Workshop on Data Analysis and Classification (DAC)
In honor of Edwin Diday

September 4, 2007
Conservatoire National des Arts et Métiers (CNAM)

Stability of Principal Axes

Ludovic Lebart,
National Center for Scientific Research (CNRS)
ENST, Paris, France.
lebart@enst.fr

Stability of Principal Axes

1 Introduction : Visualisations through principal axes and bootstrap

2 Partial bootstrap

3 Total bootstrap: principles and 3 examples

4 Other types of bootstrap

1. Introduction: visualisations through principal axes and bootstrap *(a reminder)*

- 1.1. The deadlock of analytical solutions
- 1.2. Resampling solutions

1.1 The deadlock of analytical validation

Distribution of eigenvalues. PCA case.

matrix **S** = **X'X** ($p(p+1)/2$ distinct elements)

Wishart, $W(p,n,\Sigma)$ whose density $f(\mathbf{S})$ is :

$$f(\mathbf{S}) = C(n,p,\Sigma)\,|\mathbf{S}|^{-\frac{n-p-1}{2}}\,exp\left\{-\frac{1}{2}trace(\Sigma^{-1}\mathbf{S})\right\}$$

$$C(n,p,\Sigma) = 2^{-\frac{np}{2}}\,|\Sigma|^{-\frac{n}{2}}\,\pi^{-\frac{p(p-1)}{4}}\prod_{k=1}^{p}\Gamma\left(\frac{1}{2}(n+1-k)\right)$$

Distribution of eigenvalues *(continuation)*

Distribution of Eigenvalues from a Wishart matrix:
Fisher (1939), Girshick (1939), Hsu (1939) and Roy (1939), then Mood (1951). Anderson (1958).

$$f(\mathbf{S}) = C(n,p,\mathbf{I}) \left(\prod_{k=1}^{p} l_k \right)^{-\frac{n-p-1}{2}} exp\left\{ -\frac{1}{2} \sum_{k=1}^{p} l_k \right\}$$

(If $\Sigma = I$)

$$g(\mathrm{L}) = D(n,p) \left(\prod_{k=1}^{p} l_k \right)^{-\frac{n-p-1}{2}} exp\left\{ -\frac{1}{2} \sum_{k=1}^{p} l_k \right\} \prod_{k<j}^{p} (l_k - lj)$$

Case of largest eigenvalues:
Pillai (1965), Krishnaiah et Chang (1971), Mehta (1960, 1967)

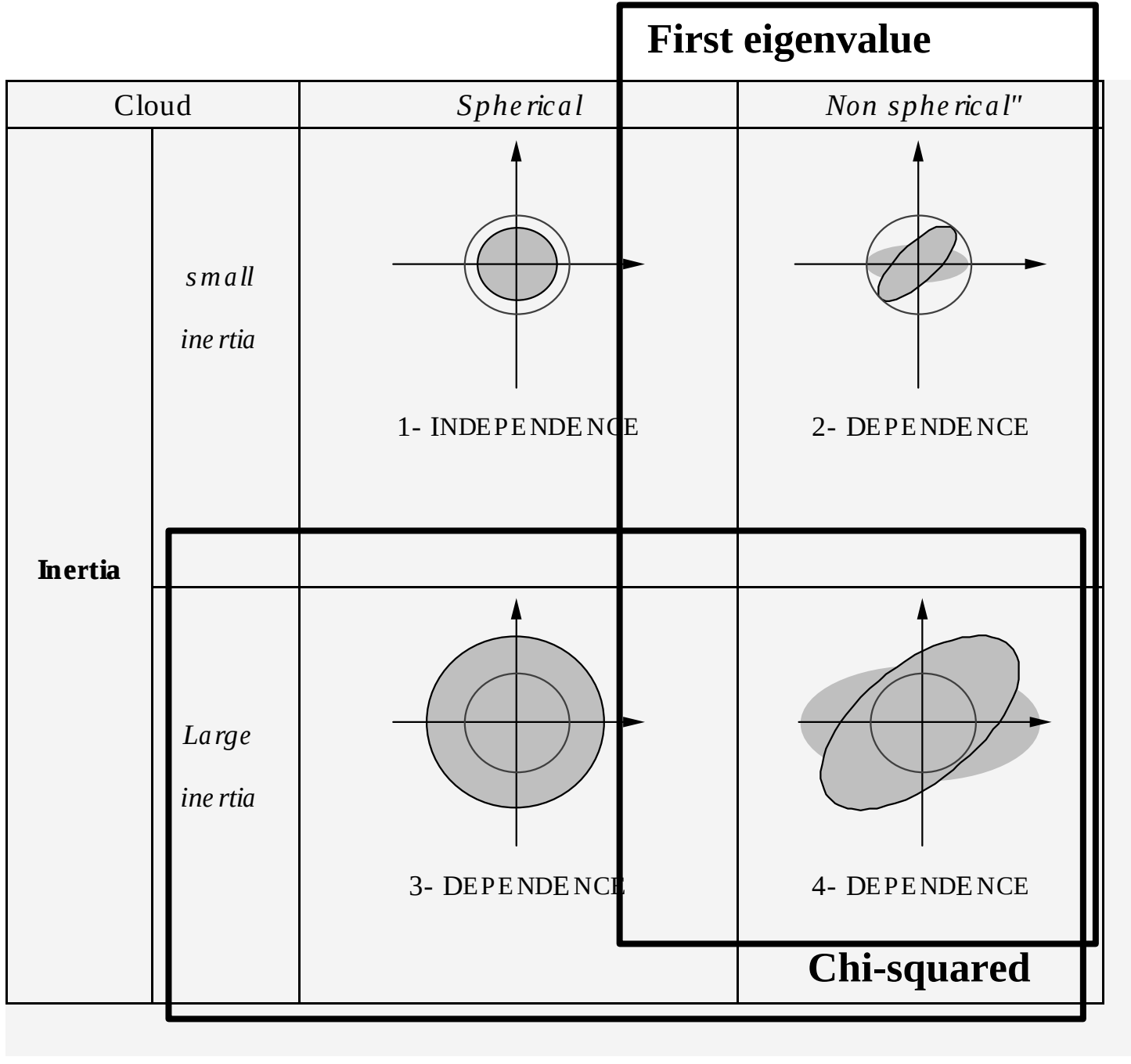
In practice, all these results are both unrealistic and unpractical

Distribution of eigenvalues. CA and MCA cases.

In Correspondence analysis, for a contingency table (n, p), the eigenvalues are those obtained from a Wishart matrix : W (n-1, p-1)

As a consequence, under the hypothesis of independence, the percentage of variance are independent from the trace, which is the usual chi-square with (n-1, p-1) degrees of freedom.

However, in the case of Multiple Correspondence Analysis, or in the case of binary data, the trace has not the same meaning, and the percentages of variance are misleading measure of information.


First eigenvalue
Cloud
Spherical
Non spherical"
Inertia
small
inertia
1- INDEPENDENCE
2- DEPENDENCE
Large
inertia
3- DEPENDENCE
4- DEPENDENCE
Chi-squared

Quality of the structural compression of data

Approximation formula

$$\mathbf{X}^{*} = \sum_{a=1}^{q} \sqrt{l_a}\ \mathbf{v}_a \mathbf{u}_a^{'} \qquad \text{with} \quad q < p$$

(Compression formula)

Measurement of the quality of the approximation

$$t_q = \frac{\sum_{a=1}^{q} l_a}{\sum_{a=1}^{p} l_a} = \frac{tr\{\mathbf{X}^{*'}\mathbf{X}^{*}\}}{tr\{\mathbf{X}^{'}\mathbf{X}\}} = \frac{\sum_{i,j=1}^{p} (\mathrm{x}_{ij}^{*})^2}{\sum_{i,j=1}^{p} (\mathrm{x}_{ij})^2}$$

Other tools for internal validation

Stability (Escofier and Leroux, 1972)

Sensitivity (Tanaka, 1984)

Confidence zones using Delta method (Gifi, 1990)
.

I.2. Resampling techniques: Bootstrap, opportunity of the method

- In order to compute estimates precision, many reasons lead to the Bootstrap method :
  - highly complex computation in the analytical approach
  - to get free from beforehand assumptions
  - possibility to master every statistical computation for each sample replication
  - no assumption about the underlying distributions
  - availability of cumulative frequency functions, which offers various possibilities

Reminder about Bootstrap Method
An example : Confidence areas in statistical mappings.

- The mappings used to visualise multidimensional data (through *Multidimensional Scaling*, *Principal Component Analysis* or *Correspondence Analysis*) involve complex computation.
- In particular, variances of the locations of points on mappings cannot be easily computed.
- The seminal paper by Diaconis and Efron in *Scientific American (1983)* *Computer intensive methods in statistics* precisely dealt with a similar problem in the framework of *Principal Component Analysis*.

2. Partial bootstrap

2.1 Reminder of bootstrap

2.2 Principle of partial bootstrap

2.3 Simple example

CA and MCA cases

Gifi (1981), Meulman (1982), Greenacre (1984) did pioneering work in addressing the problem in the context of two-way and multiple correspondence analysis.

It is easier to assess eigenvectors than eigenvalues that are much more sensitive to data coding, the replicated eigenvalues being biased replicates of the theoretical ones.

2.1 Reminder about the bootstrap

Contingency table,
592 women: Hair and eyes color.

Eye color	**Hair color**				
	black	brown	red	blond	Total
black	*68*	*119*	*26*	*7*	*220*
hazel	*15*	*54*	*14*	*10*	*93*
green	*5*	*29*	*14*	*16*	*64*
blue	*20*	*84*	*17*	*94*	*215*
Total	*108*	*286*	*71*	*127*	*592*

Source : Snee (1974), Cohen(1980)

Visualisation of associations between eye and hair color

[*Correspondence analysis*] Example of replicated tables

Original

		Hair color Black	Brown	red	blonde
eye color	black	*68*	*119*	*26*	*7*
	hazel	*15*	*54*	*14*	*10*
	green	*5*	*29*	*14*	*16*
	blue	*20*	*84*	*17*	*94*

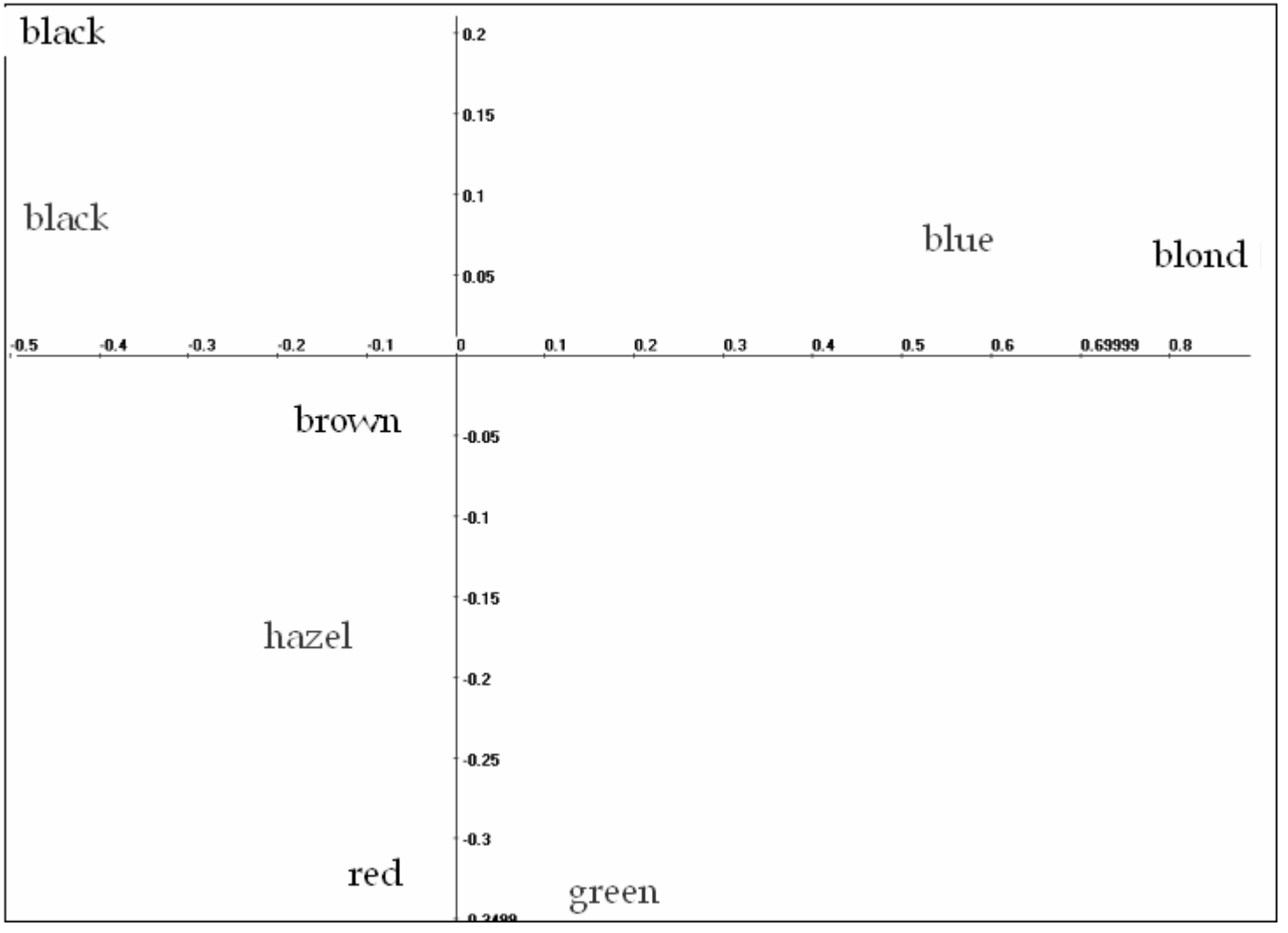
Replicate 1

eye color	black	*79*	*120*	*23*	*9*
	hazel	*14*	*60*	*15*	*12*
	green	*3*	*29*	*16*	*9*
	blue	*21*	*82*	*20*	*110*

Replicate 2

eye color	black	*72*	*111*	*32*	*7*
	hazel	*14*	*47*	*13*	*14*
	green	*5*	*30*	*15*	*19*
	blue	*20*	*89*	*16*	*98*

Principal plane (1, 2)
Snee data. Hair - Eye



2.2 Principle of partial bootstrap

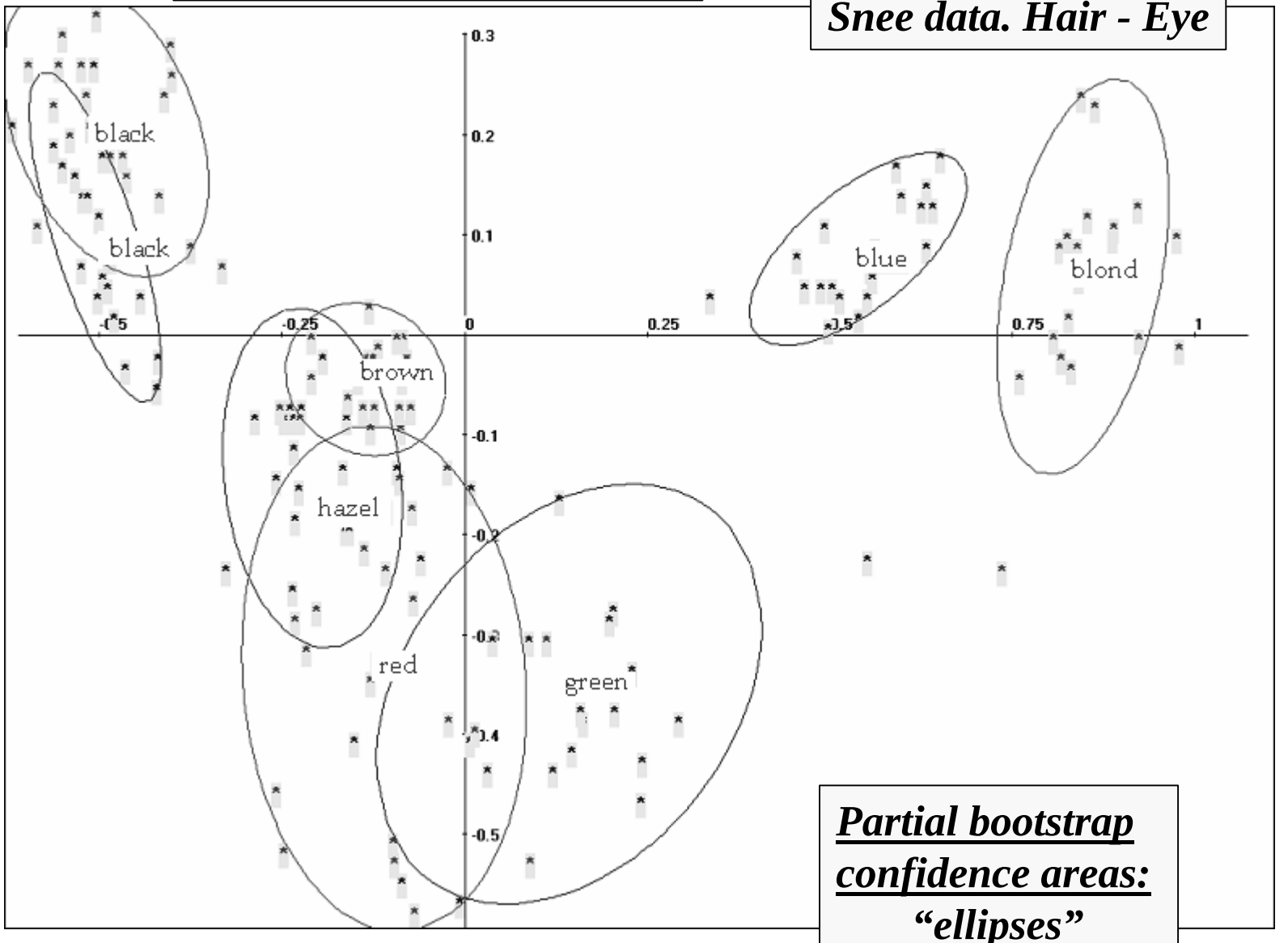
The partial bootstrap, makes use of simple *a posteriori* projections of replicated elements on the original reference subspace provided by the eigen-decomposition of the observed covariance matrix.

From a descriptive standpoint, this initial subspace is better than any subspace undergoing a perturbation by a random noise. In fact, this subspace is the expectation of all the replicated subspaces having undergone perturbations (however, the original eigenvalues are not the expectations of the replicated values).

The plane spanned by the first two axes, for instance, provides an optimal point of view on the data set.

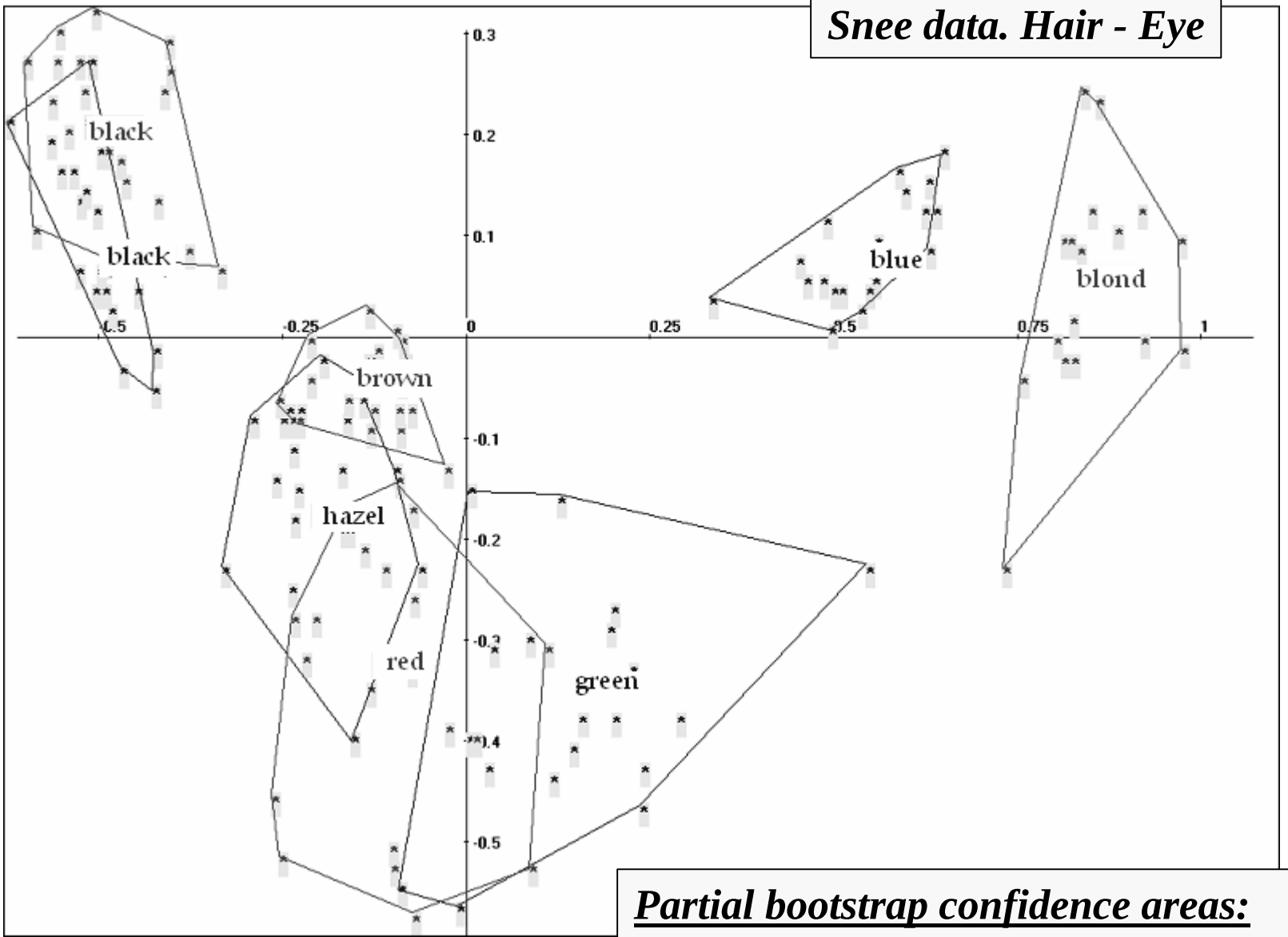
2.3 Simple example

Principal plane (1, 2)
Snee data. Hair - Eye



<u>Partial bootstrap confidence areas:</u>
"ellipses"

Principal plane (1, 2)
Snee data. Hair - Eye



<u>Partial bootstrap confidence areas:</u>
"convex hulls"

3. Total bootstrap...

3.1 Total bootstrap type 1

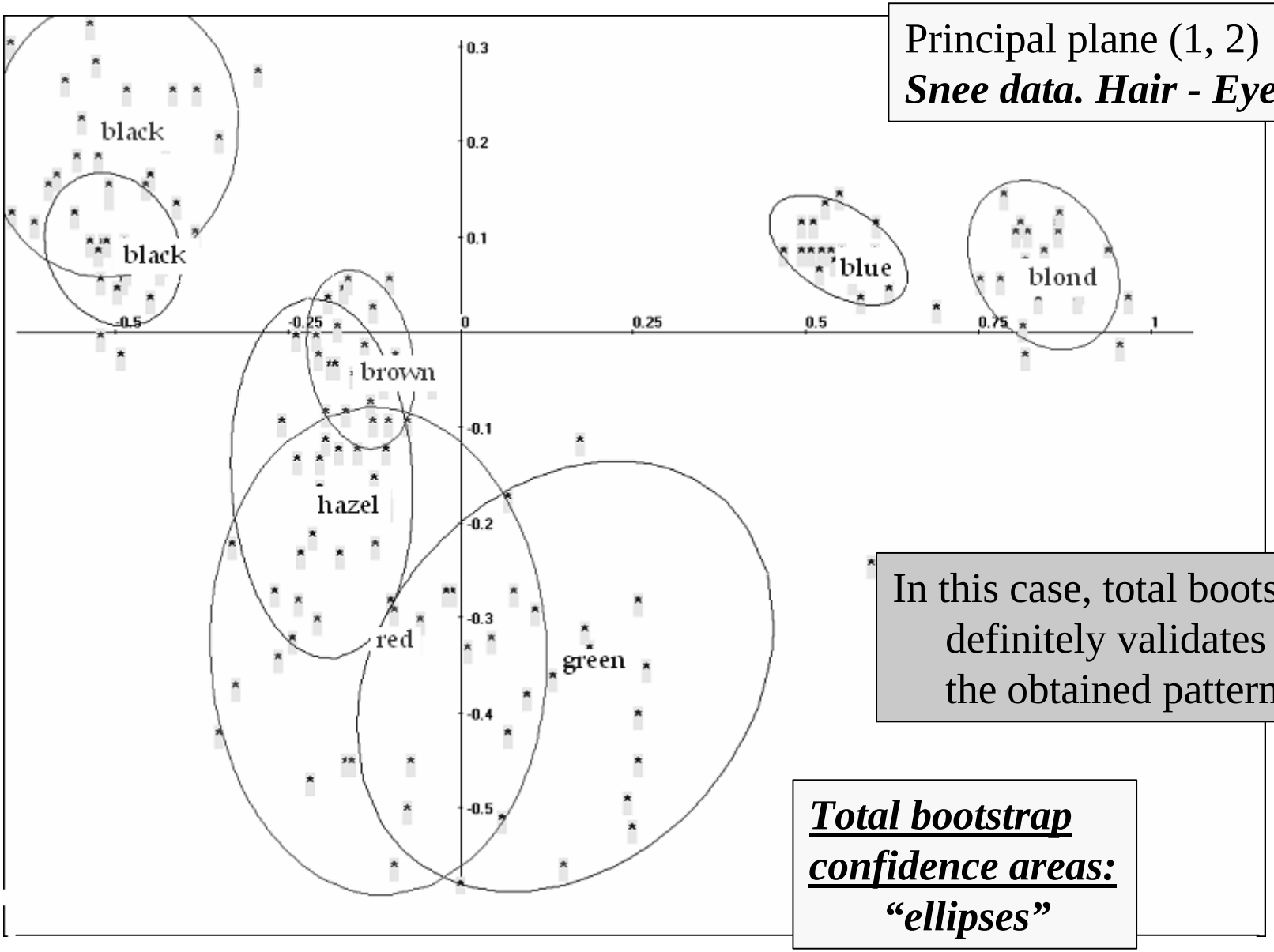
3.2 Total bootstrap type 2

3.3 Total bootstrap type 3

3.1 Total bootstrap total type 1

Total Bootstrap type 1 (very conservative) : simple change (when necessary) of signs of the axes found to be homologous (merely to remedy the arbitrarity of the signs of the axes). The values of a simple scalar product between homologous original and replicated axes allow for this elementary transformation.

This type of bootstrap ignores the possible interchanges and rotations of axes.
It allows for the validation of stable and robust structures.
Each réplication is supposed to produce the original axes with the same ranks (order of the eigenvalues).


Principal plane (1, 2)
Snee data. Hair - Eye
black
black
blue
blond
brown
hazel
red
green
0.3
0.2
0.1
-0.1
-0.2
-0.3
-0.4
-0.5
-0.5
-0.25
0
0.25
0.5
0.75
1
In this case, total bootstrap definitely validates the obtained pattern
Total bootstrap confidence areas:
"ellipses"

3.2 Total bootstrap type 2

Total Bootstrap type 2 (rather conservative) : correction for possible interversions of axes. Replicated axes are sequentially assigned to the original axes with which the correlation (in fact its absolute value) is maximum. Then, alteration of the signs of axes, if needed, as previously.

Total bootstrap type 2 is ideally devoted to the validation of axes considered as latent variables, without paying attention to the order of the eigenvalues.

3.3 Total bootstrap type 3

Total Bootstrap type 3 (could be lenient if the procrustean rotation is done in a space spanned by many axes) : a procrustean rotation (see: Gower and Dijksterhuis, 2004) aims at superimposing as much as possible original and replicated axes.Total bootstrap type 3 allows for the validtion of a whole subspace.

If, for instance, the subspace spanned by the first four replicated axes can coincide with the original four-dimensional subspace, one could find a rotation that can put into coincidence the homologous axes.
The situation is then very similar to that of partial bootstrap.

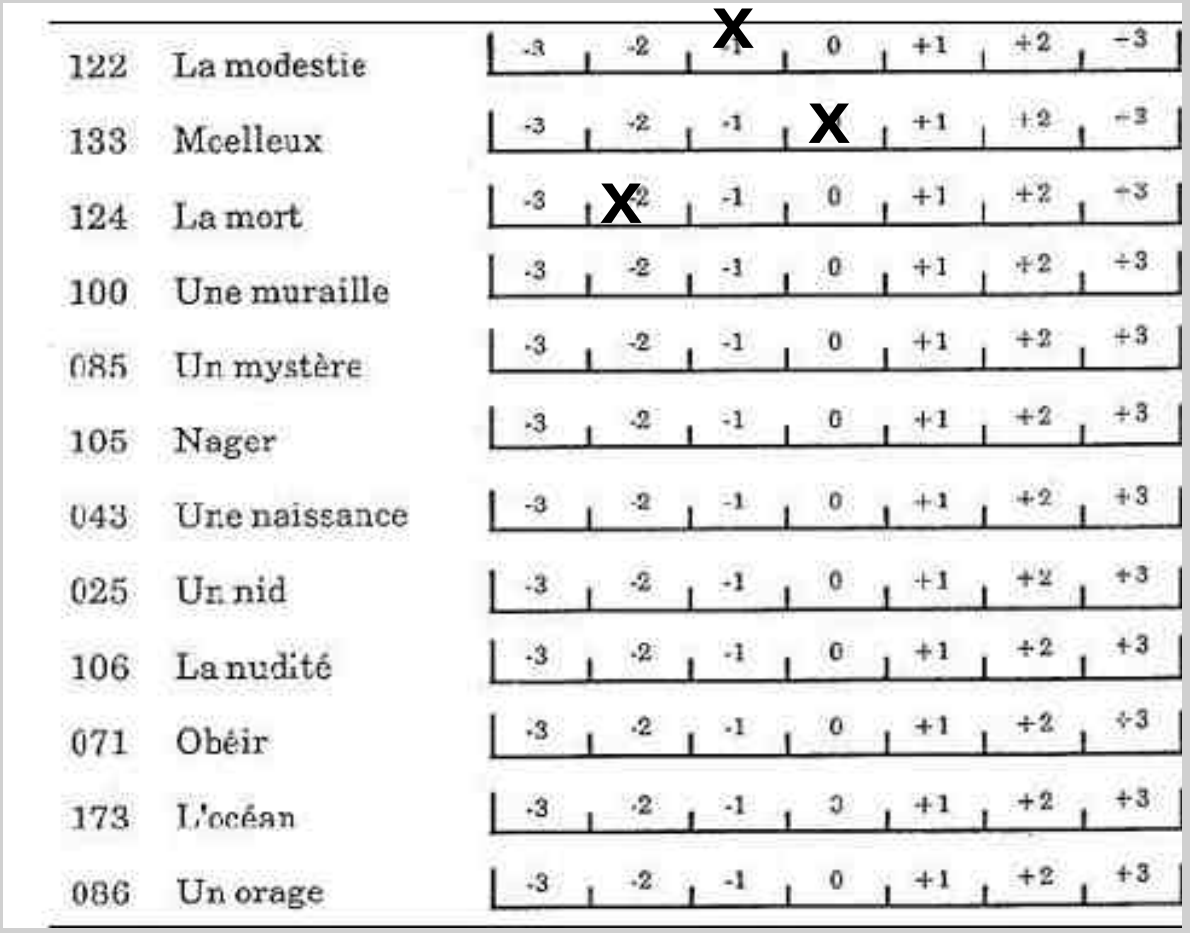
3.4 Example 1 : Validation in Semiometry

The basic idea is to insert in the questionnaire a series of questions consisting uniquely of words
(a list of 210 words is currently used, but some abbreviated lists containing a subset of 80 words could be used as well).

The interviewees must rate these words according to a seven levels scale, the lowest level (mark = 1) concerning a "most disagreeable (or unpleasant) feeling about the word", the highest level (mark = 7) concerning a "most agreeable (or pleasant) feeling" about the word.

Questionnaires in 5 languages

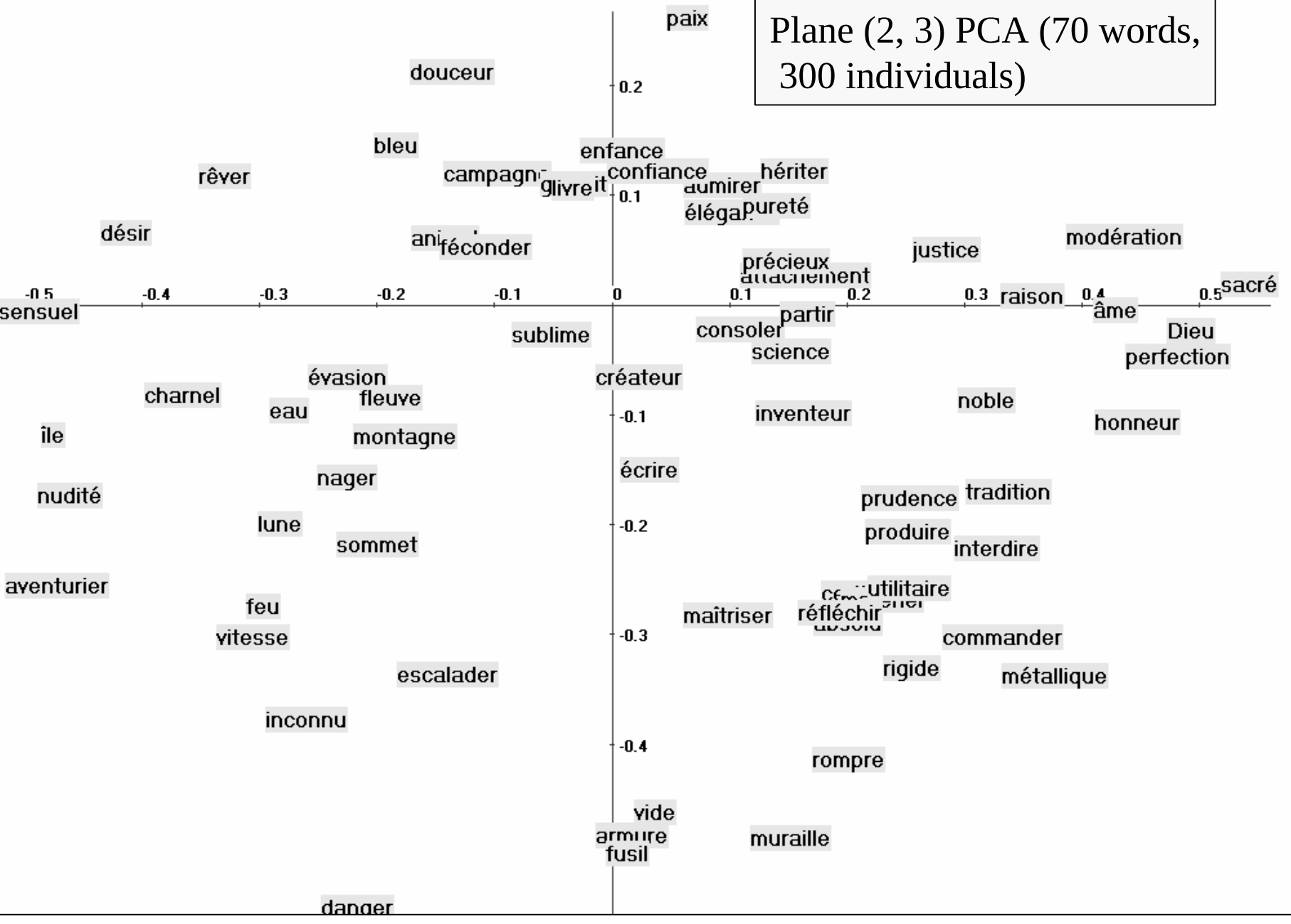
FRENCH	ENGLISH	GERMAN	SPANISH	ITALIAN
l'absolu	absolute	absolut	el absoluto	l'assoluto
l'acharnement	persistence	hartnaeckig	el empeno	l'accanimento
acheter	to buy	kaufen	comprar	comprare
admirer	to admire	bewundern	admirar	ammirare
adorer	to love	anbeten	adorar	adorare
l'ambition	ambition	der ehrgeiz	la ambicion	l'ambizione
l'âme	soul	die seele	el alma	l'anima
l'amitié	friendship	die freundschaft	la amistad	l'amicizia
l'angoisse	anguish	die angst	la angustia	l'angoscia
un animal	animal	ein tier	un animal	un animale
un arbre	tree	ein baum	un arbol	un albero
l'argent	silver	das geld	el dinero	il denaro
une armure	armour	die ruestung	una armadura	un'armatura
l'art	art	die kunst	el arte	l'arte



Facsimile of a semiometric questionnaire

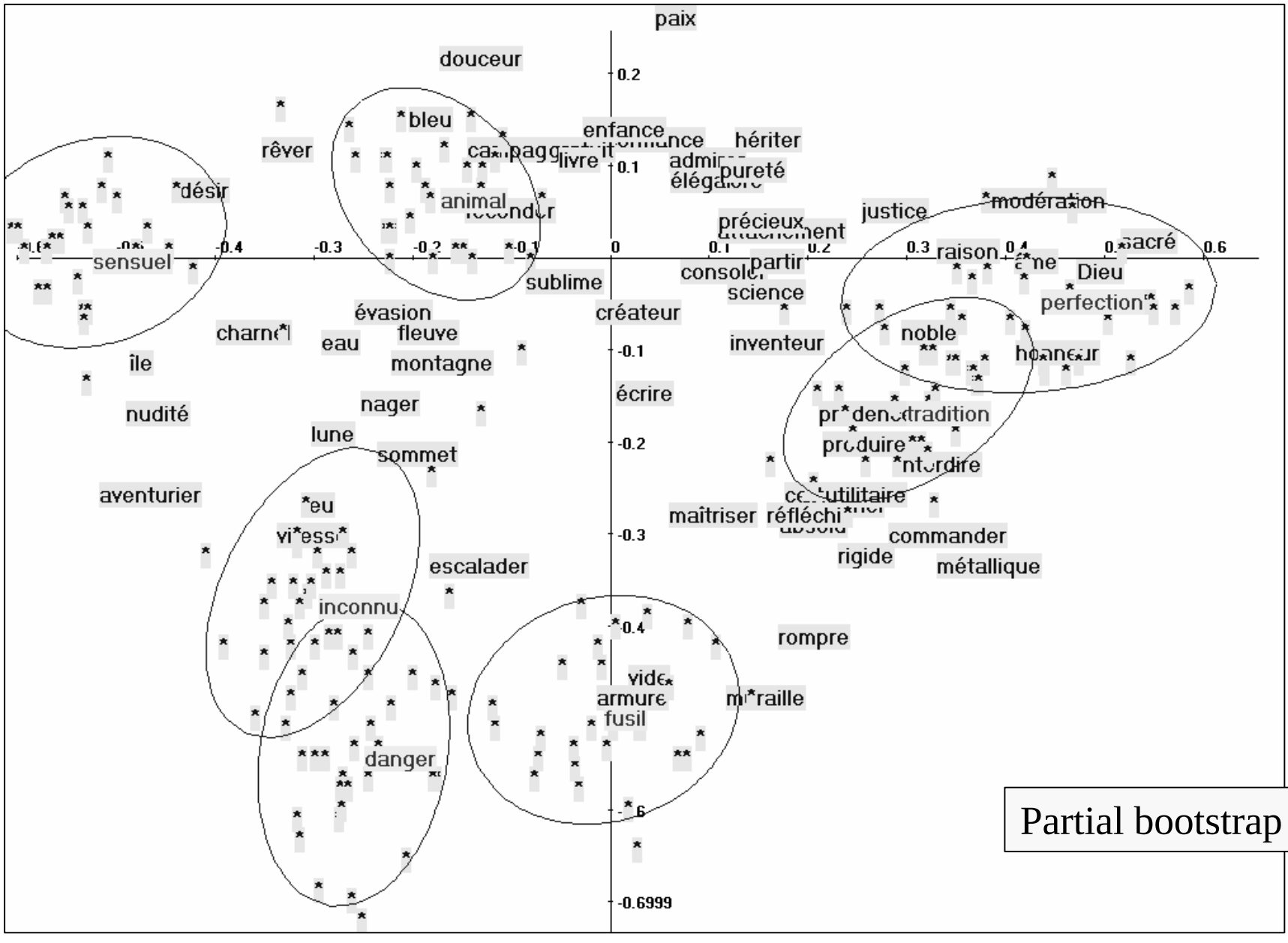
The processing of the filled questionnaires (*mainly through Principal Components Analysis*) produces a stable pattern (*up to 8 stable principal axes*).

Very similar patterns are obtained in ten different countries, despite the problems posed by the translation of the list of words.

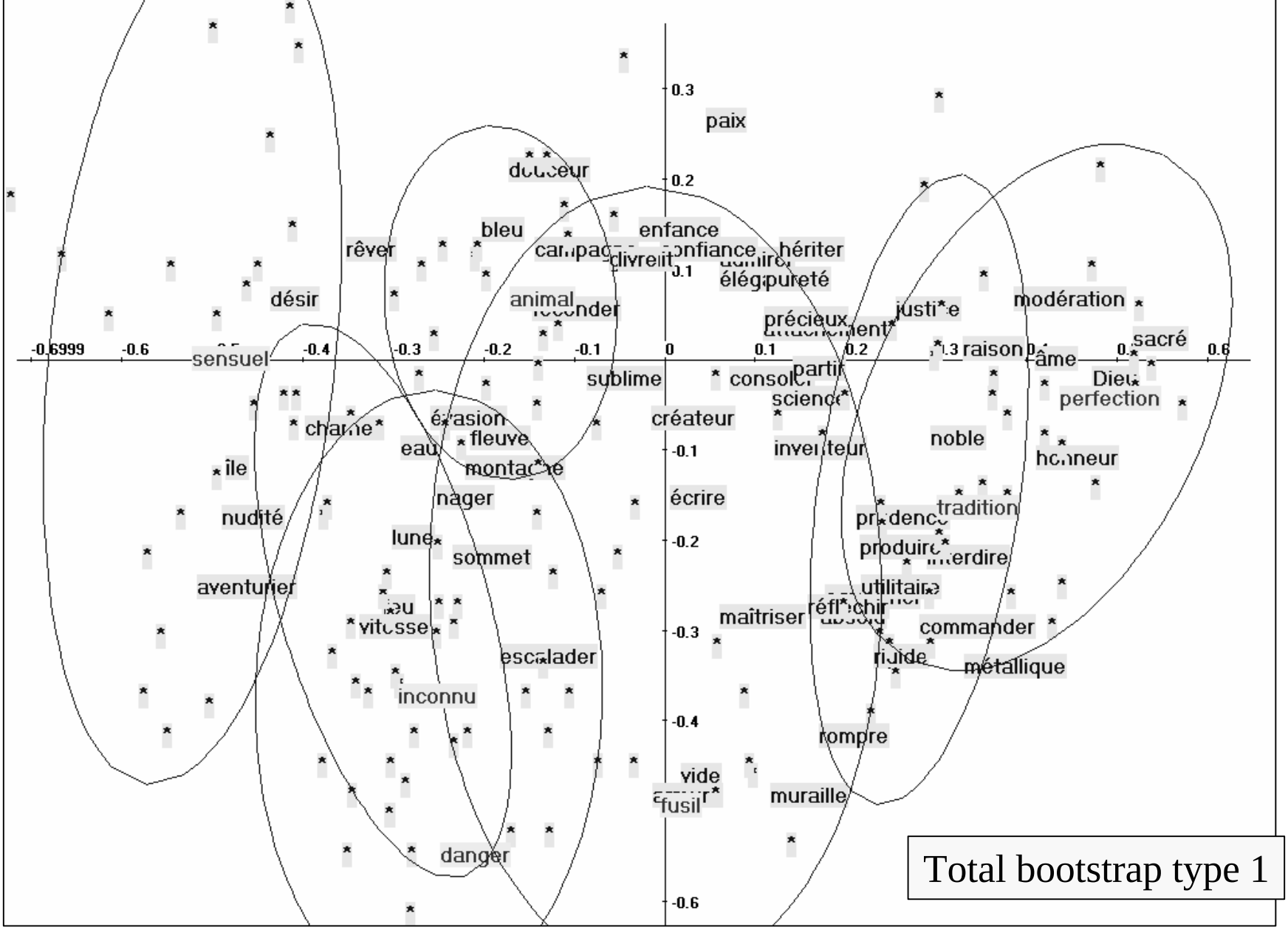

Plane (2, 3) PCA (70 words, 300 individuals)
paix
douceur
bleu
enfance
rêver
confiance
hériter
désir
pureté
modération
justice
féconder
précieux
attachement
sacré
raison
sensuel
partir
âme
consoler
Dieu
sublime
science
perfection
évasion
créateur
charnel
fleuve
noble
eau
inventeur
honneur
île
montagne
écrire
nager
prudence
tradition
nudité
lune
produire
sommet
interdire
aventurier
utilitaire
feu
maîtriser
réfléchir
vitesse
commander
rigide
métallique
escalader
inconnu
rompre
vide
armure
muraille
fusil
danger
0.2
0.1
-0.1
-0.2
-0.3
-0.4
-0.5
0
0.3
0.4
0.5

Anderson confidence intervals for eigenvalues

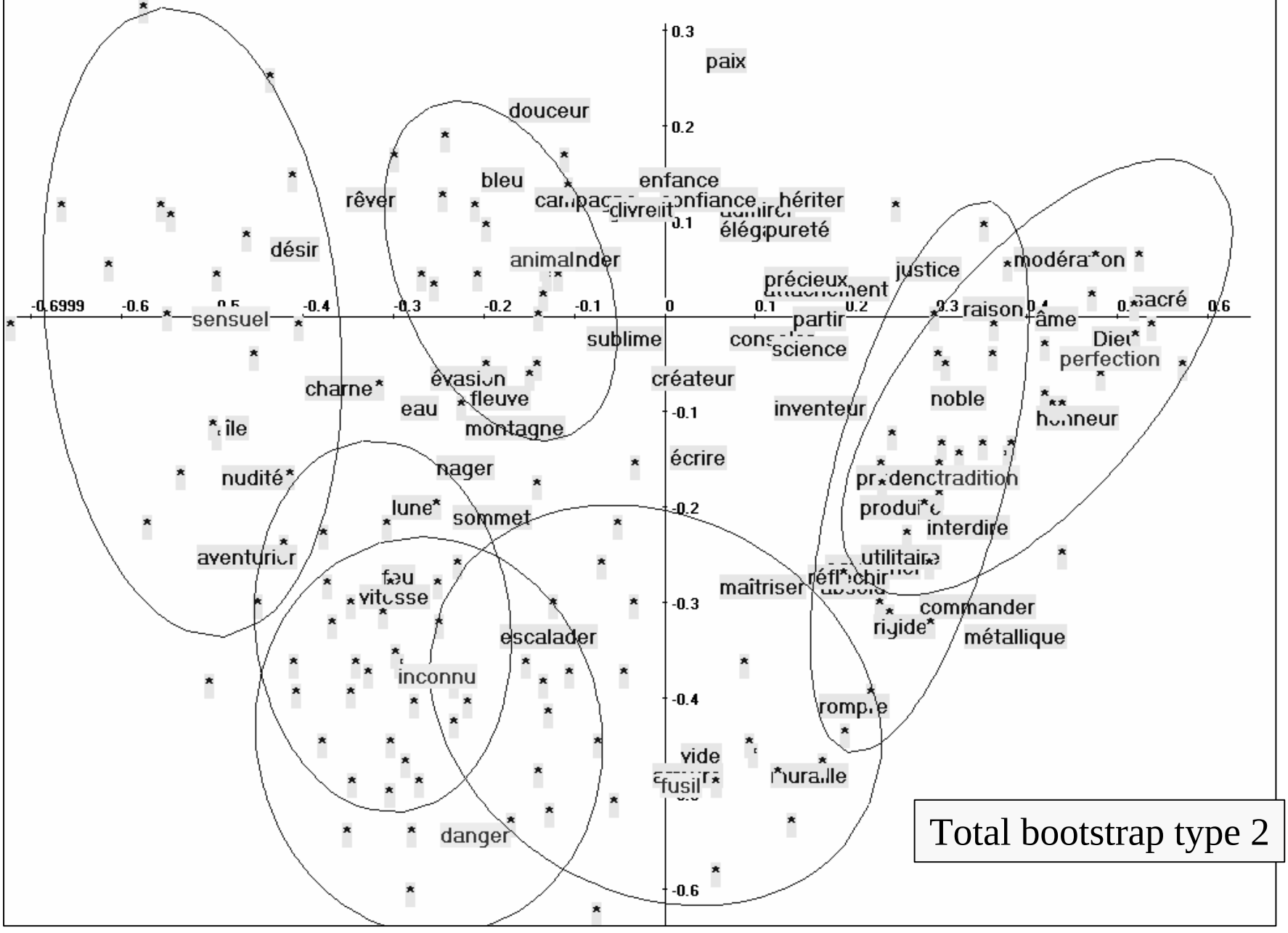
		lower bound	eigen-value	upper bound
Sample 2 000	vp1	24.00	25.35	26.77
	vp2	10.40	10.98	11.60
	vp3	8.24	8.70	9.19
	vp4	6.80	7.18	7.58
	vp5	3.80	4.01	4.23
	vp6	3.59	3.79	4.00
Sample 10 000	vp1	25.49	26.19	26.91
	vp2	10.07	10.35	10.63
	vp3	8.58	8.82	9.06
	vp4	6.82	7.01	7.20
	vp5	4.04	4.15	4.26
	vp6	3.58	3.68	3.78



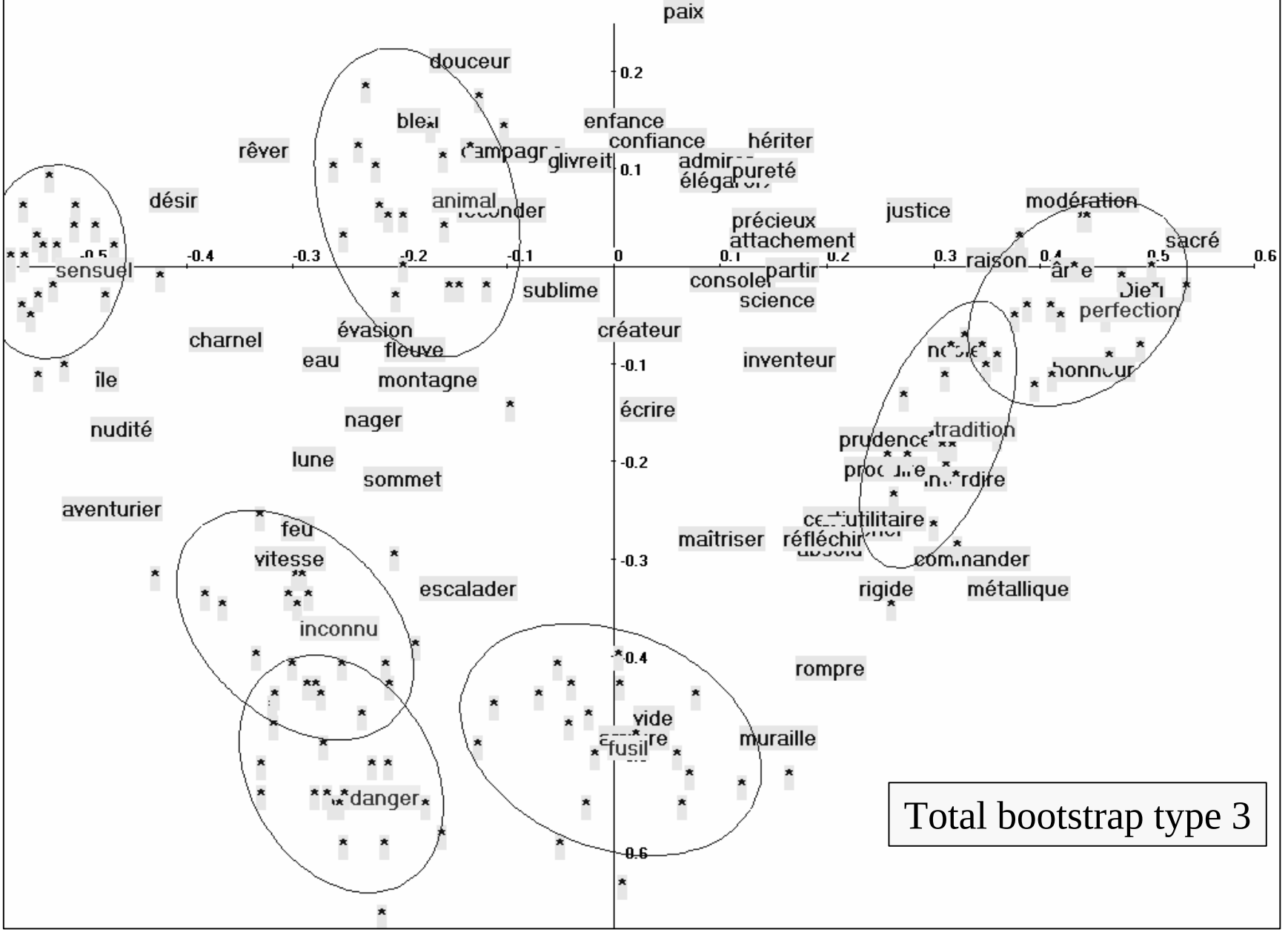
Partial bootstrap



Total bootstrap type 1



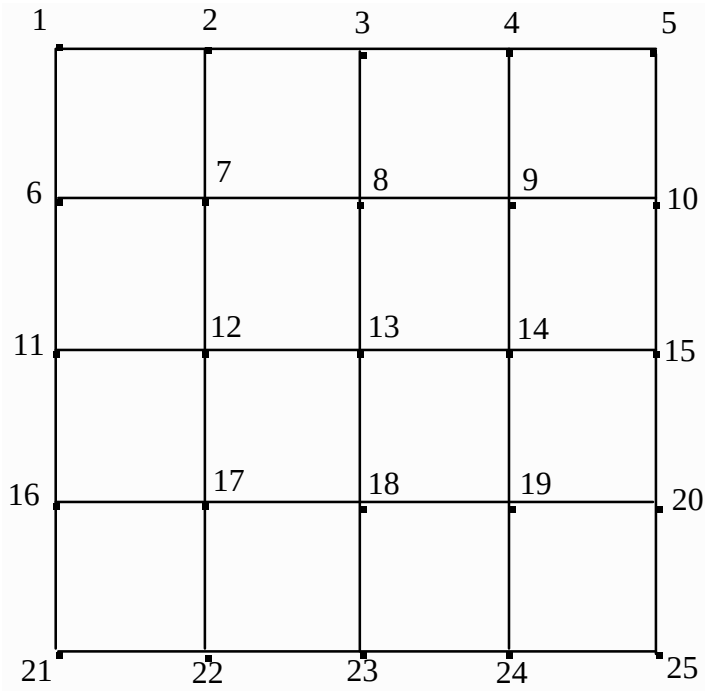
Total bootstrap type 2



Total bootstrap type 3

3.5 Example 2 : Description of graphs

Example of a graph G (***n*** = 25) associated with a squared lattice

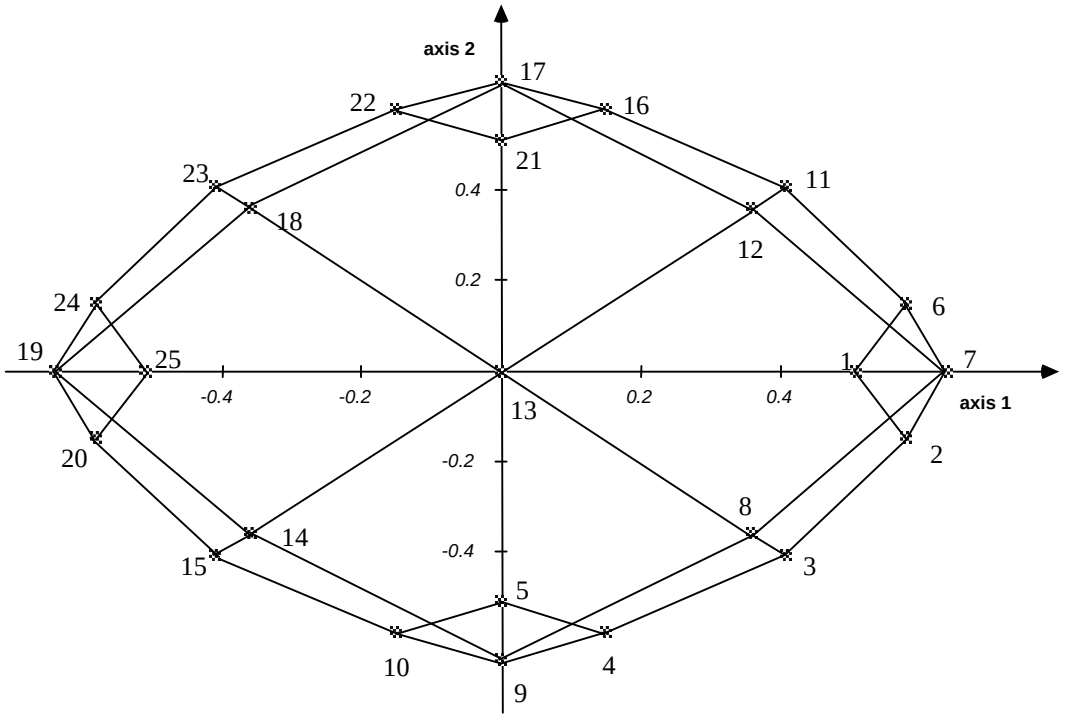


… and its associated matrix **M**

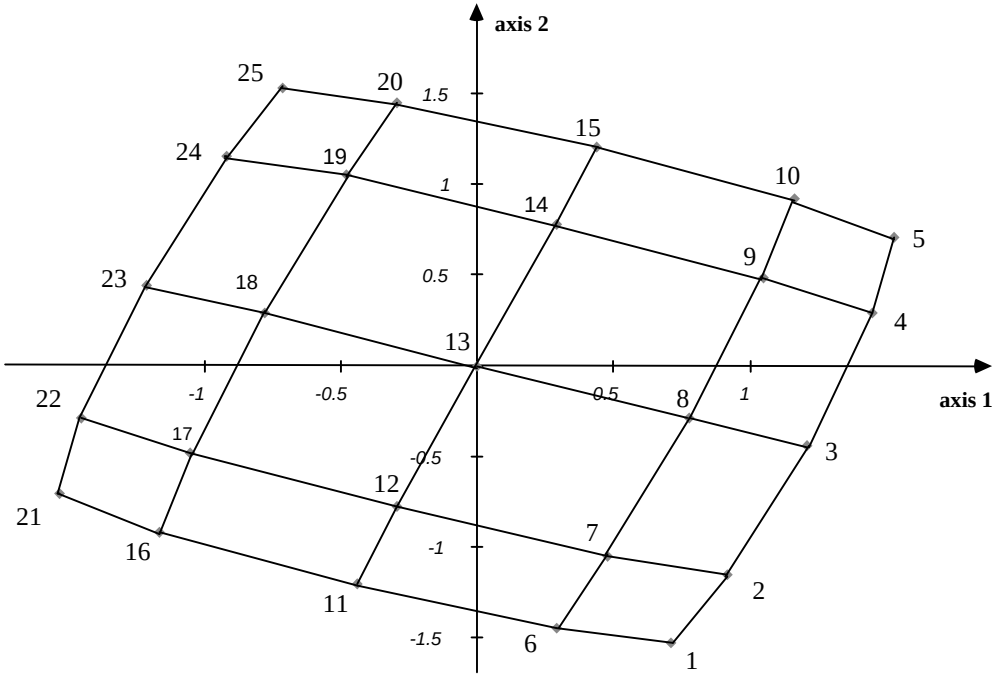
matrix: **M**

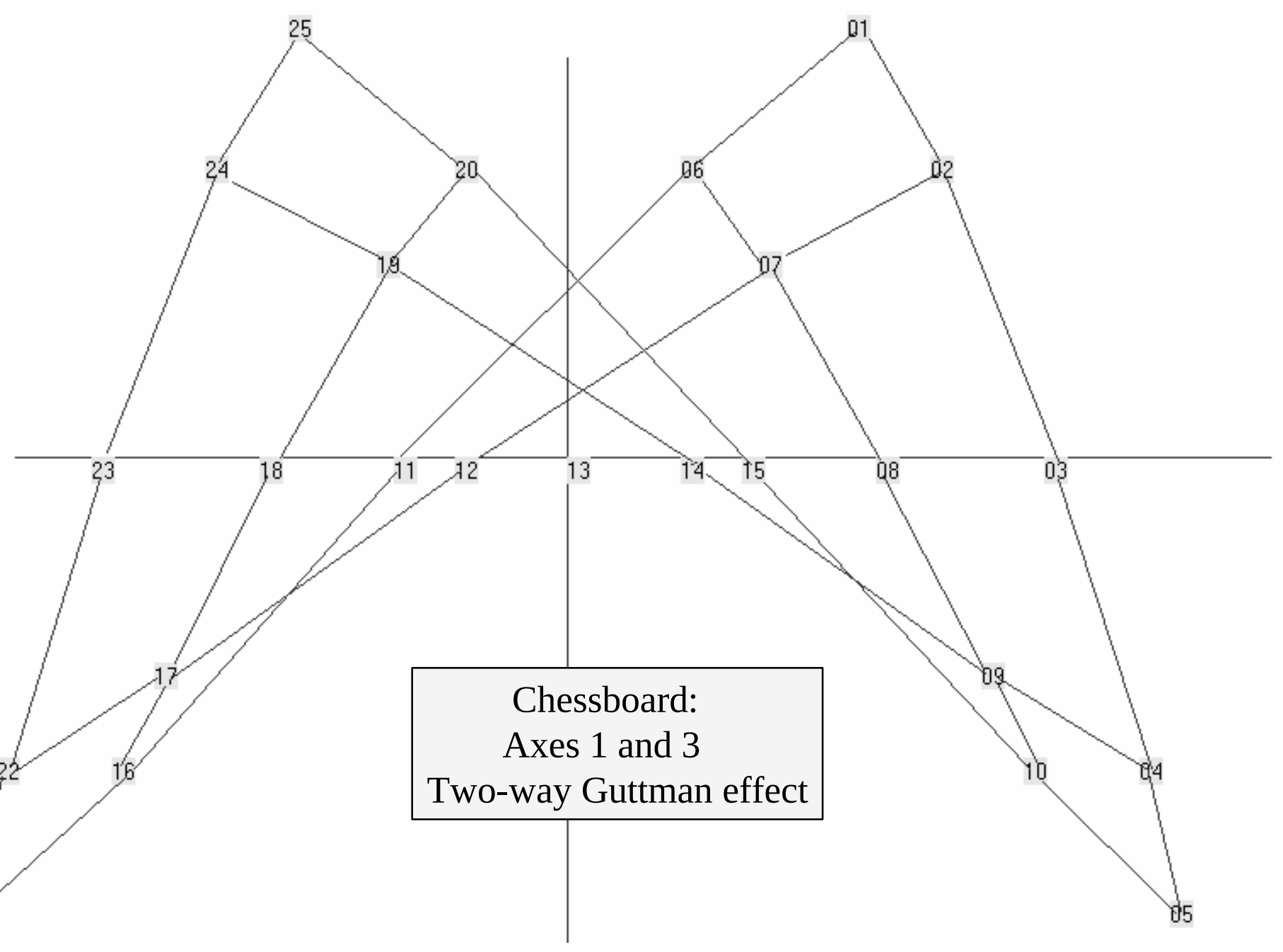
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
r01	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
r02	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
r03	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
r04	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
r05	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
r06	1	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
r07	0	1	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
r08	0	0	1	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
r09	0	0	0	1	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
r10	0	0	0	0	1	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
r11	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0
r12	0	0	0	0	0	0	1	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0
r13	0	0	0	0	0	0	0	1	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0
r14	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0
r15	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0
r16	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	1	0	0	0	0
r17	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	0	0	0	1	0	0	0
r18	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	0	0	0	1	0	0
r19	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	0	0	0	1	0
r20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	1
r21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0
r22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	0	0
r23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	0
r24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1
r25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1

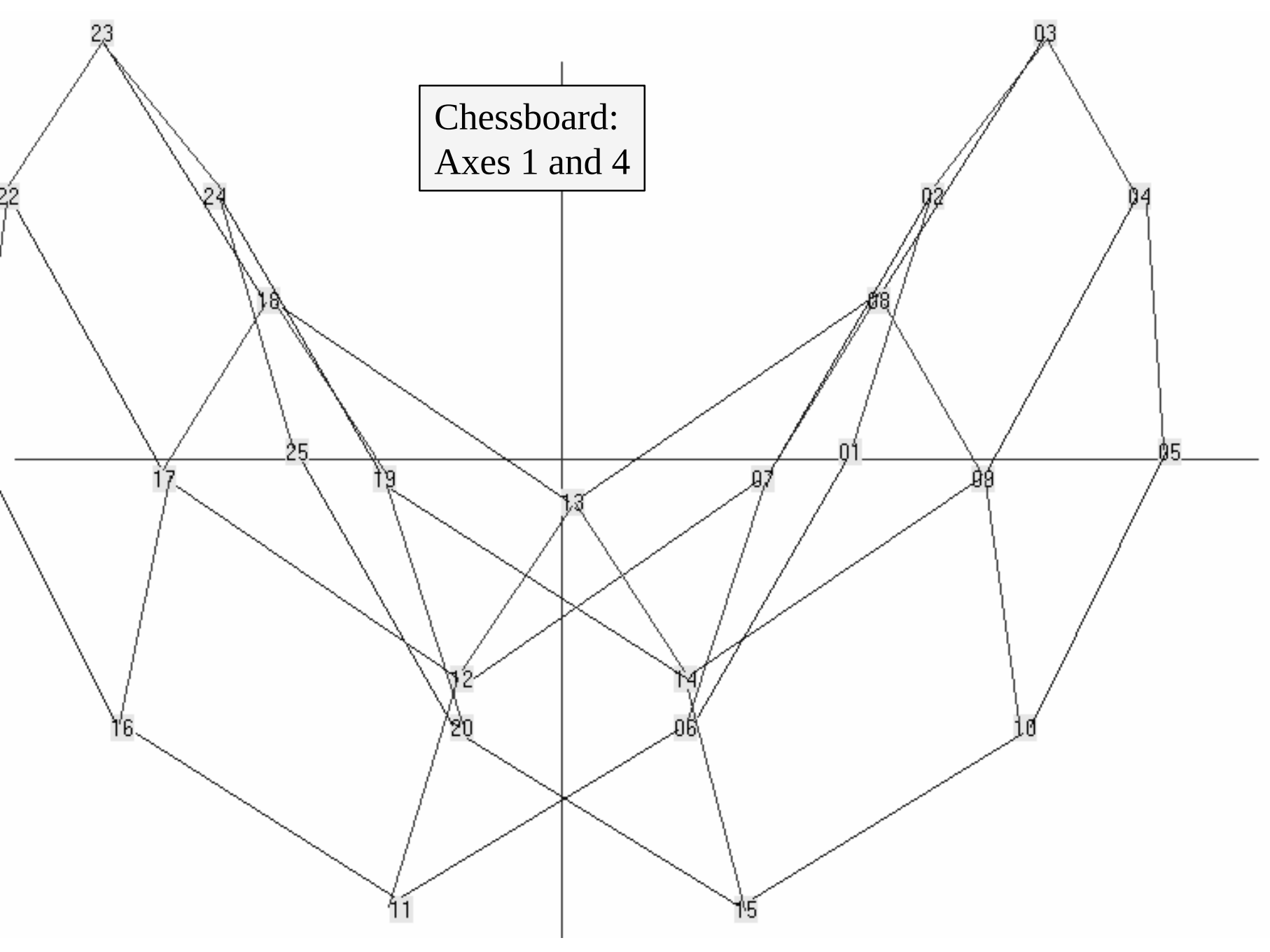
Description of G through *Principal Component Analysis* of **M**



Description of G through *Correspondence Analysis* of **M**




25
01
24
20
06
02
19
07
23
18
11
12
13
14
15
08
03
17
09
22
16
10
04
05
Chessboard:
Axes 1 and 3
Two-way Guttman effect


Chessboard:
Axes 1 and 4
23
03
22
24
02
04
18
08
25
01
05
17
19
07
09
13
12
14
16
20
06
10
11
15

Why good visualization of planar graphs are obtained?

Explanation :

Local variance = $\mathbf{y}'(\mathbf{I} - \mathbf{N}^{-1}\mathbf{M})\ \mathbf{y}$

Global variance = $\mathbf{y}'\mathbf{y}$

Bounds for c(y) = contiguity coefficient**.**

$$c(y) = \mathbf{y}'(\mathbf{I} - \mathbf{N}^{-1}\mathbf{M})\ \mathbf{y} \ / \ \mathbf{y}'\mathbf{y}$$

minimum of $c(y)$, m, is the smallest eigenvalue of:

$$(\mathbf{I} - \mathbf{N}^{-1}\mathbf{M})\ \psi = m\,\psi$$

Equivalently:

$$\mathbf{N}^{-1}\mathbf{M}\,\psi = (1 - m)\,\psi$$

transition formulae, CA of **M** : $\mathbf{N}^{-1}\mathbf{M}\,\phi = e\ddot{O}l\ \phi$

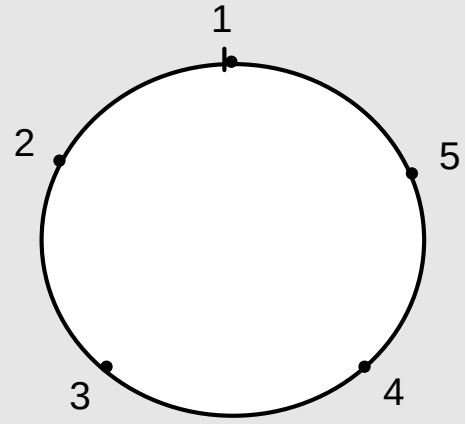
if e = +1, direct factor, if e = -1, inverse factor.

Min m = Max l , l_{max} if (e = +1).

Thus: $Min\ [\ c(y)\ ] = 1 - \ddot{O}l_{max}$

Misleading measures of information : Case of a cycle

$$\mathbf{M} = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$



$$j_a(j) = \cos\left(\frac{2jap}{n}\right) \quad \text{and} \quad y_a(j) = \sin\left(\frac{2jap}{n}\right)$$

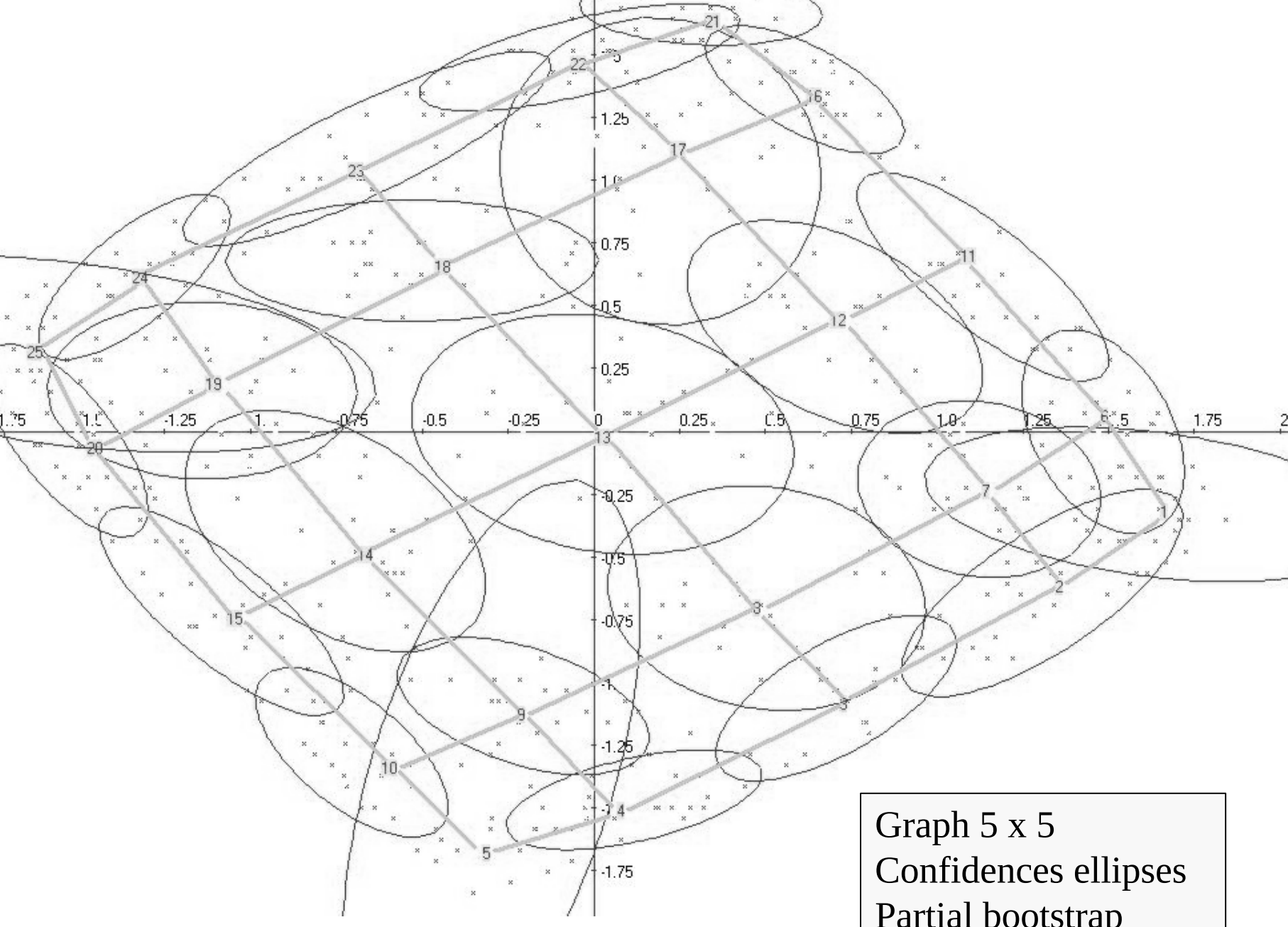
$$l_a = \cos^2\left(\frac{2ap}{n}\right) \qquad t_a = \frac{2}{n}\cos^2\left(\frac{2ap}{n}\right)$$

$$t_a \to 0 \quad when \quad n \to \infty$$

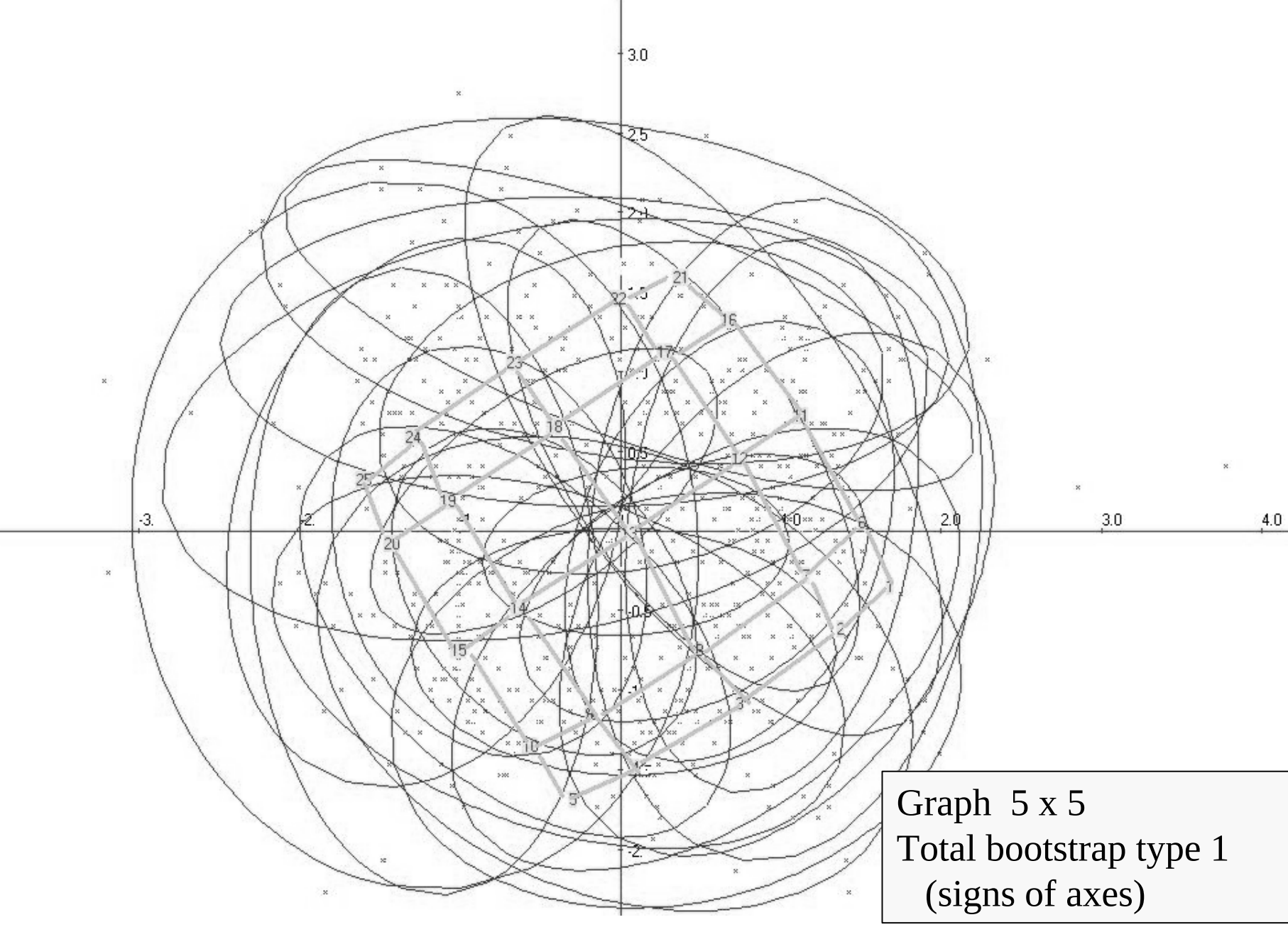
(3.5 Example 2: Continuation)

Confidence areas for the vertices of symmetric graphs

Example of a graph G (***n*** = 25) associated with a squared 5 x 5 lattice

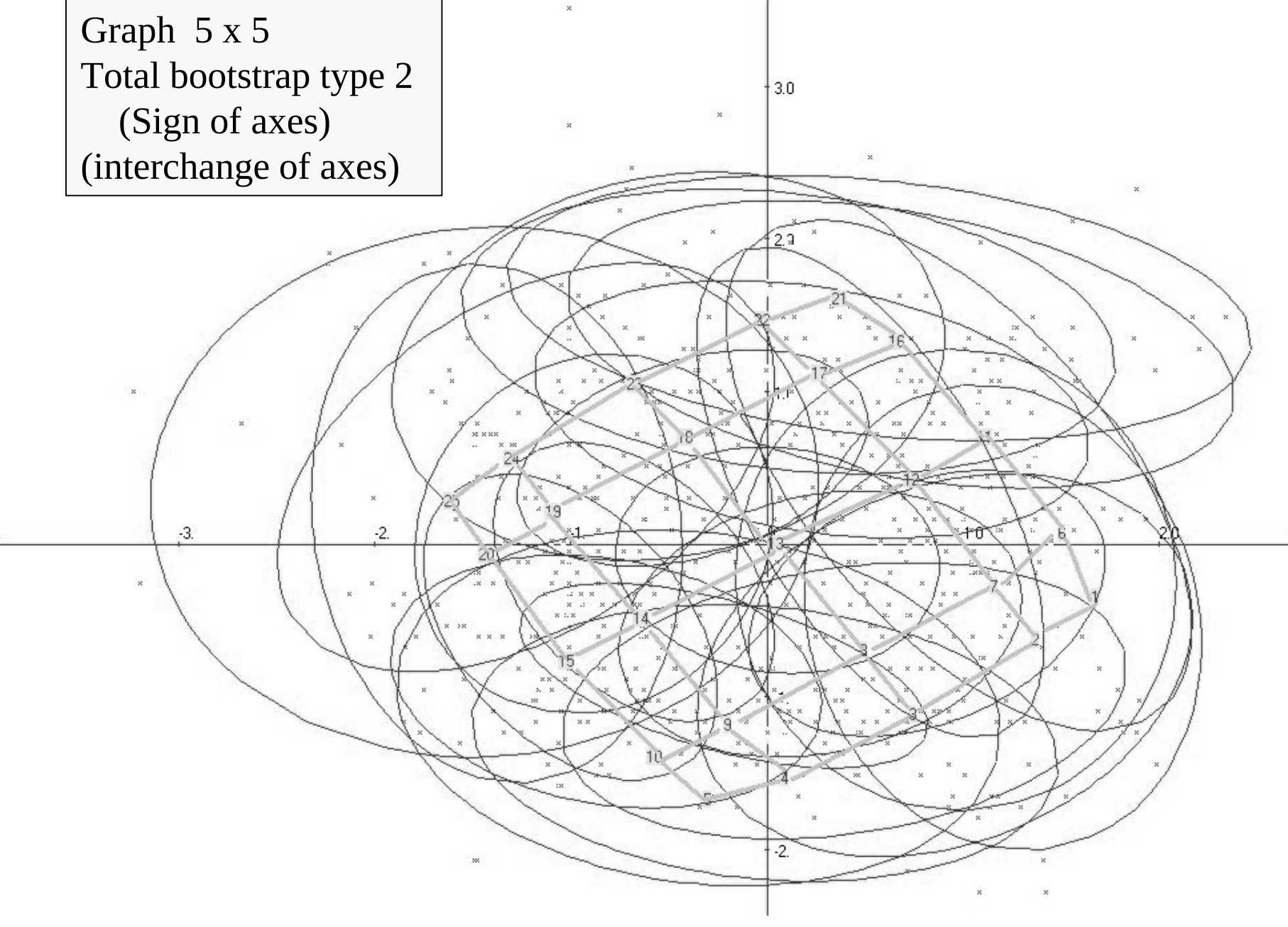


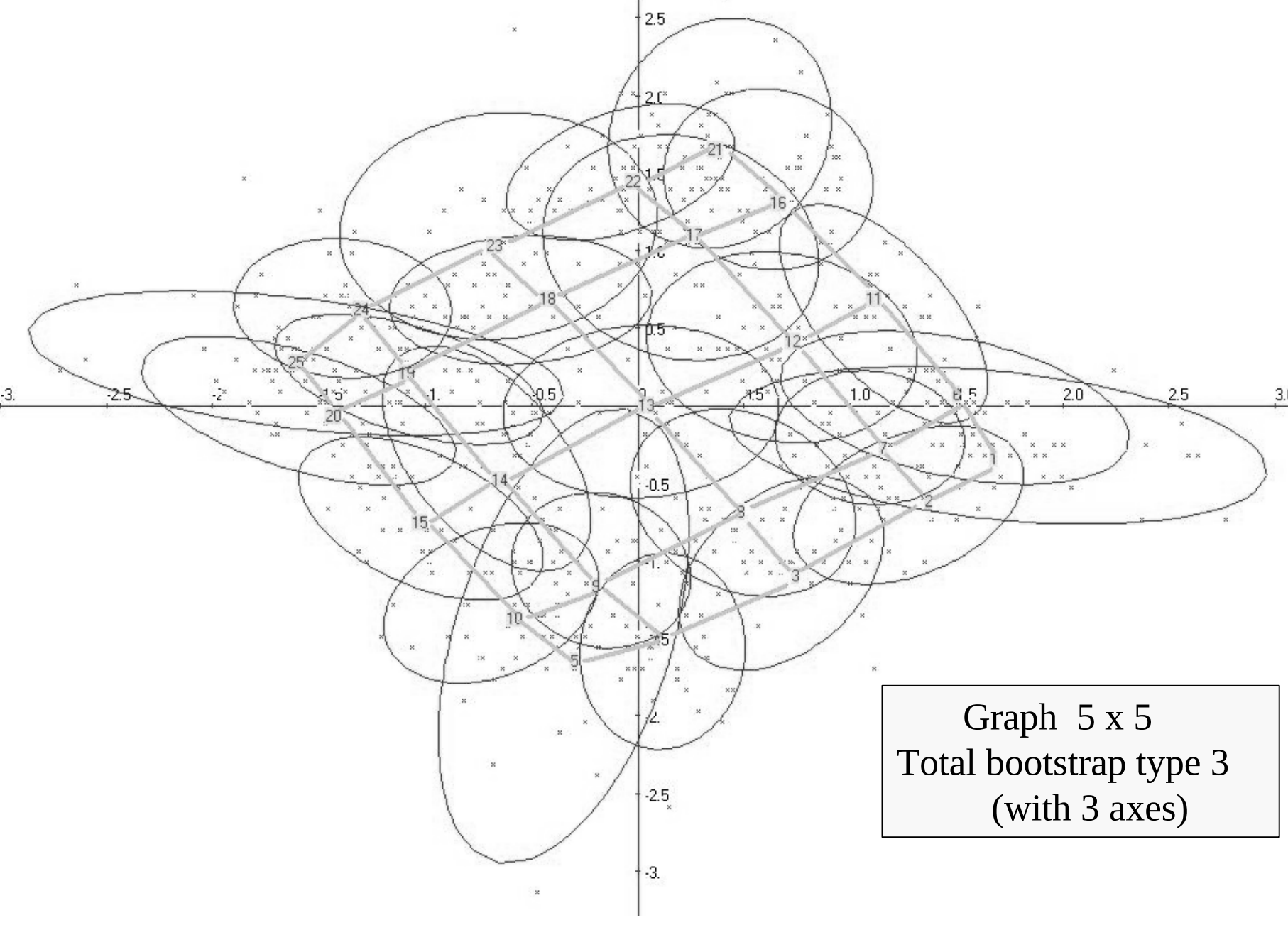
Graph 5 x 5
Confidences ellipses
Partial bootstrap



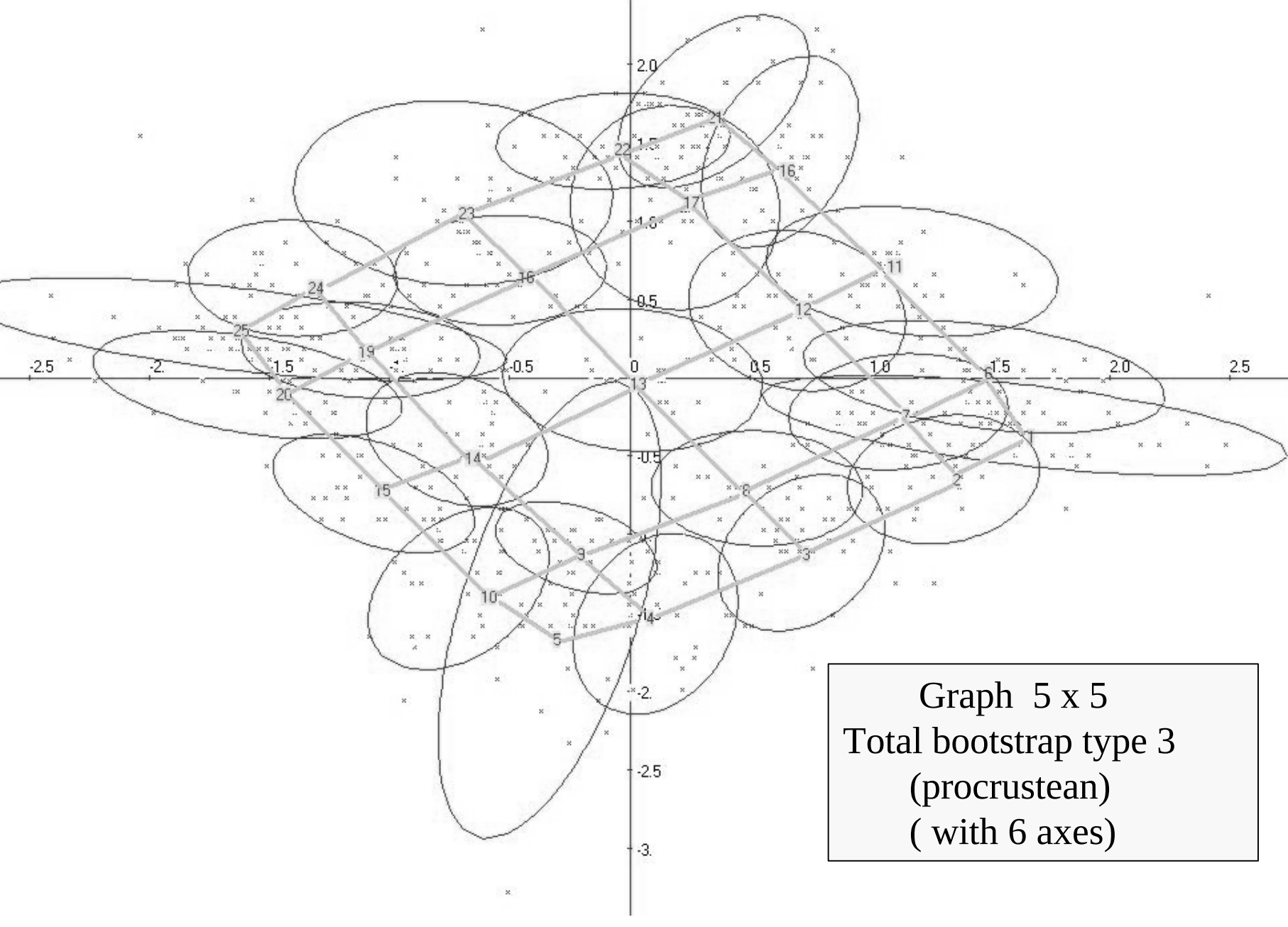
Graph 5 x 5
Total bootstrap type 1
(signs of axes)

Graph 5 x 5
Total bootstrap type 2
(Sign of axes)
(interchange of axes)

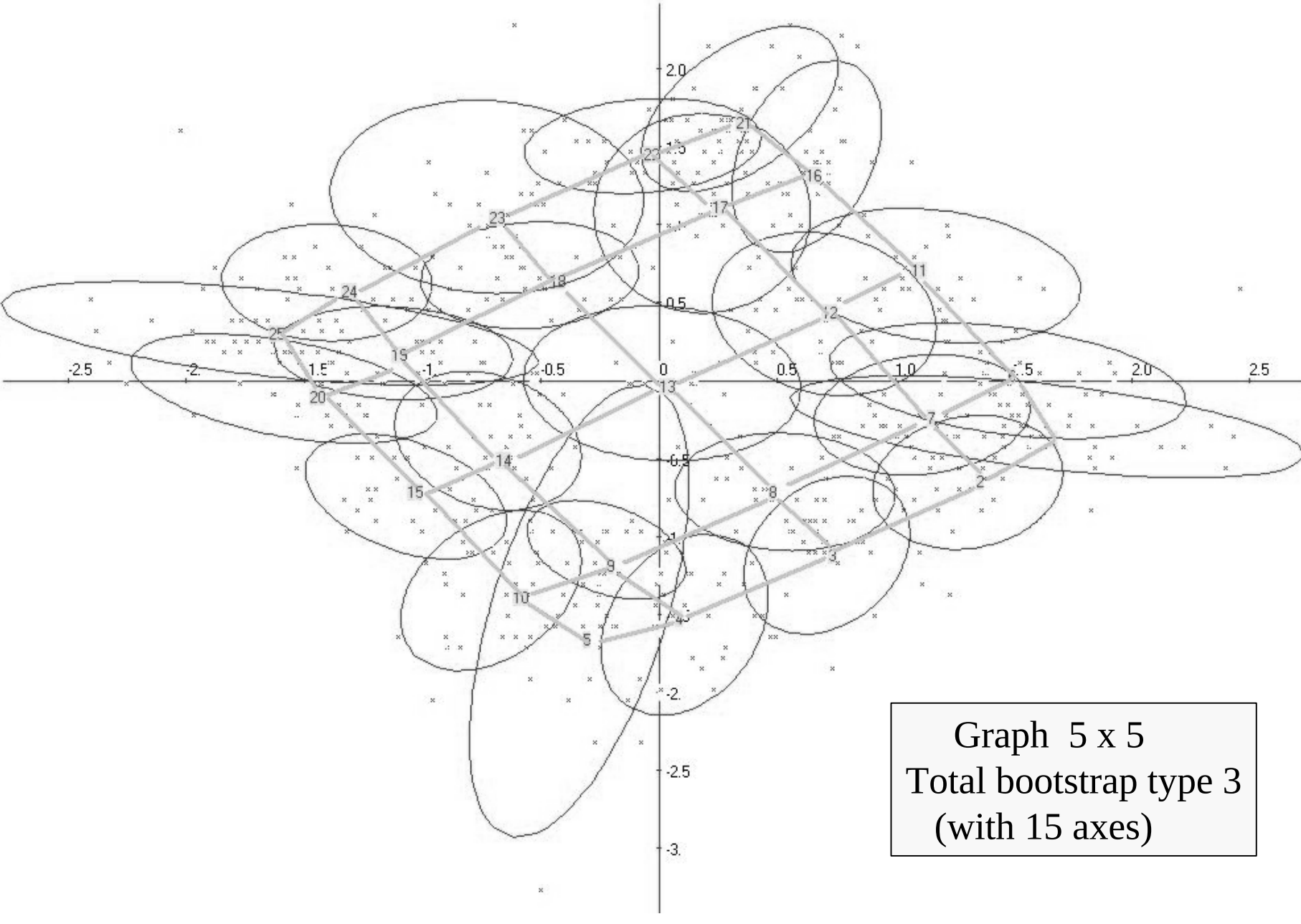




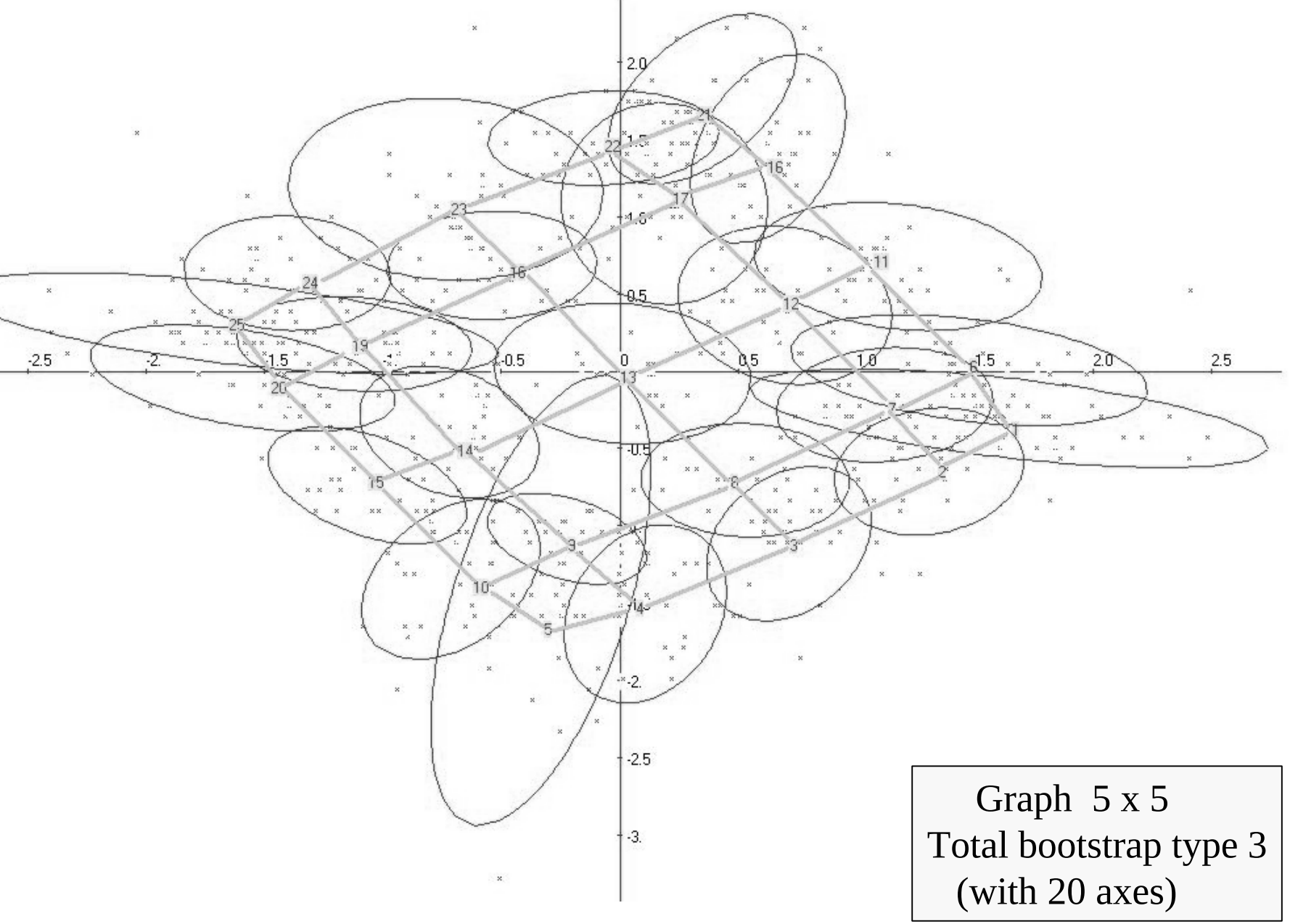
Graph 5 x 5
Total bootstrap type 3
(with 3 axes)



Graph 5 x 5
Total bootstrap type 3
(procrustean)
(with 6 axes)

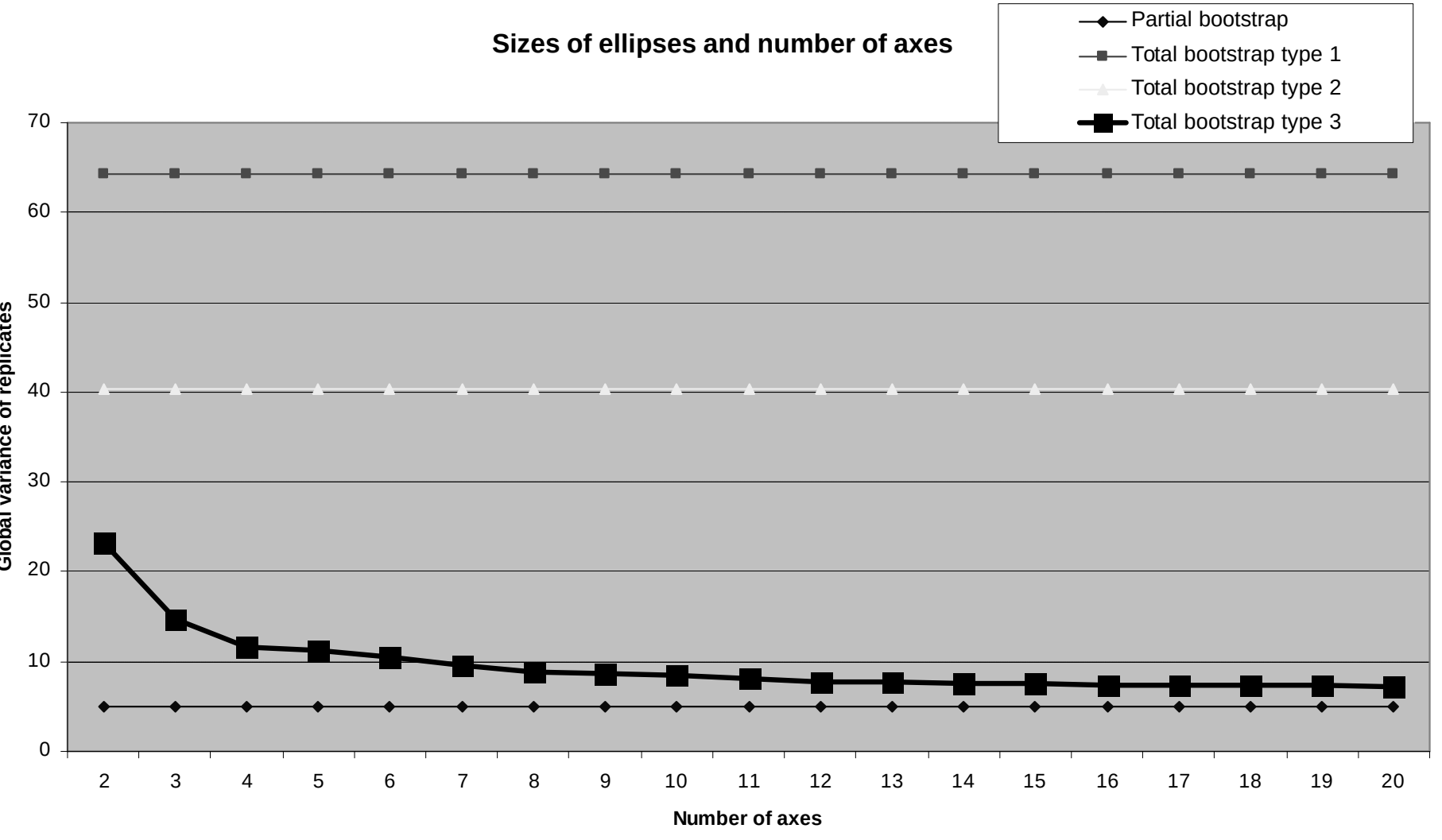


Graph 5 x 5
Total bootstrap type 3
(with 15 axes)



Graph 5 x 5
Total bootstrap type 3
(with 20 axes)

Graph 5x5: Comparison of four bootstrap techniques



3.6 Example 3: Open question in a sample surveys

The following open-ended question was asked :

"What is the single most important thing in life for you?«

It was followed by the probe:

"What other things are very important to you?".

This question was included in a multinational survey conducted in seven countries (Japan, France, Germany, Italy, Nederland, United Kingdom, USA) in the late nineteen eighties (**Hayashi *et al*., 1992**).

Our illustrative example is limited to the British sample (Sample size: *1043*).

Examples of responses to "Life" question

Gender	*Educ.*	*Age*	*Responses*
1	1	4	happiness in people around me, contented family, would make me happy
1	2	2	my own time, not dictated by other people
1	2	2	freedom of choice as to what I do in my leisure time
1	3	2	I suppose work
1	2	1	firm, my work, which is my dad's firm
2	1	6	just the memory of my last husband
2	2	6	well-being of my handicapped son
1	1	5	my wife, she gave me courage to carry on even in the bad times
2	2	3	my sons, my kids are very important to me, being on my own, I am responsible for their education
1	3	3	job, being a teacher I love my job, for the well-being of the children

Example 3, *continuation*

The counts for the first phase of numeric coding are as follows:

Out of *1043* responses, there are *13 669* occurrences (*tokens*), with *1 413* distinct words (*types*).

When the words appearing at least *16* times are selected, there remain *10 357* occurrences of these words (*tokens*), with *135* distinct words (*types*).

Example 3, *continuation*

The same questionnaire also had a number of closed-end questions (among them, the socio-demographic characteristics of the respondents, which play a major role).

In this example we focus on a partitioning of the sample into *nine* categories, obtained by cross-tabulating **age** (*three* categories) with **educational level** (*three* categories).

Example of a lexical contingency table

Partial listing of lexical table cross-tabulating 135 words of frequency greater than or equal to 16 with 9 age-education categories

	L-30	L-55	L+55	M-30	M-55	M+55	H-30	H-55	H+55
I	*2*	*46*	*92*	*30*	*25*	*19*	*11*	*21*	*2*
I'm	*2*	*5*	*9*	*3*	*2*	*1*	*0*	*0*	*0*
a	*10*	*56*	*66*	*54*	*44*	*19*	*20*	*22*	*7*
able	*1*	*9*	*16*	*9*	*7*	*4*	*4*	*5*	*0*
about	*0*	*3*	*13*	*7*	*1*	*2*	*4*	*1*	*0*
after	*1*	*8*	*11*	*3*	*1*	*2*	*0*	*0*	*0*
all	*1*	*24*	*19*	*8*	*18*	*6*	*3*	*5*	*2*
and	*8*	*89*	*148*	*86*	*73*	*30*	*25*	*32*	*13*
anything	*0*	*4*	*9*	*1*	*3*	*0*	*1*	*1*	*0*

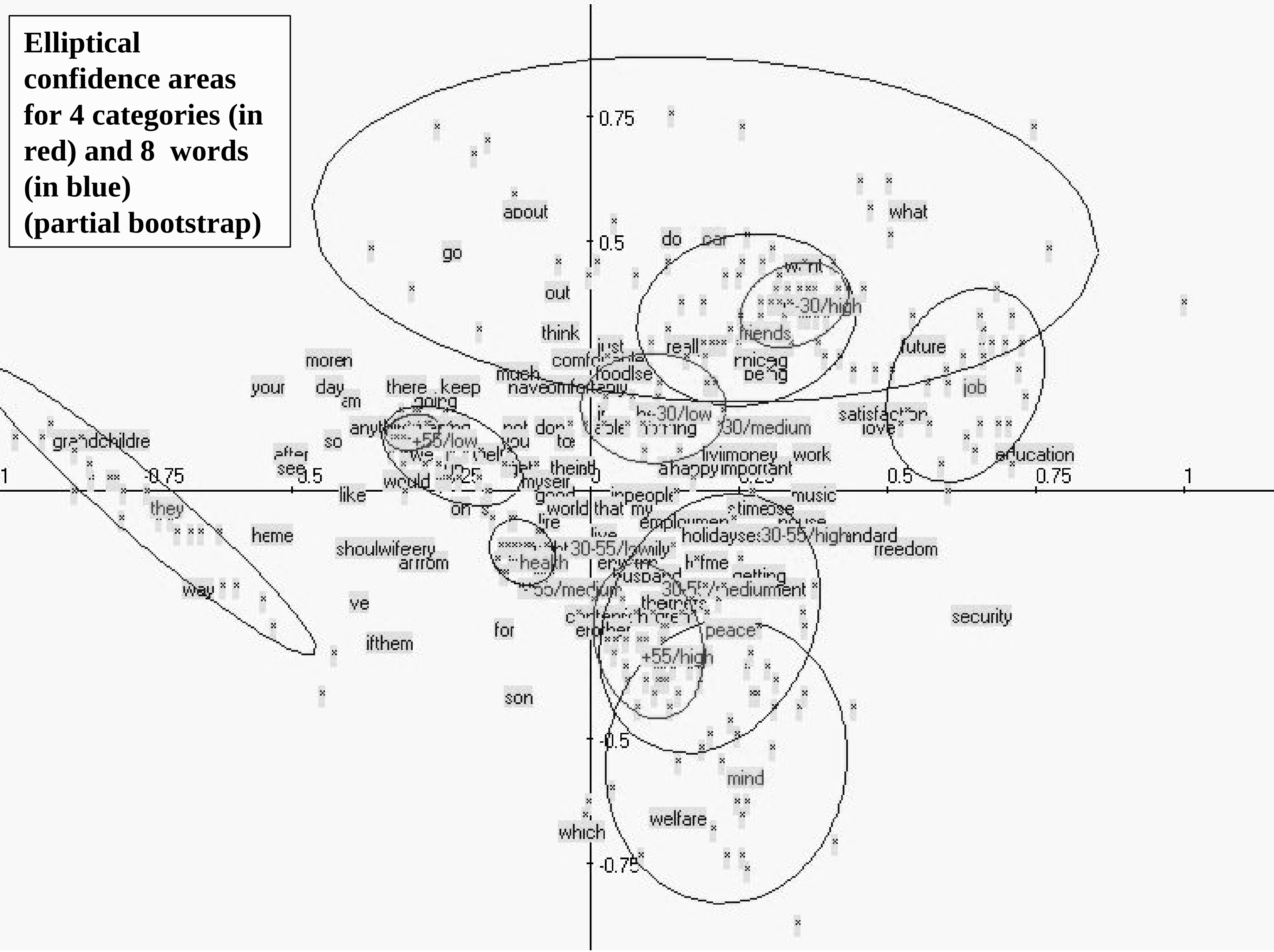
- The two forthcoming diapositives show the principal plane produced by a correspondence analysis of the previous lexical contingency table.

- Proximity between 2 category-points (columns)means similarity of lexical profiles of the 2 categories.

- Proximity between 2 word-points (rows) means similarity of lexical profiles of these words.

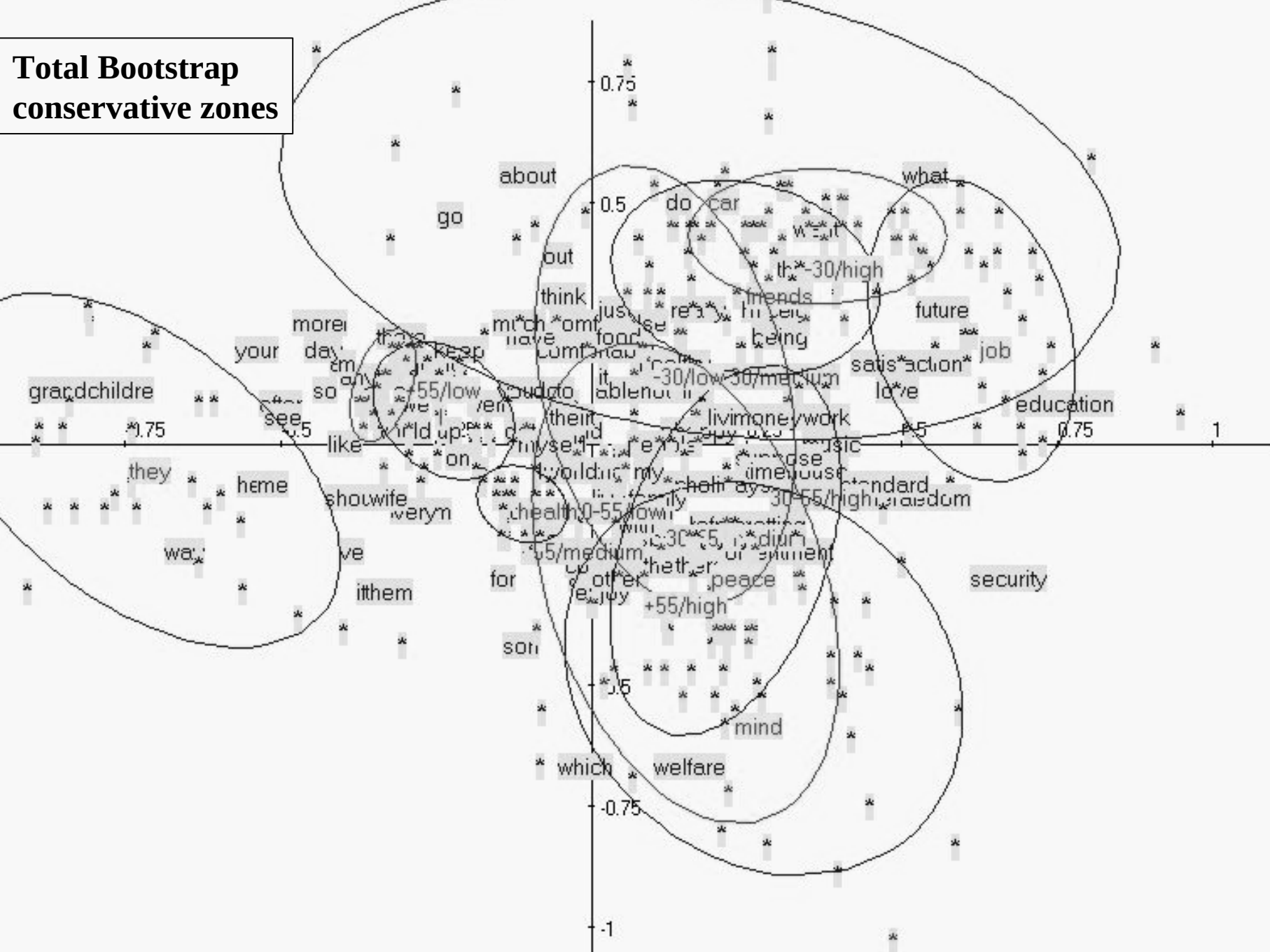
- Both ellipses and convex hulls describe the uncertainty of the location of the points.

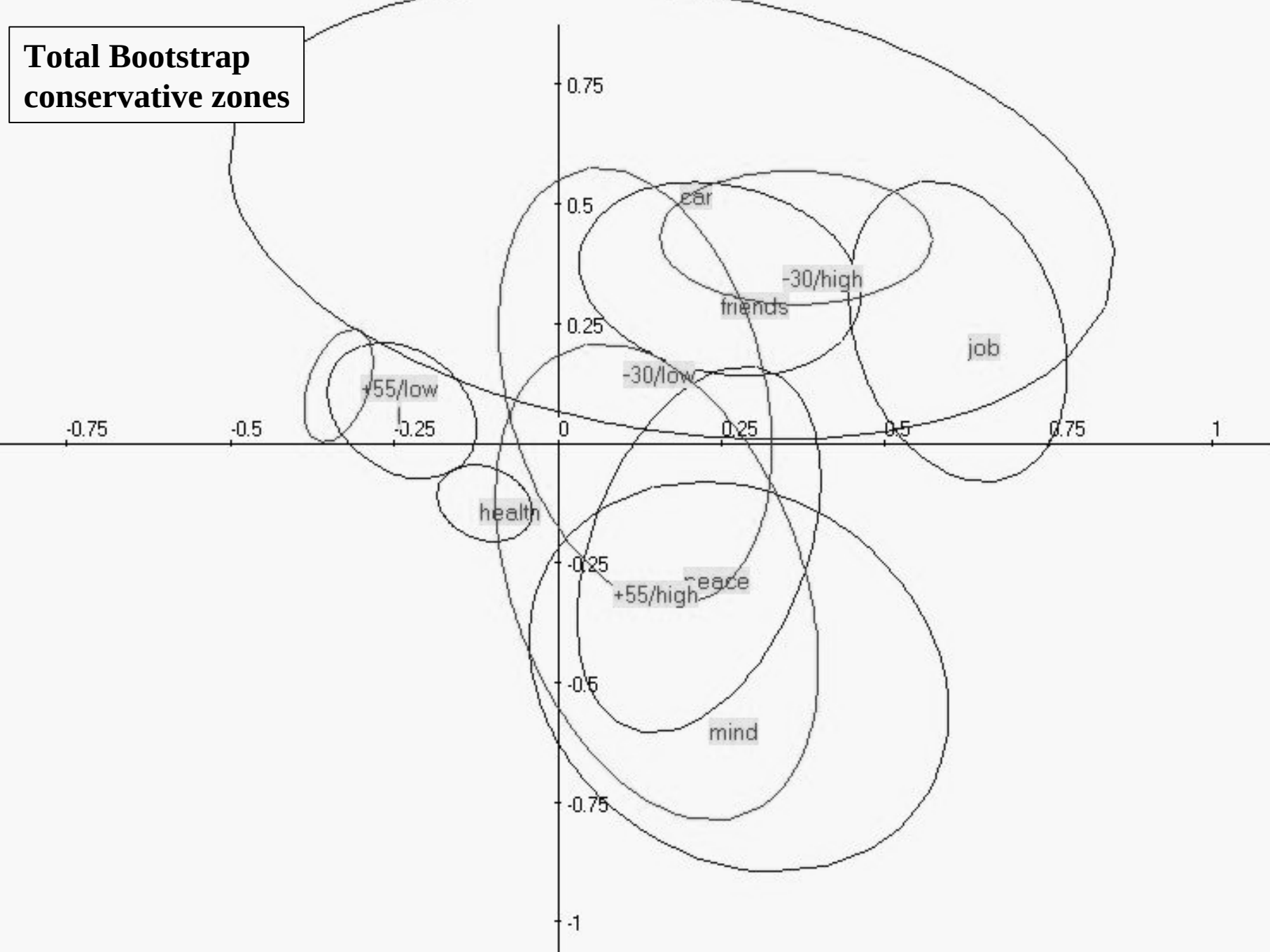
- 9 categories points, in red (all the categories, in fact)
- 6 selected word-points, in blue.



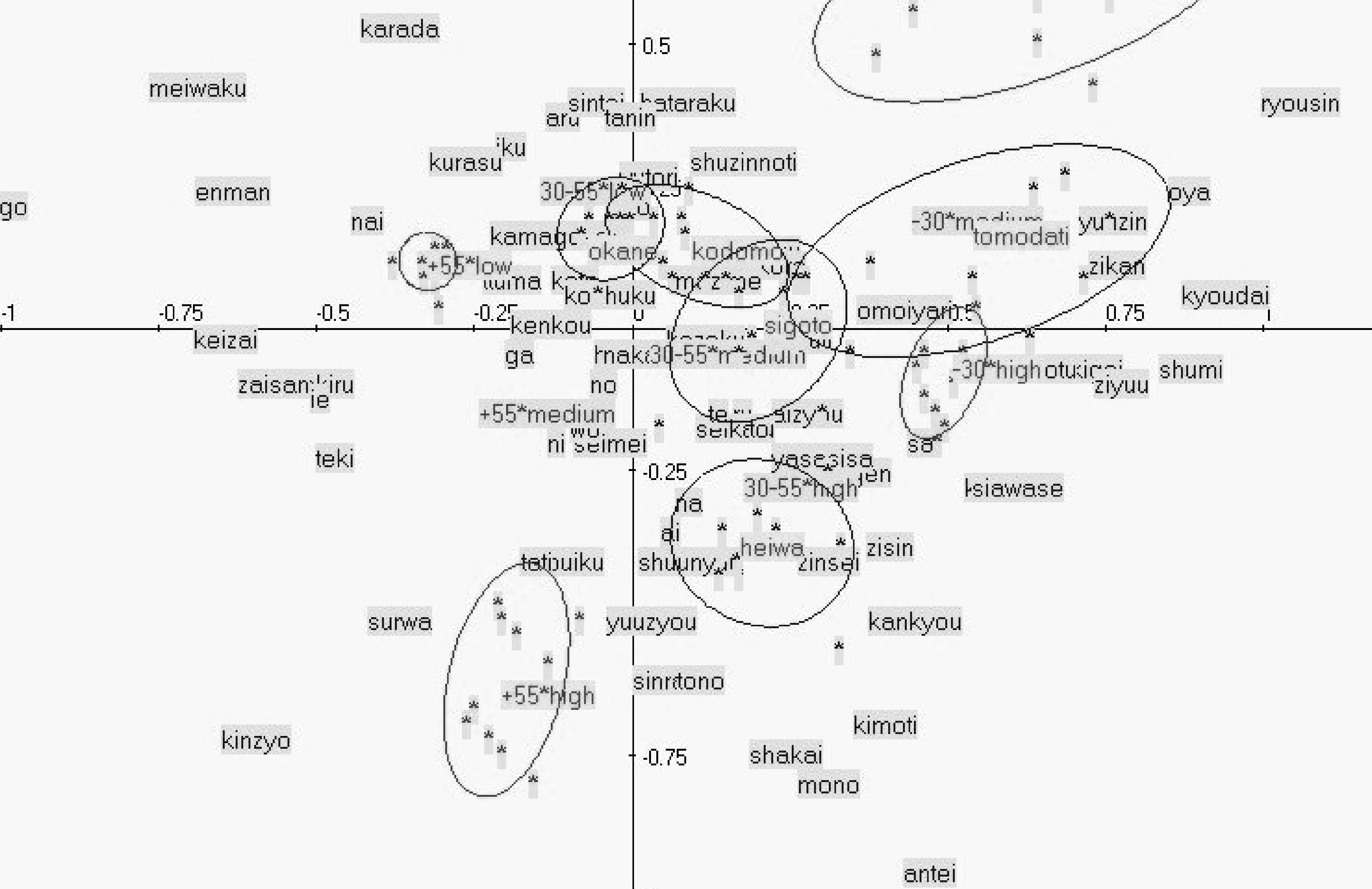
Elliptical confidence areas for 4 categories (in red) and 8 words (in blue) (partial bootstrap)

Total Bootstrap conservative zones

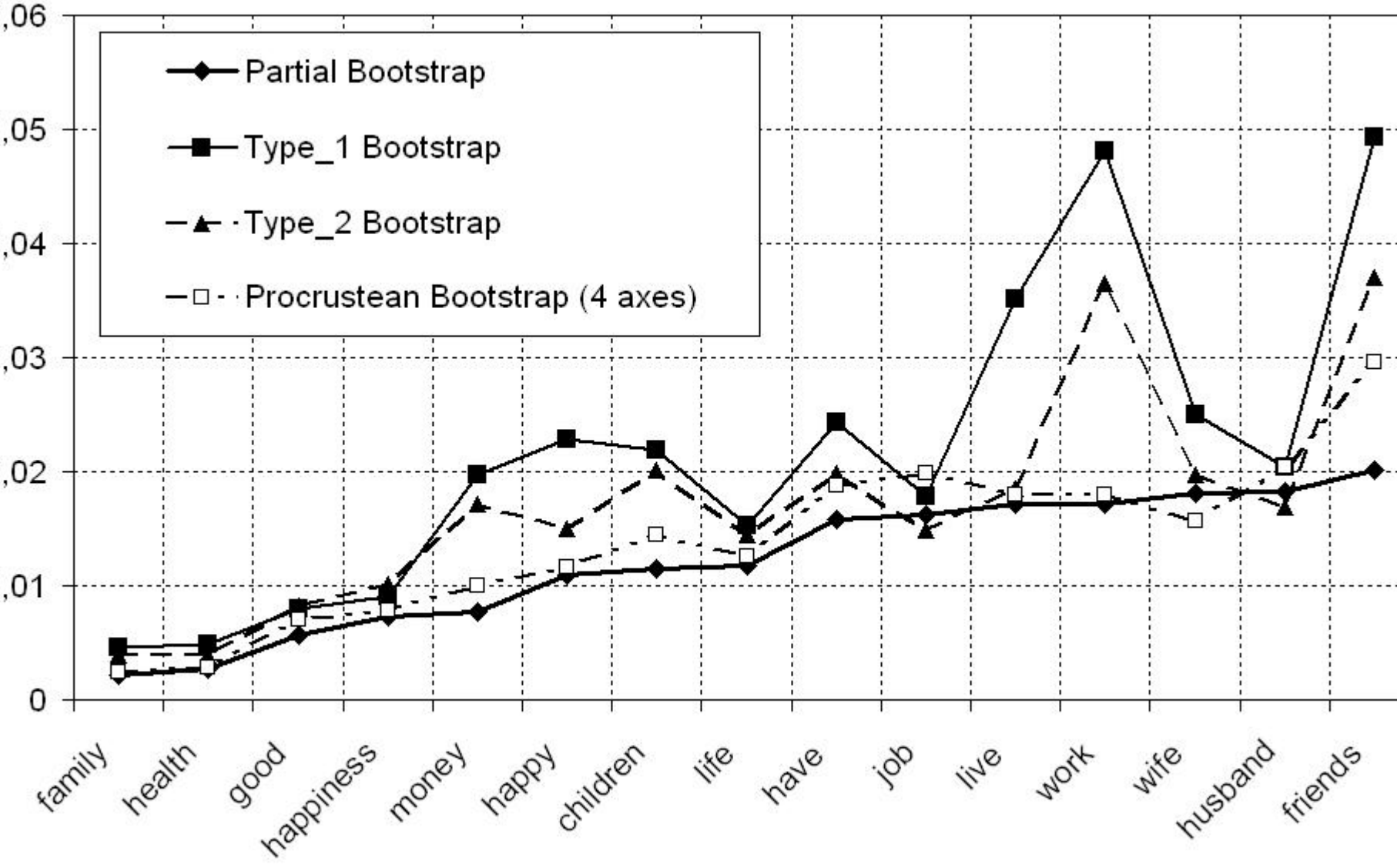



Total Bootstrap conservative zones
0.75
0.5
0.25
-0.25
-0.5
-0.75
-1
-0.75
-0.5
-0.25
0
0.25
0.5
0.75
1
car
-30/high
friends
job
-30/low
+55/low
health
+55/high
peace
mind

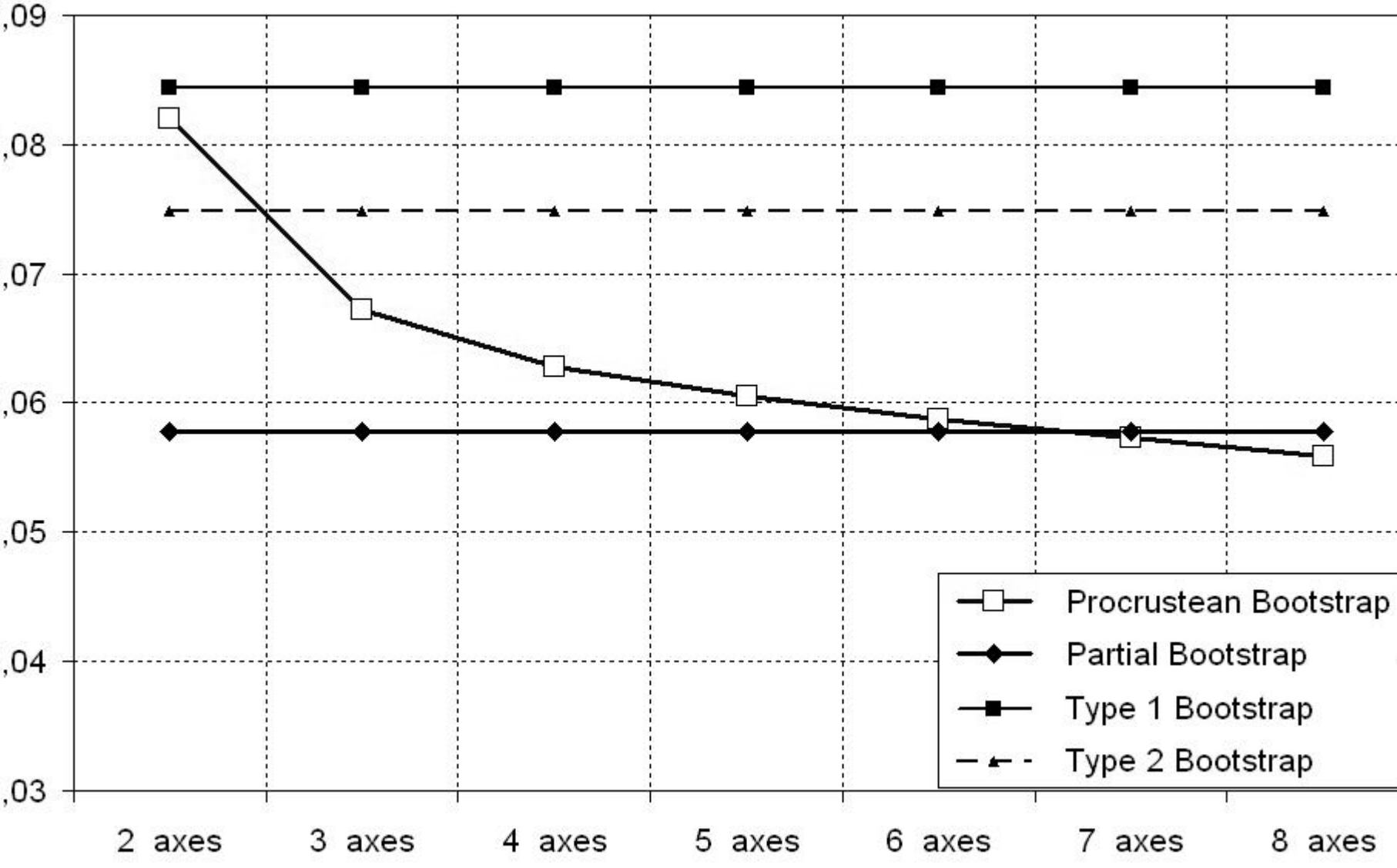
Elliptical confidence areas for 4 categories (in red) and 5 words (in blue) (Japanese sample survey)



Sizes of ellipses for 15 words



Average sizes of ellipses as a function of the dimension (2 to 8 axes)



4. Other types of bootstrap

4.1 Bootstrap on variables

4.2 Specific bootstrap (or hierarchical bootstrap)

4.1 Bootstrap on variables

Such a procedure makes sense when variables are numerous enough.

A potential universe of variables should exist, together with the concept of sample of variables

Variables could be events, moments or instants (time points), geographical stations or areas, words.

In the semiometrics example, the set of analysed words is considered as a sample of words.

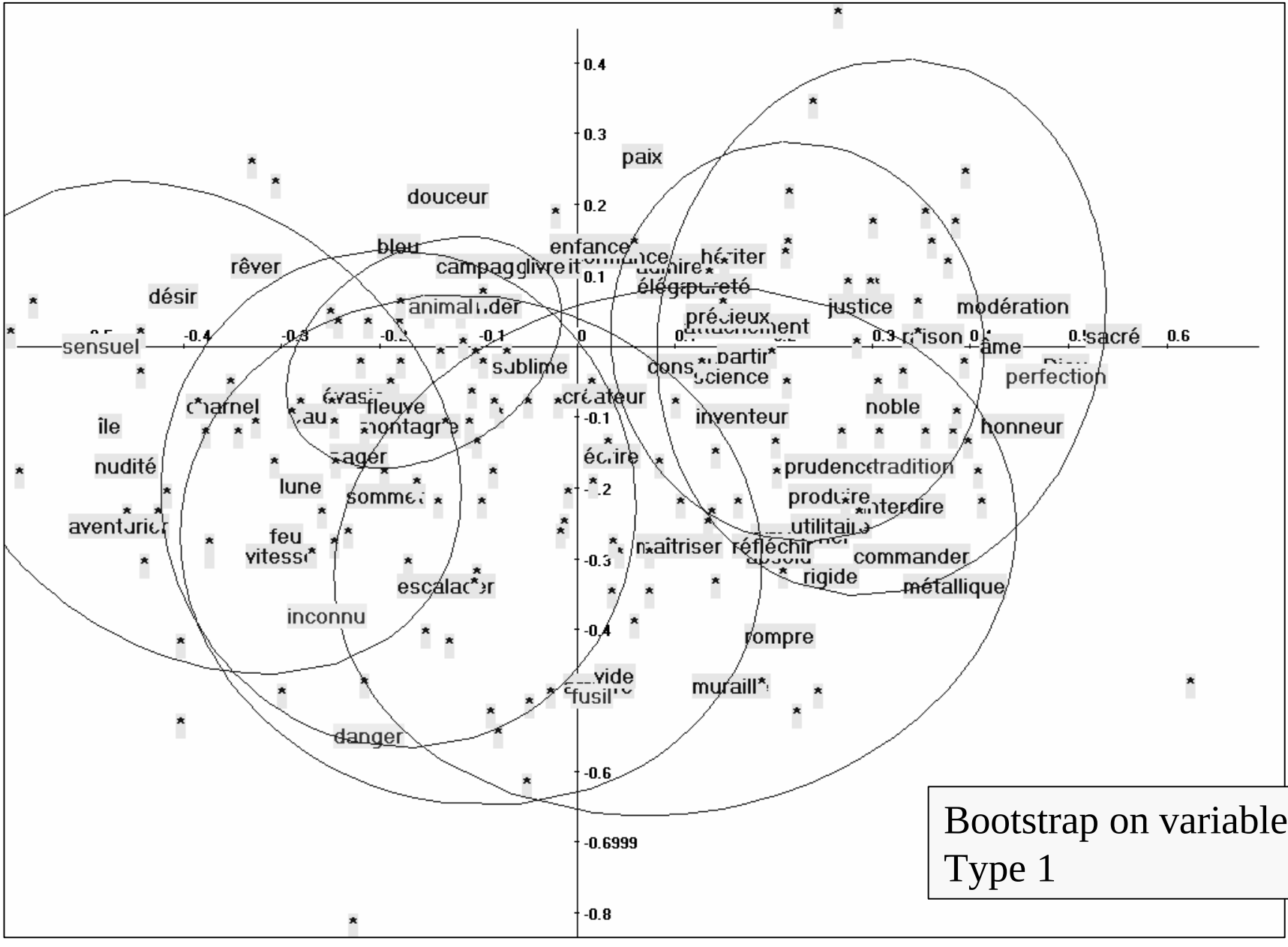
4.1 Bootstrap on variables (*continuation*)

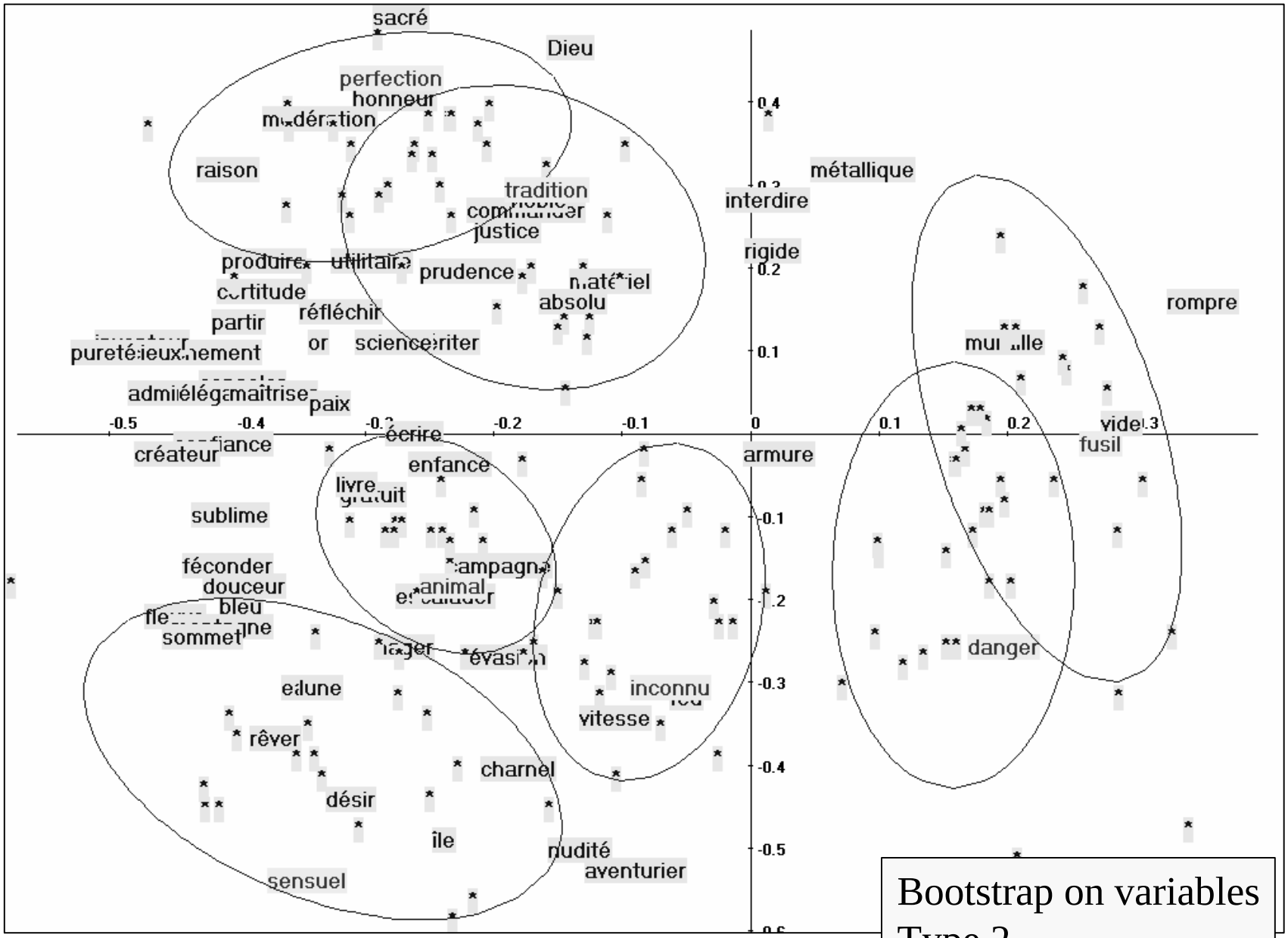
To assess the stability of structures vis-à-vis the set of variables,
the set of variables itself is replicated and analysed through total bootstrap.

Thus, the set of active variables constitutes a sample of *m* variables randomly Drawn from a larger set of potential variables..

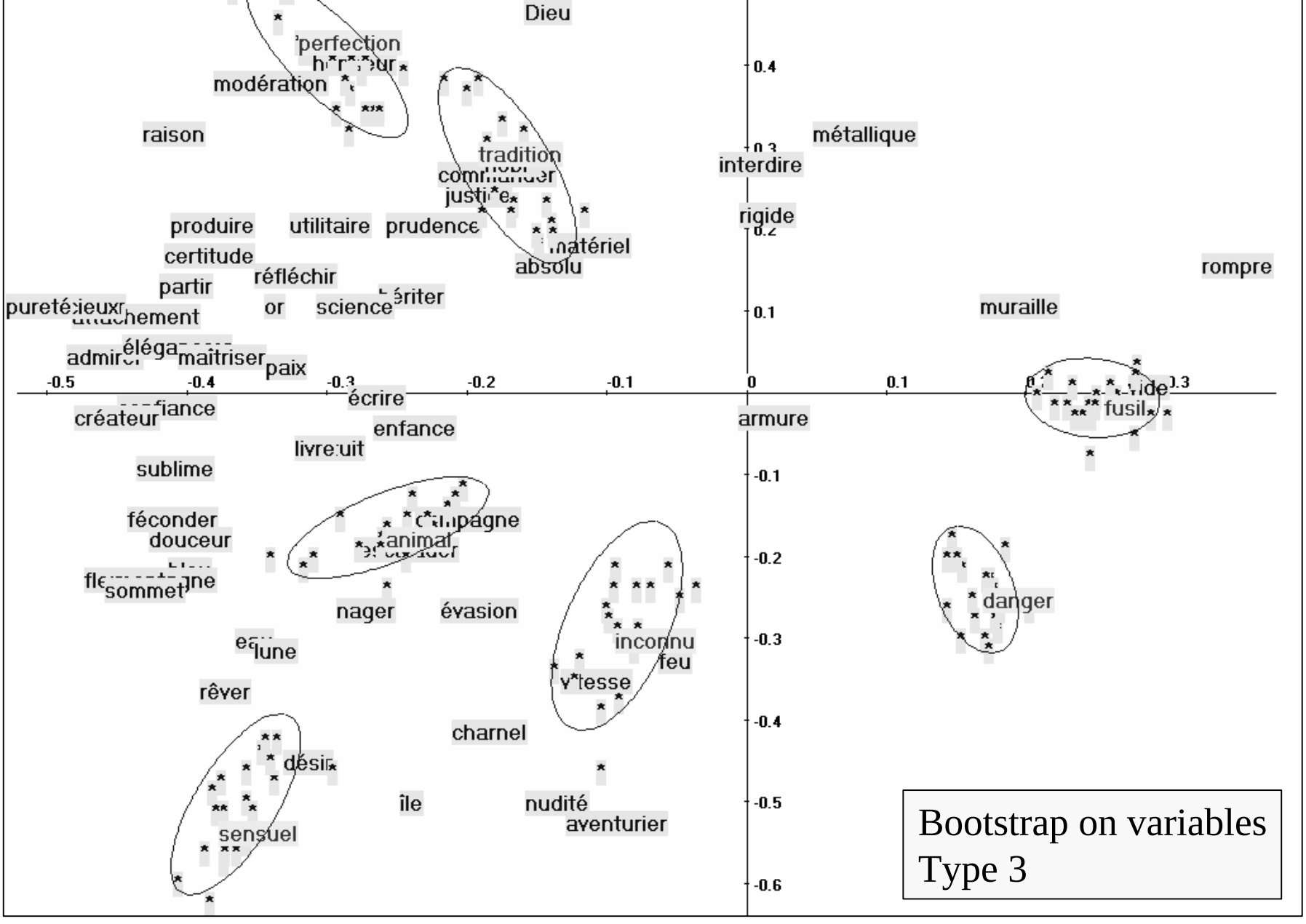
That sample will undergo the same « perturbation » than a sample of observations in the case of bootstrap.

For each replicate, the variables not drawn participate in the analysis with a weight infinitely small (supplémentary variables).


paix
douceur
bleu
enfance
rêver
désir
sensuel
sublime
créateur
fleuve
charnel
île
nudité
lune
feu
aventurier
inconnu
danger
écrire
inventeur
noble
honneur
prudence
tradition
produire
interdire
maîtriser
réfléchir
commander
rigide
métallique
rompre
vide
justice
modération
sacré
âme
perfection
0.4
0.3
0.2
0.1
-0.1
-0.2
-0.3
-0.4
-0.6
-0.6999
-0.8
-0.4
-0.3
-0.1
0
0.3
0.6
Bootstrap on variables
Type 1

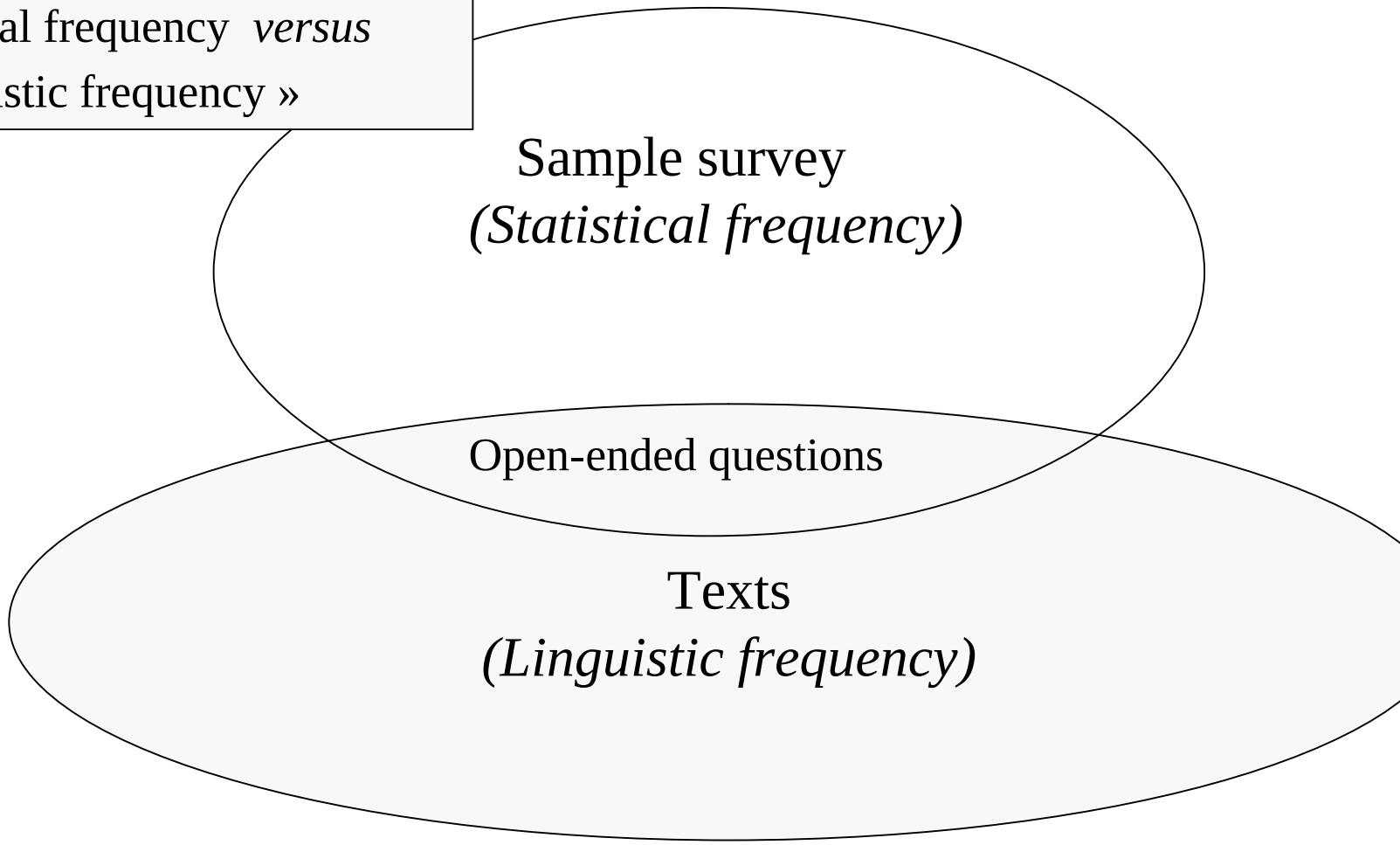


Bootstrap on variables
Type 2

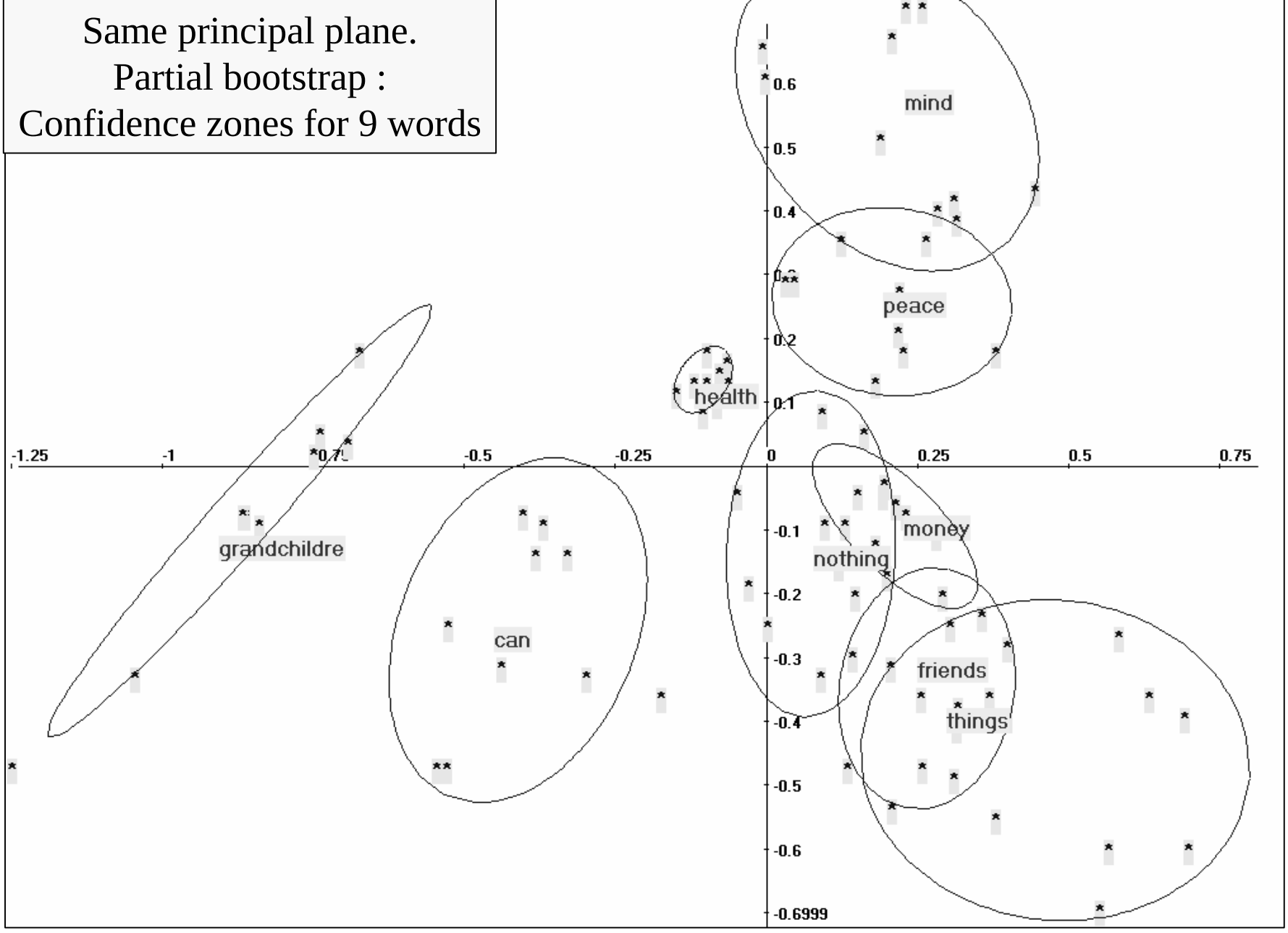

Dieu
perfection
modération
raison
tradition
justice
produire
utilitaire
prudence
matériel
absolu
certitude
réfléchir
partir
science
hériter
attachement
admirer
maîtriser
paix
écrire
métallique
interdire
rigide
rompre
muraille
vide
fusil
armure
créateur
enfance
livre
sublime
féconder
douceur
campagne
animal
sommet
nager
évasion
lune
inconnu
feu
vitesse
danger
rêver
charnel
désir
île
nudité
aventurier
sensuel
-0.5
-0.4
-0.3
-0.2
-0.1
0
0.1
0.4
0.3
0.2
0.1
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
Bootstrap on variables
Type 3

4.2 Specific (or: hierarchical) bootstrap

Textual data
Statistical frequency *versus*
« linguistic frequency »

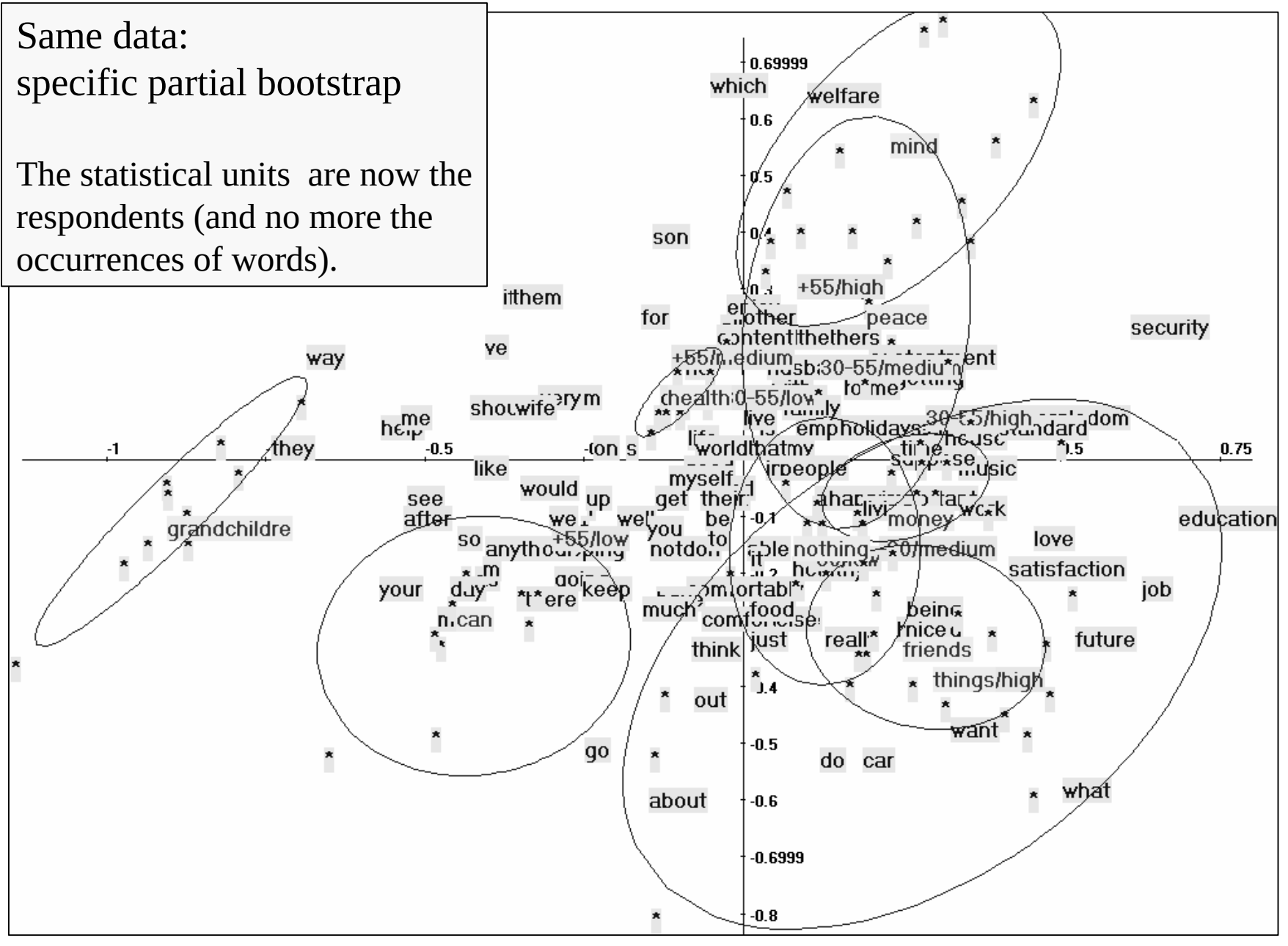


Same principal plane.
Partial bootstrap :
Confidence zones for 9 words



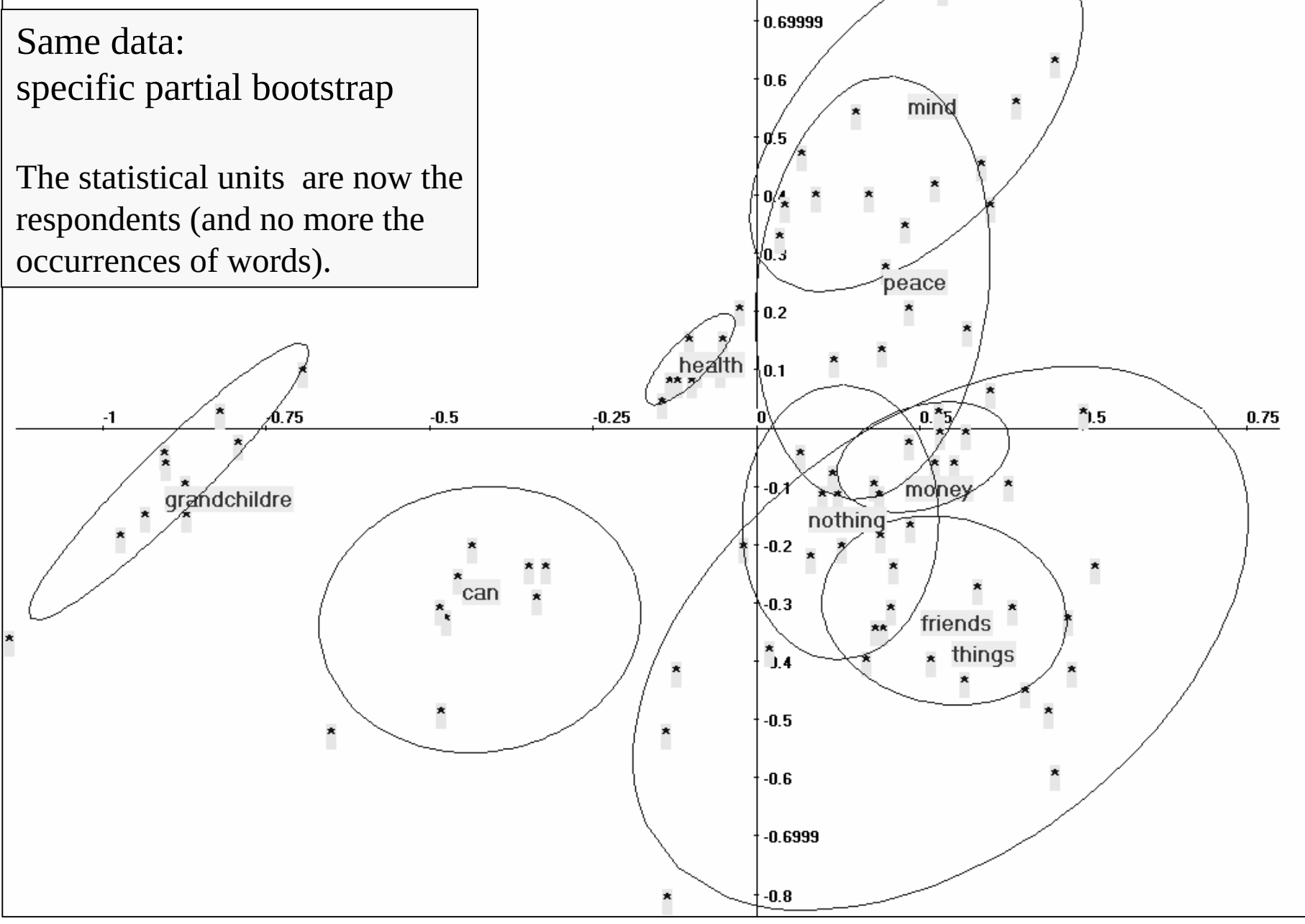
Same data:
specific partial bootstrap

The statistical units are now the respondents (and no more the occurrences of words).



Same data:
specific partial bootstrap

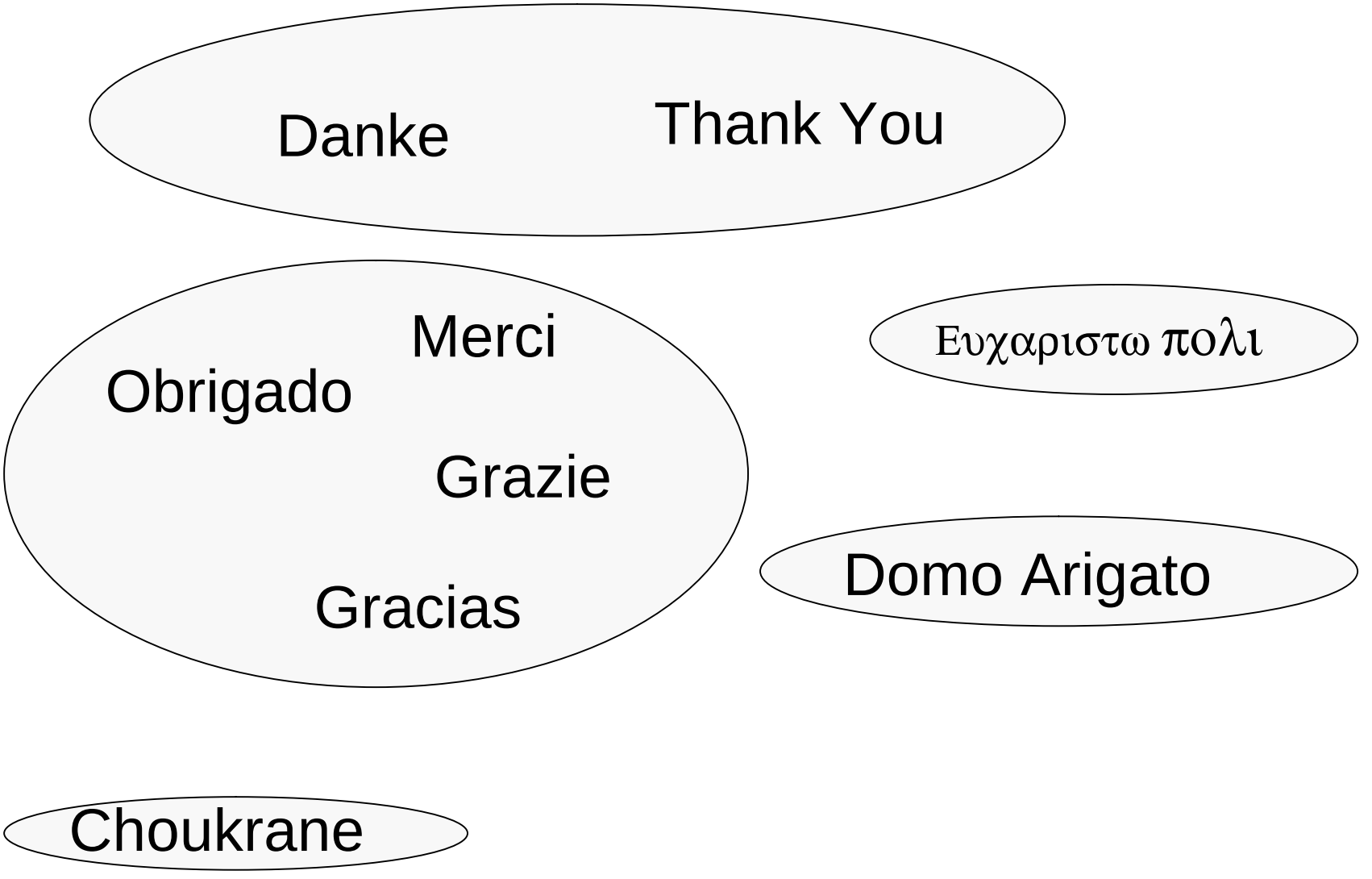
The statistical units are now the respondents (and no more the occurrences of words).



Conclusion

- Various tools, complex strategy
- Interactive implementation needed
- Toward a scientific status for visualizations ?
- Experimental statistics …

The software (DTM) together with the data sets can be freely downloaded from the website of the author.


Danke
Thank You
Merci
Obrigado
Grazie
Gracias
Ευχαριστω πολι
Domo Arigato
Choukrane