

CORRESPONDENCE ANALYSIS IN A STUDY OF APHASIC PATIENTS

Sergio Camiz

Sapienza Università di Roma
Dip. di Matematica - Piazzale Aldo Moro 5, I - 00185 Roma Italia
sergio@camiz.net

Gastão Coelho Gomes

UFRJ
DME-IM-UFRJ – C.P. 68530 – 21945-970 Rio de Janeiro – RJ – Brasil
gastão@im.ufrj.br

Fernanda Duarte Senna

UFRJ
PG-Linguística – Faculdade de Letras 21945-970 Rio de Janeiro – RJ – Brasil
fonofernandasenna@gmail.com

Christina Abreu Gomes

UFRJ
PG-Linguística – Faculdade de Letras 21945-970 Rio de Janeiro – RJ – Brasil
christina-gomes@uol.com.br

RESUMO

Dados de produção de palavras de pacientes afásicos obtidos a partir de testes de nomeação e de repetição foram avaliados em função de um índice de similaridade de segmentos e sílabas compartilhadas entre alvo e substituição. Uma amostra de 2106 observações foi submetida a uma Análise de Correspondência Múltipla (*ACM*) que revelou três agrupamentos de resultados nos mapas dos três componentes principais, a saber: ruim, intermediário e ótimo, sendo os extremos altamente concentrados em relação ao fator um no mapa dos dois primeiros fatores, e os intermediários opostos aos outros dois tipos no terceiro fator, de acordo com um padrão crescente de resposta, o que nos levou a um estudo específico dos 304 resultados intermediários utilizando o *ACM* específico para a amostra reduzida. Os resultados desta análise proporcionaram um entendimento do padrão de comportamento dos afásicos em relação aos índices de similaridade propostos.

PALAVRAS CHAVE. Análise de Correspondências, Linguística, Afasia (*EST*)

ABSTRACT

Production data of words obtained from aphasic patients were obtained through naming and repetition tests and were evaluated according to indexes of segmental and syllabic similarities shared between target and substitution. In an exploratory analysis they have been submitted to Multiple Correspondence Analysis (*MCA*) that revealed three groups of units well separated on the space spanned by the first three factors: bad, intermediate, and good answers, the extreme ones being highly concentrated and opposed on the first factor. As the intermediate answers were opposed to the other two on the third factor, but rather confused, we decided to apply a *MCA* to a reduced sample limited to them. The results of this second analysis provided a better understanding of the pattern observed in the behavior of aphasic patients in relation to the proposed indexes of similarity.

KEYWORDS. Correspondence Analysis, Linguistics, Aphasia (*STA*)

Introduction

The failure in word retrieval is a pervasive characteristic of aphasic patients. Unsuccessful attempts result in substitutions that correspond to phonological changes in the target, neologisms that retain some phonological similarities or with little relationship with the target, semantic substitution, morphological substitution and non-related substitutions. There are some evidences that there is a relationship between the amount of error type and the depth of the injury (Dell *et al.*, 1997). Studies have shown the degree of phonological overlapping between target and error observed in different types of aphasia (see Bose *et al.*, 2007). This investigation evaluates two indexes of phonological overlapping between target and error by detecting the degree of segmental and syllabic similarities, the first taking into account the linear order of these phonological units. The purpose is to capture the phonological complexity of substitutions in accessing phonological information in the lexicon. We consider for this purpose a lexicon as proposed by the Usage-based Models, organized as a network of lexical relationships based on phonetic and semantic similarities, and phonological grammar as a ladder of levels emerging from the word-forms stored in the lexicon (Pierrehumbert, 2003). Based on this model, the phonological distance between target and error will show the information kept from the activated target which can be seen as an indication of the mechanism used by the patient in the retrieval.

The *Index of Segmental Similarity (IPSseg)* considers the number of segments shared between target and substitution (N_{seg}), considering their linear order in the target, the target size (TS), and the error size (ES). The formula $IPS_{seg} = N_{seg} * 2 / (TS + ES)$ returns values that range between 0 and 1. This index is based on the *Index of Phonological Overlapping (IPO)*, a measure of phonological relatedness, proposed by Bose *et al.* (2007) to examine phonological similarities between target and neologism in jargon aphasia, by observing the segments regardless of their position. The *IPSseg* differs from *IPO* exactly by considering the linear order of the segments. The definition of *IPSseg* handles the linearity by attributing 1 to the shared segments that occur in the same position in the target and 0.25 to those that occur in a different position. A similar equation is proposed for abstract syllable shape using the same criteria of linear position. The equation for the *Index of Syllabic Similarity* is $IPS_{sy} = N_{sy} * 2 / (TS + ES)$. In this case, the size corresponds to the number of syllables of both target and error.

The data were obtained through a 70-item test of naming and repetition applied to 10 fluent and 5 non-fluent aphasic individuals at the Federal University of Rio de Janeiro Hospital. The responses were digitally recorded and later phonetically transcribed for analysis. For the purpose to establish the Index of Segmental Similarity and the Index of Syllabic Similarity for each substitution, it was considered the first attempt in naming and repetition of each picture, excluded semantic and morphological substitutions, a total of 320 substitutions.

2 Methodology

The data analysis procedure adopted to extract information from our data set follows the pathway proposed by Camiz (2001). We started by a typically exploratory analysis technique, since, instead of applying immediately a model to the data, we preferred to investigate the table structure, in the spirit of Benzécri (1973-82) that claims that “*the models should follow the data, not the inverse*” (quoted by Greenacre and Blasius, 2006). So, we submitted our qualitative multiple character data table to *Multiple Correspondence Analysis* (MCA, Benzécri et coll., 1973-82; Greenacre, 1984; Langrand and Pinzón, 2009). It is a generalization of *Simple Correspondence Analysis* (SCA, ibid.) to the case of several categorical variables that was initially developed by Guttman (1941) as *principal components of scale*. Since then, it was rediscovered many times, as *factorial analysis of qualitative data* (Burt, 1950), *second method of quantification* (Hayashi, 1956), *multiple correspondence analysis* (Benzécri, 1973-82), and *homogeneity analysis* (Gifi, 1981), according to the different facets and computational methods adopted, but eventually leading to the same technique. The 10 qualitative variables used, in MCA, are described in the next section, table 3.

We remind here briefly that the well-known exploratory multidimensional scaling techniques - *Principal Components Analysis* (PCA, Benzécri, 1973-82; Jolliffe, 2002; Langrand and Pinzón, 2009), *SCA*, and *MCA* - are all based on the so-called *singular value decomposition* (SVD, Eckart and Young, 1936) and on these authors' theorem. SVD states that any real matrix A may be decomposed as $A = UA^{1/2}V'$, with A a real quasi-diagonal matrix (that is a matrix with all off-diagonal elements equal to zero) and two real orthogonal matrices U and V (that is such that $U'U = I$ and $V'V = I$). Such SVD is unique, with U a matrix of eigenvectors of AA' , V one of $A'A$ and A the diagonal matrix whose elements are the corresponding eigenvalues of both. Thus, the elements of A may be reconstructed by the formula $x_{ij} = \sum \lambda_\alpha^{1/2} u_{i\alpha} v_{j\alpha}$. The most interesting feature of SVD results from the Eckart and Young's (1936) statement that, once the eigenelements are sorted in decreasing order, the reconstruction limited to the first r eigenelements is the best r -dimensional one in the sense of least-squares:

$x_{ij} \approx \sum_{\alpha=1}^r \lambda_\alpha^{1/2} u_{i\alpha} v_{j\alpha}$. As the total inertia of the data table is given by the sum of the squared singular values, it results that the share of total inertia explained by the r -dimensional solution is given by $\sum_{\alpha=1}^r \lambda_\alpha / \sum_{\alpha} \lambda_\alpha$.

Indeed, both *SCA* and *MCA* are currently based on the so-called *generalized singular value decomposition* (GSVD, Greenacre, 1984; Abdi, 2007), that allow to set conditions to both rows and columns of A through two positive definite square matrices. If M and W are such matrices, thus A is decomposed as $A = \tilde{U}\tilde{A}^{1/2}\tilde{V}'$ with the constraints $\tilde{U}'M\tilde{U} = I$ and $\tilde{V}'W\tilde{V} = I$. Such generalization is necessary to cope with the methods' special metrics, the chi-square, that leads, for *SCA*, to the decomposition of the chi-square statistics. GSVD is obtained from the matrix $\tilde{A} = M^{1/2}AW^{1/2}$ by performing the SVD of $\tilde{A} = UA^{1/2}V'$, and then setting $\tilde{U} = M^{-1/2}U$ and $\tilde{V} = W^{-1/2}V$ to get the desired solution. In the case of *SCA* of a contingency table $F = f..P$ crossing a m -levels character by row with a q -levels one by column, with $f..$ the table grand total and P a $m \times q$ matrix of probabilities (that $\sum \sum p_{mq} = 1$), it results in defining two diagonal matrices D_m and D_q containing the marginal row- and column-profiles respectively. Then the GSVD of F results by solving $P = \Phi A^{1/2} \Psi$, with $\Phi' D_m^{-1} \Phi = I$ and

$\Psi' D_q^{-1} \Psi = I$ or, that is the same, $\Phi' \Phi = D_m$ and $\Psi' \Psi = D_q$ respectively. As it may be proved that the highest singular value equals 1, after some manipulation the reconstruction formula of F results from that of P as $f_{ij} = f_{..} p_{ij} = n_{..} p_{i.} p_{.j} \left(1 + \sum_{\alpha=1}^{\min(m,q)-1} \lambda_{\alpha}^{-1/2} \Phi_{i\alpha} \Psi_{j\alpha} \right)$.

Prior to generalize *SCA* to more than two nominal characters, it is interesting to observe that *SCA* can be seen under (at least) two other points of view, all related among them (Baccini, 1984; Greenacre, 1984; Langrand and Pinzón, 2009). The first is the *SCA* of the $f_{..} \times (m+q)$ indicator matrix $Z = (Z_1 | Z_2)$, such that $F = Z_1' Z_2$, where the first m columns correspond to the presence-absence in the $f_{..}$ units of the m levels of the character represented by row in the table and the second q correspond to the presence-absence of the q levels of the character represented by columns. The second is the *SCA* of the so-called Burt's matrix

$B = Z'Z$, a block-square matrix, that contains four blocks $B = \begin{pmatrix} Z_1' Z_1 & Z_1' Z_2 \\ Z_2' Z_1 & Z_2' Z_2 \end{pmatrix}$ two of which equal to either F or its transpose and the two other diagonal are D_m and D_q . It can be also

shown (*ibid.*) the relation $v_{\alpha} = \frac{1 \pm \sqrt{\lambda_{\alpha}}}{2}$ that holds among the singular values of Z and those of the *SCA* of the contingency table and the relation $v_{\alpha}^2 = \mu_{\alpha}$ that holds among the singular values of Z and those of B . As a consequence, to the $\lambda_{\alpha} = 0$ of *SCA* correspond $v_{\alpha} = \frac{1}{2}$ of Z and $\mu_{\alpha} = \frac{1}{4}$ of B , whereas to each of the other λ s two values correspond, one of which larger and the other smaller than $\frac{1}{2}$ and $\frac{1}{4}$ respectively.

The *MCA* of a set of Q nominal characters corresponds to the *SCA* of either the indicator or the Burt's matrix crossing all characters. We drop here other definitions and formulas, that may be found in the quoted works. Suffice here to remind that the total inertia of Z is $I_Z = \frac{J-Q}{Q}$,

where Q is the number of variables and J the total number of levels, that is $J = \sum_{i=1}^Q l_i$ with l_i the number of levels of the i -th character and that the singular vectors of both Z and B are the same, whereas the B 's singular values are the squares of Z 's. Analogous statements concerning the inflation of the singular values of *MCA* lead to admit to limit attention in *MCA* only to those larger than their mean, that is $v_{\alpha} > \frac{1}{Q}$. This argument is discussed in detail by both Benzécri (1979) and Greenacre (1988, 2006), that suggest, in order to getting a measure of relative importance of each factor, to re-evaluate the singular values larger than their mean (that equals $1/Q$) according to the formula

$$\rho(v_{\alpha}) = \left(\frac{Q}{Q-1} \right)^2 \left(v_{\alpha} - \frac{1}{Q} \right)^2$$

Benzécri attributes the inflation to the arbitrary number of levels of a nominal character, so that he suggests considering as total inertia their sum and sets as percentage of explained inertia the ratio $\frac{\rho(v_{\alpha})}{\sum_{\alpha} \rho(v_{\alpha})}$. This results in a dramatic re-evaluation of the relative importance of the first

eigenvalues. More conservative, Greenacre suggests to divide them by the total off-diagonal inertia of the table that is the sum of squared (non-re-evaluated) eigenvalues minus the diagonal inertia, that is $\frac{Q}{Q-1} \left(I_Z - \frac{J-Q}{Q} \right)$. Experiments show that the Greenacre's reevaluation is always limited to a share of the total inertia of Burt's table even by taking into account all the singular values larger than the mean. Even though these arguments may be useful to decide the importance of each factor, they are not able to identify the true dimension of the solution.

For this task Ben Ammou and Saporta (2003) suggest to estimate the significance of the singular values of *MCA* according to their distribution. If the characters are independent,

$$\sum_{\beta=1}^{J-Q} v_{\beta} = \frac{J-Q}{Q} \quad \text{and} \quad \sum_{\beta=1}^{J-Q} v_{\beta}^2 = \frac{J-Q}{Q^2} + \frac{\sum_{i \neq j} \Phi_{ij}^2}{Q^2} \quad \text{with} \quad x_{..} \Phi_{ij}^2 \approx \chi^2_{(l_i-1)(l_j-1)}, \quad \text{thus}$$

$$E[x_{..} \Phi_{ij}^2] = (l_i-1)(l_j-1), \quad \text{so that the expected variance of the singular values (around their mean } 1/Q) \text{ is } \sigma^2 = E[S_{\lambda}^2] = \frac{1}{Q^2 (J-Q)} \sum_{i \neq j} (l_i-1)(l_j-1).$$

This allows to define a confidence interval for the mean, by assuming that the interval $\frac{1}{Q} \pm 2\sigma$ should contain about 95% of the random values. Indeed, since the kurtosis of the set of eigenvalues is lower than for a normal distribution, the actual proportion is larger than 95%.

In order to interpret the factors thus extracted, it is customary consider the *contributions* of the levels to the axes themselves: it is the share of inertia of each of them along each axis, so that a high percentage represents a strong contribution, thus the importance of each level for its formation.

3 The data

The data table concerning the aphasia problem was composed by 2106 examinations and 10 nominal characters. Indeed, both Degrees of Similarity have been recoded as qualitative indexes with six-levels, the first and last of which correspond to null or perfect response, and the intermediate result by cutting the continuous characters according to Fisher's (1958) algorithm, that minimizes the within-groups variance. In Table 1 the values of the cut-points for both Indexes are shown, based on partitions from 2 through 5 groups: for both indexes we chose to partition the intermediate values in four classes.

Table 1. The cut-points for the Indexes of Similarity according to Fisher's (1958) algorithm.

| Number of groups | Degree of Semantic Similarity (304 units) | | | | | Degree of Syllabic Similarity (188 units) | | | | |
|------------------|---|------------|-------|-------|-------|---|------------|-------|-------|-------|
| | Criterion | Cut-points | | | | Criterion | Cut-points | | | |
| 2 | 4,362 | 0,545 | | | | 1,3733 | 0,600 | | | |
| 3 | 2,063 | 0,385 | 0,684 | | | 0,6420 | 0,545 | 0,750 | | |
| 4 | 1,153 | 0,286 | 0,542 | 0,762 | | 0,3116 | 0,417 | 0,571 | 0,750 | |
| 5 | 0,706 | 0,250 | 0,462 | 0,636 | 0,783 | 0,1918 | 0,417 | 0,571 | 0,667 | 0,800 |

The 2106×10 data table was submitted to *MCA*. Its eigenstructure results in 43 non-zero eigenvalues, 19 of which are above the average (0.1) but only 12 significant according to Ben Ammou and Saporta (2003) criterion, that indicates the upper limit of confidence interval for random tables eigenvalues as 0.12738 for 5% significance level. In this paper, we limit our attention to the first three dimensions, that deserve some immediate interest. The eigenvalues,

with the inertia and both Benzécri's and Greenacre's re-evaluations are reported in Table 2.

Table 2. Eigenvalues, inertia and re-evaluated inertia of the first 3 eigenvalues of *MCA*.

| N. | Eigenvalues | | | Inertia | Benzécri's re-evaluation | | Greenacre's re-evaluation | |
|----|-------------|------|--------|----------|--------------------------|--------|---------------------------|---------|
| | Eigen. | % | Cum. % | | % | Cum. % | % | Cum. % |
| 1 | 0,4237 | 9,42 | 9,42 | 0,129390 | 45,24 | 45,24 | 34,6430 | 34,6430 |
| 2 | 0,3573 | 7,94 | 17,36 | 0,081754 | 28,59 | 73,83 | 21,8890 | 56,5320 |
| 3 | 0,2177 | 4,84 | 22,20 | 0,017116 | 5,98 | 79,82 | 4,5826 | 61,1146 |

Indeed, it may be noticed an important gap between the two first eigenvalues accounted for a very large amount of cumulated explained inertia (73.83 and 56.53, according to each criterion) and the third. Nevertheless, it revealed of immediate interest for the analysis. It may be observed that the 61.11% of Greenacre's total re-evaluation is referenced to the total off-diagonal inertia of the Burt's table. Analogous consideration concerning the reduced *MCA* led us to limit our attention to the first two main dimensions, whose cumulate re-evaluated variance according to either criterion are 55 and 39% respectively. In Table 3, the coordinates of all levels of the 10 considered characters on the first three factors of both *MCAs* are reported.

Observing the first columns of Table 3, we can state that to the first factor contribute both Index of Syllabic and Segmental Similarity, Type of response and Position in the Word, summarizing nearly 90% of the total. Looking at Figure 1, in which the Indexes trajectories are shown on the first factor plane with the Diagnostic and the Type of test, it is evident that on the first factor there is a very strong opposition between the samples with null response on the left and those with full response on the right. The same variables contribute to the second factor but this time with a major importance of the intermediate values that are situated in the upper side, in opposition to the two extremes. This is the reason that led us to take into account the third axis, since we hoped to have a better insight in this part of the data structure.

In Figure 2, the type of response, the position of words and the length in syllables and segments are reported on the same two factors. The pattern is really alike the one of Figure 1, so that we can consider a very strong agreement of the character levels, in particular in what concerns the two extreme situations of null and perfect response.

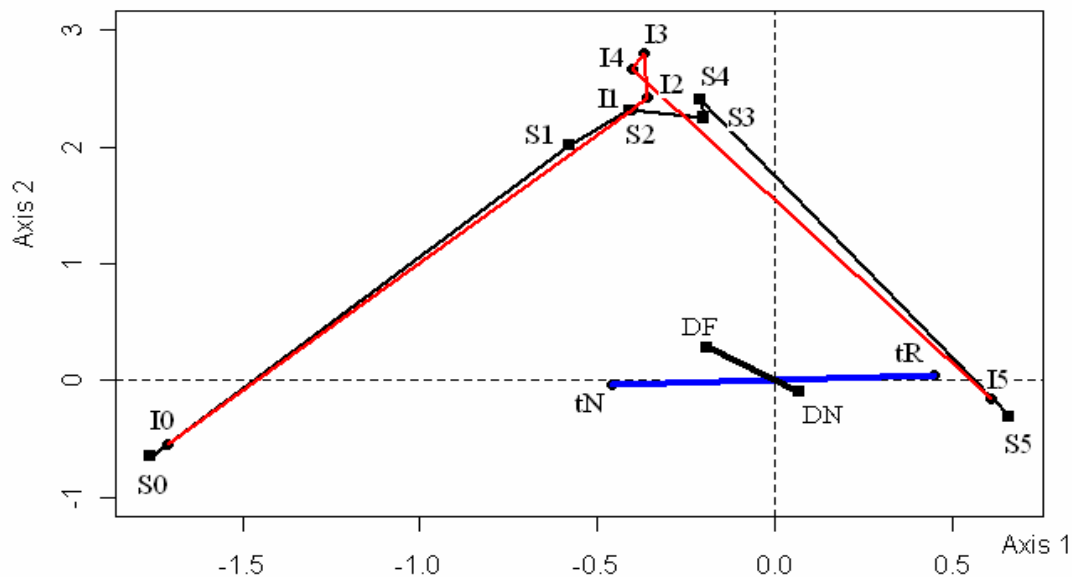


Figure 1. Indexes of similarity, diagnostic and type of test on the complete *MCA* plane spanned by the first two factors.

Table 3. Complete and reduced MCAs: the levels' coordinates on the first three factors.

| Index of Syllabic Similarity | Label | AXIS 1 | AXIS 2 | AXIS 3 | AXIS 1 | AXIS 2 | AXIS 3 |
|-------------------------------|-------|--------|--------|--------|--------|--------|--------|
| Sil Null | I0 | 1,71 | -0,55 | -0,04 | 1,51 | 0,11 | -0,04 |
| Sil < .418 | I1 | 0,40 | 2,31 | -2,64 | 0,60 | -0,03 | -2,64 |
| Sil .418 - .571 | I2 | 0,36 | 2,43 | -1,08 | 0,43 | 0,54 | -1,08 |
| Sil .572 - .750 | I3 | 0,37 | 2,80 | 0,05 | -0,16 | -0,28 | 0,05 |
| Sil > .750 | I4 | 0,40 | 2,66 | 0,58 | -0,34 | -0,10 | 0,58 |
| Sil Complete | I5 | -0,61 | -0,16 | 0,07 | -0,42 | -0,07 | 0,07 |
| Index of Segmental Similarity | Label | AXIS 1 | AXIS 2 | AXIS 3 | AXIS 1 | AXIS 2 | AXIS 3 |
| Seg Null | S0 | 1,76 | -0,65 | 0,02 | - | - | - |
| Seg < .286 | S1 | 0,58 | 2,01 | -2,95 | 1,35 | -0,26 | 0,09 |
| Seg .287 - .542 | S2 | 0,41 | 2,31 | -1,78 | 0,51 | 0,13 | 0,43 |
| Seg .543 - .762 | S3 | 0,20 | 2,25 | -0,08 | -0,01 | 0,07 | -0,19 |
| Seg > .762 | S4 | 0,21 | 2,40 | 1,50 | -0,69 | -0,04 | -0,13 |
| Seg Complete | S5 | -0,66 | -0,31 | 0,03 | - | - | - |
| Type of Test | Label | AXIS 1 | AXIS 2 | AXIS 3 | AXIS 1 | AXIS 2 | AXIS 3 |
| Naming | tN | 0,46 | -0,04 | -0,15 | 0,36 | 0,14 | 0,06 |
| Repetition | tR | -0,45 | 0,04 | 0,15 | -0,41 | -0,16 | -0,07 |
| Diagnostic | Label | AXIS 1 | AXIS 2 | AXIS 3 | AXIS 1 | AXIS 2 | AXIS 3 |
| Non fluent | Dn | 0,19 | 0,28 | 0,29 | -0,01 | -0,85 | -0,28 |
| Fluent | Df | -0,07 | -0,10 | -0,10 | 0,01 | 0,64 | 0,21 |
| Time of Injury | Label | AXIS 1 | AXIS 2 | AXIS 3 | AXIS 1 | AXIS 2 | AXIS 3 |
| < 1 year | L1 | -0,04 | 0,10 | -0,62 | 0,55 | 0,63 | 0,31 |
| 1 - 2 years | L2 | 0,05 | 0,09 | -0,32 | 0,42 | -0,55 | -0,15 |
| 2 - 5 years | L3 | -0,32 | -0,10 | 0,18 | -0,73 | 1,69 | 0,02 |
| > 5 years | L4 | 0,21 | -0,10 | 0,71 | -0,66 | -0,81 | -0,09 |
| Type of response | Label | AXIS 1 | AXIS 2 | AXIS 3 | AXIS 1 | AXIS 2 | AXIS 3 |
| R empty | R0 | 1,78 | -0,73 | 0,30 | - | - | - |
| R semantic | Rs | 1,81 | -0,65 | -0,12 | - | - | - |
| R unintelligible | Ri | 1,73 | -0,51 | -0,02 | - | - | - |
| R other | Ro | 1,72 | -0,76 | -0,48 | - | - | - |
| R circumlocution | Rc | 1,71 | -0,70 | -0,45 | - | - | - |
| R phonological | Rf | 0,20 | 2,35 | 0,96 | -0,45 | 0,06 | -0,24 |
| R pseudo-word | Rp | 0,46 | 2,14 | -2,56 | 0,91 | 0,15 | 0,42 |
| R mixed | Rm | 0,45 | 2,53 | -0,71 | 0,12 | 0,09 | 0,76 |
| R unrelated | Rnr | 0,48 | 2,03 | -2,18 | 0,91 | -0,64 | 0,33 |
| R correct | Rl | -0,66 | -0,31 | 0,03 | - | - | - |
| Position in the word | Label | AXIS 1 | AXIS 2 | AXIS 3 | AXIS 1 | AXIS 2 | AXIS 3 |
| Po not apply | P0 | 1,72 | -0,56 | -0,10 | 1,52 | -0,27 | 1,01 |
| Po initial | Pi | 0,29 | 2,47 | 0,33 | -0,12 | 0,00 | -0,05 |
| Po medial | Pm | 0,21 | 2,34 | 0,71 | -0,45 | 0,25 | -0,38 |
| Po final | Pf | 0,39 | 2,54 | -0,95 | 0,30 | -0,26 | -0,10 |
| Po complete | Pc | -0,63 | -0,22 | 0,04 | -0,33 | 0,24 | 0,24 |
| Time of therapy | Label | AXIS 1 | AXIS 2 | AXIS 3 | AXIS 1 | AXIS 2 | AXIS 3 |
| < 1 year | T1 | 0,08 | 0,07 | -0,57 | 0,62 | 0,20 | 0,13 |
| 1 - 2 years | T2 | -0,24 | 0,11 | 0,13 | -0,40 | -0,87 | -0,16 |
| > 2 year | T3 | 0,21 | -0,26 | 0,70 | -0,73 | 1,67 | 0,04 |

(continued)

Table 3. (Continuation)

| Number of segments | Label | AXIS 1 | AXIS 2 | AXIS 3 | AXIS 1 | AXIS 2 | AXIS 3 |
|---------------------|-------|--------|--------|--------|--------|--------|--------|
| 2 segments | s2 | -0,23 | -0,45 | -0,26 | 2,11 | -0,61 | -1,03 |
| 3 segments | s3 | -0,42 | -0,20 | -0,09 | 1,12 | -0,02 | -1,43 |
| 4 segments | s4 | -0,24 | -0,24 | -0,12 | 0,28 | 0,36 | -0,94 |
| 5 segments | s5 | -0,07 | -0,11 | -0,22 | 0,45 | 0,55 | -0,64 |
| 6 segments | s6 | 0,02 | 0,07 | -0,20 | 0,01 | -0,26 | 0,09 |
| 7 segments | s7 | -0,01 | -0,06 | -0,08 | 0,22 | -0,21 | 1,14 |
| 8 segments | s8 | 0,29 | 0,13 | -0,07 | -0,02 | -0,05 | 1,37 |
| 9 segments | s9 | 0,43 | 0,41 | 0,32 | -0,58 | -0,20 | 0,68 |
| 10 segments | s10 | 0,60 | 0,70 | 0,68 | -0,93 | 0,29 | 0,53 |
| 11 segments | s11 | 0,81 | 0,86 | 3,23 | -1,87 | -0,20 | -1,48 |
| 13 segments | s13 | 0,74 | 1,14 | 3,23 | -1,48 | -0,23 | -1,02 |
| Number of syllables | Label | AXIS 1 | AXIS 2 | AXIS 3 | AXIS 1 | AXIS 2 | AXIS 3 |
| 1 syllable | i1 | -0,27 | -0,32 | -0,23 | 1,72 | -0,19 | -1,29 |
| 2 syllables | i2 | -0,18 | -0,15 | -0,12 | 0,25 | 0,66 | -0,90 |
| 3 syllables | i3 | -0,03 | 0,00 | -0,21 | 0,10 | -0,38 | 0,25 |
| 4 syllables | i4 | 0,25 | 0,16 | 0,07 | -0,22 | -0,07 | 1,11 |
| 5 syllables | i5 | 0,82 | 0,94 | 0,85 | -0,97 | 0,02 | 0,36 |
| 6 syllables | i6 | 0,77 | 0,99 | 3,23 | -1,65 | -0,22 | -1,22 |

Figure 2 below shows the mapping of the categories of the variables, presented in the above table, related to the variables with type of response, number of segments, number of syllables, position in the word. Figure 3 presents the observed values for all the 2106 responses for the two main axes.

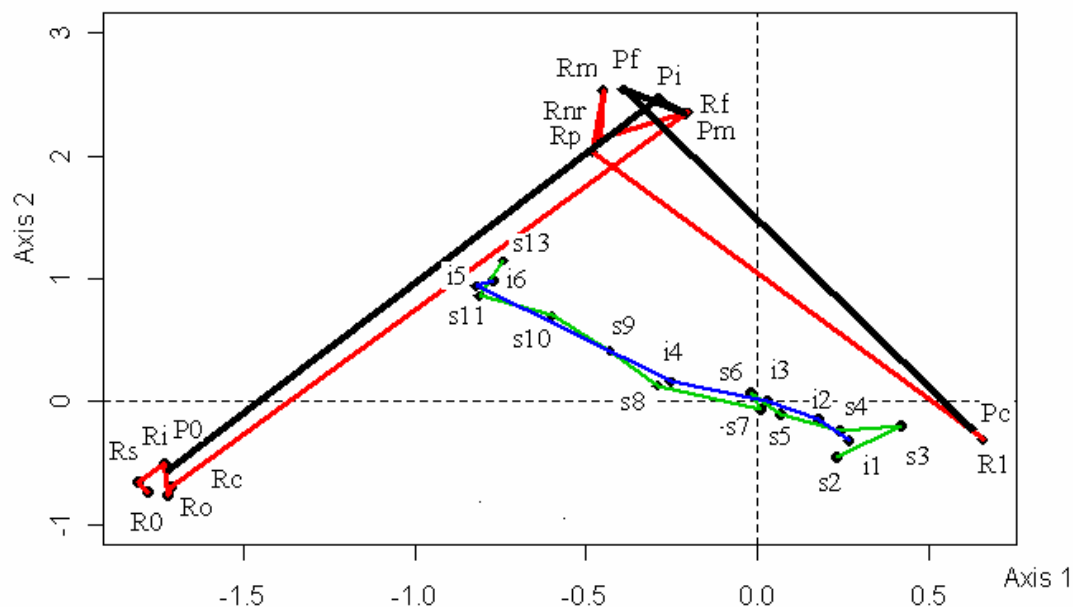


Figure 2. Type of response, position in the word, number of segments and of syllables, on the complete MCA plane spanned by the first two factors.

In Figure 3 the pattern of the units is reported on the same first factor plane. Here the partition of the units in three classes according to the null, intermediate, and complete response is clearly visible. Indeed, one may think that the different position in each class depends on the length of

the target word. Summarizing, it is evident the predominance of the extreme responses on this first factor plane: thus the examination of the third dimension may be of some interest.

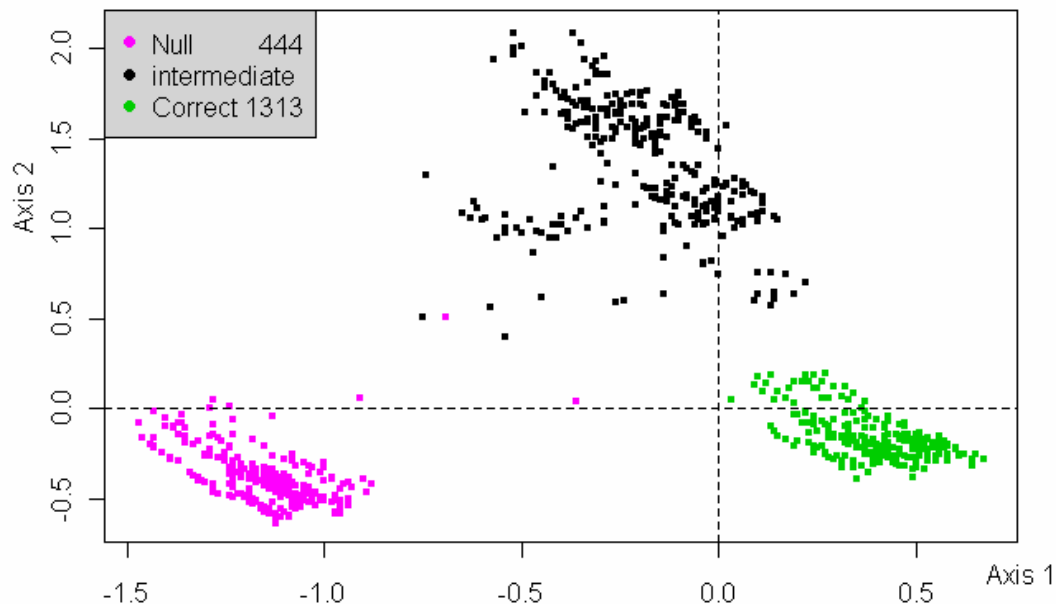


Figure 3. The units on the complete *MCA* plane spanned by the first two factors

Indeed, to the third axis contribute strongly both Indexes intermediate level, and significantly most of the characters. The inspection of Figures 4, in which the indexes levels are represented on the plane spanned by the factors 2 and 3, shows their regular displacement according to the increase of the quality of response.

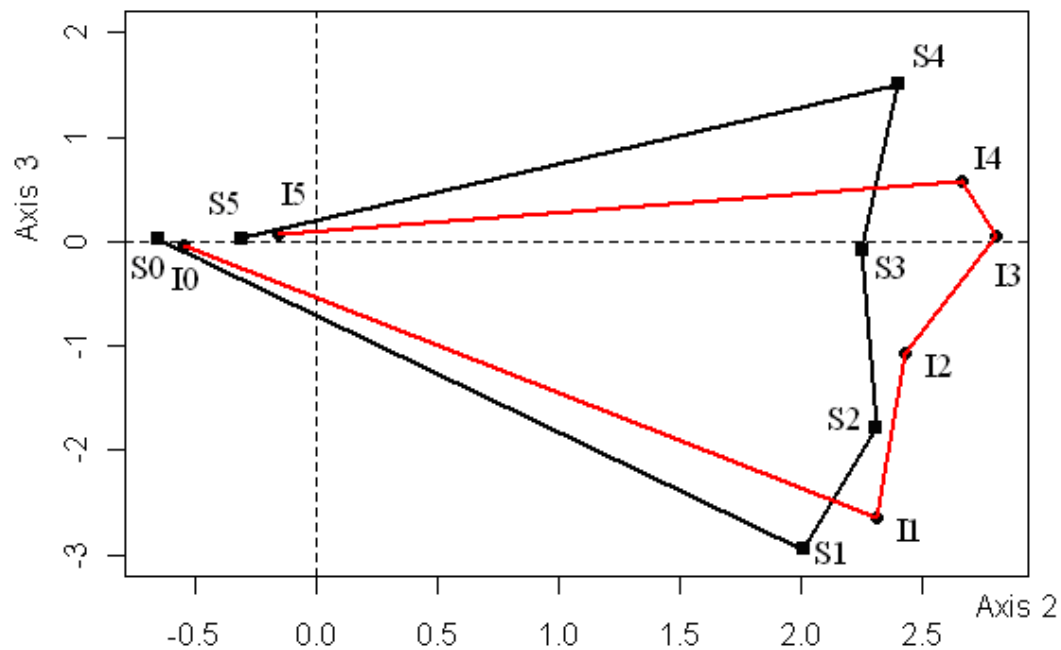


Figure 4. The trajectories of the two indexes of similarity on the complete *MCA* plane spanned by the second and third factors.

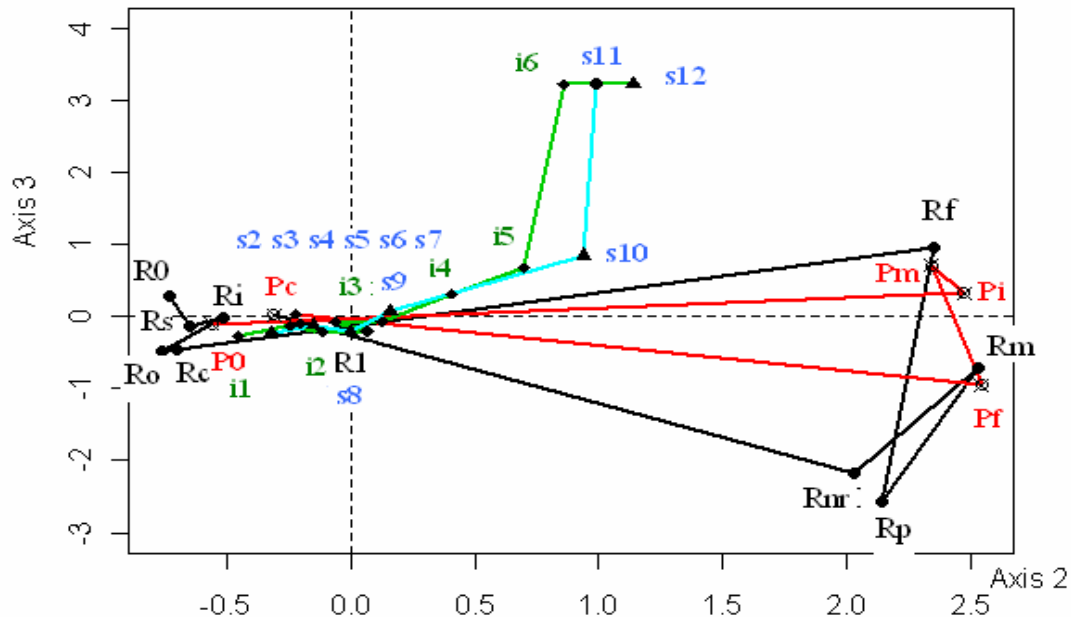


Figure 5. Type of response, position in the word, number of segments and of syllables, on the complete *MCA* plane spanned by the second and third factors.

Looking at Figure 5, in which the other characters' levels are represented, one may notice that the increase in length of the target words is in the direction of the better answers. This is an interesting feature, that may be interpreted in a capacity to better reconstruct long words than short ones. As for the complex trajectories of type of response and position in word, it is interesting to observe that the predefined sequence seems not suitable according to the position along the third axis and a better one may be considered: in order of decreasing difficulty the position results final, intermediate, and medial and the type of response would be ordered as unrelated, pseudo-word, mixed, and phonological.

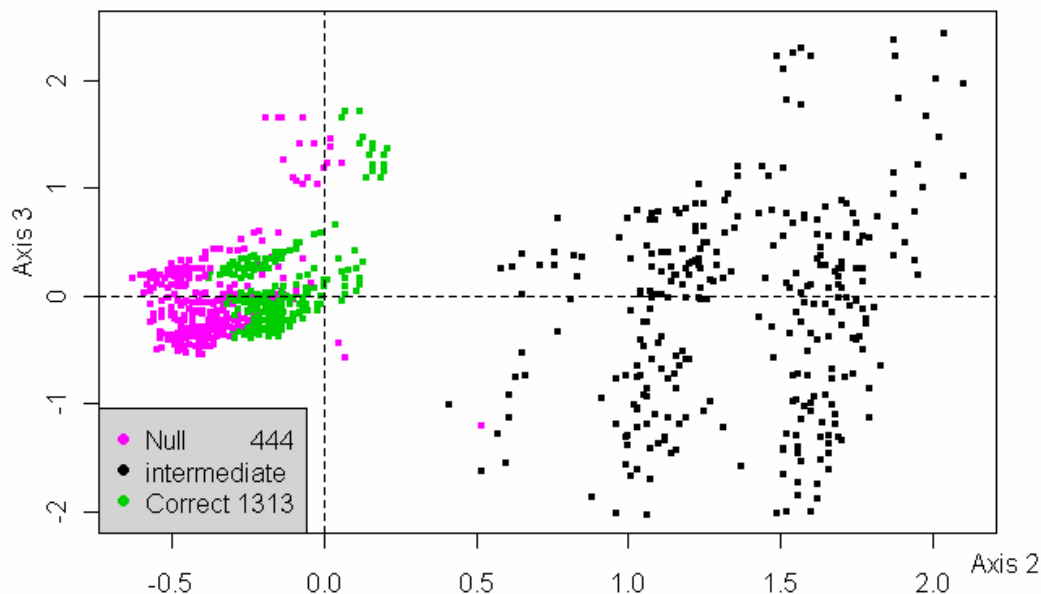


Figure 6. The units on the complete *MCA* plane spanned by the second and third factors.

In Figure 6, the scattering of the units on the same factor plane is reported. In this case, the extreme responses are confounded on the left side, whereas the intermediate ones seem set according to three layers, but its meaning is not currently clear.

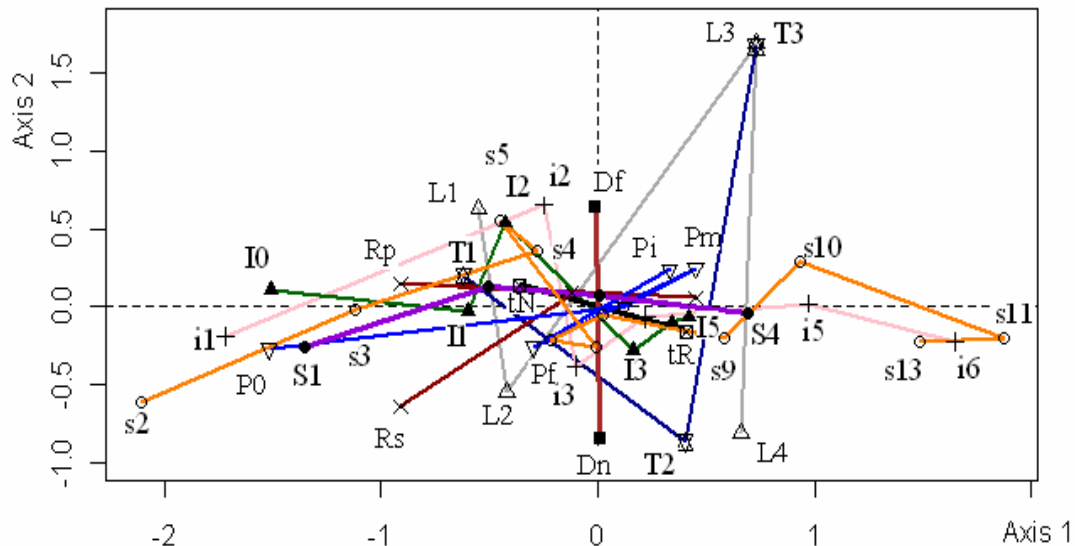


Figure 7. The distribution of the characters' levels in the first factor plane of the reduced *MCA*.

Thus, to focus on the intermediate responses, we concentrated the attention on them, by running a *MCA* on only the 304 units, corresponding to the levels 1 through 4 of the Index of Segmental Similarity.

In this case, only 16 eigenvalues are above their mean and only seven of them are larger than the upper limit of the confidence interval for the mean, that now equals .16450. Considering the re-evaluations in Table 2, the first two factors summarize 55 and 39% of the inertia, according to either criterion. To the first factor all variables contribute in a rather balanced way, excluding type of test and diagnostic. Looking at Figure 7 one can see that all variables levels follow a similar pattern. Nevertheless, it is strange that in this case the best responses correspond to the longest words both in syllables and in number of segments. The interest of the third factor seems limited to the diagnostic, with an important relation with the levels of time of injury and time of therapy, with an alternate pattern of time of injury, probably depending on the biased situation of the patients, that are too less to give a very balanced design, especially in what concerns the time of disease and of treatment.

4 Conclusion

The results obtained through the complete *MCA* show a polarization in three large groups, the intermediate of which may be studied along a third distribution, but with problems in the interpretation of the scattering according to layers. On the other side, the pattern of the character's levels according to the three main axes may be easily interpreted and constitute a nice description of the relations that exists among them.

The study of the reduced *MCA* puts in evidence some contradiction that may be understood only on the base of the too reduced number of patients (15) that took part to the experiment. For a further study we suppose that the investigation should be carried out considering a larger number of patients, in particular if a better understanding is searched of the discrimination given by the index of degree of segmental similarity in its intermediate levels.

References

- Abdi, H.** (2007), *Singular Value Decomposition (SVD) and Generalized Singular Value Decomposition (GSVD)*, in: N. Salkind (ed.), *Encyclopedia of Measurement and Statistics*, Thousand Oaks (CA), Sage.
- Baccini A.** (1984), *Étude comparative des représentations graphiques en analyses factorielles des correspondances simples et multiples*, Toulouse, Université Paul Sabatier, Laboratoire de Statistique et Probabilités, Rapport N° 02-84.
- Ben Ammou, S. and G. Saporta** (2003), «On the connection between the distribution of eigenvalues in multiple correspondence analysis and log-linear models», *REVSTAT-Statistical Journal*, 1: pp. 42-79.
- Benzécri, J. P.** (1979), «Sur les calcul des taux d'inertie dans l'analyse d'un questionnaire», *Les Cahiers de l'Analyse des Données*, 4(3): pp. 377-379.
- Benzécri, J.P., et coll.** (1973-82), *L'Analyse des données*, Tome 1, Paris, Dunod.
- Bose, A., Raza, O., Buchanan, L.** (2007), «Phonological relatedness between target and error in neologistic productions». *Brain and Language*, 103: pp. 120-121.
- Burt, C.** (1950), «The Factorial Analysis of Qualitative Data». *British Journal of Statistical Psychology*, 3: pp. 166-185.
- Camiz, S.** (2001), «Exploratory 2- and 3-way Data Analysis and Applications», *Lecture Notes of TICMI*, 2: pp. 1-37. <http://www.emis.de/journals/TICMI/Int/vol2/lecture.htm>
- Dell, G.S., Schwartz, M. F., Martin, N., Saffran, E. M., Gagnon, D. A.** (1997), «Lexical Access in Aphasic and Nonaphasic Speakers». *Psychological Review*, 104: pp. 801-838.
- Eckart, C. and G. Young** (1936), «The approximation of one matrix by another of lower rank». *Psychometrika*, 1: pp. 211-218.
- Fisher, W.D.** (1958), «On Grouping for Maximum of Homogeneity», *Journal of American Statistical Association*, 53: pp. 789-798.
- Gifi, A.** (1981), *Nonlinear Multivariate Analysis*. Leiden, Department of Data Theory FSW/RUL. New edition by DSWO-Press, Leiden, 1987.
- Greenacre, M. J.** (1984), *Theory and Application of Correspondence Analysis*, London, Academic Press.
- Greenacre, M.J.** (1988), «Correspondence analysis of multivariate categorical data by weighted least squares», *Biometrika*, 75: pp. 457-467.
- Greenacre, M.J. and J. Blasius** (Eds.) (2006), *Multiple Correspondence Analysis and Related Methods*, Dordrecht (The Netherlands), Chapman and Hall (Kluwer).
- Greenacre, M.J.** (2006), *From Simple to Multiple Correspondence Analysis*, in: Greenacre and Blasius (Eds.): pp. 41-76.
- Guttman, L.** (1941), *The Quantification of a Class of Attributes: a Theory and Method of Scale Construction*. In P. Horst (ed.) *The Prediction of Personal Adjustment*. New York, Social Science Research Council.
- Hayashi, C.** (1956), «Theory and Examples of Quantification IT», (In Japanese), *Proceedings of the Institute of Statistical Mathematics*, 4: pp. 19-30.
- Jolliffe I.T.** (2002), *Principal Component Analysis*, 2nd ed., New York, Springer, Springer Series in Statistics.
- Langrand, C. and L.M. Pinzón** (2009), *Análisis De Datos. Métodos y ejemplos*, Bogotá, Escuela Colombiana de Ingeniería Julio Garavito.
- Pierrehunbert, J.** (2003). *Probabilistic Phonology: Discrimination and Robustness*. In R. Bod, J. Hay and S. Jannedy (eds.) *Probability Theory in Linguistics*. The MIT Press, Cambridge (MA): pp. 177-228.