3

1965 International Conference on Computational Linguistics

FACTOR-ANALYSIS OF CORRESPONDENCES

Brigitte Cordier

Faculté des Sciences Rennes 35 FRANCE

Rhein. Friedr.-Wilh.-Universität Bonn
1

FACTOR-ANALYSIS OF CORRESPONDENCES

Given two finite sets I and J, we can say that there is a correspondence if the elements i I are associated with those of J by couple . There is a statistical correspondence if for each couple (i, j) there corresponds an integer $\geqslant 0$: be (i, j) . For instance, we can define the statistical correspondence on the set I of nouns appearing in a certain text and the set J of verbs appearing in the same text . The number (i, j) will be equal to the number of times that the noun i is subject of the verb j.

There is a random correspondence if there is defined on IxJ a probability measure described by the read positive function p such that $\sum_{i,j} p(i, j) = 1$. Thus p(i, j) is the paired probability of the couple (i, j) . Usually one studies a statistical correspondence by sampling a random correspondence . We therefore define p(i, j) by :

$p(i, j) = \frac{k(i, j)}{k}$ where $k = \sum_{i, j} k(i, j)$. The purpose of this study is to represent this correspondence geometrically . We are going to associate to each element i of I a point in a Euclidian space of small dimension in a manner in which the distance between these points can be counted by the qualification of the elements associated with the elements j of J i. e. that most of the elements i and i' associated in the same manner as the elements of J more than their images in Euclidian space will be near each other . We will be able to proceed in the same fashion for J and we will be able to then represent simultaneously in the same Euclidian space the sets I and J in such a manner that the i and j with a high p(i, j) be near each other . To represent the set I and the correspondence on I x J defined by the p(i, j) we proceed in the following manner : first construct a "cloud" of points in the space R^I. To each element i of I take a corresponding point e_i of R^I (all the coordinates of e_i are zero except the first which has the value 1) provided with the mass p(i) . We will define on R^I a distance by the following formula :

$$\left[d(i,\ i') \right]^2 = \sum_{i \in J} p(j) \left[\frac{p(\ j/i\ ) - p(\ j/i'\ )}{p(j)} \right]^2$$

where $p(\ i/j) = \frac{p(\ i/j\ )}{p(j)}$ conditional probability .

To obtain the resulting cloud of points representing the set I in a space of weak dimension while conserving a little close the distance between the various points one will make a factori◂ analysis of the cloud i. e. we will determine mathematically the principal directions from which the cloud develops (the problem becomes the classical one of searching for the eigenvectors in the decreasing order of the eigenvalues of a matrix) : we will represent the cloud in this new system of axis thus determined .

We can extend the method to a correspondence between a number of any sort of finite sets . We represent each set in a Euclidian space to be simultaneously all the sets in the same space . We are going to give below some examples of the application of this method . The calculations having being done on a computer .

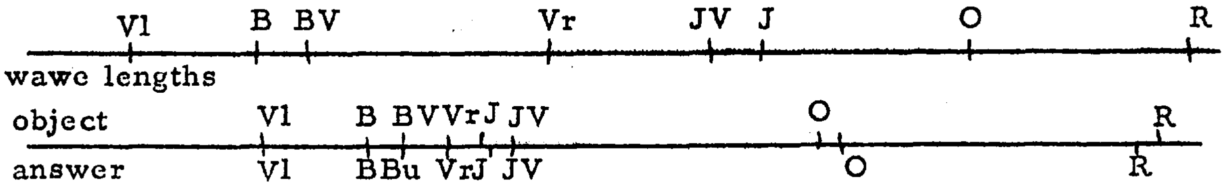
First example :

We are using the results of psychological experience explained below and carried out at the Faculty of Letters of Rennes . One presents to the subjects eight colors projected on a screen . The subjects must learn to associate with the colors eight buttons on a keyboard . We present successively these colors and after the subject has answered, he is told the correct answer . We thus obtains a statistical correspondence ; k(i, j) being the number of times that on the presentation of the color j the subject has answered the button associated with color i. We are able to represent these inputs as a matrix.

	R	O	J	JV	Vr	BV	B	Vl
R	415	45	2	8	7	4	4	3
O	32	373	16	17	8	11	12	8
J	10	12	343	70	22	20	13	10
JV	6	19	50	303	31	23	18	6
Vr	6	12	23	36	305	71	29	8
BV	10	10	15	32	91	274	38	19
B	8	11	14	6	17	60	356	36
Vl	3	5	22	13	11	13	24	403

R : rouge (Red)
O : orange (Orange)
J : Jaune (Yellow)
JV: Jaune-vert (Y - G)
Vr: Vert (Green)
BV: Bleu-Vert (B - G)
B: Bleu (Blue)
Vl : Violet (Purple)

We make a factor-analysis of this correspondence by representing on the same axis (one above, and the other below), the two sets object and response which coïncide here with the set of eight colors : the order of the points is almost perfectly that of the wawe-lengths .



Second Example :

Correspondence between the names of the colors in three languages i.e. a correspondence between three sets,

If we consider the spectrum of colors given by a prism, there is a gradation of the colors from the extremities . Following the language the spectrum is divided into a certain number of colors . For instance, in English it will be : red, orange, yellow, green, blue, purple ,

There is below a diagram representing the divisions of the spectrum in three languages [1] : in English, in a language of Rhodesia, "shona", and in a language of Liberia, "Bassa" .

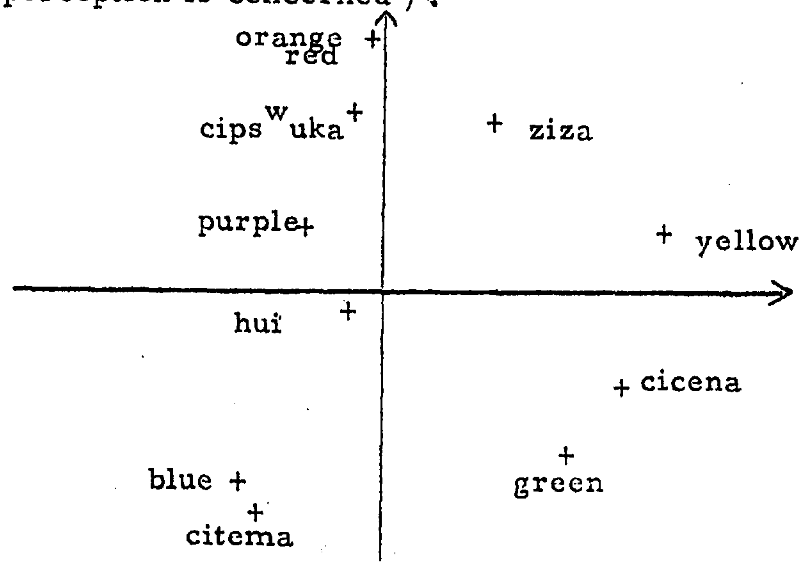
In the second case the spectrum is divided in three parts, (cips^{w}uka being found at the beginning and the end of the spectrum).

In the third case there are only two principal colors .

English	purple	blue	green	yel-low	orange	red
Shona	cips^{w}uka	citema	cicena		cips^{w}uka	
Bassa	hui			ziza		

To make a correspondence between the three set, we will give coefficients to the common length of the spectrum for the three colors in three languages.

To make the factor-analysis of this correspondence : we extract the two axis that which gives me on a plane diagram. We have represented simultaneously the three sets . We see that the colors divide themselves nearly on a circle according to the part of the spectrum which they cover (red being close to purple which is not true of wawe lengths but quite natural as far as perception is concerned) .



1 . (Data from H. A. Gleason " An introduction to descriptive linguistics" Holt, Rinehart, Winston, editors, N.Y. 1961) .

L'analyse factorielle des correspondances est une technique nouvelle qui permet de traiter des informations statistiques, notamment dans le domaine linguistique .

Si l'on connaît par exemple pour un texte donné, le nombre de fois que chacun des adjectifs de ce texte est épithète de chacun des noms de ce texte, il est naturel de considérer que d'une part deux adjectifs sont d'autant plus proches sémantiquement l'un de l'autre qu'ils s'associeront dans les mêmes proportions aux mêmes noms et d'autre part qu'un nom et un adjectif ont d'autant plus de points communs qu'ils s'associeront le plus souvent ensemble .

De même si on a les résultats d'une expérience psychologique consistant à associer des réponses à un certain nombre de stimuli, deux stimuli seront proches s'ils attirent les réponses dans les mêmes proportions, inversement deux réponses seront proches si elles s'associent aux mêmes stimuli . D'autre part, un stimulus sera proche des réponses auxquelles il sera associé le plus souvent .

Il y a bien d'autres exemples de ce type .

Le but ici, est de préciser cette notion de proximité donnée sur des ensembles quelconques par leurs relations statistiques, de définir formellement sur ces ensembles une distance qui rende compte de cette proximité, puis de dégager les diverses composantes ou facteurs de cette proximité, composantes qui dépendent de la nature des données, et enfin de représenter graphiquement les ensembles considérés munis de la distance ainsi définie .

Nous partons donc d'une correspondance statistique entre deux ensembles que l'on note I et J avec i élément de I et j élément de J . C'est-à-dire que les éléments de ces deux ensembles sont reliés par couple . La correspondance est définie par la donnée pour tout couple (i, j) d'un nombre entier positif ou nul que l'on note k(i, j) .

Ce nombre provient en général d'un corpus de couples (i, j). Pour notre premier exemple, I était l'ensemble des noms, J celui des adjectifs, le corpus le texte donné et k(i, j) le nombre de fois que dans ce corpus le nom i a pour épithète l'adjectif j .

On note $k(i) = \sum_{j \in J} k(i, j)$ et $k(j) = \sum_{i \in I} k(i, j)$

qui sont ici respectivement égaux au nombre de fois que le nom i et l'adjectif j apparaissent dans le corpus .

On note aussi $k = \sum_{i,j} k(i, j)$ l'effectif total de la correspondance .

On peut considérer cette correspondance statistique comme une correspondance aléatoire, le corpus en étant un échantillon . On aura alors estimé les probabilités d'apparition du couple (i, j), de l'élément i et de l'élément j en posant :

$$p(i,j) = \frac{k(i,j)}{k} \qquad \text{avec} \quad \sum_{i,j} k(i,j) = 1$$

$$p(i) = \frac{k(i)}{k} = \sum_{j \in J} p(i,j)$$

$$p(j) = \frac{k(j)}{k} = \sum_{i \in I} p(i,j)$$

Et $p(i/j) = \dfrac{p(i,j)}{p(j)}$ et $p(j/i) = \dfrac{p(i,j)}{p(i)}$

probabilités conditionnelles d'apparition de i, j étant fixé et de j , i étant fixé .

Nous voulons donc construire une distance qui exprime la notion de proximité dont on a parlé ci-dessus . Plus précisément la distance entre deux éléments i et i' de I doit être d'autant plus petite que ces éléments ont des probabilités conditionnelles de s'associer aux éléments de J semblables . A la limite cette distance définie uniquement par la correspondance entre I et J sera nulle si i et i' ont les mêmes probabilités conditionnelles .

On veut de plus que si l'on remplace deux éléments j_1 et j_2 de J par un seul élément j_0 tel que quel que soit i :

$$p(i,j_0) = p(i',j_1) + p(i, j_2)$$

la distance entre les éléments de I soit inchangée .

De plus cette distance ne doit pas dépendre de la fréquence ou de la rareté des apparitions respectives de i et i' (de p(i) et de p(i')).

D'où la formule :

$$d^2_{I(J)} \quad (i,i') = \sum_{j \in J} \frac{1}{p(j)} \left[p(j/i) - p(j'/i') \right]^2$$

L'ensemble J est étudié d'une façon symétrique, ce qui nous donne une formule symétrique pour la distance .

Quant à la distance entre un élément de J et un élément de J on verra par la suite comment on obtient que i soit à peu près barycentre des j affectés des masses p(i,j).

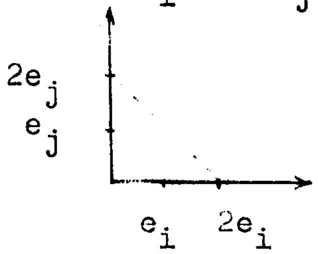
Cette distance précisée, nous voulons maintenant représenter les ensembles I et J dans un espace de petite dimension un axe et un plan par exemple en conservant autant que possible la distance ainsi définie .

Pour cela, on construit des nuages de points $\mathcal{N}_I$ pour représenter I , $\mathcal{N}_J$ pour représenter J et $\mathcal{N}_{IJ}$ pour représenter simultanément I et J dans un espace muni d'une forme quadratique telle que la distance qu'elle définit entre deux points du nuage soit égale à celle que nous avons définie ci-dessus.

Pour I on construira la nuage dans $\mathcal{R}^I$ et l'élément i sera représenté au vecteur de base e_i, il sera muni de la masse p(i) . Quant à la forme quadratique que l'on note $Q_{J(I)}$ elle a pour valeur :

$$Q_{I(J)} (e_i, e_{i'}) = \sum_{j \in J} p(j) \frac{p(i,j)}{p(i)p(j)} \frac{p(i',j)}{p(i')p(j)}$$

Pour J on a une construction symétrique et pour d^{J}_{IJ} on se place dans $R^I x R^J$ les deux axes étant orthogonaux et munis respectivement des normes associées à $Q_{I(J)}$ et $Q_{J(I)}$ et on considère le nuage des points (e_i, e_j) munis des masses p(i,j) . Le point (e_i, e_j) est au milieu du segment joignant les points $2e_i$ et $2e_j$ qui représentent respectivement i et j .



Pour représenter ces ensembles dans un espace de petite dimension on va ajuster au nuage étudié un sous espace et représenter le nuage dans ce sous espace .

Pour cela on va déterminer les directions principales dans lesquelles s'allongent le nuage . D'une façon précise , ce sera les axes principaux d'inertie extraits dans l'ordre décroissant des moments principaux d'inertie . L'espace étant muni d'une forme quadratique Q différente du produit scalaire, nous avons une méthode pour résoudre le problème dans ce cas .

Les axes principaux sont orthogonaux relativement à la forme quadratique Q . A chaque axe principal s correspond une forme linéaire ou facteur qui est la projection relativement à Q sur det axe qui est F = Q(s) . Ces axes normalisés par rapport à Q sont les vecteurs de base du sous espace dans lequel sera représenté le nuage (c'est-à-dire l'ensemble) .

Dans le cas du nuage $\mathcal{N}_I$ par exemple, si nous avons extrait deux facteurs f_1 et f_2, l'ensemble I sera représenté par un diagramme plan où l'élément i aura pour coordonnées $f_1(e_i)$ et $f_2(e_i)$. Plus on extraira d'axes, plus la distance sera représentée d'une façon précise . Mais il arrive un moment où la part du moment d'inertie du nuage non représentée par les facteurs déjà extraits est de l'ordre des aléas statistiques et les résultats obtenus ne sont plus significatifs. Pour déterminer cela, on dispose d'un test de comparaison à un χ^2 et la plupart du temps un ou deux facteurs suffisent à rendre compte de la correspondance .

Les représentations de l'ensemble I, de l'ensemble J et la représentation simultanée des deux ensembles ne sont pas indépendantes . Plus précisément, il a été montré les relations suivantes, que l'on se contente ici d'énoncer, entre les nuages $\mathcal{N}_I$, $\mathcal{N}_J$ et $\mathcal{N}_{IJ}$:

- les nuages $\mathcal{N}_I$ et $\mathcal{N}_J$ ont les mêmes moments principaux d'inertie et pour un même moment principal λ les facteurs $\mathcal{F}_I^\lambda$ et $\mathcal{F}_J^\lambda$ de $\mathcal{N}_I$ et de $\mathcal{N}_J$ se déduisent l'un de l'autre de la façon suivante :

$$\mathcal{F}_I^\lambda (e_i) = \frac{1}{\sqrt{\lambda}} \sum_j \frac{p(i,j)}{p(j)} \mathcal{F}_J (e_j)$$

- Si on note λ_q les moments principaux d'inertie de $\mathcal{N}_I$ et de $\mathcal{N}_J$, ceux de $\mathcal{N}_{IJ}$ sont tous de la forme $(\lambda_q + \varepsilon \lambda_q \sqrt{\lambda_q})$ où $\varepsilon = \pm 1$

Ceux de la forme $\lambda_q + \lambda_q \sqrt{\lambda_q}$ sont les plus grands et donc les seuls qui nous intéressent pour l'analyse factorielle du nuage $\mathcal{N}_{IJ}$.

Le facteur de $\mathcal{N}_{IJ}$, forme linéaire sur $R^I x R^J$ associé au moment $(\lambda_q + \varepsilon \lambda_q \sqrt{\lambda_q})$ est de la forme : $\mathcal{F} = \mathcal{F}_I, \varepsilon \mathcal{F}_J)$ où $\mathcal{F}_I$ et $\mathcal{F}_J$ sont les facteurs de $\mathcal{N}_I$ et $\mathcal{N}_J$ associés à λ_q.

L'image de l'élément i de I par $\mathcal{F}$ dans la représentation simultanée sera donc au point :

$$\mathcal{F}(e_i) = \mathcal{F}_I (e_i)$$

L'image de l'ensemble I dans la représentation simultanée obtenue à partir de $\mathcal{N}_{IJ}$ est donc la même que celle obtenue dans la représentation simple à partir de $\mathcal{N}_I$.

De même pour l'ensemble J .

De plus on a la relation :

$$\mathcal{F}(e_i) = \frac{1}{\sqrt{\lambda}} \sum_j \frac{p(i,j)}{p(i)} \mathcal{F}(e_j)$$

Ce qui signifie qu'au coefficient $\sqrt{\lambda}$ près , le point $\mathcal{F}(e_i)$ est le barycentre des points $\mathcal{F}(e_j)$ affectés des masses p(i,j) . Le point i sera d'autant plus proche du barycentre des j que λ est plus proche de 1 . Or le moment d'inertie λ est toujours inférieur à 1 . Il en sera donc d'autant plus proche que λ sera plus grand, c'est-à-dire que le facteur considéré rend compte d'une plus grande proportion de la variance totale .

Cette formule permet d'autre part, si l'on veut ajouter à I un nouvel élément i qui modifie peu la correspondance entre I et J, sans refaire tous les calculs de facteurs .

Ces résultats nous montrent aussi que l'étude des nuages $\mathcal{N}_I$ ou $\mathcal{N}_J$ n'a qu'un intérêt technique . Elle permet d'obtenir les résultats de l'analyse simultanée en maniant des données moindres que celles qu'exigerait l'étude de $\mathcal{N}_{IJ}$, puisque ces facteurs se déduisent les uns des autres d'une manière simple . Mais les résultats de l'étude de I ou de J seuls sont compris dans les résultats de l'analyse simultanée et nous ne montrerons donc que ces derniers dans les exemples exposés .

La méthode qui a été exposée pour une relation binaire, c'est-à-dire pour une relation entre deux ensembles I et J peut être généralisée au cas où l'on a un nombre quelconque d'ensembles en correspondance . En particulier pour une correspondance ternaire entre I,J,K définie par les nombres $p(i,j,k)$ donnés pour chaque triple .

Dans ce cas pour représenter I on étudie dans R^I le nuage de points des e_i munis des masses $p(i)$. La distance est définie par une forme quadratique notée $Q_{I(J+K)} = Q_{I(J)} + Q_{I(K)}$ où $Q_{I(J)}$ et $Q_{J(K)}$ sont les formes quadratiques des correspondances binaires entre I et J et entre I et K déduites de la correspondance ternaire :

Pour I et J : $p(i,j) = \sum_{k \in K} p(i,j,k)$

De la même façon, on aura une représentation de chacun des autres ensembles J et K .

On peut faire également une représentation simultanée de I,J,K en considérant le nuage de points des (e_i, e_j, e_k) munis des masses $p(i,j,k)$ dans $R^I x R^J x R^K$ les axes principaux R^I, R^J, R^K étant orthogonaux et munis des normes associées respectivement à $Q_{I(J+K)}$, $Q_{J(I+K)}$, et $Q_{K(I+J)}$.

Des programmes ont été écrits pour le cas binaire et pour le cas ternaire . Un certain nombre de données ont été traitées au laboratoire de calcul de RENNES sur une IBM 1620 . Les pages suivantes montrent quelques exemples des résultats que l'on peut obtenir avec cette nouvelle méthode d'analyse factorielle .

EXEMPLE 1.-

Dans un article intitulé "Statistique linguistique et histoire du vocabulaire" paru dans les cahiers de lexicologie 1960 . G. Gougenheim étudie les verbes dont le sens général est "briser" et leurs compléments dans la chanson de Roland . On donne ici sous forme de matrice les données statistiques, c'est-à-dire le nombre de fois que tel verbe à pour complément tel nom dans le poème .

Le test que l'on utilise nous montre que le premier facteur extrait est le seul significatif . Le graphique représentant les deux ensembles est donc une droite .

On sépare nettement les noms en deux classes, objets mous et objets durs . Les verbes sont aussi séparés en deux classes, chacune d'elles étant placée près de la classe de ses compléments .

Les données traitées ici étaient très simples mais on voit la possibilité, en traitant des listes plus étendues, ce qui donnerait sans doute lieu à l'apparition de plusieurs facteurs correspondant à des nuances de sens différentes, de représenter graphiquement le sens d'un nom .

	Freindre	Rompre	Froisser	Derompre	Briser	Pecoier	Depecer	casser
Ecu	15	0	3	0	0	1	0	1
Hanste	5	0	3	0	2	0	1	0
Haubert	0	5	0	3	0	0	0	0
Heaume	4	0	1	0	0	0	0	1
Durandal	3	0	1	0	2	0	0	0
Tempe	0	3	0	0	0	0	0	0
Pans du haubert	0	2	0	1	0	0	0	0
Tarje	2	0	0	0	0	0	0	0
Cités	2	0	0	0	0	0	0	0
Murs	0	0	1	0	0	1	0	0
Sangles de la selle	0	1	0	0	0	0	1	0
Ventaille du haubert	0	1	0	1	0	0	0	0

Correspondance verbes-noms : Données

EXEMPLE 2 .-

Il s'agit ici d'une correspondance ternaire dont les données ont été calculées d'après H.A. Gleason "An introduction to descriptive linguistics " .

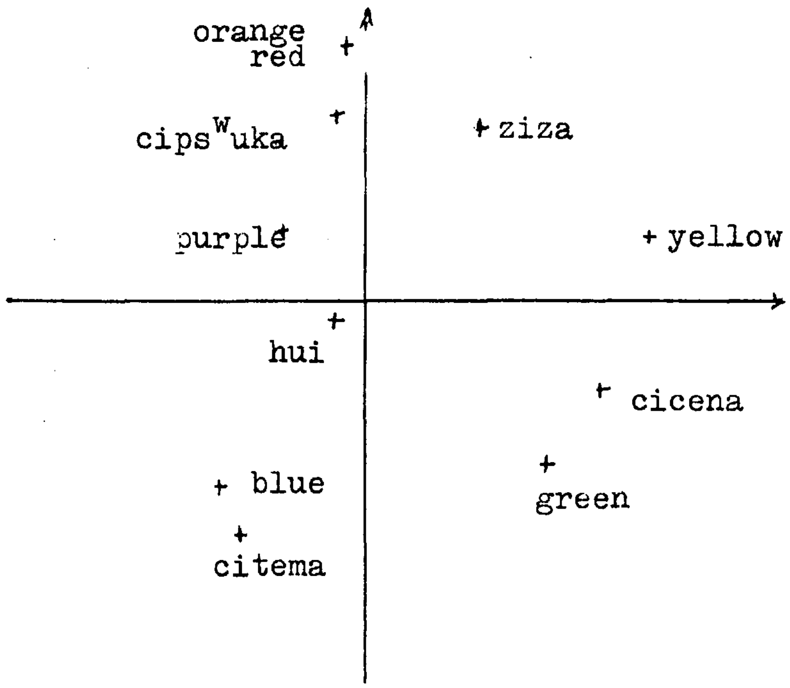
Chaque langue possède un certain nombre de noms de couleurs de base correspondant à une zone précise du spectre et le recouvrant tout entier . Les portions du spectre définies par ces couleurs ne sont pas les mêmes dans toutes les langues .

Voici d'après Gleason comment l'Anglais, le Shona (Rhodésie) et le Bassa (Liberia) divisent le spectre :

Anglais	purple	blue	green	yellow	orange	red
Shona	cips^{w}ukacitema		cicena		cips^{w}uka	
Bassa	hui			ziza		

On en a déduit une correspondance ternaire entre les trois ensembles de noms de couleurs : $\mathbb{K}$ (i,j,k) est la longueur de portion commune du spectre pour les couleurs i, j, et k . On remarque que en "Shona" les deux extrémités du spectre sont appelées par le même nom .

Les calculs faits, on a obtenu deux facteurs , donc une représentation dans le plan des 3 ensembles.



Les couleurs se rangent exactement dans l'ordre de leur fréquence moyenne autour du centre du graphique .

On voit que l'on est parvenu à retrouver la structure des ensembles étudiés .

Nous avons utilisé le spectre mais nous aurions eu des données statistiques équivalentes en présentant à trois personnes parlant respectivement chacune des trois langues une série d'objets colorés répartis uniformément dans le spectre et en notant le nombre de fois que devant un même objet les couleurs i,j, k ont été prononcées . Ceci pourrait être fait pour d'autres mots et permettrait peut-être de représenter graphiquement les sens de mots comparables dans plusieurs langues .

EXEMPLE 3.-

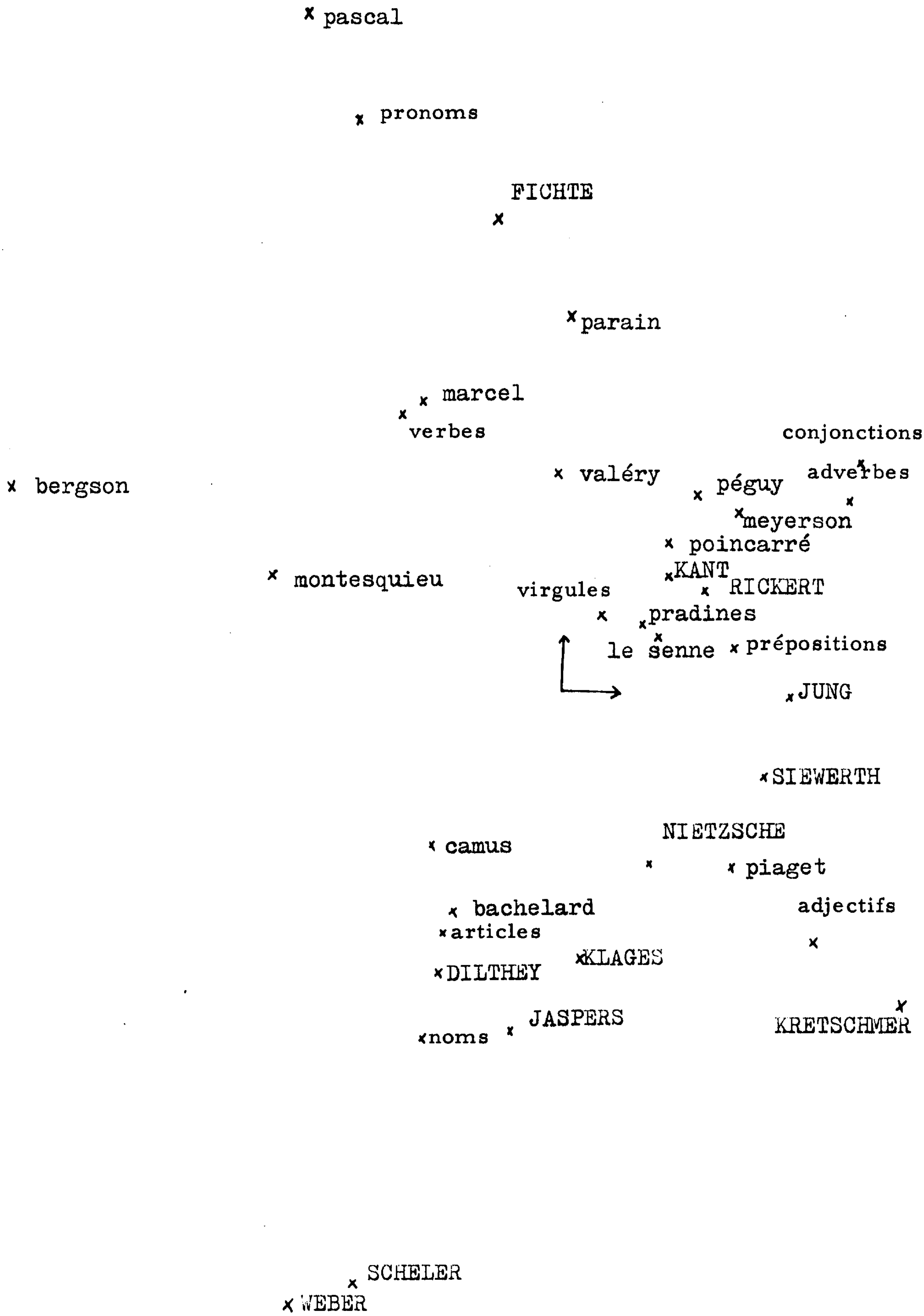
On a utilisé des résultats statistiques effectués en 1957 par J.M. Zemb et rassemblés dans un rapport de la Faculté Philosophique de Hambourg .

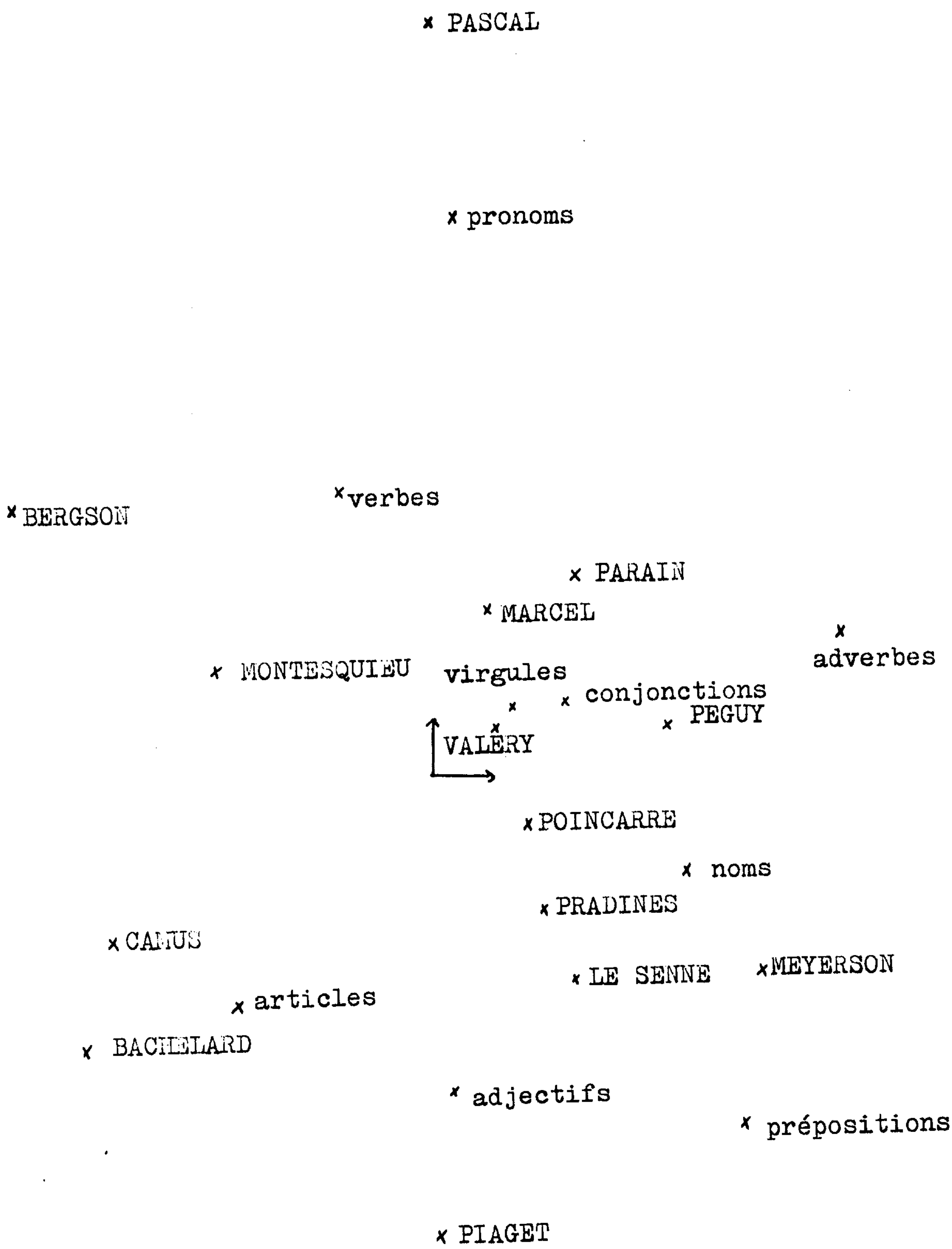
Ces données concernent l'utilisation des différentes parties du discours, noms, adjectifs, virgules etc. par une série de seize auteurs français et de auteurs allemands . Pour chaque auteur, on a le pourcentage de chacune des parties du discours qu'il utilise sur des textes de 10 000 mots en moyenne .

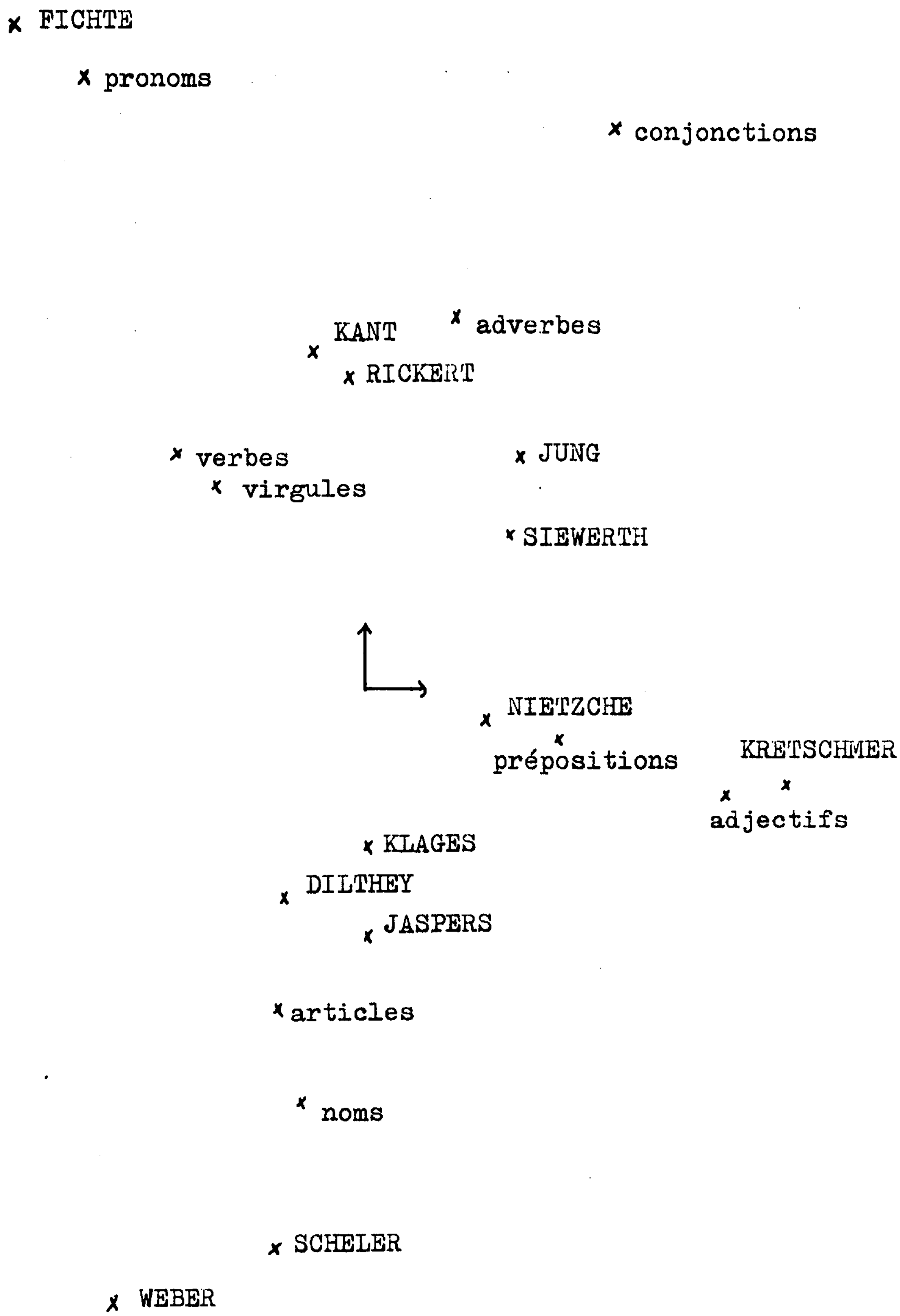
Ce qui nous donne une correspondance entre l'ensemble des auteurs étudiés et l'ensemble des parties du discours . On a fait trois analyses différentes, l'une en ne considérant que les auteurs français, l'autre que les auteurs allemands et la troisième en les considérant tous . On a représenté de plus dans cette dernière le centre de gravité des auteurs allemands et celui des auteurs français .

Le facteur principal oppose le groupe nominal (nom, adjectif) au groupe verbal (verbe, adverbe) . Le second facteur s'interprète bien chez les auteurs allemands et sur les auteurs des deux langues rassemblées (mais non en Français) . Il oppose les éléments suppressibles (adjectifs, adverbes) aux autres éléments (noms, verbes) .

On peut aussi comparer l'utilisation des parties au discours dans les deux langues . On remarque que les auteurs français se trouvent représentés du côté des "vertus", "pronoms" alors que les auteurs allemands sont représentés plus près des"noms","articles", "adjectifs" .


pascal
pronoms
FICHTE
parain
marcel
verbes
conjonctions
bergson
valéry
péguy
adverbes
meyerson
poincarré
KANT
RICKERT
montesquieu
virgules
pradines
le senne
prépositions
JUNG
SIEWERTH
NIETZSCHE
camus
piaget
bachelard
adjectifs
articles
KLAGES
DILTHEY
JASPERS
noms
KRETSCHMER
SCHELER
WEBER


PASCAL
pronoms
BERGSON
verbes
PARAIN
MARCEL
adverbes
MONTESQUIEU
virgules
conjonctions
PEGUY
VALÉRY
POINCARRE
noms
PRADINES
CAMUS
LE SENNE
MEYERSON
articles
BACHELARD
adjectifs
prépositions
PIAGET


FICHTE
pronoms
conjonctions
KANT
adverbes
RICKERT
verbes
virgules
JUNG
SIEWERTH
NIETZCHE
prépositions
KRETSCHMER
adjectifs
KLAGES
DILTHEY
JASPERS
articles
noms
SCHELER
WEBER

EXEMPLE 4.-

L'étude de la correspondance ternaire demande des calculs volumineux que nous n'avons pu encore faire pour aucune langue. Les seuls résultats précis actuellement disponibles concernent la correspondance binaire définie par les couples de phonèmes successifs des mots d'un rexte espagnol (cette langue a été choisie parce que la liste des phonèmes est assez courte). Le premier facteur sépare nettement les voyelles des consonnes .

On a :

E A O I Ú Y B G ....

toutes les consonnes

Dans notre transcription phonémique on avait distingué le yod (noté Y : PYADOSO), du i (PIO) : le yod apparaît comme la plus vocalique des consonnes .

Importance des triples :

C'est sur des triples que se basa V. Thomsen dans sa fameuse traduction des inscriptions de Orkhon : pour séparer dans l'alphabet inconnu, les consonnes des voyelles, il postula que dans les triples 121 fréquents (triples formés d'un signe entre deux identiques), 1 devait être consonne si 2 était voyelle et réciproquement . (cf e.g. O. Jespersen p. 800) .

O. Jespersen : Selected Writings of O. Jespersen : G. Allen and Unwin, London ; Senjo, TOKYO .

EXEMPLE 5.-

On utilise les résultats de l'expérience psychologique expliquée ci-dessous et faite à la section de psychologie de la Faculté des Lettres de Rennes .

On présente à des sujets huit couleurs projetées sur un écran . Les sujets doivent apprendre à associer aux couleurs les huit boutons d'un clavier . Les huit couleurs sont présentées successivement . Après que le sujet ait répondu, on lui indique la réponse exacte . On note les résultats et on établit une matrice de confusion entre couleurs dont les coefficients sont calculés ainsi : k_{ij} est le nombre de fois que sur présentation de la couleur j, le sujet a répondu la couleur i .On a fait deux matrices, l'une correspondant au début de l'apprentissage et l'autre à la fin de l'apprentissage .

Les couleurs sont représentées ainsi :

ROUGE : R	ORANGE : O	JAUNE : J
JAUNE-VERT : JV	VERT : Vr	BLEU-VERT : BV
BLEU : B	VIOLET : Vl	

On a extrait un facteur . On a représenté sur le premier axe les longueurs d'ondes des couleurs utilisées et sur le second les résultats de l'analyse factorielle . On a retrouvé à part une interversion entre deux couleurs très proches jaune et jaune-vert, la structure de l'ensemble des stimuli, celle de l'ensemble des réponses et la proximité des stimuli et des réponses .

PRESENTATION DES RESULTATS DE L'ANALYSE FACTORIELLE .

DEBUT

FIN