Automatic Identification of Mild Cognitive Impairment through the Analysis of Italian Spontaneous Speech Productions

Daniela Beltrami[1,2], Laura Calzà[1], Gloria Gagliardi[1], Enrico Ghidoni[2], Norina Marcello[2], Rema Rossini Favretti[1], Fabio Tamburini[1]
[1] Alma Mater Studiorum - Università di Bologna
[2] Arcispedale S.Maria Nuova, Reggio Emilia
E-mail: daniela.beltrami3@unibo.it, laura.calza@unibo.it, gloria.gagliardi2@unibo.it, Enrico.Ghidoni@asmn.re.it, Norina.Marcello@asmn.re.it, rema.rossini@unibo.it, fabio.tamburini@unibo.it

Abstract

This paper presents some preliminary results of the OPLON project. It aimed at identifying early linguistic symptoms of cognitive decline in the elderly. This pilot study was conducted on a corpus composed of spontaneous speech sample collected from 39 subjects, who underwent a neuropsychological screening for visuo-spatial abilities, memory, language, executive functions and attention. A rich set of linguistic features was extracted from the digitalised utterances (at phonetic, suprasegmental, lexical, morphological and syntactic levels) and the statistical significance in pinpointing the pathological process was measured. Our results show remarkable trends for what concerns both the linguistic traits selection and the automatic classifiers building.

Keywords: Pathological language, Mild Cognitive Impairment, Linguistic Features, Automatic classifiers.

1. Background

This research is part of the OPLON project ("OPportunities for active and healthy LONgevity", Smart Cities and Communities – DD 391/RIC, co-funded by the Ministry of Education as part of the Contract "Smart Cities and Communities and Social Innovation"). The project intends to propose actions and methods to prevent fragility and decline and promote the health of the elderly, designing and developing tools and networks of early diagnosis and "care & cure". Given this general project framework, the prevention of the various types of dementia appears to be one of the most challenging, nonetheless most pressing, tasks (Calzà et al., 2015).

Individuals with preclinical dementia manifest alterations in various cognitive domains: a number of longitudinal retrospective studies have already demonstrated that linguistic features could act as a prodromic marker of cognitive dysfunctions: for example, the Nun study (Snowdon, 2003), the Iris Murdoch study (Garrard et al., 2005) or the Harold Wilson project (Garrard, 2009). Deficits are seen in verbal fluency, naming and semantic knowledge (Taler & Phillips, 2008); it is also well documented that discourse alterations may be one of the earliest signs of the pathology, often measurable years before other cognitive deficits become apparent (Caramelli et al., 1998). Looking at the literature on this topic, syntactic and phonological abilities seem to be relatively preserved, even though individuals produce semantically impoverished discourse that lacks in coherence.

These linguistic complaints are definitely concomitant with neuropathological alterations and clinical manifestation, but also recognizable in the presymptomatic phases of the cognitive impairment. The investigation of this domain seems to be promising, both for early diagnosis and dementia large-scale screenings.

During the last few years, the development of new sophisticated techniques from Natural Language Processing (NLP) have been used to analyse written texts, clinically elicited utterances and spontaneous production, in order to identify signs of psychiatric or neurological disorders and to extract automatically derived linguistic features for pathologies recognition, classification and description. Computational methods have been already successfully applied to the study of linguistic cues of cerebral functional disorders: not only in the case of language disruption associated with focal brain lesions, but also for detecting dementia prodroms (Mild Cognitive Impairment) and sub-types, like Alzheimer's Disease and Fronto-Temporal Lobar Degeneration (Chapman *et al.* 2002; Peintner *et al.* 2008; Jarrold *et al.* 2010; Roark *et al.* 2011; Lehr, 2012; Satt *et al.* 2013; Fraser *et al.* 2014; Toth *et al.* 2015).

While neuropsychological tests and structured evaluations have a relevant impact on the naturalness of the subject's responses (Bucks *et al.* 2000), the analysis of spoken language productions allows to ecologically and inexpensively pinpoint language modifications in potential patients even by primary care physicians.

Inside the OPLON framework, we are working to build methods to identify cognitive frailty at very early stage by processing spontaneous language productions of Italian speakers. This instrument will be developed to be used at General Practitioner level, for frequent, low-cost and non-intrusive cognitive decline screening and cognitive status monitoring.

At the time of writing, we are not aware of any study specifically devoted to Italian performing a similar kind of automatic analysis: therefore the goal in the short-time is to test the feasibility of this approach in a controlled environment.

2. Data collection

In the whole project we plan to enrol 96 subjects: 48 healthy controls (CON) and 48 subjects with cognitive decline. The sample will be balanced by sex, age (range 50-75) and education (primary school with great intellectual stimulation throughout the life span or junior high school; high school; academic degree).

The cognitive decline refers to two categories:

1. Mild Cognitive Impairment (MCI): it causes cognitive changes that are serious enough to be assessed with neuropsychological assessment, but not severe enough to interfere with everyday activities
   a. amnestic MCI single domain (a-MCI; 16 subjects): patients who show an isolated memory deficit;
   b. multiple domain MCI (md-MCI; 16 subjects): in these individuals two or more cognitive abilities are affected (memory can be engaged or not).
2. Early Dementia (e-D; 16 subjects): these patients are affected by cognitive deficits which partially influence everyday life (however, their Mini Mental State Examination score is equal or greater than 18).

Each subject will undergo a brief neuropsychological screening composed of those traditional tests which seem to be the most sensitive to distinguish between normal subjects and people affected by MCI or dementia (Grober *et al*. 2008; Ismail *et al*. 2010; Velayuhan *et al*. 2014; Tsoi *et al*. 2015): Mini Mental State Examination – MMSE (Folstein *et al*. 1975; Measso *et al*. 1993), Montreal Cognitive Assessment – MoCA (Nasreddine *et al*. 2005; Conti *et al*. 2015), General Practitioners assessment of Cognition – GPCog (Brodaty *et al*. 2002; Pirani *et al*. 2010), Clock Drawing Test – CDT (Freedman *et al*. 1994; Mondini *et al*. 2011), Verbal fluency (phonemic and semantic; Carlesimo *et al*. 1995; Novelli *et al*., 1986). The subjects will also experience the Paired Associate Learning – PAL (subtest of the Cambridge Neuropsychological Test Automated Battery – CANTAB) which seems to be very accurate to detect the very early signs of cognitive decline (Fowler *et al*. 2002; Swainson *et al*. 2001; Blackwell *et al*. 2004; De Jager *et al*. 2005)

These tools measure those abilities that seem to be critical for an early diagnosis of cognitive decline (memory, executive functions, verbal and visuospatial abilities, attention and orientation) and form the base tools for subject classification by the neuropsychologist into one of the three considered classes (CON, MCI, e-D).

After the traditional neuropsychological assessment, we will record the spontaneous speech of the subjects during the execution of three tasks, elicited by these input sentences:

- “Describe this picture” (Ciurli *et al*., 1996);
- “Describe your typical working day”;
- “Describe the last dream you remember”.

This paper presents a pilot, but in our opinion already significant, study on 39 subjects restricting the comparison between controls (20) and MCI subjects (19); distinguishing between these two subject classes is one of the basic goals for the entire project framework.

3. Data analysis

Spontaneous speech samples are recorded in WAV files (44.1KHz, 16 bit) during test sessions. The transcriptions were produced manually from the interviews by using the Transcriber [1] software package. We chose the utterance as the processing unit, defined by using prosodic (mainly intonational) criteria. During the transcription process we annotated also a series of paralinguistic phenomena such as pauses, disfluences, lapsus, etc.

All the utterances were automatically PoS-tagged and syntactically parsed with the dependency model used by the Turin University Linguistic Environment – TULE (Lesmo, 2007), based on the TUT - Turin University TreeBank tagset (Bosco *et al.* 2000) in order to explicit all the morphological, syntactic and lexical information about texts and they were manually checked to remove all the errors introduced by the automatic tagging procedures. In this pilot study we decided to rely on carefully checked linguistic information, at least for transcription-derived features, to avoid any type of interference due to tagging errors.

With regard to the parameters derived from the speech acoustics, we used the “ssvad” Voice Activity Detector proposed by (Mak, Yu, 2014), especially developed for interview speech, to segment the recordings and identify speech vs non-speech regions, and the forced alignment system belonging to the Kaldi-DNN-ASR package[2], trained on the APASCI Italian Corpus (Angelini *et al.* 1994), for obtaining the temporally aligned phonetic transcriptions needed to compute various rhythmic features.

A multidimensional parameter computation was performed: the system conducts a quantitative analysis of spoken texts, computing rhythmic, acoustic, lexical, morpho-syntactic and syntactic features.

Both linguistic/stylometric indexes proposed in the literature and some new parameters are tested. Table 1 outlines the complete list of the features considered in this study.

Statistically relevant features will be the input for a Machine Learning (ML) classifier. The performance achieved by the system will be evaluated in terms of four metrics: accuracy, precision, recall and F-measure.

[1] http://trans.sourceforge.net
[2] http://kaldi.sourceforge.net/about.html

ACOUSTIC FEATURES		
Description	**Label**	**References**
Silence segments duration: mean, median and Std. Dev.	SPE_SILMEAN SPE_SILMEDIAN SPE_SILSD	(Satt *et al*., 2012; Satt *et al.,* 2013)
Speech segments duration: mean, median and Std. Dev.	SPE_SPEMEAN SPE_SPEMEDIAN SPE_SPESD	(Satt *et al*., 2012; Satt *et al.,* 2013)
Temporal regularity of voiced segment durations	SPE_TRVSD	(Satt *et al*., 2012; Satt *et al.* 2013)
Verbal Rate	SPE_VR	(Singh *et al*., 2001; Roark *et al*., 2007a; Roark *et al*., 2011)
Transformed Phonation Rate	SPE_TPR	(Singh *et al*., 2001; Roark *et al*., 2011)
Standardized Phonation Time	SPE_SPT	(Singh *et al.,* 2001; Roark *et al*., 2011)
Standardized Pause Rate	SPE_SPR	(Singh *et al*., 2001; Roark *et al*., 2007a; Roark *et al*., 2011)
Root Mean Square energy: mean and Std. Dev.	SPE_RMSEM SPE_RMSESD	(López-de-Ipiña *et al*., 2013)
Pitch: mean and Std. Dev.	SPE_PITCHM SPE_PITCHSD	(López-de-Ipiña *et al*., 2013)
Spectral Centroid: mean and Std. Dev.	SPE_SPCENTRM SPE_SPCENTRSD	(López-de-Ipiña *et al*., 2013)
Higuchi Fractal Dimension: mean and Std. Dev.	SPE_HFractDM SPE_HFractDSD	(López-de-Ipiña *et al*., 2013)
RHYTHMIC FEATURES		
Percentage of vocalic intervals	RHY_V	(Ramus *et al*., 2009)
Std. Dev. of vocalic and consonantal intervals	RHY_DeltaV RHY_DeltaC	(Ramus *et al*., 2009)
Pairwise Variability Index, raw and normalized	RHY_VnPVI RHY_CrPVI	(Grabe & Low*,* 2002)
Variation coefficient for ΔV and ΔC	RHY_VarcoV RHY_VarcoC	(Dellwo, 2006)
LEXICAL FEATURES		
Content Density	LEX_ContDens	(Roark *et al*., 2011)
Part-of-Speech rate	LEX_PoS	(Holmes & Singh, 1996; Bucks *et al.*, 2000; Vigorelli, 2004; Garrard *et al.*, 2005; Thomas et al., 2005; Peintner *et al.*, 2008; Cantos-Gomez *et al.*, 2009; Jarrold *et al.*, 2010; Alegria et al., 2013; Jarrold *et al.*, 2014)
Reference Rate to Reality	LEX_RefRReal	(Vigorelli, 2004)
Personal, Spatial and Temporal Deixis rate	LEX_PDEIXIS LEX_SDEIXIS LEX_TDEIXIS	(March *et al.*, 2006; Cantos-Gomez *et al.*, 2009)
Relatives pronouns and negative adverbs rate	LEX_RPNA	
Lexical Richness: Type-Token Ratio, W - Brunét's Index and R - Honoré's Statistic	LEX_TTR, LEX_BrunetW LEX_HonoreR	(Brunét, 1978; Honoré, 1979; Holmes, 1992; Holmes & Singh, 1996; Bucks *et al.*, 2000; Thomas et al., 2005)
Action Verbs rate	LEX_ACTVRB	(Gagliardi, 2014)
Frequency-of-use tagging (De Mauro/Paravia dictionary)	LEX_DM_F	(De Mauro, 1980; De Mauro, 2000; Barbagli *et al.*, 2014)
Propositional Idea Density	LEX_IDEAD	(Snowdon *et al.*, 1996; Brown *et al.*, 2008; Jarrold *et al.*, 2010; Roark *et al.*, 2011)

SYNTACTIC FEATURES		
Description	**Label**	**References**
Number of dependent elements linked to the noun, mean and Std. Dev.	SYN_NPLENM SYN_NPLENSD	
Global Dependency Distance, mean and Std. Dev.	SYN_GRAPHDISTM SYN_GRAPHDISTSD	(Roark *et al.*, 2007b; Roark *et al.*, 2011)
Syntactic complexity	SYN_ISynCompl	(Szmrecsanyi, 2004)
Syntactic embeddedness: maximum depth of the structure, mean and Std. Dev.	SYN_MAXDEPTHM SYN_MAXDEPTHSD	
Utterance length, mean and Std. Dev.	SYN_SLENM SYN_SLENSD	

Table 1: List of all the features considered in this study.

4. Experiments and results

Statistical significance (p-value < 0.05) of the features is assessed by using Kolmogorov–Smirnov nonparametric test. We chose such kind of hypothesis testing technique, compared with the T-test or the Wilcoxon-Mann-Whitney test, because of the small size of our corpus.
For each linguistic task, the features having the KS p-value < 0.10 are used as input data for three automatic classifiers available in the Orange Data Mining tool[3] (kNN 3-neighbourgs, Logistic Regression and Neural Network classifiers). The training/test sets are automatically built by the package by random sampling the entire dataset (ratio between training/test sets = 80/20%), repeating this procedure 20 times.
The statistically relevant features and the classifier performances are summarized in Table 2 for the three different tasks and in Table 3 for all tasks data together.

5. Discussion and conclusions

We are aware that building automatic classifiers using machine learning techniques with such a small amount of data may be dangerous, but we think that some provisional conclusions can indeed be drawn observing these preliminary results.
First of all, the quite good results in classification performances demonstrate that language can play a relevant role in the analysis of cognitive alterations.
Second, we tested the strength of the proposed methodology and, despite the limited dataset, the experiments pinpointed some linguistic features discriminating healthy subjects and MCI patients with a high statistical level of significance.
Looking at the most promising features in the large dataset we considered in this study, it seems that speech features are generally more reliable in distinguishing controls from MCI subjects. In particular Spectral Centroid mean (SPE_SPCENTRM) and the statistics about speech and silence duration intervals are consistently present as significant features in all tasks. Different lexical and syntactic features plays a role in the various tasks: in particular those measuring the complexity of speech production help to mark the difference between subject groups. Rhythmical features seem not to be so relevant for the studied task.
According to the literature, people presenting a progressive decline in mental abilities showed a subtle linguistic impairment even in the pre-symptomatic stages of the disease. These deficits can be successfully detected using NLP techniques. However, all these approaches are usually developed and trained on well-formed, written texts. Although pathologic language can present some hardships for these algorithms, nowadays automatic systems are sufficiently reliable for these tasks, being already able to distinguish between healthy control and patients with a fair degree of accuracy if properly set up (Roark et al. 2011). Nevertheless more work is needed to adapt these systems to adequately analyse pathologic language, increasing the overall classification performances.
At the time of writing we are finishing the collection of the whole 96 subject's interviews and their manual processing. Future works regard an in depth analysis of the whole corpus verifying the findings presented in this paper and enlarging the analysis adding more features. Moreover, we will compare the obtained results with a completely automatic interview processing (ASR, PoS-tagger, dependency parser and ML classifier) in order to build and evaluate a complete self-contained application to be distributed to General Practitioners in order to perform large-scale screenings.

[3]http://orange.biolab.si/

Task “Picture”		
		KS test p-values
Significant features	LEX_IDEAD SPE_SILSD LEX_PoS_ADJ LEX_ContDens LEX_CCW LEX_OCW SPE_TPR SPE_SILMEAN SPE_SILMEDIAN SPE_HFractDSD SPE_SPCENTRM	p = 0.046395 p = 0.044344 p = 0.040478 p = 0.022891 p = 0.022891 p = 0.022891 p = 0.021790 p = 0.019730 p = 0.007688 p = 0.006161 p = 0.000648
ML classifiers perf.	kNN : Accuracy = 0.692, Precision = 0.733, Recall = 0.579, F1= 0.647 LogR: Accuracy = 0.769, Precision = 0.727, Recall = 0.842, F1= **0.781** NeuN: Accuracy = 0.769, Precision = 0.727, Recall = 0.842, F1= **0.781**	

Task “Working Day”		
		KS test p-values
Significant features	SPE_SPEMEAN LEX_PoS_PREDET SYN_MAXDEPTHM SPE_HFractDSD SYN_ISynCompl SYN_GRAPHDISTM SYN_SLENM SPE_HFractDM SPE_SPESD LEX_PoS_INTERJ SPE_SPCENTRM	p = 0.048527 p = 0.048527 p = 0.044344 p = 0.040478 p = 0.021790 p = 0.019730 p = 0.019730 p = 0.016965 p = 0.016965 p = 0.007688 p = 0.002030
ML classifiers perf.	kNN : Accuracy = 0.619, Precision = 0.634, Recall = 0.562, F1= 0.596 LogR: Accuracy = 0.725, Precision = 0.765, Recall = 0.650, F1= **0.703** NeuN: Accuracy = 0.719, Precision = 0.769, Recall = 0.625, F1= 0.690	

Task “Dream”		
		KS test p-values
Significant features	SPE_SPCENTRM SPE_SILSD SPE_SPT SPE_SILMEDIAN SPE_TPR SPE_SPESD LEX_PoS_VERB SPE_SPR SPE_SPEMEDIAN SPE_SPEMEAN	p = 0.046395 p = 0.046395 p = 0.024040 p = 0.019730 p = 0.018767 p = 0.009050 p = 0.006886 p = 0.002030 p = 0.001910 p = 0.000531
ML classifiers perf.	kNN : Accuracy = 0.712, Precision = 0.736, Recall = 0.662, F1= 0.697 LogR: Accuracy = 0.750, Precision = 0.738, Recall = 0.775, F1= **0.756** NeuN: Accuracy = 0.743, Precision = 0.767, Recall = 0.700, F1= 0.744	

Table 2: Statistically significant features (Komolgorov-Smirnov test) and automatic classifiers performances for the different tasks considered in this study.

All tasks data together		
		KS test p-values
Significant features	LEX_PoS_VERB	p = 0.028823
	SYN_SLENM	p = 0.014911
	SPE_VR	p = 0.012840
	SYN_GRAPHDISTM	p = 0.004522
	SPE_RMSEM	p = 0.003460
	SPE_SPT	p = 0.001161
	SPE_HFractDM	p = 0.000508
	SPE_SPEMEDIAN	p = 0.000418
	SPE_SPR	p = 0.000330
	SPE_HFractDSD	p = 0.000196
	SPE_SILMEAN	p = 0.000089
	SPE_SPEMEAN	p = 0.000066
	SPE_SILSD	p = 0.000066
	SPE_SPESD	p = 0.000058
	SPE_TPR	p = 0.000041
	SPE_SILMEDIAN	p = 0.000016
	SPE_SPCENTRM	p = 0.000000
ML classifier perf.	kNN : Accuracy = 0.721, Precision = 0.727, Recall = 0.708, F1= 0.717	
	LogR: Accuracy = 0.750, Precision = 0.744, Recall = 0.766, F1= 0.753	
	NeuN: Accuracy = 0.760, Precision = 0.767, Recall = 0.754, F1= **0.759**	

Table 3: Statistically significant features (Komolgorov-Smirnov test) and automatic classifiers performances aggregating the different tasks data considered in this study.

6. Bibliographical References

Alegria, R., Gallo, C., Bolso, M., dos Santos, B., Prisco, C.R., Bottino, C., Nogueira, M.I.(2013). Comparative study of the uses of grammatical categories: adjectives, adverbs, pronouns, interjections, conjunctions, and prepositions in patients with Alzheimer's disease. In *Alzheimer's & Dementia: The journal of the Alzheimer's Association*, 9(4): 882.

Angelini, B., Brugnara, F., Falavigna, D., Giuliani, D., Gretter, R., Omologo, M. (1994). Speaker Independent Continuous Speech Recognition Using An Acoustic-Phonetic Italian Corpus, In *Proc. of ICSLP94*, 1391-1394.

Barbagli, A., Lucisano, P., Dell'Orletta, F., Montemagni, S., Venturi, G. (2014). Tecnologie del linguaggio e monitoraggio dell'evoluzione delle abilità di scrittura nella scuola secondaria di primo grado. InBasili, R., Lenci, A., Magnini, B. (eds)*, Proceeding of the First Italian Conference on Computational Linguistics (CLiC-IT 2014)*, 23–27.

Blackwell, A.D., Sahakian, B.J., Vesey, R., Semple, J., Robbins, T.W., Hodges, J.R., (2004). Detecting dementia: novel neuropsychological markers of preclinical Alzheimer's disease. In: *Dementia and Geriatric Cognitive Disorders*, 17, 42-48.

Bosco, C., Lombardo, V., Vassallo, D., Lesmo, L. (2000). Building a Treebank for Italian: a data-driven annotation schema. In *Proc. of LREC-2000*, Athens.

Brodaty, H., Pond, D., Kemp, N.M., Luscombe, G., Harding, L., Berman, K., Hupper, F.A. (2002). The GPCOG: a new screening test for dementia designed for general practice. In: *Journal of the American Geriatrics Society*, 5, 530-534.

Brown, C., Snodgrass, T., Kemper, S.J., Herman, R., Covington, M.A. (2008). Automatic measurement of Propositional Idea Density from Part-of-Speech Tagging. In *Behavior Research Methods*, 40(2): 540–545.

Brunet, E. (1978). *Le Vocabulaire de Jean Giraudoux. Structure et Evolution*. Geneve: Slatkine.

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.

Calzà, L., Beltrami, D., Gagliardi, G., Ghidoni, E., Marcello, N., Rossini Favretti, R., Tamburini, F. (2015). Should we screen for cognitive decline and dementia? In *Maturitas,* 82(1), 28-35.

Cantos-Gòmez, P. (2009). Featuring linguistic decline in Alzheimer's disease: A corpus-based approach. In Mahlberg, M., Gonzàlez Dìaz, V., Smith, C. (eds.), *Proc. of the Corpus Linguistics Conference 2009 (CL2009)*.

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*, 463–473.

Carlesimo, G. A., Caltagirone, C., Fadda, L., Gainotti, G., Gallassi, R., Lorusso, S., Marfia, G., Marra, C., Parnetti, L. (1995). Batteria per la valutazione del deterioramento mentale (parte II): standardizzazione ed affidabilità diagnostica nell'identificazione di pazienti affetti da sindrome demenziale. In: *Archivio di Psicologia, Neurologia e Psichiatria*, 56(4), 471-488.

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.

Ciurli, P., Marangolo, P., Basso, A. (1996). Esame del Linguaggio – II. Giunti Organizzazioni Speciali. Firenze.

Conti, S., Bonazzi, S., Laiacona, M., Masina, M., Vanelli Coralli, M. (2015). Montreal Cognitive Assessment (MoCA) – Italian version: regression based norms and equivalent scores. In: *Neurological Science*, 26, 209-214.

De Jager, C.A., Blackwell, A.D., Budge, M.M., Sahakian, B.J. (2005). Predicting cognitive decline in healthy older adults. In: *American Journal of Geriatric Psychiatry*, 13(8), 735-740.

Delwo, V. (2006). Rhythm and speech rate: a variation coefficient for deltaC. In Karnowski, P., Szigeti, I. (eds.), *Language and Language-Processing*, Frankfurt am Main: Peter Lang, 231-241.

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

De Mauro, T. (2000). *Il dizionario della lingua italiana*. Paravia.

Elvevåg, B., Garrard, P. (2014). Language, Computer and Cognitive Neuroscience, *Cortex*, 55.

Folstein, M. F., Folstein, S., E., McHugh, P. R. (1975) "Mini-Mental state". A practical method for grading the cognitive state of patients for the clinician, *Journal of psychiatric research*, 12(3),189-198

Fowler, K.S., Saling, M.M., Conway, E.L., Semple, J., Louis, W.J., (2002). Paired associate performance in the early detection of DAT. In: *Journal of the International Neuropsychological Society*, 8, 58-71.

Fraser, K.C., Meltzer, J.A., Graham, N.L., Leonard, C., Hirst, G., Black, S.E., Rochon, E. (2014). Automated classification of Primary Progressive Aphasia subtypes from narrative speech transcripts, Cortex, 55, 43-60.

Freedman, M., Leach, L., Kaplan, E. (1994). *Clock-drawing: a neuropsychological analysis*. New York, NY: Oxford University Press.

Gagliardi, G. (2014). *Validazione dell'ontologia dell'azione IMAGACT per lo studio e la diagnosi del Mild Cognitive Impairment*. Tesi di dottorato, Università degli Studi di Firenze.

Garrard, P. (2009). Cognitive archaeology: Uses, methods, and results. *Journal of Neurolinguistics*, 22(3), 250– 265.

Garrard, P., Maloney, L.M., Hodges, J.R., Patterson, K. (2005). The effects of very early Alzheimer's disease on the characteristics of writing by a renowned author. In *Brain*, 128: 250–260.

Grabe, E., Low, E.L. (2002). Durational variability in speech and the rhythm class hypothesis. In: Gussenhoven, C., Warner, N. (eds), *Papers in Laboratory Phonology 7*, Berlino: Mouton de Gruyter, 515-546.

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.

Holmes, D.I. (1992). A stylometric analysis of mormon scripture and related texts. In *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 155(1): 91–120.

Holmes, D.I., Singh, S. (1996). A stylometric analysis of conversational speech of aphasic patients. In *Literary and Linguistic Computing*, 11(3): 133–140.

Honoré, A. (1979). Some simple measures of richness of vocabulary. In *Association of Literary and Linguistic Computing Bulletin*, 7: 172–177.

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. *et al.* (eds.)*, Brain Informatics: International Conference* (BI 2010), 299–307.

Jarrold, W.L., Peintner, B., Wilkins, D., Vergryi, D., Richey, C., Gorno-Tempini, M.L., Ogar, J. (2014). Aided Diagnosis of Dementia Type through Computer-Based Analysis of Spontaneous Speech. In Resnik, P., Resnik,R., Mitchell, M. (eds), in: *Proc. of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, ACL - Association for Computational Linguistics, 27–37.

Lehr, M., Prud'hommeaux, E., Shafran, I., Roark, B. (2012). Fully Automated Neuropsycological Assessment for Detecting Mild Cognitive Impairment. In *Proc. InterSpeech 2012*, Portland (OR).

Lesmo, L. (2007). Il parser basato su regole del Gruppo NLP dell'Università di Torino. In *Intelligenza Artificiale*, IV(2), 46–47.

López-de Ipiña, K., Alonso, J.B., Travieso, C.M., Solé-Casals, J., Egiraun, H., Faundez-Zanuy, M., Ezeiza, A., Barroso, N., Ecay-Torres, M., Martinez-Lage, P., Martinez de Lizardui, U. (2013). On the selection of non-invasive methods based on speech analysis oriented to automatic Alzheimer disease diagnosis. In *Sensors*, 13: 6730–6745.

Mak, M.W. Yu, H.B. (2014). A Study of Voice Activity Detection Techniques for NIST Speaker Recognition Evaluations, In *Comp. Sp. & Lang.*, 28:1, 295-313.

March, E.G., Wales, R., Pattison, P. (2006). The uses of nouns and deixis in discourse production in

Alzheimer's disease. In *Jour. of Neurolinguistics*, 19: 311–340.

Measso, G., Cavarzeran, F., Zappalà, G., Lebowitz, B.D., Crook, T.H., Pirozzolo, F.J., Amaducci, L.A., Massari, D., Grigoletto, F. (1993). The Mini-mental State Examination: Normative Study of an Italian Random Sample. In: *Developmental Neuropsychology*, 9(2), 77-95.

Mondini, S., Mapelli, D., Vestri, A., Bisiacchi, P. S. (2011). *L'esame Neuropsicologico Breve 2 (ENB 2)*. Raffaello Cortina Editore, Milano.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneaus, M. P. S., Whitehead, M. S. W., Collin, I., Cummings, J. L., Chertkow, H. (2005) The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool for Mild Cognitive Impairment. *Journal of the American Geriatrics Society*, 53(4),695-699.

Novelli, G., Papagno, C., Capitani, E., Laiacona, M. (1986) Tre test clinici di ricerca e produzione lessicale. Taratura su soggetti normali. In: *Archivio di Psicologia, Neurologia e Psichiatria*, 47(4), 477-506.

Peintner, B., Jarrold, W.L., Vergyriy, D., Richey, C., Gorno-Tempini, M.L., Ogar, J. (2008). Learning diagnostic models using speech and language measures, In *Proc.International Conference of the IEEE Engineering in Medicine and Biology Society*, 4648-4651.

Pirani, A., Brodaty, H., Martini, E., Zaccherini, D., Neviani, F., Neri, M. (2010). The validation of the Italian version of the GPCOG (GPCOG-It): a contribution to cross-national implementation of a screening test for dementia in general practice. In: *International Psychogeriatric Association*, 22(1), 82-90.

Ramus, F., Nespor, M., Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. In: Cognition, 73: 265-292.

Roark, B., Hosom, J. P., Mitchell, M., Kaye, J. A. (2007a).Automatically derived spoken language markers for detecting Mild Cognitive Impairment. In *Proc. of the 2nd International Conference on Technology and Aging (ICTA)*.

Roark, B., Mitchell, M., Hollingshead, K. (2007b). Syntactic complexity measures for detecting Mild Cognitive Impairment. In Cohen K. B., Demner-Fushman, D., Frieman, C., Hirschman, L., Pestian J. (eds), *Proc. of the Workshop BioNLP 2007: Biological, translational, and clinical language processing*, ACL - Association for Computational Linguistics, 1–8.

Roark, B., Mitchell, M., Hosom, J.P., Hollingshead, K., Kaye. J.A. (2011). Spoken language derived measures for detecting Mild Cognitive Impairment. In *IEEE Transactions on Audio Speech, and Language Processing*, 19(7), 2081–2090.

Satt, A., Sorin, A., Toledo-Ronen, O. (2012). Vocal biomarkers for dementia patient monitoring. In *Proc. of Interspeech 2012*. ISCA.

Satt, A., Sorin, A., Toledo-Ronen, O., Barkan, O., Kompatsiaris, I., Kokonozi, A., Tsolaki, M. (2013). Evaluation of speech-based protocol for detection of early-stage dementia. In Proc. *Interspeech 2013*, 1692–1696.

Singh, S., Bucks, R.S., Cuerden, J.M. (2001). An evaluation of an objective technique for analysing temporal variables in DAT spontaneous speech. In *Aphasiology*, 15(6): 571–583.

Snowdon, D.A. (2003). Healthy aging and dementia: Findings from the nun study. *Annals of Internal Medicine*, 139(5), 450–454.

Snowdon, D.A., Kemper, S.J., Mortimer, J.A., Greiner, L.H., Wekstein, D.R., Markesbery, W.R. (1996). Linguistic ability in early life and cognitive function and Alzheimer's disease in late life: Findings from the nun study. In *The journal of the American Medical Association*, 275: 528–532.

Swainson, R., Hodges, J.R., Galton, C.J., Paykel, E.S., Semple, J., Michael, A., Dunn, B.D., Iddon, J.L., Robbins, T.W., Sahakian, B.J., (2001). Early detection and differential diagnosis of Alzheimer's disease and depression with neuropsychological tasks. In: *Dementia and Geriatric Cognitive Disorders*, 12, 265-280.

Szmrecsányi, B.M. (2004). On operationalizing syntactic complexity. In Purnelle, G., Fairon, C., Dister, A. (eds), *Proc. of the 7th International Conference on Textual Data Statistical Analysis*, Presses Universitaires de Louvain, 1031–1038.

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*, 30(5), 501-556.

Thomas, C., Keselj, V., Cercone, N., Rockwood, K., Asp, E. (2005). Automatic detection and rating of dementia of Alzheimer type through lexical analysis of spontaneous speech. In Gu, J., Liu, P. X. (eds), *Proc. of the IEEE International Conference on Mechatronics & Automation,* volume 3, 1569–1574.

Tóth, L., Gosztolya, G., Vincze, V., Hoffmann, I., Szatlóczki, G., Biró, E., Zsura, F., Pákáski, M., Kálmán, J. (2015), Automatic Detection of Mild Cognitive Impairment from Spontaneous Speech using ASR, In *Proc. of Interspeech 2015*, Dresden, Germany, 2015.

Tsoi, K. K. F., Chan, J. Y., Hirai, H. W., Wong, S., Y. S., Kwok, T. C. Y. (2015). Cognitive Tests to Detect Dementia. A Systematic Review and Meta-analysis. *JAMA Intern Med*, 175 (9),1450-1458.

Velayuhan, L., Ryu, S., Raczek, M., Philipot, M., Lindesay, J., Critchfield, M., Livingston, G. (2014). Review of brief cognitive tests for patients with suspected dementia. *International Psychogeriatrics*, 26(8),1247-1262.

Vigorelli, P. (ed.) (2004). *La conversazione possibile con il malato Alzheimer*, Franco Angeli.