Gloria Gagliardi, Fabio Tamburini

Linguistic biomarkers for the detection of Mild Cognitive Impairment

(doi: 10.1418/101111)

Lingue e linguaggio (ISSN 1720-9331)
Fascicolo 1, gennaio-giugno 2021

Ente di afferenza:
Università di Bologna (unibo)

LINGUISTIC BIOMARKERS FOR THE DETECTION OF MILD COGNITIVE IMPAIRMENT

GLORIA GAGLIARDI   FABIO TAMBURINI

ABSTRACT: A timely diagnosis of the prodromal stages of dementia remains a big challenge for healthcare systems: many assessment tools have been proposed over recent years, but the commonest screening instruments are largely unreliable for detecting subtle changes in cognition. The scientific literature contains a rising number of reports about language disturbances at the earliest stages of dementia, a clinical syndrome known as "Mild Cognitive Impairment" (MCI). Here we take advantage of these findings to develop a novel NLP method capable of identifying cognitive frailty at a very early stage by processing Italian spoken productions. This study constitutes a first step in the creation of an automatic tool for non-intrusive, low-cost dementia screening exploiting linguistic biomarkers. Our findings show that acoustic features (i.e., fluency indexes and spectral properties of the voice) are the most reliable parameters for MCI early identification. Moreover, lexical and syntactic features, grabbing the erosion of verbal abilities caused by the pathology, emerge as statistically significant and can support speech traits in the classification process.

KEYWORDS: Mild Cognitive Impairment, Dementia screening, Natural Language Processing, linguistic biomarkers.

1. INTRODUCTION[1]

The term "dementia", according to the World Health Organization (WHO), describes an etiologically heterogeneous clinical syndrome - usually of a progressive nature – marked by a decline in cognitive performances beyond what

[1] Corresponding author: Fabio Tamburini, fabio.tamburini@unibo.it.This work was supported by the OPLON project (Opportunities for active and healthy LONgevity, Smart Cities, Ministero Università e Ricerca, SCN_00176). The study was approved by the Ethical Committee of Azienda Ospedaliera Reggio Emilia (n.148 2013/0013438). Given the particular kind of data employed for this study and the restrictions on them from the Italian legislation, unfortunately, we cannot make the dataset publicly available.

We gratefully acknowledge Daniela Beltrami and Enrico Ghidoni for subject selection and audio files recording, Laura Calzà and Rema Rossini Favretti for all the fruitful discussions and the initial design of this project. We also thank Annalisa Raffone for proof-reading the paper. For academic requirements, G. Gagliardi takes official responsibility for Sections 1, 2, 2.1, 2.2, 2.3 and 4 and F. Tamburini for sections 2.4 and 3.

might be expected from normal aging. It refers to a group of neurodegenerative disorders characterised by an irreversible decrease of brain functions; most patients presented symptoms such as memory loss, disorientation, slowed processing speed and the inability to perform daily activities.

Cognitive dysfunctions associated with dementia have a high impact on society: in 2019, the Alzheimer's Disease International Association estimated that over 50 million people are afflicted by this pathology globally, a figure set to rise to 152 million by 2050. The increase in life expectancy is contributing to rapidly boosting this number: each year, 9.9 million new cases of dementia are identified worldwide. Namely, someone develops dementia every three seconds on average (Patterson 2018). These epidemiological trends shall be updated in the light of COVID-19 morbidity and mortality in the elderly population: according to the data gathered by the Italian National Institute of Health (Istituto Superiore di Sanità - ISS), 86.22% of deaths due to SARS-CoV-2 has occurred in patients aged $\geq$ 70 years, with an average age at death of 81 (update: March 1st, 2021, Istituto Superiore di Sanità). Nevertheless, the management of this increasing number of frail people represents a big challenge for healthcare systems.

Since many clinical trials have failed to find a cure for this ravaging condition, a paradigm shift, in kuhnian terms (Kuhn 1962), has occurred, whereby dementia is conceived as a *continuum* for which early intervention may offer the best chance of therapeutic success. As a matter of fact, the neurodegenerative process leading to dementia is thought to begin much earlier than the clinical symptoms: this "preclinical" or "prodromal" phase, called "Mild Cognitive Impairment" (MCI) (Petersen 2011), a grey area between normal aging and pathological cognitive functioning, would provide a pivotal opportunity for pharmacological treatment development and therapeutic intervention (Calzà *et al.* 2015; Ritchie *et al.* 2017). Thus, improving the rate and promptness of diagnosis has become an integral part of national and international dementia strategies.

Moreover, people living with dementia cannot always experience a complete and equitable engagement in everyday life activities because of the stigma associated with the illness: therefore, customized interventions at the very early stages of the disease might alleviate the emotional workload for patients and their caregivers. Besides, an adequate and timely risk identification may allow for implementing preventive measures such as dietary, lifestyle, and neuroprotection precautions, playing an important role in delaying the pathology onset. Despite the urgency of the issues at stake, the problem of diagnosing dementia remains controversial: an extensive literature and a considerable body of evidence exist on the possibility of a preclinical diagnosis of Alzheimer's

Disease (AD) and other types of dementia, but pre-symptomatic diagnosis raise both theoretical issues and ethical concerns (Calzà *et al.* 2015) as well as big practical problems.

Many assessment tools have been proposed over recent years, however the conventional screening instruments (e.g., "Mini Mental State Examination" - MMSE (Folstein *et al.* 1975), "Montreal Cognitive Assessment" - MoCA (Nasreddine *et al.* 2005) and "General Practitioner Assessment of Cognition" - GpCOG (Brodaty *et al.* 2004) are not too adequate for detecting the early subtle changes in cognitive abilities. It would be crucial to have highly accurate and specific psychometric tests, suitable for low-cost and large-scale use. Several initiatives and studies are in progress (Mortamais *et al.* 2017), but, at the moment, the role of these instruments is still puzzling: although suitable to detect clinically manifested cases, they are much less effective to track down the prodromal phase of cognitive frailty, such as the condition of MCI.

1.1 Language in Neurocognitive disorder

Considerable evidence is available for suggesting that linguistic deficits are present in several neurodegenerative diseases (e.g., see Boschi *et al.* 2017, for a review); that is especially the case with dementia, where language disruption is a common finding both at the earliest stages and in the full-blown pathology.

Although episodic memory and visuospatial skill impairments are the core symptom of AD, a progressive language disorder is usually also found (Ahmed *et al.* 2013; Szatloczki *et al.* 2015). Nevertheless, unlike aphasias, which are due to focal brain damage, verbal deficits usually occur in the context of multiple cognitive impairments (Forbes-McKay *et al.* 2013), which encompass executive function, reasoning and visuoconstructive abilities. Patients commonly show, among many other signs, a decline of lexical semantic knowledge, with word-finding problems (i.e., anomia and semantic paraphasias), impaired auditory and written comprehension, verbal fluency decrease and low content density (Kempler *et al.* 1987; de Lira *et al.* 2011; Catricalà *et al.* 2015; Fraser *et al.* 2016). These symptoms occur early and increase during the illness, suggesting a massive breakdown of semantic memory (i.e., the long-term memory store in which conceptual information is represented, including semantic and lexical information as well as facts about the world, Tulving 1972). At the phonetic level, temporal parameters of speech are usually altered: lower speech rate and a high number of hesitations have been reported as common findings (Hoffmann *et al.* 2010; Sajjadi *et al.* 2012). Syntactic processing tends to be relatively preserved in the early course of the disease (Cuetos *et al.* 2007; Sajjadi *et al.* 2012). Nevertheless, several studies have shown that sentence structure is correct but reduced (Kemper *et al.* 1993; de Lira *et al.* 2011; Fraser *et al.* 2016), and a greater number of inflectional errors in AD patients than in healthy persons has also been observed (Altmann *et al.* 2001; Cuetos *et al.* 2007). Deficits affect the productive and receptive discourse-level processing too: individuals generally produce shorter texts than the normal controls with less relevant information and multiple error types (e.g., incoherent/indefinite phrases, referential/temporal cohesion errors, discourse planning weakness, and inability to abstract (Ripich *et al.* 2000; Chapman *et al.* 2002; Carlomagno *et al.* 2005; March *et al.* 2006; Drummond *et al.* 2015). Conversely, repetition abilities and articulation remain relatively intact. As the AD pathology progresses, linguistic symptoms become pervasive, showing a full breakdown of speech comprehension and verbal production restricted to echolalia and stereotypy.

A progressive loss of specific language skills with relative sparing of other cognitive domains (such as memory of daily events, visual-spatial skills and behavior) marks out Primary Progressive Aphasia (PPA), a heterogeneous group of focal language-led dementias (Mesulam 2003). As a matter of fact, PPA is diagnosed when all major limitations in daily-living activities can be attributed to a language impairment for at least two years after the onset. Three subtypes are currently described: non-fluent/agrammatic variant PPA (nfvPPA), semantic variant PPA (svPPA), and logopenic variant PPA (lvPPA), each of which exhibits peculiar patterns of brain atrophy and linguistic features (Gorno-Tempini *et al.* 2011). People with non-fluent PPA show major impairments at the phonetical, phonological, and syntactic level. They usually show slow, effortful, hesitant and dysprosodic speech with prominent articulatory errors. Moreover, they exhibit both agrammatism in language production (i.e., conversation is sometimes strikingly telegraphic, with omissions of grammatical morphemes) and impaired comprehension of complex sentences (i.e., negative passive and object-relative clause) against spared single-word decoding and object knowledge. In striking contrast to non-fluent variant PPA, patients with svPPA (also known as "Semantic Dementia") exhibit a well structured and well-articulated language. The disease typically starts as a severe anomia and with the inability to express thoughts with precision. Patients' verbalizations become progressively more circumlocutory and "empty". They often use superordinate category names instead of the target name (e.g., "Border Collie" > "dog" > "animal"), but semantic deficits also affect single-word comprehension, especially for low-frequency items (e.g., "horse" instead of the less familiar "zebra"). These symptoms represent the earliest markers of a widespread conceptual knowledge degradation. The most recently described phenotype is the logopenic variant PPA (from greek, 'lack of words'): the clinical presentation usually shows a single-word retrieval difficulty and conversational lapses, from which the disease takes its name. Word-finding problems bring about a slow speech rate, but lack of frank agrammatism helps distinguish it from other subtypes. Furthermore, people suffering from the lvPPA present impaired sentence repetition and poor comprehension of complex syntactic structures.

Conversely, language disruption is not a core feature of Dementia with Lewy Bodies (LBD) (Galasko *et al.* 1996; Ash *et al.* 2011). However, naming and verbal fluency impairment, presumably connected to defective executive functioning, have been extensively reported. Also, alterations have been described both at the phonetic (e.g., speech rate, articulation errors) and pragmatic level (e.g., narrative organization, coherence and topic maintenance).

Although there is a lot of empirical evidence about language disruption in AD, PPA and LBD, less attention has been paid to language disorders in preclinical stages of cognitive frailty. The literature contains a rising number of reports about language disturbances in individuals with confirmed MCI. Reviewing the literature on the topic, verbal impairments in MCI seem to parallel those found in early/moderate stage Dementia (Taler & Phillips 2008): the most commonly reported deficit is impaired verbal fluency (Tsantali *et al.* 2013), and impaired confrontation naming (Ahmed *et al.* 2008), but pragmatic skills seem to be the most affected domains. It is also well documented that discourse alterations (i.e., semantically impoverished discourse that lacks coherence) may be one of the earliest signs of the pathology, often noticeable years before other cognitive deficits become apparent.

Language deficits have also been found to be a strong predictor of conversion from MCI to dementia: some longitudinal retrospective studies have already demonstrated that, in apparently normal elderly people, subtle measurable expression deficits reliably predict the time of clinically relevant cognitive impairment long before clinical symptoms are reported (Snowdon 2003; Oulhaj *et al.* 2009).

1.2 Natural Language Processing technologies for dementia screening

Traditional standardized neuropsychological tests (i) show a very low sensitivity to the detection of subtle changes, (ii) they do not allow for exploration of many other aspects of language, both at the segmental and suprasegmental level (e.g., prosody, rhythm), and (iii) they are time-consuming and expensive. Significant differences between the MCI and normal elderly participants have been identified with these instruments from time to time, but their clinical use is still unreliable (Taler & Phillips 2008; Szatloczki *et al.* 2015; Filiou *et al.* 2019).

Conversely, there is rising evidence of the feasibility of automatic speech analysis in addressing precisely this challenge. During the last few years, new sophisticated techniques from Natural Language Processing (NLP) and Artificial Intelligence (AI) have been used to analyze written texts, clinically elicited utterances and spontaneous speech, to identify signs of psychiatric or neurological disorders: depression (Jiang *et al.* 2017; Stasak *et al.* 2019), focal brain lesions (Fergadiotis & Wright 2011), Parkinson's disease (Arias-Vergara *et al.* 2018; Upadhya Savitha *et al.* 2019) dementia prodroms (Roark *et al.* 2011; Satt *et al.* 2013; dos Santos *et al.* 2017; Vincze *et al.* 2016; Tóth *et al.* 2018), Alzheimer's Disease (Jarrold *et al.* 2014; López-de-Ipiña *et al.* 2015; Fraser *et al.* 2016; Yancheva & Rudzicz 2016; Sirts *et al.* 2017), Primary Progressive Aphasia (Fraser *et al.* 2014) and Fronto-Temporal Dementia (Peintner *et al.* 2008)).

Within this paradigm, subtle language disruptions can act as "digital biomarkers", namely objective, quantifiable behavioral data that can be collected and measured through digital devices, allowing for a low-cost and ecologically valid assessment. Unlike traditional neuropsychological tests, which have a heavy impact on the naturalness of the subject's responses, the automatic analysis of spoken language productions could represent a non-intrusive, inexpensive technique for accurately detecting language modifications in potential patients, even by primary care physicians.

The *fil rouge* that links the referred literature is (i) the preliminary testing of a set of linguistic features to pinpoint systematic patterns of language alterations in the cohort, automatically extracted from the speech sample using NLP methods, and (ii) the building of proper algorithms for the detection of the pathology, exploiting conventional Machine Learning techniques, such as Support Vector Machines, Neural Networks, K-Nearest Neighbor, etc.

Given this complex but intriguing picture, our paper presents a novel method to analyse cognitive frailty at a very early stage by processing semi-spontaneous Italian language productions, devised within the framework of the OPLON project.[2] Despite the increasing number of international scientific papers on the topic, at the time of writing and to the best of our knowledge, no studies specifically devoted to Italian and performing a similar analysis exist.

There is a sizeable and rapidly growing scientific literature demonstrating high accuracy of systems performing binary classifications between fully developed dementia and healthy controls (e.g., Jarrold *et al.* 2014; López-de-Ipiña *et al.* 2015; Fraser *et al.* 2016; Sirts *et al.* 2017). However, this task is not helpful from a clinical perspective since available treatment may help alleviate

[2] The project was founded by the Italian Ministry of University and Research, as part of the Contract "Smart Cities and Communities and Social Innovation".

some of the symptoms only if administered in the early stages of the pathology but are almost ineffective for full-blown dementia.

The ultimate goal of the project is the development of an instrument for non-intrusive, low-cost cognitive decline screening exploiting linguistic biomarkers. This pilot study represents a first step towards creating such automatic tools by exploring the relevance of linguistic features in supporting the automatic identification of MCI patients. The paper is structured as follows: §2 describes the experimental design (i.e., corpus collection and annotation, feature extraction); §3 outlines the findings of our study; in a brief conclusion (§4), the main results are discussed in the light of previous literature and suggestions for further improvement of this work are presented.

2. MATERIALS AND METHODS

This section describes the dataset involved in the study by both providing information on the participants and some details on the language tasks (§ 2.1), and defining the annotation procedures (§2.2). It also presents a description of the linguistic features (§2.3) and the methodology for their automatic extraction from the speech samples (§2.4).

2.1 Recruitment and clinical assessment

A total of 96 subjects were enrolled: 48 healthy controls (HC) and 48 patients with cognitive decline, recruited from outpatients within the clinical services of Emilia-Romagna Region involved in the care and diagnostic evaluation of cognitive disorders and dementia. They belong to two categories:

- Mild Cognitive Impairment (MCI; 32 subjects): a syndrome defined as cognitive decline greater than expected for an individual's age and education level but not so severe to interfere with everyday activities.

- Early Dementia (e-D; 16 subjects): a clinical condition characterised by noticeable symptoms of cognitive decline, partially affecting day-to-day life.

The sample is balanced by sex, age and education. All participants underwent (i) a detailed anamnestic interview investigating personal data, occupation/retirement, social groups, family history of neurodegenerative pathologies, medical history and pharmacotherapy, and (ii) a comprehensive neuropsychological assessment, evaluating several cognitive domains (i.e., logic, memory, attention, language, visuospatial, praxic and executive functions).

The battery was composed of those tests that are most used in the clinical practice to assess cognitive decline (Tsoi *et al.* 2015), with an Italian standardization and a short administration time. Comprehensive information about the study cohort, sampling strategy as well as neuropsychological assessment can be found in Beltrami *et al.* (2018).

The semi-spontaneous speech produced by the subjects during the execution of three linguistic tasks was recorded at the end of the traditional cognitive evaluation. Verbalizations have been elicited with the following stimuli:

- Task *FIGURE*: "Could you please describe this picture?" The visual stimulus is the well-known figure provided by the "*Esame del Linguaggio II*" test (Ciurli *et al.* 1996), depicting a living room with some characters carrying out simple actions. (e.g., the father is reading a newspaper, the mother is knitting);
- Task *WORK*: "Could you please describe your typical working day?";
- Task *DREAM*: "Could you please describe the last dream you remember?".

Speech samples were recorded in a quiet room with an Olympus – Linear PCM Recorder LS-5 (as WAV files; 44.1KHz, 16 bit) placed on a table in front of the subject. Each participant gave prior informed consent before joining the study.

Our ultimate goal is to develop an algorithm for large-scale screening and early diagnosis of cognitive deterioration. Thus our analysis is focused on the discrimination between HC and MCI groups: but, even if the findings presented in this work are mainly devoted to HC/MCI distinction, the whole corpus has been fully transcribed and annotated. Unfortunately, some of the sessions could not be analyzed due to recording quality problems, mainly excessive noise: after all, only 76 sessions (44 HC, 32 MCI) over 80 have been processed.

2.2 Corpus processing

The speech samples have been transcribed by using *Transcriber*,[3] a free tool for assisting the manual annotation of speech signals (see Figure 1 for an example of transcription with the tool). Output files are exported in an XML format with the temporal alignment of the text to the signal and Unicode UTF-8 character encoding. The annotation procedure is fully compliant with the guidelines of the project, available to the transcribers.

[3] `http://trans.sourceforge.net`.

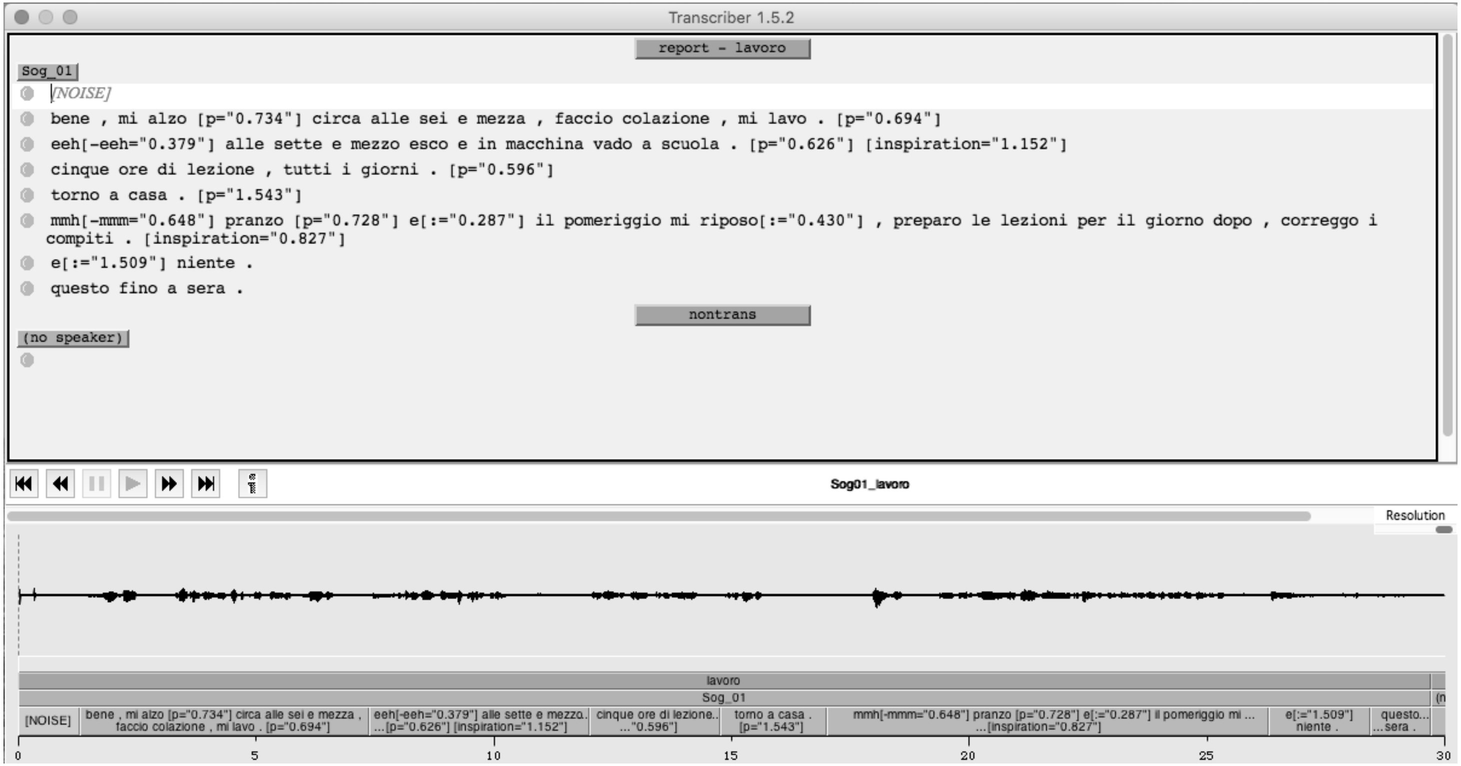

FIGURE 1: ORTHOGRAPHIC TRANSCRIPTION OF A SHORT SPEECH SAMPLE WITH THE TRANSCRIBER SOFTWARE. FROM THE TOP: TEXT EDITOR FOR SPEECH TURN ANNOTATION AND SEGMENTATION; BUTTON BAR FOR SIGNAL PLAYBACK; WAVEFORM; SEGMENTATION LINES (SYNCHRONIZED WITH THE SIGNAL).

The reference unit for the analysis of the speech flow is the utterance, defined by using pragmatic and prosodic (mainly intonational) criteria as "the linguistic counterpart to the speech act", the minimal linguistic entity that is pragmatically interpretable (Austin 1962). The identification of this unit is performed through the perception and the detection of "prosodic breaks" (Cresti & Moneglia 2018) acoustically correlating with F0 reset,[4] final lengthening, drop in intensity, pause and initial rush in the subsequent prosodic unit. As a matter of fact, the identification of breaks reaches high inter-rater agreement in annotation, also among non-expert annotators (Cohen's *k* for Italian around 0.8, Danieli *et al.* 2004), thus being a highly reliable chunking method. One or more utterances performed without interruptions by a single speaker make up a "dialogic turn".

Orthographic transcription follows the conventions of written Standard Italian; to dispel any spelling doubts, the annotators referred to the "GRADIT" dictionary (De Mauro 1999). During the transcription process, a set of paralinguistic and extralinguistic phenomena (such as empty or filled pauses, disfluencies, lapsus, hesitations/stuttering, laughs, coughs, throat clearing sounds or noises) has been annotated as well. All labels were neatly marked to allow easy removal of the tags from the corpus and reversion to the raw data (Leech 2005). The duration of the linguistic and non-linguistic events (in ms.) has been annotated too, gauging their temporal extension on the spectrogram using the Praat speech processing tool[5] (Boersma 2001).

After their automatic tokenisation,[6] transcriptions have been enriched with lexical and morphosyntactic annotations: all the utterances have been automatically part-of-speech (PoS) tagged, lemmatized and syntactically parsed with the dependency parser contained in the Turin University Linguistic Environment – TULE[7] (Lesmo 2007), which is based on the Turin University TreeBank annotation schema (Bosco *et al.* 2000). The whole process consists in associating the grammatical category with each word/token, converting each wordform to the standardised citational form from the dictionary (its corresponding lemma), and associating a dependency-based syntactic structure with the whole sentence.

[4] The fundamental frequency or F0 is the frequency at which vocal chords vibrate in voiced sounds. In particular, the term refers to the lowest frequency component in a complex sound wave. It is the fundamental acoustic parameter in studies about intonation.

[5] `http://www.fon.hum.uva.nl/praat/`.

[6] Tokenisation is the process of breaking up a text into units such as words, dates, addresses, acronyms, complex symbols and yet other elements (e.g. punctuation marks), called "tokens".

[7] `https://github.com/alexmazzei/TULE`.

As can be seen from the example *'torno a casa'* ('I get home') extracted from our corpus, the software returns, for each token, its corresponding lemma, Part of Speech, morphological description, and grammatical relation:

1	*torno*	TORNARE	VERB	MAIN_IND_PRES_INTRANS _1_SING	0	TOP
2	*a*	A	PREP	_	1	INDCOMPL
3	*casa*	CASA	NOUN	COMMON_F_SING	2	ARG
4	.	.	PUNCT	_	1	END

The syntactic structure of the sentence is described in terms of directed binary grammatical relations holding among words within the general framework of dependency grammars (Tesniére 1959). In the output of the automatic analysis, the link between words is marked through numeric indexes, and every dependency relation is labeled. For example, the root node (i.e., the head of the entire structure, namely the main verb) is marked by the index 0 and the label TOP; the token *a* is linked with the index 1, as the verb "indirect complement", while *casa* represents its "argument". This dependency-style analysis can be transposed into a graph with directed, labeled arcs, as illustrated by Figure 2.

Since syntactic analyzers are known to achieve lower accuracy scores on transcripts than on written texts, and since we had to rely on carefully annotated data, we decided to manually check the output of TULE. The revision has been carried out using the *Dependency Grammar Annotator* - DGA open-source software[8] for easy visualisation and correction of TULE mistakes at any level.

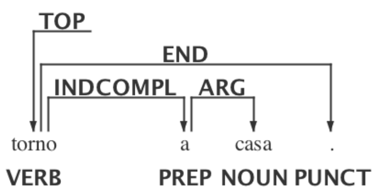

FIGURE 2: DEPENDENCY GRAPH OF THE SIMPLE SENTENCE *'torno a casa'*.

2.3 Linguistic Features as Digital Biomarkers

A multidimensional parameter analysis has been performed on the corpus: reviewing the relevant scientific literature, we identified a wide range of linguistic parameters which can potentially support the detection of MCI. In a nutshell, the verbal productions of these patients have been described as less flu-

[8] `http://medialab.di.unipi.it/Project/QA/Parser/DgAnnotator/`.

ent, syntactically reduced, and lexically poorer compared to HC (§1.1). Moreover, some studies on the English language described subtle alterations of the voice, both at the acoustic and rhythmic level. Therefore, we based our analysis on a set of 67 linguistic/stylometric indexes. Actually, we introduced a few new indexes, while some others have already been suggested by other scholars.

In particular, acoustic measures aim at probing some temporal characteristics of the speech – namely, the incidence of disfluencies and pauses – and testing some properties of the voice, such as the regularity and the complexity of the signal and the brightness of the sound. Rhythmic features capture the proportion of vocalic and consonant intervals, thus correlating with articulation deficits. Lexical indexes quantified the composition of the speaker's vocabulary (e.g., the incidence of specific PoS or linguistic elements, such as deictics or pronouns) and its richness. Finally, syntactic parameters measure the complexity of phrases and sentences.

We developed some specific algorithms for extracting these features automatically from the corpus, and performing a complete quantitative analysis of the spoken texts (§2.4). Age and Cognitive Reserve[9] (approximated using the "schooling" parameter) are among the most important risk factors for MCI (Mazzeo *et al.* 2019). Since these two parameters are available in our clinical dataset, they have been added to the analysis as "Demographic features".

Tables from 1 to 6 provide the complete list of all the features considered in this study, together with a short description and the publications that have proposed them.[10] For a thorough description, the reader is referred to Calzà *et al.* (2021).

FEATURE (REF.)	DESCRIPTION
Silence segments duration (Satt *et al.* 2013)	Silence segments of the signal (SPE_SIL_M, SPE_SIL_MD, SPE_SIL_SD).
Speech segments duration (Satt *et al.* 2013)	Speech segments of a signal (SPE_SPE_M, SPE_SPE_MD, SPE_SPE_SD).

[9] Cognitive Reserve (CR) accounts for individual differences in susceptibility to age-related brain changes, brain damage, or AD-related pathology. This construct posits that individual differences due to enriching lifetime experiences (e.g., education, occupational attainment, social and leisure activities), allow some people to cope better than others to neural network disruption.

[10] Feature names ending with '_M' refer to the mean value, those ending with '_MD' refer to the median and those ending with '_SD' to the standard deviation of the feature in the whole speech production.

Temporal regularity of voiced segments (Satt *et al.* 2013)	The measure captures the temporal structure of the voiced segments, providing information on the rate of change in the different spectrum bands (SPE_TRVS).
Verbal Rate (Singh *et al.* 2001; Roark *et al.* 2011)	The number of tokens in the sample divided by the Total Locution Time (i.e., speech time including pauses) (SPE_VR).
Transformed Phonation Rate (Singh *et al.* 2001; Roark *et al.* 2011)	It measures the ratio between the total phonation time (i.e., speech time without pauses) and the total locution time (i.e., speech time including pauses) (SPE_TPR).
Standardized Phonation Time (Singh *et al.* 2001; Roark *et al.* 2011)	The number of tokens in the sample divided by the total phonation time (i.e., speech time excluding pauses) (SPE_SPT).
Standardized Pause Rate (Singh *et al.* 2001; Roark *et al.* 2011)	The number of tokens in the sample divided by the number of pauses (SPE_SPR).
Root Mean Square energy (López-de-Ipiña *et al.* 2013)	The energy of a signal at a specific time represents the amount of effort made by the speaker in producing a specific sound (SPE_RMSE_M, SPE_RMSE_SD).
Pitch (López-de-Ipiña *et al.* 2013)	Fundamental (F0) frequency in the speech sample (SPE_PITCH_M, SPE_PITCH_SD).
Spectral Centroid (López-de-Ipiña *et al.* 2013)	The measure captures the perceptual 'brightness' of a sound (SPE_SPCENTR_M, SPE_SPCENTR_SD).
Higuchi Fractal Dimension (López-de-Ipiña *et al.* 2013)	The feature describes the complexity of the signal at a specific time (SPE_HFractD_M, SPE_HFractD_SD).

TABLE 1: ACOUSTIC FEATURES.

FEATURE	DESCRIPTION
Age	Subject's age (NPT_AGE).
Schooling	Subject's number of years at school (NPT_SCHOOL).

TABLE 2: DEMOGRAPHIC FEATURES.

FEATURE (REF.)	DESCRIPTION
Text readability (Dell'Orletta *et al.* 2011)	It is a set of four readability features as computed by the READ-IT readability assessment tool: it computes a lexical-based (REA_BASE), a morpho-syntactic (REA_MOSYN), a syntactic (REA_SYNTAX) readability indexes and a combination of the previous three indexes (REA_ALL).

TABLE 3: READABILITY FEATURES.

FEATURE (REF.)	DESCRIPTION
Percentage of vocalic intervals (Ramus *et al.* 1999)	The proportion of vocalic intervals within the utterance, that is, the sum of vocalic intervals divided by the total duration of the utterance (RHY_%V).
Std. deviation of vocalic and consonantal interval durations (Ramus *et al.* 1999)	The standard deviation of the duration of vocalic and consonantal intervals within each utterance, noted as ΔV (RHY_DeltaV) and ΔC (RHY_DeltaC).
Pairwise Variability Index (Grabe & Low 2002)	This rhythm metric takes into account the temporal succession of the vocalic and consonantal intervals (RHY_VnPVI, RHY_CrPVI).
Variation coefficient for ΔV and ΔC (Delwo 2006)	A variation coefficient ("varco") is a value describing relative variation in consonants (RHY_VarcoC) and vowels (RHY_VarcoV).

TABLE 4: RHYTMIC FEATURES.

FEATURE (REF.)	DESCRIPTION
Content Density (Roark *et al.* 2011)	The ratio of open-class tokens to closed-class tokens (LEX_ContDens).
Part-of-Speech rate (Holmes & Singh 1996; Bucks *et al.* 2000)	This class of features investigates the average rate of occurrence for each Part-of-Speech category (LEX_PoS_*).
Reference Rate to Reality (Vigorelli 2004)	The ratio of the total number of nouns to the total number of verbs (LEX_RefRReal).
Personal, Spatial and Temporal Deixis rate (March *et al.* 2006; Cantos-Gòmez 2009)	The feature probes the rate of this three kind of deictic expressions in the spoken text w.r.t. the total number of tokens (LEX_PDEIXIS. LEX_SDEIXIS, LEX_TDEIXIS).
Relative pronouns and negative adverbs rate	The rate of Relative Pronouns (LEX_RPRO) and Negative Adverbs (LEX_NEGADV) in the spoken text.

Lexical Richness (Holmes & Singh 1996; Brunet 1978; Honoré 1979)	This class of measures quantifies the richness of vocabulary/lexical diversity: *TTR, Type-Tokens Ratio* (LEX_TTR), *W - Brunet's Index* (LEX_BrunetW), *R - Honoré's Statistic* (LEX_HonoreR).
Action Verbs rate (Gagliardi 2014)	The metric probes the rate of action verbs (i.e., verbs referring to physical action, like "to put", "to run", "to eat") in the spoken text (LEX_ACTVRB).
Frequency-of-use (De Mauro 2000)	Mean frequency-of-use weight among words extracted from the De Mauro's frequency list (LEX_DM_F).
Propositional Idea Density (Snowdon *et al.* 1996; Roark *et al.* 2011)	Idea density is the number of expressed propositions (verbs, adjectives, adverbs, prepositions, and conjunctions) divided by the number of tokens. Nouns are not considered to be propositions, as the main verb and all its arguments count as one proposition (LEX_IDEAD).
Mean Number of tokens in utterances	Mean number of tokens in the speech utterances (LEX_NW).

TABLE 5: LEXICAL FEATURES.

FEATURE (REF.)	DESCRIPTION
Number of dependent elements linked to the noun	The feature explores Noun Phrase complexity, counting the number of dependent elements linked to the head, e.g. Adjectives, Relative clauses, etc. (SYN_NPLEN_M, SYN_NPLEN_SD).
Global Dependency Distance (Roark *et al.* 2007, 2011)	Given the memory overhead of long distance dependencies, the feature quantifies the difficulty in syntactic processing (SYN_GRAPHDIST_M, SYN_GRAPHDIST_SD).
Syntactic complexity (Szmrecsányi 2004)	Syntactic complexity is established by counting the linguistic tokens that can be considered to telltale signs of increased grammatical subordinateness and embeddedness, such as subordinating conjunctions, WH-pronouns, verb forms, both finite and non-finite and noun phrases. Because subordinators and WH-pronouns are the most straightforward indicators of increased embeddedness (and thus of high complexity), these features are weighted twice than verb forms and noun phrases (SYN_ISynCompl).
Syntactic embeddedness	The maximum "depth" of the dependency structure (SYN_MAXDEPTH_M, SYN_MAXDEPTH_SD).
Utterance length	The number of tokens per utterance (SYN_SLEN_M, SYN_SLEN_SD).

TABLE 6: SYNTACTIC FEATURES.

2.4 Feature extraction and data processing

To extract all the features depicted in Tables from 1 to 6, we developed a set of algorithms that analyse both raw speech recordings and linguistic annotations, and automatically compute all the described indexes.

Concerning the parameters derived from the speech acoustics, each speech sample had to be preprocessed to automatically extract all the needed information. Figure 3 outlines the various preprocessing steps. First of all, we used the "SSVAD v1.0" Voice Activity Detector proposed by Yu & Mak (2011),[11] specifically developed for interview speech, to automatically segment the recordings and identify speech (marked with 'S' in the figure) vs. non-speech regions (marked with 'h#'). These utterance segmentations provide the fundamental information for computing some acoustic features like silence (unfilled pauses) segments durations, speech segments durations, and their ratios.

We also needed the temporally-aligned phonetic transcription of our samples to compute the duration of vowels and consonants, and the ratios required to extract the rhythmic features listed in Table 4. The first step consisted in phonetically transcribing the orthographic transcription of the speech sample made manually by the annotators; we used the grapheme-to-phoneme module by Cosi *et al.* (2001) based on the SAMPA[12] phonetic alphabet for Italian. To temporally align the phonetic transcription and the acoustic signal, we implemented a forced alignment algorithm, using the Kaldi Automatic Speech Recognition package[13] trained on the APASCI Italian Corpus[14]. This aligner computes the most likely match between the acoustic features in the utterance and the phonetic transcription. The result is a reliable temporal alignment between phones and utterance segments.

Upon completing this complex chain of acoustic preprocessing steps, and the morpho-syntactic and syntactic parses described in section 2.2, the computation of acoustic, rhytmic, lexical and syntactic features simply involves counting various types of elements in the utterances (e.g. word-class frequencies, the number of tokens or types, etc.) and computing ratios between them, as detailed in the following section.

[11] `http://bioinfo.eie.polyu.edu.hk/ssvad/ssvad.htm`.
[12] `https://www.phon.ucl.ac.uk/home/sampa/`
[13] `http://kaldi-asr.org`.
[14] `http://catalogue.elra.info/en-us/repository/browse/ELRA-S0039/`

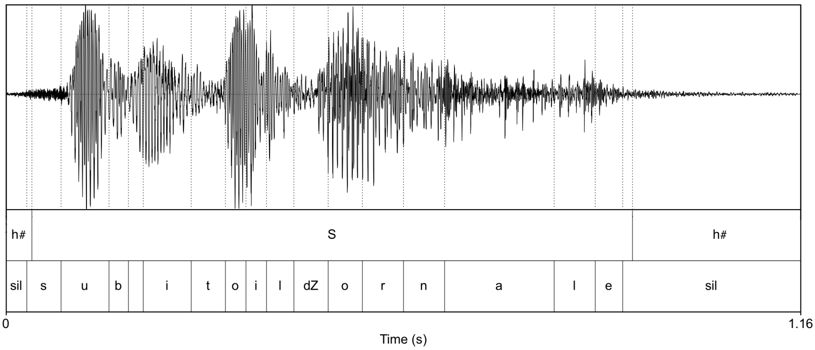

FIGURE 3: THE SPEECH PREPROCESSING STEPS NEEDED TO EXTRACT ALL THE REQUIRED INFORMATION FROM THE SPEECH WAVEFORM (ON TOP). THE SECOND TIER SHOWS THE SPEECH (S) AND NON-SPEECH (H#) SEGMENTS ALTERNATION, WHILE THE THIRD INDICATES THE TEMPORALLY-ALIGNED PHONETIC TRANSCRIPTION OBTAINED AUTOMATICALLY BY FORCED-ALIGNMENT.

3. RESULTS

We analyzed the linguistic biomarkers extracted from subjects' speech productions along two dimensions. First of all, we checked if and how strongly a single feature was able to contribute in discriminating the two groups, HC and MCI subjects. Secondly, we tried to identify a subgroup of linguistic features that, working together, contribute mostly to the classification process.

The first step involved an in-depth analysis of the proposed features by computing their individual statistical significance using the Kolmogorov- Smirnov non-parametric test, which was preferred to the Student's t-test or the Wilcoxon-Mann-Whitney test, because of the small size of our corpus. Table 7 outlines the different levels of significance for the considered features.

FEATURE	SIGNIF.	FEATURE	SIGNIF.
LEX_ACTVRB		NPT_AGE	*
LEX_BrunetW	*	NPT_SCHOOL	*
LEX_ContDens	**	REA_ALL	
LEX_DM_F		REA_BASE	*
LEX_HonoreR		REA_MOSYN	*
LEX_IDEAD		REA_SYNTAX	
LEX_NEGADV		SPE_HFractD_M	***
LEX_NW	**	SPE_HFractD_SD	***
LEX_PDEIXIS		SPE_PITCH_M	
LEX_PoS_ADJ	**	SPE_PITCH_SD	
LEX_PoS_ADV		SPE_RMSE_M	*
LEX_PoS_ART		SPE_RMSE_SD	

LEX_PoS_CONJ		SPE_SIL_M	***
LEX_PoS_INTERJ		SPE_SIL_MD	***
LEX_PoS_NOUN		SPE_SIL_SD	***
LEX_PoS_NUM		SPE_SPCENTR_M	***
LEX_PoS_PHRAS		SPE_SPCENTR_SD	
LEX_PoS_PREDET		SPE_SPE_M	***
LEX_PoS_PREP		SPE_SPE_MD	***
LEX_PoS_PRON		SPE_SPE_SD	***
LEX_PoS_VERB		SPE_SPR	***
LEX_RPRO		SPE_SPT	***
LEX_RefRReal		SPE_TPR	***
LEX_SDEIXIS		SPE_TRVS	
LEX_TDEIXIS		SPE_VR	**
LEX_TTR	*	SYN_GRAPHDIST_M	***
RHY_%V		SYN_GRAPHDIST_SD	
RHY_CrPVI		SYN_ISynCompl	
RHY_DeltaC		SYN_MAXDEPTH_M	**
RHY_DeltaV		SYN_MAXDEPTH_SD	*
RHY_VarcoC		SYN_NPLEN_M	
RHY_VarcoV		SYN_NPLEN_SD	
RHY_VnPVI		SYN_SLEN_M	***
		SYN_SLEN_SD	***

TABLE 7: LINGUISTIC FEATURES CONSIDERED IN THIS STUDY (SEE TABLES 1-6 FOR DESCRIPTIONS AND ABBREVIATIONS), AND THEIR SIGNIFICANCE IN DISTINGUISHING BETWEEN HC AND MCI SUBJECTS USING A KOMOLGOROV-SMIRNOV TEST (* 0.01<P<0.05, ** 0.001<P<0.01, *** P<0.001).

This kind of 'classical' analysis sheds light on the role played by every single linguistic parameter in supporting HC - MCI discrimination. Our results show that nearly all acoustic features (SPE_*) – directly derived from the recordings – play a central, statistically significant role in distinguishing the two classes of subjects.

Regarding lexical (LEX_*) and syntactic (SYN_*) features, some of them are highly significant and, thus, linguistically interesting. Alteration of these parameters, as widely reported in the literature, captures the most evident erosion of verbal abilities throughout the disease. In particular, lexical richness emerges as a statistically relevant parameter. It was quantified in our corpus through the Type-Token ratio (TTR) and the Brunet's Index (W). TTR is the ratio of the number of different words (or word types) the total text length. It is dependent on the text size: it is higher when texts are small and decreases as they get larger. In contrast, W - Brunet's Index quantifies lexical richness without being sensitive to text length. It is calculated according to the equation $W = N^{V^{-0.165}}$, where N stands for the total text length (i.e., the number of tokens) and V is the number of word types used by the participant. This measure generally varies between 10 and 20: the lower the value, the richer the speech./ The relevance of these parameters is in line with the slow stepwise deterioration of the lexico-semantic system described in MCI patients.

Above the word level, a set of indexes suggests a progressive syntactic simplification: the number of words per utterance gradually decreases and verbal productions show a reduced syntactic complexity, with fewer subordinate clauses and long-distance dependencies (cf. syntactic embeddedness and Global Dependency Distance features). However, it is relevant to underline that these phenomena appear in the context of a correct sentence production, characterized by an intact morphosyntactic structure (i.e., well-formed agreement and inflectional markings). All remaining features (rhythmic - RHY_*, readability - REA_* and demographic - NPT_*) seem to be irrelevant – or marginally relevant – in supporting the discrimination of the two subject groups.

One of the main advantages of applying Machine Learning (ML) techniques to NLP problems rests in the opportunity to assess complex feature interactions.

To investigate this, we leveraged an interesting ML algorithm that, once properly trained for solving a classification problem, can produce formal models that can be interpreted and visualized: “Decisions Trees" (Breiman *et al.* 1984), probably one of the very few ML models that are able to provide some insights into input feature interactions. We trained a Decision Tree model to discriminate between HC and MCI subjects by using the linguistic biomarkers described above as input features. Figure 4 shows the formal model induced from classifying our data. In a nutshell, the dataset is considered to be at root which is recursively split into branch-like segments. The hierarchy depicts the sequence of feature application, which in turn reflects the statistical relevance of each feature in distinguishing the classes.

In our model, acoustic features are highly dominant in this tentative prediction process, with 10 over 15 decision nodes represented by them. SPE_* features supporting subjects’ classification ranges from fluency measures (e.g., phonation rate, which occupies the principal node position, and means/dispersion of silence and speech segments durations) to spectral properties of the voice (e.g., Pitch, Spectral Centroid, Higuchi Fractal Dimension) and Root Mean Square Energy. These results are totally in line with the scientific literature describing the linguistic profiles of pre-clinical and clinical stages of Dementia and are often identified as the most robust digital linguistic biomarkers detecting this pathology. Lexical-syntactic simplification, namely the reduction of noun phrase complexity, a lower Content Density and the anomalous usage of personal deictics, and insidious rhythmical alteration (i.e., RHY_DeltaC) complete the picture, supporting speech features in the classification process.

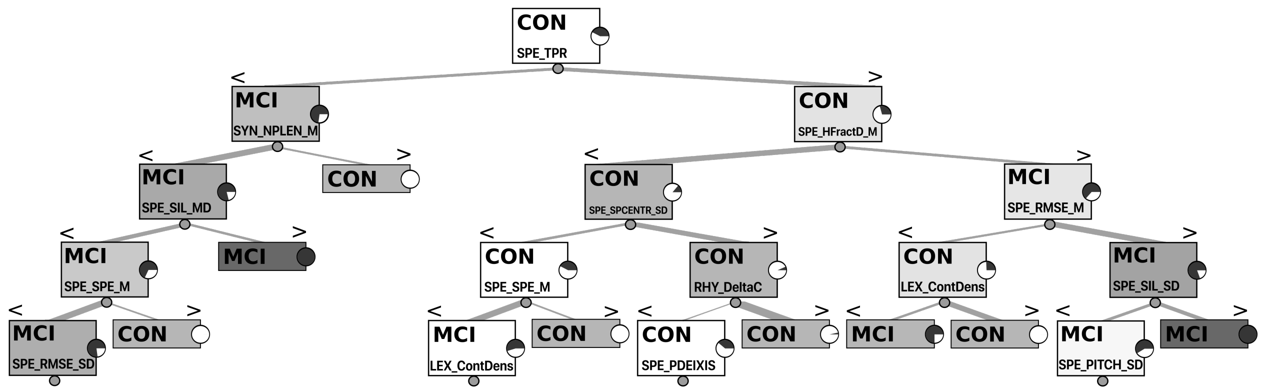

FIGURE 4: DECISION TREE MODEL AFTER TRAINING. AT EACH TREE NODE, THE MODEL HAS TO TAKE A DECISION BASED ON THE VALUE OF A LINGUISTIC FEATURE. THE HIGHER NODES IN THE TREE CORRESPOND TO MORE RELEVANT AND DISCRIMINATING FEATURES. IN THE PICTURE, WE REMOVED THE FEATURE VALUES FOR CLARITY AND LEFT THE MOST SALIENT PROPERTIES OF THE MODEL.

4. DISCUSSION AND CONCLUDING REMARKS

In previous work (Calzà *et al.* 2021), we exploited other ML algorithms to build and evaluate a complete classifier able to identify MCI subjects from HC in a reliable way. We also conducted a brief overview of related works that exploit linguistic biomarkers for MCI automatic detection in languages, with an experimental setting comparable to ours.

Table 8 summarizes the selected papers: for the sake of brevity, we included only the best classification results extracted from them, keeping the focus of our discussion on the HC vs. MCI classification tasks. Since all our experiments used the whole set of available features to avoid any bias in the evaluation phase, to allow for a fair comparison with our setting, we only report the results obtained by the other studies without any feature selection.

REFERENCE	LANGUAGE	BEST RESULTS
Vincze *et al.* (2016)	Hungarian	68.9%
Asgari *et al.* (2017)	English	71.7%
Tóth *et al.* (2018)	Hungarian	75.0%
Themistocleous *et al.* (2018)	Swedish	65.8%
Gosztolya *et al.* (2019)	Hungarian	78.3%
Fraser *et al.* (2019a)	Swedish	68.3%
Fraser *et al.* (2019b)	English	62.8%
	Swedish	71.9%
Calzà *et al.* (2021)	Italian	74.5%

TABLE 8: PREVIOUS RESULTS FOR WORKS DEVOTED TO MCI DETECTION DIRECTLY COMPARABLE TO OUR STUDY IN CALZÀ *et al.* (2021). RESULTS ARE EXPRESSED AS F1-SCORES.

Our results on Italian are in line with, or exceed the state-of-the-art for other languages, presenting a macro-averaged F1-score[15] around 75%. Moreover, upon examining the overall results, we can boldly claim that the dream of building computational tools supporting for massive screening of cognitive frailty is likely to become a concrete reality in the next few years.

The present study expands on previous works, with the same final aim of developing a novel system for the automatic detection of Mild Cognitive Impairment conditions, and investigates the relevance of specific (para-)linguistic features in supporting automatic discrimination. In particular, we tried to out-

[15] The F1-score is an overall measure of a classification model's accuracy. The higher the F1, the better the classification results.

line a communication profile of MCI patients, with a view to identifying those linguistic skills that are most vulnerable to erosion in the earliest stages of the disease.

To this aim, we applied a complex pipeline for the automatic extraction of several linguistic biomarkers (ranging from acoustic, rhythmic, lexical, and syntactic ones) from our spoken corpus. Then, we evaluated the discriminative power of the linguistic parameters, together with some readability indexes and demographic information, from a statistical point of view. We also applied a Decisions Tree ML model to the data to highlight the effect of the interaction among linguistic traits for diagnosis purposes. To the best of our knowledge, this is the first study on the Italian language examining a large set of linguistic features through NLP techniques for the early identification of cognitive decline.

Our results clearly show that acoustic features are the most reliable parameters for MCI - HC classification. As a matter of fact, verbal productions of frail patients are less fluent, and their voice undergoes subtle modifications that are not perceivable by clinicians but can be easily detected through automatic speech analysis. This result suggests that standard batteries of traditional neuropsychological tests can usefully be combined with an analysis of the acoustic parameters of a patient's spoken language productions. While neuropsychological tests and structured evaluations have a relevant impact on the naturalness of the subject's responses, the analysis of natural spoken language productions can offer an ecological, non-invasive, inexpensive and reliable test for the screening of at-risk patients even by primary care physicians.

Besides, the statistical relevance of other linguistic traits (e.g., lexical and syntactic), as shown by their presence in high nodes of our Decision Tree model, suggests that a comprehensive multidimensional analysis is better than a single-domain one.

Nonetheless, further investigations are needed to validate this hypothesis. To confirm and extend our results, our future steps will involve (i) an increase of the sample size of the current survey and (ii) extensive testing of other Machine Learning approaches to early MCI detection.

REFERENCES

Ahmed, S., R. Arnold, S.A. Thompson, K.S. Graham & J.R. Hodges (2008). Naming of objects, faces and buildings in Mild Cognitive Impairment. *Cortex*, 44(6). 746–752.

Ahmed, S., A.M.F. Haigh, C.A. De Jager & P. Garrard (2013). Connected speech as a marker of disease progression in autopsy-proven Alzheimer's disease. *Brain*, 136(12). 3727–3737.

Altmann, L.J.P., D. Kempler & E.S. Andersen (2001). Speech Errors in Alzheimer's Disease: Reevaluating Morphosyntactic Preservation. *Journal of Speech, Language, and Hearing Research*, 44. 1069–1082.

Arias-Vergara, T., J.C. Vasquez Correa, J.R. Orozco-Arroyave & E. Nöth (2018). Speaker models for monitoring Parkinson's disease progression considering different communication channels and acoustic conditions. *Speech Communication*, 101. 11–25.

Asgari, M., J. Kaye & H. Dodge (2017). Predicting Mild Cognitive Impairment from spontaneous spoken utterances. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 3(2). 219–228.

Ash, S., C. McMillan, R.G. Gross, P. Cook, B. Morgan, A. Boller, M. Dreyfuss, A. Siderowf & M. Grossman (2011). The organization of narrative discourse in Lewy body spectrum disorder. *Brain & Language*, 119. 30–41.

Austin, J.L. (1962). *How to Do Things with Words*. Oxford: Clarendon Press.

Beltrami, D., G. Gagliardi, R. Rossini Favretti, E. Ghidoni, F. Tamburini & L. Calzà (2018). Speech analysis by Natural Language Processing techniques: a possible tool for very early detection of cognitive decline? *Frontiers in Aging Neuroscience*, 10. 369.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, 5(9/10). 341–345.

Boschi, V., E. Catricalà, M. Consonni, C. Chesi, A. Moro & S.F. Cappa (2017). Connected speech in neurodegenerative language disorders: A review. *Frontiers in Psychology*, 8. 1–21.

Bosco, C., V. Lombardo, D. Vassallo & L. Lesmo (2000). Building a treebank for Italian: a data-driven annotation schema. In *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC-2000)*. ELRA - European Language Resources Association.

Breiman, L., J. Friedman, R. Olshen, & C. Stone (1984). *Classification and Regression Trees*. Belmont, CA: Wadsworth.

Brodaty, H., N.M. Kemp & L.F. Low (2004). Characteristics of the GPCOG, a screening tool for cognitive impairment. *International Journal of Geriatric Psychiatry*, 19(9). 870–874.

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

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

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

Calzà, L., G. Gagliardi, R. Rossini Favretti & F. Tamburini (2021). Linguistic features and automatic classifiers for identifying Mild Cognitive Impairment and dementia. *Computer Speech & Language*, 65. 101–113.

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

Carlomagno, S., A. Santoro, A. Menditti, M. Pandolfi & A. Marini (2005). Referential communication in Alzheimer's type dementia. *Cortex*, 41(4). 520–534.

Catricalà, E., P. Della Rosa, V. Plebani, D. Perani, P. Garrard & Stefano Cappa F. (2015). Semantic feature degradation and naming performance. Evidence from neurodegenerative disorders. *Brain & Language*, 147. 58–65.

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

Ciurli, P., P. Marangolo & A. Basso (1996). *Esame del Linguaggio - II*. Firenze: Giunti Organizzazioni Speciali.

Cosi, P., F. Tesser, R. R. Gretter & C. Avesani (2001). Festival Speaks Italian! In *Proceedings of Eurospeech 2001*. Paris: ISCA, 509–512.

Cresti, E. & M. Moneglia (2018). The illocutionary basis of Information Structure: The Language into Act Theory (L-AcT). In *Chapter 13. The illocutionary basis of information structure*, volume Information Structure in Lesser-described Languages: Studies in prosody and syntax, 360–402.

Cuetos, F., J.C. Arango-Lasprilla, C. Uribe, C. Valencia & F. Lopera (2007). Linguistic changes in verbal expression: a preclinical marker of Alzheimer's disease. *Journal of the International Neuropsychological Society : JINS*, 13(3). 433–439.

Danieli, M., M. Garrido, M. Moneglia, A. Panizza, S. Quazza & M. Swerts (2004). Evaluation of consensus on the annotation of Prosodic Breaks in the Romance Corpus of Spontaneous Speech "C-ORAL-ROM". In M.T. Lino, M.F. Xavier, F. Ferraira, R. Costa & R. Silva (eds.) *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*. Paris: ELRA, 1513–1516.

de Lira, J.O., K.Z. Ortiz, A. Carvalho Campanha, P.H. Ferreira Bertolucci & T.S. Cianciarullo Minett (2011). Microlinguistic aspects of the oral narrative in patients with Alzheimer's disease. *International Psychogeriatrics*, 23(3). 404–412.

De Mauro, T. (1999). *GRADIT: Grande dizionario italiano dell'uso*. Torino: UTET.

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

Dell'Orletta, F., S. Montemagni & G. Venturi (2011). READ–IT: Assessing readability of Italian texts with a view to text simplification. In *Proceedings of the Second Workshop on Speech and Language Processing for Assistive Technologies*. Edinburgh, Scotland, UK, 73–83.

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

dos Santos, L.B., E.A. Corrêa Jr., O.N. Oliveira Jr, D.R. Amancio, Letícia L. Mansur & S.M. Aluísio (2017). Enriching complex networks with Word Embeddings for detecting Mild Cognitive Impairment from speech transcripts. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, volume 1: Long Papers. Association for Computational Linguistics, 1284–1296.

Drummond, C., G. Coutinho, R.P. Fonseca, N. Assunção, A. Teldeschi, R. de Oliveira-Souza, J. Moll, F. Tovar-Moll & P. Mattos (2015). Deficits in narrative discourse elicited by visual stimuli are already present in patients with Mild Cognitive Impairment. *Frontiers in Aging Neuroscience*, 7(96).

Fergadiotis, G. & H.H. Wright (2011). Lexical diversity for adults with and without aphasia across discourse elicitation tasks. *Aphasiology*, 25. 1414–1430.

Filiou, R.P., N. Bier, A. Slegers, B. Houzé, P. Belchior & S.M. Brambati (2019). Connected speech assessment in the early detection of Alzheimer's disease and Mild Cognitive Impairment: a scoping review. *Aphasiology*.

Folstein, M., S.E. Folstein & P.R. McHugh (1975). Mini-Mental State: a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Reserach*, 12. 189–198.

Forbes-McKay, K., M.F. Shanks & A. Venneri (2013). Profiling spontaneous speech decline in Alzheimer's disease: a longitudinal study. *Acta Neuropsychiatrica*, 25(6). 320–327.

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

Fraser, K.C., J.A. Meltzer & F. Rudzicz (2016). Linguistic features identify Alzheimer's Disease in narrative speech. *Journal of Alzheimer's Disease*, 49. 407–422.

Fraser, K.C., K. Lundholm Fors, M. Eckerström, F. Öhman & D. Kokkinakis (2019a). Predicting MCI status from multimodal language data using Cascaded Classifiers. *Frontiers in Aging Neuroscience*, 11. 205.

Fraser, K.C., K. Lundholm Fors & D. Kokkinakis (2019b). Multilingual word embeddings for the assessment of narrative speech in Mild Cognitive Impairment. *Computer Speech & Language*, 53. 121–139.

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

Galasko, D.R., R. Katzman, D.P. Salmon & L. Hansen (1996). Clinical and neuropathological findings in Lewy Body Dementias. *Brain and Cognition*, 31. 166–175.

Gorno-Tempini, M.L., A. Hillis, S. Weintraub, A. Kertesz, M. Mendez, S.F. Cappa, J. Ogar, J.D. Rohrer, S.E. Black, B. Boeve, F. Manes, N. Dronkers, R. Vandenberghe, K. Rascovsky, K. Patterson, B.L. Miller, D.S. Knopman, J.R. Hodges, M. Mesulam & M. Grossman (2011). Classification of Primary Progressive Aphasia and its variants. *Neurology*, 76. 1006–1014.

Gosztolya, G., V. Vincze, L. Tóth, M. Pákáski, J. Kálmán & I. Hoffmann (2019). Identifying Mild Cognitive Impairment and mild Alzheimer's disease based on spontaneous speech using ASR and linguistic features. *Computer Speech & Language*, 53. 181–197.

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

Hoffmann, I., D. Nemeth, C. Dye, M. Pákáski, T. Irinyi & J. Kálmán (2010). Temporal parameters of spontaneous speech in Alzheimer's disease. *International journal of speech-language pathology*, 12(1). 29–34.

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

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

Istituto Superiore di Sanità (2021). Characteristics of COVID-19 patients dying in Italy - Report based on available data on 1st March, 2021. `https://www.epicentro.iss.it/en/coronavirus/bollettino/Report-COVID-2019_1_march_2021.pdf`.

Jarrold, W.L., B. Peintner, D. Wilkins, D. Vergryi, C. Richey, M.L. Gorno-Tempini & J. Ogar (2014). Aided diagnosis of Dementia type through computer-based analysis of spontaneous speech. In P. Resnik, R. Resnik & M. Mitchell (eds.) *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. ACL - Association for Computational Linguistics, 27–37.

Jiang, H., B. Hu, Z. Li, L. Yan, T. Wang, F. Liu, H. Kang & X. Li (2017). Investigation of different speech types and emotions for detecting depression using different classifiers. *Speech Communication*, 90. 39–46.

Kemper, S., E. LaBarge, H. Ferraro, F. Richard Cheung, H. Cheung & M. Storandt (1993). On the preservation of syntax in Alzheimer's Disease. *Archives of neurology*, 50(1). 81–86.

Kempler, D., S. Curtiss & C. Jackson (1987). Syntactic preservation in Alzheimer's disease. *Journal of Speech and Hearing Research*, 30(3). 343–350.

Kuhn, T.S. (1962). Historical structure of scientific discovery. *Science*, 136. 760–764.

Leech, G. (2005). Adding linguistic annotation. In M. Wynne (ed.) *Developing Linguistic Corpora: a Guide to Good Practice*. Oxford: Oxbow Books.

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

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

López-de-Ipiña, K., J. Solé-Casals, H. Eguiraun, J. Alonso, C. Travieso, A. Ezeiza, N. Barroso, M. Ecay-Torres, P. Martinez-Lage & B. Beitia (2015). Feature selection for spontaneous speech analysis to aid in Alzheimer's disease diagnosis: A fractal dimension approach. *Computer Speech & Language*, 30(1). 43 – 60.

March, E.G., R. Wales & P. Pattison (2006). The uses of nouns and deixis in discourse production in Alzheimer's disease. *Journal of Neurolinguistics*, 19. 311–340.

Mazzeo, S., S. Padiglioni, S. Bagnoli, L. Bracco, B. Nacmias, S. Sorbi & V. Bessi (2019). The dual role of cognitive reserve in subjective cognitive decline and Mild Cognitive Impairment: a 7-year follow-up study. *Journal of Neurology*, 266. 487–497.

Mesulam, M. (2003). Primary Progressive Aphasia: A language-based dementia. *New England Journal of Medicine*, 349(16). 1535–1542.

Mortamais, M., J.A. Ash, J. Harrison, J. Kaye, J. Kramer, C. Randolph, C. Pose, B. Albala, M. Ropacki, C.W. Ritchie & K. Ritchie (2017). Detecting cognitive changes in preclinical Alzheimer's disease: A review of its feasibility. *Alzheimer's & Dementia*, 13. 468–492.

Nasreddine, Z.S., N.A. Phillips, V. Bédirian, S. Charbonneau, V. Whitehead, I. Collin, J.L. Cummings & H. Chertkow (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for Mild Cognitive Impairment. *Journal of the American Geriatrics Society*, 53(4). 695–699.

Oulhaj, A., G.K. Wilcock, A.D. Smith & C.A. de Jager (2009). Predicting the time of conversion to MCI in the elderly: role of verbal expression and learning. *Neurology*, 73(18). 1436–1442.

Patterson, C. (2018). The World Alzheimer Report 2018. Technical report, Alzheimer's Disease International.

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

Petersen, R.C. (2011). Clinical practice. Mild Cognitive Impairment. *New England Journal of Medicine*, 364(23). 2227–2234.

Ramus, F., M. Nespor & J. Mehler (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73. 265–292.

Ripich, D.N., B.D. Carpenter & E.W. Ziol (2000). Conversational cohesion patterns in men and women with Alzheimer's disease: a longitudinal study. *International journal of language & communication disorders / Royal College of Speech & Language Therapists*, 35(1). 49–64.

Ritchie, K., M. Ropacki, B. Albala, J. Harrison, J. Kaye, J. Kramer, C. Randolph & C.W. Ritchie (2017). Recommended cognitive outcomes in preclinical Alzheimer's disease: Consensus statement from the European Prevention of Alzheimer's Dementia project. *Alzheimer's & Dementia*, 13. 186–195.

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

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

Sajjadi, S.A., K. Patterson, M. Tomek & P.J. Nestor (2012). Abnormalities of connected speech in semantic dementia vs. Alzheimer's disease. *Aphasiology*. 1–20.

Satt, A., A. Sorin, O. Toledo-Ronen, O. Barkan, I. Kompatsiaris, A. Kokonozi & M. Tsolaki (2013). Evaluation of speech-based protocol for detection of Early-Stage Dementia. In *INTERSPEECH 2013, 14th annual conference of the International Speech Communication Association*. ISCA, 1692–1696.

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

Sirts, K., O. Piguet & M. Johnson (2017). Idea density for predicting Alzheimer's disease from transcribed speech.

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

Snowdon, D.A., S.J. Kemper, J.A. Mortimer, L.H. Greiner, D.R. Wekstein & W.R. Markesbery (1996). Linguistic Ability in Early Life and Cognitive Function and Alzheimer's Disease in Late Life: Findings From the Nun Study. *The journal of the American Medical Association*, 275. 528–532.

Stasak, B., J. Epps & R. Goecke (2019). Automatic depression classification based on affective read sentences: opportunities for text-dependent analysis. *Speech Communication*, 115. 1–14.

Szatloczki, G., I. Hoffmann, V. Vincze, J. Kalman & M. Pakaski (2015). Speaking in Alzheimer's disease, is that an early sign? Importance of changes in language abilities in Alzheimer's disease. *Frontiers in Aging Neuroscience*, 7. 195.

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

Taler, V. & N.A. Phillips (2008). Language performance in Alzheimer's disease and Mild Cognitive Impairment: A comparative review. *Journal of Clinical and Experimental Neuropsychology*, 30(5). 501–556.

Tesniére, L. (1959). *Elements de syntaxe structurale*. Paris: Libraire C. Klincksieck.

Themistocleous, C., M. Eckerström & D. Kokkinakis (2018). Identification of Mild Cognitive Impairment from speech in Swedish Using Deep Sequential Neural Networks. *Frontiers in Neurology*, 9. 975.

Tóth, L., I. Hoffmann, G. Gosztolya, V. Vincze, G. Szatlóczki, Z. Bánréti, M. Pákáski & J. Kálmán (2018). A Speech Recognition-based solution for the automatic detection of Mild Cognitive Impairment from spontaneous speech. *Current Alzheimer Research*, 15. 1–10.

Tsantali, E., D. Economidis & M. Tsolaki (2013). Could language deficits really differentiate Mild Cognitive Impairment (MCI) from mild Alzheimer's disease? *Archives of gerontology and geriatrics*, 57(3). 263–270.

Tsoi, K.K., J. Chan, H. Hirai, S.Y. Wong & T. Kwok (2015). Cognitive tests to detect Dementia: a systematic review and meta-analysis. *JAMA Internal Medicine Sign In Individual Sign In Sign inCreate an Account*, 175(9). 1450–1458.

Tulving, E. (1972). Episodic and Semantic Memory. In E. Tulving & W. Donaldson (eds.) *Organization of Memory*, 381–403. New York: Academic Press.

Upadhya Savitha, S., A.N. Cheeran & J.H. Nirmal (2019). Discriminating Parkinson diseased and healthy people using modified MFCC filter bank approach. *International Journal of Speech Technology*, 22(4). 1021–1029.

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

Vincze, V., G. Gosztolya, L. Tóth, I. Hoffmann, G. Szatlóczki, Z. Bánréti, M. Pákáski & J. Kálmán (2016). Detecting Mild Cognitive Impairment by exploiting linguistic information from transcripts. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 181–187.

Yancheva, M. & F. Rudzicz (2016). Vector-space topic models for detecting Alzheimer's disease. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2337–2346.

Yu, H. & M. Mak (2011). Comparison of Voice Activity Detectors for Interview Speech in NIST Speaker Recognition Evaluation. In *Proceedings of Interspeech'11*. 2353–2356.

Gloria Gagliardi
Department of Classical Philology and Italian Studies, University of Bologna
Via Zamboni 32 - 40126 Bologna
Italy
e-mail: `gloria.gagliardi@unibo.it`

Fabio Tamburini
Department of Classical Philology and Italian Studies, University of Bologna
Via Zamboni 32 - 40126 Bologna
Italy
e-mail: `fabio.tamburini@unibo.it`