Exploring the Use of Cohesive Devices in Dementia within an Elderly Italian Semi-spontaneous Speech Corpus

Giorgia Albertin*,†, Elena Martinelli†

Alma Mater Studiorum - University of Bologna, Department of Classical Philology and Italian Studies, 32 Zamboni Street, 40126 Bologna, Italy

Abstract

The study of language disruption in dementia, aimed at individuating which features correlate with cognitive impairment, is a growing area in computational linguistic research. Still, it needs a further development in analyzing some discourse phenomena that also undergo deterioration, and can help expand our understanding of dementia-related speech and refine automatic tools. This paper explores the discourse property of cohesion by investigating three types of cohesive devices: reference, lexical iteration, and connectives. Ten features related to these categories have been defined and automatically extracted from an Italian corpus of semi-spontaneous speech collected from dementia patients and healthy controls. Some of the designed features have proven significant for the binary classification of the two groups and further quantitative analysis highlight interesting differences in the use of cohesive devices, that seem to be associated with cognitive decline.

Keywords

Cohesion, Cohesive devices, Dementia, Cognitive Impairment, Semi-spontaneous speech

1. Introduction

Linguistics deficits commonly characterized neurodegenerative diseases from their onset. In Dementia, or Major Neurocognitive Disorder (DSM-5 [1]), a syndrome of acquired and progressive impairment in cognitive function that interfere with independence in everyday life, language deterioration manifests itself within a broader framework of cognitive impairment, which could affects memory, visuo-spatial skills, executive functions and reasoning. Deficits both in verbal production and comprehension have been observed, despite the specificity of different Dementia's etiological subtypes, among which the most common is Alzheimer's Disease (AD), characterized with a primary impairment in episodic memory. In AD, for example, among the well-established linguistic deficits there are word-finding problems, which include anomia, the production of semantic paraphasias [2, 3] and the "on the-tip-of-the tongue" experience [4], low speech rate, poor word comprehension [5] and, as the disease worsen, a generalized simplification of syntax [6]. Also discourse and pragmatic level is affected by cognitive decline. Errors in referential cohesion has been registered, in particular regarding ambiguous use of pronouns [7]. Coherence is compromised, especially in spontaneous speech: the discourse appears with an abundance of irrelevant details and the overt difficulty to mention the key concept or to refer to the topic, resulting in a lack of informativeness in communication [8, 9, 10].

In recent years, speech analysis in cognitive decline has gained increasing importance in the development of low-cost and portable tools for dementia screening, also supported by the remarkable advancements in Natural Language Processing (NLP) and Machine Learning (ML) technologies [11]. The refinement of classification systems goes hand in hand with the operationalization of linguistic features computed from oral productions, that need to be adapted to different languages. Regarding Italian, the OPLON (OPportunities for active and healthy LONgevity) [2014-2016] project was devoted to the automatic extraction of an extensive group of linguistic features from acoustic, rhythmic, readability, lexical, morpho-syntactic and syntactic levels, from a speech corpus of cognitively impaired patients and healthy peers [12, 13]. Analysis of the significance of the features highlighted that the acoustics ones largely correlated with the cognitive state of the subjects [14].

Expanding the list of language levels covered to include speech properties would enrich the features used for classification and, in addition, could broaden our understanding of how cognitive decline manifests itself in verbal competence. Nevertheless, defining specific features of higher-level and complex phenomena is not trivial. Drawing inspiration from works that propose a "stratified" approach to discourse analysis, which individually considers macro-phenomena that intersect with one another [15, 16], this paper will examine cohesion, the property of the superficial form of the text to reflect its internal unity [17]. Cohesion assures continuity in dis-

CLiC-it 2024: Tenth Italian Conference on Computational Linguistics, Dec 04 — 06, 2024, Pisa, Italy

*Corresponding author.

†The contribution of each author to the paper is specified in the CRediT authorship statement declaration.

giorgia.albertin3@unibo.it (G. Albertin); elena.martinelli12@unibo.it (E. Martinelli)

https://www.unibo.it/sitoweb/giorgia.albertin3 (G. Albertin); https://www.unibo.it/sitoweb/elena.martinelli12/ (E. Martinelli)

0000-0002-5728-3473 (G. Albertin); 0009-0007-4399-6951 (E. Martinelli)

Table 1
Recruitment Criteria (age; language exposure; neurological status or diagnosis; cognitive scores: MMSE, MoCA, phonemic (PF) and semantic (SF) fluency) and Demographics (age and sex).

	Control Group	Pathological Group
Recruitment criteria	Age > 60 years Monolingual Italian L1 Absence of neurological/sensory deficits MMSE ≥ 22 MoCA > 19.262 PF ≥ 17.35 SF ≥ 7.25	Age > 60 years Monolingual Italian L1 Clinical diagnosis of dementia MMSE < 22 MoCA ≤ 19.262 PF < 17.35 SF < 7.25
Age	81 $\pm$ 6.3 (range: 63-91)	81 $\pm$ 6.9 (range: 63-92)
Sex	12F, 8M	12F, 8M

course through a network of *cohesive devices*, which are mainly words or morphemes, that contribute to maintain semantic relations occurring in the text [17]. Therefore, we proposed a method to design and formalize a set of cohesion features, with the aim of observing whether they contribute to discriminate the speech of individuals with dementia from healthy peers. Specifically, three types of elements, which Halliday & Hasan [18] indicate among the major contributors to cohesion, were taken into consideration: reference, lexical iteration and connectives. The implementation of measures based on cohesive devices is the first step towards the attempt to include discourse properties in the automatic analysis of language in cognitive decline. The study of their interaction with features of other linguistic levels is crucial to observe whether they have a positive impact on discrimination between dementia subjects and healthy subjects. The work presented in this paper, therefore, has to be intended as a preliminary analysis that will serve to pursue more sophisticated ML classification in the future.

Figure 1: *Esame del Linguaggio II* [19], stimulus figure used in the picture description task.

2. Corpus Description

In this study, we used the corpus collected within the project "Linguistic characteristics of the speech of elderly subjects with dementia" [20, 21], approved by the Bioethics Committee of the University of Bologna (Prot. N. 0072032/2022). The corpus consists of oral linguistic production of 40 Italian-speaking individuals living in Basilicata, forming two groups balanced by sex and age. Although the initial objective was to balance the cohorts also on education level, it was not possible to consider this aspect due to the lack of this information in some patients medical records. Even from a sociolinguistic perspective, it is important to advance that some participants, albeit Italian-speaking, were also exposed to dialect systems in their lives. This aspect explains the frequent occurrence of substandard linguistic expressions in the collected speech, and will be discussed in Section 4 in relation to the results of the analysis.

The Pathological Group (PG) consists of 20 patients suffering from different forms of dementia (9 cases of Alzheimer's Disease, 2 of Mixed Dementia, 5 of unspecified Dementia, 3 of Vascular Dementia, 1 of Frontotemporal Dementia), recruited at the "Universo Salute - Opera Don Uva (PZ)" rest home, and the Control Group (CG) consists of 20 subjects with neurotypical cognitive aging. Informed consent was obtained from all participants (in the case of patients, by their family members, caregivers, or legal tutors). As a first step, the recruited subjects underwent an evaluation of their cognitive status through the administration of the four following neuropsychological tests: Mini-Mental State Examination (MMSE [22]), Montreal Cognitive Assesment (MoCA [23]), and Verbal Fluency Test, both Phonemic [24, 25, 26, 27] and Semantic [28]. The Table 1 summarizes the recruitment criteria and the demographics for study participants.

Then, two narrative tasks (the story of a journey and the story of the Christmas holiday's traditions) and one picture description task (using the stimulus figure in "Lan-

Table 2
Corpus Size. Audio duration and number of tokens (of the transcriptions) are reported, both with respect to the groups (Gr. durat. and Gr. count), to the single subject (Subj. avg (st.dev)) and to the whole corpus.

	Audio		Tokens	
	Gr. durat.	Subj. avg. (sd)	Gr. count	Subj. avg. (sd)
Pathological group	04:25:26	00:12:00 (00:08:00)	23,518	1,176 (1,218)
Control group	03:23:17	00:10:00 (00:05:00)	25,745	1,287 (710)
Total	07:48:43	-	49,263	-

guage Examination II" [19], see Figure 2) were administered to collect semi-spontaneous speech, elicited with the following stimulus sentences: 1) "Do you want to tell me about a trip you took?"; 2) "How do you usually spend Christmas day?"; 3) "Could you describe this figure to me?". This protocol allowed the collection of approximately 9 hours of audio (i.e., 8 hours for the recruited groups and 1 hour for the interviewer), subsequently annotated at various linguistic levels. By using the ELAN software [29], the corpus was manually transcribed at the orthographic level, segmented into utterances (i.e., the reference unit of discursive analysis [30]), and annotated at the prosodic level (theoretical framework: The Language into Act Theory - L-AcT [31]). Table 2 summarize the size of the corpus and the average material (audio/token) collected for each patient and control subject. The total number of tokens was calculated on the orthographic transcription of the corpus (cleaned of annotation tags), and consists of 49,263 tokens (i.e., 23,518 for PG and 25,745 for CG). Finally, using the Gagliardi & Tamburini pipeline [32], tokenization, lemmatization, part-of-speech tagging, and syntactic parsing was automatically performed for the entire corpus.

3. Cohesive Devices' Features

Ten features that quantify the use of cohesive devices by the speakers were designed and formalised. The features were computed with respect to each subject, thus referring to the amount of speech produced by the single individual in the three tasks. To comprehensively address the categories of cohesive devices considered, we use the `.conll` file resulted from the data annotation as the input for our analysis. Features' automatic extraction was done via `.python` scripts. The methodology used will be described in detail in the following sections.

3.1. Reference

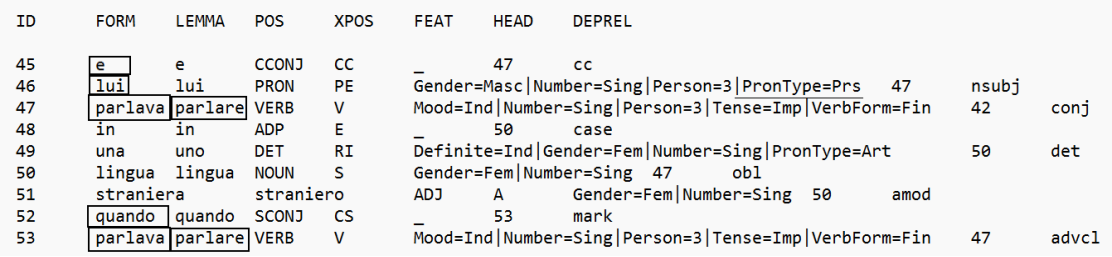
Reference is involved when an expression that requires interpretation by referring to something else occurs in the discourse [18]. This mechanism can be employed both in anaphoric and cataphoric uses, to refer respectively to something already known in the text or anticipating it. Reference functions either by repetition, which can be partial (e.g., through a synonym) or total, by semantic contiguity, or by substitution with pronouns or other elements [17]. It is this second type of referential expressions, closely linked to the textual dimension, that is investigated through the features, thus focusing on the occurrence of anaphora and cataphora.

An extensive literature review was necessary to select a relevant group of those expressions in the Italian language (see [33, 34, 35]). The group of elements collected includes pronouns, both personal (e.g., *io, tu, lei, lui*), demonstrative (e.g., *questo, quello*), indefinite (e.g., *alcuni, tutti*), and possessive, possessive adjectives (e.g., *mio, tuo*), as well as deictics (e.g., *fuori, sopra, avanti, qua, qui, dentro, dietro, giù, indietro, su, lì, avanti, oltre, ci*). The occurrences of these groups were counted and divided by the total number of tokens per subject (COE_REF). Additionally, the *pronoun density* (COE_PRON_DENS), defined as the ratio between pronouns and nouns uttered [36], was computed for each subject.

3.2. Lexical iteration

According to Halliday and Hasan [18], the iteration of a lexical item is a specific use of the repetition-type referential mechanism, which acquires cohesive force on its own because it is typically used when the referent is farther in the text. This set of features focuses on the repetition of three main open-class categories, namely nouns, (main) verbs, and adjectives. The use of words from these classes affects the richness of vocabulary, reflecting the speaker's tendency toward lexical variation. Word-finding problems occurring in cognitive decline often manifest as difficulties in retrieving forms from the lexicon. The repetition of the same words can then occur as a sort of repair mechanism, resulting in semantically impoverished speech. Conversely, the use of some types of closed-class particles, such as prepositions and auxiliaries, is bound to the syntactic structure.

Lexical iteration features were computed by separately considering word forms and lemmas of nouns, verbs, and adjectives. These features include the



ID	FORM	LEMMA	POS	XPOS	FEAT	HEAD	DEPREL		
45	e	e	CCONJ	CC	_	47	cc		
46	lui	lui	PRON	PE	Gender=Masc|Number=Sing|Person=3|PronType=Prs	47	nsubj		
47	parlava	parlare	VERB	V	Mood=Ind|Number=Sing|Person=3|Tense=Imp|VerbForm=Fin	42	conj		
48	in	in	ADP	E	_	50	case		
49	una	uno	DET	RI	Definite=Ind|Gender=Fem|Number=Sing|PronType=Art	50	det		
50	lingua	lingua	NOUN	S	Gender=Fem|Number=Sing	47	obl		
51	straniera	straniero	ADJ	A	Gender=Fem|Number=Sing	50	amod		
52	quando	quando	SCONJ	CS	_	53	mark		
53	parlava	parlare	VERB	V	Mood=Ind|Number=Sing|Person=3|Tense=Imp|VerbForm=Fin	47	advcl		

Figure 2: Example of `.conll` annotation. Occurrences of automatically extracted cohesion devices are reframed: *lui* as a referential expression (note the specification `PronType:Prs` in FEAT column), the repetition of word forms and lemma of a verb (*parlava - parlare*) and the connectives *e* and *quando*.

repetitions of elements divided by the total number of words (COE_RIP_LEM, COE_RIP_WORD), the average number of repetitions for repeated elements (COE_MEDRIP_LEM, COE_MEDRIP_WORD), and the maximum number of repetitions over the total number of iterations (COE_MAXRIP_LEM, COE_MAXRIP_WORD).

3.3. Connectives

As defined by Ferrari [37], connectives are morphologically invariable forms (e.g., conjunctions or locutions) that explicitly indicate logical relations within parts of the text and pertain to the logical level. Elements from different grammatical classes can be used as connectives and are classified based on their function, which usually reflects their meaning (e.g., temporal, causal, additive).

To compile an extensive list of connectives, we rely on the Lexicon of Italian Connectives - *LICO*[1] [38, 39]. LICO contains 173 entries, including single words (e.g., *e, se, ma, infatti, quando, quindi*), complex expressions (e.g., *a causa di, da allora*), and correlatives (e.g., *da un lato ... dall'altro*). Connectives are reported along with their lexical or orthographic variants, part of speech category, the semantic relations conveyed according to the Penn Discourse Tree Bank 3.0 schema [40], examples of usage, and alignments of connectives from other languages. A feature was devoted to compute the occurrences of connectives relative to the total number of tokens per subject (COE_TC).

Finally, the last feature was designed as an attempt to capture the overall impact of the classes of cohesive devices studied in this paper in the two cohorts of corpus speakers. Therefore, the role of cohesion elements was comprehensively measured in COE_TOT by summing referential-substitute expressions, lexical iteration items and connectives, divided by the total number of words.

Figure 3.3 shows as example an excerpt from the annotation in `.conll` format, in which some of the linguistic elements considered were highlighted.

[1] http://connective-lex.info/

Table 3
Results of Kolmogorov-Smirnov test. The cohesive devices' features are reported along with their p-value, significant ones are marked in bold. The p-values of features that resulted significant in Kolmogorov-Smirnov test but not after Bonferroni's correction are given in italic.

Features	p-value
COE_TC	*0.33*
COE_REF	1
COE_REF_DENS	1
COE_RIP_LEM	**0.04**
COE_RIP_WORD	1
COE_MEDRIP_LEM	0.81
COE_MEDRIP_WORD	*0.33*
COE_MAXRIP_LEM	1
COE_MAXRIP_WORD	1
COE_TOT	**0.04**

Table 4
Frequencies of cohesive devices by subject. The average number of occurrences of substitution-type reference items, iterations of lemmas and of word forms (of nouns, adjectives and verbs) and connectives for each subject in PG and CG is reported, along with (st. dev).

Cohesive devices	PG	CG
Reference	146.5 (152.23)	161 (90.93)
Iter. lemma	68.9 (68.00)	87.05 (42.25)
Iter. word form	74.15 (74.38)	87.8 (49.25)
Connectives	23.8 (35.15)	36.65 (26.68)

4. Results

The statistical significance of the cohesion features for the binary discrimination of PG and CG cohorts was calculated using the non-parametric Kolmogorov-Smirnov test, due to the limited sample size of the corpus. Given the number of comparisons performed, we adjusted the results with Bonferroni correction to control for Type I error. This approach involves adjusting the significance

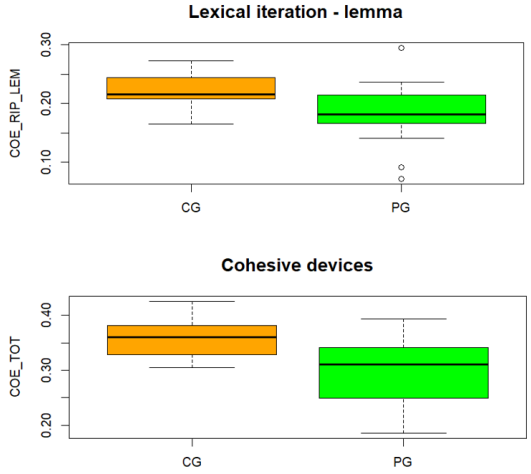


Figure 3: Distribution plots of significantly discriminative features. COE_RIP_LEM indicates the repetitions of lemmas of nouns, adjectives and verbs and COE_TOT is a comprehensive features of all the classes of cohesive devices considered.

level by dividing the conventional alpha value (0.05) by the total number of comparisons made. The results of the test, reported in Table 3, show that two of the designed features significantly contribute to differentiate the two groups: a feature related to lemmas' iteration (COE_RIP_LEM) and the comprehensive feature of cohesive devices (COE_TOT). The distribution of these features is reported in Figure 4.

The application of Bonferroni's correction caused a decrease in the p-value of two initially significant features, namely COE_TC and COE_MAXRIP_WORD. Given the exploratory nature of the experiment, which involves the formalisation of new features in order to discriminate subjects with cognitive impairment from healthy controls in Italian, we have nevertheless chosen to highlight the p-values of these features in 3.

We can observe that, compared with the control group, the speech of dementia subjects is characterized by fewer repetitions of the same noun, verb and adjective lemmas out of the total number of words uttered, captured by COE_RIP_LEM. Thus in the dataset emerges that PG group is less prone to lexical iteration of lemmas than CG. However, if we have a look to the occurrences' distributions of the cohesive elements considered, reported in Table 4, interesting trends could be noticed. Indeed, the quantitative analysis of lexical repetitions revealed a disparity between repeated lemmas and repeated word forms of the same grammatical categories (noun, adjectives and verb) between the two groups. Specifically, despite the high variability due to subjective differences, it is observed that in PG, the average repetition of forms (mean=74.15) is higher than the repetition of lemmas (mean=68.9), while the two values are very similar in CG (lemmas: mean=87.05, words: mean=87.8). This imbalance in favor of forms in the dementia patients appears to uncover lexical impoverishment compared to healthy subjects. Indeed in CG, although a higher overall number of repetitions is registered, it is combined with a more balanced distribution between lemmas and forms, suggest greater lexical variety.

An additional consideration regarding the opposing trend observed between lemmas and forms could be explained with respect to the sociolinguistic profile of the data, related to the diatopic variation of Italian language [41]. Indeed, speakers from both groups show an extensive use of dialectal terms and structures characteristic of the Italian variety spoken in the Lucanian Apennine area. As reported in Section 2, the annotation was conducted automatically using the pipeline developed by Gagliardi & Tamburini [32], which is designed to analyze standard Italian. Therefore, it is likely that the system struggled to handle some substandard expressions, which often orthographically diverge from the other words in the transcription, as can be observed in this example from a PG subject:

> ***gemm' a trua'*** *[=andammo a fare visita] a mia suocera,* ***ca*** *[=che] mio suocero è morto (...).*

It is not excluded that the presence of dialect may also have influenced the automatic extraction of other cohesive devices. Indeed, the higher frequency in CG of substitution-type reference items (mean=161) and connectives (mean=36.65) compared to PG (ref. mean=146.5, conn. mean=23.8) contrasts with what has been observed in oral production of narrative discourse in cohorts of dementia subjects and healthy controls [8]. Therefore, we consider the possibility that automatic feature extraction preceded on manually-checked annotation may yield different results than those obtained.

Nevertheless, the significance of the comprehensive feature (COE_TOT) indicates that the use of cohesive devices investigated in this paper plays a role in distinguishing dementia subjects from healthy controls. In Figure 4 it can be noted that COE_TOT shows, on average, lower values for the PG compared to the CG. This results suggests that the linguistic processing of some phenomena related to cohesion (i.e. substitution-type reference elements, lexical iteration items, and connectives) is generally affected by cognitive decline in semi-spontaneous speech. Thus, the analysis of discourse properties seems to be a promising path for studying the linguistic characterisation of neurodegenerative disorders. Therefore, we hope that our approach in the future could be applied to phenomena strictly related to cohesion - first of all, coherence - or extend to other domains, such as pragmatics, that may mask subtle clues of cognitive frailty.

5. Conclusion

In this work, we present a methodology for delineating linguistic features of cohesion to track and study changes in discourse properties in the speech of individuals with cognitive impairment compared to healthy peers. The research focused on three types of cohesive devices, i.e., reference, lexical iteration, and connectives, that were automatically extracted from a Italian corpus of semi-spontaneous speech from dementia subjects and controls, collected in Basilicata. Statistical significance for binary discrimination was computed applying the Kolmogorov-Smirnov test, and then adjusting the results with Bonferroni's method. The test shows that a feature of the repetitions of lemmas and the one related to the set of cohesive devices jointly considered contribute to distinguish the two groups. Moreover, the quantitative distribution of the cohesive devices reveals differences in the use of elements within the considered categories between PG and CG, which seem to highlight a general deterioration in discursive competencies associated with dementia. The results obtained provide a preliminary basis for further study of discourse properties in cognitive decline, with the aim of expanding the set of linguistic features that can be automatically extracted to other levels of language. This expansion is intended to refine digital systems that could be employed as support for the early diagnosis and monitoring of neurodegenerative diseases, potentially improving timely interventions for patients and their caregivers.

CRediT authorship statement declaration

GA Conceptualization, Methodology, Software (i.e. features formalization), Formal analysis, Writing (§ 1, 3, 4, 5).
EM Resources (i.e. data collection), Data curation (i.e. manual transcription), Writing (§ 2).

References

[1] D. American Psychiatric Association, D. American Psychiatric Association, et al., Diagnostic and statistical manual of mental disorders: DSM-5, volume 5, American psychiatric association Washington, DC, 2013.

[2] E. Catricalà, P. A. Della Rosa, V. Plebani, D. Perani, P. Garrard, S. F. Cappa, Semantic feature degradation and naming performance. evidence from neurodegenerative disorders, Brain and language 147 (2015) 58–65.

[3] V. Taler, N. A. Phillips, Language performance in alzheimer's disease and mild cognitive impairment: a comparative review, Journal of clinical and experimental neuropsychology 30 (2008) 501–556.

[4] E. A. Stamatakis, M. A. Shafto, G. Williams, P. Tam, L. K. Tyler, White matter changes and word finding failures with increasing age, PloS one 6 (2011) e14496.

[5] A. E. Budson, N. W. Kowall, The handbook of Alzheimer's disease and other dementias, John Wiley & Sons, 2011.

[6] S. O. Orimaye, J. S.-M. Wong, K. J. Golden, Learning predictive linguistic features for alzheimer's disease and related dementias using verbal utterances, in: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From linguistic signal to clinical reality, 2014, pp. 78–87.

[7] S. Carlomagno, A. Santoro, A. Menditti, M. Pandolfi, A. Marini, Referential communication in alzheimer's type dementia, Cortex 41 (2005) 520–534.

[8] C. Drummond, G. Coutinho, R. P. Fonseca, N. Assunção, A. Teldeschi, R. de Oliveira-Souza, J. Moll, F. Tovar-Moll, P. Mattos, Deficits in narrative discourse elicited by visual stimuli are already present in patients with mild cognitive impairment, Frontiers in aging neuroscience 7 (2015) 96.

[9] S. Ahmed, A.-M. F. Haigh, C. A. de Jager, P. Garrard, Connected speech as a marker of disease progression in autopsy-proven alzheimer's disease, Brain 136 (2013) 3727–3737.

[10] T. Bschor, K.-P. Kühl, F. M. Reischies, Spontaneous speech of patients with dementia of the alzheimer type and mild cognitive impairment, International psychogeriatrics 13 (2001) 289–298.

[11] S. De la Fuente Garcia, C. W. Ritchie, S. Luz, Artificial intelligence, speech, and language processing approaches to monitoring alzheimer's disease: a systematic review, Journal of Alzheimer's Disease 78 (2020) 1547–1574.

[12] L. Calzà, G. Gagliardi, R. R. Favretti, F. Tamburini, Linguistic features and automatic classifiers for identifying mild cognitive impairment and dementia, Computer Speech & Language 65 (2021) 101113.

[13] D. Beltrami, G. Gagliardi, R. Rossini Favretti, E. Ghidoni, F. Tamburini, L. Calzà, Speech analysis by natural language processing techniques: a possible tool for very early detection of cognitive decline?, Frontiers in aging neuroscience 10 (2018) 369.

[14] G. Gagliardi, F. Tamburini, Linguistic biomarkers for the detection of mild cognitive impairment, Lingue e linguaggio 20 (2021) 3–31.

[15] B. S. Kim, Y. B. Kim, H. Kim, Discourse measures to differentiate between mild cognitive impairment and healthy aging, Frontiers in aging neuroscience 11 (2019) 221.

[16] J. Kim, J. Shim, J. H. Yoon, Subjective rating scale for

discourse: Evidence from the efficacy of subjective rating scale in amnestic mild cognitive impairments, Medicine 98 (2019) e14041.
[17] A. Ferrari, Linguistica del testo, Principi, fenomeni, strutture, Roma, Carocci (2014).
[18] M. A. K. Halliday, R. Hasan, Cohesion in english, Routledge, 2014.
[19] P. Ciurli, P. Marangolo, A. Basso, Esame del Linguaggio II. Manuale e materiale d'esame, Giunti, Firenze, 1996.
[20] E. Martinelli, V. Garrammone, F. Mori, I. Nolè, F. Cameriero, M. Martino, G. Di Bello, G. Gagliardi, DemCorpus-basilicata: Dementia corpus, 2022. URL: http://hdl.handle.net/20.500.11752/OPEN-989, ILC-CNR for CLARIN-IT repository hosted at Institute for Computational Linguistics "A. Zampolli", National Research Council, in Pisa.
[21] E. Martinelli, G. Gagliardi, Compromissioni semantico-lessicali nei pazienti italofoni affetti da demenza: un'analisi corpus-based, ITALIANO LINGUADUE 15 (2023) 711–732. doi:10.54103/2037-3597/21986.
[22] E. Magni, G. Binetti, A. Bianchetti, R. Rozzini, M. Trabucchi, Mini-mental state examination: a normative study in italian elderly population, European Journal of Neurology 3 (1996). URL: https://api.semanticscholar.org/CorpusID:24843663.
[23] S. Conti, S. Bonazzi, M. Laiacona, M. Masina, M. V. Coralli, Montreal cognitive assessment (moca)-italian version: regression based norms and equivalent scores, Neurological Sciences 36 (2015) 209–214. URL: https://api.semanticscholar.org/CorpusID:3026657.
[24] C. Caltagirone, G. Gainotti, G. Carlesimo, L. Parnetti, L. Fadda, R. Gallassi, et al., Batteria per la valutazione del deterioramento mentale (parte I): descrizione di uno strumento di diagnosi neuropsicologica, Archivio di Psicologia, Neurologia e Psichiatria 56 (1995) 461–470.
[25] G. A. Carlesimo, C. Caltagirone, G. Gainotti, et al., Batteria per la valutazione del deterioramento mentale (parte II): standardizzazione e affidabilità diagnostica nell'identificazione di pazienti affetti da sindrome demenziale, Archivio di Psicologia, Neurologia e Psichiatria 56 (1995) 471–488.
[26] G. Carlesimo, C. Caltagirone, L. Fadda, et al., Batteria per la valutazione del deterioramento mentale (parte III): analisi dei profili qualitativi di compromissione cognitiva, Archivio di Psicologia, Neurologia e Psichiatria 56 (1995) 489–502.
[27] G. A. Carlesimo, C. Caltagirone, G. Gainotti, et al., The mental deterioration battery: Normative data, diagnostic reliability and qualitative analyses of cognitive impairment, European Neurology 36 (1996) 378–384.
[28] H. Spinnler, G. Tognoni, Standardizzazione e taratura italiana di test neuropsicologici: gruppo italiano per lo studio neuropsicologico dell'invecchiamento, Masson Italia periodici, Milano, 1987. Supplementum 8 - Italian journal of neurological sciences.
[29] ELAN (version 6.2) [computer software], 2021. URL: https://archive.mpi.nl/tla/elan.
[30] J. L. Austin, How to do things with words, Clarendon Press, Oxford, 1962.
[31] E. Cresti, M. Moneglia, The illocutionary basis of information structure: The language into act theory (l-act), in: E. Adamou, et al. (Eds.), Information Structure in Lesser-described Languages: Studies in prosody and syntax, John Benjamins Publishing Company, Amsterdam, 2018, pp. 360–402.
[32] G. Gagliardi, F. Tamburini, The automatic extraction of linguistic biomarkers as a viable solution for the early diagnosis of mental disorders, in: N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, J. Odijk, S. Piperidis (Eds.), Proceedings of the Thirteenth Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, 2022, pp. 5234–5242. URL: https://aclanthology.org/2022.lrec-1.561.
[33] M. Prandi, C. De Santis, Le regole e le scelte, Introduzione alla grammatica italiana, UTET, Torino (2006).
[34] C. Andorno, Linguistica testuale. Un'introduzione, Carocci, 2003.
[35] A. Ferrari, L. Zampese, Dalla frase al testo: una grammatica per l'italiano, Zanichelli, 2000.
[36] M. M. Louwerse, P. M. McCarthy, D. S. McNamara, A. C. Graesser, Variation in language and cohesion across written and spoken registers, in: Proceedings of the Annual Meeting of the Cognitive Science Society, volume 26, 2004.
[37] A. Ferrari, Connettivi, Enciclopedia dell'italiano (2010).
[38] A. Feltracco, E. Ježek, B. Magnini, Enriching a lexicon of discourse connectives with corpus-based data, in: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), 2018.
[39] A. Feltracco, E. Jezek, B. Magnini, M. Stede, Lico: A lexicon of italian connectives, CLiC it (2016) 141.
[40] B. Webber, R. Prasad, A. Lee, A. Joshi, A discourse-annotated corpus of conjoined vps, in: Proceedings of the 10th Linguistic Annotation Workshop held in conjunction with ACL 2016 (LAW-X 2016), 2016, pp. 22–31.
[41] G. Berruto, Sociolinguistica dell'italiano contemporaneo, Roma: Carocci (2021).