

# Représentations arborées de l'information verbatim

Hervé ABDI (\*)

Laboratoire de Psychologie  
Ancienne Faculté, 36, rue Chabot-Charny  
21000 Dijon

## I. INTRODUCTION

Cet article vise deux buts principaux : d'une part l'étude de la mémorisation de l'information lexicale, et d'autre part, la présentation d'une technique récente d'analyse descriptive multivariée : les représentations arborées.

Nous nous proposons, donc, de représenter par le dessin d'un arbre les rappels à différents intervalles d'un texte par des sujets. A cette fin, nous réduirons le corpus étudié à des ensembles de mots accompagnés de leur fréquence d'apparition dans chaque texte. Puis nous construirons des distances entre ces ensembles. Enfin, nous représentons ces distances par des arbres.

Mais avant même d'entamer l'exposé, il paraît nécessaire de justifier l'étude de la mémoire verbatim et plus précisément l'utilisation des simples formes lexicales comme variable dépendante. En effet, une grande partie des recherches insistent plus sur la mémorisation des thèmes d'un texte que sur la mémorisation de la formulation exacte d'un texte.

## II. DEFENSE ET ILLUSTRATION DE LA MEMORISATION VERBATIM

Si l'on s'accorde pour accepter l'existence d'une mémorisation verbatim dans des intervalles de temps assez courts (e.g., entre autres, Anderson et Paulson, 1977 ; pour une revue Cf. Parvella, 1979), on juge volontiers que la mémorisation à long terme porte plus sur le fond que sur la forme (Cf. Bartlett, 1932). Outre les remarquables travaux de Bartlett, cette dernière opinion est défendue par de nombreuses observations et expérimentations parmi lesquelles se trouvent les célèbres études de Sachs (1967, 1974) et Bransford (1972). Dans ces articles — comme dans la presque totalité de ceux que nous évoquerons dans ce paragraphe — les auteurs utilisent des épreuves de reconnaissance pour établir l'existence d'une mémorisa-

tion, en grande partie du fait de la plus grande sensibilité attribuée en général aux épreuves de reconnaissance face aux épreuves de rappel (les exceptions à la règle étant représentées par Kay, 1955 ; et Howe 1970).

En 1967, Sachs établit que ses Sujets n'arrivent à déceler que pendant un laps de temps très court (moins d'une trentaine de secondes) les modifications lexicales qui laissent intact le sens d'un énoncé (e.g. les Sujets ne distinguent pas « il envoya une lettre à ce propos à Galilée — le grand homme de science italien » de « il envoya à Galilée — le grand homme de science italien — une lettre à ce propos ») ; à l'inverse, les Sujets reconnaissent clairement les modifications sémantiques de la phrase : ils différencient « il envoya une lettre à ce propos à Galilée — le grand homme de Science Italien » de « Galilée — le grand homme de science Italien — lui envoya une lettre à ce propos ». Tout de même, Bransford et al. (1972 ; Cf. également Johnson et al. 1973), dans une expérience devenue classique, établit que ses Sujets ne retiennent pas la formulation exacte des phrases qu'ils entendent mais leur sens ; ainsi, ils confondent la phrase « trois tortues se prélassaient sur un tronc flottant et un poisson nageait en dessous d'elles » avec « trois tortues se prélassaient sur un tronc flottant et un poisson nageait en dessous de lui ».

Ces résultats (et bien d'autres) justifient l'intérêt porté à la mémorisation du fond plutôt que de la forme, et insistent justement — dans la lignée de Bartlett et Halbwach — sur l'aspect constructif de la mémoire. Néanmoins, d'autres travaux indiquent que l'information sur la formulation précise du matériel demeure accessible. Ainsi, dès 1955, Kay lit de

\* L'auteur tient à remercier pour leur aide précieuse et amicale : Xuan Luong, Jean-Pierre Barthélemy et Michel Fayol.

courts récits à ses Sujets et leur demande, à plusieurs reprises, de rappeler ce texte par écrit (Chaque écrit est séparé par une semaine) ; il observe une claire rétention de l'information Verbatim, et suggère que les Sujets restent plus fidèles à leur première reproduction qu'au texte d'origine (malgré une relecture du texte d'origine après chacune des six épreuves de rappel). Cette étude est reprise et étendue par Howe (1970, Cf. également Gault et Stephenson, 1967) qui, en outre, ajoute que les conclusions de Kay valent pour l'information Verbatim comme pour les « thèmes » des récits.

Par ailleurs, Begg et Paivio (1969) reprennent l'expérience de Sachs (1967) en distinguant les phrases concrètes des phrases abstraites. Ils observent alors un effet d'interaction : les modifications lexicales sont repérées — de mémoire — dans les phrases abstraites mais pas dans les phrases concrètes. Encore qu'il soit difficile de mettre à jour la Variable Indépendante responsable de cette interaction (Cf. Johnson et al., 1972 ; Pedzack et Royer, 1974 ; Kuiper et Paivio, 1977), ces travaux établissent l'existence d'une mémorisation Verbatim. Thème joué également par Hayes-Roth et Hayes-Roth (1977), qui soumettant les résultats de Sachs (1967) à une analyse détaillée, suggèrent qu'une partie de l'information lexicale est retenue. Pareillement, Reitman et Bower (1973) dans une reprise partielle de l'expérimentation de Bransford et al. (1972) utilisent comme matériel des lettres et des chiffres (plutôt que des mots en contexte) et trouvent (contrairement à Bransford) que leurs Sujets discriminent clairement les anciens items des distracteurs. Quoique ces données ne permettent pas de prétendre que les Sujets se souviennent parfaitement de la forme, elles révèlent qu'une certaine information de surface est retenue. Conclusion confirmée par Hayes-Roth et Hayes-Roth (1977) qui montrent — dans un contexte expérimental quelque peu artificiel — que l'on peut séparer des phrases effectivement lues de paraphrases de même sens. Dans un décor plus naturel — i.e. mémorisation d'un cours de psychologie — Kintsch et Bates (1977, Cf. également Goolkasian et al, 1979) retrouvent ces résultats, et remarquent, en outre, que la mémoire Verbatim encore relativement précise après un délai de deux jours décline après un délai de cinq jours. A ces auteurs s'ajoutent Abramovici (1984), Mc Koon (1977), Hayes-Roth et Thorndike (1979), Thorndike (1977), Yekovich et Thorndike (1981), Terry et Mason (1982) ainsi que Keenan et al (1977) ; ces derniers, dans une étude particulièrement astucieuse et très proche du milieu naturel démontrent que la mémorisation Verbatim peut découler d'un apprentissage incident. Ainsi, Keenan et al mesurent la reconnaissance de la formulation exacte de phrases produites spontanément lors d'une discussion informelle dans le département de Psychologie de Denver. Les participants à cette discussion ignoraient qu'elle était enregistrée et qu'ils subiraient plus tard une

épreuve de reconnaissance. Néanmoins, ils reconnaissent la formulation précise de presque la moitié des phrases pour peu qu'elles comportent un trait d'esprit, d'humour ou une critique personnelle. Ainsi, la mémoire Verbatim semble dépendre de l'intérêt accordé par le Sujet au matériel perçu (Cf. Graesser et Mandler, 1975), tout comme la mémorisation des « caractéristiques physiques » du message (e.g. timbre de la voix, etc., Cf. Fisher et Cuervo, 1983). Remarquons, au passage, que la capacité de retenir l'information lexicale peut être affectée par d'autres facteurs ; citons — entre autres — l'âge (Cf. Kail, 1979) et le niveau de lecture : les bons lecteurs retiennent mieux l'information lexicale que les lecteurs médiocres (Cf. Perfetti et Lesgold, 1979). Si l'on continue à s'éloigner du laboratoire, et si l'on décide de mettre en évidence, par une épreuve de rappel, l'existence d'une mémoire du « mot à mot », il suffit de demander à chacun d'entre nous de donner la suite de « Maître Corbeau sur un... », ou de trouver les paroles de la « Marseillaise » (Cf. les Sujets de Rubin, 1977, qui retrouvent l'Hymne national américain et le monologue de Hamlet, voir aussi Neisser, 1982 ; Hunter, 1984 ; pour une approche écologique du problème. Même des années après l'apprentissage, nous sommes capables de rappeler correctement mot à mot maints poèmes (sans parler de la table de multiplication, etc.).

Ainsi — encore que l'étude de la mémoire des récits ne saurait s'y réduire — il paraît justifié de sonder la mémoire de nos Sujets en utilisant les mots retenus comme Variable dépendante.

### III. CONSTRUCTION DE DISTANCES ENTRE TEXTES

Rappelons notre slogan : un texte est un ensemble de mots, et à chaque mot est associé sa fréquence d'utilisation. On peut, partant, pour comparer des textes construire des distances ; il en existe de nombreuses (Cf. Abdi, 1980 ; Abdi et Barthélémy, 1980), mais nous préférons, ici, n'en proposer que trois : la première (dite du chi-deux) de par sa popularité en analyse des données via l'Analyse des Correspondances (Benzecri, 1973), les deux autres (Différence Symétrique et Différence Symétrique généralisée) de par leur adéquation au modèle de représentation arborée (Cf. Tverski, 1977 ; Sattah et Tverski, Abdi et al. 1984 a et 1984 b). Commençons par la plus évidente : la distance de la différence symétrique ; considérons nos textes comme un simple ensemble de mots : un mot est ainsi présent ou absent et l'on néglige de ce fait sa fréquence (ce qui revient à poser un seuil en dessous duquel le mot est absent et au-dessus duquel le mot est présent) ; prenons deux textes, une mesure de leurs divergences peut s'obtenir simplement en comptant le nombre de mots qui appartiennent à un seul des deux textes, ce

qui revient à dénombrer les éléments de la différence symétrique des deux ensembles de mots représentant ces textes. On peut remarquer que cette distance n'est qu'un cas particulier d'une distance plus générale : si nous associons à chaque mot d'un texte sa fréquence (si un mot est absent d'un texte sa fréquence est nulle) que nous noterons  $f(m,t)$  et que nous lirons « fréquence du mot  $m$  dans le texte  $t$  »; alors, nous pouvons définir la distance du texte  $t$  au texte  $t'$  par la formule :

$$d(t, t') = \sum_m |f(m, t) - f(m, t')|$$

$$d^2(t, t') = \sum_m (f(m, t)/f(t) - f(m, t')/f(t'))^2 / f(m)$$

Avec  $f(t)$  total des fréquences des mots de  $t$ ;  
 $f(m)$  : total des fréquences du mot  $m$   
 pour l'ensemble des textes du corpus considéré.

Après l'évocation rapide des distances que nous utiliserons, nous allons maintenant examiner les représentations sous forme d'arbre d'une matrice de distance.

IV. REPRESENTATION ARBOREES

Avant tout, quelques définitions : un arbre est un graphe non-orienté connexe et sans cycle (i.e. dans un arbre un chemin et un seul relie deux sommets, la Figure 1 illustre notre définition quelque peu abstruse). Nous distinguerons les sommets terminaux de l'arbre (« les feuilles » de l'arbre) des sommets intérieurs

On appellera la distance définie par cette formule : distance de la différence symétrique généralisée (elle apparaît également en pays anglo-saxon sous le nom de distance de Hamming généralisée), on remarquera que cette formule donne la distance de la différence symétrique lorsque les mots ne prennent que les valeurs 0 ou 1.

Enfin, la distance du chi-deux (dont Benzecri et ses collaborateurs furent les ardents promoteurs) compare les textes par leur profil (i.e. leur distribution de fréquences; pour une approche intuitive claire et précise de cette distance Cf. Fénelon, 1981), nous l'obtenons par la formule :

(« les nœuds » de l'arbre). En règle générale, les textes que nous voulons décrire seront représentés par les feuilles de l'arbre. Chaque arête (« branche ») possède une certaine longueur, et la distance « sur l'arbre » d'un sommet à l'autre s'obtient par la longueur du chemin les reliant. On peut, donc, à partir d'un arbre construire une matrice de distance; à l'inverse la construction de l'arbre à partir d'une matrice de distance (ou de ressemblance, etc.) constitue un problème plus délicat, mais plusieurs algorithmes permettent d'effectuer cette tâche (Cf. Cuninghame, 1974; Carrol et Pruzansky, 1975, Sattah et Tversky, 1977; De Soete, 1983; Luong, 1983; pour plus de détails Cf. Abdi et al, 1984a; pour un historique des représentations arborées Cf. Abdi et al, 1984b, et à paraître).

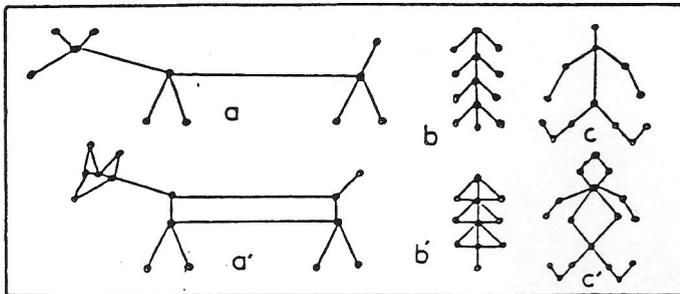


Figure 1

Figure 1 : Les graphes a, b, c sont des arbres; mais pas les graphes a', b', c'.

Lorsque la distance initiale correspond à un arbre, les différents algorithmes le retrouvent (ce qui est heureux); sinon le problème se pose de trouver une approximation de la distance de départ par une distance d'arbre (c'est sur ce point que les différents algorithmes divergent, pour notre part, nous utilisons un algorithme dû à Luong, 1983; Cf. Abdi et al, 1984). De ce fait, nous devons évaluer la « qualité de la représentation ». (Ce que les Anglo-

Saxons dénomment « Goodness of Fit »); sans dénigrer les différents candidats possibles (e.g. les différents « stress » proposés par Kruskall (1978) dans le cadre du « Multidimensional Scaling »), nous proposerons le Carré du coefficient de Corrélation entre la matrice de distance d'origine et la matrice de distance reconstituée à partir de l'arbre. Ce coefficient présente l'avantage de pouvoir s'interpréter comme la part de variance expliquée par le modèle arbo-

ré (de la même manière que le carré du coefficient de corrélation multiplie exprime la part de variance expliquée par le modèle linéaire dans la régression multiple). Par analogie avec les habitudes de l'analyse descriptive multidimensionnelle (e.g. Analyse des Correspondances et Multidimensional Scaling) nous exprimerons la part de variance expliquée par un pourcentage et nous le noterons  $\tau$  (tau).

### V. SAVEZ-VOUS PLANTER LES ARBRES (LA GUERRE DES FANTOMES...)

En 1932, Bartlett raconte à ses Sujets (qui étaient en règle générale des amis) un conte indien (intitulé « la Guerre des Fantômes » — the War of the Ghosts) et leur demande à plusieurs occasions de le raconter de mémoire. Il baptise cette manière de faire — inspirée de

Philippe (1897) — « méthode des reproductions répétées » (repeated reproductions method). On sait que Bartlett utilisa ses résultats pour mettre en avant la notion de schéma (qui se trouve abondamment illustrée dans ce volume) et pour insister sur « l'aspect constructif » de la mémoire des récits. Il fournit, dans son ouvrage classique, une quinzaine de protocoles donnée par six Sujets. Nous allons utiliser ces données à titre d'exemple et d'exercice pour nous familiariser avec les arbres. Rappelons que les textes sont ici vus comme des ensembles de mots entre lesquels nous construisons des distances (Cf. Supra). Tout d'abord, nous examinerons les résultats obtenus à partir de la distance de la différence symétrique (La Matrice de départ et la matrice reconstituée sont données en Annexe). L'arbre construit à partir de cette distance est dessiné dans la Figure 2.

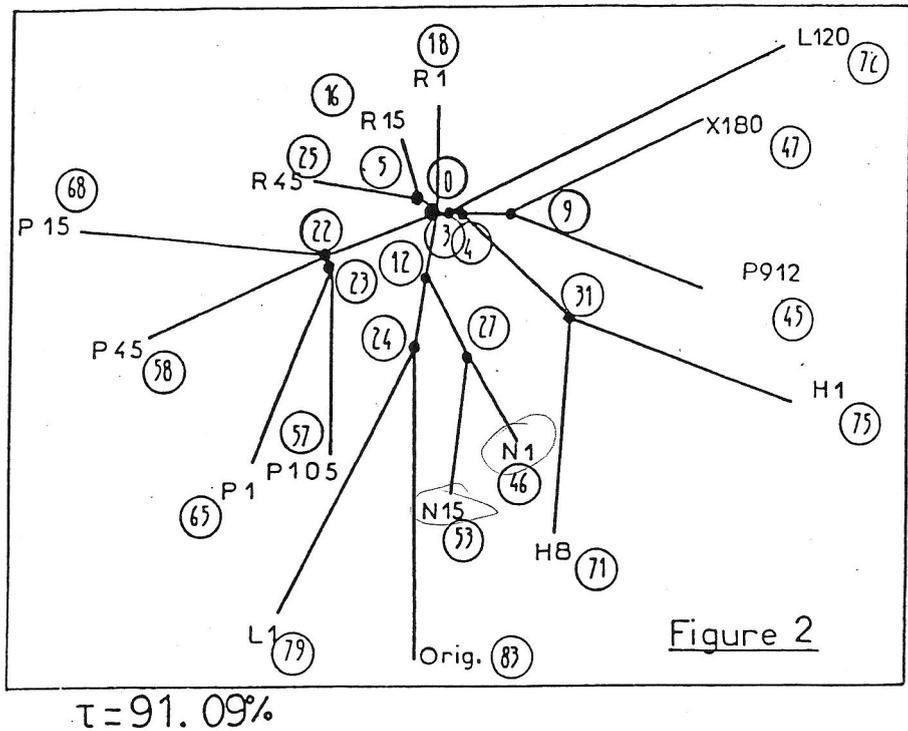


Figure 2 : Représentation arborée de la distance de la différence symétrique calculée à partir des données de Bartlett (1932).

Examinons cette figure; et commençons par le « mode d'emploi » : A chaque sommet terminal (i.e. à chaque feuille) correspond un texte repéré par un identificateur dont la première lettre indique le Sujet (i.e. H, L, N, P, R et X), les chiffres suivants donnent le nombre de jours écoulés depuis la lecture du texte d'origine; le texte d'origine est noté Orig.

A côté de chaque sommet du graphe on trouve un nombre encerclé que nous appellerons « degré d'excentricité du sommet » (que nous abrègerons souvent en « excentricité du

sommet »). Nous allons maintenant examiner la signification et l'utilité de ces nombres (ce qui suit reprend Abdi, 1984). Admettons que nos données se conforment exactement au modèle arboré (i.e. que la matrice de distance est effectivement une distance sur un arbre); et prenons deux textes, disons N15 (d'excentricité 53) et N1 (d'excentricité 46), ces deux textes se rejoignent en un sommet d'excentricité 27, cette valeur représente le nombre de mots communs à N15 et N1 qui ne font pas partie du vocabulaire commun à tous les textes

étudiés (ou si l'on préfère, on décide de négliger l'ensemble des mots employés par tous les textes, nous suivrons cette convention afin d'alléger l'exposé). De même, la longueur de l'arête qui relie N15 au sommet Union représente le nombre de mots propres à N15 et l'on obtient cette longueur simplement par la différence des degrés d'excentricité (ici :  $53-27=26$ ), la distance sur l'arbre de N15 à N1 s'interprète alors comme la somme des mots propres à N15 et des mots propres à N1, et l'on retrouve, ainsi, la distance de la différence symétrique.

Cette interprétation se généralise sans peine aux autres sommets : par exemple Orig et N1 se rejoignent en un sommet d'excentricité 12, et c'est le nombre de leurs mots communs ; la longueur du chemin de Orig à ce sommet (71) correspond au nombre de mots présents dans l'original et absents de N1, tout comme la longueur du chemin de ce sommet à N1 (34) représente le nombre de mots de N1 absents de Orig ; la longueur totale du chemin ( $71 + 34 = 105$ ) donne la valeur de la distance de la Différence symétrique. On peut trouver plus simplement le nombre de mots communs à deux textes : il suffit de tracer le chemin qui relie ces deux textes et de repérer la plus petite excentricité : c'est le nombre de mots communs. (En fait, cette manière de faire s'applique à un nombre quelconque de sommets). On peut remarquer, maintenant, un sommet particulier d'excentricité nulle, ce sommet est, ainsi, le moins « excentrique » ou le « plus central » : c'est la médiane de l'arbre (entendez : c'est le sommet à distance minimum de tous les autres). De ce fait si le chemin reliant deux textes passe par la médiane, ces textes ne possèdent pas de mots en commun (sous la réserve mentionnée ci-dessus : on néglige les mots communs à tous les textes). On peut de manière équivalente (quoique plus aisée à mettre en formule) obtenir le nombre de mots communs à deux textes (ou plus généralement : à deux sommets) en remarquant que le sommet recherché est le centre des deux sommets considérés et de la médiane (pour la notion de centre dans un graphe Cf. Hakimi et Yau, 1964, pour l'équivalence des différents points de vue, Cf. Abdi, 1984) ; de ce fait, nous pouvons calculer son excentricité à partir de la matrice de distance sur l'arbre : appelons  $x$  et  $y$  les textes et  $M$  la médiane, le nombre de mots communs à  $x$  et  $y$  s'obtient par la formule :

$$(d(x, M) + d(y, M) - d(x, y)) / 2$$

Nous pouvons, ainsi, reconstituer le tableau des mots communs (Cf. Tableau 3, en Annexe). Ici, par exemple, on note que l'original et N1 possèdent :  $(83 + 46 - 105) / 2 = 12$  mots en commun. Remarquons en outre que la somme de chaque colonne (exceptée la valeur de la diagonale) fournit un indicateur de représentativité, plus précisément : de Communauté. En effet, plus le total d'une colonne est élevé,

plus un texte possède de mots en commun avec d'autres textes, et plus, par conséquent, il pourra les représenter fidèlement ; cette notion de communauté s'oppose d'une certaine manière à la notion « d'agrégation à distance minimum » qui conduit à représenter un ensemble de textes par un texte moyen ou médian. En effet, si nous associons à chaque texte d'une part, cet indice de communauté, et d'autre part la somme des distances de ce texte aux autres textes (Cf. Tableau 4 en annexe), nous voyons clairement que ces deux indices ne se confondent pas. En particulier, le critère de « communauté maximale » conduit à représenter notre ensemble de textes par « P1 » ou « P105 » ; mais le critère « du minimum de la somme des distances » élit comme représentant « R45 » (qui est le texte le plus proche de la Médiane de l'arbre). Remarquons, au passage, que si nous cherchons le texte « optimisant simultanément les deux critères », nous retrouvons le texte d'origine. On pourrait, par ailleurs, relier cette discussion au problème de la typicalité évoquée dans le contexte de la structuration des Concepts par Rosch (1973, 1975, 1983, pour une revue. Cf. Cordier et Dubois, 1981 ; Medin et Smith, 1984) : le problème de la catégorisation des concepts ainsi défini conduisant à rechercher des candidats pour représenter un ensemble de concepts en leur demandant « d'optimiser » simultanément divers critères. Mais après ce détour technique, revenons à nos résultats.

L'arbre dessiné dans la Figure 2 « rend compte » de 91 % de la variance de la matrice de distance de la différence symétrique ; cet arbre offre, ainsi, un résumé de qualité. A sa vue, l'étonnante fidélité des Sujets à eux-mêmes s'impose d'emblée : les différentes reproductions d'un Sujet se retrouvent en général proches les unes des autres. Relevons, néanmoins, deux exceptions : d'une part, les quatre premières reproductions de P (de 1 à 105 jours) sont remarquablement groupées et constituent à elles seules une branche de l'arbre, cependant deux ans et demi plus tard, le texte de la cinquième reproduction de P se perche sur une autre branche ; d'autre part, le Sujet L passe de la « fidélité maximale » le premier jour à « l'infidélité maximale » quatre mois plus tard.

Les autres distances (Différence symétrique généralisée et chi-deux) conduisent aux mêmes conclusions et nous ne les détaillerons pas ; nous donnons simplement ci-dessous les arbres obtenus. L'arbre construit à partir de la distance de la Différence symétrique généralisée donne le degré d'extrencité de chaque sommet ; l'excentricité s'interprète de manière identique à celle que nous avons détaillée ci-dessus mais en prenant en compte le fait que chaque mot est pondéré par sa fréquence ; ainsi, par exemple, le nombre de mots communs à deux textes devient le minimum des fréquences de chaque mot (cela revient à généraliser les concepts classiques des ensembles en termes de « sous-ensembles flous » ou de « mesures molles »).

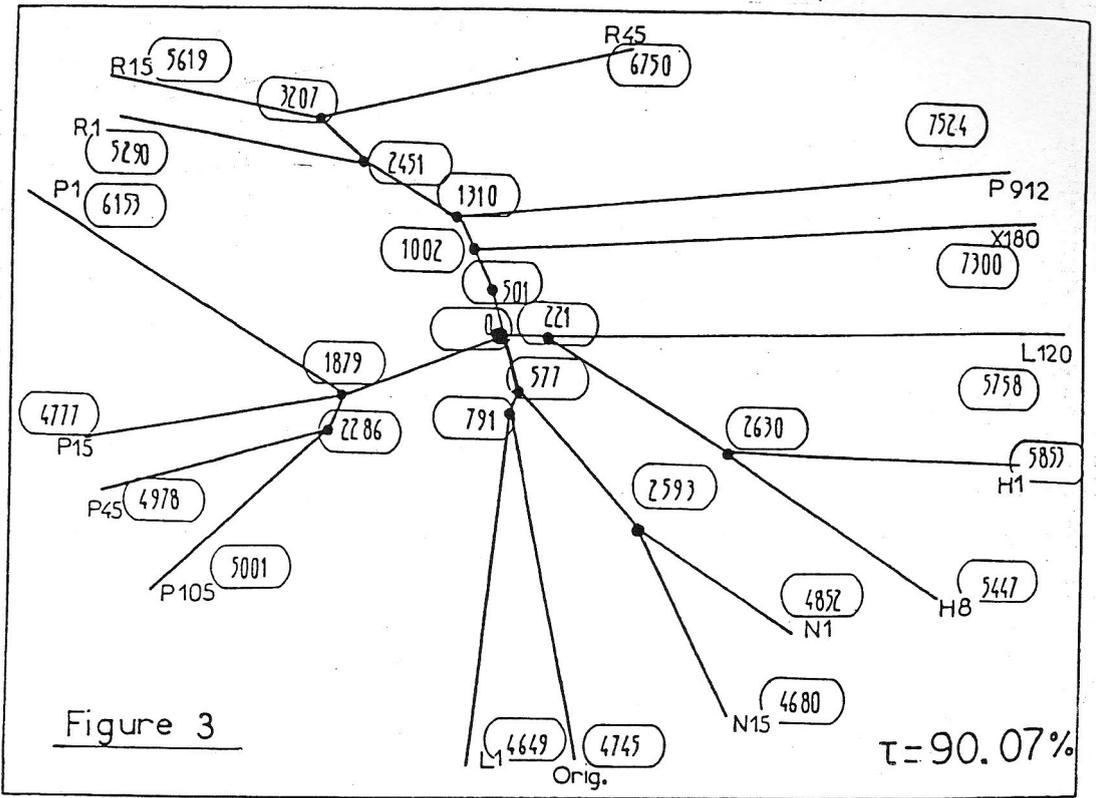


Figure 3 : Représentation arborée de la distance de la Différence symétrique généralisée calculée à partir des données de Bartlett (1932).

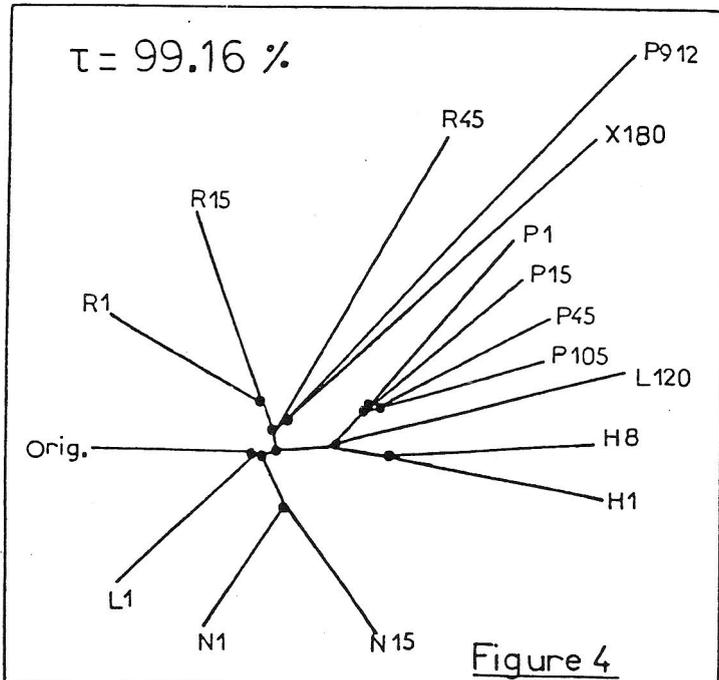


Figure 4 : représentation arborée de la distance du Chi-deux calculée à partir des données de Bartlett (1932).

VI. DISCUSSION

Ainsi, les représentations arborées du vocabulaire des textes recueillis par Bartlett révèlent une fidélité d'ensemble des Sujets à eux-mêmes (avec, néanmoins, les réserves élevées ci-dessus). Mais deux interprétations — au moins — s'accordent avec cette fidélité : soit les Sujets restent proches de leur première reproduction du texte, soit leurs habitudes linguistiques les poussent à utiliser le même vocabulaire lors des différentes reproductions du texte. Le premier terme de cette alternative correspond à une certaine rétention verbatim à long terme (Via le filtre de l'endocage lors de la lecture originelle); le second terme de l'alternative transforme l'apparente fidélité des Sujets en épiphénomène. Par malheur, les données recueillies par Bartlett ne permettent pas de trancher. En effet, pour séparer ces deux interprétations il faudrait pouvoir obtenir au moins deux séries de reproductions par Sujet; l'hypothèse de constance mnésique prédit que les reproductions par un même Sujet de deux textes différents devront peu se ressembler (i.e. les reproductions des deux textes devront se placer sur des branches distinctes); à l'inverse l'hypothèse des « habitudes linguistiques » prédit que les reproductions d'un même Sujet devront se ressembler indépendamment du texte d'origine (i.e. les productions de chaque Sujet doivent se placer sur une même branche). En outre, si les Sujets retiennent effectivement une certaine information verbatim alors les reproductions de textes différents doivent se séparer en autant d'arbrisseaux qu'il y a de textes (chaque arbrisseau correspondant aux différentes reproductions d'un texte).

Afin de décider entre ces hypothèses concurrentes, nous avons mis en place une petite expérience que nous détaillons ci-dessous.

VI. PAPILLONS ET SCIENCE-FICTION

Nous tenons à remercier J. Barbieri, A. Champenois et M. Lacaze, Psychologues Scolaires en Formation à Besançon, pour leur participation à cette expérimentation).

Conformément à la discussion ci-dessus, nous avons demandé à quelques Sujets de lire deux textes différents et de les reproduire de mémoire un jour, huit jours et quinze jours plus tard. Le premier texte « La Dame aux Papillons » est adapté d'un conte indien, le second « Monsieur Mead » est adapté d'une nouvelle de Science Fiction inspirée de 1984. (Ces textes, ainsi que les reproductions des Sujets sont à disposition du lecteur intéressé). Les Sujets étaient prévenus de notre volonté de sonder leur mémoire aux intervalles indiqués, mais sans leur préciser explicitement notre intérêt pour une mémorisation verbatim. Nous donnons ici les résultats obtenus par quatre Sujets. L'arbre dessiné ci-dessous est obtenu à partir de la distance de la Différence symétrique généralisée; nous avons préféré cette distance qui, d'une part, s'interprète aussi aisément que la distance de la Différence symétrique (elle est ainsi profondément en accord avec le modèle arboré), et d'autre part, prend en compte l'information apportée par la fréquence d'usage des mots du vocabulaire (mais les conclusions cueillies sur cet arbre fleurissent également sur les arbres construits avec les autres distances). Ajoutons encore un détail technique: pour de (viles) raisons informatiques nous n'avons retenu que les mots apparaissant au moins quatre fois dans l'ensemble des textes. Les codes des textes de la Figure 5 obéissent au code suivant : La première lettre donne l'initiale du Sujet (i.e. A, C, M., P), la seconde repère le texte d'origine (P ou M), le nombre terminal indique le délai (en jours) séparant la lecture du texte original de la reproduction.

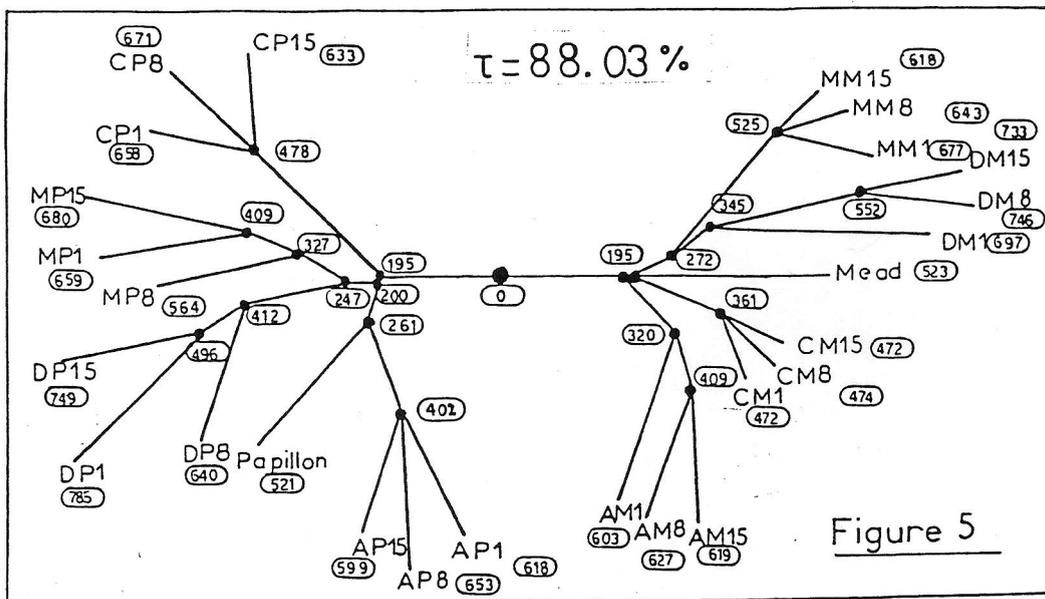


Figure 5 : représentation arborée des « reproductions répétées » de deux textes par quatre Sujets (distance de la Différence symétrique généralisée).

Un simple regard à la Figure 5 permet de trancher entre les hypothèses concurrentes : l'arbre se sépare clairement en deux arbrisseaux bien distincts. Chacun de ces arbrisseaux correspond à un texte d'origine et aux différentes reproductions de ce texte, et les reproductions d'un même texte par un même Sujet se regroupent en une branche. Ainsi s'élimine l'hypothèse des « habitudes linguistiques » : la fidélité des Sujets à leur première reproduction ne reflète pas la constance de leur style. En outre, la réunion en une même branche des différentes reproductions d'un Sujet s'accorde avec l'idée d'une fidélité des Sujets à leur première reproduction plus qu'à l'original (ainsi que le proposaient déjà Kay, 1955 ; et Howe, 1970). On remarquera, de surcroît, que les différentes reproductions d'un même Sujet sont plus proches entre elles qu'elles ne le sont de l'Original, ainsi que la fréquente tendance des textes d'un même Sujet à se réunir en un

seul sommet (en dessinant une « patte d'oie ») : cette configuration suggère la possibilité d'une trace mnésique « réinterprétée » à chaque rappel.

## VII. CONCLUSION

Ainsi, les représentations arborées obtenues à partir de distances construites sur le vocabulaire des textes offrent un résumé graphique de bonne qualité de l'ensemble des résultats.

Dans un premier temps (Cf. Bartlett), ces arbres mettent en évidence une surprenante fidélité d'ensemble des Sujets à eux-mêmes. Puis, les représentations arborées construites à partir des résultats de notre petite expérience confirment l'hypothèse d'une rétention de l'information lexicale, et permettent d'éliminer la possibilité que cette conclusion ne soit le simple fruit d'un artéfact.

## BIBLIOGRAPHIE

- ABDI (H.) 1980. — *Approches préordinales en analyse des données Psychologiques* ; Aix-en-Provence : Thèse.
- ABDI (H.) 1984. — Comment lire une représentation arborée ; *Cinquièmes Journées « Mathématiques discrètes et Sciences Sociales »* ; Dijon.
- ABDI (H.), BARTHELEMY (J.P.) 1980. — Analyse de contenu, analyse ordinale des données ; *Bulletin de Psychologie de F.C.*, 1-18.
- ABDI (H.), BARTHELEMY (J.P.), LUONG (X.) 1984 a. — Tree representations of associative structures in semantic and episodic memory research ; In DEGREEF E., VAN BUGGENHAUT J., (Eds.), *Trends in Mathematical Psychology* ; New York : Elsevier.
- ABDI (H.), BARTHELEMY (J.P.), LUONG (X.) 1984 b. — Représentations arborées, Catégories et concepts ; *Actes du colloque S.F.P. : Les apports de l'intelligence artificielle à la Psychologie*, Grenoble, (15 p.).
- ABDI (H.), BARTHELEMY (J.P.), LUONG (X.) (à paraître). — Représentation arborée et catégories naturelles ; In TIBERGHEN (G.), BONNET (C.), HOC (J.M.). — (Eds.), *Psychologie et intelligence artificielle* ; Bruxelles : Mardaga.
- ABRAMOVICI (S.) 1984. — Lexical information in reading and memory ; *Reading Research Quarterly*, XIX (2), 173-83.
- ANDERSON (J.R.), PAULSON (R.) 1977. — Representation and retention of verbatim information ; *Journal of Verbal Learning and Verbal Behavior*, 16, 439-451.
- BARTLETT (F.C.) 1932. — *Remembering* ; Cambridge : C.U.P.
- BATES (E.), MASLING (M.), KINTSCH (W.) 1978. — Recognition memory for aspects of dialogue ; *Journal of Experimental Psychology : Human Learning and Memory*, 4, 187-97.
- BEGG (I.), PAIVIO (A.) 1969. — Concreteness and imagery in sentence meaning ; *Journal of Verbal Learning and Verbal Behavior*, 8, 821-827
- BENZECRI (J.P.) 1973. — *L'analyse des données* (2 Vol.) ; Paris : Dunod.
- BRANSFORD (J.D.), BARCLAY (J.R.), FRANKS (J.J.) 1972. — Sentence memory : a constructive versus interpretive approach ; *Cognitive Psychology*, 3, 193-209.
- CARROL (J.D.), PRUZANSKY (S.) 1975. — Fitting of hierarchical tree structure models ; paper presented at the *U.S.-Japan Seminar on M.D.S.* ; San Diego.
- CORDIER (F.), DUBOIS (D.) 1981. — Typicalité et représentation cognitive ; *Cahiers de Psychologie Cognitive* ; 3 (1), 299-334.
- CUNNINGHAM (J.P.) 1974. — Finding and optimal tree realization of a proximity matrix ; paper presented at the *Mathematical Psychology Meeting* ; Ann Arbor.
- DE SOETE (G.) 1983. — A least squares algorithm for fitting additive trees to proximity data ; *Psychometrika*, 48 (4), 621-6.
- FENELON (J.P.) 1981. — *Qu'est-ce que l'analyse des données* ; Paris : LEFONEN.

FISHER (R.P.), CUERVO (A.) 1983. — Memory for physical features of discourse as a function of their relevance; *Journal of Experimental Psychology: Learning Memory and Cognition*, 9 (1), 130-8.

GAULT (A.), STEPHENSON (G.M.) 1967. — Some experiments relating to BARTLETT's Theory of remembering; *British Journal of Psychology*, 58 (1 et 2), 39-49.

GOOLKASIAN (P.), TERRY (W.S.), PARK (D.C.) 1979. — Memory for lectures: Effect of delay and distractor type; *Journal of Educational Psychology*, 71, 465-70.

GRAESSER (A.), MANDLER (G.) 1975. — Recognition memory for the meaning and surface structure of sentences; *Journal of Experimental Psychology: Human Learning and Memory*, 1 238-48.

HAKIMI (S.L.), YAU (S.S.) 1964. — Distance matrix of a graph and its realizability; *Quarterly of Applied Mathematics*; 22; 305-7.

HALBWACHS (M.) 1952. — *Les cadres sociaux de la mémoire*; Paris: P.U.F.

HAYES-ROTH (B.), HAYES-ROTH (F.) 1977. — The prominence of lexical information in memory representation of meaning; *Journal of Verbal Learning and Verbal Behavior*, 16, 119-36.

HAYES-ROTH (G.), THORNDIKE (P.) 1979. — Integration of knowledge from text; *Journal of Verbal Learning and Verbal Behavior*, 18, 91-108.

HOWE (M.J.A.) 1970. — Repeated presentations and recall of meaningful prose; *Journal of Educational Psychology*, 61 (3), 214-9.

HUNTER (I.M.L.) 1984. — Lengthy Verbatim recall (LVR) and the mythical gift of tape recorder memory; In LAGERSPETZ (K.M.J.), NIEMI (P.), (Eds.), *Psychology in the 1990's*; New York: Elsevier.

JARVELLA (R.J.) (1979). — Immediate memory and discourse processing; In BOWER G.H. (Ed.), *The psychology of Learning and motivation* (Vol. 13); New York: Academic Press.

JOHNSON (M.K.), BRANSFORD (J.D.), NYBERG (S.F.), CLEARY (J.J.) 1972. — Comprehension factors in interpreting memory for abstract and concrete sentences; *Journal of Verbal Learning and Verbal Behavior*, 11, 451-4.

JOHNSON (M.K.), BRANSFORD (J.D.), SOLOMON (S.) 1973. — Memory for tacit implications of sentences; *Journal of Experimental Psychology*, 98, 203-5.

KAIL (R.) 1979. — *The development of Memory in children*; San Francisco: Freeman.

KAY (H.) 1955. — Learning and retaining verbal material; *British Journal of Psychology*, 44, 81-100.

KEENAN (J.M.), MacWHINNEY (B.), MAYHEW (D.) 1977. — Pragmatics in memory: a study of natural conversation; *Journal of Verbal Learning and Verbal Behavior*, 16, 549-60.

KINTSCH (W.), BATES (E.) 1977. — Recognition memory for statements from a classroom lecture; *Journal of Experimental Psychology: Human Learning and Memory*; 3, 150-9.

KRUSKAL (J.B.), WISH (M.) 1978. — *Multi-dimensional Scaling*; Beverly Hills: Sage Publications.

KUIPER (N.A.), PAIVIO (A.) 1977. — Incidental

recognition memory for concrete and abstract sentences equated for comprehensibility; *Bulletin of the Psychonomic society*, 9, 247-9.

LUONG (X.) 1983. — Voisinage lâche; score et famille scorante; Besançon: Multigraphié.

McKOOON (G.) 1977. — Organisation of Information in text memory; *Journal of Verbal Learning and Verbal Behavior*, 16, 247-260.

MEDIN (D.L.), SMITH (E.E.) 1984. — Concepts and concept formation; *Annual Review of Psychology*; 35, 113-38.

NEISSER (U.) 1982. — *Memory observed: remembering in natural contexts*; San Francisco: Freeman.

PEDZEK (K.), ROYER (J.M.) 1974. — The role of comprehension in learning concrete and abstract sentences; *Journal of Verbal Learning and Verbal Behavior*, 13, 551-8.

PERFETTI (C.A.), LESGOLD (A.M.) 1979. — Coding and comprehension in skilled reading and implication for reading instruction; In RESNICK (L.R.), WEAVER (P.A.), (Eds.), *Theory and practice of early reading* vol. 1; Hillsdale: L. Erlbaum.

PHILIPPE (M.) 1897. — Sur les transformations de nos images mentales; *Revue Philosophique*; XLII, 481-93.

REITMAN (J.S.), BOWER (G.H.) 1973. — Storage and later recognition of exemplars of concepts; *Cognitive Psychology*, 4, 194-206.

ROSCH (E.) 1973. — On the internal structure of perceptual and semantic categories; In MOORE (T.E.), (Ed.), *Cognitive development and the acquisition of language*; New York: Academic Press.

ROSCH (E.) 1975. — Cognitive representations of semantic categories; *Journal of Experimental Psychology: General*; 104, 192-233.

ROSCH (E.) 1983. — Prototype classification and logical classification: the two systems; In SCHOLNIK (E.K.), (Ed.), *New trends in conceptual representations*; Hillsdale: Erlbaum.

RUBIN (D.C.) 1977. — Very long-term memory for prose and verse; *Journal of Verbal Learning and Verbal Behavior*, 16, 611-21.

SACHS (J.S.) 1967. — Recognition memory for syntactic and semantic aspects of connected discourse; *Perception and Psychophysics*, 2 (9), 437-442.

SACHS (J.S.) 1974. — Memory in reading and listening to discourse; *Memory and Cognition*, 2 (1A), 95-100.

SATTAH (S.), TVERSKI (A.) 1977. — Additive similarity tree; *Psychometrika*; 42 (3); 319-45.

SMITH (E.E.), MEDIN (D.L.), 1981. — *Categories and concepts*; Harvard: H.U.P.

TERRY (W.S.), MASON (H.M.) 1982. — Recognition of sentences from prose; *Bulletin of the Psychonomic Society*; 19 (1), 7-10.

TVERSKI (A.) 1977. — Features of similarity; *Psychological Review*, 84, 327-352.

TVERSKI (A.), GATI (I.) 1978. — Studies in similarity; In ROSCH (E.), LLYOD (B.B.), (Eds.), *Cognition and categorisation*; Hillsdale: L. Erlbaum.

YEKOVICH (F.R.), THORNDYKE (P.W.) 1981. — An evaluation of alternative functional models of narrative schemata; *Journal of Verbal Learning and Verbal Behavior*, 20, 454-69.

ANNEXE

	ORIG	H1	H8	H1	H15	L1	L120	X180	P1	P15	P45	P105	P912	R1	R15	R45
ORIG	0	158	154	103	103	113	153	138	158	143	138	142	144	103	112	127
H1	158	0	84	126	131	149	141	112	140	147	140	134	112	103	102	107
H8	154	84	0	120	127	147	135	116	136	129	130	128	108	101	102	113
H1	103	126	120	0	43	107	121	90	112	119	106	108	98	65	68	79
L1	103	131	127	43	0	110	126	101	131	132	113	121	101	74	75	80
L1	113	149	147	107	110	0	136	135	143	140	137	135	137	114	117	130
L120	153	141	135	121	126	136	0	113	137	142	129	127	117	106	103	106
X180	138	112	116	90	101	135	113	0	112	123	108	106	74	73	64	67
P1	158	140	136	112	131	143	137	112	0	89	92	76	104	105	100	109
P15	143	147	129	119	132	140	142	123	89	0	83	81	119	106	99	110
P45	138	140	130	104	113	137	129	108	92	83	0	70	104	91	92	105
P105	142	134	128	108	121	135	127	106	74	81	70	0	96	99	98	107
P912	144	112	108	98	101	137	117	74	104	119	104	96	0	67	60	61
R1	103	103	101	65	74	114	106	73	105	104	91	99	67	0	29	48
R15	112	102	102	68	75	117	103	64	100	99	92	98	60	29	0	31
R45	127	107	113	79	80	130	106	67	109	110	105	107	61	48	31	0

TABLEAU 1: Matrice de distance de la différence symétrique calculée à partir des données de BARTLETT (1932)

orig	H1	H8	H1	H15	L1	L120	X180	P1	P15	P45	P105	P912	R1	R15	R45	17	18	19	20	21	22	23	24	Med	27	28	
Orig	0	158	154	105	112	113	154	130	148	150	141	139	128	101	98	107	58	114	87	106	87	71	92	104	83	86	85
H1	158	0	84	121	128	154	142	115	140	142	133	132	113	93	90	100	99	44	103	98	80	87	77	97	75	71	72
H8	154	84	0	117	124	151	138	111	136	139	129	128	109	90	87	96	96	40	99	94	76	83	73	93	71	67	69
H1	105	121	117	0	43	101	118	93	111	113	104	103	91	64	61	70	46	77	18	69	50	34	55	48	44	49	48
H15	112	128	124	43	0	109	125	100	118	120	111	110	98	71	68	78	54	84	25	76	57	41	62	75	53	57	55
L1	113	154	151	101	109	0	151	127	145	147	138	136	125	98	95	104	65	110	83	102	84	68	89	101	79	83	82
L120	154	142	138	118	125	151	0	114	137	139	130	129	112	90	87	97	96	98	100	95	76	84	76	94	72	71	69
X180	130	115	111	93	100	127	114	0	113	115	106	104	74	66	63	72	72	71	75	70	52	59	38	69	47	44	45
P1	148	140	136	111	118	145	137	113	0	89	60	76	110	84	81	90	90	94	93	42	70	77	74	43	65	69	68
P15	150	142	139	113	120	147	139	115	89	0	82	81	113	86	83	92	92	99	95	47	72	79	77	46	68	71	70
P45	141	133	129	104	111	138	130	106	80	82	0	72	103	77	74	83	83	89	84	36	63	70	68	36	58	62	61
P105	139	132	128	103	110	136	129	104	76	81	72	0	102	75	72	82	81	88	85	34	61	69	66	35	57	61	59
P912	128	113	109	91	98	125	112	74	110	113	103	102	0	64	61	70	70	69	73	68	50	57	36	67	45	42	43
R1	101	93	90	64	71	98	90	66	84	86	77	75	64	0	34	43	43	49	44	41	23	30	28	40	18	22	21
R15	98	90	87	61	68	95	87	63	81	83	74	72	61	34	0	31	40	47	43	39	11	27	25	37	16	19	18
R45	107	100	96	70	78	104	97	72	90	92	83	82	70	43	31	0	49	56	52	48	20	37	34	46	25	28	27
17	58	99	96	46	54	55	96	72	90	92	83	81	70	43	40	49	0	56	29	48	29	13	34	46	24	28	27
18	114	44	40	77	84	110	98	71	96	99	89	88	69	49	47	56	56	0	59	54	36	43	33	53	31	27	29
19	87	103	99	18	25	83	100	75	93	95	86	85	73	46	43	52	29	59	0	51	32	16	37	50	28	31	30
20	106	98	94	69	74	102	95	70	42	47	38	34	68	41	39	48	48	54	51	0	28	35	32	1	23	27	26
21	87	80	76	50	57	84	76	52	70	72	63	61	50	23	11	20	29	34	32	28	0	14	14	26	5	8	7
22	71	87	83	34	41	68	84	59	77	79	70	69	57	30	27	37	13	43	16	35	16	0	21	34	12	16	14
23	92	77	73	55	62	89	76	38	74	77	68	66	36	28	25	34	34	33	37	32	14	21	0	31	9	6	7
24	104	97	93	68	75	101	94	69	43	46	36	35	67	40	37	46	46	53	50	1	26	34	31	0	22	25	24
25	83	75	71	46	53	79	72	47	65	68	58	57	45	18	16	25	24	31	28	23	5	12	9	22	0	4	3
27	86	71	67	49	57	83	71	44	69	71	62	61	42	22	19	28	28	27	31	27	8	14	6	25	4	0	1
28	85	72	69	48	55	82	69	45	68	70	61	59	43	21	18	27	27	29	30	26	7	14	7	24	3	1	0

TABLEAU 2: Matrice de distance de la différence symétrique reconstituée à partir de la représentation arborée (d'après les données de BARTLETT 1932); les sommets repérés par des numéros (de 17 à 28) correspondent aux sommets intérieurs de l'arbre.

