ANALYSIS OF TEXTUAL DATA USING DUAL SCALING

Nicole Arnal, Direction Générale des Télécommunications
Ludovic Lebart, Centre National de la Recherche Scientifique

The needs for analysing textual data are very large in many fields like sociology, psychology and even economy and forecasting. We present a stastistical method which permits to treat open-ended questions in surveys, depth interviews, meetings texts... This method has been developed at first on responses to open-ended questions in socio-economic surveys. In order to give a better understanding of the manner in which the data are treated, we will explain here this first application, then we will give other kinds of examples.

In socio-economic surveys, the sets of response-items are often very large, diversified and sometimes not completely known. When the response-items are multiple, one open-ended question can replace several closed-ended questions. But open-ended questions are specially helpful when the matter of the survey is new and when the questionnaire is long. The respondent may express his personal opinion and it seems that less fatigue appears than in the case of closed-ended questions.

The statistical programs presented here (S.P.A.D., 1982) are applied to answers computerized in their textual form without coding. The complete process involves three steps : usual lexical statistics, graphical visualization using correspondence analysis (BENZECRI (1969), method also named : dual scaling (NISHISATO, 1981), reciprocal averaging (HIIL, 1974)) and a selection of characteristic responses for each group of individuals.

I - SELECTING PRINTINGS AND LEXICAL STATISTICS.

A basic record is composed of the number of the respondent and his complete answer to a given open-ended question. For each question, the first treatment consists in regrouping the responses according to a relevant partition for the problem under study. For example, we can analyze the living conditions in relation with occupational groups. So the sets of responses of workers, employees, project engineers... are printed. Regrouping of responses can lead to the appearance of homogeneous discourses owing to the repetition of certain topics.

In fact, responses to open-ended questions are statistical elements which can be treated as "sparse vectors". Let us consider a matrix T with n rows (n=number of individuals) and p columns (p=number of different utilized words). Each response can be described by a row of 0, 1, 2, 3... according to the absence or the presence, one, two, three... times of the words in the response. In the example, with 15 occupational groups, the initial matrix T becomes a matrix C with 15 rows and p columns. The element C (i, j) represents the number of times, the word j appears in the group i. The problem is to recognize, compute and classify all the words used. Each line of text is read letter after letter, the words being delimited by separators like blanks, commas, periods... The procedure is fast because memory space is reserved beforehand according to the length of words, up to 16 letters in this program.

The histogram shows the distribution according to the number of letters of the words in a corpus of about 40 000 French words. For other languages, it could be necessary to change these parameters.

Number of letters	Number of different words
1	30
2	90
3	150
4	250
5	350
6	420
7	450
8	420
9	350
10	250
11	150
12	100
13	60
14	50
15	40
16	20
Total :	3 180

Thus we obtain the matrix C wich is a contingency table that enables us to calculate the number of different words for all the individuals and for each group. We can compute simultaneously usual lexical statistics. Of course it is possible not to take into account the words used very few times (less than twice for example) and the words with only one or two letters. But we note that even the words like "because", "then"... are important because they often indicate argumentative replies (as opposed to short replies without argumentation).

II - FACTORIAL ANALYSIS.

In this step a correspondence analysis is performed on the contingency table C. The correspondence analysis is a data analysis technique, similar to principal component analysis.

An n x p contingency table $K = \{k_{ij}\}$ specifies the counts of joint occurrences of two discrete variables. In a contingency table, the rows and the columns represent two partitions of the same population. In order that the distances between the row elements and the column elements could be interpreted, the percentage distributions within a row or a column (i.e profiles) are used.

We note

$k = \sum_i \sum_j k_{ij}$ the total of the contingency table

$f_{ij} = k_{ij} / k$ relative frequencies

$f_{i.} = \sum_j f_{ij}$
$f_{.j} = \sum_i f_{ij}$
marginal relative frequencies

In the space R^p of n elements, the coordinates of the element i are :

$\{(f_{ij} / f_{i.}) ; j = 1, 2 \ldots p\}$ with the weight $f_{i.}$

Because we deal with profiles in the spaces R^p and R^n (the formula are symetric), we use the CHI - Square distance. (GUTTMAN (1941), BENZECRI (1969)).

In R^p, the distance between i and i'is given by:

$$d^2(i, i') = \sum_{j=1}^{p} 1 / f_{.j} (f_{ij} / f_{i.} - f_{i'j} / f_{i'.})^2$$

The solution of correspondence analysis is equivalent to a singular value decomposition problem and thus, is solved by extracting the eigen values of a positive semi-definite symetric matrix. Correspondence analysis can be generalized to process multi-way contingency tables : multiple correspondence analysis can also provide scaling of categories (of more than two questions) and individuals.

Thus, the graphical representation enables us to visualize the proximities between words, between groups and between words and groups. Nevertheless the proximities depend on the groups of responses, i. e., the choice of the initial partition. We shall give other possibilities of partition in § 4. We can also directly analyze the responses without before hand grouping (LEBART, 1982).

On the other hand, we work on words and the sentence around is not taken into account. In order to insert the words in their contexts, we introduce the concept of "modal responses".

III - MODAL OR CHARACTERISTIC RESPONSES

In analyzing the matrix C, we do not have the association between words within responses (or sentences). This notion exists in the matrix T which crosses the individuals (or responses) and the words used. But this matrix is very large and sparse and it is difficult to visualize directly all the responses of all individuals. We use an interesting property of Correspondence analysis : the procedure is performed on the aggregated matrix C and the elements of the complete matrix T are printed as "supplementary elements" on the graphics. This can be done because the matrices T and C have the same column elements.

Let us note x_i the abscissa of the response i on the axis α corresponding to the eigen value λ_α.

The response i contains $t_{i.} = \sum_j t(i,j)$ words

On this axis α, the abscissa of the word j is φ^j_α

Then the usual formula is

$$x_i = (1 / \sqrt{\lambda_\alpha}) \sum_j [t(i, j) \varphi^j_\alpha / t_{i.}]$$

It is the transition relationship or reciprocal average relationship.

Response-points and group-points are in the same space (the space spanned by the words) and so we may interpret the proximities between groups and initial responses. As the complete visualization is impossible due to the large number of responses, we select for each group, the nearest response-points of the group. The distance between response and group is calculated as CHI-Square (χ^2) distance between the rows of the matrices C and T. The distance between the response i and the group k is given by :

$$d^2(i, k) = \sum_j t../t.j (t_{ij} / t_{i.} - c_{kj} / c_{k.})^2$$

Note that t.j = C.j represents the frequency of the word j. For each k, these distances are computed and arranged in order of magnitude. The responses corresponding to the smallest distances are printed.These modal responses are real responses using the largest number of characteristic words of the group.

Of course, other criteria can be used to select modal responses. For example, the most utilized words of each group can be used or the relative frequency of a word in a group of individuals can be compared to its relative frequency in the population.

IV - INSTRUMENTAL PARTITIONS

Because of its size, the initial matrix T has to be aggregated and we need to choose a partition of the individuals. One simple way is to consider a partition by means of a qualitative variable like occupational groups for example. If no partition has to be privilegied a priori, we perform a cluster analysis with the maximum of factual variables. The constraint is a minimum number of individuals in each group. The clusters are considered as an instrument, but not as a result in itself. However, the problem of groups do not exist in every application. In the analysis of textual data of meetings, the number of groups are actually small.

V - PRESENT RESEARCH

The research is now being conducted to compare the effects of different modes of communication upon the user's behaviour. The corpus concern in-person meetings, audio and video teleconferences. Now, most of the available corpus are real meetings (face to face audio and video) wich were recorded in companies. Five experiments were also conducted with teachers in the primary school.

For each meeting, there was at least one cine-camera in every room (for example, at least two for an audioconference). For each corpus, we have the complete transcription of the texts from one or two video-tapes.

At first, the statistical programs are applied to the textually transcripted data in order to compare the effects of different media on verbal communication. The automatic process enables us to treat a large number of corpus and in so doing a large number of tasks

(problem-solving, information-exchange). Thus it seems possible to cut out the effects due to the personality of individuals or to the peculiarity of tasks.

The usual lexical statistics are here the speakers'turns, the simultaneous speech events, the numbers of words of each turn, of each speaker. For the visualization through correspondence analysis, the groups can be the different media or the different speakers in order to know if they have specific influence upon the pattern of meetings.

We equally plan to search for a convenient coding in order to analyze non verbal part of the communication (gestures, body posturing, head-nods). In fact, the principle is the same when we consider the succession of codes as textual data. The best thing would be to analyze both the verbal and the non verbal communication in a parallel direction.

REFERENCES.

BENZECRI, J.P. (1969). Statistical analysis as a tool to make patterns emerge from data. In Methodologies of Pattern Recognition (S. Watanabe, ed.), pp. 35-60. New York : Academic.

GUTTMAN, L. (1941). The quantification of a class of attributes : a theory and method of scale construction. In The Prediction of Personal Adjustment (P. Horst, ed.), pp. 251-364. New York : Social Science Research Council.

HILL, M.O. (1974). Correspondence Analysis : A Neglected Multivariate Method. Institute of Terrestrial Ecology, Bangor Research Station. Appl. Statist., pp. 340-354.

LEBART, L. (1982). Exploitation analysis of large sparse matrices with application to textual data. In Proceedings in Computational Statistics. (Compstat, Sept. 1982, Toulouse.) Wien : Physica-Verlag.

NISHISATO,S. (1980). Analysis of categorical data : dual scaling and its applications. University of Toronto Press.

S.P.A.D. (1982). (Portable system for data analysis) by L. LEBART and A. MORINEAU. Program available at CESIA, 82, rue de SEVRES 75007 PARIS.

A SHORT EXAMPLE

ANALYSIS OF MICRO-COMPUTER BRANDS

(advertisings in newspapers and magazines)

```
** TANDY TRS - 80 **
1     1  FINANCIAL PLANNING AND FORECASTING MADE EASY
1     2  NORMAL PRICE L 1766 95
1     3  A COMPLETE DESKTOP BUSINESS COMPUTER INCLUDING FINANCIAL PLANNING SOFTWARE
1     4  SOFTWARE AVAILABLE FOR ACCOUNTING WORD PROCESSING DATA PROCESSING
1     5  BUILT IN EXTRAS INCLUDE TWO DISK DRIVES UPPER LOWER CASE MONITOR AND MUCH MORE
1     6  FORECAST THE EFFECTS OF BUSINESS CHANGES IN SECONDS INSTEAD OF HOURS
1     7  USING THE MICROCOMPUTER
1     8  IT STIMULATES A GIANT COLUMNAR PAD CHANGE ANY VARIABLE
1     9  AND ALL CALCULATED ENTRIES ARE AUTOMATICALLY AND INSTANTLY UPDATED
1    10  AND YOU CAN ADD A PRINTER ANYTIME
1    11  YOU CAN ALSO USE YOUR MODEL FOR A WIDE VARIETY OF USES SUCH AS WORD PROCESSING
1    12  ACCOUNTING REPORTING TIME ACCOUNTS INFORMATION MANAGEMENT AND MUCH MUCH MORE
1    13  PHONE ONE OF OUR CENTRES LISTED BELOW FOR A DEMONSTRATION

** IBM PERSONAL COMPUTER **
2     1  HOW TO CHANGE A FORECAST
2     2  TO HELP WEATHER THE STORM OF ECONOMIC VARIABLES
2     3  A PERSON COULD USE THE PERSONAL COMPUTER
2     4  WITH SOFTWARE LIKE REALLY AN ELECTRONIC WORKSHEET
2     5  YOU CAN CALCULATE UP TO 63 COLUMNS AND 254 ROWS OF NUMBERS
2     6  IMPLEMENTING FORMULAS AND CHANGING LABELS AS YOU GO
2     7  YOU CAN ALSO PLAN ON THE PERSONAL COMPUTER TO HELP CREATE A SALES FORECAST
2     8  SPOT A TREND
2     9  TEST A BUDGET
2    10  AND AID YOU ON THE QUEST FOR THE RIGHT ANSWER TO WHAT IF
2    11  NOW DON T WAIT FOR A RAINY DAY TO VISIT AN AUTHORIZED PERSONAL COMPUTER DEALER
2    12  YOU LL LEARN THAT THE QUALITY POWER AND PERFORMANCE OF THIS TOOL
2    13  ARE WHAT YOU D EXPECT FROM
2    14  A TOOL FOR MODERN TIMES

** CADO **
3     1  ALL SMALL COMPUTERS LOOK ALIKE BUT
3     2  ALL DESKTOP BUSINESS AND PERSONAL COMPUTERS HAVE A KEYBOARD VIDEO TERMINAL
3     3  A PROCESSOR AND DISK DRIVES
3     4  MOST OFFER A SINGLE TASKING OPERATING SYSTEM
3     5  AND NOW WE RE ALL PRICED ABOUT THE SAME
3     6  CONFUSED YOU MUST BE
3     7  SOLUBLE OF COURSE
3     8  NOW READ ON
3     9  THE PRICE ISN T
3    10  IT ISN T LIMITED TO FLOPPY DISKS LIKE MOST SYSTEMS
3    11  IN FACT WITH OPTIONAL DESK TOP STORAGE DEVICES
3    12  YOU CAN HAVE UP TO 15 MILLION CHARACTERS OF HIGH SPEED COMPACTED FILES
3    13  THAT S EQUIVALENT TO 45 MILLION CHARACTERS OF CONVENTIONAL FILES
3    14  TERMINALS OPERATE AT A SPEED OF 19 2 KBPS TWICE THE INDUSTRY NORM
3    15  ONLY OFFERS THE ENGLISH LANGUAGE INQUIRY METHOD
3    16  THAT LETS YOU GET REPORTS FROM YOUR COMPUTER WITHOUT EXPENSIVE PROGRAMMERS
3    17  THIS FEATURE ALONE WILL SAVE THOUSANDS OVER THE LIFE OF YOUR SYSTEM
3    18  ONLY IT HAS COMPUTER AIDED TUTORIALS
3    19  THAT TEACH BASIC ACCOUNTING AND WORD PROCESSING
3    20  WHILE YOUR PEOPLE LEARN TO USE THE SOFTWARE
3    21  PROFESSIONALS OFFER YOU SALES SERVICE AND SUPPORT IN ONE LOCATION
3    22  AND THEY CAN PROVIDE SOFTWARE PACKAGES FOR HUNDREDS OF DIFFERENT BUSINESSES
3    23  AND PROFESSIONS FIELD PROVEN IN OVER 10 000 OFFICES WORLDWIDE
3    24  LIKE IT OR NOT ATTEMPTING TO INSTALL A BUSINESS SYSTEM
3    25  WITHOUT PROFESSIONAL HELP IS AKIN TO DO IT YOURSELF BRAIN SURGERY
3    26  THE BOTTOM BOTTOM LINE
3    27  THE PURCHASE PRICE OF MOST DESKTOP SYSTEMS IS ONLY A DOWN PAYMENT AN ENTRY FEE
3    28  COSTS WILL CONTINUE TO MOUNT FOR YEARS
3    29  A SYSTEM IS THE LOWEST COST MOST RESPONSIVE SYSTEM AVAILABLE TODAY
3    30  ITS UNIQUE DESIGN FEATURES MAKE IT POSSIBLE FOR ITS MICRO LOGIC
3    31  TO COMPETE WITH SYSTEMS AS LARGE AS SYSTEM AND WIN
3    32  WANT TO KNOW MORE SEND IN THE COUPON NOW


** SORD M23 MARK III ***
4     1  SUPER EASY EVEN THE FIRST TIME THE PORTABLE
4     2  IF YOU THINK SPECIAL TRAINING IS NECESSARY TO BECOME A COMPUTER OPERATOR
4     3  THINK AGAIN
4     4  MICROCOMPUTER HAS ORIGINAL PAN INFORMATION PROCESSING SYSTEM SOFTWARE BUILT IN
4     5  THIS ELIMINATES THE NEED TO PROGRAM AND MAKES YOU AN EXPERT IN ALMOST NO TIME
4     6  NO PROGRAMMING SOFTWARE IN THE COMPACT
4     7  LIGHTWEIGHT MEANS
4     8  ALL STAFF MEMBERS CAN EASILY HANDLE ROUTINE BUSINESS DATA PROCESSING
4     9  SIMPLY TYPE OUT EASY COMMANDS AND DOES THE REST
4    10  ENTERS NEW DATA EDITS CALCULATES RETRIEVES SEARCHES
4    11  SORTS AND GRAPHIC DISPLAY OF ANALYSIS DATA
4    12  MAIN FEATURES
4    13  LARGEST STANDARD MEMORY 128 BYTE WITH 64 BIT DYNAMIC CHIPS
4    14  COMPACT PORTABILITY PERMITS OPERATION ANYWHERE
4    15  WITH OPTIONAL CRYSTAL DISPLAY CARTRIDGES AND BATTERY PACK
4    16  GRAPHIC FUNCTION
4    17  WIDE RANGE SOFTWARE AND WORD PROCESSING
4    18  VIRTUALLY UNLIMITED EXPANDABILITY
4    19  A CHOICE OF MICROFLOPPY MINIFLOPPY 8 FLOPPY AND 5 HARD DISKS AND MANY CARTRIDGES

** COMPUTERLAND **
5     1  SOMEDAY YOU LL OWN A COMPUTER
5     2  SOMEDAY MAY BE SOONER THAN YOU THINK
5     3  BECAUSE AFFORDABLE EASY TO OPERATE SMALL COMPUTERS ARE HERE NOW
5     4  AND THE USES OF THESE AMAZING MACHINES
5     5  IN BUSINESS AND SCIENCE IN THE HOME IN EDUCATION
5     6  ARE AS BOUNDLESS AS THE HUMAN IMAGINATION
5     7  NO ONE KNOWS SMALL COMPUTERS BETTER THAN THE EXPERTS AT COMPUTER LAND
5     8  TO LEARN HOW A COMPUTER COULD CHANGE THE WAY YOU LIVE AND WORK VISIT TODAY
5     9  COMPUTER LAND WE KNOW SMALL COMPUTERS LET US INTRODUCE YOU

** ICL PERSONAL COMPUTER **
6     1  YOU LL BE GLAD YOU CHOSE A PERSONAL COMPUTER
6     2  WITH POWER ENOUGH TO GROW WITH YOUR NEEDS
6     3  CHOOSING THE WRONG PERSONAL COMPUTER CAN WORK OUT VERY COSTLY
6     4  BECAUSE THOUGH THEY GROW TO MEET YOUR NEEDS THEY JUST DON T GROW ENOUGH
6     5  THE NEW PERSONAL COMPUTER DIVES YOU MORE
6     6  IT S A VERSATILE PROFESSIONAL PERSONAL COMPUTER
6     7  IT CAN START WITH A SINGLE VISUAL DISPLAY UNIT NATURALLY
6     8  AND IT HAS A WIDE RANGE OF FUNCTIONS TO MEET YOUR INCREASING REQUIREMENTS
6     9  AND IT CAN GROW LARGER THAN MOST OTHER PERSONAL COMPUTER SYSTEMS
6    10  BECAUSE ITS CAPACITY FOR ADDITIONAL HARDWARE IS GREATER
6    11  THE PERSONAL COMPUTER PROVIDES A RANGE OF OPTIONS
6    12  ENABLING YOU TO HAVE A SYSTEM TAILORED TO MEET YOUR CHANGING NEEDS
6    13  FOR EXAMPLE BY ADDING A FIXED DISC
6    14  YOU CAN HAVE EIGHT TIMES THE ORIGINAL STORAGE CAPACITY
6    15  AND SUPPORT UP TO 4 VISUAL DISPLAY UNITS
6    16  AND THERE IS AN EXTENSIVE RANGE OF PERSONAL COMPUTER HARDWARE
6    17  WIDE RANGE OF APPLICATION SOFTWARE AVAILABLE PROVIDING LIMITLESS POSSIBILITIES
6    18  SO YOUR SECRETARY CAN DO HER WORD PROCESSING YOU CAN DO YOUR FORECASTS
6    19  AND YOUR ACCOUNTANT CAN DO THE INVOICING ALL AT THE SAME TIME
6    20  THAT S WHAT MAKES THE PERSONAL COMPUTER MORE THAN JUST A PERSONAL COMPUTER
6    21  AND BECAUSE IT IS EUROPE S LEADING COMPUTER COMPANY
6    22  WITH THOUSANDS OF SATISFIED USERS IN OVER 80 COUNTRIES WORLDWIDE
6    23  WE CAN OFFER TRADER POINT SERVICE BACK UP THAT S SECOND TO NONE
6    24  WHAT MORE COULD YOU ASK FOR APART FROM A DEMONSTRATION

** SINCLAIR ZX81 **
7     1  FOR 5 99 95 YOU CAN HAVE A FULL POWERED PERSONAL COMPUTER
7     2  MOST PEOPLE KNOW BY NOW
7     3  THAT IT IS THE LOWEST PRICED PERSONAL COMPUTER IN THE WORLD
7     4  BUT SERIOUS PROGRAMMERS ARE LOOKING FOR MORE THAN A LOW PRICE
7     5  THEY RE LOOKING FOR TRUE COMPUTER POWER
7     6  AND THAT S WHERE IT SURPRISES A LOT OF PEOPLE
7     7  JUST LOOK AT THE KEYBOARD AND YOU LL GET SOME IDEA OF THE POWER
7     8  IT HAS MORE THAN 60 COMMANDS 20 GRAPHIC SYMBOLS
7     9  AND COMPLETE MATHEMATICAL FUNCTIONS
7    10  AND THERE S EVEN MORE POWER THAT YOU CAN T SEE
7    11  A BREAKTHROUGH IN PERSONAL COMPUTERS
7    12  IT OFFERS FEATURES FOUND ONLY ON COMPUTERS COSTING TWO OR THREE TIMES AS MUCH
7    13  JUST LOOK AT WHAT YOU GET
7    14  CONTINUOUS DISPLAY INCLUDING MOVING GRAPHICS
7    15  MULTI DIMENSIONAL STRING AND NUMERICAL ARRAYS
7    16  MATHEMATICAL AND SCIENTIFIC FUNCTIONS ACCURATE TO 8 DECIMAL PLACES
7    17  UNIQUE ONE TOUCH ENTRY OF KEY WORDS LIKE PRINT RUN AND LIST
7    18  AUTOMATIC SYNTAX ERROR DETECTION AND EASY EDITING
7    19  RANDOMIZE FUNCTION USEFUL FOR BOTH GAMES AND SERIOUS APPLICATIONS
7    20  BUILT IN INTERFACE FOR PRINTER
7    21  1K OF MEMORY EXPANDABLE TO 16K
7    22  A COMPREHENSIVE PROGRAMMING GUIDE AND OPERATING MANUAL
7    23  ORDER NOW AND TRY IT OUT FOR 10 DAYS
7    24  SIMPLY SEND THE COUPON ALONG WITH A CHECK OR MONEY ORDER
```

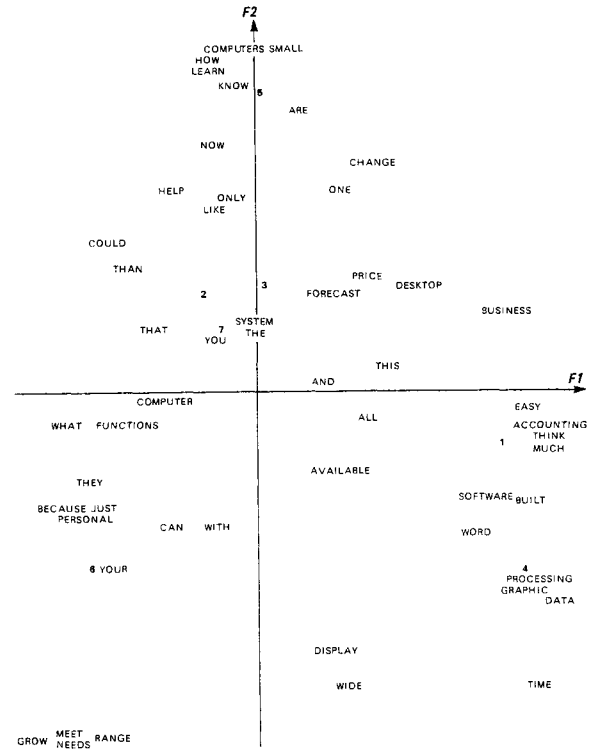
CONTINGENCY TABLE

		1	2	3	4	5	6	7
1	AND	6.	4.	8.	7.	3.	6.	11.
2	FOR	3.	3.	3.	0.	0.	3.	6.
3	THE	2.	6.	10.	5.	5.	7.	5.
4	ALL	1.	0.	3.	1.	0.	1.	0.
5	ARE	1.	1.	0.	0.	2.	0.	1.
6	YOU	2.	6.	4.	2.	4.	7.	4.
7	CAN	2.	2.	2.	1.	0.	0.	2.
8	USE	1.	1.	1.	0.	0.	0.	0.
9	ONE	1.	0.	1.	0.	1.	0.	1.
10	HOW	0.	1.	1.	0.	1.	0.	0.
11	NOW	0.	1.	2.	0.	1.	0.	2.
12	GET	0.	0.	1.	0.	0.	0.	2.
13	HAS	0.	0.	1.	1.	0.	1.	1.
14	ITS	0.	0.	2.	0.	0.	1.	0.
15	OUT	0.	0.	0.	1.	0.	1.	1.
16	EASY	1.	0.	0.	2.	1.	0.	1.
17	WORD	2.	0.	1.	1.	0.	1.	0.
18	DATA	1.	0.	0.	3.	0.	0.	0.
19	MUCH	3.	0.	0.	0.	0.	0.	1.
20	MORE	2.	0.	1.	0.	0.	3.	3.
21	YOUR	1.	0.	3.	0.	0.	7.	0.
22	WIDE	1.	0.	0.	1.	0.	2.	0.
23	TIME	1.	0.	0.	2.	0.	1.	0.
24	HELP	0.	2.	1.	0.	0.	0.	0.
25	WITH	0.	1.	2.	2.	0.	4.	1.
26	LIKE	0.	1.	2.	0.	0.	0.	1.
27	WHAT	0.	2.	0.	0.	0.	2.	1.
28	THAT	0.	1.	3.	0.	0.	2.	3.
29	THIS	0.	1.	1.	1.	0.	0.	0.
30	FROM	0.	1.	1.	0.	0.	1.	0.
31	LOOK	0.	0.	1.	0.	0.	0.	2.
32	HAVE	0.	0.	2.	0.	0.	2.	1.
33	MOST	0.	0.	4.	0.	0.	1.	1.
34	ONLY	0.	0.	3.	0.	0.	0.	1.
35	OVER	0.	0.	2.	0.	0.	1.	0.
36	THEY	0.	0.	1.	0.	0.	2.	1.
37	KNOW	0.	0.	1.	0.	1.	0.	1.
38	THAN	0.	0.	0.	0.	2.	2.	2.
39	GROW	0.	0.	0.	0.	0.	4.	0.
40	MEET	0.	0.	0.	0.	0.	3.	0.
41	JUST	0.	0.	0.	0.	0.	2.	2.
42	PRICE	1.	0.	2.	0.	0.	0.	1.
43	BUILT	1.	0.	0.	1.	0.	0.	1.
44	COULD	0.	1.	0.	0.	1.	1.	0.
45	LEARN	0.	1.	1.	0.	1.	0.	0.
46	POWER	0.	1.	0.	0.	0.	1.	3.
47	TIMES	0.	1.	0.	0.	0.	1.	1.
48	SMALL	0.	0.	1.	0.	3.	0.	0.
49	OFFER	0.	0.	2.	0.	0.	1.	0.
50	THINK	0.	0.	0.	2.	1.	0.	0.
51	RANGE	0.	0.	0.	1.	0.	4.	0.
52	NEEDS	0.	0.	0.	0.	0.	3.	0.
53	CHANGE	1.	1.	0.	0.	1.	0.	0.
54	SYSTEM	0.	0.	6.	1.	0.	1.	0.
55	PEOPLE	0.	0.	1.	0.	0.	0.	2.
56	DESKTOP	1.	0.	2.	0.	0.	0.	0.
57	SYSTEMS	0.	0.	3.	0.	0.	1.	0.
58	GRAPHIC	0.	0.	0.	2.	0.	0.	1.
59	DISPLAY	0.	0.	0.	2.	0.	2.	1.
60	BECAUSE	0.	0.	0.	0.	1.	3.	0.
61	BUSINESS	2.	0.	2.	1.	1.	0.	0.
62	COMPUTER	1.	3.	2.	1.	4.	10.	3.
63	SOFTWARE	2.	1.	2.	3.	0.	1.	0.
64	FORECAST	1.	2.	0.	0.	0.	0.	0.
65	PERSONAL	0.	3.	1.	0.	0.	9.	3.
66	FEATURES	0.	0.	1.	1.	0.	0.	1.
67	AVAILABLE	1.	0.	1.	0.	0.	1.	0.
68	COMPUTERS	0.	0.	2.	0.	3.	0.	2.
69	FUNCTIONS	0.	0.	0.	0.	0.	1.	2.
70	ACCOUNTING	3.	0.	1.	0.	0.	0.	0.
71	PROCESSING	3.	0.	1.	3.	0.	1.	0.

Number of letters	1	2	3	4	5	6	7	8	9	10	11	12	13
Number of different words used	0	0	48	86	87	55	67	45	43	23	12	7	7

Total number of words	957
Number of different words	480
Number of different words analysed (words of three letters or more and words used twice at least)	71

FACTORIAL PLANE 1-2



The figure above shows the factorial plane corresponding to the two first axes. This display points out the association between words, between brands and between words and brands. Only the words with the most important contributions (scores) were printed.

The first axis (horizontal, 24 % of total variance) opposes two kinds of professional advertisings. On the lower left hand corner, the advertising insists on the fact that it is "your personal computer"and it can "grow to meet your needs". On the lower right hand corner, the advertising deals with the easiness and the rapidity of training and with the uses for "accounting word processing, data processing". The second axis (vertical, 21 % of total variance) opposes the professional use to the public at large use. In this case (upper part), the discourse is "to learn how computers could change the way you live".The underlined sentences are precisely part of modal responses.

We note that two different lexical forms of a word can play different roles according to the context. It is the case for "computer" (your computer at work) and "computers" (the computers, generally speaking).

The example presented here to illustrate the program, is very short but this statistical program is of great use specially for large corpus.