DOCUMENT RESUME

ED 352 405 TM 019 324

AUTHOR	Helmersson, Helge
TITLE	Main Principles for Perspective Text Analysis via the PC-system PERTEX. No. 41.
INSTITUTION	Lund Univ. (Sweden). Cognitive Science Research.
REPORT NO	ISSN-0281-9864
PUB DATE	92
NOTE	57p.; Paper presented at the International Congress of Psychology (25th, Brussels, Belgium, July 19-24, 1992).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Cluster Analysis; Coding; *Computer Software; Matrices; *Microcomputers; Reader Text Relationship; *Structural Analysis (Linguistics); *Text Structure
IDENTIFIERS	*Perspective Text Analysis; *PERTEX Computer Program; Text Organization

ABSTRACT

This document describes how the main principles of Perspective Text Analys.s are implemented in the PC-system PERTEX, concentrating on the main steps of the analysis. The analysis starts with normal text and ends in a topological representation of the mentality that the text presents. The text material is processed in the following main steps: (1) coding of function words by means of a special dictionary; (2) design and coding of blocks according to the AaO (Agent-verb-Objective) paradigm; (3) supplementation of A- and O-dummies; (4) generation of A/O matrices; (5) cluster analysis based on generated matrices; and (6) topological presentation cf outcomes. PERTEX gives an integration of all the steps in the analysis, and the user is offered numerous comprehensive functions for automatic coding and control of syntax. By a multilingual design, PERTEX can operate on texts in different languages. The user can select different menu-languages for the interaction with PERTEX. The technical output of the system is illustrated in the appendix with the complete 17-page printout from analysis of a classic text. (Author/SLD)

Main Principles for Perspective Text Analysis via the PC-system PERTEX

Helge Helmersson

1992

No. 41

KOGNITIONSVETENSKAPLIG FORSKNING

Cognitive Science Research

Lund University
Sweden

Main Principles for
Perspective Text Analysis
via
the PC-system PERTEX

Helge Helmersson

1992 **No. 41**

Cognitive Science Research
Lund University
Sweden

Postal address

Cognitive Science Research
P.O. Box 7080
S-220 07 LUND
SWEDEN

Editor

Bernhard Bierschenk
Department of Psychology
Lund University

Adm. editor

Helge Helmersson
Department of Business Adm.
Lund University

Abstract

The principles for Perspective Text Analysis has been implemented into the PC-system PERTEX. The analysis starts from normal text and ends up in a topological representation of the mentality the text presents. The text material is processed by the following main steps: (1) Coding of function words by means of a special dictionary, (2) design and coding of blocks according to the AaO-paradigm, (3) supplementation of A- and O-dummies, (4) generation of A/O-matrixes, (5) cluster analysis based on generated matrixes, (6) topological presentation of outcomes. PERTEX gives an integration of all the steps in the analysis and the user is offered a lot of comprehensive functions for automatic coding and control of syntax. By a multilingual design PERTEX can operate on texts in different languages. The user can select different menu-languages for the interaction with PERTEX.

This paper describes how the main principles of Perspective Text Analysis (PTA), Bierschenk & Bierschenk (1986 a, b, c) are implemented in the PC-system PERTEX. The description is concentrated on the main steps in the analysis without technical details of system design and programming. Each step in the analysis is first described i general terms and then partly illustrated by small examples. The complete printer output from the analysis of a text is enclosed in the appendix. The example in appendix will be used as a basis for illustrations of different aspects in the analysis.

The paper does not present the great number of screen layouts and parameter options for the integrated handling of all the steps in the analysis. As user of PERTEX you have control of all the steps in the analysis and you can stop and restart the process at different stages.

PERTEX is built to realize the following main steps of PTA:

(1) *Coding of function words for verbs, prepositions, sentence openers, clause openers.* PERTEX has a specially developed dictionary and language dependent routines for identification of function words in the text.

(2) *Design and coding of blocks according to the AaO-paradigm.* A block is based on a verb, the a-component in AaO. The block also consists of an agent, the A-component, and an Objective, the O-component. The O-component is differentiated by prepositions used in the text. The limits of a block are in the general case set by sentence- or clause openers, e. g. full stop or comma.

(3) *Supplementation of A- and O-dummies.* In normal text variables for the A- and O-components are sometimes omitted. All the implicit references to and from A- and O-components are made explicit in this step. Even different forms of self references are handled by PERTEX.

(4) *Generation of A/O-matrixes.* The block oriented connection between A and O is a corner-stone in PTA. All such connections between the unique A- and O-components in the text are organized in different binary A/O-matrixes.

(5) *Cluster analysis based on A/O-matrixes.* Ward's method for clustering is used i PERTEX. By clustering the O-rows in a A/O-matrix, with the A's as variables, we extract the structural relations of the Objective in the text. When, in the transposed matrix, the A-rows are clustered, with the O's as variables, we extract the structural relations of the text producer's perspective on the Objective.

(6) *Topological presentations of outcomes.* The user of PERTEX has to select the significant number of clusters in every cluster analysis. Here PERTEX offers not only Ward's ESS-values but also different t-tests. The number of selected clusters and the text content of every cluster are presented for the user's naming of the clusters. The clusters are then organized according to the cluster tree and the user can fulfil the investigation of the text by following the synthesis from the clusters to the root of the cluster tree.

This short exposé of PTA via PERTEX is only meant as a preliminary frame. Now every step will be presented with more details and illustrative examples. First, however, some comments on import and export of normal text to and from PERTEX.

Normal text to and from PERTEX

The input text to Perspective Text Analysis is normal text produced in any purposeful context. Such a text can be imported to PERTEX from a text file on diskette. It is also possible to write/edit the text by using PERTEX's text editor. A text imported from a text file can also be edited by this editor. A normal text in PERTEX can be exported to an external text file for use in other systems, e. g. a system for word processing.

The handling of normal text in PERTEX is illustrated on page A3 in appendix. The text has running numbers per row. This numbers can be used by the text editor in searching specific parts of the text. Numbers for reference to rows are also used in coding and supplementation, see page A4-A8. In the automatic functions for control of syntax these numbers are used for indication of where errors are located.

Under the head "Label", see page A3, the user can insert labels indicating the start of a specific section of the text. By labels it is possible, later on in the process, to select only some part or parts of the text as basis for an analysis. If, for example, a text reports a discussion between two or more people labels can by used to select the text from one of the speakers. The use of labels is optional. If no label is used the entire text is used in the analysis.

In different phases of the analysis PERTEX use different formats in handling the text. The format can be strings of single words, blocks, Agents and Objectives. Starting from normal text on a file the text is first made up for dictionary coding. In this format, see page A4, every single word, full stop, comma, colon and semi-colon is placed separately on one row. In this formatting of the text PERTEX uses special rules for English. Short forms like *it's* and *I'm* are transformed into the ordinary forms *it is* and *I am*. These transformations are necessary because PERTEX must operate on single words in coding the text.

Dictionary coding

The dictionary coding is the most language dependent phase of the analysis. Strings for single words shall be coded as verbs, prepositions, sentence openers or clause openers. For every language there is a unique dictionary. At the moment PERTEX is practically operative on English and Swedish texts. The English dictionary has about 6.000 items and the Swedish about 3.000 items. For German, French, Danish, Norwegian, Finnish and Latin PERTEX has dictionary "embryos" . It is also possible to use PERTEX on a text without specification of any language. In that case PERTEX is only used as an administrative instrument with control functions and the user has to do all the coding manually.

PERTEX use language dependent dictionaries for coding single rows in the text formatted according to page A4. The codes used are as follows:

00	sentence opener	60	preposition for ground
01	clause opener	6?	uncertain preposition for ground
		70	preposition for mean
40	active verb	7?	uncertain preposition for mean
40P	passive verb	80	preposition for goal
4?	uncertain verb	8?	uncertain preposition for goal

These codes are placed by PERTEX in the Code column, see page A5. In this automatic coding each row (one word) is matched against the dictionary. If the text string on the row is found in the dictionary the corresponding code is set at the row. Verbs can be identified irrespective of the conjugation of verbs in the text. The dictionary contains the stem of the verb plus codes for the conjugation. For English texts PERTEX operates with six different patterns for conjugation of verbs. Each pattern represent four different forms of the verb, e. g. *walk, walks, walked, walking*. Verbs that have irregular paradigms are also included in the dictionary.

For Swedish texts PERTEX operates with ten separate patterns for conjugation of verbs. Here each pattern includes 12-14 different forms for the verb. In Swedish a verb is often combined with a lot of different prefixes. All this combinations as separate verbs would have resulted in a dictionary too large to be handled. Instead PERTEX can combine about 50 Swedish prefixes with the items in the Swedish dictionary.

Dependent on the situation, specially in English, a specific word is a verb or not a verb, a preposition or not a preposition. When we read a text it is obvious which words are verbs and prepositions. This flexibility in language can not be managed automatically by simple matching of text strings. After a lot of experiments with language dependent rules for identification of verbs and prepositions PERTEX can eliminate most of the 4?, 6?, 7? and 8? codes generated from the matching of text to dictionary. At the moment PERTEX has 22 such rules for handling the primary result from dictionary coding of English texts. A short example illustrates this important stage of the automatic coding of text.

A sentence like *I want to go to the town by bike to buy a present for my friend's birthday.* has five words that by matching against the dictionary are coded 4? or 6?. This preliminary result is processed by special rules in PERTEX so that the final result from dictionary coding of the sentence will not include any 6? or 4? code.

Only matching to dictionary		Final result from dictionary coding	
	I		I
4?	want	40	want
6?	to	01	to
40	go	40	go
6?	to	60	to
	the		the
	town		town
70	by	70	by
	bike		bike
6?	to	01	to
40	buy	40	buy
	a		a
4?	present		present
70	for	70	for
	my		my
	friend's		friend's
	birthday		birthday
00	.	00	.

If the dictionary coding ends up with any ?-code left, PERTEX's control of syntax will indicate where these ?-codes are. The user of PERTEX then has to decide how this words shall be coded. Without elimination of all the ?-codes it is not possible to continue the analysis to the next stage. The text in appendix, page A5, is automatically coded without any ?-code.

PERTEX has an optional function for logging all corrections the user does of the coding automatically produced in dictionary coding. This code journal can be used for improvement in the dictionary and the routines for dictionary coding.

Coding of blocks

In this phase the whole text is formatted and coded in blocks. According to the AaO-paradigm a block consists of three types of components: Agent (A), action (a) and Objective (O). The a-components are already coded from the dictionary coding. The simple main rule for creating a block says that the text strings before the a-component are the Agent and the text strings after the a-component are the Objective. In practice it is, however, a complicated process to create the blocks from the dictionary coding of the text. First the start row and stop row for every block must be identified. Then the following codes are used to give a code to every row, see page A6.

01	Boundary for blocks
30	Agent (A)
40	action (a), active verb
40P	action (a), passive verb
50	Objective (O), Figure
60	Objective (O), Ground
70	Objective (O), Mean
80	Objective (O), Goal

Bierschenk & Bierschenk (1986 b) present a set of 18 process rules for coding of blocks. This rules are used and to some extent transformed into a transaction design in PERTEX. This means that the overall frame of reference, naturally present when using the original rules manually, is replaced by a more computer oriented design of strict transactions based on certain combinations of codes and text strings.

In PERTEX the block coding is processed in nine separate steps operating with 46 different transactions. A simple example on such a transaction can be illustrated by the following situation. A row with a 40-code has a succeeding row not yet coded. The code for that succeeding row is then set to 50. Another transaction is used when a 50-row is followed by a row not yet coded. The code for that following row will be set to 50. From these two examples of transactions we see that it is necessary to organize the use of transactions in a purposeful way. By the design of block coding in the nine separate steps this requirement is fulfilled. The transaction design of block coding has proved to be a practical and flexible design for the development of new and language specific rules for computerized block coding.

The whole process of block coding is too comprehensive to be reported here in detail. The sentence used for illustration of dictionary coding can be used here as well, to give an example of some of the main principles for block coding.

Final result from dictionary coding		Final result from block coding			
		01		*	
	I		30	I	(Start block 1)
40	want		40	want	
			50	*	
01	to	01		to	
			30	*	(Start block 2)
40	go		40	go	
60	to		60	to	
	the		60	the	
	town		60	town	
70	by		70	by	
	bike		70	bike	
01	to	01		to	
			30	*	(Start block 3)
40	buy		40	buy	
	a		50	a	
	present		50	present	
80	for		80	for	
	my		80	my	
	friend's		80	friend's	
	birthday		80	birthday	
00	.	00		.	
		01		*	

The boundaries for the three blocks are the 01-codes from dictionary coding. The first block is based on the verb *want*. The text before *want* is coded 30 as an A-component. After *want* there is originally no text before the boundary between block 1 and block 2. As a block per definition must have an Objective we insert a dummy Objective indicated by '*' and coded by 50.

In block 2 there is originally no text before the a-component *go*. As a block must have an Agent we insert a dummy Agent indicated by '*' and coded by 30. The Objective in block 2 is specified by 60-code from *to* and 70-code from *by*. Strings following a preposition are generally coded by the code from the preposition.

In block 3 another dummy for the Agent is inserted. After the verb *buy* the Figure code 50 is used for *a present* and the 80-code from *for* gives the codes for the remaining words in the block.

The sentence *I want to go to the town by bike to buy a present for my friend's birthday.* generates three blocks based on one verb in each block. If explicit Agents and Objectives are omitted in a block dummies are inserted. Here it is important to notice that it is the **functional position** in the block, and not any classification of words, that governs the identification of Agents and Objectives. This means that a single word can be an Agent in one block and an Objective in another block.

This little example gives only an indication of some of the details in PERTEX's automatic block coding. The appendix, page A5 and A6, demonstrate some other details. First we can notice that a 01-code not always is a boundary for blocks, see block 1 on row 2-29. If codes for prepositions are located in text strings for an Agent, the 30-code eliminates former 60-, 70- and 80-codes, see block 1. This is an example of substitution of codes during the process of block coding.

On page A6 row 141, '70 *' is inserted as an O-dummy. This dummy is initiated by '70 with' on row 140 as the last word in that block. This 'with' refers to something that will be made explicit in the supplementation phase. The same situation can be found for a '60 *' on row 197.

On page A5, row 109 and 110 are both coded 01 from dictionary coding. Tow such 01-codes close to each other, are interpreted as a boundary for a sentence and the first one of the 01-codes is replaced by a 00-code, see row 128 on page A6.

If the first block in a sentence has a word as clause opener and first boundary, and this word is just before the verb, the general X-Agent is inserted. This Agent is also differentiated by the clause opener, see rows 64-65 on page A5 and rows 69-72 on page A6.

Passive verbs, coded as 40P, can be found on page A6 row 59, 88, 118 and 169. In block coding for an English text the passive form is identified by the verb *be* followed by another verb. This two verbs are transformed to one row, e. g. row 59 *be_driven*. The identification of Agent and Objective in a block with a passive verb is handled in a special way, see page A6. In Swedish the passive form is normally indicated via the s-form of the verb. Such passive verbs are coded 40P already in dictionary coding.

The system has a function for strict control of the block syntax. In case of any syntax error the process can not continue to supplementation before the user has eliminated all errors. After a lot of experiments with different texts and automatic block coding, PERTEX is now operating without syntax errors in this complicated phase of the analysis. The interactive design of PERTEX makes it possible for the user to manipulate all the codes, and even do all the coding manually.

Supplementation

In this phase of the analysis all the A- and O-dummies will be supplemented by explicit text strings. The main rules in PTA say that an A-dummy is supplemented by the Agent in the preceding block and an O-dummy is supplemented by the Agent and the Objective in the succeeding block. If a block has the Agent 'it', then that Agent is supplemented by the Agent and the Objective from the preceding block. This main rules seems very simple. But in practice the supplementation is very complicated because of chained relations between several blocks and self references between dummies.

In Bierschenk & Bierschenk (1986 b) a set of 18 rules stipulates how to handle the supplementation. These rules operate with reference numbers for blocks. With chained relations between blocks these reference numbers must be handled in complex chains with numbers embedded at several levels. These rules are not copied directly to a program in PERTEX. By transformation of the rules to a strict transaction design, cf. the design of block coding, the supplementation in PERTEX can be done without chains of embedded reference numbers. Instead PERTEX operates in an iterative way by using a set of circa 30 transactions for specific combinations of codes and text strings. The number of iterations depends on how complicated the text is according to A- and O-dummies and the relations between the dummies.

In supplementation it is no longer strings for single words that are handled. It is now the entire Agent, 50-Objective, 60-Objective, 70-Objective and 80-Objective that are manipulated. Because of this PERTEX starts the supplementation by making up the text from block coding into a format for supplementation, see page A7. This transformation of the sentence used for demonstration of block coding, and the supplementation of that sentence, are as follows:

Made up for supplementation format			Supplementation ready		
01		*	01		*
	30	I		30	I
	40	want		40	want
	50	*		50	I to the town by bike
01		to	01		to
	30	*		30	I
	40	go		40	go
	60	to the town		60	to the town
	70	by bike		70	by bike
01		to	01		to
	30	*		30	I
	40	buy		40	buy
	50	a present		50	a present
	80	for my friend's birthday		80	for my friend's birthday
00		.	00		.
01		*	01		*

In this little example a supplementation is first done for the A-dummy (30 *) in block 2. The Agent 'I' from block 1 is used as Agent in block 2 instead of '*'. Then PERTEX continues with the A-dummy in block 3, and this dummy is replaced with the Agent from block 2, which now is 'I' as a result from the preceding supplementation transaction. All three blocks have got the same agent 'I'. We also see that it was necessary to supplement the Agent in block 2 before the Agent in block 3. The supplementation of an Agent requires an explicit Agent in the preceding block. As illustrated in this little example the supplementation of Agents runs from the first to the last block of the text.

The supplementation of O-dummies runs from the last to the first block. In the example above there is only one O-dummy, see block 1. After the supplementation of all the A-dummies this O-dummy is supplemented by the Agent and the Objective from block 2.

After one Agent-run from top to bottom and one Objective-run from bottom to top all the dummies in the example are eliminated. For a real text this is usually not the case. Even after an iterative use of Agent-runs and Objective-runs, as illustrated above, some dummies remain and can not be supplemented in this way. The reason can be that the Agent in block *n* is of type 'it'. This Agent requires an explicit Agent and an explicit Objective in block *n-1*. If the Objective in block *n-1* is an O-dummy the situation indicates a simple self reference between A- and O-dummies. The O-dummy in block *n-1* requires an explicit Agent in block *n*. But the agent in block *n* is the 'it'-dummy we started with. PERTEX breaks this form of reference loop when the supplementation can not continue according to the main rules illustrated above. The Agent in block *n-1* and the Objective in block *n* will be the supplement for the O-dummy in block *n-1*. After that '30 it' in block *n* can be supplemented without problem. Self references between A- and O-dummies can be more complicated and involve several blocks.

Different forms of more sophisticated supplementation indicates the need for an iterative process in supplementation. PERTEX works with an iterative change between supplementation of A-dummies and O-dummies in as many runs as needed for elimination of all dummies. The system always brings this iterative process to a normal stop with all the dummies supplemented. The number of iterations reported on screen also include administrative iterations in supplementation. When text strings are combined for new Agents and Objectives the result will sometimes end up with the same text string duplicated in an Agent or an Objective. All such copies of text are automatically eliminated.

For the text example in appendix we have five '30 *', one '30 it', two '50 *' one '60 *' and one '70 *', ten dummies in all, see page A7. The result of the supplementation can be found on page A8. This text does not illustrate any particular complications in supplementation.

Generation of A/O matrixes

In Perspective Text Analysis the block oriented relation between Agent and Objective is a corner-stone. PERTEX also calculates how many unique Agent/Objective combinations there are in the text. These A/O-combinations are found in the supplemented version of the text and presented in binary A/O matrixes.

Here it is important to notice that the generation of A/O-matrixes is not based on frequencies for the Agents and Objectives. As the analysis is built on affinity, and not on frequencies, it is the number of unique Agents and Objects that are of interest. It is the block-wise combinations of these unique Agents and Objectives that are organized in the binary matrixes. PERTEX produces four types of such matrixes.

Matrix	Represents
50/30	Figure
60/30	Ground
70/30	Mean
80/30	Goal

As indicated by the matrix type, it is the four different aspects of the Objective component that are used for a separation into four different types of A/O-matrixes.

The same (only one) Agent in a block is used in all the four possible A/O-combinations in one block. By definition a block has minimum one and maximum four types of A/O-combinations. The number and kind of A/O-combinations and A/O-matrixes is only dependent on the text and can not be governed from outside the text.

The sentence used earlier for illustration of coding has three blocks but only one unique Agent, 'I'. This means that all matrixes generated from the supplementation of the sentence will be of vector type. With the two unique 50-Objectives (*I to the town by bike* from block 1 and *a present* from block 3) the binary matrix of type 50/30 has the dimension 2x1 and has a binary '1' in each of the (two) cells. As can be seen from supplementation, the 60/30, 70/30 and 80/30 matrixes will have the dimension 1x1 with one binary '1' each. This little example, primarily used for some basic illustrations of coding, is too small for any realistic demonstration of A/O-matrixes. Therefore we now leave that sentence.

In appendix, page A9, is reported the 17 unique 50-Objectives and the 11 unique Agents that are combined in the blocks from supplementation of the text. The unique text strings are marked by simple numbers for identification. These numbers are created successively when PERTEX is processing the text from the first to the last block. All the unique 50-Objectives are reported but only the unique Agents from blocks which have a 50-Objective are reported.

The 17 unique 50-Objectives and the 11 unique Agents define a 50/30-matrix of size 17x11. On page A9 this matrix is technically described in three different ways. The Agent coordinates (column index) are given for every 50-Objective. The same matrix is also described by the 50-Objective coordinates (row index) for every Agent. Finally the matrix is also reproduced in explicit form with a binary '1' according to the coordinates. Here it is important to notice that a binary '1' only indicates that the combination of the corresponding Agent and 50-Objective is found in at least one block. The matrix does not give any indication of how many times this combination is used in the text. As said before, it is not frequencies of A/O-combinations but affinity for the unique A/O-combinations that are in focus for the analysis. This focus will be dealt with in the forthcoming cluster analysis based on the binary matrixes.

On page A10 is reported the unique text strings and coordinates for the 60/30-matrix. Here we see that all the unique text strings indicating Ground begin with a preposition coded 60 in dictionary coding. Among the unique Agents we find for example that the first one, *the style of their ships*, is not an Agent combined with 50-Objectives, see page A9. In the second block of the text, see page A8 row 8-18, we find the basis for this situation. In block 3, page 8 rows 13-17 we see how the Agent *there X* is combined with both a 50-Objective and a 60-Objective.

On page A11 is reported some statistics for the text and the matrixes. The TEXT STATISTICS are for the present in focus at the front line in the scientific interpretation of results from PTA and will not be discussed in this paper. In the section for MATRIX STATISTICS we find a short summary of the generation of the matrixes discussed here.

The 50/30-matrix is based on 17 (81%) of all the blocks in the text. The dimension 17/11 has been discussed earlier. The measure for Density is a standardized value (0.0-1.0) indicating the proportion of binary '1' in the matrix. Density=0.0 means that there is no extra binary '1' in the matrix. Density=1.0 means that all unique Objectives are combined with all the Unique Agents. A Density of 1.0 is not expected to be found for real texts. For the 50/30-matrix the Density is set to 0.0 because there are only 17 cells marked '1' in the matrix. This is the minimum number of unique combinations marked '1' in a matrix with maximum number of rows or columns equal 17. The interpretation of the Density value is still under investigation. Tests on different texts indicate that the structural relations are more complex (see cluster analysis and sub-trees discussed in the next section) for matrixes with relatively high Density value.

On page A11 is reported a 70/30-matrix with dimension 4/4 and the comment *Diagonal, no structure*. This means that the unique combinations between Agents and 70-Objectives generates a 4x4 matrix with four binary '1' placed on the diagonal. A cluster analysis based on such a matrix is not meaningful because the clustering process is here completely arbitrary. For all such matrixes PERTEX signals this type of comment to the user. It is technically possible to fulfil an analysis based on a matrix with binary '1' only in the diagonal, but it will be done on the users responsibility. The 70/30-matrix from this text is not handled any further here.

The fourth type of matrixes, 80/30, is reported on page A11 as a 1x1 matrix, which, as the comment says, is *No matrix*.

The matrixes generated from the text discussed here represent a general pattern found for many texts. The 50/30-matrix, for Figure, is the largest matrix. Then the matrixes for Ground, Means and Goal are decreasing in size, or perhaps not present at all. This observations are only technical remarks and do not say anything about the final results of the analysis.

Cluster analysis based on A/O-matrixes

The binary matrixes discussed in the previous section are used for cluster analysis. First the Agents are interpreted as variables and the 50/30-, 60/30-, 70/30- and 80/30-matrixes are set up as conventional data matrixes with the dimension $n \times p$. By clustering these types of matrixes we expose the structural relations for the Objectives in the text. When clustering the transpose of these matrixes we expose the text producers perspective on the Objectives.

Ward's method, Ward(1963), is used for clustering. This is a robust and well-known method which generates valid and interesting results in PERTEX. In Ward's method the ESS-value (Error Sum of Squares) is used as the clustering criteria. The ESS-value for one variable and n items is calculated by the formula:

$$ESS = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2$$

For a binary variable the ESS-formula can be transformed to:

$$ESS_B = \frac{NB_1 * NB_0}{n}$$

where NB_1 is the number of binary '1'-values for the variable and NB_0 is the number of binary '0'-values for the same variable. By definition $NB_1 + NB_0 = n$. The total ESS-value for a binary data matrix is calculated by accumulation of the ESS_B-values for all the variables.

By using the ESS_B-formula in PERTEX the calculation of ESS-values can be done in a very effective way. The simple idea involved is the fact that all values in a binary data matrix are '1' or '0'. The ESS-value for a cluster of *m* items and *p* variables is in the general case based on *m* x *p* values. With binary variables the calculation of ESS_B for such a cluster can be based on only *p* values. Each of these *p* values marks the number of binary '1' in the cluster for that variable.

The clustering process in PERTEX is also built according to the special form of the binary data matrixes used in this phase of the analysis of the text. As illustrated in appendix, page A9, the binary matrix has significant more '0'-elements than '1'-elements. Therefore it is effective to base the calculation of ESS_B-values on the '1'-elements. If we know the coordinates for all the '1'-elements, then we also know the whole matrix. This specific character of the binary A/O-matrix is used in PERTEX to optimize the comprehensive searching for the next fusion in the clustering process and for updating a similarity matrix used in that process.

Despite the special things made in the design of the clustering process, the results produced in PERTEX is exactly according to Ward's method. The overall purpose for the specific implementation of Ward's method in PERTEX is the optimization of calculation to minimize the need for computer time. After a lot of improvements the algorithm can now handle rather big matrixes in acceptable amount of time. A binary matrix of dimension *132* x *74* is for example clustered in 48 seconds on a 386-PC.

The result from a cluster analysis of the 50/30-matrix in appendix is reported on page A12. The whole clustering process going from 17 to 2 clusters is documented with fused rows and ESS-value for every step. The ESS-values are reported both as the increase of ESS in every step and as the accumulated ESS so far in the process.

A decision of number of significant clusters can be based on a cut-off point in the ESS-values. To improve the support for the users decision of number of significant clusters t-tests can be used. The ESS-values are tested in one-side t-tests, both for the step by step values and the accumulated total values. Two tests with different degrees of freedom can be ordered. The test of step by step values and DF=(number of items - 2) is the same testing as proposed in Wishart(1987). The test of step of step values with DF=(present number of clusters - 1) has been proposed in Bierschenk & Bierschenk (1986 c) as a more conservative way of testing ESS-values in connection with Perspective Text Analysis. Practical experience so far indicates that the second way of testing seems to be a robust way to transform the cut-off point for ESS-values to standardized value of significance. The corresponding t-tests for the total ESS-values are under observation and the relevance of these tests will be evaluated on the basis of future experience. The different t-tests are optional and can easily be ordered by the user. An automatic cut-off for useless values of significance greater than 0.5 saves some computer time.

In the example reported on page A12 both the cut-off point for the ESS-values step by step and the corresponding t-tests indicate that seven clusters seem to be the greatest number of significant clusters in this example. The cluster tree can be printed with the dynamic scale based on the step by step ESS-values or the total ESS-values. When a specific number of clusters are selected a dotted line indicates the cut-off in the tree.

A hierarchical cluster tree can be organized and presented in many different ways. For a tree with n items clustered, the orientation of the branches at the nodes can be combined in $2^{(n-1)}$ different ways. All the cluster trees produced i PERTEX are organized and printed according to the following two rules. The rules are expressed under the assumption of a horizontal representation of the tree, see the trees on page A12.

- If the number of clustered items are not equal on both the branches at a node, that branch with the greatest number of items will be oriented upwards.

- If the number of clustered items are equal on both branches at a node, that branch with the lowest item number, *Row no*, will be oriented upwards.

The items clustered, here unique text strings for 50-Objectives, are indicated by their *Row no* to the left on page A12. *Row no* refer to rows in the 50/30-matrix.

These two rules for design of the cluster tree are very important and the technical key for the interpretation of the clustering as a result of a process and not only as a hierarchical organization of certain items. More about this will be outlined in the next section.

With this design of the cluster tree we normally get a special cluster at the top. This top cluster is created at the lowest ESS-value according to the cut-off line and consists of 'residual' items not fused to other clusters because of dissimilarity. The top cluster for CLUSTER TREE 5030 on page A12 consists of the items 1,2,7,12 and 15. The 50/30-matrix on page A9 shows that just these five items are the only items, of totally seventeen items, that have no binary '1'-connection to any other item via a common Agent. Item 5 and item 10, for example, have such a connection via Agent 4. In the cluster tree on page A12 we also find that item 5 and item 10 create a cluster, the third cluster from the top. The clusters are identified by number of sequence from top to bottom along the cut-off line. This numbers are used as reference numbers in the next section.

So far we have discussed clustering based on the 50/30-matrix. This analysis gives the structural relations for the Figure. The interpretation of this cluster analysis, specially the synthesis from clusters according to the cluster tree, will be explained in the next section.

The transposed 50/30-matrix, the 30/50-matrix, is also used in a cluster analysis, see page A14. The Agents are here clustered with the 50-Objectives used as variables. This clustering is done exactly in the same way as described for the 50/30-matrix. Seven significant clusters are created in this analysis. This text is to some extent special as we find the same number of significant clusters for Objectives and for Agents. Normally, and particularly for longer texts, we get more O-clusters than A-clusters. The use of the result from the clustering of the 30/50-

matrix will be demonstrated in the next section. An important connection between the clustering of Objectives and Agents will also be discussed in the next section.

The Ground matrix 60/30 and its transpose 30/60 are also clustered in exactly the same way as the 50/30- and 30/50-matrixes. The results from the 60/30-clustering are reported on page A16, and the results from the 30/60-clustering on page A18. In both cases four significant clusters are created.

Naming of clusters and synthesis from clusters

This is the last phases in Perspective Text Analysis. Here the user of PERTEX has to do the important part of the job, and the system is mainly used as an instrument for documentation and organization of text material. The basic idea is that all the preceding steps of analysis will now end up in a description, topological in nature, of the structural relations uncovered by the clustering of Objectives and the text producers perspective on the Objectives uncovered by the clustering of Agents.

The seven significant clusters from the 50/30-analysis are reported on page A13. Here the unique 50-Objective strings are printed for each cluster. The naming of clusters is now an important task for the user. Based on the text strings in the cluster a prototypic naming/description of the cluster must be created. This task can be more or less difficult. Different texts, and alternative numbers of significant clusters selected, will require a varying amount of intellectual effort in finding appropriate names for clusters.

For cluster 1 with the unique 50-Objective strings 1, 2, 7, 12 and 15, the prototypic name is set to *Preparedness*. Cluster 1 is the cluster that collects the 'residual' items mentioned earlier. The name of such a cluster has to be relatively broad to sum up a general frame present in the text and expressed by the text strings in the cluster. Generally one or more synonyms can be found for naming of a cluster. The situation described in cluster 1 indicates different forms of threats. Without these threats no preparedness. Threats and preparedness can be seen as alternatives for characterising the general frame expressed in cluster 1.

For a small text, like the actual one, some clusters only consist of a few text strings and the naming must eventually be based on one 'dominating' item. In cluster 2 item 4 'it' is of no help in naming the cluster. The proposed name *Direction* is easily found as an general expression for what is prototypic for item 3. Naming of the remaining five clusters are found on page A13 and will not be discussed here in detail.

It is important to notice that the clusters are named independently of each other. The separate text strings for each cluster are presented on screen when PERTEX is used interactively for the naming task. Names for clusters can easily be changed on screen.

So far the analysis has, in all the phases, from dictionary coding to naming of clusters, manipulated certain and mainly rather small parts of the text. Now we are ready, finally, to see how all these isolated manipulations will end up in a synthesis that grasp the intention of the entire text. This is practically possible mainly because of the unique AaO-paradigm used as a theoretical base. In a more technical sense, it is of great importance that, in the clustering process every fusion is based on calculations for all possible fusions. By this approach we are now prepared to see how a synthesis from the named clusters, according to the structure of the cluster tree, will uncover the structure of the intention the text presents.

On page A13, the right part, PERTEX has printed a box-like version of the cluster tree. This box-tree is read from top to bottom. The box at the top represents the well-known cluster 1 from clustering the 50/30-matrix. To the left in this box-tree we find all the seven clusters from that analysis. The names given to the clusters are printed in the boxes. The number identification for each box is the cluster number before colon. After colon is marked the number for the first item in the cluster. These numbers are useful as reference numbers between the box-tree, the naming of clusters and the ordinary cluster tree.

When running the box-tree on screen for the first time only the cluster names are present. The final task is to follow the arrows indicating a flow, according to the structure of the fusions in the cluster tree, to build up a synthesis based on the cluster names. The first cluster *Preparedness* will be transformed by cluster 2 *Direction*. The result from the fusion when *Preparedness* is transformed by *Direction* is *Determination*. In the next step *Determination* is transformed in a new fusion by cluster 3 *Regime*. The result is formulated as *Domination*. So the synthesis process continues until the last box is filled with *Safety* after the transformation of *Violation of Order* by *Constraint*.

In the *Safety*-box all the 50-Objective items, the Figure items, are assembled. Therefore we can say that the Figure described in the text can be characterized by *Safety*. A remarkable thing is that this conclusion can be made without really reading the text. The result is based on a comprehensive, strict, formal and computerized, analysis and a synthesis based on certain elements from that analysis. To check this result it is of course necessary to read the text. When doing so it is a bit astonishing to notice how well the whole text can be characterized by *Safety*. What would be a better alternative?

In the box-tree it is not only the final box that is of interest. The structural relations indicated by the arrows between the boxes demonstrate how, and from what, the Figure *Safety* is created. The process from cluster 1 to the final box is unique for every text. Dependent on the complexity in the text there can be one or more sub-trees developed. A sub-tree handles a separate 'line of thinking' in text, e. g. a question or a theme, that is elaborated to the extent that the cluster analysis is influenced. In the cluster tree it is possible to follow how such discussions are linked together via fusions.

The synthesis from the 50-Objective clusters is only one part of the results for the Figure component. The other part is the text producers perspective on the Figure. This perspective is handled via clustering of the 30/50-matrix. The clusters of Agents are created according to the connection between the Agents and the Figure items. Every Agent cluster has its own specific collection of such connections to the Figure. Therefore, the Figure items connected to a Agent cluster is interpreted as an indication of the perspective on the Figure. Here we must notice that we are not going to describe an Agent cluster as a group of unique Agents. Instead it is the Figure items, as variables, which motivate such a cluster, that are of interest. The clustering of Agents can be said to generate a number of 'agent positions' from which the Figure is 'observed'. This positions define the perspective for observing the Figure. By clustering the Agents connected to the Figure we uncover the structural relations for the text producers focus on the Figure.

The clustering of the 30/50-matrix is done in the same way as was described for the 50/30-matrix, see page A14. The interpretation of the Agent clusters is, however, already prepared by the naming of Figure clusters. On page A15, middle of the page, is reported a matrix for connections between 50- and 30-clusters. This matrix is nothing but an aggregated form of the binary matrix from page A9. The aggregation is based on the clustering of both Figure items and Agent items. The size of the matrix is set by the number of Figure clusters and the number of Agent clusters, here 7 x 7. The numbers for identification of clusters discussed earlier is used as identification of rows (Figure clusters) and columns (Agent clusters). The cells in the matrix mark the number of connections between unique Figure items and unique Agent items on cluster level. Notice that the sum of these numbers of connections is the same as the sum of binary '1'-values in the matrix on page A9, and the same as the numbers of coordinates in MATRIX STATISTICS on page A11.

By using the connections on cluster level we easily can handle the naming of Agent clusters by references to the corresponding Figure cluster(s). In this example the connection matrix is a diagonal matrix. This means that there is an exact correspondence between clusters on item level. Therefore the naming of all the Agent clusters can be copied from the Figure clusters, see the right part on page A15. The box-tree for the perspective on Figure is just the same tree as the tree for Figure. In this text the perspective on the Figure is the same as the Figure. The text is not written to put some aspects in focus, all aspects are emphasized at the same level.

The diagonal form of the connection matrix is often found for shorter texts and texts primarily reporting facts. When a text has more of discussion or argumentation we normally get more Figure clusters than Agent clusters. In that case the connection matrix can not be a diagonal matrix. Instead the matrix usually describes how a number of the first Figure clusters are all connected to the first Agent cluster. This means that the process in the Figure tree, from cluster 1 to cluster x, is concentrated in one Agent cluster. It is the point reached so far in the Figure tree that is in focus and will be taken as name for the first Agent cluster. From this starting point for the perspective on the Figure the remaining Agent clusters are handled according to the connection matrix. In some cases the elaboration of the perspective on the Figure will imply that the perspective represents a transformation of the direction in a part of the process in the Figure tree. This means that in the perspective, Figure clusters can be picked up and transformed in a process that is not only a compressed copy of the Figure process.

The perspective on the Figure is a unique quality in Perspective Text Analysis. The use of the connection matrix is a key instrument for uncovering the structural relation in the perspective. So far PERTEX has no automatic routines for the use of the connection matrix. The matrix is automatically produced and can also be used to reproduce the binary matrix for connection between unique Figure items and unique Agent items. If the number of Figure clusters selected is set to the number of unique Figure items, and the same is done for unique Agent items, then the connection matrix will reproduce the binary matrix. This new matrix has, however, the rows and columns organized according to the specific sequence of items received from the organization of the cluster tree. The pattern of '1'-values in the matrix will be more like a diagonal than in the original binary matrix. Such a matrix can be useful for experiments and special studies of combined Figure and Agent clustering.

The Ground, Mean and Goal components are technically handled in PERTEX in the same way as has been described for the Figure. The result for naming of Ground clusters and the synthesis from these clusters are reported on page A17 and the perspective on page A19.

These results do not require any technical explanations as the results do not introduce any new complications in using PERTEX or understanding Perspective Text Analysis.

Summary

This paper has demonstrated the main steps in Perspective Text Analysis when using PERTEX. As has been demonstrated, this kind of new and unique text analysis is not one single type of analysis. PERTEX is built to cover all the different kinds of analysis, and synthesis, involved. The technical output from the system is illustrated in appendix with the complete printer output from analysis of a classic text. All information in this printer output is also available on screen during the interactive use of PERTEX.

The different phases in the analysis, as discussed for the example in appendix, are easily handled in the interactive use of PERTEX. The process described for analysis of a text can be stopped and restarted at any stage. The strict control of syntax prevents the user to leave one level of analysis before all errors are eliminated. Special indicators on screen mark the status of the analysis for every text. A lot of parameters are automatically set for this purpose. By other parameters the user has a lot of options to design his personal use of PERTEX. By using PERTEX it is now possible to practise Perspective Text Analysis on a text to get a quite unique insight into the structural relations of the mentality the text presents.

References

Bierschenk, B. & Bierschenk I. (1986 a). Concept formulation. Part I. The phenomenon of cognition. *Cognitive Science Research* (10).

Bierschenk, B. & Bierschenk I. (1986 b). Concept formulation. Part II. Measurement of formulation processes. *Cognitive Science Research* (11).

Bierschenk, B. & Bierschenk I. (1986 c). Concept formulation. Part III. Analysis of mentality. *Cognitive Science Research* (10).

Hutton, M. (1958). *Publius Cornelius Tacitus: Germania*. London & Cambridge, MA: The Loeb Classical Library. (Original work published 1920)

Ward, J.H. (1963). Hierarchical grouping to optimize an objective function. *Journal of American Statistical Association, 58*, 263-244.

Wishart, D. (1987). *Clustan user manual*. (Fourth Edition). St Andrews: University of St Andrews, Computing Laboratory.

Author Notes

This paper has been presented in the symposium "Computerized Research Methods for Psychologists" chaired by Wim G. B. Liebrand at the 25th International Congress of Psychology, held in Brussels July 19-24, 1992. The paper was also distributed at a four days exhibition and demonstration of the PERTEX system. The demonstration was arranged on site with the kind assistance of the Congress President Géry d'Ydewalle.

APPENDIX: PRINTER OUTPUT FROM PERTEX

The appendix is organized in the following sections:

The content of this appendix is a complete printer output of an analysis via PERTEX. The text selected for illustration of output from PERTEX is an English translation of Tacitus' classical text on the Suiones, Chapter 44, sections two and three, Hutton(1958):

Beyond these tribes the states of the Suiones, not on, but in, the ocean, possess not merely arms and men but powerful fleets: the style of their ships differs in this respect, that there is a prow at each end, with a beak ready to be driven forwards; they neither work it with sails, nor add oars in banks to the side: the gearing of the oars is detached as on certain rivers, and reversible as occasion demands, for movement in either direction. Among these peoples, further, respect is paid to wealth, and one man is accordingly supreme, with no restrictions and with an unchallenged right to obedience; nor is there any general carrying of arms here, as among the other Germans: rather they are locked up in charge of a warder, and that warder a slave. The ocean forbids sudden inroads from enemies; and, besides, bands of armed men, with nothing to do, easily become riotous: it is not to the king's interest to put a noble or a freeman or even a freedman in charge of the arms.

The purpose of using this text here is not to discuss historical or other aspects of the Suiones or Tacitus' text compared to other texts. The text is only used as an illustration of how a text is processed by PERTEX in Perspective Text Analysis. In a forthcoming article Bernhard Bierschenk will discuss the outcome from PERTEX-analysis of Tacitus' Latin text as well as analysis of translations of the text into five other languages.

During the interactive use of PERTEX you have the possibility to select just that printer output you want to have. All the information in printer output is also available on the PC-screen.

TEXT

TEXT EDITED IN PERTEX

Row Label Text

Beyond these tribes the states of the Suiones, not on, but in, the ocean, possess not merely arms and men but powerful fleets: the style of their ships differs in this respect, that there is a prow at each end, with a beak ready to be driven forwards; they neither work it with sails, nor add oars in banks to the side: the gearing of the oars is detached as on certain rivers, and reversible as occasion demands, for movement in either direction. Among these peoples, further, respect is paid to wealth, and one man is accordingly supreme, with no restrictions and with an unchallenged right to obedience; nor is there any general carrying of arms here, as among the other Germans: rather they are locked up in charge of a warder, and that warder a slave. The ocean forbids sudden inroads from enemies; and, besides, bands of armed men, with nothing to do, easily become riotous: it is not to the king's interest to put a noble or a freeman or even a freedman in charge of the arms.

TEXT
MADE UP FOR DICTIONARY CODING

Row	Code	Dic.-coding: Made up
	00	****** START OF TEXT
		Beyond
		these
		tribes
5		the
		states
		of
		the
		Suiones
10		,
		not
		on
		,
		but
15		in
		,
		the
		ocean
		,
20		possess
		not
		merely
		arms
		and
25		men
		but
		powerful
		fleets
		:
30		the
		style
		of
		their
		ships
35		differs
		in
		this
		respect
		,
40		that
		there
		is
		a
		prow
45		at
		each
		end
		,
		with
50		a
		beak
		ready
		to
		be
55		driven
		forwards
		they
		neither
60		work
		it
		with
		sails
		,
65		nor
		add
		oars
		in
		banks
70		to
		the
		side
		:
		the
75		gearing
		of
		the
		oars
		is
80		detached
		as
		on
		certain
		rivers
85		,
		and
		reversible
		as
		occasion
90		demands
		,
		for
		movement
		in
95		either
		direction
		.
		Among
		these
100		peoples
		,
		further
		,
		respect
105		is
		paid
		to
		wealth
		,
110		and
		one
		man
		is
		accordingly
115		supreme
		,
		with
		no
		restrictions
120		and
		with
		an
		unchallenged
		right
125		to
		obedience
		;
		nor
		is
130		there
		any
		general
		carrying
		of
135		arms
		here
		,
		as
		among
140		the
		other
		Germans
		:
		rather
145		they
		are
		locked
		up
		in
150		charge
		of
		a
		warder
		,
155		and
		that
		warder
		a
		slave
160		.
		The
		ocean
		forbids
		sudden
165		inroads
		from
		enemies
		;
		and
170		,
		besides
		,
		bands
		of
175		armed
		men
		,
		with
		nothing
180		to
		do
		,
		easily
		become
185		riotous
		:
		it
		is
		not
190		to
		the
		king's
		interest
		to
195		put
		a
		noble
		or
		a
200		freeman
		or
		even
		a
		freedman
205		in
		charge
		of
		the
		arms
210	00	.
	01	****** END OF TEXT

CODING

DICTIONARY CODING

Row	Code	Dic.-coding: Ready
	00	****** START OF TEXT
	60	Beyond
		these
		tribes
5		the
		states
	60	of
		the
		Suiones
10	01	,
		not
	60	on
	01	,
	01	but
15	60	in
	01	,
		the
		ocean
	01	,
20	40	possess
		not
		merely
		arms
	01	and
25		men
	01	but
		powerful
		fleets
	01	:
30		the
		style
	60	of
		their
		ships
35	40	differs
	60	in
		this
		respect
	01	,
40	01	that
	01	there
	40	is
		a
		prow
45	60	at
		each
		end
	01	,
	70	with
50		a
		beak
		ready
	01	to
	40	be
55	40	driven
		forwards
	01	;
		they

Row	Code	Dic.-coding: Ready
	01	neither
60	40	work
		it
	70	with
		sails
	01	,
65	01	nor
	40	add
		oars
	60	in
		banks
70	60	to
		the
		side
	01	:
		the
75	40	gearing
	60	of
		the
		oars
	40	is
80	40	detached
	01	as
	60	on
		certain
		rivers
85	01	,
	01	and
		reversible
	01	as
		occasion
90	40	demands
	01	,
	80	for
		movement
	60	in
95	01	either
		direction
	00	.
	60	Among
		these
100		peoples
	01	,
		further
	01	,
		respect
105	40	is
	40	paid
	60	to
		wealth
	01	,
110	01	and
		one
		man
	40	is
		accordingly
115		supreme
	01	,

Row	Code	Dic.-coding: Ready
	70	with
		no
		restrictions
120	01	and
	70	with
		an
	40	unchallenged
		right
125	60	to
		obedience
	01	;
	01	nor
	40	is
130	01	there
		any
		general
	40	carrying
	60	of
135		arms
		here
	01	,
	01	as
	60	among
140		the
		other
		Germans
	01	:
		rather
145		they
	40	are
	40	locked
	60	up
	60	in
150		charge
	60	of
		a
		warder
	01	,
155	01	and
	01	that
		warder
		a
		slave
160	00	.
		The
		ocean
	40	forbids
		sudden
165		inroads
	60	from
		enemies
	01	;
	01	and
170	01	,
	60	besides
	01	,
		bands
	60	of

Row	Code	Dic.-coding: Ready
175	40	armed
		men
	01	,
	70	with
		nothing
180	01	to
	40	do
	01	,
		easily
	40	become
185		riotous
	01	:
		it
	40	is
		not
190	60	to
		the
		king's
		interest
	01	to
195	40	put
		a
		noble
	01	or
		a
200		freeman
	01	or
	01	even
		a
		freedman
205	60	in
		charge
	60	of
		the
		arms
210	00	.
	01	******* END OF TEXT

CODING
BLOCK CODING

Row	Code	Block-coding: Ready
	00	****** START OF TEXT
	01	*
	30	Beyond
	30	these
5	30	tribes
	30	the
	30	states
	30	of
	30	the
10	30	Suiones
	30	,
	30	not
	30	on
	30	,
15	30	but
	30	in
	30	,
	30	the
	30	ocean
20	30	,
	40	possess
	50	not
	50	merely
	50	arms
25	50	and
	50	men
	50	but
	50	powerful
	50	fleets
30	01	:
	30	the
	30	style
	30	of
	30	their
35	30	ships
	40	differs
	60	in
	60	this
	60	respect
40	60	,
	00	that
	01	there
	30	there
	30	X
45	40	is
	50	a
	50	prow
	60	at
	60	each
50	60	end
	60	,
	70	with
	70	a
	70	beak
55	70	ready
	01	to
	30	to
	30	X

Row	Code	Block-coding:
	40P	be_driven
60	50	forwards
	50	;
	50	they
	01	neither
	30	*
65	40	work
	50	it
	70	with
	70	sails
	00	,
70	01	nor
	30	nor
	30	X
	40	add
	50	oars
75	60	in
	60	banks
	60	to
	60	the
	60	side
80	01	:
	30	*
	40	gearing
	60	of
	60	the
85	60	oars
	01	*
	30	X
	40P	is_detached
	50	as
90	50	on
	50	certain
	50	rivers
	50	,
	50	and
95	50	reversible
	60	on
	60	certain
	60	rivers
	60	,
100	60	and
	60	reversible
	01	as
	30	occasion
	40	demands
105	50	for
	50	movement
	50	in
	50	either
	50	direction
110	80	for
	80	movement
	80	in
	80	either
	80	direction
115	00	.
	01	*

Row	Code	Block-coding:
	30	X
	40P	is_paid
	60	Among
120	60	these
	60	peoples
	60	,
	60	further
	60	,
125	60	respect
	60	to
	60	wealth
	00	,
	01	and
130	30	one
	30	man
	40	is
	50	accordingly
	50	supreme
135	50	,
	70	with
	70	no
	70	restrictions
	70	and
140	70	with
	70	*
	01	*
	30	*
	40	unchallenged
145	50	right
	60	to
	60	obedience
	00	;
	01	nor
150	30	nor
	30	X
	40	is
	50	*
	01	there
155	30	any
	30	general
	40	carrying
	60	of
	60	arms
160	60	here
	60	,
	60	as
	60	among
	60	the
165	60	other
	60	Germans
	01	:
	30	X
	40P	are_locked
170	50	rather
	50	they
	60	up
	60	in
	60	charge

Row	Code	Block-coding:
175	60	of
	60	a
	60	warder
	60	,
	60	and
180	60	that
	60	warder
	60	a
	60	slave
	00	.
185	01	*
	30	The
	30	ocean
	40	forbids
	50	sudden
190	50	inroads
	60	from
	60	enemies
	60	;
	60	and
195	60	,
	60	besides
	60	*
	01	,
	30	bands
200	30	of
	40	armed
	50	men
	50	,
	70	with
205	70	nothing
	01	to
	30	*
	40	do
	50	*
210	01	,
	30	easily
	40	become
	50	riotous
	01	:
215	30	it
	40	is
	50	not
	50	to
	50	the
220	50	king's
	50	interest
	60	to
	60	the
	60	king's
225	60	interest
	01	to
	30	*
	40	put
	50	a
230	50	noble
	50	or
	50	a

Row	Code	Block-coding: Ready
	50	freeman
	50	or
235	50	even
	50	a
	50	freedman
	60	in
	60	charge
240	60	of
	60	the
	60	arms
	00	.
	01	******* END OF TEXT *******

CODING
MADE UP FOR SUPPLEMENTATION

Row	Code	Supplementation: Made up
	00	****** $TART OF TEXT ******
	01	*
	30	Beyond these tribes the states of the Suiones, not on, but in, the
	30	ocean,
5	40	possess
	50	not merely arms and men but powerful fleets
	01	:
	30	the style of their ships
	40	differs
10	60	in this respect,
	00	that
	01	there
	30	there X
	40	is
15	50	a prow
	60	at each end,
	70	with a beak ready
	01	to
	30	to X
20	40P	be_driven
	50	forwards ; they
	01	neither
	30	*
	40	work
25	50	it
	70	with sails
	00	.
	01	nor
	30	nor X
30	40	add
	50	oars
	60	in banks to the side
	01	:
	30	*
35	40	gearing
	60	of the oars
	01	*
	30	X
	40P	is_detached
40	50	as on certain rivers, and reversible
	60	on certain rivers, and reversible
	01	as
	30	occasion
	40	demands
45	50	for movement in either direction
	80	for movement in either direction
	00	.
	01	*
	30	X
50	40P	is_paid
	60	Among these peoples, further, respect to wealth
	00	,
	01	and
	30	one man
55	40	is
	50	accordingly supreme,
	70	with no restrictions and with
	70	*

Row	Code	Supplementation: Made up
	01	*
60	30	*
	40	unchallenged
	50	right
	60	to obedience
	00	;
65	01	nor
	30	nor X
	40	is
	50	*
	01	there
70	30	any general
	40	carrying
	60	of arms here, as among the other Germans
	01	:
	30	X
75	40P	are_locked
	50	rather they
	60	up in charge of a warder, and that warder a slave
	00	.
	01	*
80	30	The ocean
	40	forbids
	50	sudden inroads
	60	from enemies ; and, besides
	60	*
85	01	,
	30	bands of
	40	armed
	50	men,
	70	with nothing
90	01	to
	30	*
	40	do
	50	*
	01	,
95	30	easily
	40	become
	50	riotous
	01	:
	30	it
100	40	is
	50	not to the king's interest
	60	to the king's interest
	01	to
	30	*
105	40	put
	50	a noble or a freeman or even a freedman
	60	in charge of the arms
	00	.
	01	******* END OF TEXT *******

BEST COPY AVAILABLE

Code	Supplementation: Ready
00	****** START OF TEXT ******
01	*
30	Beyond these tribes the states of the Suiones, not on, but in, the
30	ocean,
40	possess
50	not merely arms and men but powerful fleets
01	:
30	the style of their ships
40	differs
60	in this respect,
00	that
01	there
30	there X
40	is
50	a prow
60	at each end,
70	with a beak ready
01	to
30	to X
40P	be_driven
50	forwards ; they
01	neither
30	to X
40	work
50	it
70	with sails
00	,
01	nor
30	nor X
40	add
50	oars
60	in banks to the side
01	:
30	nor X
40	gearing
60	of the oars
01	*
30	X
40P	is_detached
50	as on certain rivers, and reversible
60	on certain rivers, and reversible
01	as
30	occasion
40	demands
50	for movement in either direction
80	for movement in either direction
00	.
01	*
30	X
40P	is_paid
60	Among these peoples, further, respect to wealth
00	,
01	and
30	one man
40	is
50	accordingly supreme,
70	with no restrictions and with one man right to obedience
01	*
30	one man
40	unchallenged
50	right
60	to obedience
00	:
01	nor
30	nor X
40	is
50	any general of arms here, as among the other Germans
01	there
30	any general
40	carrying
60	of arms here, as among the other Germans
01	:
30	X
40P	are_locked
50	rather they
60	up in charge of a warder, and that warder a slave
00	.
01	*
30	The ocean
40	forbids
50	sudden inroads
60	from enemies ; and, besides bands of men, with nothing
01	,
30	bands of
40	armed
50	men,
70	with nothing
01	to
30	bands of
40	do
50	easily riotous
01	,
30	easily
40	become
50	riotous
01	:
30	easily riotous
40	is
50	not to the king's interest
60	to the king's interest
01	to
30	easily riotous
40	put
50	a noble or a freeman or even a freedman
60	in charge of the arms
00	.
01	******* END OF TEXT *******

50/30-matrix   17 unique 50-rows

Row	Text
1	not merely arms and men but powerful fleets
2	a prow
3	forwards ; they
4	it
5	oars
6	as on certain rivers, and reversible
7	for movement in either direction
8	accordingly supreme,
9	right
10	any general of arms here, as among the other Germans
11	rather they
12	sudden inroads
13	men,
14	easily riotous
15	riotous
16	not to the king's interest
17	a noble or a freeman or even a freedman

30/50-matrix   11 unique 30-rows

Row	Text
1	Beyond these tribes the states of the Suiones, not on, but in, the ocean,
2	there X
3	to X
4	nor X
5	X
6	occasion
7	one man
8	The ocean
9	bands of
10	easily
11	easily riotous

50/30-matrix   Dimension: (17/11)

Row	Coordinates
1	1
2	2
3	3
4	3
5	4
6	5
7	6
8	7
9	7
10	4
11	5
12	8
13	9
14	9
15	10
16	11
17	11

30/50-matrix   Dimension: (11/17)

Row	Coordinates		
1	1		
2	2		
3	3	4	
4	5	10	
5	6	11	
6	7		
7	8	9	
8	12		
9	13	14	
10	15		
11	16	17	

30
Agent

50 Figure	1	2	3	4	5	6	7	8	9	10	11
1	1										
2		1									
3			1								
4			1								
5				1							
6					1						
7						1					
8							1				
9							1				
10				1							
11					1						
12								1			
13									1		
14									1		
15										1	
16											1
17											1

```
60/30-matrix    12 unique 60-rows
 Row Text
---- ----------------------------------------------------

   1 in this respect,
   2 at each end,
   3 in banks to the side
   4 of the oars
   5 on certain rivers, and reversible
   6 Among these peoples, further, respect to wealth
   7 to obedience
   8 of arms here, as among the other Germans
   9 up in charge of a warder, and that warder a slave
  10 from enemies ; and, besides bands of men, with nothing
  11 to the king's interest
  12 in charge of the arms
```

```
30/60-matrix    8 unique 30-rows
 Row Text
---- ---------------------------

   1 the style of their ships
   2 there X
   3 nor X
   4 X
   5 one man
   6 any general
   7 The ocean
   8 easily riotous
```

```
60/30-matrix    Dimension: (12/8)
 Row Coordinates
---- ---------------------------

   1    1
   2    2
   3    3
   4    3
   5    4
   6    4
   7    5
   8    6
   9    4
  10    7
  11    8
  12    8
```

```
30/60-matrix    Dimension: (8/12)
 Row Coordinates
---- ---------------------------

   1    1
   2    2
   3    3    4
   4    5    6    9
   5    7
   6    8
   7   10
   8   11   12
```

30
Agent

60 Ground	1	2	3	4	5	6	7	8
1	1							
2		1						
3			1					
4			1					
5				1				
6				1				
7					1			
8						1		
9				1				
10							1	
11								1
12								1

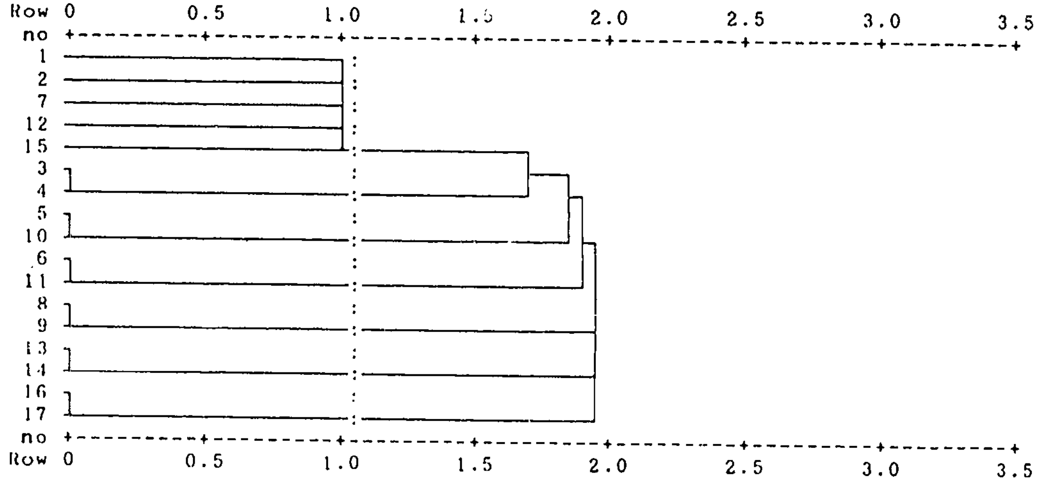
TEXT STATISTICS		Signs		Nr of		Value
	Text	775	Words	180	Extension (e)	4.0841E+06
	. ! ?	3	Blocks	21	Flow (v)	1.3614E+06
	, : ;	26	A-dum.	6	Momentum (M)	2.5990E-08
	Space	179	O-dum.	4	Force (S)	3.5381E-02
		------			Power (P)	1.1794E-02
	Total	983			Energy (E)	8.2556E-02

MATRIX STATISTICS

Matrix		Blocks Nr of	%	Dimension	Density	Comments
Figure	50/30	17	81	17/11	0.00000	17 coordinates
Ground	60/30	12	57	12/8	0.00000	12 coordinates
Means	70/30	4	19	4/4	0.00000	Diagonal, no structure
Goal	80/30	1	5	1/1		No matrix

FIGURE
CLUSTER ANALYSIS

CLUSTER TREE 5030 STEP-SCALE



5030 CLUSTERING of 17 elements by Ward's method

No of clusters (n)	Fused rows		==== STEP BY STEP ===== ESS	--- t-test --- DF=n-1	DF=15	======== TOTAL ======== ESS	--- t-test --- DF=n-1	DF=15
2	1	16	1.96	0.1706	0.0001	15.29	0.1066	0.0000
3	1	13	1.95	0.0892	0.0001	13.33	0.0513	0.0000
4	1	8	1.93	0.0521	0.0002	11.38	0.0107	0.0001
5	1	6	1.90	0.0334	0.0002	9.45	0.0533	0.0010
6	1	5	1.84	0.0250	0.0004	7.56	0.1111	0.0190
7	1	3	1.71	0.0275	0.0013	5.71	0.2902	0.1955
8	1	15	1.00	0.4433	0.4186	4.00	>0.5	>0.5
9	1	12	1.00	0.4396	0.4186	3.00		
10	1	7	1.00	0.4362	0.4186	2.00		
11	1	2	1.00	0.4329	0.4186	1.00		
12	16	17	0.00	>0.5	>0.5	0.00		
13	13	14	0.00			0.00		
14	8	9	0.00			0.00		
15	6	11	0.00			0.00		
16	5	10	0.00			0.00		
17	3	4	0.00			0.00		

CLUSTER TREE 5030 TOTAL-SCALE

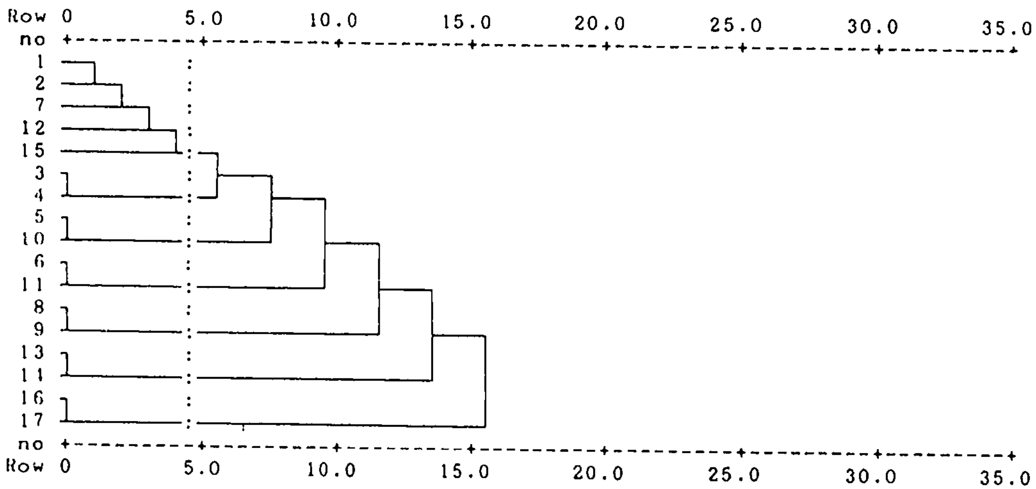


FIGURE
CLUSTERS
SYNTHESIS FROM CLUSTERS

7 clusters 5030

CLUSTER: 1 Preparedness

1 not merely arms and men but powerful fleets
2 a prow
7 for movement in either direction
12 sudden inroads
15 riotous

CLUSTER: 2 Direction

3 forwards ; they
4 it

CLUSTER: 3 Regime

5 oars
10 any general of arms here, as among the other Germans

CLUSTER: 4 Flexibility

6 as on certain rivers, and reversible
11 rather they

CLUSTER: 5 Sovreignty

8 accordingly supreme,
9 right

CLUSTER: 6 Distress

13 men,
14 easily riotous

CLUSTER: 7 Constraint

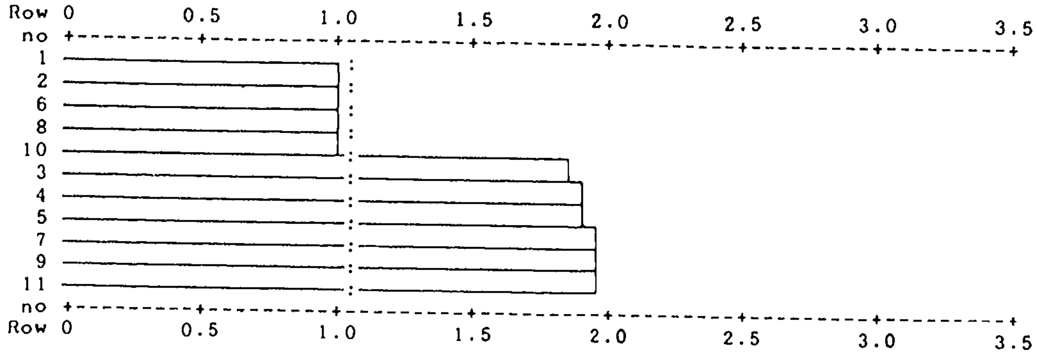
16 not to the king's interest
17 a noble or a freeman or even a freedman

7 clusters 5030

SYNTHESIS FROM CLUSTERS

1:1 Preparedness	
2:3 Direction	Determination
3:5 Regime	Domination
4:6 Flexibility	Vigilance
5:8 Sovreignty	Gating
6:13 Distress	Violation of Order
7:16 Constraint	Safety

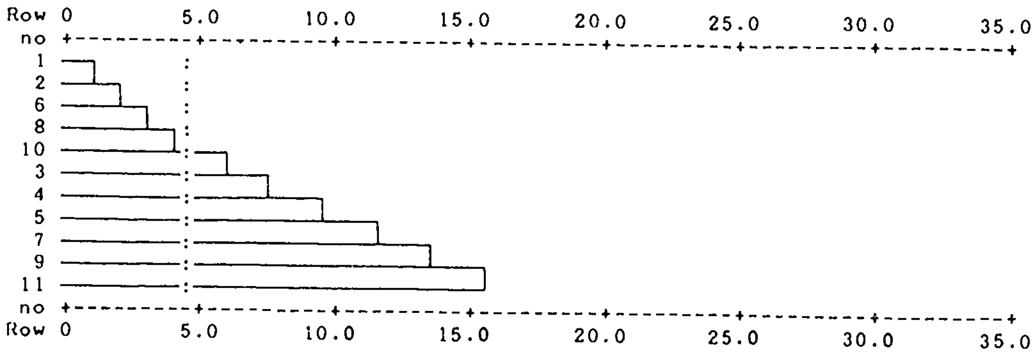
CLUSTER TREE 3050 STEP-SCALE



3050 CLUSTERING of 11 elements by Ward's method

No of clusters (n)	Fused rows		==== STEP BY STEP ==== ESS	--- t-test --- DF=n-1	DF=9	======== TOTAL ======== ESS	--- t-test --- DF=n-1	DF=9
2	1	11	1.95	0.2174	0.0113	15.45	0.1319	0.0003
3	1	9	1.94	0.1399	0.0126	13.50	0.0846	0.0019
4	1	7	1.93	0.1001	0.0147	11.56	0.0973	0.0136
5	1	5	1.91	0.0789	0.0183	9.63	0.1863	0.0949
6	1	4	1.88	0.0706	0.0253	7.71	0.4364	0.1164
7	1	3	1.83	0.0784	0.0426	5.83	>0.5	>0.5
8	1	10	1.00	>0.5	>0.5	4.00		
9	1	8	1.00			3.00		
10	1	6	1.00			2.00		
11	1	2	1.00			1.00		

CLUSTER TREE 3050 TOTAL-SCALE



PERSPECTIVE ON FIGURE
CONNECTIONS TO FIGURE CLUSTERS
SYNTHESIS FROM CLUSTERS

7 clusters 5030

SYNTHESIS FROM CLUSTERS

1:1 Preparedness	
2:3 Direction	Determination
3:5 Regime	Domination
4:6 Flexibility	Vigilance
5:8 Sovreignty	Gating
6:13 Distress	Violation of Order
7:16 Constraint	Safety

50: 7 clusters from 17 elements
30: 7 clusters from 11 elements

CONNECTIONS BETWEEN 50- AND 30-CLUSTERS

50 \ 30	1 1	2 3	3 4	4 5	5 7	6 9	7 11
1:1	5						
2:3		2					
3:5			2				
4:6				2			
5:8					2		
6:13						2	
7:16							2

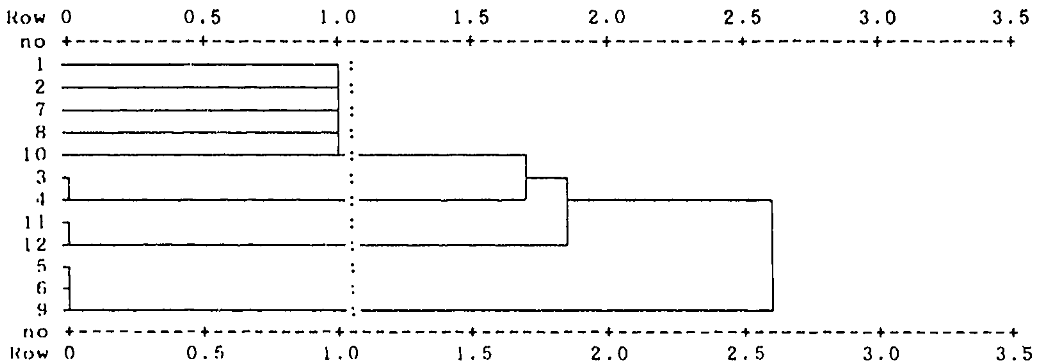
7 clusters 3050

SYNTHESIS FROM CLUSTERS

1:1 Preparedness	
2:3 Direction	Determination
3:4 Regime	Domination
4:5 Flexibility	Vigilance
5:7 Sovreignty	Gating
6:9 Distress	Violation of Order
7:11 Constraint	Safety

GROUND
CLUSTER ANALYSIS

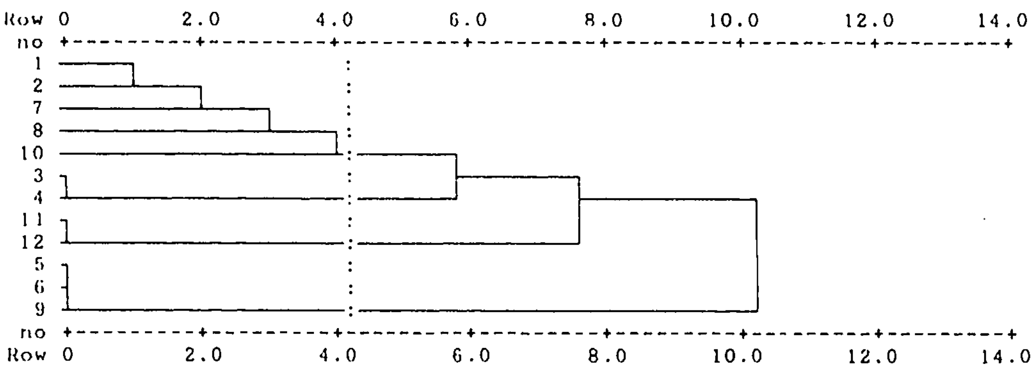
CLUSTER TREE 6030 STEP-SCALE



6030 CLUSTERING of 12 elements by Ward's method

No of clusters (n)	Fused rows		==== STEP BY STEP ==== ESS	--- t-test --- DF=n-1	DF=10	======== TOTAL ======== ESS	--- t-test --- DF=n-1	DF=10
2	1	5	2.61	0.1125	0.0000	10.17	0.1060	0.0000
3	1	11	1.84	0.1065	0.0031	7.56	0.0770	0.0008
4	1	3	1.71	0.0853	0.0069	5.71	0.1112	0.0115
5	1	10	1.00	0.4284	0.3906	4.00	0.2856	0.1911
6	1	8	1.00	0.4207	0.3906	3.00	>0.5	>0.5
7	1	7	1.00	0.4137	0.3906	2.00		
8	1	2	1.00	0.4073	0.3906	1.00		
9	11	12	0.00	>0.5	>0.5	0.00		
10	5	9	0.00			0.00		
11	5	6	0.00			0.00		
12	3	4	0.00			0.00		

CLUSTER TREE 6030 TOTAL-SCALE



4 clusters 6030

CLUSTER: 1 Governance

1 in this respect,
2 at each end,
7 to obedience
8 of arms here, as among the other Germans
10 from enemies ; and, besides bands of men, with nothing

CLUSTER: 2 Arrangement

3 in banks to the side
4 of the oars

CLUSTER: 3 Protection

11 to the king's interest
12 in charge of the arms

CLUSTER: 4 Order

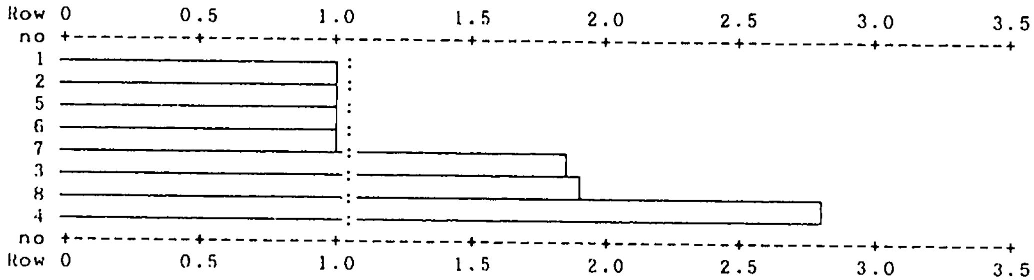
5 on certain rivers, and reversible
6 Among these peoples, further, respect to wealth
9 up in charge of a warder, and that warder a slave

4 clusters 6030

SYNTHESIS FROM CLUSTERS

1:1 Governance	
2:3 Arrangement	Regime
3:11 Protection	Defence
4:5 Order	Barrier

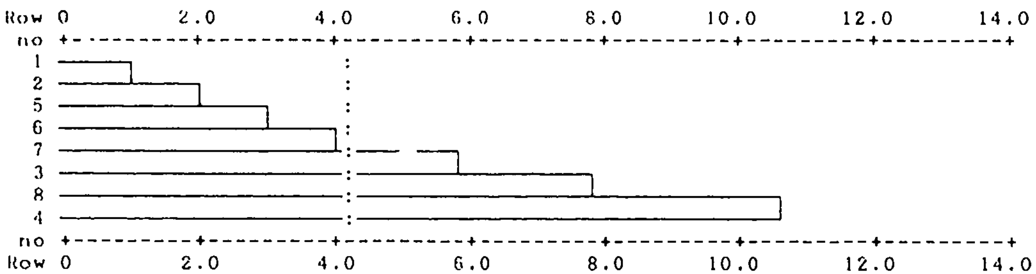
CLUSTER TREE 3060 STEP-SCALE



3060 CLUSTERING of 8 elements by Ward's method

No of clusters (n)	Fused rows		==== STEP BY STEP ===== ESS	--- t-test --- DF=n-1	DF=6	======== TOTAL ======== ESS	--- t-test --- DF=n-1	DF=6
2	1	4	2.79	0.1164	0.0014	10.50	0.1271	0.0022
3	1	8	1.88	0.2217	0.0990	7.71	0.1403	0.0333
4	1	3	1.83	0.2045	0.1261	5.83	0.3028	0.2378
5	1	7	1.00	>0.5	>0.5	4.00	>0.5	>0.5
6	1	6	1.00			3.00		
7	1	5	1.00			2.00		
8	1	2	1.00			1.00		

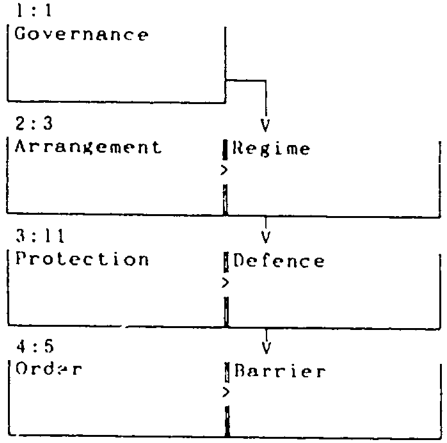
CLUSTER TREE 3060 TOTAL-SCALE



PERSPECTIVE ON GROUND
CONNECTIONS TO GROUND CLUSTERS
SYNTHESIS FROM CLUSTERS

4 clusters 6030

SYNTHESIS FROM CLUSTERS



60: 4 clusters from 12 elements
30: 4 clusters from 8 elements

CONNECTIONS BETWEEN 60- AND 30-CLUSTERS

60 \ 30	1:1	2:3	3:8	4:4
1:1	5			
2:3		2		
3:11			2	
4:5				3

4 clusters 3060

SYNTHESIS FROM CLUSTERS