

*Marche linguistiche come indicatori del
decadimento cognitivo*

Daniela Beltrami
Università di Bologna / Arcispedale S.Maria Nuova (RE)
daniela.beltrami3@unibo.it

Laura Calzà
Università di Bologna
laura.calza@unibo.it

Gloria Gagliardi
Università di Bologna
gloria.gagliardi2@unibo.it

Enrico Ghidoni
Arcispedale S.Maria Nuova (RE)
Enrico.Ghidoni@asmn.re.it

Norina Marcello
Arcispedale S.Maria Nuova (RE)
marcello.norina@asmn.re.it

Rema Rossini Favretti
Università di Bologna
rema.rossini@unibo.it

Fabio Tamburini
Università di Bologna
fabio.tamburini@unibo.it

Introduzione

Nello spettro eterogeneo di sintomi correlati alle fasi iniziali del deterioramento cognitivo associato a patologie neurodegenerative a carattere progressivo si evidenzia spesso un declino subliminale delle facoltà linguistiche (Caramelli, 1998; Taler & Phillips, 2008). Di solito tali deficit non hanno ripercussioni evidenti sulla correttezza formale delle produzioni, e non intaccano le capacità linguistico-comunicative di base come la comprensione in un contesto ecologico, la partecipazione a conversazioni di

routine, le funzioni strumentali di lettura e scrittura.

Tali cambiamenti, presenti già diversi anni prima della fase clinica della malattia, ma difficilmente identificabili mediante i test neuropsicologici tradizionali, rappresentano una delle manifestazioni prodromiche del processo patologico in corso e possono, in quanto tali, essere utilizzati come indici per lo *screening* e la diagnosi precoce.

Nella letteratura specifica degli ultimi anni si segnalano diversi lavori che si occupano dell'identificazione automatica dei vari stadi e tipi di demenze a partire dall'analisi statistica della produzione linguistica spontanea dei pazienti (Chapman et al., 2002; Jarrold et al., 2010; Lehr, 2012; Roark et al., 2011; Satt et al., 2013). Tali prospettive di analisi computazionale si sono dimostrate in grado di rilevare pattern latenti e regolarità nel linguaggio di queste popolazioni di pazienti, potenzialmente utili all'identificazione precoce, classificazione e descrizione del decadimento cognitivo (Elvevåg & Garrard, 2014).

Il progetto OPLON ("OPportunities for active and healthy LONgevity", Smart Cities and Communities – DD 391/RIC) si pone l'obiettivo di sviluppare strumenti di diagnosi precoce della fragilità finalizzati alla prevenzione del declino e alla promozione della salute dei soggetti anziani. Entro tale linea generale di azione si propone, in particolare, di elaborare tecniche semplici e facilmente somministrabili di raccolta e analisi di campioni di linguaggio spontaneo per lo *screening* sistematico della popolazione potenzialmente a rischio, grazie all'integrazione delle competenze di neurologi, psicologi, linguisti ed esperti informatici.

A medio termine, questo studio si propone di dimostrarne la fattibilità in ambiente controllato, visto che in letteratura non risultano al momento studi simili sulla lingua italiana.

Raccolta dati

Il progetto prevede il reclutamento di 96 soggetti bilanciati per sesso, età (range 50-75) e scolarità (licenza media o licenza elementare con buona stimolazione intellettuale nel corso della vita, diploma di scuola superiore; laurea), di cui 48 soggetti di controllo e altrettanti affetti da declino cognitivo. Tale declino fa riferimento a due diverse categorie:

1. *Mild Cognitive Impairment* (MCI): deterioramento cognitivo lieve non in grado di interferire in modo significativo con le attività della vita quotidiana, riguardante una o più funzioni cognitive.
 - a. MCI con isolato deficit di memoria (*a-MCI single domain*; n = 16);
 - b. MCI con deficit di funzioni cognitive ma non di memoria (*non a-MCI single or multiple domain*; n = 16).
2. Demenza in fase iniziale (*early-D*; n = 16): pazienti con deficit cognitivi che influenzano, sebbene ancora solo parzialmente, le autonomie nelle attività della vita quotidiana (ad es. assunzione autonoma della terapia farmacologica, cura personale, cura dell'ambiente domestico, ecc.). Al MMSE (Mini Mental State Examination) presentano un punteggio grezzo maggiore o uguale a 18.

Ai soggetti sono stati somministrati i test neuropsicologici più efficaci nella discriminazione tra soggetti normali, con MCI e con Demenza, ovvero

MMSE, MoCA, GPCog, CDT Fluenze Verbali (fluenza fonemica e semantica), Paired Associate Learning (PAL, subtest del Cambridge Neuropsychological Test Automated Battery – CANTAB).

Dopo la valutazione tradizionale, i soggetti sono stati sottoposti ad una registrazione dell'eloquio spontaneo durante alcuni *task*:

- “Descrivi questa immagine”;
- “Descrivi una giornata lavorativa”;
- “Descrivi l'ultimo sogno che ricordi”.

Questo contributo presenta i risultati preliminari su un insieme parziale di locutori: 19 MCI (in questa fase non ulteriormente suddivisi) e 20 controlli.

Metodologia

I campioni di linguaggio parlato raccolti nel corso delle sedute di test nella forma di file audio .

WAV vengono sottoposti a trascrizione ortografica manuale utilizzando il *software* Transcriber². L'unità di allineamento scelta è l'enunciato, definito e identificato sulla base di criteri intonativi (Cresti, 2014). Contestualmente alla trascrizione vengono annotati anche i principali fenomeni paralinguistici (pause, disfluenze, lapsus, etc.).

Vengono quindi isolati i turni di parola del soggetto sottoposto a test: gli enunciati selezionati, emendati manualmente dalle disfluenze, sono sottoposti ad annotazione morfosintattica semi-automatica, ovvero subiscono *Part-of-Speech (PoS) Tagging* e *Parsing* automatico mediante il *parser* a dipendenze TULE (Lesmo, 2007). Il *tagset* adottato è quello del TUT - Turin University TreeBank (Bosco et al., 2000).

Per ciascun livello di analisi vengono calcolati numerosi indici linguistici e stilo-metrici descritti in letteratura e risultati significativi negli studi condotti su lingue diverse dall'italiano, o creati *ex-novo*. I parametri acustici vengono calcolati direttamente sul segnale vocale: attraverso il *Voice Activity Detector* “adintool” contenuto nel *package* Julius-ASR³ il file audio viene segmentato e vengono identificati i segmenti che contengono *speech* da quelli che contengono rumori o silenzio. Per ottenere una trascrizione fonetica temporalmente allineata, necessaria per il calcolo degli indici ritmici, viene utilizzato il sistema di *forced alignment* del pacchetto Kaldi-DNN-ASR⁴, addestrato sul *corpus* APASCI (Angelini et al., 1994).

ACOUSTIC FEATURES
<i>Silence segments duration: mean, median and Std. Deviation</i> [SPE_SILMEAN, SPE_SILSD, SPE_SILMEDIAN]
<i>Speech segments duration: mean, median and Std. Deviation</i> [SPE_SPEMEAN, SPE_SPESD, SPE_SPEMEDIAN]
<i>Temporal regularity of voiced segment durations</i> [SPE_TRVSD]
<i>Verbal Rate</i> [SPE_VR]
<i>Transformed Phonation Rate</i> [SPE_TPR]

² <http://trans.sourceforge.net>

³ http://julius.osdn.jp/en_index.php

⁴ <http://kaldi.sourceforge.net/about.html>

<i>Standardized Phonation Time</i> [SPE_SPT]
<i>Standardized Pause Rate</i> [SPE_SPR]
<i>Root Mean Square energy: mean and Std. Deviation</i> [SPE_RMSEM, SPE_RMSESD]
<i>Pitch: mean and Std. Deviation</i> [SPE_PITCHM, SPE_PITCHSD]
<i>Spectral Centroid, mean and Std. Deviation</i> [SPE_SPCENTRM, SPE_SPCENTRSD]
<i>Higuchi Fractal Dimension, mean and Std. Deviation</i> [SPE_HFractDM, SPE_HFractDSD]
RHYTHMIC FEATURES
<i>Percentage of vocalic intervals</i> [RHY_V]
<i>Standard deviation of vocalic and consonantal intervals</i> [RHY_DeltaV, RHY_DeltaC]
<i>Pairwise Variability Index, raw and normalized</i> [RHY_VnPVI, RHY_CrPVI]
<i>Variation coefficient for ΔV and ΔC</i> [RHY_VarcoV, RHY_VarcoC]
LEXICAL FEATURES
<i>Lexical density: Open-class ratio</i> [LEX_OCW]
<i>Content Density</i> [LEX_ContDens]
<i>PoS rate</i> [LEX_PoS_X] (es. LEX_PoS_ADJ, LEX_PoS_VERB)
<i>Reference Rate to Reality</i> [LEX_RefRReal]
<i>Personal, Spatial and Temporal Deixis rate</i> [LEX_PDEIXIS, LEX_SDEIXIS, LEX_TDEIXIS]
<i>Relatives pronouns and negative adverbs rate</i> [LEX_RemaW]
<i>Lexical Richness: Type-Token Ratio, W - Brunét's Index and R - Honoré's Statistic</i> [LEX_TTR, LEX_BrunetW, LEX_HonoreR]
<i>Action Verbs rate</i> [LEX_ACTVRB]
<i>Frequency-of-use tagging (De Mauro/Paravia dictionary)</i> [LEX_DM_F]
<i>Propositional Idea Density</i> [LEX_IDEAD]
SYNTACTIC FEATURES
<i>Number of dependent elements linked to the noun, mean</i> [SYN_NPLENM]
<i>Mean Global Dependency Distance</i> [SYN_GRAPHDISTM]
<i>Syntactic complexity</i> [SYN_ISynCompl]
<i>Syntactic embeddedness: maximum depth of the structure, mean and Std. Deviation</i> [SYN_MAXDEPTHM, SYN_MAXDEPTHSD]
<i>Utterance length, mean and Std. Deviation</i> [SYN_SLENM, SYN_SLENSD]

L'effettiva significatività statistica per la lingua italiana degli indici e delle possibili correlazioni tra variabili è stata valutata attraverso test e metriche statistiche standard (parametriche e non parametriche) e i parametri risultati significativi sono stati ulteriormente verificati mediante algoritmi di *feature selection*. Tramite metodologie di *machine learning* gli indici selezionati vengono integrati in un classificatore automatico, le cui performance sono valutate in termini di *accuracy*, *precision*, *recall* e *F-*

measure.

Nella fase conclusiva del progetto i risultati del test linguistico verranno unificati in un punteggio di *scoring* sintetico, in grado di esprimere un indicatore di impoverimento del linguaggio dell'individuo sottoposto a test, da integrare nel modello OPLON di predizione del rischio di insorgenza di fragilità nell'anziano.

Risultati preliminari

L'effettiva fattibilità della metodologia è stata verificata mediante un *Proof of Concept* su un campione pilota di 39 soggetti, 19 MCI e 20 controlli.

Per ciascun *task* del test linguistico sono state identificate le *feature* statisticamente significative ($p\text{-value} < 0.05$) applicando il test non parametrico di Kolmogorov-Smirnov, in ragione della ridotta dimensione del campione.

Gli indici selezionati sono stati integrati in un classificatore automatico *kNN 3-neighbours*. Il classificatore esegue un campionamento casuale dei testi e 5 iterazioni delle fasi di *training* e *testing*, con proporzioni (80%/20%).

È evidente che la ridotta quantità di dati non permette l'applicazione produttiva di queste metodologie, se non per ricercare tendenze e informazioni al fine di predisporre metodologie più adeguate in vista del campione completo.

Le *feature* significative e le prestazioni del classificatore (in relazione a ciascun *task* e globali) sono sintetizzate nelle tabelle che seguono.

TASK "FIGURA"		
Tratti selezionati	RHY_DeltaC	KS test p = 0.048527
	LEX_PoS_ADJ	p = 0.040478
	LEX_ContDens	p = 0.022891
	LEX_CCW	p = 0.022891
	SPE_SILMEAN	p = 0.018767
	SPE_HFractDSD	p = 0.006515
	SPE_SPCENTRM	p = 0.000648
	Prestazioni del Classificatore	<i>Accuracy</i> = .825 <i>Precision</i> = .882 <i>Recall</i> = .750 <i>F-measure</i> = .811
TASK "LAVORO"		
Tratti selezionati	SPE_SPEMEAN	KS test p = 0.048527
	LEX_PoS_PREDET	p = 0.048527
	SPE_TPR	p = 0.040478
	SYN_SLENM	p = 0.019730
	SPE_HFractDM	p = 0.016965
	LEX_PoS_INTERJ	p = 0.007688
	LEX_DM_F	p = 0.006886
	SPE_SPCENTRM	p = 0.002030
Prestazioni del	<i>Accuracy</i> = 1.0	

Classificatore	<i>Precision</i> = 1.0 <i>Recall</i> = 1.0 <i>F-measure</i> = 1.0
----------------	---

TASK "SOGNO"			
Tratti selezionati	SYN_ISynCompl	D = 0.439474	KS test
	SPE_SPCENTRM		p = 0.048527
	SPE_SPEMEDIAN		p = 0.046395
	LEX_PoS_VERB		p = 0.042373
			p = 0.006886
Prestazioni del Classificatore	<i>Accuracy</i> = .725 <i>Precision</i> = .765 <i>Recall</i> = .65 <i>F-measure</i> = .702		

TUTTI TASK			
Tratti selezionati	SPE_SPEMEDIAN		KS test
	LEX_PoS_VERB		p = 0.035847
	LEX_DM_F		p = 0.028823
	SPE_RMSEM		p = 0.026520
	SYN_GRAPHDISTM		p = 0.019976
	SYN_SLENM		p = 0.017278
	SPE_SILSD		p = 0.014911
	SPE_VR		p = 0.013232
	SPE_SPR		p = 0.011725
	SPE_SILMEDIAN		p = 0.008088
	SPE_SPESD		p = 0.007837
	SPE_SPEMEAN)		p = 0.006269
	SPE_TPR		p = 0.002212
	SPE_HFractDM		p = 0.000966
	SPE_SILMEAN		p = 0.000488
	SPE_HFractDSD		p = 0.000344
SPE_SPCENTRM	p = 0.000212		
	p = 0.000000		
Prestazioni del Classificatore	<i>Accuracy</i> = .758 <i>Precision</i> = .772 <i>Recall</i> = .733 <i>F-measure</i> = .752		

I risultati presentati, ancorché preliminari, appaiono incoraggianti. Pur utilizzando un set di dati estremamente ridotto, l'esperimento ha permesso di saggiare la validità della metodologia proposta e ha fornito indicazioni preliminari sulla significatività che alcuni degli indici linguistici presentano per la lingua italiana nei testi in esame.

Nelle fasi prodromiche delle malattie dementigene è già presente un deficit linguistico, seppur subliminale; tale deficit, identificabile con strumenti automatici, non è circoscrivibile ad un singolo livello ma, come si può notare dai tratti risultati statisticamente significativi nei diversi *task*, interessa il piano ritmico/acustico, lessicale e sintattico.

A causa della ridotta dimensione del campione è stato possibile testare unicamente l'algoritmo di classificazione kNN (algoritmo *memory based*). Ulteriori metodologie di classificazione (es. *neural network*, *Support Vector Machine*, *Quantum Classifiers*) verranno valutate sul *corpus* completo.

Bibliografia

- 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*, pp. 1391-1394.
- Bosco, C., Lombardo, V., Vassallo, D. & Lesmo, L. (2000). Building a Treebank for Italian: a data-driven annotation schema. In *Proc. of LREC-2000*.
- 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*, pp. 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. *Alzheimer Disease & Associated Disorders*, 16(3), pp. 177-186.
- Cresti E. (2014). Syntactic properties of spontaneous speech in the language into act theory: Data on Italian complements and relative clauses. In Raso T. & Mello H. (eds.) *Spoken Corpora and Linguistic Studies*, Amsterdam: Benjamins, pp. 365-410.
- Elvevåg, B. & Garrard, P. (2014). *Language, Computer and Cognitive Neuroscience: Cortex Special Issue*, 55.
- 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)*, pp. 299-307.
- Lehr, M., Prud'hommeaux, E., Shafran, I. & Roark, B. (2012). Fully automated neuropsychological assessment for detecting Mild Cognitive Impairment. In *Proc. of InterSpeech 2012*.
- Lesmo, L. (2007). Il parser basato su regole del Gruppo NLP dell'Università di Torino. *Intelligenza Artificiale*, IV(2), pp. 46-47.
- Roark, B., Mitchell, M., Hosom J. P., Hollingshead, K. & Kaye, J. A. (2011). Spoken language derived measures for detecting Mild Cognitive Impairment. *IEEE Transactions on Audio Speech, and Language Processing*, 19(7), pp. 2081-2090.
- 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*, pp. 1692-1696.
- Taler, V. & Phillips, N. A. (2008). Language performance in Alzheimer's disease and mild cognitive impairment: a comparative review. *Journal of Clinical and Experimental Neuropsychology*, 30(5), pp. 501-556.