we can by-pass Rule 10, which says that a PP can change its state. In general the rule is correct, but of subordinate interest in this context. A sentence such as *He grew plants* is coded 30 + 40 + 50. We can assume that the plants subsequently became larger and that implication exists in the verb without our needing to state in a relative clause *so that the plants grew larger* or suchlike. We would treat the sentence *He pleased me* in the same way. We do not know what the action consisted of nor do we need to know in order to be able to represent the concepts in codes.

The difference between Schank's instrumental case and ours can be said to be that the instrument is coded by us as a syntactical instrument if it is explicit. This means that it is not considered to be a necessary part of the AaO paradigm.

Hereby we have come to the question of how the conceptually necessary parts can be coded most suitably. Therefore an account will be given here of how we consider the conceptual level can be made explicit for our purposes. The natural language as a means of information is characterized by an economy which means among other things that references are expressed by pronouns. Box 2 showed how we code a personal pronoun with the referent in brackets. There are also other ways of expressing a sentence with a complete content, without all the necessary parts being explicit. In a conversation between two people, for example, all the parts need not be included, since in that context the recipient of the information is aware of them. The principle can be illustrated with this simple question-answer example: *A: Do you want the big or the little apple? B: The little one.* In fact B means: *I want the little apple.* Transferred to the dependency theory discussed earlier, this means that in coding B's answer we first code the answer as a dependent concept (PA) belonging to the independent (PP) *the apple.* In addition the answer contains the syntactical function, the object (°), and that does not suffice as conceptualization. The verb is *want* (we must work from what has been said earlier and not what B might have said instead). But want + object (ACT $\overset{\circ}{\leftarrow}$ PP) are not enough either; the agent (or subject) *I* is missing.

That which we can call vertical dependency here applies to the supplementation of the concept complex, so that the main words can be made explicit in the coding. If the main word is the concept that is said, we can consider the concept complex as being independent, but in the cases where attributes or modifiers stand alone in the clause, we supplement them with the concepts that are independent and that are to be found in the preceding context. There can also be cases where the concepts that are main words do not sufficiently explain the whole context, e.g. *lärarhögskolan* (the school of

education). Here we can work from the definite article *n* (the). The concept must then be further specified, e.g. *i Malmö* (in Malmö), which then becomes a postpositive qualifier with the figure 3 as the second figure in the code. We have worked out rules for how our supplementations are to be made (I. Bierschenk, 1974) but they are not yet complete.

The dependency that is expressed through the relation of independent concepts to each other in a syntactical sense could be called horizontal dependency. Like Schank, we work from the verb as being the central concept. The "direction" of the verb decides which codes the main words involved are assigned. This means that we cannot perform a syntactical coding, where the first main word (PP) is subject irrespective of its relation to the other main words. Our AaO paradigm guides us and the PP that is agent need not always be the first nominal in a clause. We describe the governing concept as the agent ("action centre"), regardless of whether it is abstract or concrete. At the same time this means that we cannot state in advance whether the verb is an action or a state. On the other hand we differentiate between copula and other verbs, thereby making a distinction between substantives that carry out an action or are in a state and those that are the objects of evaluation or classification on the part of the speaker. The determination of the verb's degree of activity takes place through the scaling (see Chap. 4) and only after that can we hope to be able to form categories of verbs. Guided by these results, it will also be possible to group the substantives in accordance with a content theory on an empirical basis. Thus we limit the content of our codes to the conceptual function, which is very similar to Schank's theory of cases and is based on the syntactical role.

The importance of the verb as the key to how the rest of the concepts are to be coded is illustrated in Box 4.

Box 4. Coding of the direction in the passive voice

Text	Translation	Codes
Det här (X)	This (X)	50
har	has	40
aldrig	never	- -
undersökts	been investigated	40

The code - - denotes negation

Har undersökt (has investigated) is the active voice. There is no agent to the verb and the agent is here so undefined (poss. *by researchers*) that it has not

been supplemented in. On the other hand, it is plain that the object is *X* and that this cannot be the agent. The passive voice of the verb makes that impossible.

Agentless clauses with passive verbs are relatively common, when the speaker does not consider the agent to have any essential part in this connection. No agent then exists in the context. Our paradigm is not complete here, but the sentence is coded all the same. We have considered it important to put in as much text as possible and we can then work on the different kinds of clauses later. The information we want is decided by the linkages of the concepts available. Thus the verb decides whether the paradigm can be coded in its entirety or not. In this case the form is the evidence. In other cases the coding can depend on the meaning of the verb. Box 5 shows two types of verb, one of which requires a supplementation of the type we have called goal, i.e. corresponding to Schank's recipient case, while the other is a type of verb concept, where the main element is really a noun.

Box 5. Coding the complete paradigm

Text	Translation	Codes
Vi (XYZ-projektet)	We (the XYZ-project)	30
skickade in	sent in	40
en projektansökan	a project application	50
(till Riksbanksfonden)	(to the Bank of Sweden Tercentenary Fund)	70
Jag	I	30
var handledare	was tutor	40
för lärarkandidater	for student teachers	50
på ämneslärarlinjen	in the subject teacher course	53

The verb *skicka* (send) contains implicitly a *goal* or recipient. When this concept is known in the text we supplement it so that a complete conceptualization is formed. The second example given here is an illustration of the fact that we do not make a complete syntactical analysis. Such an analysis would namely not have considered *handledare* (tutor) as a verb and the link between *jag* (I) and *lärarkandidater* (student teachers) would not have emerged. We see *var handledare* (was tutor) as a verb corresponding to *handleda* (supervise) to which there must be an object. On the other hand there is here no goal for the action supervise, which does not express direction. (The way in which we define verbs can be seen in I. Bierschenk, 1974, p. 57.)

3.4 Assignment of codes to relations between conceptualizations

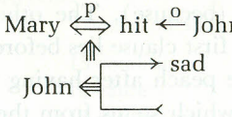
The relationship between dependent and independent concepts has been discussed, together with the relationship between such concept complexes in a complete conceptualization. A complete conceptualization can in its turn however, be related to another and this relationship corresponds to a complex sentence. Schank calls the connection between concepts dependency, while the connection between conceptualizations is called a relationship.

The most important conceptual relation is, in Schank's (1973, p. 202) opinion, that expressing causality, which is stated symbolically:

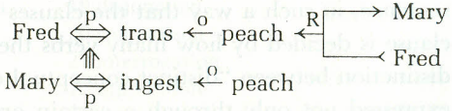
Rule 11a: ⇔ ⇑ ⇔ 11b: ⇔ ⇑ →

Causal relation is expressed by the arrow between the two clauses. Schank (1973, p. 203) gives a few examples of how causality is expressed in English:

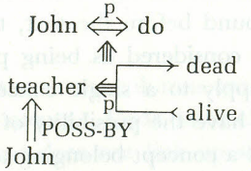
John was sad because Mary hit him.



When Fred gave Mary a peach she ate it.



John killed his teacher.



(INGEST is a category of ACT as well as TRANS.)

The verb *kill* also implies a change of state that leads to a result in accordance with Rule 10. The action in *kill* could be *by shooting*, which is then realized as e.g. *propel bullets via a gun to the teacher's head.* Thus *kill* is a class of transitive verb that Schank calls "pseudo-state verbs". They have the property that the object of the verb is the actor in the dependent conceptualization (the teacher dies). Often the verb is ACT in the dependent clause.

Sam flew his plane to San Francisco would be interpreted *Sam acted in such a way that his plane flew him to S.F.* Thus it is a kind of disguised causal, which is also discussed in the example *John comforted Mary,* where John causes a state in Mary (Rule 10). As has been seen Rule 10 is less important for us and the third example above is therefore not taken up in the following comparison.

Rule 11 will be compared below with ANACONDA's coding rules.

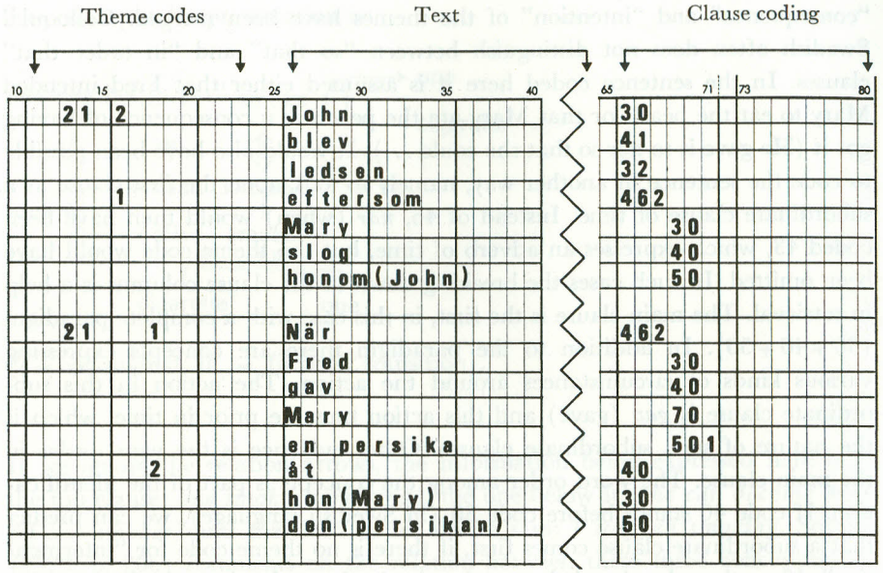
Starting from the above example of causality, the question must be put: What is meant here by causality? The splitting of the rule into 11a and 11b makes it possible to explain this as a more general phenomenon than that the relation should apply to such clauses as would traditionally be analyzed as causal clauses. The causality expresses a relation between clauses, in which two (or more) actions are separated in time. Schank (1973, p. 205) says:

"In other words, we want to be sure to distinguish distinct conceptual events in the real world."

The first example would be analyzed as a sentence with a causal clause introduced by *eftersom* (because). The other sentence expresses more of a prerequisite, where the first clause lies before the second one in time. The relation is that Mary ate the peach after having got it. The third example expresses a "disguised causal", which stems from the meaning of the verb.

In the ANACONDA system every clause is coded separately within its sentence, in such a way that the clauses take up two columns. Thus what is a clause is decided by how many verbs the sentence has, which is the same as a distinction between "distinct conceptual events". Relations between clauses are expressed not only through a certain order or a certain number, but also by means of the theme according to which the whole sentence is coded. Theme codes are to be found before the text, together with the identification codes and are therefore considered as being predominant to the concept codes. A theme does not apply to a single concept but is coded consistently to each concept. We then have the possibility of using a code at the beginning of the punch card to sort a concept belonging to a specific theme. The way in which Schank's first two examples of causal relation can be coded is presented in Figure 4.

The codes that can occur in the theme columns (15—23) are either 1 or 2. One means that the clause is prior in time in relation to the second clause. The syntactical order is retained. The clause coding shows how the clauses go in curves. (For the practical consequences of this system, see the detailed account given in connection with Fig. 5.) One figure following the concept code shows in which column the sentence continues. Both *eftersom* (because) and *när* (when) state a new clause and are called clause markers. When punching



English translation according to Swedish word order

First sentence:		Second sentence:		Theme codes in column:
John	John	När	When	15 condition
blev	was	Fred	Fred	16 cause
ledsen	sad	gav	gave	17 concession
eftersom	because	Mary	Mary	18 consequence/intention
Mary	Mary	en persika	a peach	19 disjunction
slog	hit	åt	ate	20 comparison
honom	him	hon	she	21 interrogation
		den	it	22 supposition
				23 volition

Figure 4. Coding of relations

is carried out, the code is repeated in the theme columns until it is broken by a new code (does not apply at end of sentence). Code 46 as clause marker is an indicator of some form of relation that does not link up single concepts but whole clauses. Schank (1973, p. 206) expresses this thus:

"A relation is used to relate dependencies not concepts."

The interpretation of the first clause is simple, since the causal relation is expressed explicitly, so it need not be discussed further. The other sentence can be more difficult, however. It is not equally unambiguous. The interpretation in the figure means that the sentence states an action that has a consequence in a new action. In parenthesis it should be mentioned here that the

"consequence" and "intention" of the themes have been merged. Colloquial Swedish often does not distinguish between "so that" and "in order that" clauses. In the sentence coded here it is assumed either that Fred intended Mary to eat the peach or that Mary ate the peach as a consequence of having got it (He gave it to her so that she could . . .). It would also have been possible to code the sentence in another way, namely to look upon the first clause as a subordinate clause of time. Instead of 46, *när* (when) would then have been coded 43, which expresses an adverb of time, but the theme code would have been omitted. In such cases the breaking down in the clause columns is a help in retrieval. The main clause is the first, in this case with a complete paradigm (30 + 40 + 50). In addition to the paradigm there are concepts expressing various kinds of circumstances around the action. The action in this subordinate clause is *gav* (gave) and this action must be prior in time, which is the nature of such subordinate clauses. Other evidence is the word order in the main clause. The word order among the concepts is part of the identification. If code 40 stands before code 30 (in Swedish language), we can predict that a subordinate clause comes first, if there is no theme code for "interrogative". If on the other hand the word order in the main clause is the opposite, it is not possible to make such a prediction, but we must then use code 43. The problem with this type of relation need not be very great. Either we code the clause theme or have explicitly a time marker representing the relation; in both cases it is a question of the time aspect of the actions. The example shows, however, that there are many circumstances to take into consideration in the development of a computer-based system for content analysis. The AaO paradigm alone would have been insufficent when it comes to coding clause relations.

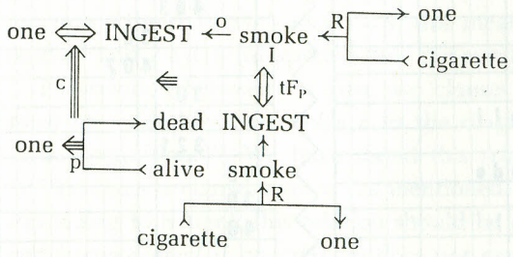
Thus we have codes for those aspects of a sentence that cannot be expressed by our paradigm. These aspects will be described in more detail in the continuing comparison.

Figure 8 states the place for modifications referring to aspects of the verb. In Figure 4 tense is coded in column 13 and mood in column 14. Schank (1973, pp. 206—207) says:

"Any conceptualization can be modified by certain *conceptual tenses* of which 'p' for past is one. /. . ./ These tenses modify a conceptualization as a whole."

In the same way as we code themes according to the meaning of the clause relation, we code tense and mood from the point of view of the verb. Schank gives an example of the importance of distinguishing relations from the "*conceptual tenses*" that can exist in ACT in the sentence: *Since smoking can kill you, I stopped.*

The conceptual representation is:



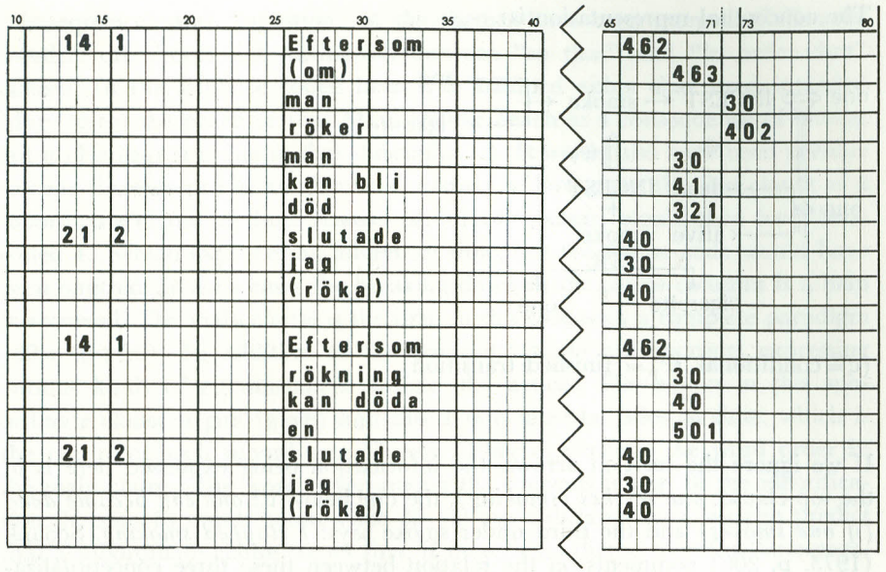
(c = conditional, tF_p = finished transition).

If we ignore the relation arrows, the information being expressed here is in the top clause: *one smokes (smoking)*, the one below it: *one can become dead (if one smokes)* and the third under smoke says: *I stopped smoking.* Schank (1973, p. 206) comments on the relation between these three conceptualizations as follows:

"This sentence contains two conceptualizations related by a causal and a causal relating that causal to a third conceptualization. Such a thing is nearly impossible to handle in more traditional lingustic representations."

We do not wish to let this last statement pass untested, so we shall try to analyze the relations according to the ANACONDA system, as shown in Figure 5.

The three conceptualizations *man röker* (one smokes), *man kan bli död/dö* (one can become dead/die) and *jag slutade röka* (I stopped smoking) can be expressed in codes in three columns, as shown in the top example in Figure 5. Schank's C-diagram contains a deeper conceptualization of smoking than we would give. For the sake of exemplification we have here divided the concept *rökning* (smoking) into the two concepts that are necessary for a clause to be syntactically complete. The second clause says the same thing as Schank's Rule 10, i.e. a change of state from alive to dead. It could also have been expressed via 30 + 40, i.e. *man kan dö* (one can die), without changing the import. In our case *bli* (become) stands for "change of state" and the result is *död* (dead), which is a description of an agent, thus a dependency concept (32). The 30 + 40 variant involves no difference, since the action contained in *dö* (die) has "itself" as a result. The third clause has been made complete by supplementation of the action indicated by the verb *sluta* (stop). We have the same rule here as Schank (1973, p. 207) expresses with the words:



English translation according to Swedish word order

First sentence:		Second sentence:	
Eftersom	Since	Eftersom	Since
(om)	(if)	rökning	smoking
man	one	kan döda	can kill
röker	smokes	en	one
man	one	slutade	stopped
kan bli	can become	jag	I
död	dead	(röka)	(smoking)
slutade	stopped		
jag	I		
(röka)	(smoking)		

Authors' comments:

The translation follows Schank's representation. Therefore the Swedish *man* corresponds to *one* (and not to *you*). The same goes for the text in connection with this Figure.

Figure 5. Coding of relations

"The English word 'stop' for example is actually an instance of the conceptual tense 't_F' and thus predicts an ACT. That ACT was unstated . . ."

As far as tense is concerned, we have not made all the distinctions that would be possible, e.g. start of action, ongoing action, terminated action, etc. *Börja* (start), *bruka* (be used to), *sluta* (finish) etc. are coded as verbs and are part of the whole verb complex. In the example these are separated but belong together through having the same code. There is a programme that places the verbs together in a string for further processing. We have not yet decided where we stand regarding the shades of meaning mentioned above, but it is quite possible to determine content in retrospect, when we have a sufficient number of examples of auxiliary verbs or "incomplete" verbs; they could then be

specified in the coding rules. Tense is coded in column 13 with code 1 for the present, 2 for the past and 3 for the future. If (as in our example) we have in columns 13 and 14 the codes 1 + 4, this means "for the speaker present time, modal" and 2 + 1 is interpreted "for the speaker past time, in reality".

The relation between the first two clauses in the top sentence in Figure 5 must be made clear. Schank's *c* in the diagram with the dependency arrow states "conditional". We have stated this relation with an "if". This kind of "if" becomes a clause marker (as mentioned earlier) and is assigned code 46. According to Schank this relation should be a causal, but in what way? *One can become dead if one smokes* does not express a conditional relation. The condition for considering oneself dead is not necessarily that one has smoked. Conditions must be expressed as a real relation. The reason for one's death need not be smoking either, but it *can* be. In other words, it indicates a potential causal connection. We do not code any theme for that particular relation, but we state connection by means of code 46. Instead we express mood by means of code 4 in column 14. This relation, the potential cause, is then connected with *jag slutade röka* (I stopped smoking). The clause marker for this is *eftersom* (since), which consequently has code 46. The import of the whole sentence must be interpreted as: *Since there is a potential causal connection between smoking and death, I stopped smoking.* This is the reason why I stopped smoking and this later causal connection must be expressed by a theme code, namely in column 16. The fact that I stopped smoking is a consequence (2) of the fact that I have thought about the risks (1).

Hopefully, this description has shown that code 46 indicates a clause, forwards or backwards. This is stated with figures for the order in which the clauses come in the sentence. In computers the reading process is arranged so that each concept (punch card) is read separately in the order in which they stand. It can happen that in our analyses we wish to work with the single concepts without identification, theme or position in sentence, by linking concepts. The loop code exists so that the concept codes will not "hang in the air". To give an example of the loop code, we can make the following search. We want to know how different verbs are related (we count 41 + 32 as corresponding to 40) and the machine collects three verbs from this sentence in the order in which they stand. We have no other concepts to guide us. The first verb is coded in columns 72 and 73, which is shown on the punch card. Then we wonder: How are the verbs connected? The machine is given an instruction in column 74, saying, "go to number two clause" (loop). There it meets code 46. In that clause is the second verb. Then we can agree that the second verb is dominant over the first one. After the second comes an instruction to go to clause number one. There a new code 46 states a connection with the third verb. Without "knowing" which sentence is involved the computer can in this way group concepts that form a structure.

Now we go back to the example used by Schank: *Since smoking can kill you, I stopped. Rökning* (Smoking) was the same as *man röker* (one smokes). The third clause in the first example is thus contracted to one concept, *rökning*. Then we find ourselves in clause two by looping. *Kan bli död* (Can become dead) was the same as *kan dö* (can die). How is *man kan dö* (one can die) connected with smoking? Well, looping says that there is a causal connection: *one can die because of smoking*, i.e. *smoking can kill one*. Example 2 in Figure 5 expresses this. We then see the concept *smoking* as agent and *one* as the victim (*en* = oblique form). Nothing in the conceptualization has really changed; the theme coding is the same. As the last example shows, ANACONDA would have represented the sentence. We do not consider that we need to code the fact that smoking means that *one inhales smoke by a cigarette (or rather smokable object) transporting the smoke from itself to one.* The way in which the concept smoking should be specified semantically in a dictionary is a much later question.

We have tried to show here that ANACONDA can handle complex relations despite its relatively simply constructed code system. There is still one point to be made, however, concerning the coding discussed here. We do not obtain the fact that is implicit in *I stopped smoking*, namely *I smoked (or have smoked)*, which must mean that I have *started and continued smoking, and done so at certain intervals.* But on the other hand Schank has not discussed this either.

According to Schank, the other conceptual relations are "time" and "location". Two rules symbolize the time relation, one of which refers to a concept and the other one states that a conceptualization is a time aspect of another conceptualization. The third indicates that an event must take place somewhere. These three rules are symbolized in Box 6, together with the corresponding ANACONDA codes.

Box 6. Comparison between the C-diagram and ANACONDA: Time and place relations

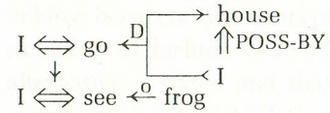
C-diagram	ANACONDA
Rule 12: T ↓ ⟺	43
Rule 13: ⟺ ↓ ⟺	43 (+ loop + clause)
Rule 14: PP ⇓ ⟺	44

T in Rule 12 stands for a time concept, as e.g. *yesterday, at 12 o'clock*. The concept is related to a whole conceptualization. Schank (1973, p. 207) says:

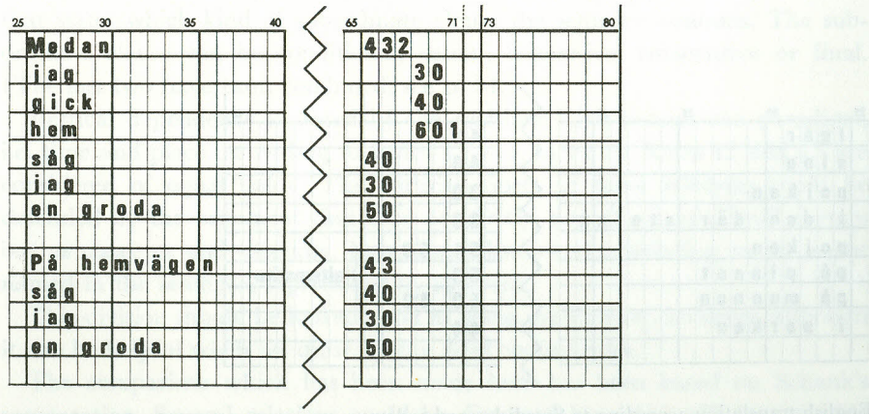
"The time of something modifies the entire conceptualization and not any particular item in it."

As in the case of location, time does not refer to ACT, like a case, which must exist explicitly in certain verbs. For practical reasons we have adapted these codes to traditional designations of these concepts, namely adverbs of time and place. Therefore the codes state a dependency on the verb. Since the verb occupies a central position in a conceptualization, these codes can be defended.

Rule 13 is explained by the example, *While going home I saw a frog*, which is represented



The symbol for "time" (see Box 6) expresses two actions, which are really difficult to separate in time, so as to get one preceding the other (cf. the discussion around Figure 4). The top sentence expresses time, encompassing a direction goal and the lower one denotes event. The time clause could be contracted to *på hemvägen* (on the way home). Our coding of Schank's example according to rules 12 and 13 are shown in Figure 6.



English translation according to Swedish word order

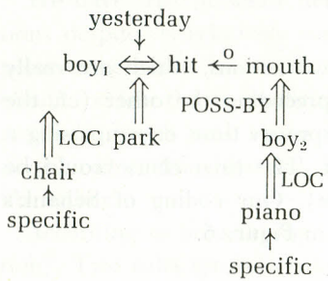
First sentence:				Second sentence:	
Medan	While	såg	saw	På hemvägen	On the way home
jag	I	jag	I	såg	saw
gick	was going	en groda	a frog	jag	I
hem	home			en groda	a frog

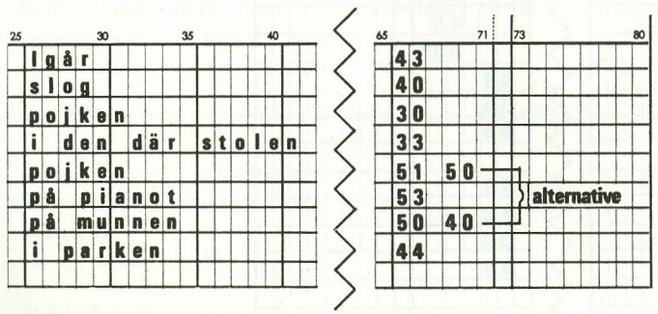
Figure 6. Coding of time concept and relation

The top example is coded according to rule 13 and follows Schank's representation. In Rule 12 (Box 6) we can see that the arrow for time is not double, i.e. should not contain a PP. Our *hemväg* (way home) could not be interpreted as a concrete noun *väg* (road) by the code for time, either, so therefore no risk would be involved in contracting the clause to one concept.

Rule 14 is interesting, since there are different kinds of location definitions. This rule does not refer to such definitions as are dependent on a PP (according to Rule 5, Box 1). Schank gives an interesting example of these differences. This example will be given here, immediately followed by ANACONDA's coding. The sentence is:

Yesterday, the boy in that chair hit the boy on the piano in the mouth in the park.





English translation according to Swedish word order:

Igår	Yesterday	pojken	the boy
slog	hit	på pianot	on the piano
pojken	the boy	på munnen	in the mouth
i den där stolen	in that chair	i parken	in the park

Figure 7. Theoretical and practical representation of locality concepts

We can see in Figure 7 that the two LOC are dependent PP and they are coded with vertical dependency codes (33, 53). *Park* stands between the agent and the action and can therefore be seen as dependent on the action that took place; *i parken* (in the park) is coded with 44. So far it is easy to divide up the sentence according to our method. What makes the example interesting is the coding of *på munnen* (in the mouth). As Schank states, boy number 2 is the owner of the mouth hit by boy number 1. Then according to our suggestion, *munnen* (the mouth) should be coded as main word and *pojken* (the boy) as qualifier in the genitive sense. The whole concept complex is then interpreted as *pojkens på pianot mun* (the mouth of the boy on the piano), which must be correct. But really our coding does not say that exactly. Code 53 is to be a supplement to the main word in the complex and then we would get *på munnen på pianot* (in the mouth on the piano). In such cases when the linking between two concepts is not semantically meaningful, the search is extended to include code 51, where the dependency relation emerges. One alternative is stated and that is that it is also conceivable to state the whole action (40 + 50) as ACT, i.e. code 40, which is then regarded as the idiomatic expression *slog på käften* (punched on the jaw), where *käften* (the jaw) need not be interpreted literally. The first coding is probably the most adequate and follows Schank's model.

Rule 11a and b could be expressions of a horizontal dependency, since they express that two conceptualizations have a causal or other connection with each other on the time level, so that the one presupposes the other in time. These relations are coded in our system with 46 + loop, together with a theme that states which kind of subordinate clause the sentence contains. The subordinate clause can be conditional, causal, concessive, consecutive or final. (The last two have been combined, see p. 54.)

Vertical dependency is stated by the relations lying outside the paradigm, i.e. time and place, according to Rules 12—14, in which Rules 12 and 13 are considered of equal value. The vertical aspect in these relations could be defended by the argument that they are firstly, coded not as overall themes but as concepts and secondly, have dependency codes referring to the event named in the sentence, represented by the verb.

This division should be looked upon as being preliminary, however, although it can be helpful when we draw up the final coding rules.

The comparison which has been made here has been based on Schank's presentation. Several relations could be discussed here, but will be set aside in this context. What we have tried to show is the way in which a theory about language in the form of symbols and underlying thought structures can be represented in the form of figures for input into computers. The 14 rules for concepts and relations are fundamental. In his presentation Schank goes on to "conceptual semantics". We are not yet prepared to undertake a com-

parison with this, for the reason that we base the "content" in the concepts on an empirical testing in which the statistical analyses are not ready.

3.5 Control of coder agreement

If we are to be able to develop a CCA method, it will be necessary for us to create a system of rules that two or more independent coders can use with a high degree of mutual agreement.

As has been mentioned, researchers at institutions for educational and psychological research in Sweden have been interviewed about their research situation, ideas for research problems, strategies and techniques in carrying out research tasks and their method of gathering information about the problem area concerned. A detailed description of the plan of the investigation and its execution are to be found in B. Bierschenk (1974). Data have been collected via interviews with both open-ended questions and statements with response categories of the Likert-type. The open-ended questions have resulted in a set of material covering 4000 pages of text. Such a large amount of text can naturally not be used in the development of ANACONDA, and so about 10 % of the material has been processed.

By means of a random table, four interview subjects (31, 2, 40 and 33) were picked out from the interviewed sample of 40 researchers from a population of 126. From the respective interviews, four interview questions (5, 6, 7 and 8) concerning information and documentation have been chosen. It can be assumed that the information that will be extracted from the text will be relatively concrete and consequently easy to interpret. This should be an advantage in the development of a new technique.

The interview questions were to be coded in their entirety, so that the context of the discussion could be used in supplementation. Spreading the selection of text over the entire text or over all the subjects has been considered an unsuitable method of procedure. The intercoder agreement was examined with regard to

1. segmentation of concepts. A check is made of whether both coders have supplemented and deleted identical elements (words).
2. segmentation of clauses. A check is made of whether the coders have identical clauses.
3. assignment of codes to concepts. A check is made of whether both coders have assigned identical codes to one and the same concept.
4. assignment of codes to themes. A check is made of whether both coders have assigned identical codes to one and the same theme in a sentence.

Table 1. Summary of intercoder agreement in applying ANACONDA

Steps in the analysis		Interview person No. 31	2	40*	33
(1) Segmentation of concepts	z	3.92	2.20	—.58	3.21
	i	.88	.86	.82	.86
	N	799	1098	237	1255
(2) Segmentation of clauses	z	2.82	2.64	.67	—2.76
	p			.75	
	i	.94	.93	.92	.84
	N	165	227	47	246
(3) Assignment of codes to concepts	z	7.64	9.42	1.16	8.51
	i	.91	.92	.83	.90
	N	841	1089	222	1190
(4) Assignment of codes to themes: source, time, mood	z	7.33	5.51	1.40	4.37
	i	.98	.93	.93	.93
	N	320	397	83	422
Segmentation of concepts before check on comparable text	z	—9.89	—13.08	—4.71	—17.60
	i	.77	.76	.76	.74
	N	1013	1377	272	1673
Assignment of codes to concepts before check on comparable concepts	z	—2.47	—4.52	—6.23	—10.40
	i	.83	.82	.73	.78
	N	992	1328	283	1549

z test value, binomial test
p probability: $p < .05$ states that the criterion .80 has not been achieved
i Osgood's index for agreement
N total number of assessments

* Ip 40 has given oral comments to question 5. Questions 6 and 7 were answered by filling in a questionnaire, while the Ip did not comment on question 8

All the comparisons are of the same type, i.e. either there is agreement or not. The number of common judgements has been noted. In addition the total number of judgements and the number of judgements each coder has made separately have been calculated. An extremely detailed scrutinization and comprehensive documentation may be found in Berg (1974). Here, however, only a summarized table will be presented, with the values for points 1—4 above. The values have been compiled from Berg (1974, p. 30).

The checks of the intercoder agreement in the steps of the analysis carried out so far show that segmentation can be done with a satisfactorily high level of agreement. As Table 1 shows, Osgood's index for agreement is between .74 and .98. Spiegelman, Terwilliger & Fearing (1953, p. 175) give as the mini-

mum requirement an index value that is equal to or greater than .75, irrespective of the method by which the intercoder agreement has been estimated. Osgood et al. (1956, p. 59), report index values of between .64 and .88. Our result is by comparison very satisfactory, since the analysis this report is dealing with is much more detailed and comprehensive. In addition the interview material contains for natural reasons greater variations, while at the same time it is less complete than Osgood's printed material.

The binomial test shows, however, that the critical value .80 could not be established in every case. As is shown in Table 1, neither the "segmentation of concepts before a check on comparable text" nor the "assignment of codes to concepts before a check on comparable concepts" has resulted in satisfactory values. This is caused by the lack of unequivocal rules. If, for example, one coder uses the term "researcher" while another describes the same person as a "behaviourist", this leads to differences in supplementation. This difference can, however, be nullified by e.g. rewriting the rules for supplementation, appropriate construction of dictionairies and facetting. All the supplementations are marked in parenthesis, which makes it possible for us to analyze the material both with and without supplementations and thus investigate the extent to which this leads to different results.

The index values reported above the line are comparable with the results that we would have got by limiting concepts in written text. As can be seen from Table 1, the agreement is good, though with the exception of "Segmentation of clauses" in interview No. 33. This is probably a result of there being a large number of unsupplemented clauses (see Berg, 1974, p. 23).

Attributes and adverbs have obviously caused most of the deviations in the coding. The agreement for attributes is admittedly over 80 % but some of the deviations could be explained by the confusion that has occurred between the two categories. Thus, the coding of e.g. "Researcher A in Malmö" has partly been coded as an adverb of place "in Malmö" and partly as a postpositive attribute. In addition there has been confusion between adverbs of time and degree. In the clause "I read daily", the word "daily" has been coded both as a statement of time (adverb of time) and as a statement of frequency (adverb of degree). For the examples presented here, the rules will be improved.

3.6 Computer input of text

The text material that is to be processed by means of ANACONDA has been written down from a tape-recording in as authentic a state as possible, which means that we must first treat the text before any computer processing of the text can take place. Thus the text must be cleaned up, so that it can be broken down into sentences and clauses. This means that there must be rules for

Box 7. Example of authentic text in the interview material and treatment prior to coding

I: Could you say anything about how the search for information should be planned in order to create ideal conditions for the research process?

Ip: I'm a tremendously bad researcher when it comes to things like that. I'm bloody uninterested in so to speak, I think this kind of problem is also, well you refute this by pointing to X's paper, or her thesis, but my argument is that you, that important tip you gave, that was a tip at the right moment, one would have been sure to hear about hers all the same. But otherwise I am tremendously unsystematic in my whole way of searching and e.g. I've never once gone to any of these collections of references, but on the other hand, I've sometimes tried to be systematic when looking through journals and suchlike, but I soon abandon it, but on the other hand on some points I read up on an area or a journal extremely well, but that means that I have partly made very little use of that type of compendium, abstracts and so on, and partly I have very few recommendations. Hell, didn't someone write somewhere about that being a symptom of an intellectual crisis or something, that you have to reduce information to be able to absorb it or something. But I don't really remember what kind of . . ., but you see to be able to use information, you have to destroy it, you see. — — —

I: Could you say anything about how the search for information should be planned in order to create ideal conditions for the research process?

Ip: //I'm a tremendously bad researcher when it comes to things like that.// ~~I'm bloody uninterested in so to speak, I think this kind of problem is also, well you refute this by pointing to X's paper, or her thesis, but my argument is that you, that important tip you gave, that was a tip at the right moment, one would have been sure to hear about hers all the same~~//But otherwise I am tremendously unsystematic in my whole way of searching//and e.g. I've never once gone to any of these collections of references,//but on the other hand, I've sometimes tried to be systematic when looking through journals and suchlike,//but I soon abandon it,// but on the other hand on some points I read up on an area or a journal extremely well,//but that means that I have partly made very little use of that type of compendium, abstracts and so on,//and partly I have very few recommendations.//Hell, didn't someone write somewhere about that being a symptom of an intellectual crisis or something, that you have to reduce information to be able to absorb it or something.//~~But I don't really remember what kind of . . .~~//but you see to be able to use information, you have to destroy it, you see.//— — —

Irrelevant analysis text
Deletion within analysis text
Segmentation of sentences //

coders (see I. Bierschenk, 1974, p. 50). In order to illustrate how our authentic material can look, Box 7 presents a section of text, followed by the same text after treatment.

Our analysis unit is the *sentence* and not the individual word (element). Each sentence is analyzed by itself in clauses and the parts contained there, i.e. *concepts*, which are coded according to the function in a clause. Thus the computer-input is concepts, which can consist of one or more linguistic elements. If we had transferred the entire text on to punch-cards without prior segmentation, we would have given ourselves a lot of work with the recognition of the different parts of the text. But we have never intended to develop such a method of analysis.

The text that is to be analyzed consists of interviews, in which the speech of the interviewer is not to be included in the analysis. Only the opinions of the interviewee are reflected in the flow of the text. But it can happen that the person interviewed refers to what someone else has said or expressed and we have wanted to distinguish such information. Therefore we have created a special column that we call "source of information" with two alternative codes, 1 for the interviewee's own opinions and 2 for the opinions of others.

We are also convinced that the theme is essential; it is namely there that the whole sentence gets its meaning (the counting of nouns as a measure of a certain content can hardly be reliable information, if a clause or a sentence expresses supposition, negation or suchlike). By coding the theme we can both compile dictionaries containing concepts, and gain access to themes of clauses. This means that we can work directly with parts of the text, e.g. the different themes on their own, without needing to search through the whole text for words like e.g. "not", as expressions of negation.

The coding is done on data forms, designed so that certain columns are specified for certain types of codes. Figure 8 shows a coded sentence, where numerical signs in the form of identification, theme, syntactical (or function) codes can be seen.

The writing down of the text is made as time-saving as possible, so that the same code is not repeated down the columns. In the punching the codes are repeated unless a new code breaks in. Figure 8 also shows clearly how each card can be referred to the text by repeating the codes in the punching. As far as the syntactical codes are concerned, we shall here only say that different clauses are marked by looping to other columns. The loop codes (Figure 8) mark in which clause column the sentence continues.

3.7 Control of punch cards

It is important that the text material that is to form the basis for the further development of ANACONDA is faultless. Otherwise it would be very difficult

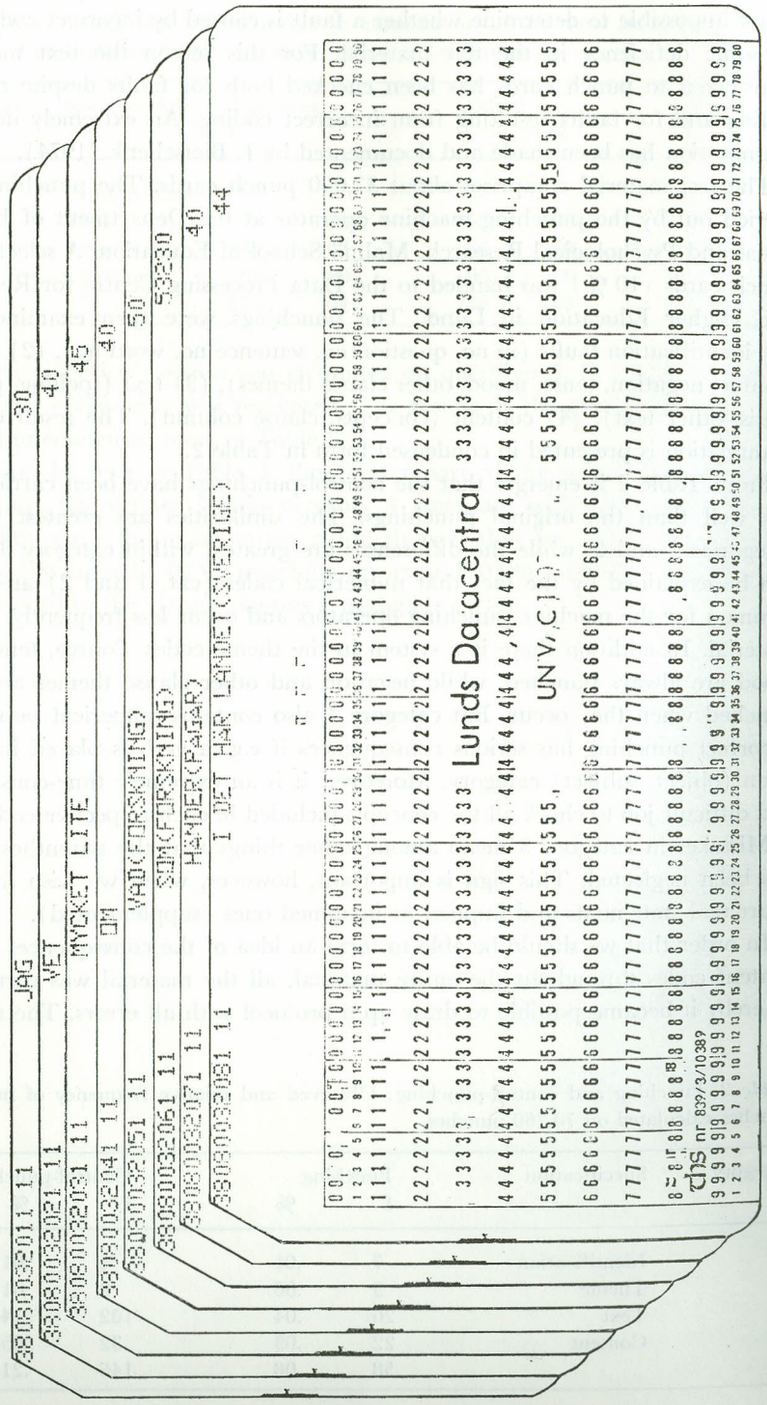


Figure 8. Punch card: Example of coded text

if not impossible to determine whether a fault is caused by incorrect coding or by some deficiency in the test material. For this reason the text material transferred to punch cards has been checked both for faults despite correct coding and for faults resulting from incorrect coding. An extremely detailed examination has been made and documented by I. Bierschenk (1974).

The test material comprises about 37,000 punch cards. The punching was carried out by the punching machine operator at the Department of Educational and Psychological Research, Malmö School of Education. A selection of punch cards (10 %) was handed to the Data Processing Centre for Research and Higher Education in Lund. The punchings were then examined for (1) identification faults (ip no, question no, sentence no, word no), (2) theme (source, negation, tense, mood, other clause themes), (3) text (spelling, parenthesis, other text), (4) content (concepts, clause column). The result of this examination is presented in condensed form in Table 2.

From Table 2 it emerges that the control-punchings have been carried out less well than the original punchings. The similarities are greatest within categories 1 and 2, while the differences are greatest within category 3. This can be explained by the fact that numerical codes (cat. 1 and 2) are more common for the machine punching operators and occur less frequently in this material. In addition there is a system in the theme codes. Source, tense and mood are always punched, while negation and other clause themes are only punched when they occur. But category 4 also contains numerical punching. Incorrect punching has serious consequences if e.g. a verb is placed in some noun (object, subject) category. Moreover, it is an extremely time-consuming and difficult job to check all the concepts included in each respective code.

Mistakes in category 3 mean among other things that the parenthesis sign has been neglected. This sign is important, however, when we wish to keep apart real statements and implied or imagined ones (supplemented).

In order that we should be able to form an idea of the consequences of the content codes throughout the entire material, all the material was corrected. Thereby it became possible to draw up a protocol with all errors. The text of

Table 2. Punching and control-punching: Observed and relative frequency of incorrect punches calculated on 70,260 punches

Category	Specification	Punching		Control-punching	
		f	%	f	%
1	Identification	7	.01	4	.01
2	Theme	3	.00	8	.01
3	Text	26	.04	102	.14
4	Content	22	.03	32	.05
Σ		58	.08	146	.21

Table 3. Punching and coding errors in examination of the total punched material: Observed and relative frequency calculated on 702,600 punches

Category	Specification	Punching		Coding		Σ	
		f	%	f	%	f	%
1	Identification	—		—		—	
2	Theme	6	.00	53	.00	59	.00
3	Text	182	.03	34	.00	216	.00
4	Content	88	.00	99	.00	187	.03
Σ		276	.04	186	.03	462	.07

all forty interview persons was examined on questions 5, 6, 7 and 8, card by card, and every error was registered. The results of the examination are presented in condensed form in Table 3.

As can be seen in Table 3, the greater part of the errors depend on the punching. Corrections within category 4 covariate with alternations in the text. But since the examination made showed that we only need calculate with approximately .04 % punching errors and .03 % coding errors, they are with regard to the clause columns a negligible factor.

4. Construction of dictionairies and quantification of concepts

In the code system concepts are divided up according to function (and often position) in the clause. As in traditional clause analysis this means that a verb can stand in the subject position, but that a concept that strictly grammatically is a noun, can have a verb function even though such changes between parts of speech are not the most common form of representation. An account will be given below of the way in which we can extract our concepts from the material put into the computer.

If we want to know how many and which adjectives there are, we order from our material the codes standing for attribute, with the import description or classification of a noun. All actions are registered in verb codes. Nouns can be found in the agent or objective codes, but also in codes lying outside the AaO paradigm, e.g. as place qualifier. Since we have primarily been interested in studying the codes within the paradigm, the specification of nouns will be limited to agent or object.

As has been said earlier, a concept can consist of several words, which often means a string of words, in which a noun is surrounded by articles and prepositions. This applies above all to the nouns. We wanted to obtain a basis for a register of the concepts that occur within the different codes. The first stage was to have the codes we wanted written out. After the particles had been removed the concepts were sorted. By different kinds of truncation, it became possible to search for these concepts, both in different inflections and combinations and in different codes where they might occur. The way in which concepts can be limited by truncation is described below.

4.1 Truncation of adjectives and verbs

Computer outputs have been compiled in order to construct suitable dictionaries. The concepts have then been truncated so that as few endings as possible need to be registered. Endings are counted as the element(s) that if occurring after an asterisk (e.g. *behaviour*, behaviour**s) give(s) the concept another meaning (in gender, comparison or tense). In simple terms we can say that a concept is truncated at the point up to which it is spelt the same.

In cases when the stem of the concept is mutated, i.e. has another vowel, the concept is registered as many times as the number of stem vowels. Box 8 shows the principles for truncation of adjectives and verbs.

Box 8. Examples of truncation of adjectives and verbs

Truncation		Outcome/endings
Adjectives		
stor*	(big)	-t, -a
stör*	(big, mutation)	-re, -st, -ste/a
fin*	(fine)	-are, -ast, -aste, -a, -t
hjälpsam*	(helpful)	-mare, -mast, -maste, -ma/e, -t
Verbs		
arbeta*	(work)	-r, -de, -t, -d[1], -s, -ts, -des
bjud*	(invite, offer)	-a, -it, -en[1], -s, -es, -its
bjöd*	(invited)	-s
slå*	(strike)	-r, -s
slog*	(struck)	-s
slag*	(stroke)	-it, -its, -et[1], -en

[1] Participle form (sometimes used as adjective)
The s-forms mean that we also allow for passive verbs (it is possible to say "he was invited". Swedish: "han bjöds")

The endings presented in Box 8 form our ending file. With its help we can use the truncation procedure to search through the material for concepts and concept combinations. In the next phase of the construction, the tense forms will be combined into one concept representing the others. This simplification is possible since we have theme codes, which specify tense in each clause.

4.2 Truncation of nouns

In Swedish there are more endings to nouns than to adjectives and verbs, because of the various inflection patterns for the many declensions and the definite and indefinite forms in different genders. In addition we also have the s-genitive, which can occur in combination with all the other forms. Plural inflection with another stem vowel also occurs and is treated in the same way as in the case of adjectives and verbs.

One of the most common derivative suffixes in Swedish is *-ning*. This is not regarded as an ending here, since in that case we would get several meanings

for a truncation, which could span over both abstract and concrete concepts, causing difficulties in building up dictionaries. We can take as an example the Swedish words *bokförsäljare* (bookseller), *bokförsäljning* (bookselling): a truncation *bokförsälj** does not only overlap with regard to inflectional endings. We regard ending in the same way as we did adjectives and verbs. Examples are given of the importance of derivative suffixes compared to inflectional endings. Thus Box 9 describes noun truncation that is safeguarded against derivative differences.

Box 9. Example of truncation of nouns: Derivation compared to inflectional endings

Truncation		Outcome/endings
arkivarbet*	(archive work)	1) -e, -et, -ets, -en, -ena, -enas
	(archive worker)	2) -are, -ares, -arens, -arna, -arnas
arkivarbete*		-t, -ts, -s, -n, -na, -nas
arkivarbetar*		-e, -es, -ens, -na, -nas

It should be obvious that only the latter type is satisfactory. Similar cases could also be discussed, in which the derivative ending appears to make very little difference, e.g. *anmälan* (application) and *anmälning* (application). There is a slight difference, however, and here we have consistently assumed that it is a question of concepts with different meanings.

Hitherto each concept has been regarded as standing alone and being different from every other concept. We have not yet clarified which are to be considered the same, i.e. having the same empirical meaning and consequently by means of a truncation procedure only needing to be represented once in a dictionary. We have tried to determine the empirical meaning by means of Osgood's "Semantic Differentials". Briefly the principle is that adjectives and verbs are assigned a value on a 7-point scale after assessment made by a panel of researchers. The weighted means of the dependent concepts connected with an independent concept comprise the "value" which is used for defining the empirical content of an argument, i.e. independent concept. This means that we consider nouns as *mnemonics* and adjectives and verbs as referents which give them their empirical meaning.

In our analysis the concepts are context-bound, in accordance with a theory about relations between dependent and independent concepts and role functions. It is in accordance with these conditions that coding takes place and further analysis may be possible. The coding of natural language is a listening phenomenon. What the computer is to achieve is not an understanding of text but a structurization of the text in agreement with the way in which we have specified the computer input.

4.3 Scaling of properties, actions and states

The import of our *theoretical constructions*, the purpose of which is to represent a phenomenon that evades direct observation, can only be studied by means of more or less sophisticated analysis models. As can be seen from Figure 2 (p. 23), there are on the latent level two different basic elements, namely *mnemonic* and *reference*. While the relations that exist between mnemonics define memory structure based on the individual's "impression formation", the relations existing between references define the individual's frame of reference. The latter arises through the ability of the individual to relate references to an object of perception. Both types have often been studied in relative isolation. It is above all Asch's (1946, pp. 258—290) article "Forming impression of personality" that has given rise to many experimental concept formation studies, aiming at mapping the cognitive structure of the individual. Cognitive structures have been studied from the point of view of perceived relations between adjectives, which define a property in the structure of a perception object. Wishner (1960, pp. 96—112) shows in an experiment that Asch's "central" and "periferal" traits can be predicted by starting from the intercorrelations that have been obtained independently of each other for both object lists and adjective lists, i.e. "stimulus list" and "check lists". Thus an experimentally proven relation exists between the structure of mnemonics and the structure of references that define an object. The cognitive structure of a particular individual can be defined by means of the perceived relation that exists between the properties defining an object. If these relations can be quantified by means of the values representing the covariation of these properties, it is also possible that we shall be able to determine the weights that each property should be given in a prediction of an object. Presumably the cognition of an object, i.e. assignment of properties to an object, presupposes a multi-variate processing of information, since different scaling experiments have shown that adjectives have multi-dimensional content (see van der Kloot, 1975, p. 23). If one wishes to study both memory and reference structures, it becomes necessary for the object and properties to be scaled in isolation from each other. In this case we have a situation in which both two or more objects and two or more adjectives are permitted to vary and the analysis model could be a multiple regression or correlation analysis.

It is above all van der Kloot's (1975, pp. 60—68) experiment that provides empirical evidence of the importance of object and adjective in the formation of the individual's cognitive structure. van der Kloot used adjective-occupation combinations for the purpose of determining the configuration by means of which occupations, adjectives and adjective-occupation combinations could be depicted. In the main the analysis models he used were canonical analysis, discriminant, and multiple regression analysis. Unfortunately multivariate

analysis has only been used as a post-control. By scaling the adjectives and the objects (occupations) independently of one another, he made it possible to analyze the importance of adjective and object respectively in the cognition process. Occupation-adjective combinations were predicted with two models ("summation" and "averaging" models) as the starting point. The analysis model used was a multiple regression analysis in order to find an optimal empirical combination rule for the occupations and adjectives. The comparison shows that the summation model gives the better prediction. The weights in the regression equation show that the adjectives play the dominating role in optimal prediction of personality features (in van der Kloot's case). The importance of adjectives becomes even more apparent if one compares the coefficients for the first and second dimension in van der Kloot's (1975, p. 67) analysis. For the first dimension ("evaluation") they are (.137) for the occupations and (.973) for the adjectives. For the second dimension ("dominance") they are (.022) for the occupations and (.909) for the adjectives.

The four method studies carried out by van der Kloot also show that

1. the addition of an adjective changes the order among the occupations in the direction stated by the loading for each respective adjective
2. dispersion of the occupations with fixed adjectives is less than for the occupations without an adjective
3. the occupations that are very similar to one another but that are presented together with different adjectives display very large dispersions along both dimensions.

An investigation was also made as to whether and to what extent the original "occupational stereotypes" of the individual survive when the individual is given additional information in the form of adjectives. The result of this study (van der Kloot, 1975, p. 80) shows that "occupational stereotypes" disappear when adjectives are added to the description.

To sum up, the results that have been presented support the assumption that it is the adjectives and not the nouns that form the base for the conceptualization and that this can be described by three dimensions that are on the whole independent of one another. Using a statistical elimination of the evaluation, we ought in addition to be able to show a more differentiated factorial structure of the models governing the actions of the researcher.

A method often used for a quantitative description of properties and processes or state is Osgood's Semantic Differentials. The method is an attempt to study the individual's reactions to different types of object. The result of these attempts shows that one seldom needs more than three dimensions, namely "Evaluation" (E), "Potency" (P) or dominance and "Activity" (A).

Many people have tried to interpret the meaning of semantic differentials and discussed the usefulness of this technique. It is above all the constancy of

the factorial structure (E—P—A) and the psychological implications given to it that have been debated. Usually the discussion concerns the denotative and connotative or affective implications of the scales. But irrespective of the theoretical standpoint one adopts, one cannot ignore the fact that this factor structure exists (see Miron, 1969, p. 189). Osgood (1969, pp. 194—199) claims that there is a fundamental agreement between this structure and Wundt's (1918, p. 100) "Gefühle als dreidimensionale Mannigfaltigkeit", namely (1) "Lust—Unlust", (2) "Erregung—Beruhigung" and (3) "Spannung—Lösung". Kuusinen (1969, pp. 181—188) analyzed 59 adjective scales that have been used to describe the individual's personality. A factor analysis and varimax-rotation were carried out, partly on the basis of the product-moment correlations between them, partly on the basis of partial correlations. Since Kuusinen (1969, p. 185) partialized out twelve semantic scales that measure evaluation, the mean correlation was reduced from .559 to .336, which shows that there was sufficient analyzable variance left. By means of this statistical manipulation, the evaluation effect of the adjective (see Asch, 1946, p. 259) is kept constant. The result shows that an elimination of the adjective's evaluation leads to factors that from the point of view of a psychology of personality provide more meaningful dimensions than was the case when the evaluation effect was not partialized out.

Rosenberg, Nelson & Vivekananthan (1968, pp. 283—294) found in a multi-dimensional scaling of 60 adjectives that these form a mainly two-dimensional space that could be described by means of an oblique rotation. The first dimension was described as "good—bad" and the second dimension was designated "hard—soft". A third, though weaker, dimension emerged and was designated "active—passive".

Another problem in the scaling of adjectives concerns the exact formulation of rules by means of which a multi-dimensional content of an object is related to the individual's decision to choose an adjective or to state a particular definite assessment regarding the property in question.

A few pairs of opposites are not associated with each other irrespective of which is the stimulus (e.g. "old" is not associated with "birth" and "young" but often quite the contrary). These form opposites irrespective of the order in which they are offered. They are called "true opposites" (Deese, 1965, pp. 181—212).

Considering these results, we have chosen to scale our adjectives and verbs by means of seven-point assessment scales, the bi-polar terminals of which are described by the pairs of adjectives (1) positive—negative, (2) active—passive and (3) strong—weak.

We use adjectives as stimulus objects since we wish to obtain detailed information about the implicit models used by the individual in order to form an opinion.

While researchers making use of Semantic Differentials have usually used adjectives to describe an object, we consider that the boundaries between adjectives and verbs are vague and that in principle all "dependent" concepts should be utilized in the description of a phenomenon. Ross (1969, pp. 352—353) refers to the well-known language theoreticians Postal and Lakoff and states:

"... the parts of speech which have traditionally been called *verbs* and *adjectives* should really be looked upon as two subcategories of one major category, predicate. /.../... adjectives and verbs are members of the same lexical category. /.../ It should be obvious, however, that to accept this claim is not to maintain that verbs and adjectives behave identically in any respects, but only that their deep similarities outweigh their superficial differences in syntactic behavior."

Since in our analysis we take in account "syntactic behavior" and regard both adjectives and verbs as descriptive concepts, we have chosen to scale adjectives and verbs. Adjectives directly describe a noun. Verbs describe the object more indirectly through the process in which a noun is involved. In the coding of the interview material, consideration has been taken to the qualifications of adjectives and verbs, but how these are to be scaled constitutes a separate problem.

The scaling method that is to be developed for a quantitative description of of interview text is based on the assumption that the import of verbal material can be described by means of three main dimensions. Further it is assumed that a number of judges can give a reliable description or assessment of properties and processes against the empirical background they have. If twelve or more judges are used for these assessments, the reliability is as high as for the more valuable of the objective tests (see e.g. Guilford, 1954, pp. 251—256; Cattell, 1973, p. 250).

Scaling adjectives and verbs means that we abandon the classical way of using Semantic Differentials since we create assessments that are independent of a particular object of assessment. By using such a procedure, we can avoid the problem that the semantic structure in the selected adjective scales is changed as a function of different categories of objects.

4.4 Procedure for scaling adjectives and verbs

Following the discussion hitherto and the results presented, we decided that in the first phase the scaling should apply only to adjectives and verbs. Adverbs and articles were removed. All adjectives were included as basic forms in the adjective lists which were to form the basis of the scaling. The verbs were changed so that the infinitive form represented different variants. The material was treated in this way in order to create files that would be constructed as

economically as possible. Each adjective and verb that came to be included in the respective files would be given three different values corresponding to the three dimensions that are considered to describe a semantic space. Pre-studies showed that the persons participating found it easier to assess an adjective or verb by means of a scale graded from 1 to 7, than with the typical scale, where the minus sign is used together with the figures 1—3 and with the zero value as the middle point on the scale. The middle point is given with the figure 4 on the seven-point scale (1—7).

In the project's interview study (see B. Bierschenk, 1974, p. 33) the population of researchers has been defined. It would have been desirable for the researchers participating in the interviews to have assessed their own adjectives and verbs, but for several reasons this method of approach could not be employed:

1. It was considered impractical to let all 40 interviewees assess all the adjectives and verbs extracted (a total of 1453), not least considering that these persons had already participated extensively in the investigation.
2. If each individual interviewee assessed only the adjectives and verbs that occurred in his own interview, we would admittedly have got this person's assessment, but since only a few words are common to the majority of the interviews, it would have been difficult to create an assessment base that all the interviewees could have in common.

The method of approach which should create assessment values for all the adjectives and verbs extracted in the dimension concerned (8718 assessments) and which in addition it should be possible to generalize to the researchers interviewed, is panel assessment. Since the persons included in the assessment panel are covered by our definition of "researcher", they are assumed to have the same background of experiences and thus the same reference system as the interviewed researchers are assumed to have. One possible limitation to generalization is the fact that all those in the panel come from the south of Sweden, from the Malmö—Lund area. But on the other hand there is nothing in the evaluated material to indicate any regional effects. In order to achieve maximal certainty in the assessments, it was decided that all the researchers from our population who had not taken part in the interview study should be included in the assessment panel. These form a random sample, since the interviewees were chosen by means of a random table. The total number was 20. Four of these were excluded, however, two because of commitments abroad, one as a result of ill-health and the first author of this book, who was considered too involved in the material to be able to take part under the same conditions as the others.

The remaining 16 researchers were asked personally if they were willing to

participate and after all sixteen had replied positively the material was sent by post to each researcher's home address, together with instructions for handling it. Since those participating work at very varying hours, it would probably have been unrealistic to try to carry out the assessments at a given place and time. Instead each one could use his time as he thought best, although preferably within a limit of approximately two weeks from the date of posting.

Thus for practical reasons, it has been impossible to check certain factors, such as the time taken for the assessments and the time of day when the assessments have been made. Nor can we know with certainty if the assessors have followed the given order of work stated in the instructions. Other factors could be checked, however. In order to avoid some concepts being liable to a tiredness effect, the order of the words has for each individual assessor been determined by the generator of random numbers, i.e. 16 different random orders of sequence were generated, one for each researcher. In addition the three dimensions have been separated in order to counteract any mixing of the individual scales, which can easily happen when they are to be assessed together. This means that each person received six different random orders of sequence.

This arrangement has been possible only because we have had access to the Computing Center of Lund University. The programming has been done by Fil. Kand. Leif Robertsson. The computer print-outs have saved much time and are more valuable from the point of view of legibility than the typed version. It has also been a great advantage that it has only been necessary to check the punch cards. Finally it must be said that our time limit proved to be far too optimistic. The last assessments were received in May, as opposed to the anticipated date, March 15, 1975. One of the assessors has made no assessments at all, despite strong and persistent pressure.

4.5 Processing and description of data

The assessment panel made its assessments during the spring term of 1975. When all assessors (except one) had returned our computer print-outs, these were checked for possible non-response. In general all the assessments had been made with great thoroughness and there is no non-response in the assessments apart from 3 and 5. The non-response here seems to be a consequence of their having had difficulties in managing the computer print-outs. For assessor 3 the non-response in the assessment of adjectives ($n = 570$) is for evaluation 15.8 %, activity 10.5 % and potency 5.7 %. In the assessment of verbs ($n = 883$), the non-response in assessor 3 is for evaluation 9.3 % and potency 3.7 %, while the non-response in assessor 5 is for potency 11.8 %.

Considering that the non-response for the other 13 assessors is practically non-existent, the percentages given appear relatively high. But since we can be

sure that no systematic non-response has occurred, we have decided to replace the missing scores by estimating mean assessments of the group.

Following this design for the panel study means that only the numerical values of the assessments needed to be transferred to punch cards. After these checks, the first statistical description of the material was made. There is a frequency statistic for each individual word, namely the number of assessments for each alternative answer on the seven-point bipolar scale, the number of assessments, mean and standard deviation. There is also a frequency statistic for each individual assessor's assessments with regard to evaluation, activity and potency. It states the number and proportion of assessments per alternative answer, non-response, mean and standard deviation. Since the material is so extensive, it is for practical reasons very difficult to give an account of the basic material. Any reader who is interested in the basic material may obtain access to it through the authors.

If we are to measure property combinations, these combinations must be assessed directly. The computer analysis concerns this analysis of property structures. The first question we need to answer is: Has the assessment panel with satisfactory reliability in the assessments been able to assess adjectives and verb in accordance with the three assumed dimensions?

In order to study this question, a component analysis was made for adjectives and verbs. The observation values obtained were ranked according to the following covariation schedule: Measuring object 1(1)570 and 1(1)883 respectively, variables 1(1)15 and scales 1(1)3. A separate component analysis was carried out for each scale.

In order to obtain a coefficient for maximal reliability for the respective scales, each position on the seven-point scale has been weighted according to the component analysis. The coefficient for maximal reliability was introduced by Lord (1958, pp. 291—296). This coefficient is a simple function of the largest characteristic root of a correlation matrix for the variables forming the scale. This coefficient (α_{max}) is well-known and the random sample characteristics of the coefficient have recently been presented by Joe & Woodward (1975, pp. 93—98). In the evaluation of the assessment panel's assessments each position on the scale has been weighted.

4.6 Analysis of data

We assume that independent assessors can state the meaning of a particular adjective or verb. This means in our case that each individual word can be assessed with regard to three different characteristics, namely what evaluation a particular word expresses, what activity it states and what potency it has. Some of the assessors included in the assessment panel have commented on this task. In order to illustrate the reactions produced by the assessments, a few

comments will be given here, made by assessors with relatively different basic views on the research process. The following points of view were given by telephone:

"Swedish has a lack of words, which means that the same word can have many meanings. It took me a long time to find out which it should be. The verbs were more difficult to assess than the adjectives. This can probably be seen in the assessment through there being fewer extremes. They are not equally distinct. It was easier to assess the adjectives since one has a better idea of what they mean."

Another assessor gave the following written comment:

"Difficulty in keeping separate (1) the face value of the words and (2) their psychological performance, such as 'be familiar with'. The face value is 'weak', the psychological expression is 'strong'. Example: *Suffer a lack of*. The face value is 'strong', the psychological expression 'weak'. I have (although probably not consistently) looked for category 2."

A third assessor wrote:

"Am not sure that I have maintained the same attitude to the scale throughout each section, but have tried to."

A fourth written comment was:

"Have done my bit but also want to express my considerable doubts about the whole procedure. /. . ./ Against this background our scales become a game of "Blind Man's Buff" with reality."

A fifth and final comment on the assessments:

"/. . ./ Mostly the work is easy. There is usually no difficulty in making the assessments. Sometimes it gets tough, however. Some examples follow. /. . ./ The language is obviously used carelessly. Possibly the choice of words would have been plainer, if one had been given whole expressions, not just single words."

These comments on the scaling procedure chosen show how different assessors have experienced difficulties in assessing adjectives and verbs detached from their context. Since the persons on our assessment panel are our "measuring instruments", the question arises of whether there is any empirical (objective) basis for the doubts expressed, or whether they are simply more subjective, casual or rather unsystematic observations. The first measure to be taken is to test whether the error variance in the assessment panel ("measuring instruments") and the error variance originating from the conditions under which the assessments have been made exceeds systematic variance, i.e. the variance that is constant over a number of repeated measurements. If, as a result of the fluctuations in the assessors' assessments, the variance is low, we can establish that there is a high degree of reliability in the assessment of the different dimensions that characterize adjectives and verbs.

Table 4. ANOVA design for assessment of intraclass correlations

Index	Adjectives		Verbs	
	Words	Assessors	Words	Assessors
No. of levels	570	15	883	15
Size of population	∞	∞	∞	∞

An estimation of the reliability in the assessment panel's assessments of the respective dimensions can be made by means of the variance components in an ANOVA model (see Winer, 1971, pp. 283—289). The ANOVA design is presented in Table 4.

This model assumes that the measurement errors (e) in the assessment (a) of one word (i) by an assessor (j) are uncorrelated. Consequently repeated assessments of the same word by the same or by comparable assessors (a_i) are assumed to remain constant, while e_{ij} is assumed to vary. If the systematic variance does not differ from the error variance, there is no evident correlation between the assessors' assessments of adjectives and verbs. A source of error that is often mentioned in connection with panel assessments is the influence of what is called the "halo" effect. Halo effects can be defined statistically as interaction effects between assessor and object of assessment (see Guilford, 1965, p. 299). Thus if there are any marked halo effects, this would lead to an increase of the variance that is calculated for the word × assessor interaction, which should reduce the size of the F value both for the word factor (W) and for the assessor factor (A).

The result of the analysis of variance design is presented in Tables 5—10.

Table 5. ANOVA for adjectives: Evaluation

Source of variation	df	MS	F	$\hat{\omega}^2$	f^2	f	g	r_I
W	569	16.99	26.43					
A	14	13.42	20.87	.012	.012	.110	.84	.952
WA	7966	.64						

Table 6. ANOVA for adjectives: Activity

Source of variation	df	MS	F	$\hat{\omega}^2$	f^2	f	g	r_I
W	569	9.11	9.86					
A	14	57.95	62.71	.059	.063	.250	>.99	.984
WA	7966	.92						

Table 7. ANOVA for adjectives: Potency

Source of variation	df	MS	F	$\hat{\omega}^2$	f^2	f	g	r_I
W	569	8.62	7.30					
A	14	69.74	59.02	.085	.093	.305	>.99	.983
WA	7966	1.18						

Table 8. ANOVA for verbs: Evaluation

Source of variation	df	MS	F	$\hat{\omega}^2$	f^2	f	g	r_I
W	882	9.33	17.96					
A	14	40.20	77.39	.034	.035	.187	>.99	.987
WA	12348	.53						

Table 9. ANOVA for verbs: Activity

Source of variation	df	MS	F	$\hat{\omega}^2$	f^2	f	g	r_I
W	882	11.50	13.26					
A	14	134.05	154.58	.057	.060	.246	>.99	.994
WA	12348	.87						

Table 10. ANOVA for verbs: Potency

Source of variation	df	MS	F	$\hat{\omega}^2$	f^2	f	g	r_I
W	882	5.87	6.54					
A	14	58.30	64.90	.047	.493	.222	>.99	.985
WA	12348	.89						

df Degrees of freedom
F F ratio computed
f^2 Effect size index indicates standard deviations for standardized means, when the independent variable is known
g Denotes the power in the results obtained
MS Mean square
r_I Intraclass correlation coefficient
$\hat{\omega}^2$ Random sample assessment of the proportional reduction in variance of the dependent variable given the independent

As can be seen from these results, the reliability of the assessments is at a high level.

Since we shall be making use of the mean assessment of the assessors and all adjectives and verbs respectively that define a particular noun, these values undoubtedly express reliable assessments. In addition these results imply that doubts expressed by individual assessors as to the reliability of the assessments and the possibility of being able to assess at all adjectives and verbs separated from their surrounding text are nothing other than subjective judgements lacking an objective foundation.

Another way of studying the agreement in the assessments made by the panel is to study their structure (see Guilford, 1954, pp. 253—254). This can be done by making a factor analysis or a reduced component analysis. The method assumes that the assessors' assessments are not defined by only one source of variation, but by several. This means that we can study the variance that different assessments, who are independent of each other, have in common. If the judges agree on their assessment (a particular position on the scale) of a word (adjective, verb) with regard to any of the three dimensions we are working with (E-A-P), this means that the assessment is based on the same underlying dimension. If, on the other hand, they do not agree with each other in their assessments, this may depend partly on the assessment being based on different dimensions, partly on their assigning different importance (weights) to the same dimension.

Thus the latent dimensions that influence our 15 assessors in the same way give rise to what are called common factors or components. The part of the common variance that the respective assessors contribute can be seen in the communality values. The part of the common variance that derives from a certain arrangement of assessments, on the other hand, can be seen from the correlation between the respective assessors and a particular known component. If the assessors' assessments correlate with only one dimension, we can state that they are of the same opinion. However, if there are two or more dimensions that are independent of each other in the material, this means that the assessors can be divided into different groups depending on how they load on each dimension. Such a result means that there are different opinions.

For the purpose of studying the assessments from the point of view of structure analysis, six component analyses were carried out. The correlations and components are presented in Tables 11 and 12. The pattern in the panel's assessments of the evaluation dimension of the adjectives shows that there is only one component. This result implies that the assessors are of the same opinion in their evaluation of adjectives. The assessment of the other two dimensions shows, however, that two components are needed to explain the relation pattern. This result can be interpreted as showing that the assessors are not as unanimous in their perception of activity and potency in the adjec-

tives as they were in their assessment of the evaluation dimension. For the activity dimension the varimax rotation shows that the assessors group themselves in three different clusters (loading $>.30$), namely cluster 1 (1, 2, 4, 7), cluster 2 (8, 9, 13) and cluster 3 (3, 5, 6, 10, 11, 12, 14, 15). Only the numbers and not names are given here, since we only wish to present the groupings and the size of the groups. No interpretation of their import was intended, nor is it possible, since the assessments were made anonymously. As can be seen from Table 13, however, the first component in the activity assessment of the adjectives is responsible for 86.08 % of the common variance extracted.

For the potency dimension the varimax rotation shows that here too the assessors can be divided into three clusters, namely cluster 1 (9, 10, 11, 13, 14), cluster 2 (1, 3, 6, 7, 12, 15) and cluster 3 (2, 4, 5, 8). A comparison with the clusters for the activity assessments shows that the composition of the individual clusters varies. The first component in the potency assessments is responsible for 63.82 % of the common variance extracted.

If the calculation of the reliability is based on weighted assessments, α_{max} for the respective summation variables proves to be (.965) for the evaluation, (.917) for the activity and (.877) for the potency.

The panel's assessments of the evaluation aspect of the verbs show that one component is sufficient to reproduce the relation pattern. The same applies to the assessment of the activity dimension of the verbs. This can be interpreted as indicating that the assessors are of the same opinion in their assessment of these two dimensions. There appear to be differences of opinion, however, in the assessment of the potency dimension of the verbs. The varimax rotation shows that three clusters can be distinguished, namely cluster 1 (1, 3, 4, 6, 7, 8, 12, 15), cluster 2 (9, 10, 13, 14) and cluster 3 (2, 5, 11). The first component in the potency assessments is responsible for 68.43 % of the common variance extracted.

The calculation of α_{max} shows for the assessments of the different dimensions of the verbs high reliability scores, namely (.951) for evaluation, (.931) for activity and (.859) for potency.

A comparison between the reliability scores based on intraclass correlations shows that they lead to a certain degree of overestimation, since the assessors are regarded as "identical measuring instruments". This overestimation arises through an unrealistic assumption on which the model is based. This results in the model being insensitive both to differences in the variation between different assessors on the panel and to differences in the reliability level between different assessors (see Jackson & Messick, 1967, p. 232).

If the original set of variables "assessors" is transformed into a new exact zero-correlated set of variables, i.e. to components that give a good approximation of the original set of data, the loadings can be used for weighting each

Table 11. Product-moment correlations for 15 assessors: Adjectives

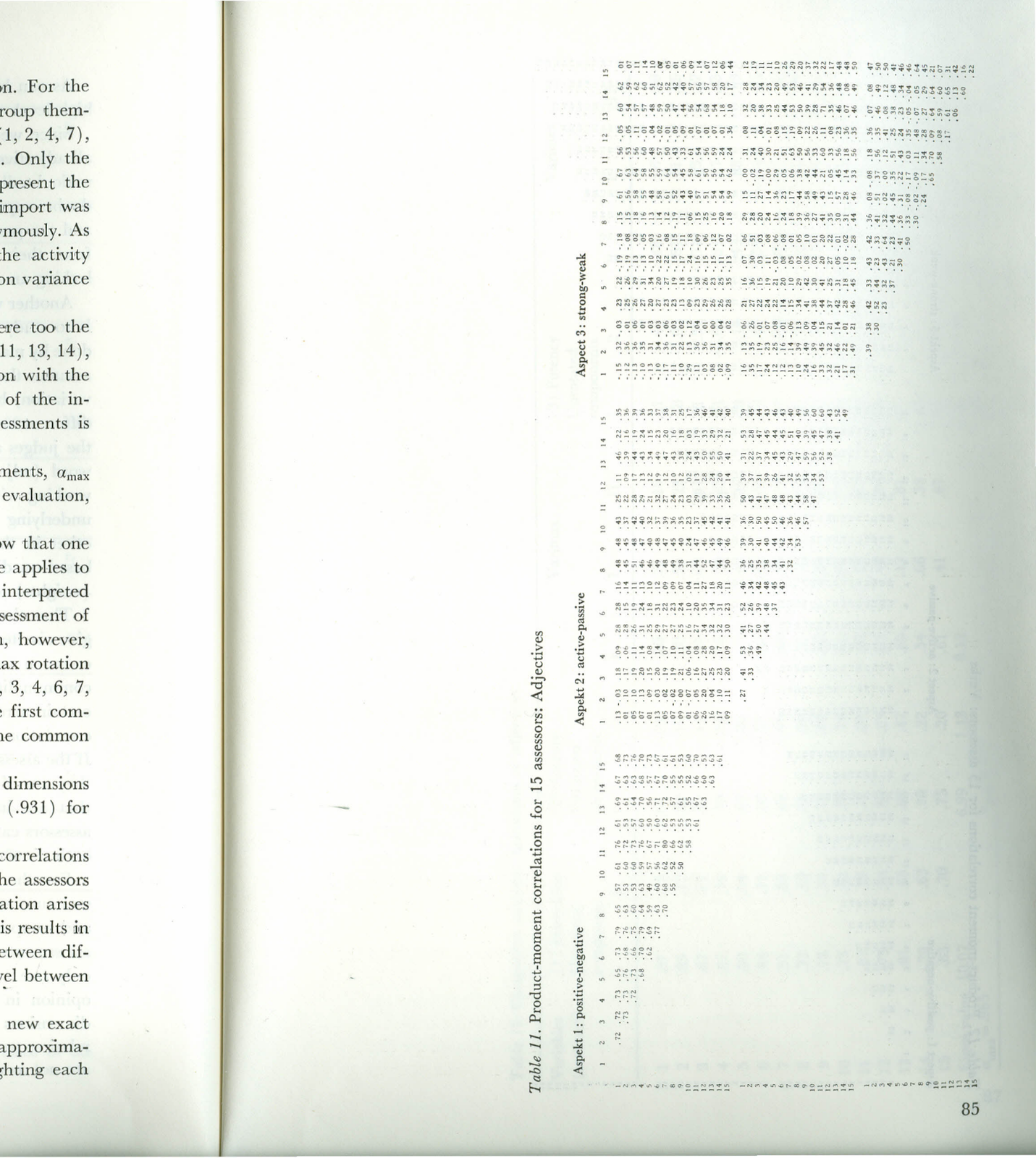


Table 12. Product-moment correlations for 15 assessors: Verbs

Aspect 1: positive-negative

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1		.62	.55	.65	.49	.64	.65	.60	.62	.50	.67	.58	.63	.60	.66
2			.53	.66	.59	.59	.65	.55	.56	.51	.67	.49	.54	.57	.68
3				.52	.53	.51	.54	.44	.51	.55	.55	.44	.57	.48	.55
4					.57	.68	.71	.56	.60	.48	.73	.57	.57	.56	.67
5						.51	.57	.46	.44	.54	.55	.38	.49	.43	.56
6							.62	.54	.62	.45	.70	.60	.57	.53	.60
7								.55	.62	.54	.72	.56	.59	.56	.63
8									.52	.39	.53	.49	.50	.48	.54
9										.45	.66	.56	.58	.56	.55
10											.48	.40	.47	.43	.47
11												.59	.62	.61	.63
12													.51	.54	.48
13														.56	.57
14															.54
15															

Aspect 2: active-passive

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	.12	.24	.25	.04	.24	.18	.06	.16	.19	.26	.17	.15	.26	.09	.19
2	.02	.21	.16	.02	.28	.09	.01	.01	.16	.19	.13	.10	.17	.08	.17
3	.03	.15	.18	.04	.25	.12	.04	.04	.15	.21	.10	.08	.19	.01	.12
4	-.03	.14	.11	-.03	.25	.07	-.01	.03	.10	.13	.13	.07	.15	.02	.09
5	-.02	.08	.06	-.03	.34	-.01	-.00	-.01	.07	.11	.08	.00	.12	-.00	.11
6	.02	.13	.15	.01	.23	.11	.02	.04	.11	.16	.14	.13	.17	.00	.12
7	.03	.15	.18	.02	.22	.09	.05	.03	.14	.16	.11	.08	.18	.06	.11
8	.13	.22	.22	.11	.28	.21	.11	.23	.22	.27	.19	.16	.23	.15	.22
9	.09	.19	.23	.07	.24	.18	.11	.10	.21	.25	.19	.22	.27	.12	.17
10	-.08	-.05	.02	-.08	.14	-.01	-.09	-.02	-.03	.07	-.04	-.05	.03	-.09	-.03
11	-.03	.11	.12	-.04	.25	.07	-.01	-.01	.09	.12	.15	.11	.18	.02	.10
12	.16	.23	.28	.12	.23	.24	.12	.13	.27	.29	.22	.30	.26	.10	.21
13	.15	.22	.26	.12	.28	.22	.13	.14	.24	.28	.21	.18	.40	.15	.24
14	.19	.28	.26	.15	.28	.24	.18	.17	.27	.28	.26	.25	.33	.18	.26
15	.09	.23	.19	.06	.32	.14	.06	.09	.21	.24	.17	.16	.21	.10	.21
1		.58	.58	.66	.24	.49	.68	.46	.62	.56	.51	.47	.56	.61	.60
2			.52	.50	.36	.46	.50	.39	.60	.56	.47	.43	.55	.50	.57
3				.53	.24	.44	.53	.36	.59	.60	.43	.45	.52	.50	.48
4					.22	.51	.65	.41	.54	.52	.53	.44	.49	.56	.53
5						.27	.25	.17	.35	.38	.34	.28	.35	.24	.35
6							.50	.42	.51	.47	.44	.48	.45	.41	.44
7								.38	.59	.56	.54	.44	.51	.56	.56
8									.41	.37	.33	.35	.37	.32	.36
9										.61	.53	.51	.54	.55	.58
10											.50	.45	.55	.48	.55
11												.50	.46	.47	.52
12													.41	.42	.44
13														.49	.55
14															.51
15															

Aspekt 3: strong-weak

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	-.03	.10	-.03	-.07	.13	-.23	-.08	-.01	.47	.43	.07	-.09	.42	.40	-.05
2	-.10	.09	-.09	-.08	.20	-.21	-.06	-.06	.46	.41	.09	-.13	.37	.37	-.04
3	-.01	.05	.03	-.06	.13	-.11	-.02	-.00	.40	.40	.08	-.01	.32	.27	-.03
4	-.09	-.02	-.08	-.10	.15	-.21	-.07	-.06	.45	.34	.08	-.13	.34	.31	-.12
5	-.12	-.03	-.10	-.10	.25	-.18	-.10	-.04	.36	.38	.09	-.10	.27	.24	-.08
6	-.09	-.04	-.08	-.15	.16	-.29	-.08	-.07	.41	.31	.08	-.14	.33	.34	-.12
7	-.08	-.03	-.09	-.13	.15	-.24	-.10	-.08	.49	.39	.06	-.13	.38	.37	-.12
8	-.02	.07	.00	-.01	.21	-.13	-.02	.08	.36	.34	.10	-.02	.35	.32	-.05
9	-.06	.03	-.08	-.09	.15	-.27	-.04	-.04	.51	.37	.12	-.14	.40	.34	-.07
10	-.16	-.08	-.13	-.15	.09	-.19	-.15	-.12	.36	.35	-.05	-.12	.23	.18	-.17
11	-.06	-.02	-.11	-.11	.18	-.23	-.07	-.10	.52	.35	.12	-.13	.39	.37	-.12
12	-.03	.04	-.04	-.06	.14	-.23	-.03	-.00	.43	.35	.11	-.07	.40	.39	-.10
13	-.01	.11	-.01	-.00	.17	-.19	-.05	.00	.51	.42	.14	-.07	.52	.40	.02
14	.02	.14	.00	-.01	.18	-.14	-.02	.03	.44	.37	.15	-.05	.44	.48	-.01
15	-.03	.12	.02	-.04	.21	-.20	.00	.00	.45	.45	.13	-.08	.40	.41	.01
1	.34	.40	.19	.22	.16	.06	.14	.19	.11	.27	.33	.16	.41	.36	.37
2	.31	.56	.29	.30	.27	.09	.25	.24	.22	.37	.34	.21	.51	.44	.40
3	.17	.31	.20	.11	.14	-.04	.10	.12	.21	.33	.20	.12	.40	.36	.21
4	.32	.35	.18	.27	.16	.08	.18	.21	.10	.18	.32	.20	.32	.29	.31
5	.26	.28	.17	.21	.40	.06	.17	.18	.29	.35	.36	.18	.40	.36	.21
6	.34	.42	.28	.28	.23	.17	.27	.29	.15	.18	.34	.26	.36	.28	.26
7	.30	.36	.17	.21	.18	.05	.23	.17	.10	.20	.32	.18	.34	.30	.30
8	.26	.36	.30	.20	.19	.19	.22	.33	.06	.13	.24	.25	.27	.18	.28
9	.31	.44	.23	.24	.25	.06	.22	.20	.26	.34	.39	.22	.48	.40	.30
10	.25	.36	.19	.20	.25	-.03	.18	.18	.30	.46	.31	.14	.49	.39	.26
11	.35	.31	.19	.27	.29	.10	.26	.23	.24	.25	.56	.21	.46	.37	.28
12	.38	.37	.28	.27	.26	.11	.33	.26	.15	.19	.46	.33	.37	.33	.25
13	.28	.41	.23	.21	.23	.03	.21	.20	.30	.36	.33	.14	.64	.45	.31
14	.20	.31	.13	.16	.13	.01	.11	.12	.12	.23	.28	.07	.35	.37	.28
15	.30	.45	.21	.30	.28	.06	.22	.25	.24	.33	.41	.18	.46	.39	.42
1		.43	.44	.48	.28	.38	.53	.38	.03	.11	.46	.42	.28	.12	.36
2			.48	.39	.26	.28	.44	.32	.13	.22	.37	.37	.39	.27	.45
3				.38	.22	.40	.52	.38	.00	.10	.30	.47	.17	.14	.33
4					.26	.34	.39	.37	.03	.15	.39	.38	.26	.15	.35
5						.14	.28	.25	.20	.25	.41	.25	.29	.20	.19
6							.39	.34	-.13	-.07	.22	.41	.02	-.06	.29
7								.38	.02	.10	.41	.48	.20	.11	.33
8									.00	.11	.30	.37	.17	.14	.31
9										.45	.20	-.05	.50	.40	.01
10											.25	.03	.49	.38	.14
11												.31	.39	.26	.35
12													.13	.03	.29
13														.53	.25
14															.18
15															

Table 13. Component analyses: Assessment of adjectives

Variables (assessors)	(1) Evaluation Unrotated component 1	h	(2) Activity Unrotated components 1	2	h	Varimax I	II	(3) Potency Unrotated components 1	2	h	Varimax I	II
1	.87	.75	.67	—.38	.60	.75	.19	.51	.46	.48	.06	.69
2	.85	.72	.52	—.17	.30	.50	.23	.78	—.05	.61	.61	.49
3	.87	.75	.72	—.12	.53	.60	.41	.51	.51	.51	.03	.72
4	.87	.76	.70	—.31	.58	.72	.25	.71	—.06	.51	.57	.43
5	.80	.64	.67	—.00	.45	.49	.46	.64	.14	.43	.38	.53
6	.84	.71	.68	—.13	.48	.58	.38	.38	.63	.53	—.14	.72
7	.91	.83	.64	—.41	.58	.75	.15	.50	.66	.68	—.07	.82
8	.77	.59	.64	.30	.51	.26	.66	.61	.17	.41	.34	.54
9	.73	.53	.72	.40	.68	.24	.79	.64	—.57	.74	.86	.01
10	.73	.53	.73	.29	.61	.33	.71	.51	—.63	.66	.81	—.12
11	.87	.76	.77	.05	.60	.52	.57	.74	.43	.73	.83	—.18
12	.73	.54	.60	—.07	.37	.49	.36	.48	.41	.39	.08	.62
13	.81	.65	.68	.47	.69	.17	.81	.59	—.54	.65	.80	—.00
14	.79	.63	.70	.22	.54	.66	.33	.66	.49	.68	.82	.09
15	.83	.70	.75	.20	.61	.41	.67	.68	.43	.65	.21	.78
Eigen value	10.07		6.99	1.13	8.12			5.54	3.14	8.68		
	$\alpha_{max}=.965$		$\alpha_{max}=.917$					$\alpha_{max}=.877$				

Table 14. Component analyses: Assessment of verbs

Variables (assessors)	(1) Evaluation Unrotated component		(2) Activity Unrotated component		(3) Potency Unrotated components		Varimax		
	1	h	1	h	1	2	h	I	II
1	.83	.68	.82	.67	.72	—.18	.55	.73	.12
2	.80	.65	.75	.56	.70	.04	.49	.63	.32
3	.74	.54	.73	.54	.67	—.25	.51	.71	.05
4	.83	.69	.76	.58	.66	—.13	.45	.65	.15
5	.70	.49	.44	.20	.57	—.15	.34	.46	.36
6	.80	.64	.68	.46	.51	—.46	.48	.66	—.22
7	.83	.69	.78	.61	.70	—.24	.55	.74	.07
8	.70	.49	.56	.31	.59	—.19	.38	.61	.06
9	.77	.59	.81	.65	.22	.74	.59	—.10	.76
10	.66	.43	.77	.59	.34	—.64	.53	.06	.73
11	.85	.72	.71	.50	.67	.14	.46	.55	.40
12	.71	.51	.66	.43	.62	—.34	.50	.70	—.06
13	.76	.58	.73	.54	.53	.63	.68	.23	.79
14	.73	.54	.72	.52	.37	.62	.53	.09	.72
15	.80	.64	.76	.57	.58	—.08	.35	.57	.17
Eigen value	8.88		7.72		5.05	2.33	7.38		
	$\alpha_{max} = .951$		$\alpha_{max} = .931$		$\alpha_{max} = .859$				

individual assessor's assessment in agreement with systematic variance that is explained by the first component. This is responsible for a maximum of the variance from the original variables.

In order to create a weighted summation variable, the weights are used from the first unrotated component. Each assessment is multiplied by the weight for the respective assessor and dimension. Then the sums of these products were formed. In order to create weighted means for each adjective and verb, the totals have been divided by the sum of the weights. If any assessment has been dropped, the corresponding weight has been subtracted when the sum of the weights was formed.

In order to study the connection between unweighted and weighted means for the respective summation variables, a correlation analysis was carried out in Tables 15 and 16.

As the results show, the correlations between the means of the weighted and unweighted summation variables are exceptionally high.

These correlations imply that there are only very small fluctuations resulting from the different frames of reference of the assessors, i.e. in the assessments that have not been adjusted for each assessor's individual contribution to the systematic variance.

Table 15. Correlation of means for adjectives

	Unweighted aspect			Weighted aspect		
	1	2	3	1	2	3
1 Evaluation		.439	.414	.999	.464	.507
2 Activity			.586	.435	.993	.618
3 Potency				.416	.581	.988
1 Evaluation					.461	.509
2 Activity						.619
3 Potency						

All scores are significant with $\alpha = .001$

Table 16. Correlations of means for verbs

	Unweighted aspect			Weighted aspect		
	1	2	3	1	2	3
1 Evaluation		.238	.130	.998	.245	.085
2 Activity			.615	.229	.998	.607
3 Potency				.130	.618	.994
1 Evaluation					.237	.085
2 Activity						.609
3 Potency						

All scores are significant with $\alpha = .001$

5. Design of search logics

According to the model presented in Figure 2 (p. 23), perception means a *representative* sampling of data which undergo continuous grouping and classification processes. This is a necessary condition if we are to be able to use our language to communicate our experiences. The dependencies that exist between nouns and adjectives, and the relations that exist between nouns and verbs are assumed to reflect the relations that connect phenomena with one another. The usual way of stating a relation is to specify a rule that says what are to be regarded as elements, pairs etc. If, for example, we wish to state for our sample space (S) that A is the set of all nouns and B is the set of all adjectives modifying A, this relation can be stated more formally in the following way:
$S = \{(a, b) \in A \times B \,||\, a \text{ modified by } b\}$.
In light of Figure 2, we assume that the relations that exist between *mnemonic* and reference express relations between form and empirical content. We assume further that a noun functions as a form that gets an empirical content through the adjective and/or verb connected to it. Quantitative empirical relations can, in other words, be established by the use of scaled adjectives. If perceived similarities or covariations between different properties are defined, we can carry out multivariate analysis for the purpose of determining the position of a certain property in a number of latent dimensions.

Control and systematic variation are the strategies that make it possible to find relations and define them, so that they can become functional relations. A relation is called a functional relation or a function if every element in a domain (set of sets) is paired with only and exclusively one element of the range, i.e. an adjective with a weighted mean.

5.1 Statement of search questions

The discovery of constancy between widely different phenomena, or in other words laws, is often considered to be one of the fundamental scientific goals. It is this hope that accounts for the work of so many behavioural scientists being focussed on discovering ("objectifying") in the concrete object of

investigation general principles that govern all conceivable systems. These researchers claim that the only a priori requirement of science is *objectivity*.

If we are to successfully overcome the difficulties with such complex problems as those dealt with in behavioural science research, a large measure of openness and willingness to use new research strategies is required. It should, however, be possible to define, specify and make these strategies explicit. This requires a purposeful communication behaviour in the researcher and a purposeful handling of plans of investigation, models of analysis, statistical methods and techniques, measuring instruments and technical aids, such as computers.

Our fundamental assumption is that every step in the process of problem perception, structurization and definition is governed by three essential prerequisites, namely the researcher's (1) motivation, (2) idiosyncratic strategies of behaviour and (3) structures of organization (environment). Consequently these components have been given a central role (see B. Bierschenk, 1974, pp. 4—27). Thus each individual researcher is steered by different motives in this process and each individual develops his own specific strategies of behaviour as a result of his perception of the problems and his search for information on problem structures. He uses different methods and means to realize his strategy of problem formulation within the constraints defined by the structure of the research organization. The researcher is a component in this organization, which means that there are different reference systems influencing him and the other persons associated with the organizations in question. There are probably people within the reference system who function as promotors of certain ideas. But this type of influence must be highly dependent on how their supporters perceive the problems and assess their relevance. Starting from the model that has guided the collection of data (see B. Bierschenk, 1974, p. 13), we hope eventually to be able to answer questions concerning: Motivation, perception, selection of problem, choice of research methods, the importance of the frame of reference and the organizational structure of the system.

5.2 Formulation of hypotheses

The first measure taken in constructing a dictionary has been to use all the linguistic elements grouped under the codes in Figure 3 (p. 40) to build up files. These will eventually be replaced by *structured dictionaries*. The fact that many different search logics could be developed, depending on which search questions are stated and which hypotheses formulated, arises from our having no formalized theory. At the same time this means that our decisions must of necessity be arbitrary and consequently require verification. By means of logically meaningful connections, we intend to extract information step-wise from the interview material. Thus we must be able to formulate and test a

number of hypotheses before we obtain material suited to *statistical analysis.* This method of approach can be illustrated by an example.

Hypothesis 1
Linguistic elements in codes 30, 50, 60, 70 and 80 only get their empirical content when they are linked to linguistic elements in codes 32, 52, 62, 72 and 82 respectively.

Hypothesis 1 is to be tested with the element "researcher". Working from our psychological process model, we assume that a particular person knows that "researcher" refers to a certain category of objects denoting persons. This means that this noun has got its empirical content from innumerable experiences that the interviewees have had. Our individual wishes, however, to communicate a particular message, which means that the range of the variation in the information, i.e. in the listener's anticipated interpretation possibilities, must be limited. Since the interviewee makes use of different modifiers, a *zooming-in* process occurs, i.e. different modifiers function in the same way as a mobile lens in a TV camera. If by means of Boolean algebra, volition (code 23) or condition (code 15) are connected to noun (codes 30, 50, 60, 70, 80), copula (code 41) and adjective (codes 32, 52, 62, 72, 82), we obtain an evaluation of researchers as shown in Table 17.

The scores presented in Table 17 are only meant to illustrate how quantitatively defined concepts are built up. Cliff (1969, p. 158) considers that adjective-noun combinations usually have the properties of both adjective and noun (but cf. van der Kloot's experiment, p. 90), and that consequently the combination rule ought to be some form of addition. The same rule should apply to a combination of adjectives.

The purpose of the example presented in Table 17 has been to show that

Table 17. Example of a logic clause and its outcome

Content	Evaluation mean	Activity mean	Potency mean
Volition clause (code 23)			
established & researcher	4.43	3.56	4.51
responsible & researcher	4.63	4.80	4.67
Condition clause (code 15)			
inexperienced & researcher	2.75	3.40	2.93
must be orientated & researcher	5.01	4.62	4.31
must be open & researcher	5.50	4.77	4.71

&: logical *and*

even a few simple connections produce meaningful and interesting results. But only through more complex statistical analyses can we discover what latent structures there are in the material.

The hypothesis of Oller & Sales (1969, p. 229) is that in a given context modifiers are arranged in accordance with the limiting effect they have. This hypothesis has been verified experimentally. The same assumption has formed the foundation for the development of ANACONDA, although this was not formulated equally explicitly from the start.

Hypothesis 2
Limiting modifiers group themselves concentrically around linguistic elements. The most limiting modifier is to be found in the periphery. Thus each new modifier creates a new division. Cliff's (1969, pp. 143—160) study shows that certain verbs in combination with adjectives function multiplicatively, i.e. adverbs of degree have the function of multipliers for the adjectives they modify. Cliff (1969, pp. 157—158) writes:

"... adverbs and adjectives of specifiable types combine according to a multiplicative rule. /. . ./ In a very real sense 'extremely good' may be said to be about one-and-a-half times as good as 'good'."

The fact that adverbs can modify adjectives suggests that adjectives should be treated in the same way as verbs, i.e. modifiers such as *manner* and *degree* should be used to differentiate shades of meaning in the adjectives. Thus in order to build up a system for system analysis that can handle differentiated content in a text requires that modalities can be specified and that suitable combination rules can be developed.

Different linguistic elements form the building blocks of a concept, irrespective of whether it is dependent or independent. In this type of analysis a word that has earlier been regarded as an adjective or verb with a varying lexical meaning is re-defined. Adjectives and adverbs become modifiers and verbs state the implication of a class of events, i.e. they define the context of a clause. Since they indicate modifications and/or events with regard to the agent and/or object, they have a temporary nature, i.e. they form an intermediate stage in the building up of a concept. This is the way in which we build up the concepts that are to form the basis for a statistical analysis of researchers' cognitive and emotional structure, which is assumed to steer the perception and evaluation of the initial phase of the research process.

6. Data processing

The example presented in Table 17 aims at stating what type of values will form the basis for a set of data matrices. In an empirical study of relations between linguistic elements or between concepts, methods of bi-variate and multi-variate relation analyses could be used. The most direct of the bi-variate methods that has been used is *subjective scaling*. This means in a linguistic context that the meaning in a linguistic element is assessed and given a score in accordance with this assessment. This method was applied by Messick (1969, pp. 161—167) in a study of certain metrical properties (mainly the equidistance of the intervals) in semantic differentials. But Cliff (1969, pp. 143—160) also made use of this method in scaling adverbs of degree.

Another and perhaps the most well-known method in the bi-variate tradition is the association method. It is based on the theory of association and assumes that the similarity between two linguistic elements can be expressed as a relation between intersection and union of the distribution of these two elements. The technique has been used by e.g. Deese (1965) for the purpose of building up an "associative dictionary". Word associations can admittedly be studied by means of such a method, but since the method is sensitive to syntactic and phonetic associations, it is difficult to interpret the result. Miller (1967, p. 54) writes:

"Attempts have been made to classify associates as either syntagmatic or paradigmatic, but the results have been equivocal, e.g. if *storm* elicits *cloud* or *flower* elicits *garden,* is the response to be attributed to paradigmatic semantic similarity or to a familiar sequential construction?"

A third method is a combination of subjective scaling and association. This has become known as Semantic Differentials. The underlying theory is the theory of association. A closely related theory has become known as sentence supplementation. This method is a semantic test and is based on the assumption that the individual can replace words in a particular given context or that all contexts that fit a given word can be stated. This phenomenon is sometimes also called "privilege of occurrence". Miller (1967, p. 54) writes:

"In terms of a theory of semantic markers, some such relation would be expected, since the semantic features of words in any meaningful sentence are interdependent."

This technique has been used by e.g. Oller & Sales (1969) for the purpose of studying "conceptual restrictions" in English.

7. Data analysis and inference

A development of the methods mentioned above has led to what is known as multidimensional scaling (MDS). Like the factor analysis or component analysis model, this is based on a linear space model. Thus the basis of both models consists of a geometric representation of a metric space in which the measuring object is represented as points on coordinates on k orthogonal dimensions. From a formal point of view, therefore, both models can be considered comparable. The above mentioned analysis by Rosenberg et al. (1968) shows that as far as the scaling of adjectives is concerned, they lead to the same result.

Our purpose is to study the dimensionality of the interview material. This requires that we choose a model for assessment of semantic distance or in other words a metric space of low dimensionality. Distance can be related to similarities, which means that by measuring the distance between different concepts we can come to some conclusion about the relations between the concepts. Another argument for the choice of this method is that assessments are easy to make and thus suitable for a large amount of empirical material, while MDS is not.

It has already been stated repeatedly that multivariate analysis techniques seem to provide the best answer to our intention of describing the structure of the interview material (or random sample of persons) as economically as possible. The assessment scores forming the basis of a description of the interview material can be arranged in accordance with the following general scheme of covariation:

K: Scales 1(1) 3; *V:* Variables 1(1) m; *P:* Persons 1(1) 40.

Working with the concepts that occur in the text from a particular interviewee, our next step in the development of a structured dictionary will be a number of cluster analyses. But we have also planned to study the relation pattern by means of factor analyses and eventually a multivariate discriminant analysis. If, for example, we start with a complete AaO relation, we can define "agents" as measuring objects and "objects" as variables. These together with

the scales form three groups for assessment. In this way we would be able to carry out forty different discriminant analyses, one for each person interviewed. If the data matrices are collapsed in the context of different criteria, we can then study the material from different angles. We could, for example, make analyses of the common agents and objects, in which either the interviewees or concepts are the measuring object and variables respectively. By means of this technique we can study the linear combinations that must be formed in order for the variance between e.g. the scales to be maximized in relation to the intragroup variance. By using a multiple discriminant analysis we can, as Abelson (1960, p. 171) writes:

"... distinguish in each given case the *objects* of discrimination, the *agents* of discrimination, and the *modes* of discrimination".

A discriminant analysis in which we investigate the importance of agents, objects and scales for a particular individual means that we must study the co-variances between the scales. The interaction between agents and objects functions as a basis for the assessment of the error variance. In this case our hypothesis is that the differences between the objects are stable with regard to the agents. The interaction between the objects and the scales, on the other hand, forms the basis for the identification of the structure in the discriminating functions, i.e. the factor structures. Depending on how the model is defined, therefore, different classes are homogenized (persons, objects or scales).

8. Construction of theories and models: A recapitulation

Both manual and computer-based content analyses presume that the researcher can formulate problems and define his measuring objects. This means that the development of our method for a computer-based content analysis, called ANACONDA, has a dual background: a set of concrete material, namely the interview text, and a psycholinguistic model. This states a number of functional limitations. We try to pay consideration to experience and to develop a number of inference rules that we assume to be necessary to explain human behaviour. The psychological model on which ANACONDA is based makes the assumption that every utterance is based on concepts that form the basis for the key-words in a clause. Further it is assumed that a clause does not come into being as a result of words being combined at random, but that it is the result of active organizing principles. These assumptions have resulted in ANACONDA being based on only two types of concept, namely dependence and independence and only two role functions.

We have shown the way in which a model containing symbolic representations of concepts and relations can be used on an empirical material. In our presentation two different kinds of dependence and independence are distinguished: vertical, i.e. between dependent and independent concepts, and on a syntactically horizontal level, i.e. between several independent concepts. We also discuss dependent and independent relations, i.e. the relation of clauses to each other. The first kind of dependence refers to Schank's (1973) Rules 1—5. Independent concepts are those that are main words in a complex, consisting of e.g. attributive qualifiers to this concept. The main concept has a code number ending in 0, while the last figure in dependent concepts is 1, 2, 3, . . ., 6. Rules 6—9 concern so-called conceptual cases and refer to a horizontal dependence, insofar as one takes the meaning of the verb to specify how many independent concepts there must be on the sentential level. These case-relations, *objective, recipient, instrumental* and *directive* appear in Figure 3 (p. 40).

We also take the theory of necessary parts into account in another way. Since language is economical in relation to the thoughts behind the utterance,

not all necessary parts are included in a speaker-listener situation. The necessary concept apparatus exists in the listener and a syntactically incomplete sentence is understood all the same. But the computer does not have this understanding, so therefore we supplement in the parts that are missing. It is our task to code complete complexes and complete AaO paradigms. Then when the verb has a built-in object, this concept is supplemented in. In the paradigm, on the other hand, no instrument is included as being necessary if it does not exist explicitly. One essential difference between Schank's theory and our practical coding in connection with it concerns how far we should go in the representation of conceptual rules (thought structures). For Schank the causal "instrument" is a necessary part of the verb's meaning, so that e.g. the verb *eat* must mean roughly "with cutlery" as instrument. As has been said earlier, we code no instruments that have not been named in the text. But Schank also says himself that there are certain concepts that are so well-known to the listener that it is of no interest to specify the instrument. We do not think of it consciously. For the same reason, Rule 10 is irrelevant.

Schank's rules 11 (a, b)-14 concern what he calls relations, i.e. the relation of clauses to each other within a sentence. It is possible for us to express relations by means of our system of numbers in combination with overall coding of a so-called clause theme.

Here too a difference is reflected in the way of representing the verb. Schank symbolizes explicitly the result built into certain verbs, e.g. *kill*, which he calls pseudo-state verbs. We stop at coding the verb and do not state any possible result. This means that rule 11b is not relevant to our work.

Relations emerge through a clause marker. In addition clause dependence is coded with codes for cause, intention, etc. A loop system makes it possible to differentiate main clause and subsidiary clause or which of the clauses is prior to the other one in time. This is what Schank calls causality. There are also other conceptual relations, namely time and place relations. Time as a single concept or as a qualifier in the form of a subordinate clause is looked upon as a modifier to a whole clause and therefore there is no causality. We have thought that the most practical way of stating modification of a whole clause is to give these concepts a dependency code to the verb, since it is the verb that is the most essential part of the clause. The fact that there is no causality is stated by the clause not being given a clause theme code like the others. Rules 12—14 are therefore reflected in our code system.

The content analysis method that we intend to develop should be able to approximate the interviewees', i.e. researchers', implicit models that are assumed to steer the perception and evaluation of the initial phase of the research process. As a first measure in building up concepts with an empirical root, all adjectives and verbs have been scaled by means of semantic differentials. It is namely to some extent these linguistic elements that form the

building blocks for a concept. The programme flow-chart described (Fig 3, p. 40) shows the way in which concepts could be built up.

The scaling of adjectives and verbs is based on certain assumptions and experimental results. These claim that it is the dependent concepts and not the independent ones that form the basis for conceptualization and that this can be described in main by three dimensions: (1) evaluation, (2) activity, and (3) potency. The scaling has been done in the form of panel assessments. In order to achieve maximal reliability in the assessments, it was decided that all the researchers in our population who had not participated in the interview study should be included in the panel. Out of 20 persons, 15 finally completed the desired assessments. The results of this panel assessment show high reliability scores ($\alpha_{max} = .859—.965$).

The relations that are assumed to exist between concepts are implicative or inferential and they are intended to be operationalized by means of analysis models based on perceived covariations or correlations. In order to be able to study the dimensionality of the interview material, in the next phase we shall apply a number of different models of analysis by means of which we can estimate semantic distance or represent content as a metric space of low dimensionality. Since distance can be related to similarities, we hope to be able to say something about latent structures and build structured dictionaries.

The development of complex programmes or programme packages is an experimental activity, since only carefully planned programmes, supplied with complete descriptions can be expected to produce both the desired results and indications of what is right or wrong in our method of approach.

Thus the essential factor in this work is that we can develop preliminary versions of individual system components and that by an interactive process we can improve these continuously. For this reason individual components (sub-programmes) in the system are subjected to constant revision and reformulation. The development of a method for computer-based content analyses and the construction of suitable systems is a long-term goal. Therefore this method refers only to the development of a system that is adapted to our particular interview material. Ultimately, however, our goal is to develop a more general system.

9. References

Abelson, R. P. Scales derived by consideration of variance components in multi-way tables. In: Gulliksen, H. & Messick, S. J. (Eds.) *Psychological scaling: Theory and applications.* New York: Wiley, 1960. Pp. 169—181.

Abelson, R. P. Psychological implication. In: Abelson, R. P., Aronson, E., McGuire, W. J., Newcomb, T. M., Rosenberg, M. J. & Tannenbaum, P. H. (Eds.) *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally, 1968. Pp. 112—139.

Abelson, R. P. The structure of belief systems. In: Schank, R. C. & Colby, K. M. *Computer models of thought and language.* San Francisco: Freeman, 1973. Pp. 287—339.

Asch, S. E. Forming impressions of personality. *Journal of Abnormal and Social Psychology,* 1946, *41,* 258—290.

Berg, M. Reliabilitetsprövning av en metod för innehållsanalys av intervjutext. /Reliability testing of a method of content analysis applied to interview texts./ *Testkonstruktion och testdata,* No 26, 1974.

Bertalanffy, L. *Organismic psychology and systems theory.* Barre, Mass.: Clark University Press, 1968.

Bierschenk, B. Att mäta subjekt—objekt-relationer i externt förmedlade självkonfrontationsprocesser via intern television: Presentation av ett kategorisystem. /The measurement of subject—object relationships in externally mediated self-confrontation processes via closed circuit television. Presentation of a category system./ *Testkonstruktion och testdata,* No 6, 1972.

Bierschenk, B. Perception, strukturering och precisering av pedagogiska och psykologiska forskningsproblem på pedagogiska institutioner i Sverige. /The perception, structuring and definition of educational and psychological research problems at the Departments of Education in Sweden./ *Pedagogisk-psykologiska problem,* No 254, 1974.

Bierschenk, I. Konstruktion av ett regelsystem för en datorbaserad innehållsanalys av intervjutext: Preliminärmanual och några utprövningsresultat. /Construction of rules for a computer-based content analysis of interview texts: A preliminary manual and some evaluation data./ *Testkonstruktion och testdata,* No 25, 1974.

Boulding, K. E. *The image.* Ann Arbor: The University of Michigan Press, 1956.

Cattell, R. B. *Personality and mood by questionnaire.* San Francisco: Jossey-Bass Publishers, 1973.

Chomsky, N. *Syntactic structures.* The Hague: Mouton, 1957.

Cliff, N. Adverbs as multipliers. In: Snider, J. G. & Osgood, C. E. *Semantic differential technique.* Chicago: Aldine, 1969. Pp. 143—160.

Cronbach, L. J., Gleser, G. C., Nanda, H. & Rajaratnam, N. *The dependability for scores and profiles.* New York: Wiley, 1972.

Deese, J. *The structure of association in language and thought.* Baltimore: The John Hopkins Press, 1965.

Gibson, J. J. A theory of direct visual perception. In: Royce, J. R. & Rozeboom, Wm. W.

(Eds.) *The psychology of knowing.* New York: Gordon & Breach, 1972. Pp. 215—250.
Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1954.
Guilford, J. P. *Fundamental statistics in psychology and education.* New York: McGraw-Hill, 1965.
Hunt, E. The memory we must have. In: Schank, R. C. & Colby, K. M. *Computer models of thought and language.* San Francisco: Freeman, 1973. Pp. 343—371.
Jackson, D. N. & Messick, S. (Eds.) *Problems in human assessment.* New York: McGraw-Hill, 1967.
Joe, G. W. & Woodward, J. A. An approximate confidence interval for maximum coefficient alpha. *Journal of Multivariate Behavioral Research,* 1975, *10* (1), 93—98.
van der Kloot, W. A. *A cognitive structure approach to person perception.* Leiden: University of Leiden, 1975. (Mimeographed doctoral dissertation.)
Krippendorff, K. Models of messages: Three proto-types. In: Gerbner, G., Holsti, O. R., Krippendorff, K., Paisley, W. J. & Stone, P. J. *The analysis of communication content. Developments in scientific theories and computer techniques.* New York: Wiley, 1969. Pp. 69—106.
Kuusinen, J. Affective and denotative structures of personality ratings. *Journal of Personality and Social Psychology,* 1969, *12* (3), 181—188.
Lord, F. M. Some relations between Guttman's principal components of scale analysis and other psychometric theory. *Psychometrika,* 1958, *23* (4), 291—296.
Luria, A. R. *The mind of a mnemonist.* London: Jonathan Cape, 1969.
Messick, S. J. Metric properties of the semantic differential. In: Snider, J. G. & Osgood, C. E. *Semantic differential technique.* Chicago: Aldine, 1969. Pp. 161—167.
Miller, G. A. Empirical methods in the study of semantics. In: Arm, D. L. (Ed.) *Journeys in science: Small steps—great strides.* Albuquerane: University of New Mexico Press, 1967. Pp. 51—73.
Miller, G. A., Galanter, E. & Pribram, K. H. *Plans and the structure of behavior.* New York: Holt, 1960.
Miron, M. S. What is it that is being differentiated by the semantic differential. *Journal of Personality and Social Psychology.* 1969, *12* (13), 189—193.
Monod, J. *Slump och nödvändighet. /Le hasard et la nécessité./* Stockholm: Aldus, 1972.
Oller, J. W. & Sales, B. D. Conceptual restrictions on English: A psycholinguistic study. *Lingua,* 1969, *23,* 209—232.
Osgood, C. E., Saporta, S. & Nunnally, J. C. Evaluative assertion analysis. *Litera,* 1956, *3,* 47—102.
Osgood, C. E. On the whys and wherefores of E, P, and A. *Journal of Personality and Social Psychology,* 1969, *12* (3), 194—199.
Piaget, J. *The origins of intelligence in children.* New York: Norton, 1963.
Piaget, J. *On the development of memory and identity.* Barre, Mass.: Clark University Press, 1968.
Piaget, J. *Genetic epistemology.* New York: Columbia University Press, 1970.
Premack, D. Language: On the difference between training a chimp to press a lever and telling it to press the lever. (Invited address.) Washington: American Psychological Association Meetings, 1969.
Premack, D. On the assessment of language competence in the chimpanze. In: Schrier, A. M. & Sollnitz, F. (Eds.) *Behavior of nonhuman primates, Vol. 4.* New York: Academic Press, 1971.
Pribram, K. Neurological notes on knowing. In: Royce, J. R. & Rozeboom, Wm. W. (Eds.) *The psychology of knowing.* New York: Gordon and Breach, 1972. Pp. 449—480.

Quine, W. V. Epistemology naturalized. In: Royce, J. R. & Rozeboom, Wm. W. (Eds.) *The psychology of knowing.* New York: Gordon and Breach, 1972. Pp. 9—23.
Reid, L. S. Toward a grammar of the image. *Psychological Bulletin,* 1974, *81* (6), 319—334.
Rosenberg, S., Nelson, C. & Vivekananthan, P. S. A multidimensional approach to the structure of personality impression. *Journal of Personality and Social Psychology,* 1968, *9* (4), 283—294.
Ross, J. R. Adjectives as noun phrases. In: Reibel, D. A. & Schane, S. A. (Eds.) *Modern studies in English.* Englewood Cliffs, N.J.: Prentice-Hall, 1969. Pp. 352—360.
Rozeboom, Wm. W. Problems in the psycho-philosophy of knowledge. In: Royce, J. R. & Rozeboom, Wm. W. (Eds.) *The psychology of knowing.* New York: Gordon & Breach, 1972. Pp. 25—109.
Salton, G. (Ed.) *The SMART retrieval system.* Englewood Cliffs, N.J.: Prentice-Hall, 1971.
Schank, R. C. Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology,* 1972, *3* (4), 552—631.
Schank, R. C. Identification of conceptualizations underlying natural language. In: Schank, R. C. & Colby, K. M. *Computer models of thought and language.* San Francisco: Freeman, 1973. Pp. 187—247.
Simmons, R. F. Semantic networks: Their computation and use for understanding English sentences. In: Schank, R. C. & Colby, K. M. *Computer models of thought and language.* San Francisco: Freeman, 1973. Pp. 63—113.
Skinner, B. F. *Verbal learning.* New York: Appleton—Century—Crofts, 1957.
Spiegelman, M., Terwilliger, C. & Fearing, F. The reliability of agreement in content analysis. *Journal of Social Psychology,* 1953, *37,* 189—203.
Uttal, W. R. *The psychobiology of sensory coding.* New York: Harper & Row, 1973.
Waterman, D. A. & Newell, A. Protocol analysis as a task for artificial intelligence. *Artificial Intelligence,* 1971, *2,* 285—318.
Watson, J. D. *The double helix.* New York: The New American Library, 1968.
Wearing, A. J. Remembering complex sentences. *Quarterly Journal of Experimental Psychology.* 1972, *24,* 77—86.
Wiener, N. *Cybernetics or control and communication in the animal and the machine.* New York: The M.I.T. Press, 1948.
Wilson, C. *Nya frågeställningar inom psykologin. Maslow och den post-freudianska revolutionen.* /New pathways in psychology. Maslow and the post-Freudian revolution./ Staffanstorp: Cavefors, 1974.
Winer, B. J. *Statistical principles in experimental design.* (2nd ed.) New York: McGraw-Hill, 1971.
Wishner, J. Reanalysis of "Impressions of personality". *Psychological Review,* 1960, *67,* 96—112.
Wundt, W. *Grundriss der Psychologie* (13. Aufl.) Leipzig: Alfred Kröner Verlag, 1918.