USING EXPLORATORY ANALYSIS AND DISTANCE TO EVALUATE THE COMPONENTS OF A NAMING TEST FOR STUDYING APHASIA

Gastão Coelho Gomes[a,1], Sergio Camiz[b,2], Christina Abreu Gomes[c,3], Fernanda Duarte Senna[c,3]

[a] *DME–IM–UFRJ, Caixa Postal 68530 cep: 21945-970, RJ*
gastao@im.ufrj.br
[b] *Dipartimento di Matematica Guido Castelnuovo Sapienza Università di Roma, Piazzale Aldo Moro 5, I - 00185 Roma Italia*
sergio.camiz@uniroma1.it
[c] *Departamento de Linguística, UFRJ, Cidade Universitária, RJ*
christina-gomes@uol.com.br
fonofernandasenna@gmail.com

Abstract

A set of exploratory analyses were used to analyze two components of a naming test for studying lexical access in aphasic patients, that is, the naming agreement of images and age of acquisition of their corresponding words, taken from the Snodgrass and Vanderwarts (1980) original test. To get the test reliable, the images should be easily and inequivocally named by the same word for any person. Theoretical assumptions about word learning state that words acquired later tend to be the first lost in the situation of brain damage in aphasia. Thus, these two indices are important predictors of the patients word retrieval. Thus, the images were submitted to a set of judges, then the given scores have been studied through exploratory factor analyses. Eventually, 161 images could be selected and the primitiveness of words in terms of their age of acquisition could be estimated.

[1] Partially supported by Sapienza Università di Roma
[2] Partially supported by FAPERJ.
[3] Partially supported by CNPq.

Resumo: Análises exploratórias de dados Multidimensionais foram usadas para avaliar dois componentes de um teste de nomeação para o estudo do acesso lexical em pacientes afásicos, ou seja, o acordo de nomeação de imagens e a idade de aquisição de suas palavras correspondentes, tomada a partir do teste proposto por Snodgrass e Vanderwarts (1980) . Para o estudo ser confiável as imagens devem ser inequivocamente reconhecidas por qualquer indivíduo usando a mesma palavra/imagem. Pressupostos teóricos sobre as palavras afirmam que palavras adquiridas mais tarde na aprendizagem tendem a serem as primeiras a serem perdidas devido a um dano cerebral em afasia. Assim, estas duas variáveis são importantes preditoras da recuperação de palavras. Primeiro selecionamos as imagens de reconhecimento com juízes normais; então, as classificamos com relação a sua primitividade. Estas imagens foram submetidas a dois conjuntos de juízes, que tiveram que responder de acordo com duas diferentes escalas. Os dados foram analisados com técnicas de Análises Exploratórias Multidimensionais, incluindo Análises de Correspondências Simples e Múltiplas, Análise de Componentes Principais e Análise Fatorial Múltipla. Eventualmente, 161 imagens poderiam ser seleccionada e a primitividade destas palavras, em termos da sua idade de aquisição podem ser estimada.

Este trabalho foi revisto durante a visita do prof. Camiz a UFRJ, em abril 2014, apoiada pela FAPERJ, processo APV-E-26/110.019/2014, foi enviado para submissão na revista "Discrete Applied Mathematics - special edition, DGA'2013.

Keywords: Aphasia, Exploratory Factor Analysis, Linguistics, Principal Component Analysis, Correspondence Analysis, Multiple Factor Analysis.

1. Introduction

This study concerns the evaluation of figures to be used in a test for studying lexical access in aphasia (the loss of some abilities related to language production and/or comprehension due to brain damage). The extent to what the disease affects word retrieval is measured through a test, in which aphasic patients must verbalize a series of images that represent familiar objects. In order to be reliable, a selection of images internationally used for this test (Snodgrass and Vanderwarts, 1980) must be performed, based on, at least, two criteria: the images should be easily and unequivocally recognizable and the primitiveness of the word regarding its age of acquisition should be checked.

Indeed, theoretical assumptions about word learning state that words acquired later tend to be the first to be lost in the situation of brain damage in aphasia. The age of words acquisition was first used as a variable to evaluate lexical access by Caroll and White (1973). According to the authors, the age at which the name related to an object is acquired is more important in determining the speed of lexical access than word frequency. Brown and Watson (1987) state a difference in the phonological representation of words acquired earlier in life, as they are stored in memory in a more complete form, while a different mechanism occurs for the representation of late-acquired words. So the inclusion of age of acquisition (*primitiveness*) in this research is legitimated by a long tradition of lexical access studies. The procedures were adapted from Cuetos et al. (2002).

Thus, these two indices, the ability of an image to be recognized and its primitivity, are important predictors of the patients word retrieval. This work is based on the set of 260 images taken from Snodgrass and Vandewarts (1980) and describes how we proceeded to the estimation of these indices. In the first study, we aimed at selecting the images with the highest agreement of naming and in the second we tried to build a primitivity index, to be given to each of the selected images. For both tasks, we carried out a survey involving panels of judges randomly chosen and both selecting the images and giving the primitivity index according to a set of exploratory analyses.

2. The data

The basis of the current study is the set of images found in Snodgrass and Vanderwarts (1980). The authors proposed a set of 260 pictures for use in experiments investigating differences and similarities in the processing of pictures and words. The pictures had been selected according to a set of rules that provide consistency of pictorial representation. In Snodgrass and Vanderwatts experiment, 4 groups of judges were presented to four different sequences of the 260 slides and performed four different tasks in order to evaluate the drawings. They were asked to identify each picture as briefly and unambiguously as possible by writing only one name. They also judged the familiarity of each picture "according to how usual or unusual the object is in your realm of experience." (p. 183), evaluating the visual complexity of the image by using a scale of 5-points in which 1 indicated very simple and 5 very complex, with the complexity defined as the amount of detail or intricacy of lines in the picture. The judges provided information about the image agreement, in order to establish how closely each picture resembled their mental image of the object. As a result the pictures were standardized, according to four characters of central relevance to memory and cognitive processing: name agreement, image agreement, familiarity, and visual complexity, and have been used in different tasks that associated pictures and words. Indeed, we had to adapt the test to the specific brazilian reality, both on the point of view of the identification and of their naming. Thus,

- for the selection of the images, all of them (261) were submitted to a panel of 38 judges randomly selected among non-affected people. The answers have been coded as 1 = recognized, 0 = not recognized. From this selection, 161 images resulted;

- to measure primitiveness, we asked 128 other non-affected judges to estimate how primitive were the 161 represented objects, according to their personal experience. This estimation was based on two different scales: *i)* the first panel, with 60 judges, labeled E, was asked to measure the age of acquisition on a scale from 1 to 7 according to how early in their life each word was first known, without specifically mentioning the age; thus, 1 corresponds to very early in life and 7 to most late; *ii)* the second panel, with 68 judges, labeled I, was asked to measure based on a scale 1-7 this time based on age classes: 1=0-2 years, 2=2-4 years, 3=4-6 years, 4=6-8 years, 5=8-10 years, 6=10-12 years and 7=13 and further.

3. Theoretical Framework

The limits of the surveys prevented the use of statistical tests, since neither the randomness nor the representativeness requirements could be fulfilled. On the opposite, we could take advantage of the tools available in the framework of exploratory analysis. By this term, due to Tukey (1977), we refer to all mathematical, statistical, and computer science methods that may concur to study the structure of a data set in order to extract the contained information in an interpretable way. Their use, started by Benzécri (1973-82) and his French school, that claim that “the models should follow the data, not the inverse”, may be roughly considered a cognitive activity, nowadays referred as data mining in the Anglo-Saxon world. The exploratory work is not exhaustive (Tukey, 1977): it is a detective work (Diggle et al., 1994) adopted to ease the acquisition of information contained into the data, with the drawback that this information may not be inferred statistically to a population. Thus, for this task, more classical statistical tests are required, to be performed accordingly. In this work we adopte one of the two main sets of tools that are available in multidimensional exploratory data analysis, corresponding to the two big frames currently taken into account (Benzécri, 1973-82), the ordination, as the other, the classification, did not result of interest for our current purposes. Ordination aims at finding uni- or multi-dimensional orderings in the data, that may be used to sort them accordingly. The identification of such orderings, usually independent from each other, sometimes allows identifying them with factors that influence the objects at hand, thus causing the found data diversity. We adopted different ordination techniques, according to the kind of data at hand, as each one allows some analysis tools and prevents some other, depending upon their specific features.

3.1. Ordination

The identification of orderings, also known as gradients or factors, is based on the well known *Singular Value Decomposition* (*SVD*, Eckart and Young, 1936; Greenacre, 1983; Abdi, 2007) and on Eckart and Young's theorem. *SVD* states that any real matrix A may be decomposed according to $A = U\Lambda^{1/2}V'$, with a real diagonal matrix Λ (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 up to the ordering of elements, with U a matrix of eigenvectors of AA', V one of $A'A$ and Λ the diagonal matrix whose elements are the corresponding eigenvalues of both. To get the solution unique, it is sufficient to sort all matrices according to the decreasing order of the eigenvalues. Thus, the elements of A may be reconstructed by the formula

$$x_{ij} = \sum \sqrt{\lambda_\alpha}\, u_{i\alpha} v_{i\alpha}.$$

The most interesting feature of *SVD* results from the Eckart and Young's (1936) statement that, the reconstruction limited to the first r largest eigenvalues is the best r-dimensional one in the sense of least-squares: $x_{ij} \approx$

$\sum_{\alpha=1}^{r} \sqrt{\lambda_\alpha}\ u_{i\alpha} v_{i\alpha}$. 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 $\frac{\sum_{\alpha=1}^{r} \lambda_\alpha}{\sum_\alpha \lambda_\alpha}$.

If we take a real data matrix X, with n rows and p columns, we can consider both of them as vectors spanning respectively the n- and p-dimensional spaces R^n and R^p. Thus, implicitly, both rows may be represented as points in the space spanned by the others. Due to this representation, it results that *SVD* is strictly tied to the diagonalisation of the scalar products defined in these spaces by XX' and $X'X$ respectively. It may be proved that the eigenvectors are orthogonal and that in decreasing order they maximize their scalar product with all the vector of the basis. It may be also proved that the straight lines spanned by each one maximize the inertia of the orthogonal projection of the clouds of points on them. This leads to an important paradigm of exploratory ordination, that is the higher importance of the directions of maximum inertia in respect to the others.

In this work, we used four different methods of ordination, Principal Component Analysis (*PCA*), and Simple (*SCA*) and Multiple Correspondence Analysis (*MCA*), according to the nature of the data: *PCA* for measures, *SCA* for frequencies, and *MCA* for qualitative data. In addition, we adopted Multiple Factor Analysis (*MFA*) to compare the results of two different scales of judgements. Indeed, the ordinal scale scores given by the judges may not easily be dealt with, as no easy ordination technique is currently available. Thus we used both *PCA*, considering the scales as true measures, and *MCA*, losing the scale nature, but earning the possibility to locate each scale's levels on the factor space.

3.2. Transformation of the Data Matrix

An interesting feature of the analyses that we adopted is that they are all based on *SVD* of some transformation of the original data matrix $T : X \rightarrow A = T(X)$. The data transformations, according to the different methods may be described as follows:

PCA $\quad x_{ij} \rightarrow z_{ij} = \frac{x_{ij} - \overline{x}_j}{\sqrt{n}\sigma_j}$ $\quad$ standardization

MFA $\quad x_{ij_k} \rightarrow z_{ij_k} = \frac{x_{ij_k} - \overline{x}_{j_k}}{\sqrt{\lambda_k^1}\sqrt{n}\sigma_{j_k}}$ $\quad$ std. adjusted to group's coherence

$$\textbf{SCA} \quad x_{ij} \rightarrow s_{ij} = \frac{x_{ij}}{\sqrt{x_{i.}x_{.j}}} - \frac{\sqrt{x_{i.}x_{.j}}}{x_{..}} \quad \text{deviation from independence}$$

$$\textbf{MCA} \quad x_{ij_q} \rightarrow s_{ij_q} = \frac{1}{\sqrt{Q}}\left(\frac{x_{ij_q}}{\sqrt{x_{i.}}} - \frac{\sqrt{x_{i.}}}{x_{..}}\right) \quad \text{deviation from average profile}$$

(where Q is the number of variables).

3.3. Principal Component Analysis (PCA)

Principal Component Analysis (Benzécri, 1973-82; Gower and Hand, 1996; Jolliffe, 2002; Langrand and Pinzón, 2009) is the most known ordination method suitable for measure data. It gives principal components as directions along which the maximum of data inertia results. Considered as new variables, the principal components are maximally correlated with all the variables that form the vector space's basis. This leads to a nice interpretation of these components, that are built as linear combinations of the variables and that represent a component common to all of them. The reciprocal relations between variables and principal components derive from the eigenvectors and result in absolute (the share of principal component variation due to the unit variation of a variable) and relative contributions (the share of the variable variation due to the unit variation of the principal component). This method will be used for the scales scores, albeit they only roughly approximate a measure.

3.4. Simple Correspondence Analysis (SCA)

Simple Correspondence Analysis (Benzécri et coll., 1973-82; Greenacre, 1983) was developed in the framework of the study of the contingency data tables crossing two qualitative variables, but it is nowadays applied to any table of positive values. We do not step here into discussion of such choices: suffice here to say that some of its features find their rationale (and their utility) in the decomposition of the chi-square in independent components that loses much of its sense once applied to tables in which the chi-square statistics may not be applied. It may be shown that the reconstruction formula in this case becomes

$$f_{ij} = f_{..}\, p_{ij} = n_{..}\, p_i\, p_j \left(1 + \sum_{\alpha=1}^{min(m,q)-1} \frac{1}{\sqrt{\lambda_\alpha}}\, \Phi_{i\alpha}\Psi_{j\alpha}\right)$$

with $\Phi_{i\alpha}$, $\Psi_{j\alpha}$ the *SCA* factors.

3.5. Multiple Correspondence Analysis (MCA)

Multiple Correspondence Analysis (Benzécri et coll., 1973-82; Greenacre, 1983; Gower and Hand, 1996) is considered a PCA for qualitative data. It is also a generalization of Correspondence Analysis to the case of several categorical variables, as it shares its same chi-square metrics. In Camiz et al. (2010) the relations between MCA and SCA applied to both the indicator matrix Z derived from X and the so-called Burt's matrix $B = X'X$ are described. Indeed, the two methods give related solutions. We shall adopt this method when dealing with the scales scores, losing their scale nature, but keeping the ability to represent all levels on the factor spaces. This could reveal an improvement in respect to PCA, albeit an important Guttman effect may result. Indeed, this could be interpreted as a seriation and, as such, taken into account.

3.6. Multiple Factor Analysis (MFA)

A special case of data tables is the one in which the same units are observed according to either several sets of variables, that for some reason one wants to keep separated, or the same variables repeatedly observed on the same units in different occasions. These multiple tables deserve being analysed in a special way, in order to distinguish the structure of each table/occasion and show the differences among the tables. For this task, the current two-way methods are not suitable, because it is neither possible to identify the structure of each subtable nor show the variation of the pattern from one table to the other. Indeed, the structure of each table may be investigated by a specific PCA, but the following comparison of the results may be cumbersome and not effective. For a simultaneous study, Multiple Factor Analysis (Escofier and Pagès, 1997) was introduced, based on multiple tables with the same units and different sets of variables or the same variables observed in different occasions. Technically speaking, it is but a weighed PCA that is a PCA of a pooled table, in which all the values of each table are divided by the square root of the inverse of the first eigenvalue of its individual PCA, but this allows very useful developments. This rescaling is introduced to equilibrate the importance of all tables in this pooled analysis: in fact, without it, tables with stronger inner structure that is with more correlated variables, would have a higher influence on the first largest factors. In this way, instead, they are normally balanced. As well, this rescaling allows to compare the inertia of each table projected on the factors with the usual interpretation of contributions and quality of representation not only of the variables but also of each table as a whole. The comparison of these tables inertias on the factors allows to study the interstructure, that is the mutual relations among the tables; the representations on the *MFA* factor spaces allows both the representation of the compromise, that is the representation of the units due to all tables, and of the intrastructure, that is the common representation of all variables and the individual factors; eventually, on this same space, the partial units, that is those seen by the different tables, may be individually projected, allowing individual comparisons and the drawing of their trajectories.

Exploratory multidimensional factor analyses were used throughout the investigation. According to the data at hand, the analyses have been submitted to *SCA*, to identify both judges and items with critical behavior, *PCA* to check the agreement among judges, with *MCA* to identify the special behaviour of some of these in respect to others. Both *MCA* and *PCA* were used to evaluate the primitiveness of the words, the former to show the different pattern due to the two judgement scales and the latter to provide a primitivity index.

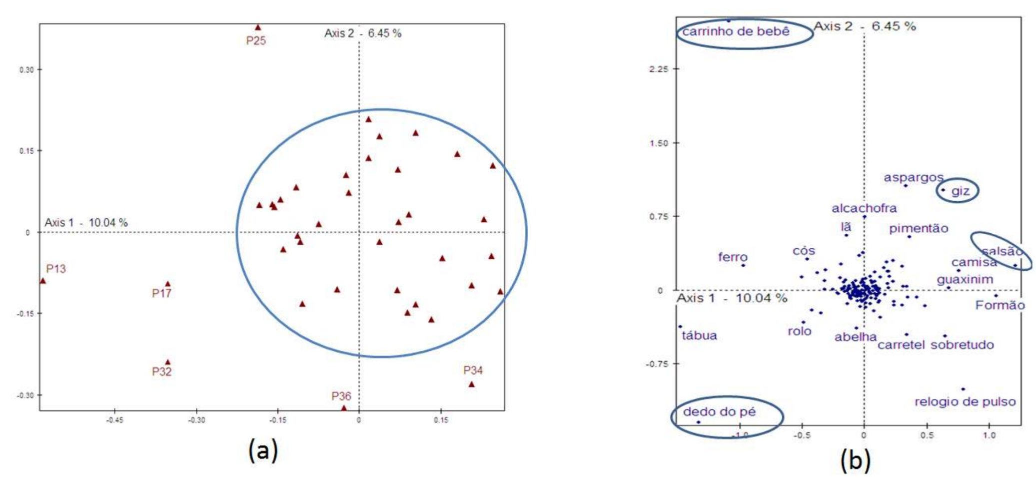

Figure 1: *Analysis for the selection of the images. The items on the first factor plane of SCA: (a) The judges, (b) The names.*

4. Results

4.1. Selecting the images

A selection of the images proposed by Snodgrass and Vanderwarts (1980) was necessary to cope with the Brazilian environment in which the test would be carried out, both in respect to the familiarity and the verbalization. Thus, we tried to identify which of the 260 images, with their associated name, could be used for the test, aiming at evaluating the ability of aphasic patients to identify and correctly name the submitted images. Thus we first submitted the images to a panel of 38 judges and built a data table, which entries were one, if the name given by the judge was the same expected by the researcher, and zero otherwise. At first glance, 66 items were recognized by all judges and 10 were never: whereas the former were automatically retained, the latter were immediately rejected. Therefore, we applied *SCA* to the remaining images to get a graphical representation of the pattern of both judges and names on the factor planes. This could allow to identify possible deviant judges to withdraw. According to Figure 1*a* below, six judges, *P13, P17, P25, P32, P34*, and *P36*, appear further from the origin than all others, whose pattern results rather homogeneous. The inspection of the scatter of the names on the same first plane (Figure 1*b*) shows a pattern in which those further apart from the central cloud correspond to items very rarely identified. Some of them, *carrinho de bebê* (baby stroller), *dedo do pé* (toe), *salsão* (celery), and *giz* (chalk), were identified by no more than 5 judges and many others were rarely identified. They are located at the border of the cloud as can be seen on Figure 1*b*.

Thus, we could say that on the *SCA* first factor plane, both judges and items are far from the center depending upon the reduced number of their correct identification. We re-ran *SCA* with only 32 judges, having removed the 6 outliers and also all whose frequency of correct identification is lower than 50%. Thus only 55 items were taken into account, their correct identification ranging within 17 and 31 judges. In this *SCA* the pattern of the judges on the first factor plane resulted much more homogeneous, without evident outliers, so that we could conclude that no further removal of judges seemed necessary.

Based on these analyses, we limited attention to the tests carried out by the 32 selected judges and decided to keep all the images that were correctly identified by at least 90% of them. Thus, 97 images were identified by all of them, 26 by only 31 (97%), 24 by 30 (94%), 14 by 29 judges (91%), so that a total of 161 images were selected.

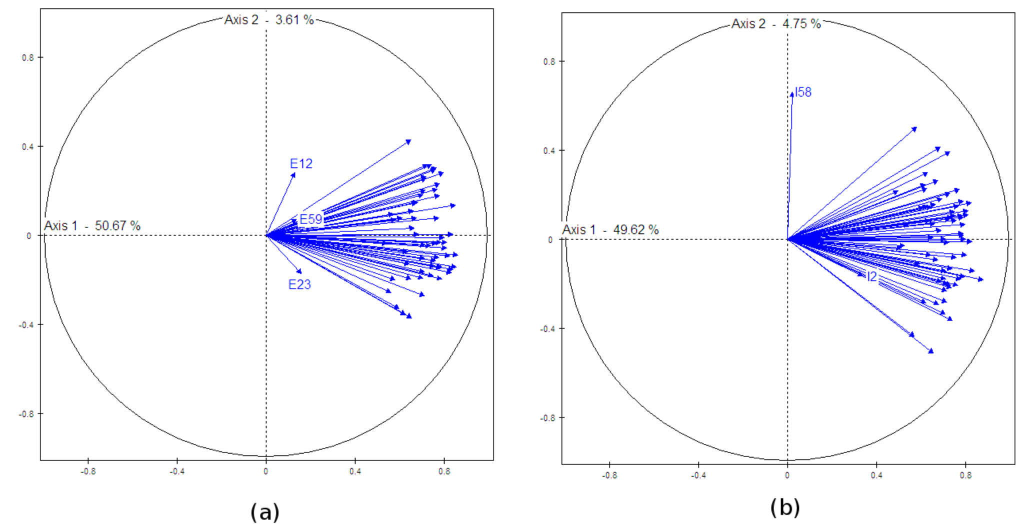

Figure 2: *Analysis of age of acquisition judgements. The judges on the plane spanned by the first two principal components of PCA on the respective data: (a) the E scale-free, (b) the I age-based scale. Only the outliers are labeled.*

4.2. Defining the words' primitiveness

The 161 words selected from the previous study had to be evaluated to define a primitivity index, that is, their estimated age of acquisition. For this task, 128 judges (different from the previous ones) were asked to estimate the degree of primitiveness of these words using two different kind of scales: *i)* the first panel, with 60 judges, labeled E, was asked to measure the age of acquisition on a scale from 1 to 7 according to how early in their life each word was first known, without specifically mentioning the age; here 1 corresponds to very early in life and 7 to most late; *ii)* the second panel, with 68 judges, labeled I, was asked to indicate in which class of years of their life they acquired them, on a scale 1-7 this time corresponding to age classes, with 1=0-2 years, 2=2-4 years, 3=4-6 years, 4=6-8 years, 5=8-10 years, 6=10-12 years and 7=13 and further.

In order to first examine the homogeneity of the judges, we started by running two separate PCAs on the tables with words in rows as units and judges in column as variables.

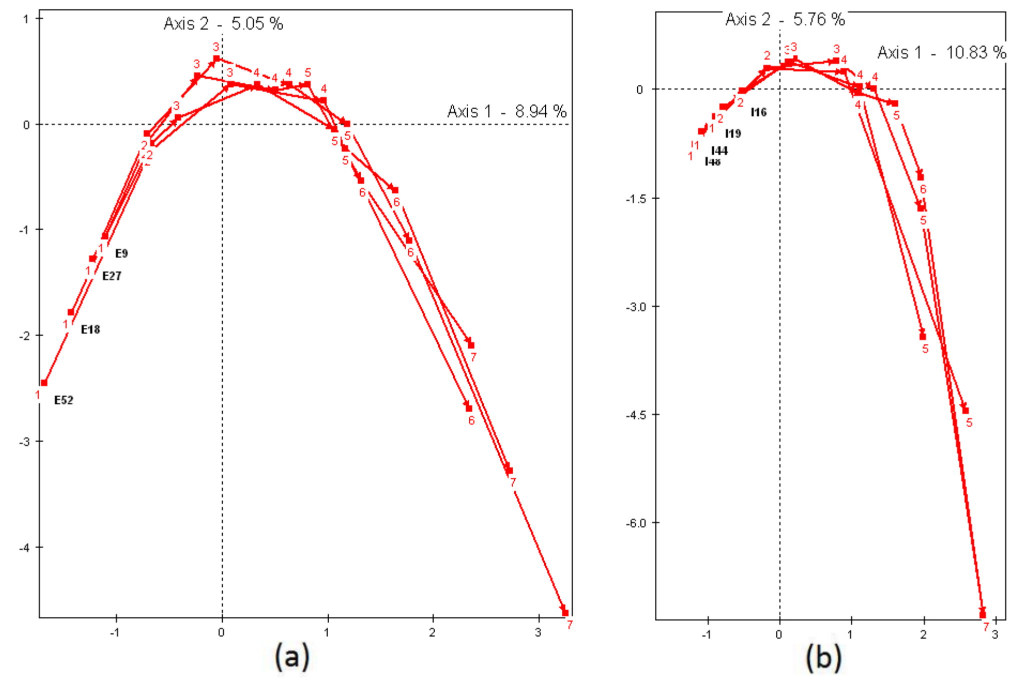

Figure 3: *Analysis of the primitiveness judgements. The judges' trajectories represented on the first factor plane of MCA: (a) free judgements, (b) judgements based on age intervals.*

Thus, the judges with a non-homogeneous behaviour would result in an outlier position in respect to the others. Indeed, five of them, *E12, E23, E59, I2* and *I58* (figure 2a and b), appeared little or no correlated with the others, thus were removed from the data set.

Indeed, in the figure the five outliers result very poorly correlated with the first axis and much more with the following ones. On the opposite, no outlier word resulted evident from the inspection of the principal plane and their contribution to the following factors (not shown).

To better understand in what the behaviour of the 5 said judges was different, we ran two *MCA*s, in which the age of acquisition was taken as a qualitative level. *MCA* allowed us to represent the trajectories of the evaluation scales used by the judges. In the Figures 3a and b, they appear on the plane spanned by the first two factors: all trajectories pattern appear pretty regular along an arch in both sets, but the outliers (not represented here), indeed a Guttman effect due to their co-occurrence. It is noteworthy a comparison of the two images: it is evident that the two panels of judges used the scales in different ways: the “free” ones adopted the scale in a more variate way than those “constrained” by the age classes. In particular, the scale’s first steps has been used much more from the free than by the constrained judges. Thus, the free scale seems more adapt for this task.

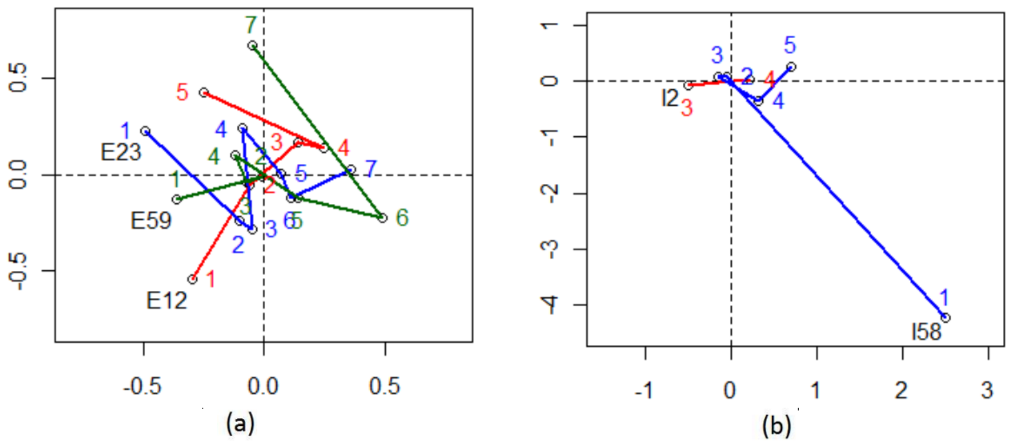

Figure 4: *Analysis of the primitiveness judgements. The outlier judges’ trajectories represented on the first factor plane of the respective MCA: (a) free judgements, (b) judgements based on age intervals.*

Unlike the other judges’, these outliers’ trajectories result very short and irregular. In Figures 4*a* and *b* their very strange pattern is represented: it is noteworthy how short they are in respect to the others. Indeed, also these five judges were eventually removed.

As a final step, we wanted to check to what extent the two judgement scales could affect the evaluation of primitiveness. Thus, we ran a *MFA*, considering the two groups of reduced judges (57 that used the free scale: E) and 66 with age-scale (I). A specific advantage of *MFA* in respect to *PCA* it its ability to represent on factor planes not only the global units, but also the partial ones, that is, in our case, the projection of the words seen by either group of judges. Indeed, the total word is situated on the centroid of the two partial words.

Therefore, distances between partial words are a measure of their dissimilarity according to the two sets of measurements and they may be decomposed according to the different axes. The words with highest negative differences along the first factor are *burro, gravata, lâmpada, mala*, and *patins*

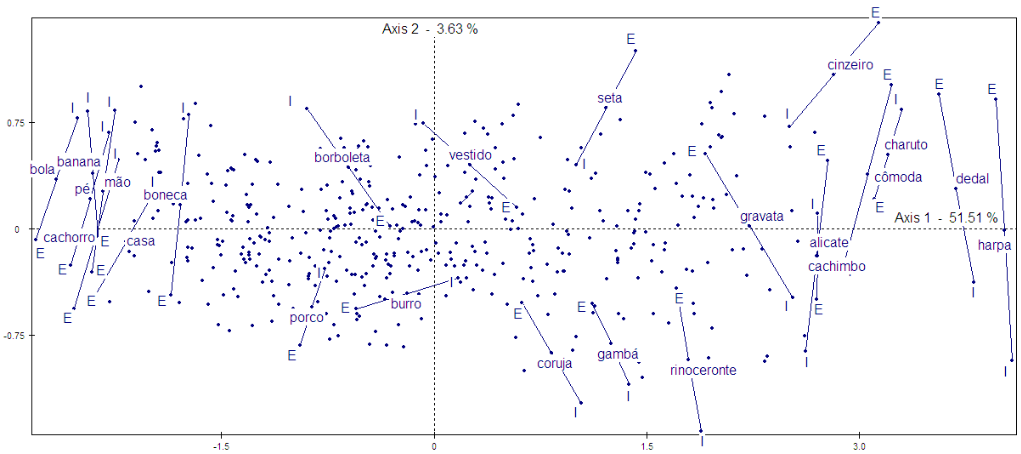

Figure 5: *Analysis of the age of acquisition judgements. All the words represented on the plane spanned by the first two factors of MFA. Only the words with the largest trajectories are labeled, with the word in the compromise position and either E or I, the partial ones.*

and those with highest positive ones are *borboleta, cigarro, cinzeiro, escada, galinha, ônibus*, and *vestido*. Thus, the first might be words judged more primitive by the free-scale judges, whereas the second might be judged more primitive by the age-scale ones. Here, we deal only with the first axis that clearly represents primitivity of words (51.51% of total inertia), since the following explain too little inertia to deserve being taken into account (the second only 3.64%). In Figure 5 all words are represented both totally and partially, with the total units at the centroid of the respective partials. Looking at the extreme of the first axis it is interesting to find the words with the largest differences on the second axis and in particular a reverse behaviour: this reflects the small rotation of the first factors of partial tables, but does not deserve a true interest for our purposes.

As the partial first factors of the two tables where most correlated among each other (.98) and with the *MFA* one (over .99), we decided to merge the two data sets, so that as measure of the words' primitivity was taken the first principal component of this unified table's *PCA*.

5. Conclusions

The study aiming at both selecting images with high naming agreement and measuring the degree of primitiveness of their correspondent words, has been carried out using only exploratory multidimensional data analyses. For the first task, this allowed to withdraw judges with a clearly biased behaviour in respect with the others, thus getting a much more homogeneous evaluation of the selected images, corresponding to those almost unanimously recognised by a more consistent panel. For the second task, once again the exploration allowed to identify outlying judges and to withdraw their evaluations from the data set. Then, the difference among the scales could be investigated, resulted in a better performance of the free scale than the other, since it allowed a more instinctive estimate, but not enough to cancel a whole panel of judges. Eventually, the first principal component of the overall *PCA* could te taken as an index and as such to be used in the test to which aphasic patients may be submitted.

References

[1] 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.

[2] Benzécri J.P. and coll. (1973-82), *L'analyse des données*, 2 voll., Paris, Dunod.

[3] Brown G.D.A., Watson F. L. (1987). "First in, first out: Word learning age and spoken word frequency as predictors of word familiarity and word naming latency". *Memory and Cognition*, 15: 208-216.

[4] Carroll J.B., White M.N. (1973a). "Word frequency and age of acquisition as determiners of picture naming latency". *Quarterly Journal of Experimental Psychology*, 25: 85-95.

[5] Cuetos F., Aguado G., Izura C., Ellis A.W. (2002). "Aphasic naming in spanish: predictors and errors". *Brain and Language*, 82: 344-365.

[6] Diggle P.J., Liang K.Y., Zieger S.L. (1994), *Analysis of Longitudinal Data*, Oxford, Clarendon Press.

[7] Eckart C., Young G. (1936), "The approximation of one matrix by another of lower rank", *Psychometrika*, 1: 211-218.

[8] Escofier B., Pagès J. (1998). *Analyses factorielles simples et multiples*, 3e ed., Paris, Dunod.

[9] Camiz S., Gomes G.C., Senna F.D., Gomes C.A. (2010). "Correspondence Analysis in a Study of Aphasic Patients", *XLII SBPO 2010*, Bento Gonçalves (RS) Brazil, 30/8-3/9. http://www.sobrapo.org.br/sbpo2010/xliisbpo_pdf/72251.pdf

[10] Gower J.C., Hand D.J. (1996), *Biplots*, London, Chapman and Hall.

[11] Greenacre M. (1983), *Theory and Applications of Correspondence Analysis*, London, Academic Press.

[12] Guttman L. (1953), "A Note on Sir Cyril Burt's Factorial Analysis of Qualitative Data", *British Journal of Statistical Psychology*, 6: 2124.

[13] Jolliffe I.T. (2002), *Principal Components Analysis*, Berlin, Springer.

[14] Rust R.T., Cooil B. (1994), "Reliabilty Measures for Qualitative Data: Theory and Implications". *Journal of Marketing Research*, 31: 1-14.

[15] Snodgrass J.G., Vanderwart M. (1980), "A standardized set of 260 pictures: Norms for name agreement, familiarity and visual complexity". *Journal of Experimental Psychology: Human Learning & Memory*, 6(2): 174-215.