Additive Tree Representations of Verbatim Memory

Hervé ABDI[1]

Abstract

First a review of the literature shows that lexical forms can be used as dependent variable in the study of memory for texts. Next some distances are defined between texts seen as sets of words (the symmetric difference distance, the generalized symmetric difference distance, and the so-called "chi-square" distance). Additive trees representations are then constructed from the data of the classic Bartlett's "war of the ghosts" study. These results are used to detail the interpretation of a tree and to show how the vertices and the edges of the tree can be interpreted in terms of the original data matrix (i.e., a words by texts matrix) in terms of set-operations, and their generalization to sets with weighted elements (e.g., "fuzzy-sets"). It is also shown that these trees can be seen as a kind of regression analysis (where weights are estimated from data), or as a factorial analysis (where a score is decomposed into specific and common components). The main result that emerges from this analysis is a striking fidelity of the subjects to themselves in terms of memory for lexical forms. But this fidelity can be explained either by some memory for lexical forms or by the constancy of the subjects linguistic habits. To decide between these two hypotheses an experiment was run. Four subjects were asked to recall two different texts 1, 8 and 15 days after the learning phase. The hypotheses predict distinct tree structures for the results of this experiment (leading to a "confirmatory analysis"). The data and the tree used to represent them support the verbatim memory hypothesis.

---

[1] Laboratoire de Psychologie, Ancienne Faculté, 36 rue Chabot-Charny, 21000 Dijon, France.
I wish to thank Xuan Luong, Michel Fayol, Alice J. O'Toole, and David C. Rubin for help and comments.

DEFENSE & ILLUSTRATION OF VERBATIM MEMORY.

The existence of a verbatim memory for very short period of time is generally agreed upon (e.g., between others, Anderson & Paulson, 1977; for a review see Jarvella, 1979). By contrast, the abstractive aspect of long term memory is, in general, emphasized in the literature. Besides the classic (and hit) work of Bartlett (1932), this contention is supported by numerous observations and experiments among which the well-known studies of Sachs (1967, 1974) and Bransford *et al.* In these papers — as in almost all those reviewed here — the authors used a recognition paradigm, partly because of the greater sensitivity of recognition compared to recall. Actually, only a small number of papers used a recall paradigm (Cofer, 1941, 1943; Kay, 1955; Howe, 1970).

Sachs (1967) tested subject's ability to detect lexical modifications of sentences. When the modifications left unchanged the meaning of the statement, Sachs found that subjects detected the lexical modification for only a short period of time (some thirty seconds). Thus, her subjects did not distinguish "*he sent a letter about it to Galileo, the great Italian scientist*" from "*he sent Galileo, the great Italian scientist, a letter about it.*" Conversely, her subjects detected clearly the semantic modification of a statement. They were able to differentiate easily between "*he sent a letter about it to Galileo, the great Italian scientist*" and "*Galileo, the great Italian scientist, sent him a letter about it.*" In the same vein, Bransford *et al* (1972, cf. also Johnson *et al.*, 1973), in a classic experiment, showed that their subjects did not keep the precise formulation but the meaning of the sentence they heard. For example, the subjects of their experiment confounded the sentence "*three turtles rested* **on** *a floating log, and a fish swam beneath* **them**" and "*three turtles rested* **on** *a floating log, and a fish swam beneath* **it**" but did not confound the two sentences when the preposition *on* is replaced by *beside* (because the two sentences are not congruent in this latter case).

These results (and a lot of others) fully justify the stress put on the memory for gist as opposed to verbatim memory, and properly emphasize — in the tradition of Bartlett (1932) and Halbwach (1952) — the constructive aspect of memory. Nevertheless, some other studies point out that some information on the precise formulation may be accessible. For example, Kay (1955) read some short stories to his subjects and asked them on six different intervals (from week to week) to recall this text by writing. Kay (1955), then clearly observed some memory for verbatim information for the original text. the results suggested, also, that subjects may show an even greater fidelity to their first reproduction that to the original text, and so albeit a re-reading of the original after each reproduction. These results have been replicated and extended by Howe (1970, cf. also Cofer, 1941, 1943; Gauld & Stephenson, 1967), who added that the conclusions of Kay (1955) are valid for verbatim memory as well as for gist memory. Furthermore, Begg & Paivio (1969) have replicated the experiment of Sachs (1967), but they separated the concrete sentences from the abstract ones in the data analysis. They observed an interaction effect: the lexical modifications are detected — from memory — in the abstract sentences but not in the concrete ones. Although the independent variable is not clearly identified (due to possible confounded factors; cf. Johnson *et al*, 1972; Pedzeck & Royer, 1974; Kuiper & Paivio, 1972), this work clearly establishes the possibility of some verbatim memory.

Along the same line, Hayes-Roth & Hayes-Roth, after a fine re-analysis of Sachs' original data (1967), concluded that part of the lexical information is kept. Moreover, Reitman & Bower (1973), in a partial replication of the Bransford *et al.* (1972) experiment, used as experimental material letters and digits (instead of contextualized words) and found (contrary to Bransford *et al.*) that their subjects clearly discriminated the old items from the new distractors.

Although none of these data could defend the contention of a perfect "tape-memory," all support the existence (in certain cases) of some verbatim memory. This conclusion is reinforced by Hayes-Roth & Hayes-Roth (1977) who showed — in a somewhat artificial experimental context — that subjects can to distinguish sentence actually read from synonymous paraphrases.

In a more natural setting (i.e. memorization of a lecture in psychology), Kintsch & Bates (1977, see also Goolkasian *et al*, 1979) re-assert these results, and remarked that verbatim memory, still relatively precise after two days, decreases sharply after five days. Several studies confirm and complete the preceding ones (e.g., Abramovici, 1984; McKoon,1977; Hayes-Roth & Thorndike, 1979; Yerkovitch & Thordnike, 1981; Terry & Mason, 1982; and Keenan *et al.*, 1977). These latter authors showed — in a very clever and naturalistic study — that verbatim memory could be the result of incident learning. They measured the recognition of the exact formulation of sentences spontaneously uttered in an informal discussion in the psychology department in Denver. The persons taking part in the discussion did not know that their discussion was tape recorded, and that they would later be asked to participate in a recognition task. Nevertheless, they were able to recognize the precise formulation of almost half the sentences when they were humorous or implied some personal involvement (e.g., criticisms or praise). So, verbatim memory seems to depend upon the degree of interest or on the "perceived relevance or importance" of the stimulus (cf. Alba & Hasher, 1983; Cofer, 1941, 1943; Graesser & Mandler, 1975; Hasher & Griffin, 1978), in the same manner that the memory of "physical features" of the message does (cf. Fisher & Cuervo, 1983). Note, incidentally, that the memory of lexical information is apparently affected by some other factors, namely age (cf. Kail, 1979) and competence in reading (good readers being better memorizers than are bad readers, cf. Perfetti & Lesgold, 1979).

To continue to go outside the laboratory, and to reveal the existence of a "word for word" memory, suffice to ask American subjects to recite the monologue of Hamlet or the national anthem (as witness by the results of Rubin, 1977; cf. also Smith, 1933), or French subjects to recite the "fables de la Fontaine" or "La Marseillaise" (cf. also Neisser, 1982, and Hunter, 1984, for an "ecological" approach of the problem, and the links between verbatim memory and literacy). Even years after learning, we are able to recite a lot of poems, sing songs, or drone multiplication tables...

To sum up, the use of the lexical forms as a dependent variable is clearly justified by the evidence from the literature (although it would be foolish to pretend to reduce the study of memory for texts to verbatim memory).

CONSTRUCTION OF DISTANCE BETWEEN TEXTS

Recall the *motto* of this paper: a text is seen as a set of words, and with each word is associated its frequency of occurences in the text. Consequently, the texts can be compared by computing distances between sets. There is a great number of such distances (cf. Abdi, 1980; Abdi & Barthélemy, 1980), but, here, only three of them will be considered:

- the first distance (the so "so-called" chi-squared distance) because of its popularity in multidimensional analysis (at least on the french scene, cf. Benzécri, 1973; Hill, 1974; Nishisato, 1980; Greenacre, 1984),

- the other two distances (symmetric difference, and generalized symmetric difference), because of their close relations with the additive tree representation (cf. Tversky, 1977; Sattath & Tversky, 1977; Abdi *et al.*, 1984; Abdi, 1988).

We will begin with the most obvious of the three: the symmetric difference. A text is defined as a simple set or words (i.e., a word is present or absent). That is, its frequency in the text is neglected. This is equivalent to deciding the threshold of the word. Now, consider two texts, a straightforward measure of disagreement is simply the number of words belonging to *only one* text (i.e. the cardinality of the symmetric difference between the two sets of words). This distance is obviously a particular case of a more general one. If to each word is associated its frequency (if a word is absent in a text its frequency is zero), then the distance between two texts, say t say t' is computed by the formula:

$$d(t,t') = \sum_m |f_{m,-t} - f_{m,t'}|$$

With $f_{m,t}$ frequency of word m in text t. This distance is called the generalized symmetric difference distance. It could be described within the fuzzy sets theory framework as a "fuzzy" generalization of the symmetric difference distance. This distance is known also as the "city-block distance," or Hamming distance. It is clear that if the frequency of a word may is replaced by the values 0 (if absent) or 1 (if present) then this formula computes the symmetric difference distance. Note, also, that the generalized symmetric difference distance is in turn a particular case of a more general family of distances: the well-known Minkowski's distances.

Finally the so-called chi-square distance (cf. Benzécri, 1973; Greenacre, 1984) compares the texts by using their frequency profiles The chi-square distance is computed with the following formula:

$$d^2(t, t') = \sum_m \frac{\left(\frac{f_{m,t}}{f_{\cdot t}} - \frac{f_{m,t'}}{f_{\cdot t'}}\right)^2}{f_{m\cdot}}$$

With:

$$f_{\cdot t} = \sum_m f_{m,t}\,, \qquad f_{m\cdot} = \sum_t f_{m,t}$$

Although words scored by a strictly verbatim criterion are used here, it is obviously possible to use the same general method with propositions, or words recoded for meaning, or any semantic units provided an appropriate coding schema.

Comment on these distances

Obviously, these distances do not take into account the structure of the text. Consequently, they are not sensitive to any "re-ordering" of the text material. To caricature, these distances will find no difference between a text and any random permutation of the words of this text. Although it is obvious that subjects never give a random recollection of a text, it implies that these distances may be quite insensitive to some subtle differences. However, these distances are clearly sensitive to any change in surface formulation (which is the main point of this research).

All in all, the previous remarks suggest that these distances are reliable when they reveal a difference between texts, even if they may be less reliable when they reveal a similarity between texts.

	Orig	H1	H8	N1	N15	L1	L120	X180	P1	P15	P45	P105	P912	R1	R15	R45
Orig	0	158	156	106	103	113	153	138	158	143	138	142	144	103	112	127
H1		0	84	126	131	149	141	112	140	147	140	134	112	103	102	107
H8			0	120	127	147	135	116	136	129	130	128	108	101	102	113
N1				0	43	107	121	90	112	119	106	108	98	65	68	79
N15					0	110	126	101	131	132	113	121	101	74	75	80
L1						0	136	135	143	140	137	135	137	114	117	130
L120							0	113	137	142	129	127	117	106	103	106
X180								0	112	123	108	106	74	73	64	67
P1									0	89	92	76	104	105	100	109
P15										0	83	81	119	106	99	110
P45											0	70	104	91	92	105
P105												0	96	99	98	107
P912													0	67	60	61
R1														0	29	48
R15															0	31
R45																0

Table 1: Symmetric difference distance matrix obtained from Bartlett's data (1932).

ADDITIVE TREE REPRESENTATIONS

In the previous chapter, some distances between texts have been defined. By using these distances, it is possible to compute a distance matrix between a given set of texts. These data matrices, however, are not easy to "read". As a consequence it may be convenient to represent these matrices in a more "readable" way. The aim of the present paper is to represent texts as "leaves" on a tree (i.e., terminal vertices of the tree). Each edge of the tree has a definite length. The distance between two vertices of the tree is the length of the path on the tree between these two vertices. The tree distance is computed to approximate the original distance matrix (see the different papers in this volume, and Abdi, Barthélemy & Luong, 1984; Abdi, 1988). The quality of this approximation (i.e., the goodness of fit) is given by the squared correlation coefficient between the original data matrix and the tree distance matrix. This coefficient is usually denoted by τ. It can be interpreted as the proportion of variance of the original distance matrix explained by the tree model.[2]

SOME TREES IN THE WAR OF THE GHOSTS

In the mid thirties, Bartlett (1932) recited a story called the "war of the ghosts" to some subjects, and asked them to recall it on different occasions. As most of his subjects were friends, they were often asked to give their recollections on occasional meetings, and the interval of time separating two such meetings were not carrefully standardized. Bartlett named his method[3] the *repeated reproduction method*. As it is well known, Bartlett

[2] The measure of goodness of fit used in the MultiDimensional Scaling literature (the "Stress"), is adequate only for data measured on an ordinal scale. But, as both the original data matrix and the tree distance matrix are actually measured on an interval scale, a coefficient of correlation is adequate to assess the goodness of fit.

[3] inspired by Philippe (1987).

	Orig	H1	H8	N1	N15	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	88	85
H1		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			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
N1				0	43	101	118	93	111	113	104	103	91	64	61	70	46	77	18	69	50	34	55	68	46	49	48
N15					0	109	125	100	118	120	111	110	98	71	68	79	54	84	25	76	57	41	62	75	53	57	55
L1						0	151	1127	145	147	138	136	125	98	95	104	55	110	83	102	84	68	89	101	79	83	82
L120							0	114	137	139	130	129	112	90	87	97	96	98	100	95	76	84	76	94	72	71	69
X180								0	113	115	106	104	74	66	63	72	72	71	75	70	52	59	38	69	47	44	45
P1									0	89	80	76	110	84	81	90	90	96	93	42	70	77	74	43	65	69	68
P15										0	82	81	113	86	83	92	92	99	95	47	72	79	77	46	68	71	70
P45											0	72	103	77	74	83	83	89	86	38	63	70	68	35	57	61	59
P105												0	102	75	72	82	81	88	85	34	61	69	66	36	58	62	61
P912													0	64	61	70	70	69	73	68	50	57	36	67	45	42	43
R1														0	34	43	43	49	46	41	23	30	28	40	18	22	21
R15															0	31	40	47	43	39	11	27	25	37	16	19	18
R45																0	49	56	52	48	20	37	34	46	25	28	27
17																	0	56	29	48	29	13	34	46	24	28	27
18																		0	59	54	36	43	13	53	31	27	28
19																			0	51	32	16	37	50	28	31	30
20																				0	28	35	32	1	23	27	28
21																					0	16	14	26	5	8	7
22																						0	21	34	12	16	14
23																							0	31	9	7	6
24																								0	22	25	24
Med																									0	4	3
26																										0	1
27																											0

Table 2: Tree reconstitution of the symmetric difference matrix obtained from Bartlett's data (1932). The vertices labelled with numbers 17 to 27 are the interior vertices. *Med* is the median of the tree.

interpreted his results using the notion of *schema* abstraction. He put the stress on the importance of the constructive aspect of memory for meaningful material. In his classic book, Bartlell (1932) gave the reproductions obtained from six subjects. These data will be taken to introduce the tree analysis approach. Recall that the distances computed between the texts are simply distances between sets of words. The first analysis deals with the symmetric difference distance. The distance matrix and the tree distance matrix are given in Table 1 and Table 2. The tree is displayed in Figure 1.

Each "leaf" on the tree of Figure 1 represents a text identified by a letter followed by a number. The letter (*H*, *L*, *N*, *P*, *R*, *X*) stands for the subjects's name and the number indicates the number of days separating the first reading of the "war of the ghosts" and the reproduction. For example, *R*45 stands for the reproduction given by subject *R*, 45 days after the original reading. The "war of the ghost" original text is designated by *Orig*. Next to every vertex of the tree, the number in a circle is called the *eccentricity* of the vertex.[4]

Use and meaning of the "eccentricity"

To grasp the meaning of the eccentricity of a vertex, it will help to suppose that the original data matrix is already a tree distance matrix. Consider two texts, say *N*15 (eccentricity 53) and *N*1 (eccentricity 46). These two texts "merge" in a vertex with an eccentricity of 27. Call this vertex the *intersection vertex* of the vertices (cf. Abdi, 1988). This value means that the two texts have 27 words in common *excluding* the words common to all the texts.[5] The length of the edge from *N*15 to the intersection vertex represents the number of words that belong *only* to *N*15. This length can be obtained directly as the difference between the eccentricity of the two vertices (i.e., 53 − 27 = 26). The distance on the tree between vertices (say *N*1 and *N*15) could then be interpreted as the sum of the words belonging only to *N*1 and the words belonging only to *N*15. That defines the symmetric difference distance.

The interpretation of the distance on the tree can be generalized without difficulty to any pair or any *n*-uple of vertices. For example, *Orig* and *N*1 join into a vertex of eccentricity 12, and this is the number of their common words. The path length from this intersection vertex to *Orig* represents the number of words present in *Orig* and absent from *N*1. The

[4] *eccentricity* is to be taken here in its etymological sense of "far from a center."

[5] To take the words common to all the texts into account is equivalent to adding a (positive) constant to the eccentricity of each vertex. The general interpretation will remained unchanged.

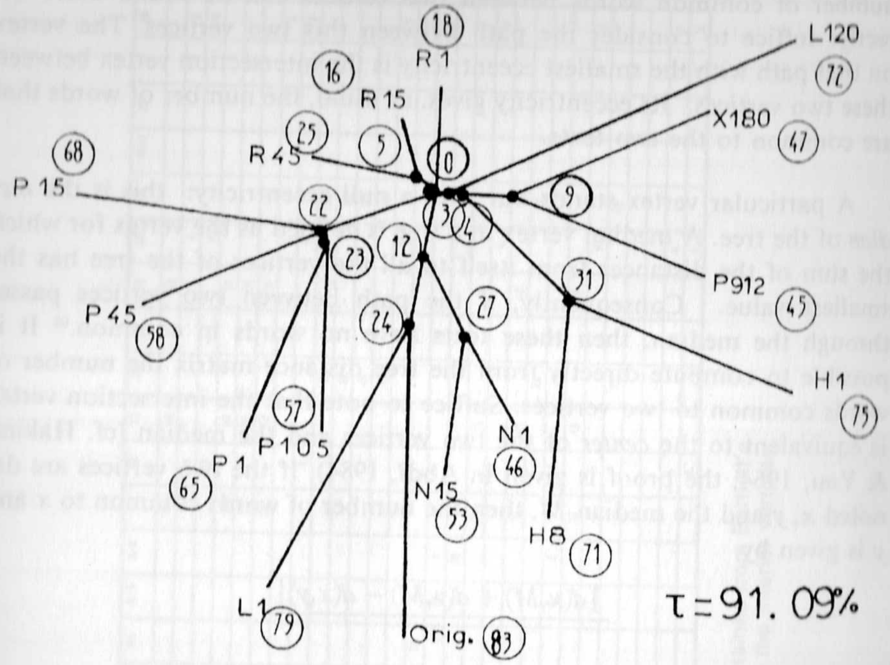

Figure 1. : Additive tree representation of the symmetric difference matrix computed from Bartlett's data (1932).

number of common words between two vertices can be found more directly: suffice to consider the path between this two vertices. The vertex on that path with the smallest eccentricity is the intersection vertex between these two vertices. Its eccentricity gives, as usual, the number of words that are common to the two texts.

A particular vertex stands out with a null eccentricity: this is the *median* of the tree. A median vertex in a tree is defined as the vertex for which the sum of the distances from itself to all the vertices of the tree has the smallest value. Consequently, if the path between two vertices passes through the median, then these texts have no words in common.[6] It is possible to compute directly from the tree distance matrix the number of words common to two vertices. Suffice to note that the intersection vertex is equivalent to the *center* of the two vertices and the median (cf. Hakimi & Yau, 1964; the proof is given in Abdi, 1984). If the two vertices are denoted x, y and the median M, then the number of words common to x and y is given by

$$\frac{|d(x,M) + d(y,M) - d(x,y)|}{2}$$

Accordingly, it is possible to reconstitute the table of common words. This is done in Table 3., where it can be seen, for example that *Orig* and *N1* have

$$\frac{83 + 46 - 105}{2} = 12$$

words in common.

Note that the sum of each column of Table 3 is an index of representativity of a vertex. This idea can be expressed by using a notion from the factorial analysis tradition, namely the term of *communality*. Indeed, the greater the total of a column, the greater is the number of words

---

[6] except the words common to all the text, cf. previous note.

	Orig	H1	H8	N1	N15	L1	L120	X180	P1	P15	P45	P105	P912	R1	R15	R45	17	18	19	20	21	22	23	24	Med	26	27	28
Orig	83			12	12	24											24		12			12						
H1		75	31				3	4					4					31					4				4	3
H8		31	71				3	4					4					31					4				4	3
N1	12			46	28	12											12		28			12						
N15	12			28	53	12											12		28			12						
L1	24			12	12	29											24		12			12						
L120		3	3				72	3					3					3					3				3	3
X180		4	4				3	47					9					4					9				4	3
P1									65	22	22	23								23				22				
P15									22	68	22	22								22				22				
P45									22	22	58	22								22				22				
P105									23	22	22	57								23				22				
P912		4	4				3	9					45					4					9				4	3
R1														18														
R15															16	5					5							
R45															5	25					5							
17	24			12	12	24											24		12			12						
18		31	31				3	4					4					31					4				4	3
19	12			28	28	12											12		28			12						
20									23	22	22	23								23				22				
21															5	5					5							
22	12			12	12	12											12		12			12						
23		4	4				3	9					9					4					9				4	3
24									22	22	22	22								22				22				
Med																												
26																												
27		4	4				3	4					4					4					4				4	3
28		3	3				3	3					3					3					3				3	3

Table 3: Tree reconstitution of the common words matrix obtained from Bartlett's data (1932). The vertices labelled with numbers 17 to 28 are the interior vertices. *Med* is the median of the tree.

shared by the text described by that column and the other texts. Consequently, the greater the communality of a text, the more representative this text is for the whole set of texts.

Texts	Orig	H1	H8	N1	N15	L1	L120	X180	P1	P15	P45	P105	P912	R1	R15	R45
Sums of the distances	1937	1845	1791	1415	1515	1983	1863	1503	1657	1692	1562	1541	1471	1135	1056	1214
Communality	48	42	42	52	52	48	12	20	67	66	66	67	20	0	5	5

Table 4: Sum of the distances to the leaves of the tree of Figure 1, and communality. The communality of a text is the number of words the text has in common with the other texts according to the tree model. Data from Bartlett (1932).

The notion of communality is to be contrasted with the notion of the representation of a set of objects by an exemplar that minimizes some kind of distance to the whole set of objects (e.g., a mean, or more adequately for a tree, a median). Suffice to compare, for each vertex of the tree, the communality and the sum of the distance to the other vertices. It is shown in Table 4 that these two indices do not agree.[7] Specifically, the criterion of "maximal communality" leads to the choice of of *P1* or *P105* for the most representative vertex; but the criterion of "minimal sums of distance" leads to the choice of *R45* (which is the text closest to the median of the tree).

Back to the results

The tree displayed in Figure 1 "explains" 91% of the variance of the original data distance matrix. Thus, the tree can be seen as a very good ap-

[7] Actually, It can be proved that these two indices agree only when the tree is a star.

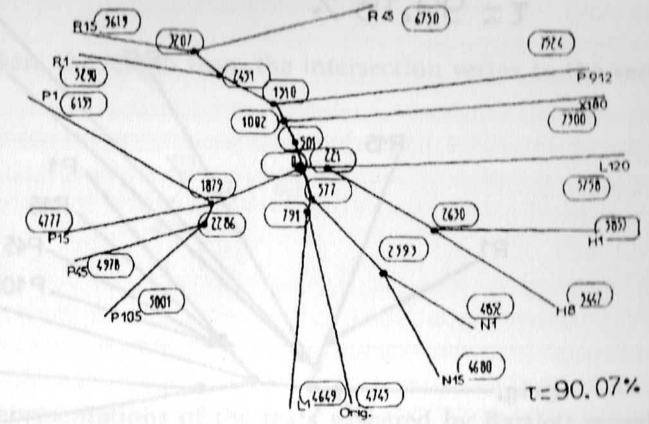

Figure 2. : Additive tree representation of the generalized symmetric difference distance matrix computed from the results of Bartlett's data (1932).

proximation of the original data. The fidelity of the subjects to themselves is very clear when looking at the tree. The different reproductions of a same subject are, in general, very close to each other. Nevertheless, two exceptions are noticeable:

- the first four reproductions of subject *P* (from *P1* to *P105*) are very well clustered. They constitute by themselves a "branch" of the tree. However, the reproduction given by the same subject two years and a half later (*P192*) is placed very far away from this branch.
- Subject *L* moves from the "maximal fidelity" for the reproduction given on the first day to the "maximal infidelity" for the reproduction given four months later.

The trees built from the two other distances lead to essentially the same conclusions as before (but note the striking goodness of fit index for the tree in Figure 3). Figure 2 will be used to introduce the rules of inter-

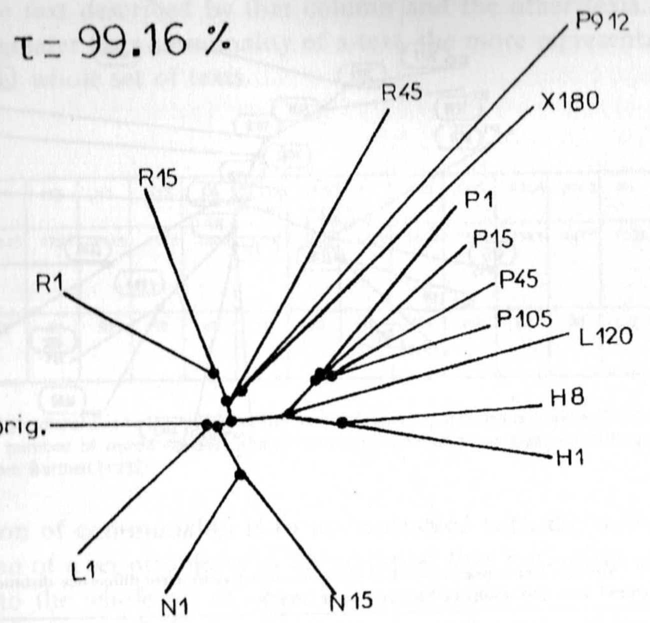

Figure 3. : Additive tree representation of the chi-square distance matrix computed from Bartlett (1932) repeated reproductions.

pretation of a tree when the underlying distance is the generalized symmetric difference distance. The eccentricity of the vertices keeps the same meaning as previously, but takes into account the fact that a word is weighted by its frequency. Thus the eccentricity of an "intersection vertex" gives the "generalized intersection" of the two vertices that merge to create it. More formally, if A and B are two texts, and if a_i (resp. b_i) denotes the weight of the i-th word in text A (resp. B), then the eccentricity of the intersection vertex is:

$$\sum_i \min\{a_i, b_i\}$$

This expression is equivalent to the fuzzy generalization of the set intersection.

By the same token, the length from the intersection vertex to the vertex A is obtained by:

$$\sum_{a_i > b_i} (a_i - b_i)$$

DISCUSSION

The additive tree representations of the texts gathered by Bartlett reveal a global fidelity of the subjects to themselves (but with some reserves). However, at least two interpretations can explain these results; either the subjects stay close to their first reproduction, or they keep their linguistic habits and consequently they tend to use the same vocabulary in the different reproductions. The first proposition corresponds to the existence of some verbatim memory (*via* the encoding filter of the first reading). The second proposition states that the apparent memory for lexical forms is an *artifact*.

Unfortunately, it is not possible to choose between these two hypotheses with the data from Bartlett (1932). Indeed, to do so at least two series of reproductions have to be given by each subject. The first hypothesis (fidelity to the first reproduction) predicts that the reproductions by the same subject of two different texts must be clustered according to the texts of origin. On the contrary, the hypothesis of the constancy of the linguistic habits predicts that the reproductions of one subject must be similar to each other independently of the text of origin. Moreover, if the subjects do have some kind of verbatim memory, then each set of reproductions coming from the same original text must appear as a "sub-tree" of the tree. To decide between these two rival hypotheses, an experiment was run.

BUTTERFLY & SCIENCE-FICTION

Method and Subjects

Four adult subjects (2 males and 2 females) participated to the experiment.[8] All subjects were native French speakers. They were not students in psychology. Subjects were asked to read two different texts and to reproduce them from memory by writing. Each subject was asked to recall each text one, eight, and fifteen days after the original encoding. The learning sessions (i.e., between learning the first and the second text) were separated by at least one month, and at most three months. The learning parts of the experiment took place in the same room as the recollections. The subjects were informed of the intention of obtaining the reproductions and of their timing. However, no mention of an interest for lexical forms memory was made. Each text was read twice. The subjects were not allowed to take notes. The subjects were given unlimited time to give their reproductions. The mispellings were corrected for the data analysis.

Material

The first text used was an adaptation of an indian fairy tale (*the woman with the butterfly*). The second text was adapted from a short story more or less inspired by Orwell's novel 1984 (*Mister Mead*). The original texts as well as the reproductions are available on request.

Results

The data were analyzed using the generalized symmetric difference distance, because of its close relation with additive trees. However, the same

[8] I would like to thank J. Barbieri, A. Champenois and M. Lacaze for their help in running the experiment.

general pattern of results emerged when the data analysis was conducted using different distances. Due to computer memory space shortage, only words with an absolute frequency greater than three have been kept for the final analysis. Basically, this procedure eliminates idiosyncrasies. These idiosyncrasies would express themselves by a lengthening of the "terminal branches" (i.e., the edges just before a leaf). As a consequence, the shape of the tree will remain unchanged by addition of these rare words to the corpus.

The results of the tree analysis are given in Figure 4. The leaves of the tree in Figure 4 are labelled as follows. The first letter stands for the subject's name (*A*, *C*, *M*, *P*). The second letter stands for the original text (*P* for "Papillon",[9] *M* for "Mead"). The number at the end indicates the interval in days (1, 8, 15) between the first reading and the reproduction.

Discussion

A simple glance at Figure 4 is sufficient to decide between the two rival hypotheses. The tree is clearly made up of two distinct sub-trees, and the different reproductions of a same subject are found on a same branch. In this way, the "linguistic consistency hypothesis" is eliminated. The fidelity of the subjects to themselves cannot be explained by the constancy of their vocabulary. Moreover, the fact that the different texts are clustered together is in close agreement with the hypothesis of a greater fidelity of the subjects to their first reproduction rather than to the original text (as pointed out previously by Kay, 1955; Howe, 1970). Of interest, also, is the clear tendency for the reproductions of a same text given by one subject to join in a same vertex, (i.e., to form a "crow-foot" configuration). This suggests the hypothesis of the existence of a "mnesic-trace" relatively permanent (at least for the short intervals of time studied here). This trace is then activated for each reproduction, and some "re-interpretation" of the

[9] Butterfly in french

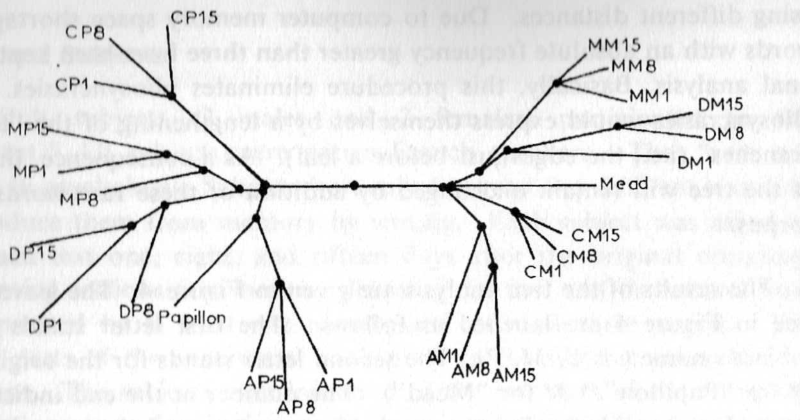

Figure 4. : Additive tree representations of the "repeated reproduction" of two texts ("Papillon" and "Mead") by four subjects. The first letter indicates subject's name. The second letter indicates the text learned. The number indicates the days separating learning from recall. The distance data matrix is computed using the generalized symmetric difference distance. The percentage of variance explained by the tree model is $\tau = 88.03$. See text for explanations.

trace takes place (e.g., the subject tries to fill in the blanks left in the trace, and that is done independently for each reproduction).

The central position of the original texts is worth noticing. Specifically, *Papillon* is the median leaf for the sub-tree composed of its reproductions; and *Mead* is almost the median in its sub-tree. This confirms that subjects memorize some verbatim information, and rules out a possible alternative interpretation for the observed configuration of the tree. Namely, the hypothesis that the theme of the story puts sufficient constraints on the vocabulary used by the subjects is not in agreement with this configuration.

Although the linguistic habit hypothesis is clearly eliminated in its strong form, some weak versions of this hypothesis, however, are compatible with the data. In particular, the co-existence of verbatim memory *and* linguistic habits is fully compatible with the observed configuration. Along the same line, the co-existence of verbatim memory and reconstructive processes as proposed by schema theories is compatible also with

the results (cf. Alba & Hasher, 1983, for a similar point of view). The tree in Figure 4 is also in agreement with some previous studies of Cofer (1941, 1943; see also Cofer & Musgrave, 1963), who proposed that subjects memorized prose material in a "telegraphic abbreviated form". The important or more informative words in a text constituting a "core version" of the text. Following this model, the subjects add some new (unimportant) words or variants to the core on each recall of the story. This hypothesis could be supported by an analysis where the words are qualified by their subjective importance in the text.

To summarize, the trees derived from that "micro-experiment" allows the elimination of the linguistic consistency hypothesis as the sole explanation for Bartlett's data. However some different hypotheses remain compatible with the observed tree structure. But all these hypotheses imply at least some form of verbatim memory.

The tree in Figure 4 gives a hierarchy of the sources of variance. The major source of variance are the texts, the second one are the subjects (this result may look trivial, but the converse order would have looked plausible as well). Interestingly enough, the time elapsed between the reproductions constitutes only a negligible source of variance. This counter intuitive result needs to be confirmed by further experiments.

CONCLUSION

The additive tree representations, obtained from distances computed on texts seen as set of words, give a good description of the data obtained from memory experiments. These trees can be interpreted in a meaningful manner. They can be interpreted either as a kind of regression analysis, or as a variety of factor analysis. A *post-hoc* justification of the tree-analysis of texts can be found in the impressive quality of the tree model as a descriptive tool. Specifically, the part of variance explained by the tree model varies from 88 to 99 percent. As suggested previously, this may be interpreted as an indication of a pronounced hierarchical structure in the set of memory reproductions. All in all, such a goodness a fit justifies in itself the use of additive tree representations as a *descriptive tool* for sets of texts recollections.

In sum, the trees analysis of the present data revealed a striking fidelity of the subjects to themselves (cf. Bartlett's data). When used in a confirmatory manner, the tree analysis helped to decide between rival hypotheses and confirmed the existence of a verbatim memory. Moreover, the tree analysis offers a hypothesis about memory trace and its use by subjects.

REFERENCES

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. (1986). La mémoire sémantique: une fille de l'intelligence artificielle et de la Psychologie. In G. Tiberghien G., C. Bonnet, J.M. Hoc (Eds.), *Psychologie et intelligence artificielle.* Bruxelles: Mardaga.

Abdi, H. (1988). Additive tree representations. In A. Dress, A.V. Haessler (Eds.), *Hierarchical Structures.* New York: Springer Verlag.

Abdi, H., Barthélemy, J.P. (1980). Analyse de contenu, analyse ordinale des données. *Bulletin de Psychologie de F.C.*, **4**, 1 – 18.

Abdi, H., Barthélemy, J.P., Luong, X. (1984). Tree representations of associative structures in semantic and episodic memory research. In E. Degreef E., J Van Buggenhaut, (Eds.), *Trends in Mathematical Psychology.* New York: Elsevier.

Abramovici, S. (1984). Lexical information in reading and memory. *Reading Research Quaterly*, **19**, 173 – 83.

Alba, J., Hasher, L. (1983). Is memory schematic? *Psychological Bulletin*, **31**, 203 – 131.

Anderson, J.R., Paulson, R. (1977). Representation and retention of verbatim information. *Journal of Verbal Learning & Verbal Behavior*, **16**, 439 – 451.

Bartlett, F.C. (1932). *Remembering.* Cambridge: C.U.P.

Begg, I., Paivio, A. (1969). Concreteness and imagery in sentence meaning. *Journal of Verbal Learning & Verbal Behavior*, **8**, 821 – 827.

Benzécri, 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 interpretative approach. *Cognitive Psychology*, **3**, 193 – 209.

Cofer, C.N. (1941). A comparison of logical and verbatim learning of prose passage of different lenghts. *American Journal of Psychology*, **54**, 1 – 20.

Cofer, C.N. (1943). An analysis of errors made in the learning of prose materials. *Journal of Experimental Psychology*, **32**, 399 – 410.

Cofer, C.N., Musgrave, B.S. (Eds.) (1963). *Verbal behavior and learning: Problems and processes.* New York: McGraw-Hill.

Christiaanson, R. (1980). Prose memory: forgetting rates for memory codes. *Journal of Experimental Psychology: Human learning & Memory*, **6**, 611 – 9.

Fisher, R.P., Cuervo, A. (1983). Memory for physical features of discourse as a function of their relevance. *Journal of Experimental Psychology: Learning Memory & Cognition*, **9**, 130 – 8.

Gauld, A., Stephenson, G.M. (1967). Some experiments relating to Bartlett's Theory of remembering. *British Journal of Psychology*, **58**, 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 & Memory*, **1**, 238 – 48.

Greenacre, M.J. (1984). *Theory and applications of correspondence analysis.* London: Academic Press.

Hakimi, S.L., Yau, S.S. (1964). Distance matrix of a graph and its realizability. *Quaterly of Applied Mathematics*, **22**, 305 – 7.

Halbwachs, M. (1952). *Les cadres sociaux de la mémoire*. Paris: P.U.F.

Hasher, L., Griffin, M. (1978). Reproductive and reconstructive processes in forgetting. *Journal of Experimental Psychology: Human Learning & Memory*, **4**, 318 – 330.

Hayes-Roth, B., Hayes-Roth, F. (1977). The prominence of lexical information in memory representation of meaning. *Journal of Verbal Learning & Verbal Behavior*, **16**, 119 – 36.

Hayes-Roth, G., Thorndike, P. (1979). Integration of knowledge from text. *Journal of Verbal Learning & Verbal Behavior*, **18**, 91 – 108.

Hill, M.O. (1974). Correspondence analysis, a neglected multivariate method. *Applied Statistics*, **23**, 237 – 49.

Howe, M.J.A. (1970). Repeated presentations and recall of meaningful prose. *Journal of Educational Psychology*, **61**, 214 – 9.

Hunter, I.M.L. (1984). Lenghty Verbatim recall (LVR) and the mythical gift of tape recorder memory. In K.J.M. Lagerspetz , P. Niemi, (Eds.), *Psychology in the 1990's*. New York: Elsevier.

Hunter, I.M.L. (1985). Lengthy verbatim recall: the role of text. In A.W. Ellis (Ed.). *Progress in the psychology of language*. Hillsdale: Erlbaum.

Jarvella, R.J. (1979). Immediate memory and discourse processing. In G.H. 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 & 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 & Verbal Behavior*, **16**, 549 – 60.

Kintsch, W., Bates, E. (1977). Recognition memory for statements from a classroom lecture. *Journal of Experimental Psychology: Human Learning & Memory*, **3**, 150 – 9.

Kruskal, J.B., Wish, M. (1978). *Multidimensional 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: Multigraphed.

McKoon, G. (1977). Organisation of Information in text memory. *Journal of Verbal Learning and Verbal Behavior*, **16**, 247 – 260.

Neisser, U. (1982). *Memory observed: remembering in natural contexts*. San Francisco: Freeman.

Nishisato, S. (1980). *Analysis of categorical data: dual scaling and its applications*. Toronto: U.T.P.

Pedzek, K., Royer, J.M. (1974). The role of comprehension in learning concrete and abstract sentences. *Journal of Verbal Learning & Verbal Behavior*, **13**, 551 – 8.

Perfetti, C.A., Lesgold, A.M. (1979). Coding and comprehension in skilled reading and implication for reading instruction. In L.R. Resnick, P.A. Weaver (Eds.), *Theory and practice of early reading (Vol.1)*. Hillsdale: Erlbaum.

Philippe, M. (1897). Sur les transformations de nos images mentales. *Revue Philosophique*, **LII**, 481 – 93.

Reitman, J.S., Bower, G.H. (1973). Storage and later recognition of exemplars of concepts. *Cognitive Psychology*, **4**, 194 – 206.

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 & Psychophysics*, **2**, 437 – 442.

Sachs, J.S. (1974). Memory in reading and listening to discourse. *Memory & Cognition*, **2**, 95 – 100.

Sattath, S., Tversky, A. (1977). Additive similarity trees. *Psychometrika*, **42**, 319 – 45.

Smith, M.E. (1935). Delayed recall of previously memorized material after twenty years. *Journal of Genetic Psychology*, **47**, 477 – 481.

Steiner, B.A., Allen, G.A. (1983). Verbatim recall of ambiguous sentences. *Psychological Reports*, **52**, 711 – 8.

Terry, W.S. (1983). Recognition of sentences from text after massed versus spaced reading. *Journal of General Psychology*, **109**, 67 – 71.

Terry, W.S., Mason, H.M. (1982). Recognition of sentences from prose. *Bulletin of the Psychonomic Society*, **19**, 7 – 10.

Thorndike, D.W., Yekovich, F.R. (1980). A critique of schema-based theories of human story memory. *Poetics*, **9**, 23 – 49.

Tversky, A. (1977). Features of similarity. *Psychological Review*, **84**, 327 – 352.

Tversky, A., Gati, I. (1978). Studies in similarity. In E. Rosch, B.B. Llyod, (Eds.), *Cognition and categorisation*. Hillsdale: Erlbaum.

Yekovich, F.R., Thorndike, P.W. (1981). An evaluation of alternative functional models of narrative schemata. *Journal of Verbal Learning & Verbal Behavior*, **20**, 454 – 69.

The Use of Ordered Trees for Uncovering Strategies in Free-Recall

Stephen C. Hirtle* Edward J. Crawley**

ABSTRACT

Complex strategies in free-recall tasks are uncovered using a technique based on the pairwise comparison of recall strings. The technique is used to parse recall trials into distinct sets, thus identifying trials in which the subject used a distinct strategy. The parsing is based on an index of dissimilarity between all pairs of recall strings, where the index is based on the structure of ordered trees. Using this technique, it was possible to identify multiple outliers, circular patterns of recalls, and multiple strategies. The implications of multiple recall strategies on free-recall experiments are discussed.

The structure of semantic memory has been studied using a variety of cluster based models (e.g., Friendly, 1977; Reitman & Rueter, 1980; Shepard & Arabie, 1979). Many of these techniques are based on building tree models of symmetric distance matrices of similarity data (e.g., Sattath & Tversky, 1977; Shepard & Arabie, 1979). These methods can result in either a single tree (Sattath & Tversky, 1977), or multiple trees (Carroll & Pruzansky, 1986, 1980). However, methods based on distance

* University of Pittsburgh

** State University of New York at Albany

The authors would like to thank to Judith Reitman Olson, John Jonides, and James Jaccard for comments and suggestions during several stages of this project and Duckhyun Kim for assistance with the data collection. This work was supported by NSF grant BNS8617732 to the first author. Portions of this work were presented at the annual meeting of the Society for Mathematical Psychology in Berkeley, CA, August 1987.

Correspondence concerning the article should be addressed to Stephen C. Hirtle, Interdisciplinary Department of Information Science, 752 LIS Building, University of Pittsburgh, Pittsburgh, PA 15260, USA.