

The Automatic Extraction of Linguistic Biomarkers as a Viable Solution for the Early Diagnosis of Mental Disorders

Gloria Gagliardi, Fabio Tamburini

FICLIT-University of Bologna, Italy

{gloria.gagliardi, fabio.tamburini}@unibo.it

Abstract

Digital Linguistic Biomarkers extracted from spontaneous language productions proved to be very useful for the early detection of various mental disorders. This paper presents a computational pipeline for the automatic processing of oral and written texts: the tool enables the computation of a rich set of linguistic features at the acoustic, rhythmic, lexical, and morphosyntactic levels. Several applications of the instrument - for the detection of Mild Cognitive Impairments, Anorexia Nervosa, and Developmental Language Disorders - are also briefly discussed.

Keywords: Mental Disorders, Automatic Speech and Language Analysis, Digital Linguistic Biomarkers

1. Digital Linguistic Biomarkers: a novel NLP tool for diagnostic purposes

Natural Language Processing methods and tools are becoming increasingly crucial in the medical domain (Wang et al., 2020), with a broad range of applications ranging from direct patient care, to diagnostics, clinical coding, and patient-facing services (Locke et al., 2021). In particular, a growing interest surrounded the possibility to exploit the automatic analysis of speech and language as a sensible early clue of pathological processes.

Language is a complex cognitive function, which relies on broadly distributed brain networks (Catani et al., 2012). Its neurobiological basis is not limited to the left-dominant perisylvian language network (i.e., the neural loop located around the lateral sulcus, composed by the arcuate tract connecting Broca’s, Wernicke’s and Geschwind’s territories, cf. Catani et al. (2005)), but includes a complex of cortical and subcortical brain structures related to motor and sensory-related representations, non-verbal memory skills, emotional processing, and executive functions (Hagoort, 2017; Hertrich et al., 2020). Consequently, even minor brain changes due to mental health issues (e.g., reversible or progressive cerebral atrophy) can result in subtle language alterations.

The term “Digital Linguistic Biomarkers” (henceforth referred to as “DLBs”) indicates the possibility to automatically extract objective indications of the medical state of the patients directly from their verbal productions. These highly accurate and reproducible measurements should allow “for a low-cost pathology detection, classification, and monitoring” (Gagliardi et al., 2021).

In the last ten years, this approach has gained popularity among researchers and clinicians seeking fast, replicable, and objective proxy-measures of mental disorders (Gagliardi, under revision). Beside Dementia, which has represented the earliest area of application,

this approach has been applied to several clinical developmental and acquired pathologies (e.g., Autism Spectrum Disorder, Parkinson’s Disease, and Progressive Supranuclear Palsy) with encouraging results.

In the following pages, we will present our NLP pipeline for DLBs extraction from Italian oral and written texts. The paper is structured as follows: §2 presents the overall structure of the pipeline (i.e., pre-processing steps and DLBs computation), §3 illustrates successful case studies on the Italian language (e.g., cognitive frailty due to dementia, Anorexia Nervosa, and Developmental Language Disorder detection), and §4 discusses the main challenges arising from this technology. In §5, we will address future development directions.

2. The DLB Computational Pipeline

The need for an efficient method to extract a large set of DLBs for studying possible correlations between linguistic features and various mental disorders in Italian has led us to devise an appropriate computational pipeline. This instrument must be capable of performing some basic NLP tasks on spoken or written productions to extract the whole set of DLBs. In particular, the pipeline is structured as a complex set of interacting and flexible processing modules. Figure 1 illustrates the whole pipeline structure.

The tool can process three different kinds of inputs: spoken recordings (as a WAV audio file), raw written texts (TXT), or preprocessed texts in the CoNLL-U format¹ containing morphosyntactic and syntactic analyses.

Table 1 lists the complete set of DLBs produced by the pipeline. They are subdivided into six groups of biomarkers:

- Speech-derived features: i.e., acoustic (SPE) or rhythmic (RHY) DLBs.

¹<https://universaldependencies.org/format.html>

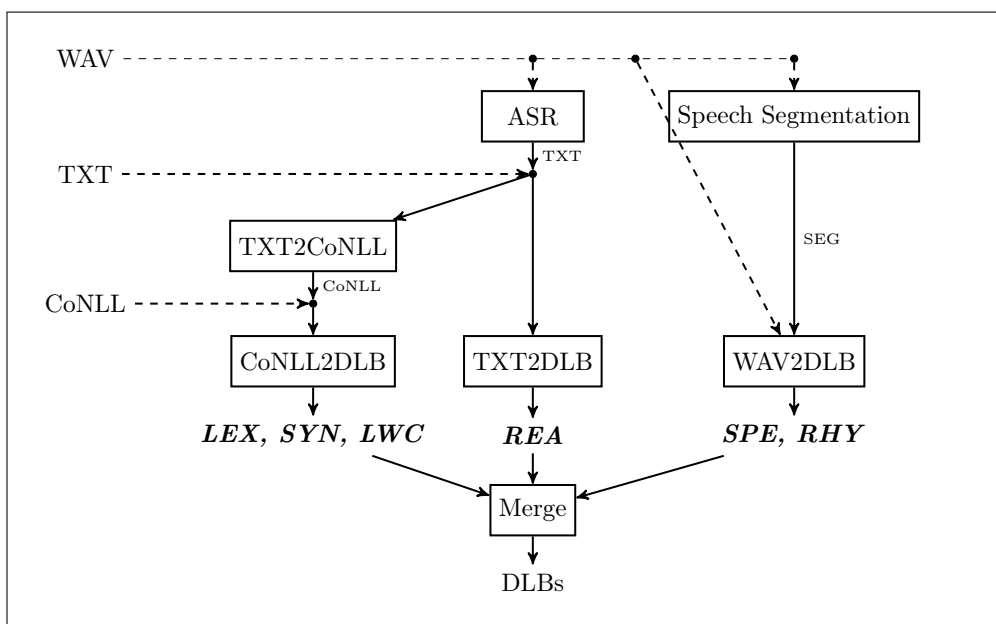


Figure 1: The whole structure of the Pipeline. Inputs can be provided as WAV, raw text, or CoNLL files. The modules are described in detail in Sect. 2.

- Text-derived features: i.e., lexical (LEX), syntactic (SYN), LIWC-based (LWC), and readability (REA) DLBs.

Due to space limits, it is not possible to depict each of these features in detail. For an extensive description and calculation details, the reader can refer to Calzà et al. (2021).

Given a specific input type, either a WAV, TXT, or CoNLL file, the pipeline computes all the DLBs that can be derived from it. The larger set is obtained by providing the speech recording, alone or with the manual transcription (to bypass any mistake produced by the ASR module).

The following subsections will describe in detail all the pipeline modules.

2.1. Pre-Processing Steps

To compute the DLBs, the input data must be pre-processed by applying basic speech analysis and NLP tools.

2.1.1. Speech to Text

If the input consists of a recording only, speech transcription represents a basic requirement for computing reliable text-related DLBs. In this respect, Automatic Speech Recognition, aka ‘speech-to-text’, is the fundamental task.

To this aim, we inserted a specific ASR module into the pipeline, testing two different options:

- The first possibility involves the usage of some pre-trained cloud ASR service provided by IT giants (i.e., Google, Microsoft, IBM, Amazon). We produced the speech transcriptions by leveraging the Google Cloud ASR system, a reliable

instrument for Italian when dealing with non-pathological language. To get an idea about the actual performance of this tool when applied to atypical speech, we ran an evaluation experiment comparing the automatically-derived transcriptions with the manual counterparts, obtaining a Word Error Rate (WER) of 27.78%. Considering the complex nature of the task, this WER seems rather acceptable.

- The second option implies the development of an offline ASR system for Italian. Tamburini (2021) presented a similar tool using the NVIDIA NeMo package. The system shows good performance (WER<10%) when applied to non-pathological speech.

An in-depth focus on the issues raised by ASR will be presented in Section 4.2.

2.1.2. Speech Segmentation

To directly extract the features from speech, the samples have to be preprocessed. First, we used the ‘‘SS-VAD v1.0’’ Voice Activity Detector (VAD) proposed by Yu and Mak (2011)² - tailored for interviewed speech - to automatically segment the recordings and identify speech vs. non-speech regions. These segmentations provide crucial information for computing some acoustic features, such as silence segments durations, speech segments durations, and their ratios.

We also need the temporally-aligned phonetic transcription of the samples to calculate the duration of vowels, consonants, and the ratio of their intervals, for

²<http://bioinfo.eie.polyu.edu.hk/ssvad/ssvad.htm>.

Feature	References
Acoustic Features (SPE)	
Silence segments duration (M, MD, SD)	(Satt et al., 2013)
Speech segments duration (M, MD, SD)	(Satt et al., 2013)
Temporal regularity of voiced segments	(Satt et al., 2013)
Verbal Rate	(Singh et al., 2001; Roark et al., 2011)
Transformed Phonation Rate	(Singh et al., 2001; Roark et al., 2011)
Standardized Phonation Time	(Singh et al., 2001; Roark et al., 2011)
Standardized Pause Rate	(Singh et al., 2001; Roark et al., 2011)
Root Mean Square energy (M, SD)	(López-de-Ipiña et al., 2013)
Pitch (M, SD)	(López-de-Ipiña et al., 2013)
Spectral Centroid (M, SD)	(López-de-Ipiña et al., 2013)
Higuchi Fractal Dimension (M, SD)	(López-de-Ipiña et al., 2013)
Rhythmic Features (RHY)	
Percentage of vocalic intervals - %V	(Ramus et al., 1999)
SD of vocalic, ΔV , and consonantal, ΔC , interval durations	(Ramus et al., 1999)
Pairwise Variability Index, raw, rPVI, and normalized, nPVI	(Grabe and Low, 2002)
Variation coefficient for ΔV and ΔC	(Delwo, 2006)
Lexical Features (LEX)	
Content Density	(Roark et al., 2011)
Part-of-Speech rate	(Holmes and Singh, 1996; Bucks et al., 2000)
Reference Rate to Reality	(Vigorelli, 2004)
Personal, Spatial and Temporal Deixis rate	(March et al., 2006; Cantos-Gòmez, 2009)
Relative pronouns and negative adverbs rate	-
Lexical Richness: TTR, Brunet's and Honoré's Indexes	(Holmes and Singh, 1996)
Action Verbs rate	(Gagliardi, 2014)
Frequency-of-use tagging	(De Mauro, 2000)
Propositional Idea Density	(Snowdon et al., 1996; Roark et al., 2011)
Mean Number of words in utterances	-
Linguistic Inquiry and Word Count Features (LWC)	
Language Metrics (e.g., words per sentence, words > 6 letters)	(Agosti and Rellini, 2007)
Function Words (e.g., pronouns, articles, auxiliary verbs)	
Affect Words (e.g., positive/negative emotion)	
Cognitive Processes (e.g., insight, certainty, tentativeness)	
Perceptual processes (e.g., seeing, hearing, feeling)	
Biological processes (e.g., body, health/illness, ingesting)	
Personal concerns (e.g., work, leisure, money, religion, death)	
Social Words (e.g., family, friends)	
Punctuation (e.g., periods, commas, colons, question marks)	
Readability Features (REA)	
READ-IT features for readability evaluation (at the lexical, morpho-syntactic, syntactic, and global levels)	(Dell'Orletta et al., 2011)
Syntactic Features (SYN)	
Number of dependent elements of the nouns (M, SD)	-
Global Dependency Distance (M, SD)	(Roark et al., 2007; Roark et al., 2011)
Syntactic complexity	(Szmrecsányi, 2004)
Syntactic embeddedness: maximum tree depth (M, SD)	-
Utterance length (M, SD)	-

Table 1: The list of Digital Linguistic Biomarkers extracted by the pipeline. Most of these features are computed as means (M), medians (MD), and standard deviations (SD). Please refer to Calzà et al. (2021) for the descriptions and computation details.

the extraction of the rhythmic features listed in Table 1. The approach pursued in the current version of the pipeline involves: i) the grapheme-to-phoneme conversion of the orthographic transcription, exploiting the grapheme-to-phoneme module by Cosi et al. (2001) (based on the SAMPA³ phonetic alphabet); ii)

the temporal alignment of the phonetic transcription and the acoustic signal. To this aim, we implemented a forced alignment algorithm, using the Kaldi Automatic Speech Recognition package⁴ trained on the APASCI

³<https://www.phon.ucl.ac.uk/home/>

sampa/

⁴<http://kaldi-asr.org>.

Italian Corpus⁵. This aligner predicts the most likely match between the acoustic features in the utterance and the phonetic transcription. The result is a fairly reliable temporal alignment between phones and the utterance segments, thus, it is possible to join vocalic and consonantal intervals and compute the required rhythmic features, starting from that segmentation.

We are currently trying to enhance the computation of vowel/consonant intervals by experimenting with algorithms for the detection of speech landmarks (Stevens et al., 1992; Liu, 1995; Howitt, 2002), in particular applying the approach described in SpeechMark⁶ (Ishikawa et al., 2017).

2.1.3. Text Analysis

The analysis of texts - whatever they derive from written productions or speech transcriptions - usually requires some preprocessing steps to derive basic linguistic features, endowing more complex estimations.

For the extraction of text-related DLBs, we leveraged Stanford STANZA, a collection of efficient neural-network-based tools for the linguistic analysis of many languages, organized into a coherent text processing pipeline (Qi et al., 2020). In particular, we exploit STANZA v1.3.0 for tokenization, lemmatization, part-of-speech tagging, and parsing employing a dependency model trained on the Universal Dependencies⁷ ISDT Italian corpus v2.5.

To get a linguistically sound dependency parse structure for Italian (and in general for any romance language) a proper treatment of enclitics pronoun clusters (e.g., *it. aiutarlo*, en. ‘to help him’; *it. dargliela*, en. ‘to give it to him/her/them’) must be considered. In other terms, it is crucial to ‘break’ the clusters, allowing the correct retrieval of the direct/indirect object arguments of the verb. We applied a powerful Italian morphological analyzer (Tamburini and Melandri, 2012) for identifying enclitic forms and tokenizing them accordingly. Figure 2 shows an Italian sentence after enclitics preprocessing and STANZA analysis.

2.2. Computing DLBs

Table 1 lists the whole set of DLBs computed by the pipeline. Some are strictly connected with speech articulation and thus intimately related to phonetic processing. Others involve text-related features and can be computed on both written productions and transcripts.

2.2.1. Linguistic Biomarkers from Speech

The *Acoustic and Rhythmic features (SPE and RHY)* can be broadly divided into two classes of DLBs:

DLBs based only on Speech Segmentation

All the Acoustic and Rhythmic DLBs broadly involve some unit duration measures, their ratio, or are computed only in speech segments. Then, they are based

on the temporal segmentation of speech samples into homogeneous regions. Acoustic DLBs relies on the results provided by the VAD preprocessing: they are based on various kinds of measures of sample sections containing silence (unfilled pauses) or spoken segments. On the contrary, Rhythmic DLBs are built on the speech segmentation into vocalic and consonantal intervals provided by the Kaldi aligner described before.

DLBs based on Speech Signal analysis

The last four SPE features in Table 1 (i.e., Root Mean Square energy, Pitch, Spectral Centroid, and Higuchi Fractal Dimension) involve some signal processing of the speech sections. Pitch-related measures are computed by using the SWIPE’ pitch tracking algorithm (Camacho, 2007) as implemented into the Speech Signal Processing Toolkit (SPTK)⁸; the others are directly extracted by the pipeline following the techniques described in the cited references.

2.2.2. Linguistic Biomarkers from text

Four types of textual DLBs are currently computed by the pipeline:

Lexical Features (LEX)

Lexical DLBs are computed using frequency measures of particular combinations of tokens, types, and their PoS-tags. In some cases, lists of words - which identify specific linguistic phenomena, like deixis and negation - are exploited. We briefly describe here the most opaque of them. *Content Density* measures the ratio of open-class words over closed-class words. *Reference Rate to Reality* is the ratio of the number of nouns over verbs. The three *Lexical Richness* features (i.e., *Type-Token Ratio*, *Brunet’s*, and *Honoré’s Indexes*) are quite standard in text analysis literature, while the *Propositional Idea Density* deserves a further description. As defined by Snowdon et al. (1996) and Roark et al. (2011), it is the number of expressed propositions (i.e., distinct facts or notions contained in a text) divided by the number of words. It is a measure of the extent to which the speaker is making assertions (or asking questions) rather than just referring to entities. Propositions correspond to verbs, adjectives, adverbs, prepositions, and conjunctions. Conversely, nouns are not considered propositions, since the main verb and its arguments count as one proposition.

Syntactic Features (SYN)

This group of features includes DLBs that quantify the complexity of the syntactic structures produced by STANZA. Most of them bear self-descriptive names. However, two features deserve some further explanation. Given the memory overhead due to long-distance dependencies, the *Global Dependency Distance* quantifies the difficulty in syntactic processing. Instead, *Syntactic complexity*, as defined by Szmrrecsanyi (2004), is established by counting the linguistic

⁵<http://catalogue.elra.info/en-us/repository/browse/ELRA-S0039/>

⁶<https://speechmrk.com>

⁷<https://universaldependencies.org/>

⁸<http://sp-tk.sourceforge.net/>

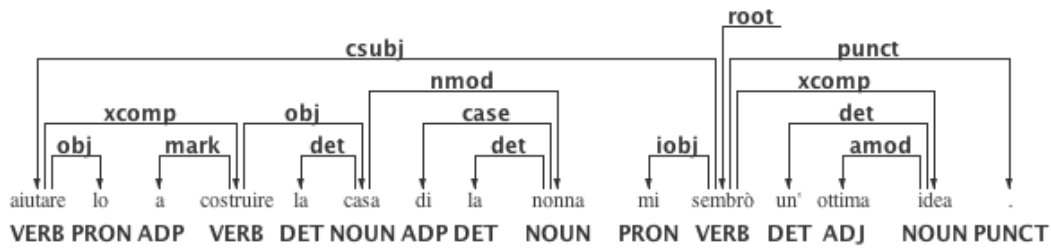


Figure 2: Dependency parse tree for an Italian sentence. The graph shows the management of enclitic pronouns and articulated prepositions (it.: “aiutarlo a costruire la casa della nonna mi sembrò una buona idea”, en. ‘to help him build the grandmother’s house seemed to be a good idea to me’).

tokens that can telltale increased grammatical subordinateness and embeddedness (i.e., subordinating conjunctions, WH-pronouns, finite and non-finite verb forms, 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.

Readability Features (REA)

This set includes four readability features as assessed by the READ-IT assessment tool (Dell’Orletta et al., 2011) at the lexical, morpho-syntactic, syntactic, and global levels. For computing these features, we directly rely on the original tool implemented as API calls to the Dylan server at the ILC-CNR in Pisa⁹. It is relevant to note that, concerning Italian, READ-IT is considered the most reliable reference tool to evaluate readability indexes both on single sentences and whole texts.

Features derived from “Linguistic Inquiry and Word Count” - LIWC Analysis (LWC)

This kind of analysis estimates the incidence of words that fall into one or more semantic categories reflecting emotions, thinking styles, social concerns, and affective processes¹⁰ (Chung and Pennebaker, 2007). The assumption is that simple words of everyday speech is a hint of an underlying psychological state, i.e., “the words we use in daily life reflect what we are paying attention to, what we are thinking about, what we are trying to avoid, how we are feeling, and how we are organizing and analyzing our worlds” (Tausczik and Pennebaker, 2010, p. 30). In our work, we exploited the Italian dictionary by (Agosti and Rellini, 2007) and computed the LWC DLBs accordingly.

3. Successful Case Studies

In the last five years, we successfully applied an early version of the pipeline to study the communicative profiles of various mental disorders.

For example, we extensively investigated the linguistic correlates of Mild Cognitive Impairment and Dementia (American Psychiatric Association, 2013; Petersen,

2004) with diagnostic purposes. In particular, the paper by Beltrami et al. (2018) describes i) the extraction of DLBs from the semi-spontaneous speech of patients and healthy matched controls, and ii) the statistical evaluation of their discriminative power. The study pinpointed several subtle modifications (at the acoustical, lexical, and syntactic levels) that are promising indicators of preclinical stages of cognitive decline. Following on this, the studies by Calzà et al. (2021) and Gagliardi & Tamburini (2021) tested several Machine Learning algorithms (i.e., Support Vector Machine, Random Forrest, and Decision Tree) to automatically distinguish between healthy and MCI subjects, reaching high F1 scores (i.e., around 75%, the state-of-the-art performance for this specific task).

We apply the same methodology to the written text produced by female teenagers with a clinical diagnosis of Anorexia Nervosa (American Psychiatric Association, 2013) and normal-weight peers (Cuteri et al., 2021). We hypothesized that the peculiar psychological features of the disorder (e.g., disturbances in self-perceived body image, inflexible and obsessive thinking, and anxious or depressive traits) result in altered linguistic patterns. Here too, the syntactic level (i.e., sentence length, noun phrase structure, and global syntactic complexity) represents a pivotal domain. We ascribed this peculiar pattern of linguistic erosion to the severe (but reversible) metabolic impairment affecting the central nervous system in Anorexia Nervosa.

Finally, we are currently exploiting DLBs for profiling the communicative skills of preschoolers diagnosed with Developmental Language Disorder (American Psychiatric Association, 2013; Bishop et al., 2017). In this pilot, our priority was to measure the discriminative power of acoustic and rhythmic cues, supporting DLD/controls classification. To this goal, we manually transcribed 1h 57’41” of recorded speech collected from a balanced cohort of sixteen monolingual infants (eight DLD children with expressive deficits and eight peers without language, hearing, or cognitive impairments) to avoid the degradation of ASR results due to young voices. Then we applied the pipelines to derive DLBs from speech. The statistical analysis demonstrates that, even after therapeutic remediation, some spectral characteristics of the voice can distinguish

⁹<http://www.ilc.cnr.it/dylanlab/apps/texttools/>

¹⁰<http://liwc.wpengine.com/>

language-impaired children from peers. In our opinion, these preliminary results are particularly relevant since the significant DLBs are not audible to the human ear, falling outside the possibilities of conventional paper-and-pencil neuropsychological tests (Gagliardi et al., under revision). To date, the actual relation between specific linguistic cues and clinical symptoms of these pathologies is not clear. A larger body of evidence is needed to shed light on this point.

4. Pipeline Issues

The construction of our DLB pipeline posed some tricky questions that deserve an in-depth discussion.

4.1. Storing and Sending Sensitive Data

One of the major issues for the large-scale application of this approach is data scarcity. As a matter of fact, DLBs require large linguistic resources, mainly corpora of atypical language (i.e., balanced collections of verbal production written/uttered by patients and matched healthy controls). However, the gathering, storage, and sharing of these data are particularly difficult due to ethic constraints (de la Fuente Garcia et al., 2020): given their personally-identifying nature, speech recordings and DLBs pertain to “special category of personal data” according to EU law (*Regulation EU 2016/679 - General Data Protection Regulation*)¹¹, and are subjected to strict privacy rules.

Several international initiatives are tackling this challenge (e.g., the DELAD project - *Database Enterprise for Language And speech Disorders* (Nautsch et al., 2019; Lee et al., 2021), which is linked up with the CLARIN’s *Knowledge Centre for Atypical Communication Expertise*)¹². However, currently, the issue is far from being resolved.

For the same reasons, to devise computational infrastructures that safely collect and store biometric and sensible data (i.e., voice samples and personal information) on the web looks very problematic. An NLP pipeline like the one described in this paper should provide this service over the net (e.g., by offering API calls). However, the free exchange of these sorts of data among different organizations is avoided by current regulations¹³.

4.2. ASR Challenging Problems

As argued in the previous sections, working on a good transcription of the subjects’ speech is a foremost issue for extracting reliable DLBs.

Unfortunately, the development of high-performance ASR tools for transcribing spontaneous verbal productions of individuals with communication disorders is very challenging, considering all the intrinsic phenomena of connected speech (e.g., disfluencies such as filled pauses, repetitions, and hesitations due to planings difficulties), which are present in typical verbalizations but often amplified by pathological states. State-of-the-art systems are often developed by big IT companies for their products and are not freely available for research, but delivered through cloud services for a small, tolerable amount of money. However, this access mode presupposes the possibility of sending the speech sample to a remote server, collecting and storing the transcription results. As discussed in Section 4.1, according to the current EU legislation, this is not permitted or very limited.

Developing a custom offline ASR system appears to be the only viable solution, especially for languages different from English. Nevertheless, the lack of large datasets for training these tools and the huge computational power needed to process a considerable amount of data prevent - *de facto* - the actual implementation of such key instruments. Crowdsourcing initiatives like Mozilla Common Voice¹⁴ try to fill the data gap between big IT companies and research groups, but we only are at the beginning of the journey.

Therefore, specific models are needed for an appropriate handling of pathological speech. Considering these aspects, it can be stated that the ‘big IT’ cloud solutions can not provide acceptable performances, leaving this challenging problem almost totally unexplored, at least for less-resourced languages.

5. Conclusions

In this work we presented the development of an NLP pipeline for the automatic extraction of DLBs from speech samples and written texts. The system architecture foresees a set of integrated modules (i.e., ASR, speech segmentation, syntactic parsing, DLBs calculation) that allows the computation of 6 groups of linguistic parameters: acoustic, rhythmic, lexical, syntactic, LIWC-based, and readability indexes. At the moment, the pipeline is exclusively for internal use but we plan to make it freely available in the near future.

Forthcoming works will go towards a further improvement in the algorithm’s robustness and enrichment of linguistic features. In particular, at the lexical level, we plan to revise the Italian LIWC dictionary, adding new lemmas and psychologically-relevant categories (e.g., LIWC2015 classes not present in Agosti and Rellini’s version, like “Core Drives and Needs”). We will also enlarge the list of acoustical features (e.g., Formant Trajectories, Jitter, Shimmer, and Harmonic-to-Noise Ratio - HNR) following the valuable reviews provided by de la Fuente Garcia et al. (2020), Petti et al. (2020), and Voletti et al.(2020).

¹¹<https://eur-lex.europa.eu/eli/reg/2016/679/oj>

¹²<https://ace.ruhosting.nl/>

¹³A possible solution to this issue is the expression of a positive opinion by the competent ethical committee, followed by the conclusion of an agreement between the organizations. However, most of the time, this is not a viable option, due to long lead times.

¹⁴<https://commonvoice.mozilla.org>

Acknowledgements

The authors are grateful to Rema Rossini Favretti, Laura Calzà, Enrico Ghidoni, Daniela Beltrami, Antonia Parmeggiani, Vittoria Cuteri, Giulia Minori, and Milivia Innocenti, who directly collaborate to the studies mentioned in section 3.

6. Bibliographical References

- Agosti, A. and Rellini, A. (2007). *The Italian LIWC dictionary. Technical report*. LIWC.net, Austin (TX).
- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. American Psychiatric Association, Arlington (VA), fifth edition. edition.
- Beltrami, D., Gagliardi, G., Rossini Favretti, R., Ghidoni, E., Tamburini, F., and Calzà, L. (2018). Speech analysis by natural language processing techniques: A possible tool for very early detection of cognitive decline? *Frontiers in Aging Neuroscience*, 10:369.
- Bishop, D., Snowling, M., Thompson, P., Greenhalgh, T., and THE CATALISE-2 CONSORTIUM. (2017). Phase 2 of catalise: a multinational and multidisciplinary delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, 58(10):1068–1080.
- Bucks, R. S., Singh, S., Cuerden, J. M., and Wilcock, G. K. (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., Gagliardi, G., Rossini Favretti, R., and Tamburini, F. (2021). Linguistic features and automatic classifiers for identifying Mild Cognitive Impairment and dementia. *Computer Speech & Language*, 65:101–113.
- Camacho, A. (2007). *SWIPE: A sawtooth waveform inspired pitch estimator for speech and music*. Ph.D. thesis, University of Florida.
- Cantos-Gòmez, P. (2009). Featuring linguistic decline in Alzheimer’s disease: A corpus-based approach. In Michaela Mahlberg, et al., editors, *Proceedings of the Corpus Linguistics Conference 2009 (CL2009)*.
- Catani, M., Jones, D. K., and ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Annals of Neurology*, 57(1):8–16.
- Catani, M., Dell’Acqua, F., Bizzi, A., Forkel, S. J., Williams, S. C., Simmons, A., Murphy, D. G., and Thiebaut de Schotten, M. (2012). Beyond cortical localization in clinico-anatomical correlation. *Cortex*, 48(10):1262–1287.
- Chung, C. and Pennebaker, J. (2007). The psychological functions of function words. In K. Fiedler, editor, *Social Communication*, page 343–359. Psychology Press, New York (NY).
- Cosi, P., Tesser, F., R. Gretter, R., and Avesani, C. (2001). Festival Speaks Italian! In *Proceedings of Eurospeech 2001*, pages 509–512. ISCA, Paris.
- Cuteri, V., Minori, G., Gagliardi, G., Tamburini, F., Malaspina, E., Gualandi, P., Rossi, F., Moscano, M., Francia, V., and Parmeggiani, A. (2021). Linguistic feature of anorexia nervosa: a prospective case–control pilot study. *Eating and Weight Disorders - Studies on Anorexia, Bulimia and Obesity*.
- de la Fuente Garcia, S., Ritchie, C. W., and Luz, S. (2020). Artificial intelligence, speech, and language processing approaches to monitoring alzheimer’s disease: A systematic review. *Journal of Alzheimer’s Disease*, 78:1547–1574.
- De Mauro, T. (2000). *Il dizionario della lingua italiana*. Paravia, Torino.
- Dell’Orletta, F., Montemagni, S., and Venturi, G. (2011). READ–IT: Assessing readability of Italian texts with a view to text simplification. In Norman Alm, editor, *Proceedings of the Second Workshop on Speech and Language Processing for Assistive Technologies*, pages 73–83. ACL, Stroudsburg (PA).
- Delwo, V. (2006). Rhythm and speech rate: a variation coefficient for deltaC. In P. Karnowski et al., editors, *Language and Language-Processing*, pages 231–241. Peter Lang, Frankfurt am Main.
- Gagliardi, G. and Tamburini, F. (2021). Linguistic biomarkers for the detection of Mild Cognitive Impairment. *Lingue e Linguaggio*, XX:3–31.
- Gagliardi, G., Kokkinakis, D., and Duñabeitia, J. A. (2021). Editorial: Digital linguistic biomarkers: Beyond paper and pencil tests. *Frontiers in Psychology*, 12:752238.
- Gagliardi, G., Innocenti, M., and Tamburini, F. (under revision). Can speech-based measures support Developmental Language Disorder identification? An explorative study.
- 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.
- Gagliardi, G. (under revision). Natural language processing techniques for studying language in healthy and pathological aging: A methodological survey.
- Grabe, E. and Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven et al., editors, *Papers in Laboratory Phonology 7*, pages 515–546. Mouton de Gruyter, Berlin.
- Hagoort, P. (2017). The core and beyond in the language-ready brain. *Neuroscience and biobehavioral reviews*, 81(Pt B):194–204.
- Hertrich, I., Dietrich, S., and Ackermann, H. (2020). The margins of the language network in the brain. *Frontiers in Communication*, 5:93.
- Holmes, D. I. and Singh, S. (1996). A stylometric analysis of conversational speech of aphasic patients. *Literary and Linguistic Computing*, 11(3):133–140.
- Howitt, A. W. (2002). Vowel landmark detection.

- The Journal of the Acoustical Society of America*, 112(5):2279–2279.
- Ishikawa, K., MacAuslan, J., and Boyce, S. (2017). Toward clinical application of landmark-based speech analysis: Landmark expression in normal adult speech. *The Journal of the Acoustical Society of America*, 142(5):441–447.
- Lee, A., Bessell, N., van den Heuvel, H., Saalasti, S., Klessa, K., Müller, N., and Ball, M. (2021). The latest development of the delad project for sharing corpora of speech disorders. *Clinical Linguistics & Phonetics*.
- Liu, S. A. (1995). *Landmark detection for distinctive feature-based speech recognition*. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge (MA).
- Locke, S., Bashall, A., Al-Adely, S., Moore, J., Wilson, A., and Kitchen, G. B. (2021). Natural language processing in medicine: A review. *Trends in Anaesthesia and Critical Care*, 38:4–9.
- López-de-Ipiña, K., Alonso, J.-B., Travieso, C. M., Solé-Casals, J., Egiraun, H., Faundez-Zanuy, M., Ezeiza, A., Barroso, N., Ecaz-Torres, M., Martínez-Lage, P., and Martínez de Lizardui, U. (2013). On the Selection of Non-Invasive Methods Based on Speech Analysis Oriented to Automatic Alzheimer Disease Diagnosis. *Sensors*, 13:6730–6745.
- March, E. G., Wales, R., and Pattison, P. (2006). The uses of nouns and deixis in discourse production in Alzheimer’s disease. *Journal of Neurolinguistics*, 19:311–340.
- Nautsch, A., Jasserand, C., Kindt, E., Todisco, M., Trancoso, I., and Evans, N. (2019). The GDPR & speech data: Reflections of legal and technology communities, first steps towards a common understanding. In *Proceedings of Interspeech 2019*, pages 3695–3699, Grenoble. ISCA.
- Petersen, R. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of internal medicine*, 256(3):183–194.
- Petti, U., Baker, S., and Korhonen, A. (2020). A systematic literature review of automatic alzheimer’s disease detection from speech and language. *Journal of the American Medical Informatics Association*, 27(11):1784–1797.
- Qi, P., Zhang, Y., Zhang, Y., Bolton, J., and Manning, C. D. (2020). Stanza: A Python natural language processing toolkit for many human languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. ACL, Stroudsburg (PA).
- Ramus, F., Nespore, M., and Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73:265–292.
- Roark, B., Mitchell, M., and Hollingshead, K. (2007). Syntactic complexity measures for detecting Mild Cognitive Impairment. In Kevin Bretonnel Cohen, et al., editors, *Proceedings of the Workshop BioNLP 2007: Biological, translational, and clinical language processing*, pages 1–8. ACL, Stroudsburg (PA).
- Roark, B., Mitchell, M., Hosom, J.-P., Hollingshead, K., and Kaye, J. A. (2011). Spoken Language Derived Measures for Detecting Mild Cognitive Impairment. *IEEE Transactions on Audio Speech, and Language Processing*, 19(7):2081–2090.
- Satt, A., Sorin, A., Toledo-Ronen, O., Barkan, O., Kompatsiaris, I., Kokonozi, A., and Tsolaki, M. (2013). Evaluation of Speech-Based Protocol for Detection of Early-Stage Dementia. In *Proceedings of Interspeech 2013*, pages 1692–1696. ISCA, Grenoble.
- Singh, S., Bucks, R. S., and Cuerden, J. M. (2001). An evaluation of an objective technique for analysing temporal variables in DAT spontaneous speech. *Aphasiology*, 15(6):571–583.
- Snowdon, D. A., Kemper, S. J., Mortimer, J. A., Greiner, L. H., Wekstein, D. R., and Markesbery, W. R. (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.
- Stevens, K. N., Manuel, S. Y., Shattuck-Hufnagel, S., and Liu, S. (1992). Implementation of a model for lexical access based on features. In *The Second International Conference on Spoken Language Processing, ICSLP 1992*, pages 499–502. ISCA.
- Szmrecsányi, B. M. (2004). On Operationalizing Syntactic Complexity. In G. Purnelle, et al., editors, *Proceedings of the 7th International Conference on Textual Data Statistical Analysis*, pages 1031–1038. Presses Universitaires de Louvain, Louvain-la-Neuve.
- Tamburini, F. and Melandri, M. (2012). AnIta: a powerful morphological analyser for Italian. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC’12)*, pages 941–947. European Language Resources Association (ELRA), Paris.
- Tamburini, F. (2021). Playing with NeMo for building an automatic speech recogniser for Italian. In Elisabetta Fersini, et al., editors, *Proceedings of the Eighth Italian Conference on Computational Linguistics - CLiC-it 2021*. CEUR-WS, 3033.
- Tausczik, Y. and Pennebaker, J. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1):24–54.
- Vigorelli, P. (2004). *La conversazione possibile con il malato Alzheimer*. Franco Angeli, Milano.
- Voleti, R., Liss, J., and Berisha, V. (2020). A review of automated speech and language features for assessment of cognitive and thought disorders. *IEEE Journal of Selected Topics in Signal Processing*, 14(2):282–298.
- Wang, J., Deng, H., Liu, B., Hu, A., Liang, J., Fan, L.,

- Zheng, X., Wang, T., and Lei, J. (2020). Systematic evaluation of research progress on natural language processing in medicine over the past 20 years: Bibliometric study on pubmed. *Journal of Medical Internet Research*, 22(1):e16816.
- Yu, H. and Mak, M. (2011). Comparison of voice activity detectors for interview speech in nist speaker recognition evaluation. In *Proc. of Interspeech'11*, pages 2353–2356.