DOI: 10.17469/O2103AISV000015

LORENZO CIAURELLI, DANIELA BELTRAMI

Spontaneous speech in patients with early-stage dementia and Mild Cognitive Impairment: The role of age of acquisition

The aim of this study is evaluating differences in the characteristics of words produced by patients with early-stage dementia(e-D) and Mild Cognitive Impairment (MCI) and healthy control (HC). We used two different corpora to obtain an age of acquisition (AoA) and word frequency values for words produced by subjects in two semi-spontaneous speech tasks. The results are in line with the tendencies found in previous studies for English and Italian (Silveri *et al.*, 2002; Forbes-McKay *et al.*, 2005; Rodríguez-Ferreiro *et al.*, 2009; Cuetos, Herrera & Ellis, 2010), that is the words produced by e-D patients have a lower age of acquisition value than those produced by MCI or HC. Concerning word frequency value, no significant difference was found between words produced by these different populations.

Key words: Early-stage dementia, Alzheimer's disease, Mild Cognitive Impairment, age of acquisition, spontaneous speech.

1. *Introduction*

Among the heterogeneity of symptoms related to dementia, language impairments are already present at the very early stage of cognitive decline (Taler, Phillips, 2008; Olney, Spina & Miller, 2017). These deficits vary along the course of the disease. Moreover, in the initial phase there is a prominent decline at lexical-semantic level, whereas phonological and syntactic abilities are relatively well preserved (Caramelli, Mansur & Nitrini, 1998). Because of their involvement in the preclinical phase of the disease, the deficits affecting the linguistic abilities can be used as clues for early diagnosis and dementia large-scale screenings. A number of studies, also based on the new sophisticated techniques from Natural Language Processing (NLP), have already demonstrated that linguistic features can be used for detecting and classifying dementia prodroms (Snowdon, Greiner & Markesbery, 2000; Chapman *et al.*, 2002; Jarrold *et al.*, 2010; Beltrami *et al.*, 2016).

The word retrieval impairment is one of the deficits affecting the language abilities in patients with dementia and it seems to follow the rule of "last-in, first-out" (i.e. the words acquired later are more vulnerable to cognitive decline). This process is also an effect of aging (Hodgson, Ellis, 1998), but it appears more severe in the cognitive decline (Holmes, Fitch & Ellis, 2006). For this reason, several studies have successfully used age of acquisition of words (i.e. the age at which a word is acquired)

to assess the severity of cognitive decline (Silveri *et al.*, 2002; Forbes-McKay *et al.*, 2005; Rodríguez-Ferreiro *et al.*, 2009; Cuetos, Herrera & Ellis 2010). Furthermore, most of these studies have revealed that age of acquisition has an effect independent of word frequency (Morrison, Ellis, 1995; Räling, Schröder & Wartenburger, 2016). Such experiments usually rely on verbal fluency or picture naming tasks and use two different values of AoA: a) objective (directly from children) and b) subjective (by adult rating).

In this paper, we present a pilot study aimed to evaluate the role of AoA in the speech production of 48 subjects belonging to three different populations (early-stage dementia-e-D, Mild Cognitive Impairment – MCI and healthy control – HC). To this end, we analyzed the spontaneous speech of subjects in a semi-automatic way. Notwithstanding the spontaneous speech has a more intra- and inter-personal variability compared to structured evaluations like verbal fluency task, it allows a more naturalistic assessment of language abilities (Bucks *et al.*, 2000).

We expected that the subjects affected by cognitive decline would have produced a higher number of words with a lower value of AoA and that the more severe the impairment, the more evident the trend. Furthermore, we wanted to determine if the AoA value has an effect independent of word frequency.

2. *Data collection*

The sample was composed of 96 subjects (48 male and 48 female, age range 50-75, mean 63.3, SD 7.2): 48 healthy controls and 48 with cognitive decline previously evaluated with medical and neuropsychological assessment. The group with cognitive decline was composed of 32 with MCI and 16 with e-D. All subjects were native Italian speakers. Each subject underwent a neuropsychological screening composed of some of those tests considered the most reliable for the diagnosis of MCI or dementia (Grober *et al.*, 2008; Ismail, Raji & Shulman 2010): Mini Mental State Examination – MMSE, Montreal Cognitive Assessment – MoCA, General Practitioners assessment of Cognition – GPCog, Clock Drawing Test – CDT, Verbal fluency (phonemic and semantic) and the Paired Associate Learning (PAL, subtest of the Cambridge Neuropsychological Test Automated Battery – CANTAB).

The spontaneous productions of subjects were collected during the execution of three tasks, elicited by these input sentences:

a) "Describe this picture";
b) "Describe your typical working day";
c) "Describe the last dream you had or remember".

This data was collected, transcribed and POS-tagged under the OPLON project ("OPportunities for active and healthy LONgevity", Smart Cities and Community – DD 391/RIC, co-funded by Ministry of Education) (see Beltrami *et al.*, 2016 for a more detailed description of data collection procedures).

3. *Data analysis*

To assign an AoA value to the words produced by the subjects in the three different tasks we used two corpora available online:

– *AoAObj*. It contains 223 Italian nouns acquired by children up to 11 years old. The AoA value was objective, that is obtained directly from the children (Lotto, Surian & Job, 2010);
– *AoASubj*. It contains626 Italian nouns acquired by children up to 13 years old. The AoA value was subjective, that is obtained by adult rating (Barca *et al.*, 2002).

The two corpora were different in their composition[1]. They shared 79 tokens, that is 35% of the total of AoAObj and 12% of the total of AoASubj. Moreover, 79% of the lemmas of AoAObj and 99% of AoASubj were drawn from *Vocabolario di Base della Lingua Italiana* (De Mauro, 1980).

Table 1 - *Means number of the words and nouns produced by the subjects in each task and the nouns/words and matched-nouns/nouns ratio grouped by task, group and corpora*

Corpora	*Group*	*Task*	*Tot Words (mean)*	*Tot Nouns (mean)*	*Nouns / Tot Words (%)*	*Matched Nouns / Tot Nouns (%)*
AoAObj	CON	Picture	64.48	32.90	51.03	14.13
AoAObj	CON	Work	89.71	35.19	39.23	8.13
AoAObj	MCI	Picture	44.95	26.16	58.20	15.90
AoAObj	MCI	Work	67.00	26.26	39.20	9.82
AoAObj	e-D	Picture	29.50	16.63	56.36	18.83
AoAObj	e-D	Work	35.00	12.50	35.71	22.00
AoASubj	CON	Picture	64.48	32.90	51.03	37.05
AoASubj	CON	Work	89.71	35.19	39.23	29.89
AoASubj	MCI	Picture	44.95	26.16	58.20	41.06
AoASubj	MCI	Work	67.00	26.26	39.20	30.46
AoASubj	e-D	Picture	29.50	16.63	56.36	41.38
AoASubj	e-D	Work	35.00	12.50	35.71	36.56

After a preliminary analysis, the task "dream" was excluded from further analysis, due to low variation of lexicon.

We automatically matched the nouns contained in the corpora with the lemmas extracted from each subject production by using a script made in Python. The Table 1 shows the average number (in percentage) of the nouns mapped by this procedure,

[1] As regards to *AoAObj* corpus, in order to obtain an age of acquisition value for each noun, a set of 223 drawings were presented to a group of children aged 2-11 years and split in bands (6-months age bands for children up to 6 years old and 12-months band for children up to 11 years old). The AoA value assigned to a noun were the median value of the youngest age-level group able to reach the 75% of success in naming the drawings. The age of acquisition values contained in the *AoASubj* corpus were obtained from subjective adult ratings (on a 7-point scale) by presenting to the subjects 626 nouns printed in a booklet.

namely the nouns produced by the subjects for which we had an AoA value. The AoA and word frequency values were then given to the nouns thus obtained and for each subject and task we calculated the mean of the AoA and word frequency values. We grouped the data by task and then we performed a Kolmogorov-Smirnov non-parametric test to assess the statistical significance (p-value < 0.05) of the AoA and word frequency features.

4. *Results*

The results presented here are from a subset of 48 subjects (21 CON, 19 MCI, 8 e-D) balanced by sex and age (range 49-75, mean 62, SD 7.1). All subjects spoke Italian as their first language.

Figure 1 - *Distribution of the AoA values for the matched nouns of the task "work" using the AoAObj Corpus*

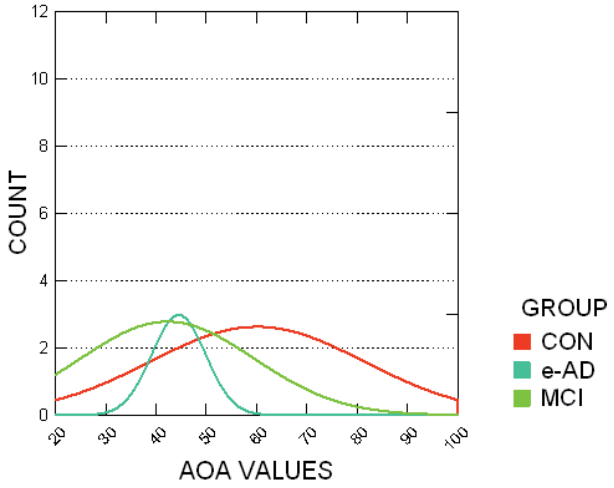


Figure 1 shows the distribution of AoA values assigned to the nouns matched using the AoAObj corpus in the task "work". The Kolmogorov-Smirnov nonparametric test was performed on AoA value with groups as a factor. A significant difference was found between the control group and the e-D group (d = 0.71; p-value = 0.042) in the task "work". A word frequency value was also assigned to the matched nouns and the Kolmogorov-Smirnov non-parametric test was also performed, but no significant difference was found among groups.

Figures 2 and 3 show the distribution of AoA values assigned to the nouns matched using the AoASubj corpus in the tasks "picture" and "work" respectively.

Figure 2 - *Distribution of the AoA values for the matched nouns of the task "picture" using the AoASubj Corpus*

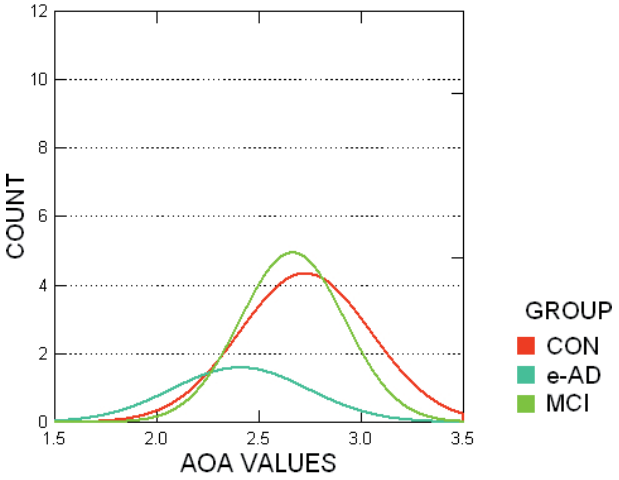
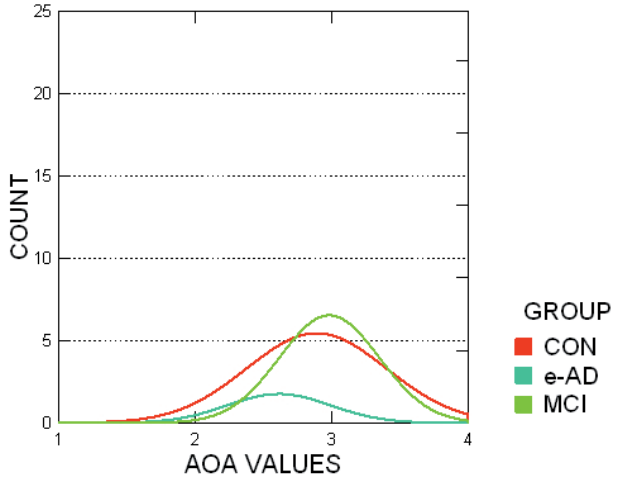


Figure 3 - *Distribution of the AoA values for the matched nouns of the task "work" using the AoASubj Corpus*



The Kolmogorov-Smirnov nonparametric test shows that AoA value can significantly differentiate the control group from e-D group (d = 0.57; p-value = 0.027) in the task "picture". Another significant difference was found between the MCI and e-D groups (d = 0.63; p-value = 0.048) in the task "work". Finally, a word frequency value was also assigned to the matched nouns and the Kolmogorov-Smirnov non-parametric test also performed, but no significant different was found among groups.

5. *Discussion and Conclusion*

This preliminary study seems to confirm for Italian the tendencies found in previous studies for English. In fact, using the AoAObj corpus we were able to differentiate the control group from the e-D group in the task "work" and, as expected, the control group had greater AoA values than the e-D group. A similar trend was found using the AoASubj corpus as the control group had greater AoA values than the e-D group in the task "picture" while the MCI group had greater AoA values than the e-D group in the task "work". No significant differences were found between MCI and control group in any tasks.

The choice of analyzing only the nouns is supported by the literature: it seems that the ability to refer to action (verbs) is relatively well preserved compared to object naming performance (nouns) (Williamson, Adair, Raymer & Hellman, 1998; Fung *et al.*, 2001).

In trying to interpret these results, we have to take into account some limitations of the experimental design. First, this kind of experiment generally relies on the verbal fluency task because spontaneous speech has more intra-personal variability. Moreover, the corpora used to match the nouns and to assign them an AoA value were too small (223 and 626 tokens respectively); indeed, the rate of mapped nouns was low. Finally, given that the years of education are significantly lower in the e-D group than in the CON and MCI groups, our results may have been confounded by this variable. This limit did not interfere with the neuropsychological results since most of the cognitive tests were adjusted for years and education according to the respective standardizations.

The two corpora differ in their composition (that is the percentage of nouns drawn from Vocabolario di Base della Lingua Italiana) and in the kind of AoA value assigned to the lemmas (objective vs. subjective). Comparing the two different corpora is out of the aim of this work. We can just assert, in line with the literature (Morrison, Chappel & Ellis, 1997; Lotto *et al.,* 2010), that the adult rating seems to be a reliable method to estimate the AoA value of acquired words. Due to these restrictions, this study must be considered as a pilot study aiming to assess the feasibility of utilizing available corpora to automate the analysis of a large sets of spontaneous speech samples. Nevertheless, some conclusions can be drawn.

The cognitive decline seems to be accompanied by loss of the words acquired later in life. In this process, the word frequency has no role and this could be due to the different origin of AoA and word frequency effects. In fact, AoA correlates highly with semantic variables (for example imageability or concreteness), but less with lexical variables such as word frequency (Morrison *et al.*, 1997; Räling *et al.*, 2016). So, it is likely that in the earliest phases of cognitive decline due to dementia the brain areas underlying the semantic process are more impaired than the areas responsible for lexical process (Taler, Phillips, 2008). Semantic impairment is also well documented in literature for MCI. So, the lack of statistical significance between MCI and control groups is more difficult to be accounted for. Given the heterogeneity in MCI populations and his fluctuating cognitive dysfunction (Feldam,

Jacova, 2005), it is possible that analysis by group fails to pinpoint a common, subtle deficit. Another possible explanation may arise from the fact that the data on age of acquisition value drawn from spontaneous speech fails to capture the difference between intentional and automatic access to semantic memory. In fact, while it is well documented that even in the early-stage dementias like Alzheimer's disease or FTD both intentional and automatic access are impaired (Taler, Phillips, 2008; Olney, Spina & Miller, 2017), it seems that in Mild Cognitive Impairment only automatic access is inhibited (Duong *et al.*, 2006; Taler, Phillips, 2008).

A further study aimed to correlate type and duration of hesitation phenomena with the age of acquisition value of the nouns involved in the hesitations itself could enlighten the nature of deficit in semantic access in Mild Cognitive Impairment.

Bibliography

Barca, L., Burani, C. & Arduino, L.S. (2002). Word naming times and psycholinguistic norms for Italian nouns. In *Behavior Research Methods, Instruments, & Computers*, 34, 424-434.

Beltrami, D., Calzà, L., Gagliardi, G., Ghidoni, E., Marcello, N., Rossini Favretti, R. & Tamburini, F. (2016). Automatic identification of Mild Cognitive Impairment through the analysis of Italian spontaneous speech productions. In Calzolari, N., Choukri, K., Declerck, T., Goggi, S., Grobelnik, M., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J. & Piperidis, S. (Eds.), *Proceedings of the 10th International Conference on Language Resources and Evaluation*. Portorož, Slovenia, 2086-2093.

Bucks, R.S., Singh, S., Cuerden, J.M. & Wilcock, G.K. (2000). Analysis of spontaneous, conversational speech in dementia of Alzheimer type: Evaluation of an objective technique for analysing lexical performance. In *Aphasiology*, 14(1), 71-91.

Caramelli, P., Mansur, L. & Nitrini, R. (1998). Language and communication disorders in dementia of the Alzheimer type. In Stemmer, B., Whitaker, H.A. (Eds.), *Handbook of neurolinguistics*. San Diego: Academic, 463-473.

Chapman, S.B., Zientz, J., Weiner, M.F., Rosenberg, R.N., Frawley, W.H. & Burns, M.H. (2002). Discourse changes in early Alzheimer disease, Mild Cognitive Impairment, and normal aging. In *Alzheimer Disease & Associated Disorders*, 16(3), 177-186.

Cuetos, F., Herrera, E. & Ellis, A.W. (2010). Impaired word recognition in Alzheimer's disease: The role of age of acquisition. In *Neuropsychologia*, 48, 3329-3334.

De Mauro, T. (1980). *Guida all'uso delle parole*. Roma: Editori Riuniti.

Duong, A., Whitehead, V., Hanratty, K. & Chertkow, H. (2006). The nature of lexico-semantic processing deficits in mild cognitive impairment. In *Neuropsychologia*, 44, 1928-1935.

Forbes-Mackay, K.E., Ellis, A.W., Shanks, M.F. & Venneri, A. (2005). The age of acquisition of word produced in a semantic fluency task is highly predictive of early Alzheimer's disease. In *Neuropsychologia*, 43, 1128-1137.

Fung, T.D., Chertkow, H., Murtha, S., Whatmough, C., Peloquin, L., Whitehead, V. & Templeman, F.D. (2001). The spectrum of category effects in object and action knowledge in dementia of the Alzheimer's type. In *Neuropsychology*, 15, 371-379.

Grober, E., Hall, C.H., Lipton, R.B., Zonderman, A.B., Resnick, S.M. & Kawas, C. (2008). Memory impairment, executive dysfunction, and intellectual decline in preclinical Alzheimer's disease. In *Journal of the International Neuropsychological Society*, 14(2), 266-278.

Hodgson, C., Ellis, A.W. (1998). Last in, first to go: Age of acquisition and naming in the elderly. In *Brain and Language*, 64, 146-163.

Holmes, S.J., Fitch, F.J. & Ellis, A.W. (2006). Age of acquisition affects object recognition and naming in patients with Alzheimer's disease. In *Journal Clinical and Experimental Neuropsychology*, 28, 1010-1022.

Ismail, Z., Rajji, T.K., Shulman, K.I. (2010). Brief cognitive screening instruments: an update. In *International Journal of Geriatric Psychiatry*, 25(2), 111-120.

Jarrold, W.L., Peintner, B., Yeh, E., Krasnow, R., Javitz, H.S. & Swan, G.E. (2010). Language Analytics for Assessing Brain Health: Cognitive Impairment, Depression and Pre-symptomatic Alzheimer's Disease. In Yao, Y., Sun, R., Poggio, T., Liu, J., Zhong, N. & Huang, J. (Eds.), *Proceedings of Brain Informatics: International Conference*. Toronto, Canada, 299-307.

Lotto, L., Surian, L. & Job, R. (2010). Objective Age of Acquisition for 223 Italian Words: Norms and Effects on Picture Naming Speed. In *Behavior Research Methods*, 42, 123-133.

Morrison, C.M., Ellis, A.W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. In *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 21, 116-133.

Morrison, C.M., Chappell, T.D. & Ellis, A.W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. In *Quarterly Journal of Experimental Psychology*, 50A, 528-559.

Olney, N.T., Spina, S. & Miller, B.L. (2017). Frontotemporal Dementia. In *Neurologic Clinics*, 35, 339-374.

Räling, R., Schröder, A. & Wartenburger, I. (2016). The origin of age of acquisition and typicality effects: Semantic processing in aphasia and the ageing brain. In *Neuropsychologia*, 86, 80-92.

Rodríguez-Ferreiro, J., Davies, R., González-Nosti, M., Barbón, A. & Cuetos, F. (2009). Name agreement, frequency and age of acquisition, but not grammatical class, affect object and action naming in Spanish speaking participants with Alzheimer's disease. In *Journal of Neurolinguistics*, 22, 37-54.

Silveri, M.C., Cappa, A., Mariotti, P. & Puopolo, M. (2002). Naming in patients with Alzheimer's disease: influence of age of acquisition and categorical effects. In *Journal of Clinical and Experimental Neuropsychology*, 24, 755-764.

Snowdon, D.A., Greiner, L.H. & Markesbery, W.R. (2000). Linguistic Ability in Early Life and the Neuropathology of Alzheimer's Disease and Cerebrovascular Disease: Findings from the Nun Study. In *Annals of the New York Academy of Sciences*, 903, 34-38.

Taler, V., Phillips, N.A. (2008). Language performance in Alzheimer's disease and mild cognitive impairment: a comparative review. In *Journal of Clinical and Experimental Neuropsychology*, 305(5), 501-556.

Williamson, D.J., Adair, J.C., Raymer, A.M. & Hellman, K.M. (1998). Object and action naming in Alzheimer's disease. In *Cortex*, 34, 601-610.