

Revue Internationale de

ISSN 0980-1472

systemique

NUMÉRO SPÉCIAL
« BIOLOGIE ET SYSTÉMIQUE »

Vol. 2, N° **4**, 1988

afcet

Dunod

AFSCET

Revue Internationale de
systemique

Revue
Internationale
de Sytémique

volume 02, numéro 4, pages 479 - 504, 1988

Sur la recherche des filiations en phylogénèse

Pierre Delattre

[Numérisation Afscet, janvier 2016.](#)



Creative Commons

- (39) WEINBERG A., «On the relation between information and energy systems», *Interdisciplinary Science Reviews*, 7, 47-52, 1982.
- (40) FROHLICH H., «Evidence for Bose condensation-like excitation of coherent modes in biological systems», *Physics Letters*, 1975, 51A, 1, 21-22.
- (41) STAHL W.R., «Dimensional analysis in mathematical biology. I. General Discussion», *Bull. Math. Biophysics*, vol. 23, p. 355-376, 1961.

ARCHIVES

SUR LA RECHERCHE DES FILIATIONS EN PHYLOGENESE ¹

Pierre DELATTRE

On considère un ensemble d'individus en nombre quelconque, chaque sujet étant repéré par la lettre s indicée :

$$s_1, s_2, s_3 \text{ --- } s_u \text{ ----}$$

Chaque sujet est affecté d'un certain nombre de caractères :

$$A, B, C, D, \text{ ---- } N$$

et chacun de ces caractères peut prendre r états repérés par les entiers successifs :

$$1, 2, 3, \text{ ---- } r$$

On admet les hypothèses suivantes :

Hypothèse 1 : *Pour chaque caractère, la succession des états possibles (mutations) est celle de la suite repérée ci-dessus par les entiers successifs de 1 à r, sans omission et sans retour en arrière.*

Hypothèse 2 : *Les divers caractères considérés sont indépendants les uns des autres, du point de vue des mutations qu'ils subissent.*

1. Nous publions dans la rubrique «Archives» ce qui est en fait un manuscrit inédit de P. Delattre. Il s'agit d'une de ces nouvelles orientations suivies par cet auteur qui lui permettaient d'embrasser un champ toujours plus vaste de la Biologie Théorique, sa démarche inflexible et féconde le conduisant naturellement de l'axiomatique des systèmes de transformation aux interrogations les plus insistantes de notre temps, qu'il s'agisse de l'Evolution ou, dans ses ultimes publications, du problème du cancer. A ce titre il ne semble pas erroné de penser que le travail ainsi publié dans la Revue Internationale de Systémique en tant qu'article original puisse être simultanément consacré en tant qu'archive ! un mécanisme qui n'est d'ailleurs pas loin de certains processus de mémorisation. Signalons enfin qu'il convient de se reporter, après la lecture de ce travail, à l'annexe de l'article de Y. Bouligand concernant précisément cet inédit (E.B.-W.).

Par ailleurs, pour les considérations qui vont suivre, on adoptera la définition suivante :

Définition : On appellera «distance mutationnelle» entre deux sujets quelconques, le nombre de mutations permettant de passer de l'un à l'autre, en considérant tous les caractères présents dans les deux sujets considérés.

Exemple : Soient les deux sujets définis par :

$$\left. \begin{aligned} s_1 &= A_{i_1} B_{j_1} C_{k_1} D_{l_1} \text{-----} N_{m_1} \\ s_2 &= A_{i_2} B_{j_2} C_{k_2} D_{l_2} \text{-----} N_{m_2} \end{aligned} \right\} \quad (1)$$

avec $i_1 \leq i_2$; $j_1 \leq j_2$; $k_1 \leq k_2$; $l_1 \leq l_2$; ----- ; $m_1 \leq m_2$ (2)

La distance mutationnelle $d(s_1 \rightarrow s_2)$ entre s_1 et s_2 sera définie par :

$$\begin{aligned} d(s_1 \rightarrow s_2) &= (i_1 - i_2) + (j_1 - j_2) + (k_1 - k_2) + (l_1 - l_2) + \text{-----} + (m_1 - m_2) \\ &= (i_1 + j_1 + k_1 + l_1 + \text{-----} + m_1) - (i_2 + j_2 + k_2 + l_2 + \text{-----} + m_2) \\ &= \Sigma(\text{ind. } s_1) - \Sigma(\text{ind. } s_2) * \end{aligned} \quad (3)$$

(c'est-à-dire : somme des indices des caractères de s_1 , moins somme des indices des caractères de s_2).

On notera que la distance n'est définissable que si les relations (2) sont respectées, ce qui est lié au fait de l'irréversibilité des mutations.

A. Problèmes relatifs aux distances mutationnelles entre sujets

Dans la recherche des filiations on peut être conduit à se poser divers types de questions relatives aux distances entre sujets d'une population, ou entre ces sujets et leurs ancêtres possibles. On va examiner quelques-unes de ces questions, sans prétendre à l'exhaustivité.

* Pour une raison que j'ignore, P. Delattre introduit des distances négatives en raison des relations (2). Elles seraient normalement positives si l'on remplaçait les parenthèses par des barres verticales de valeur absolue. Le signe positif de ces distances est rétabli à partir de la formule (9). (Y. Bouligand).

Question 1 : Quel est le nombre d'ancêtres communs possibles (ACP) d'une population donnée de sujets.

Soit une population de sujets :

$$\left. \begin{aligned} s_1 &= A_{i_1} B_{j_1} C_{k_1} D_{l_1} \text{-----} N_{m_1} \\ s_2 &= A_{i_2} B_{j_2} C_{k_2} D_{l_2} \text{-----} N_{m_2} \\ s_3 &= A_{i_3} B_{j_3} C_{k_3} D_{l_3} \text{-----} N_{m_3} \\ &\text{-----} \\ s_u &= A_{i_u} B_{j_u} C_{k_u} D_{l_u} \text{-----} N_{m_u} \end{aligned} \right\} \quad (4)$$

et soit :

$$\left. \begin{aligned} \alpha_1 &= \min. i_1, i_2, i_3 \text{-----} i_u \\ \alpha_j &= \min. j_1, j_2, j_3 \text{-----} j_u \\ \alpha_k &= \min. k_1, k_2, k_3 \text{-----} k_u \\ \alpha_l &= \min. l_1, l_2, l_3 \text{-----} l_u \\ &\text{-----} \\ \alpha_m &= \min. m_1, m_2, m_3 \text{-----} m_u \end{aligned} \right\} \quad (5)$$

$$\text{Le sujet } S_a = A_{\alpha_1} B_{\alpha_j} C_{\alpha_k} D_{\alpha_l} \text{-----} N_{\alpha_m} \quad (6)$$

est un ancêtre commun possible de tous les sujets définis par (4) puisque tous ses caractères sont affectés d'indices égaux ou inférieurs à ceux qui affectent les mêmes caractères dans les sujets définis par (4).

Dans certains cas particuliers, il est possible que S_a se confonde avec l'un des sujets (4), si un tel sujet a tous ses indices respectant les relations (5). L'ensemble (4) comporte alors un sujet pouvant être ancêtre commun de tous les autres.

Mais il existe aussi, en général, beaucoup d'autres ACP des sujets (4), en plus de S_a . Ce sont en effet tous les sujets définis par :

$$S_a = A_{\alpha'_i} B_{\alpha'_j} C_{\alpha'_k} D_{\alpha'_l} \dots N_{\alpha'_m} \quad (7)$$

$$\text{avec : } \left. \begin{array}{l} 1 \leq \alpha'_i \leq \alpha_i \\ 1 \leq \alpha'_j \leq \alpha_j \\ 1 \leq \alpha'_k \leq \alpha_k \\ 1 \leq \alpha'_l \leq \alpha_l \\ \dots \\ 1 \leq \alpha'_m \leq \alpha_m \end{array} \right\} \quad (8)$$

En effet, tous ces sujets (7) ont tous leurs indices inférieurs aux indices correspondants des sujets (4).

Le dénombrement de ces ACP est immédiat. On notera d'abord que la seconde inégalité dans chacune des relations (8) étant non stricte (égalité possible), l'expression (7) inclut le sujet S_a défini par (6). On remarque alors que α'_i peut prendre α_i valeurs (de 1 à α_i) ; de même α'_j peut prendre α_j valeurs ; α'_k peut prendre α_k valeurs ; etc... Ainsi, à chaque combinaison des caractères BCD... N dans (7) correspondra α_i sujets S_a ; de même, à chaque combinaison des caractères ACD... N dans (7) correspondra α_j sujets S_a ; etc... Le nombre total d'ACP définis par (7) est donc égal au produit $\alpha_i \cdot \alpha_j \cdot \alpha_k \cdot \alpha_l \dots \alpha_m$. Dans le cas particulier où le sujet (6) est compris dans (7), il faut diminuer le nombre précédent d'une unité si l'on veut dénombrer seulement les ACP extérieurs à la population de sujets considérés. Donc :

Conclusions 1 : Les ancêtres communs possibles extérieurs à une population définie par les relations (4) sont au nombre total de $(\alpha_i \cdot \alpha_j \cdot \alpha_k \cdot \alpha_l \dots \alpha_m - 1)$ ou de $(\alpha_i \cdot \alpha_j \cdot \alpha_k \cdot \alpha_l \dots \alpha_m)$, selon que le sujet $A_{\alpha_i} B_{\alpha_j} C_{\alpha_k} D_{\alpha_l} \dots N_{\alpha_m}$ fait partie ou non de la population (4).

Question II : Quelle est la distance entre un ancêtre commun possible (ACP) d'une population et un sujet donné de cette population ? Y a-t-il des chemins de longueurs différentes menant de cet ancêtre commun au sujet considéré ?

Soit un sujet quelconque (par exemple s_u) de la population (4), et soit un des ACP de cette population, ancêtre défini par (7). Compte tenu de la définition de la distance donnée par (3), on a :

$$d(S_a, \rightarrow s_u) = (i_u + j_u + k_u + l_u + \dots + m_u) - (\alpha'_i + \alpha'_j + \alpha'_k + \alpha'_l + \dots + \alpha'_m) \quad (9)$$

Cette distance ne dépendant que des indices affectant les caractères des sujets considérés, elle est indépendante des chemins suivis pour passer de l'un à l'autre.

Conclusion II : La distance entre un sujet d'une population donnée et l'un quelconque des ancêtres communs de cette population est indépendante du chemin suivi pour passer de l'ancêtre au sujet considéré, c'est-à-dire indépendante de l'ordre dans lequel les mutations sont supposées intervenir.

Question III : Y a-t-il un ACP unique à égale distance de tous les sujets de la population considérée ?

La relation (9) permet de répondre immédiatement par la négative dans le cas général. En effet, pour un ACP donné, la seconde parenthèse de (9) est fixée. Pour obtenir la même valeur de la distance pour tous les sujets s_u ($u = 1, 2, 3, \dots$) de la population, il faut se trouver dans le cas particulier où la somme des indices des caractères est identique pour tous ces sujets.

Conclusion III : Un ACP quelconque d'une population donnée est en général situé à des distances variées de chacun des sujets de cette population. Pour que tous les sujets de la population soient à même distance de l'ACP considéré, il est nécessaire que la somme des indices des caractères soit la même pour tous les sujets. Dans ce cas, l'égalité de distance est obtenue quel que soit l'ACP considéré ; mais la valeur de cette distance commune dépend bien entendu de l'ACP considéré.

Question IV : Comment se situent les écarts de distance respectifs entre divers sujets de la population considérée et un de leurs ACP ?

Soient 2 sujets s_1 et s_2 de la population (4) et un de leurs ACP défini par (7). On a, d'après (9) :

$$d(S_a, \rightarrow s_1) = (i_1 + j_1 + k_1 + l_1 + \dots + m_1) - (\alpha'_i + \alpha'_j + \alpha'_k + \alpha'_l + \dots + \alpha'_m)$$

$$d(S_a, \rightarrow s_2) = (i_2 + j_2 + k_2 + l_2 + \dots + m_2) - (\alpha'_i + \alpha'_j + \alpha'_k + \alpha'_l + \dots + \alpha'_m)$$

Et l'écart $\Delta(s_1 \rightarrow s_2)$ est :

$$\Delta(s_1 \rightarrow s_2) = d(S_a, \rightarrow s_2) - d(S_a, \rightarrow s_1) = (i_2 + j_2 + k_2 + l_2 + \dots + m_2) - (i_1 + j_1 + k_1 + l_1 + \dots + m_1)$$

écart qui ne dépend que des indices des caractères sur les sujets considérés, et non de l'ACP considéré.

Conclusion IV : *Les écarts de distances respectifs entre divers sujets d'une population et un de leurs ACP ne dépendent que des différences des sommes d'indices des caractères sur les sujets considérés, et non de l'ACP considéré. Cet écart de distances pour un couple de sujets quelconque est donc le même pour tous les ACP de la population.*

Question V : *Quel est l'ACP le plus proche de tous les sujets d'une population donnée ? Peut-il exister plusieurs ACP présentant cette même proximité ?*

Comme on l'a vu à partir des relations (5) et (6), l'ACP le plus proche est celui qui a été dénommé S_a et qui est défini par (6). Si l'on ne pose aucune restriction particulière sur l'éventuelle appartenance de S_a à la population des sujets considérés, cet ACP est unique puisqu'il résulte des relations (5). Si au contraire on souhaite connaître l'ACP le plus proche, mais qui soit extérieur à la population considérée, et si S_a ne remplit pas cette condition, alors il faut considérer les ACP définis par (7) et pour lesquels l'un au moins des indices est inférieur d'une unité à l'indice correspondant dans l'expression (6). Ce qui revient à dire que l'on doit choisir l'une quelconque des conditions :

$$\alpha'_i = \alpha_i - 1 ; \text{ ou } \alpha'_j = \alpha_j - 1 ; \text{ ou } \alpha'_k = \alpha_k - 1 ; \text{ ou } \alpha'_1 = \alpha_1 - 1 ; \text{ --- ; ou } \alpha'_m = \alpha_m - 1 \quad (10)$$

il y a autant de solutions possibles que de caractères A, B, C, D, --- N considérés dans la définition des sujets de la population. Il y a donc N ancêtres communs possibles, qui présentent tous la même proximité minimale vis-à-vis des divers sujets de la population. Cela n'est toutefois vrai que si les relations (10) donnent toutes des valeurs non nulles des α' ; sinon ces α' nuls sont à éliminer et le nombre N est diminué d'autant d'unités qu'il existe d'indices $\alpha = 1$ dans l'expression (6). **Conclusion V :** *Si l'on admet que l'ACP le plus proche des sujets d'une population fasse éventuellement partie de cette population, cet ACP est unique. Si l'on refuse cette éventualité, il y a en général (N-b) ancêtres communs possibles présentant la même proximité minimale vis-à-vis des sujets de la population N représentant le nombre de caractères considérés, et b le nombre d'indices $\alpha_i, \alpha_j, \alpha_k, \alpha_1, \text{ --- } \alpha_m$ égaux à 1 dans l'expression (6).*

Exemples : On va illustrer à partir de populations arbitraires les questions qui viennent d'être évoquées.

1) Soit la population :

$$s_1 = A_4 B_4 C_3 D_2 ; s_2 = A_3 B_4 C_2 D_3 ; s_3 = A_3 B_3 C_2 D_4$$

$$s_4 = A_2 B_4 C_3 D_2 ; s_5 = A_3 B_2 C_2 D_1$$

On a ici (voir relations (5)) :

$$\alpha_i = \min. \{ 4, 3, 3, 2, 3 \} = 2$$

$$\alpha_j = \min. \{ 4, 4, 3, 4, 2 \} = 2$$

$$\alpha_k = \min. \{ 3, 2, 2, 3, 2 \} = 2$$

$$\alpha_1 = \min. \{ 2, 3, 4, 2, 1 \} = 1$$

Le plus proche ACP est donc (formule (6)) :

$$s_a = A_2 B_2 C_2 D_1$$

et cet ancêtre est extérieur à la population considérée

$$\{ s_1, s_2, s_3, s_4, s_5 \} .$$

Le nombre total d'ACP pour cette population est :

$$\alpha_i \cdot \alpha_j \cdot \alpha_k \cdot \alpha_1 = 2 \cdot 2 \cdot 2 \cdot 1 = 8$$

On vérifie facilement que ces 8 ACP sont les suivants :

$$A_2 B_2 C_2 D_1 ; A_2 B_1 C_2 D_1 ; A_1 B_2 C_2 D_1 ; A_1 B_1 C_2 D_1$$

$$A_2 B_2 C_1 D_1 ; A_2 B_1 C_1 D_1 ; A_1 B_2 C_1 D_1 ; A_1 B_1 C_1 D_1$$

La distance mutationnelle entre le s_2 et l'ACP le plus proche s_a est :

$$d(S_a \rightarrow s_2) = (3 + 4 + 2 + 3) - (2 + 2 + 2 + 1) = 12 - 7 = 5$$

Entre s_5 et S_a , on a :

$$d(S_a \rightarrow s_5) = (3 + 2 + 2 + 1) - (2 + 2 + 2 + 1) = 8 - 7 = 1$$

L'écart des distances respectives entre S_a et s_2 d'une part, et S_a et s_5 d'autre part, est :

$$d(S_a \rightarrow s_2) - d(S_a \rightarrow s_5) = 5 - 1 = 4$$

Cet écart est égal à la distance entre s_2 et s_5 , indépendamment de l'ACP choisi :

$$d(S_5 \rightarrow s_2) = (3 + 4 + 2 + 3) - (3 + 2 + 2 + 1) = 12 - 8 = 4$$

2) Soit la population précédente à laquelle on ajoute l'élément $s_6 = A_2 B_2 C_2 D_1$. Il est facile de constater que l'ACP le plus proche est toujours le même

$$S_a = A_2 B_2 C_2 D_1$$

et qu'il se confond donc avec s_6 . Si l'on cherche les ACP les plus proches de la population, mais extérieurs à celle-ci, ce sont (expression (7) dans laquelle on respecte successivement les relations (10) :

$$\alpha'_i = \alpha_i - 1 ; \alpha'_j = \alpha_j - 1 ; \text{etc...} :$$

$$S_{a'_1} = A_1 B_2 C_2 D_1 ; S_{a'_2} = A_2 B_1 C_2 D_1 ; S_{a'_3} = A_2 B_2 C_1 D_1$$

Ils sont bien au nombre de $N-b$ puisqu'ici $N = 4$ (caractères) et $b = 1$ (indice de D dans S_a).

La distance d'un sujet quelconque de la population par rapport à ces divers ACP est constante. Par exemple, pour $s_3 = A_3 B_3 C_2 D_4$:

$$d(S_{a'_1} \rightarrow s_3) = (3 + 3 + 2 + 4) - (1 + 2 + 2 + 1) = 12 - 6 = 6$$

$$d(S_{a'_2} \rightarrow s_3) = (3 + 3 + 2 + 4) - (2 + 1 + 2 + 1) = 12 - 6 = 6$$

$$d(S_{a'_3} \rightarrow s_3) = (3 + 3 + 2 + 4) - (2 + 2 + 1 + 1) = 12 - 6 = 6$$

B. Etablissement d'un diagramme de filiations

Le principe même d'irréversibilité des mutations implique que tous les sujets situés sur une même ligne de filiation présentent, dans le sens ascendance-descendance, des indices toujours croissants (ou éventuellement stationnaires pour certains d'entre eux) d'un sujet au suivant, et cela pour tous les caractères considérés. Lorsque cette condition n'est pas respectée elle donne lieu à des bifurcations des lignes de filiations.

Par exemple, les quatre sujets :

$$s_1 = A_{i_1} B_{j_1} C_{k_1} D_{l_1} ; s_2 = A_{i_2} B_{j_2} C_{k_2} D_{l_2} ; s_3 = A_{i_3} B_{j_3} C_{k_3} D_{l_3} ;$$

$$s_4 = A_{i_4} B_{j_4} C_{k_4} D_{l_4}$$

ne pourront être placés sur une même ligne de descendance que si les relations suivantes sont simultanément respectées :

$$i_1 \leq i_2 \leq i_3 \leq i_4 ; j_1 \leq j_2 \leq j_3 \leq j_4 ; k_1 \leq k_2 \leq k_3 \leq k_4 ;$$

$$l_1 \leq l_2 \leq l_3 \leq l_4$$

Si l'une au moins des inégalités n'est pas respectée, il devra y avoir bifurcation, et par conséquent apparition de plusieurs branches de filiations. Par exemple, si l'on a :

$$i_1 \leq i_2 \leq i_3 \leq i_4 ; j_1 \leq j_2 \leq j_3 \leq j_4 ; k_1 \leq k_2 \leq k_3 > k_4 ;$$

$$l_1 \leq l_2 \leq l_3 \leq l_4$$

les sujets s_1, s_2, s_3 peuvent être placés dans une même ligne de filiations ; mais s_4 doit être situé dans une ligne s_1, s_2, s_4 qui bifurque en s_2 , ou en s_1 , selon les valeurs respectives de k_1 et k_4 . Les conditions seront vues plus aisément dans l'exemple qui sera traité plus loin.

Lorsque le nombre de sujets de la population considérée est faible, une mise en ordre «à vue» est souvent possible sans grandes difficultés. Mais on peut aussi utiliser une méthode graphique que nous allons illustrer sur un exemple.

Exemple : Soit la population déjà considérée précédemment :

$$s_1 = A_4 B_4 C_3 D_2 ; s_2 = A_3 B_4 C_2 D_3 ; s_3 = A_3 B_3 C_2 D_4 ;$$

$$s_4 = A_2 B_4 C_3 D_2 ; s_5 = A_3 B_2 C_2 D_1$$

Portons sur des axes verticaux les valeurs des indices $\alpha_i, \alpha_j, \alpha_k, \alpha_l$ pour les divers sujets, en commençant par l'ACP le plus proche qui est ici :

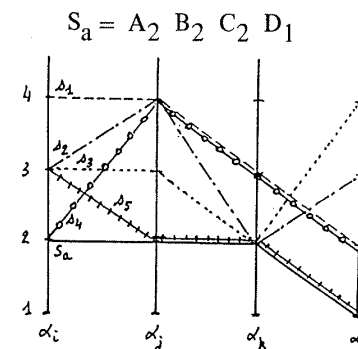


Figure 1

Pour chaque sujet joignons les points correspondant à ses divers indices, par une ligne individualisée selon les symboles suivants :

—————	: S_a
-----	: s_1
-. - . - . - . -	: s_2
.....	: s_3
-o-o-o-o-	: s_4
-:-:-:-:-	: s_5

Le principe énoncé plus haut quant à la relation d'ordre des indices pour tous les sujets d'une même ligne de filiation se traduit par le fait que, dans le schéma précédent, seules les lignes brisées représentatives ne se coupant pas correspondent à des sujets pouvant être placés dans une même ligne de filiation.

On constate ainsi facilement que :

— S_a , s_4 , s_1 peuvent être placés dans une même ligne de filiation (toutes les lignes représentatives des autres sujets s_2 , s_3 , s_5 recoupent l'une ou l'autre des lignes représentatives de s_4 et s_1 ; s_2 recoupe s_4 et s_1 ; s_3 recoupe s_4 et s_1 ; s_5 recoupe s_4).

— S_a , s_5 et s_3 peuvent également être placés dans une même ligne de filiation

— S_a , s_5 et s_2 peuvent également être placés dans une même ligne de filiation

— S_a , s_5 et s_1 peuvent également être placés dans une même ligne de filiation.

On aboutit ainsi au diagramme suivant :

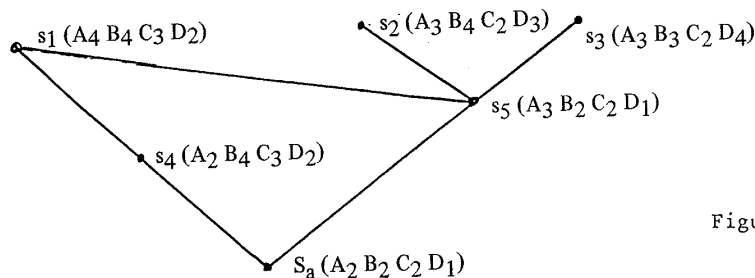


Figure 2

Figure 2

Ce diagramme est unique, mais on doit noter qu'il est conditionné par les seuls sujets *connus*, c'est-à-dire observés effectivement. Supposons par exemple que l'on découvre ensuite un sujet $s_6 = A_3 B_3 C_2 D_2$. Il devra être situé à la fois dans la ligne de filiation $S_a s_5 s_6 s_3$ et dans la ligne de filiation $S_a s_5 s_6 s_2$, et le diagramme correspondant sera comme ci-dessous :

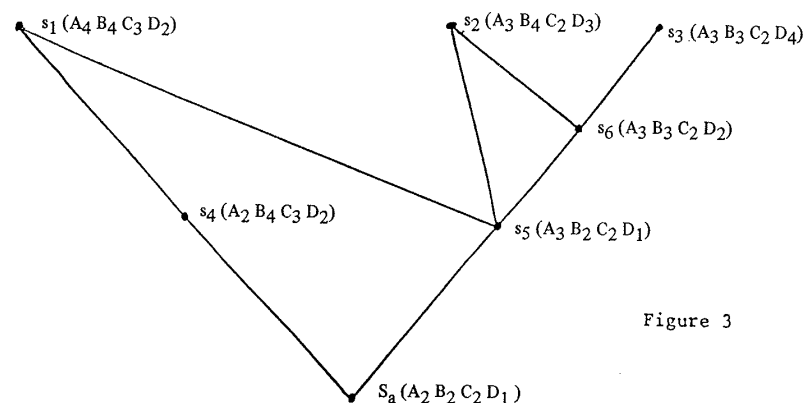


Figure 3

Figure 3

La méthode graphique proposée ci-dessus peut être facilement informatisée si le nombre de sujets de la population considérée devient trop grand pour une exploitation «manuelle» aisée.

C. Remarques

1 — Il est important de se souvenir que toutes les conclusions qui précèdent ne sont valables que dans le cadre des deux hypothèses admises initialement (ordre donné des mutations possibles sur un caractère donné, et indépendance des caractères du point de vue des mutations qu'ils subissent).

2 — Il est important de noter que les diagrammes construits selon

la méthode précédente ne sont pas des *arbres* au sens mathématique du terme. Ils peuvent contenir des parties fermées correspondant à l'existence de plusieurs chemins susceptibles de conduire d'un sujet à un autre (par exemple, partie $S_1 s_4 s_1 s_5$ de la figure 2, correspondant à deux chemins possibles de S_a à s_1 : $S_a s_4 s_1$ et $S_a s_5 s_1$; de même, pour les parties $s_5 s_6 s_2$ et $S_a s_4 s_1 s_5$ de la figure 3). Nous verrons un peu plus loin comment se présente ce problème des chemins multiples.

3 — Pour un ACP choisi, la distance mutationnelle entre cet ancêtre et un sujet donné de la population est parfaitement déterminée, puisqu'elle ne dépend que de la somme des indices du sujet et de l'ACP. Il n'y a donc pas de chemin minimal conduisant de l'un à l'autre (cf. Conclusion II). Pour introduire une notion de «chemin minimal», il faudrait adopter une définition de la distance mutationnelle différente de celle qui a été donnée ici, ou encore admettre que certains états des caractères peuvent être omis dans la succession des mutations (modification de l'hypothèse 1).

D. Inventaire des chemins possibles entre couples de sujets et problème des critères de choix

Nous avons constaté précédemment la possibilité de chemins multiples entre les deux sujets d'un couple quelconque. Cette possibilité peut déjà apparaître dans le diagramme que nous appellerons *initial **, construit à partir des seuls sujets effectivement observés. Mais on peut aussi se demander quels sont les chemins théoriquement possibles entre deux sujets quelconques du diagramme initial, même lorsque n'apparaît qu'un seul chemin dans cette première construction (par exemple entre s_5 et s_2 de la figure 2, ou entre S_a et s_4 dans les figures 2 et 3).

Pour ne pas traîner tout au long du calcul des formules trop lourdes, nous traiterons le cas de sujets présentant seulement trois caractères. La généralisation à un nombre quelconque de caractères se fera ensuite sans difficulté.

Soient donc deux sujets quelconques :

$$\begin{array}{ll} s_2 = A_{\alpha_2} B_{\beta_2} C_{\gamma_2} & \alpha_2 \geq \alpha_1 \\ & \text{avec } \beta_2 \geq \beta_1 \\ s_1 = A_{\alpha_1} B_{\beta_1} C_{\gamma_1} & \gamma_2 \geq \gamma_1 \end{array}$$

* La dénomination de diagramme réticulé, proposée par S. Tillier, conviendrait aussi tout à fait.

entre lesquels on veut évaluer le nombre de chemins théoriquement possibles, compte tenu des postulats de départ (indépendance des caractères, et croissance des indices sans omission ni retour en arrière).

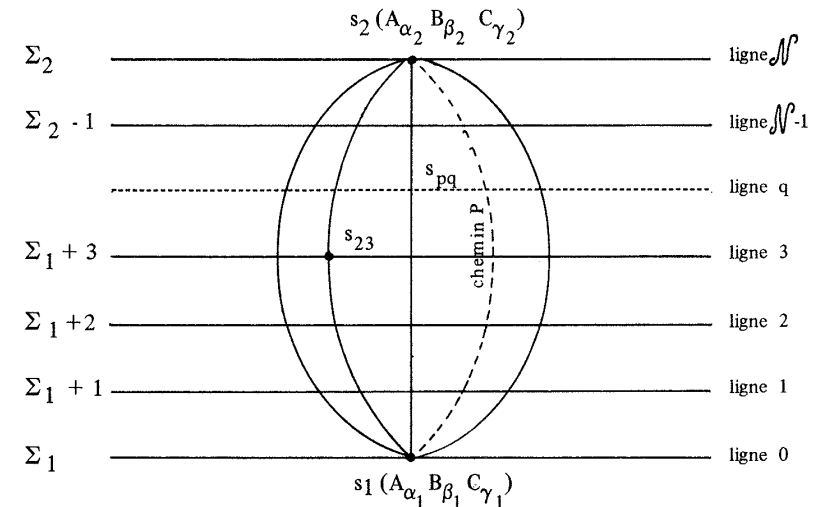


Figure 4

Sur la figure 4, le chemin pointillé symbolise l'ensemble des chemins non effectivement représentés.

Les légendes horizontales repèrent la valeur de la somme des indices pour chacun des sujets situés à l'intersection de la ligne correspondante et des divers chemins possibles.

Appelons Σ cette somme d'indices ; à s_1 correspond :

$$\left. \begin{array}{l} \Sigma_1 = \alpha_1 + \beta_1 + \gamma_1 \\ \text{et à } s_2 : \Sigma_2 = \alpha_2 + \beta_2 + \gamma_2 \end{array} \right\} \quad (11)$$

A partir de Σ_1 , chaque ligne horizontale correspond à la croissance d'une unité de Σ pour chaque sujet situé sur cette ligne, c'est-à-dire à la croissance d'une unité de l'un de ses indices. Par exemple,

pour le sujet s_{23} situé à l'intersection du 2e chemin et de la 3e ligne, la somme des indices sera $\Sigma_{23} = \Sigma_1 + 3$. (et il en sera de même pour tous les sujets situés sur la même ligne horizontale).

Le nombre d'intervalles définis par les lignes horizontales, entre s_1 et s_2 est :

$$\mathcal{J}^p = \Sigma_2 - \Sigma_1 = (\alpha_2 + \beta_2 + \gamma_2) - (\alpha_1 + \beta_1 + \gamma_1) \quad (12)$$

Les lignes horizontales sont numérotées comme indiqué sur la figure 4. La dernière ligne, qui passe par s_2 , est donc la $\mathcal{J}^p$ e ligne.

Soit un sujet s_{pq} situé à l'intersection du chemin p et de la ligne q ; écrivons ses indices :

$$\begin{aligned} \alpha_{pq} &= \alpha_{p(q-1)} + x_{\alpha q}^{(p)} ; \beta_{pq} = \beta_{p(q-1)} + x_{\beta q}^{(p)} ; \\ \gamma_{pq} &= \gamma_{p(q-1)} + x_{\gamma q}^{(p)} \end{aligned} \quad (13)$$

Les variables x sont des variables booléennes, qui ne peuvent prendre que les valeurs 0 ou 1. En effet, à chaque progression d'une ligne à la suivante, un seul des coefficients α, β, γ est augmenté d'une unité, ce qui signifie que parmi les 3 variables x des relations (13) une seule prendra la valeur 1, et les deux autres la valeur zéro. La seule condition à remplir, dans la progression le long d'un chemin, c'est que chacun des coefficients atteigne finalement les valeurs $\alpha_2, \beta_2, \gamma_2$ caractérisant le sujet s_2 .

Si l'on écrit cette condition pour chacun des coefficients, et pour le chemin p , on obtient le système d'équations :

$$\begin{aligned} x_{\alpha 1}^{(p)} + x_{\alpha 2}^{(p)} + x_{\alpha 3}^{(p)} + \dots + x_{\alpha \mathcal{J}^p}^{(p)} &= \alpha_2 - \alpha_1 \\ x_{\beta 1}^{(p)} + x_{\beta 2}^{(p)} + x_{\beta 3}^{(p)} + \dots + x_{\beta \mathcal{J}^p}^{(p)} &= \beta_2 - \beta_1 \\ x_{\gamma 1}^{(p)} + x_{\gamma 2}^{(p)} + x_{\gamma 3}^{(p)} + \dots + x_{\gamma \mathcal{J}^p}^{(p)} &= \gamma_2 - \gamma_1 \end{aligned} \quad (14)$$

Et il y aura autant de systèmes (14) qu'il y aura de chemins p possibles ; c'est précisément ce nombre de chemins qu'il faut évaluer.

On remarquera que la somme des termes situés sur une même verticale, dans les premiers membres des relations (14), doit toujours être égale à 1 puisqu'elle correspond à l'accroissement total de la somme des termes α, β, γ lorsque le sujet passe, sur le chemin p , de la ligne $(q-1)$ à la ligne q .

La résolution du système (14) revient donc à considérer un ta-

bleau de N lignes (N = nombre de caractères, ici $N = 3$) et de $\mathcal{J}^p$ colonnes (cf. expression (12) pour $\mathcal{J}^p$), et à y placer des 0 et des 1 de telle manière que les sommes des termes des lignes soient respectivement égales à $(\alpha_2 - \alpha_1), (\beta_2 - \beta_1), (\gamma_2 - \gamma_1)$, et que la somme des termes dans chaque colonne soit égale à 1. La figure 5 en donne un exemple pour $N = 3 ; \alpha_2 - \alpha_1 = 2 ; \beta_2 - \beta_1 = 3 ; \gamma_2 - \gamma_1 = 1$ (c'est-à-dire $\mathcal{J}^p$

$$= \Sigma_2 - \Sigma_1 = 2 + 3 + 1 = 6 \text{ colonnes}).$$

$$\begin{array}{lcl} \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} & \rightarrow & \begin{array}{l} \text{Somme} = \alpha_2 - \alpha_1 = 2 \\ \text{Somme} = \beta_2 - \beta_1 = 3 \\ \text{Somme} = \gamma_2 - \gamma_1 = 1 \end{array} \\ \text{-----} & & \\ 1 & 1 & 1 & 1 & 1 & 1 & \leftarrow & \text{Somme des colonnes} \end{array} \quad (15)$$

Figure 5

Le nombre d'arrangements possibles des 0 et des 1 respectant les conditions imposées, dans le tableau considéré, fournit le nombre de chemins distincts possibles entre s_1 et s_2 . L'évaluation de ce nombre se fait facilement à partir de la formule des permutations avec répétition :

Pour la 1ère ligne : il faut déterminer le nombre de façons de placer les termes 1 et 0 dans les $\mathcal{J}^p$ positions possibles, en tenant compte du fait que parmi les $\mathcal{J}^p$ termes de la ligne il y a :

$$\begin{aligned} &(\alpha_2 - \alpha_1) \text{ terme } 1 \\ \text{et } \mathcal{J}^p &- (\alpha_2 - \alpha_1) \text{ terme } 0 \end{aligned}$$

La formule des permutations avec répétitions donne ce nombre de façons :

$$R_1 = \frac{\mathcal{J}^p !}{(\alpha_2 - \alpha_1) ! [\mathcal{J}^p - (\alpha_2 - \alpha_1)] !} \quad (16)$$

Pour la 2e ligne : à cause de la condition sur la somme des termes dans chaque colonne, on ne peut placer de termes 1 en dessous des emplacements déjà occupés par des 1 dans la 1ère ligne. Il reste donc seulement $[\mathcal{J}^p - (\alpha_2 - \alpha_1)]$ positions possibles, et parmi celles-ci il faut placer :

$(\beta_2 - \beta_1)$ termes 1

et $[\mathcal{N} - (\alpha_2 - \alpha_1) - (\beta_2 - \beta_1)]$ termes 0

On obtient ainsi pour le nombre de façons de placer ces termes :

$$R_2 = \frac{[\mathcal{N} - (\alpha_2 - \alpha_1)]!}{(\beta_2 - \beta_1)! [\mathcal{N} - (\alpha_2 - \alpha_1) - (\beta_2 - \beta_1)]!} \quad (17)$$

Pour la troisième ligne : en raisonnant comme précédemment, on obtient :

$$R_3 = \frac{[\mathcal{N} - (\alpha_2 - \alpha_1) - (\beta_2 - \beta_1)]}{(\gamma_2 - \gamma_1)! [\mathcal{N} - (\alpha_2 - \alpha_1) - (\beta_2 - \beta_1) - (\gamma_2 - \gamma_1)]!} \quad (18)$$

Finalement, puisqu'à chaque organisation d'une ligne correspondent toutes les organisations possibles de la ligne suivante, le nombre total de manières d'organiser le tableau en respectant les conditions imposées, c'est-à-dire le nombre de chemins possibles de s_1 à s_2 , est égal à :

$$n = R_1 \cdot R_2 \cdot R_3 \quad (19)$$

A partir des expressions (16) (17) (18) et en remarquant que (formule 12)) :

$$[\mathcal{N} - (\alpha_2 - \alpha_1) - (\beta_2 - \beta_1) - (\gamma_2 - \gamma_1)]! = 0! = 1 \quad (20)$$

on voit immédiatement que n s'écrit explicitement :

$$n = \frac{\mathcal{N}!}{(\alpha_2 - \alpha_1)! (\beta_2 - \beta_1)! (\gamma_2 - \gamma_1)!} = \frac{[(\alpha_2 - \alpha_1) + (\beta_2 - \beta_1) + (\gamma_2 - \gamma_1)]!}{(\alpha_2 - \alpha_1)! (\beta_2 - \beta_1)! (\gamma_2 - \gamma_1)!} \quad (21)$$

Pour un nombre quelconque de caractères, d'indices respectifs $\alpha, \beta, \gamma, \dots, \delta$, on aura :

$$n = \frac{[(\alpha_2 - \alpha_1) + (\beta_2 - \beta_1) + (\gamma_2 - \gamma_1) + \dots + (\delta_2 - \delta_1)]!}{(\alpha_2 - \alpha_1)! (\beta_2 - \beta_1)! (\gamma_2 - \gamma_1)! \dots (\delta_2 - \delta_1)!} \quad (21 \text{ bis})$$

On remarque que le nombre de chemins dépend uniquement des écarts entre les indices correspondants de s_1 et s_2 , et non des valeurs absolues de ces indices.

Exemples numériques :

1 - Soient $s_1 = A_1 B_1$ $s_2 = A_3 B_3$ ($N=2$)

$$\mathcal{N} = \Sigma_2 - \Sigma_1 = 4; \quad \alpha_2 - \alpha_1 = 2; \quad \beta_2 - \beta_1 = 2$$

$$\text{on trouve : } n = \frac{[2+2]!}{2! 2!} = \frac{4!}{2! 2!} = \frac{24}{4} = 6$$

2 - Soient $s_1 = A_2 B_3 C_1$ $s_2 = A_4 B_6 C_2$ ($N=3$)

$$\mathcal{N} = \Sigma_2 - \Sigma_1 = 6; \quad \alpha_2 - \alpha_1 = 2; \quad \beta_2 - \beta_1 = 3; \quad \gamma_2 - \gamma_1 = 1$$

$$\text{on trouve : } n = \frac{6!}{2! 3! 1!} = \frac{1.2.3.4.5.6}{1.2.1.2.3.1} = \frac{120}{2} = 60$$

Le nombre de sujets différents entre $s_1 = A_{\alpha_1} B_{\beta_1} C_{\gamma_1}$ et $s_2 = A_{\alpha_2} B_{\beta_2} C_{\gamma_2}$ y compris ces deux sujets extrêmes s_1 et s_2 est égal à :

$$r = (\alpha_2 - \alpha_1 + 1) \cdot (\beta_2 - \beta_1 + 1) \cdot (\gamma_2 - \gamma_1 + 1) \quad (22)$$

(on remarque que si s_1 est l'ancêtre de base $A_1 B_1 C_1$ c'est-à-dire si $\alpha_1 = \beta_1 = \gamma_1 = 1$, la formule (22) redonne bien le nombre, ici $\alpha_2 \cdot \beta_2 \cdot \gamma_2$, des ancêtres communs à s_2 , y compris s_2 , ainsi qu'on l'a vu dans la conclusion I de la partie A).

Le nombre de sujets intermédiaires distincts entre s_1 et s_2 , non compris ces deux sujets extrêmes, est donc égal à :

$$r' = (\alpha_2 - \alpha_1 + 1) \cdot (\beta_2 - \beta_1 + 1) \cdot (\gamma_2 - \gamma_1 + 1) - 2 \quad (23)$$

Or le nombre r'' de sujets intermédiaires représentés sur l'ensemble des chemins possibles entre s_1 et s_2 (fig. 4) est égal au produit du nombre de chemins (n) par le nombre de lignes horizontales intermédiaires entre s_1 et s_2 , c'est-à-dire $(\mathcal{N} - 1)$:

$$r'' = n \cdot (\mathcal{N} - 1) \quad (24)$$

Le nombre r'' est en général beaucoup plus grand que r' , ce qui signifie que de nombreux sujets intermédiaires seront représentés à de nombreux exemplaires sur les chemins possibles entre s_1 et s_2 .

Exemples :

$$1 - \text{ Soient } s_1 = A_1 B_1 \quad s_2 = A_2 B_2$$

$$\mathcal{N} = \Sigma_2 - \Sigma_1 = 4 ; \alpha_2 - \alpha_1 = 2 ; \beta_2 - \beta_1 = 2$$

On a vu que $n = 6$

Les expressions (23) et (24) donnent respectivement :

$$r' = 3.3 - 2 = 7$$

$$r'' = 6.3 = 18$$

$$2 - \text{ Soient } s_1 = A_2 B_3 C_1 \quad s_2 = A_4 B_6 C_2$$

$$\mathcal{N} = \Sigma_2 - \Sigma_1 = 6 ; \alpha_2 - \alpha_1 = 2 ; \beta_2 - \beta_1 = 3 ; \gamma_2 - \gamma_1 = 1$$

On a vu que $n = 60$

Les expressions (23) et (24) donnent respectivement :

$$r' = 3.4.2 = 24$$

$$r'' = 60.5 = 300$$

Il s'ensuit des considérations précédentes qu'une question intéressante consiste à se demander à combien d'exemplaires est représenté un sujet intermédiaire donné s_u , dans l'ensemble des chemins possibles entre s_1 et s_2 . Pour évaluer ce nombre, on peut raisonner comme ceci : il y a d'abord autant d'exemplaires de s_u qu'il y a de chemins distincts possibles entre s_u et s_2 ; mais chacun de ces exemplaires est l'aboutissement d'un nombre de chemins égal à celui de tous les chemins possibles entre s_1 et s_u (ce point se voit beaucoup mieux si l'on trace les chemins sous forme d'arbres comme dans la figure 6, plutôt que sous la forme des chemins explicitement séparés comme dans la figure 4).

Par exemple, dans le cas de la figure 6, si $s_1 = A_1 B_1$; $s_2 = A_3 B_3$; $s_u = A_3 B_3$, il existe un seul chemin de $A_3 B_2$ à $A_3 B_3$, mais il en existe trois qui, partant de s_1 aboutissent à s_u . De même, si $s_u = A_2 B_2$, il existe deux chemins de s_u à s_2 ($A_2 B_2 \rightarrow A_2 B_3 \rightarrow A_3 B_3$ et $A_2 B_2 \rightarrow A_3 B_2 \rightarrow A_3 B_3$), et deux chemins de s_1 à s_u ($A_1 B_1 \rightarrow A_1 B_2 \rightarrow A_2 B_2$ et $A_1 B_1 \rightarrow A_2 B_1 \rightarrow A_2 B_2$).

D'une manière générale, s'il existe n_{u2} chemins de s_u à s_2 , et n_{u1} chemins de s_1 à s_u , le nombre n_u d'exemplaires de s_u dans l'ensemble des chemins possibles entre s_1 et s_2 est donné par :

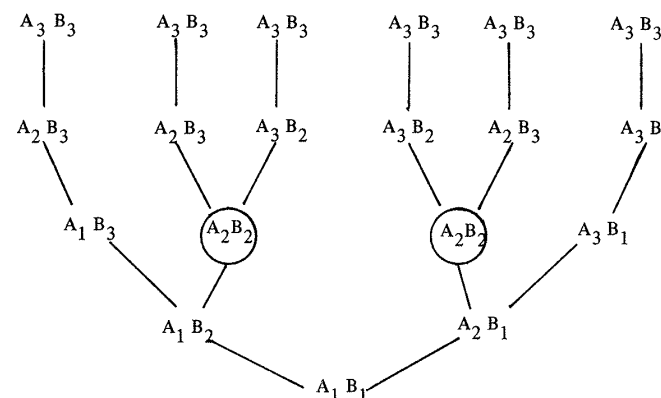


Figure 6

$$n_u = n_{u1} \cdot n_{u2} \quad (25)$$

Or les valeurs de n_{u1} et n_{u2} sont données par la relation (21) :

$$n_{u1} = \frac{[(\alpha_u - \alpha_1) + (\beta_u - \beta_1) + (\gamma_u - \gamma_1)]!}{(\alpha_u - \alpha_1)! (\beta_u - \beta_1)! (\gamma_u - \gamma_1)!}$$

$$n_{u2} = \frac{[(\alpha_2 - \alpha_u) + (\beta_2 - \beta_u) + (\gamma_2 - \gamma_u)]!}{(\alpha_2 - \alpha_u)! (\beta_2 - \beta_u)! (\gamma_2 - \gamma_u)!}$$

et finalement

$$n_u = \frac{[(\alpha_u - \alpha_1) + (\beta_u - \beta_1) + (\gamma_u - \gamma_1)]!}{(\alpha_u - \alpha_1)! (\beta_u - \beta_1)! (\gamma_u - \gamma_1)!} \cdot \frac{[(\alpha_2 - \alpha_u) + (\beta_2 - \beta_u) + (\gamma_2 - \gamma_u)]!}{(\alpha_2 - \alpha_u)! (\beta_2 - \beta_u)! (\gamma_2 - \gamma_u)!} \quad (26)$$

Pour un nombre quelconque de caractères d'indices respectifs $\alpha, \beta, \gamma \dots \delta$, on aura :

$$n_u = \frac{[(\alpha_u - \alpha_1) + (\beta_u - \beta_1) + (\gamma_u - \gamma_1) + \dots + (\delta_u - \delta_1)]!}{(\alpha_u - \alpha_1)! (\beta_u - \beta_1)! (\gamma_u - \gamma_1)! \dots (\delta_u - \delta_1)!}$$

$$\frac{[(\alpha_2 - \alpha_u) + (\beta_2 - \beta_u) + (\gamma_2 - \gamma_u) + \dots + (\delta_2 - \delta_u)]!}{(\alpha_2 - \alpha_u)! (\beta_2 - \beta_u)! (\gamma_2 - \gamma_u)! \dots (\delta_2 - \delta_u)!} \quad (26 \text{ bis})$$

Exemples :

1 — Soient $s_1 = A_1B_1$; $s_2 = A_3B_3$; $s_u = A_1B_3$

$$\alpha_u - \alpha_1 = 0 ; \beta_u - \beta_1 = 2$$

$$\alpha_2 - \alpha_u = 2 ; \beta_2 - \beta_u = 0$$

$$n_u = \frac{2!}{0!} \cdot \frac{2!}{2!} = 1 \text{ (} A_1B_3 \text{ est représenté en un seul exemplaire)}$$

Si $s_u = A_2B_2$, avec les mêmes s_1 et s_2 , on a :

$$\alpha_u - \alpha_1 = 1 ; \beta_u - \beta_1 = 1$$

$$\alpha_2 - \alpha_u = 1 ; \beta_2 - \beta_u = 1$$

$$n_u = \frac{[1+1]!}{1!1!} \cdot \frac{[1+1]!}{1!1!} = 2!2! = 4$$

(A_2B_2 est représenté à 4 exemplaires sur l'ensemble des chemins distincts).

2 — Soient

$$s_1 = A_1B_1 ; s_2 = A_5B_5 ; s_u = A_4B_4$$

$$\alpha_u - \alpha_1 = 3 ; \beta_u - \beta_1 = 3$$

$$\alpha_2 - \alpha_u = 1 ; \beta_2 - \beta_u = 1$$

$$n_u = \frac{6!}{3!3!} \cdot \frac{2!}{1!1!} = 40 \text{ (} A_4B_4 \text{ est présent en 40 exemplaires)}$$

3 — Soient

$$s_1 = A_2B_2C_1 ; s_2 = A_4B_6C_2 ; s_u = A_3B_4C_2$$

$$\alpha_u - \alpha_1 = 1 ; \beta_u - \beta_1 = 2 ; \gamma_2 - \gamma_u = 1$$

$$\alpha_2 - \alpha_u = 1 ; \beta_2 - \beta_u = 2 ; \gamma_2 - \gamma_u = 0$$

$$n_u = \frac{[1+2+1]!}{1!2!1!} \cdot \frac{[1+2+0]!}{1!2!0!} = \frac{4!}{2!2!} = 36$$

Dans le prolongement de ce qui précède, on peut encore se poser la question suivante : parmi tous les sujets s_u intermédiaires possibles entre s_1 et s_2 , quel est celui, ou ceux, qui sont représentés par le plus grand nombre d'exemplaires ?

Si, à titre d'exemple, on considère des sujets à trois caractères, le nombre d'exemplaires du sujet s_u est donné par l'expression (26) ; il s'agit donc ici, s_1 et s_2 étant donnés, de déterminer s_u (c'est-à-dire $\alpha_u, \beta_u, \gamma_u$) de telle manière que n_u soit maximum. Il faut donc chercher les conditions qui maximisent le numérateur de (26) et qui minimisent son dénominateur.

On remarquera d'abord que les deux factorielles du numérateur de (26) peuvent aussi s'écrire :

$$F_1! = (\Sigma_u - \Sigma_1)!$$

$$F_2! = (\Sigma_2 - \Sigma_u)!$$

Or la somme :

$$F_1 + F_2 = \Sigma_2 - \Sigma_1$$

est fixée par les valeurs des indices des deux sujets s_1 et s_2 . Lorsque s_u change, si F_1 est diminué d'une certaine quantité, F_2 est augmenté de la même quantité. Il est facile de voir que le produit $F_1! F_2!$ sera maximal lorsque les termes F_1 et F_2 seront aussi différents que possible l'un de l'autre, c'est-à-dire lorsque leur différence $F_2 - F_1$ sera maximale. En effet supposons que :

$$F_1 = F_2 = w$$

$$\text{alors : } F_1! = 1.2.3. \dots w ; F_2! = 1.2.3. \dots w$$

$$F_1! F_2! = (1.2.3. \dots w)^2$$

Si F_1 diminue de x unités, F_2 augmente de x unités, et l'on a :

$$F_1' = 1.2.3. \dots (w-x)$$

$$F_2' = 1.2.3. \dots (w-x)(w-x+1) \dots w.(w+1) \dots (w+x)$$

$$F_1! F_2! = [1.2.3. \dots (w-x)]^2 \cdot (w-x+1) \dots w \cdot (w+1) \dots (w+x)$$

alors que $F_1! F_2!$ peut aussi s'écrire :

$$F_1! F_2! = [1.2.3. \dots (w-x)]^2 \cdot [(w-x+1) \dots w]^2$$

et puisque $(w-x+1) \dots w \cdot (w+1) \dots (w+x) > [(w-x+1) \dots w]^2$ quel que soit x , on a :

$$F_1! F_2! > F_1! F_2!$$

Autrement dit, le produit des factorielles du numérateur de (26) est d'autant plus grand que les F_1 et F_2 sont plus différents.

Si l'on considère maintenant les termes du dénominateur de (26), on constate que l'on peut également les regrouper en produits de deux factorielles ; par exemple :

$$\left. \begin{array}{ll} (\alpha_u - \alpha_1)! & (\alpha_2 - \alpha_u)! \\ (\beta_u - \beta_1)! & (\beta_2 - \beta_u)! \\ (\gamma_u - \gamma_1)! & (\gamma_2 - \gamma_u)! \end{array} \right\} \text{ ou toute autre combinaison des } \\ \text{termes pris 2 à 2}$$

Pour chacun de ces produits de deux factorielles, on peut refaire le même raisonnement que précédemment. Mais il s'agit maintenant de minimiser chacun de ces produits, de manière à minimiser le dénominateur de (26), afin de maximiser n_u . Il faut donc maintenant rendre les termes entrant dans chacun des produits de factorielles aussi voisins que possibles, c'est-à-dire, par exemple, tendre vers les relations :

$$\left. \begin{array}{l} \alpha_u - \alpha_1 = \alpha_2 - \alpha_u \\ \beta_u - \beta_1 = \beta_2 - \beta_u \\ \gamma_u - \gamma_1 = \gamma_2 - \gamma_u \end{array} \right\} \text{ ou tout autre jeu d'égalités entre } \\ \text{les expressions des premiers et } \\ \text{des seconds membres}$$

Mais de telles relations ont aussi pour effet de tendre vers l'égalité des termes F_1 et F_2 du numérateur de (26), c'est-à-dire de minimiser celui-ci.

Les conditions qui maximisent le numérateur de (26) et celles qui minimisent son dénominateur ne sont donc pas strictement compatibles, et le maximum de n_u ne peut être trouvé que dans un compromis entre les deux conditions.

Je n'ai pas démontré l'optimalité de ce compromis, mais on peut faire à ce sujet la conjecture suivante :

Les sujets les plus représentés sont parmi ceux qui se trouvent à la plus petite distance mutationnelle d'un des extrêmes (s_1 ou s_2) et dont les écarts d'indices avec chacun de ces extrêmes, pour les divers caractères, sont aussi voisins que possible.

Autrement dit, étant donnée une valeur de $\Sigma_2 - \Sigma_1 = F_1 + F_2$, les sujets les plus représentés sont à chercher parmi ceux qui réalisent la plus grande différence $F_1 - F_2$, et pour lesquels les grandeurs

$$(\alpha_u - \alpha_1), (\beta_u - \beta_1), (\gamma_u - \gamma_1) \text{ d'une part}$$

$$\text{et } (\alpha_2 - \alpha_u), (\beta_2 - \beta_u), (\gamma_2 - \gamma_u) \text{ d'autre part}$$

sont aussi voisines que possible.

Exemples

$$1 - \text{ Soit } \Sigma_2 - \Sigma_1 = 6$$

On choisit les sujets s_u tels que, par exemple :

$$F_1 = 1 ; F_2 = 5$$

$$(\alpha_u - \alpha_1) = 0 ; (\beta_u - \beta_1) = 1 ; (\gamma_u - \gamma_1) = 0$$

$$(\alpha_2 - \alpha_u) = 2 ; (\beta_2 - \beta_u) = 1 ; (\gamma_2 - \gamma_u) = 2$$

$$\text{On a alors : } n_u = \frac{1!}{1! 0! 1!} \cdot \frac{5!}{2! 1! 2!} = 30$$

Avec la même valeur de $\Sigma_2 - \Sigma_1$, si l'on choisit :

$$F_1 = 2 ; F_2 = 4$$

$$(\alpha_u - \alpha_1) = 1 ; (\beta_u - \beta_1) = 0 ; (\gamma_u - \gamma_1) = 1$$

$$(\alpha_2 - \alpha_u) = 2 ; (\beta_2 - \beta_u) = 2 ; (\gamma_2 - \gamma_u) = 0$$

$$\text{On a : } n_u = \frac{2!}{1! 0! 1!} \cdot \frac{4!}{2! 2! 0!} = 12$$

Et si, en conservant $F_1 = 2 ; F_2 = 4$, on choisit :

$$(\alpha_u - \alpha_1) = 1 ; (\beta_u - \beta_1) = 0 ; (\gamma_u - \gamma_1) = 1$$

$$(\alpha_2 - \alpha_u) = 2 ; (\beta_2 - \beta_u) = 1 ; (\gamma_2 - \gamma_u) = 1$$

$$\text{On a : } n_u = \frac{2!}{1! 0! 1!} \cdot \frac{4!}{2! 1! 1!} = 24$$

$$2 - \text{ Avec } \Sigma_2 - \Sigma_1 = 8$$

$$\text{Si l'on choisit : } F_1 = 1 ; F_2 = 7$$

$$\text{et } (\alpha_u - \alpha_1) = 1 ; (\beta_u - \beta_1) = 0 ; (\gamma_u - \gamma_1) = 0$$

$$(\alpha_2 - \alpha_u) = 2 ; (\beta_2 - \beta_u) = 2 ; (\gamma_2 - \gamma_u) = 3$$

$$\text{On a : } n_u = \frac{1!}{1! 0! 0!} \cdot \frac{7!}{2! 2! 3!} = 210$$

$$\text{Avec } (\alpha_u - \alpha_1) = 1 ; (\beta_u - \beta_1) = 0 ; (\gamma_u - \gamma_1) = 0$$

$$(\alpha_2 - \alpha_u) = 1 ; (\beta_2 - \beta_u) = 2 ; (\gamma_2 - \gamma_u) = 4$$

$$\text{On a : } n_u = \frac{1!}{1! 0! 0!} \cdot \frac{7!}{1! 2! 4!} = 105$$

$$\text{Avec } F_1 = 2 ; F_2 = 6$$

$$\text{et } (\alpha_u - \alpha_1) = 1 ; (\beta_u - \beta_1) = 1 ; (\gamma_u - \gamma_1) = 0$$

$$(\alpha_2 - \alpha_u) = 2 ; (\beta_2 - \beta_u) = 2 ; (\gamma_2 - \gamma_u) = 2$$

$$\text{On a : } n_u = \frac{2!}{1! 1! 0!} \cdot \frac{6!}{2! 2! 2!} = 90$$

E. Conclusions

Les résultats décrits dans cette Note peuvent aider à répondre à un certain nombre de questions, mais pour qu'ils soient vraiment utiles, encore faut-il que ces questions ne présupposent pas des postulats mal fondés.

S'il s'agit simplement de classer les sujets observés, en admettant la validité des deux hypothèses de départ, ce que l'on a appelé le *diagramme initial* répond à la question.

S'il s'agit de préciser les échelons manquants entre les sujets observés, alors apparaît le problème du choix entre tous les chemins possibles, et ce choix implique la définition de critères convenablement justifiés. Parmi ces critères, celui de la fréquence d'apparition d'un sujet intermédiaire dans l'ensemble des chemins possibles ne paraît pas épistémologiquement très fondé. Il est en effet important de ne pas confondre cette fréquence avec la probabilité de réalisation effective de l'état correspondant au cours de l'Evolution. L'évaluation de la fréquence ne tient pas compte des probabilités relatives des diverses transformations théoriquement possibles à partir d'un état donné, alors que l'Evolution réelle dépend de ces probabilités respectives qui peuvent éventuellement être différentes.

Le véritable problème se ramène donc en définitive à essayer de confirmer ou de réfuter les hypothèses de départ. Par exemple, on peut très bien imaginer que, pour des raisons de cohérence physiologique, les mutations sur les différents caractères ne peuvent pas toujours se faire de manière indépendante *. Des arguments en faveur ou contre cette hypothèse peuvent être recherchés dans l'examen des sujets effectivement observés. Ainsi, si l'on constate que l'on a toujours $\beta > \alpha$, on peut en conclure à la vraisemblance du fait que le caractère B doit toujours être « en avance » de une ou plusieurs mutations sur le caractère A. De telles contraintes réduisent évidemment le nombre des chemins possibles entre tout couple de sujets observés.

On peut aussi se demander si la correspondance admise entre les divers états mutationnels d'un caractère et la suite des nombres entiers est bien conforme à l'ordre d'apparition obligé de ces états au cours de l'Evolution. Par exemple, si l'on permute les états correspondants à A_3 et à A_4 , on a pour le diagramme évolutif l'équivalent du remplacement de la transformation $A_3 \rightarrow A_4$ par la transformation $A_4 \rightarrow A_3$.

* Dans cet ordre d'idées, voir par exemple : L.L. Whyte — Internal factors in Evolution — Tavistock Publications, London, 1965.

ce qui change en général la structure dudit diagramme. Il serait intéressant d'examiner les types de modifications subies par ce diagramme lorsque l'on opère les divers changements raisonnablement envisageables dans l'ordre temporel d'apparition des états sur les différents caractères. La question serait alors de savoir si l'on peut définir des critères de choix entre les diagrammes ainsi obtenus (ce pourrait être, par exemple, l'absence de plusieurs chemins entre tout couple de sujets effectivement observés).

Tout cela demande à être discuté, mais il est assez clair que la véritable question de fond est là, c'est-à-dire dans l'ordre d'apparition des états sur chaque caractère et dans les contraintes éventuelles entre les états respectifs des différents caractères à un instant donné. La meilleure façon d'aborder ce problème me semble résider dans les propositions qui viennent d'être évoquées : relevé des relations d'ordre éventuelles apparaissant entre les états des divers caractères sur l'ensemble des sujets effectivement observés, et argumentation physiologique sur les compatibilités ou incompatibilités entre les états simultanés de ces caractères.

ERRATA

Dans le texte de présentation de l'article de Georges Th. Guilbaud, «Pilotes, stratèges et joueurs : vers une théorie de la conduite humaine», paru dans le numéro 3 du volume 2 (rubrique Archives), il faut lire : page 321, ligne 3 du second paragraphe : Beaufrère (et non Beaufiles) et page 322, ligne 7 : auraient (et non avaient).

COMPTE RENDU

Réunion Annuelle de l'International Society for General System Research (mai 1988).

The 32nd. Annual Meeting and Conference of the International Society for General Systems Research was held in St Louis, Missouri, USA, on May 23-27th. 1988 under the presidency of the out-going President Dr. Russell L. Ackoff. The meeting was honored by the presence of Drs. C. West Churchman (President-Elect), K. Boulding, P.B. Checkland, J.J. Miller, A. Rapoport, T. Cowan, J. Warfield, and L. Troncale.

One of the main events of the annual meeting —the Ludwig von Bertalanffy Lecture— was delivered by Dr. Anatol Rapoport. He spoke about the role of General System Theory in the guidance of human affairs. He questioned the use of science and technology to seek knowledge, without questioning the value of the goals pursued. As citizens of the world, Rapoport urged us to rethink having a blind belief in the benefits of science. General Systems Theory must counter blind faith in science and idolatry of technology. These orientations are the driving forces behind the arms race. It is imperative that we uncover the false reasoning of those who push for an increase in armaments to improve security. The war machinery functions like an organism. It fights extinction by adaptation. It owes its survival and continued existence to successful adaptation to hostile environments. War is the product of creative science research devoid of ethics. As Rapoport asserted, war is an institution of modern societies which functions like a cancer : «It is a malignant growth» that is impossible to eradicate. It cannot be cured through «a technological fix» which leads to more armaments in the name of deterrence, itself a superstition. Rapoport noted the fallacies and paradoxes which justify the continued existence of the war machine. Only a change of thinking will save humanity from utter destruction. The systems movement can show the absurdity in the thinking that pervades the arms race. Is there hope and can we remain optimistic that the cancer will be exorcised ?