

COMPROMISSIONI SEMANTICO-LESSICALI NEI PAZIENTI ITALOFONI AFFETTI DA DEMENZA: UN'ANALISI CORPUS-BASED

Elena Martinelli, Gloria Gagliardi¹

1. I CORRELATI LINGUISTICI DEL DETERIORAMENTO COGNITIVO: STATO DELL'ARTE

Gli studi di linguistica clinica hanno da tempo evidenziato i vantaggi della profilazione linguistica delle patologie: fin dai primi lavori al riguardo (Crystal, 1980), l'identificazione di caratteristiche verbali significativamente divergenti tra una data coorte di pazienti e la popolazione normotipica, infatti, si è rivelata utile non solo per la progettazione di strumenti in supporto alla pratica clinica, ma anche per accrescere le conoscenze scientifiche in merito al funzionamento normotipo della facoltà di linguaggio (Gagliardi, 2019).

L'ausilio che tale tecnica può garantire alla prassi clinica è ormai consolidato soprattutto nel caso della demenza, una sindrome clinica che si manifesta in seguito all'insorgenza di molteplici patologie (es. malattie neurodegenerative, patologie cardiovascolari e/o metaboliche, ecc.) e che si caratterizza per il progressivo e multiplo deterioramento delle funzioni cognitive, da cui tipicamente scaturisce la perdita della personale autonomia funzionale (APA, 2013).

Le stime epidemiologiche circa l'incidenza e mortalità della patologia nella popolazione anziana sono allarmanti: nel 2020 sono stati identificati più di 50 milioni di casi, se ne stimano circa 82 milioni nel 2030 e 152 milioni nel 2050 (almeno 9,9 milioni di nuovi casi diagnosticati ogni anno, ovvero 1 ogni 3,2 secondi; cfr. ADI *et al.*, 2015; 2018; 2021). In unione all'ingente dispendio di risorse sanitarie, socioeconomiche e assistenziali richiesto, tali previsioni spiegano il motivo per il quale la demenza sia stata definita una priorità mondiale di salute pubblica dall'Organizzazione Mondiale della Sanità (WHO, 2012).

Condurre *screening* di massa della popolazione è dunque urgente: allo stato attuale, tuttavia, i biomarcatori più affidabili – ovvero quelli di *neuroimaging* – sono piuttosto costosi e invasivi, per cui il loro impiego su vasta scala è particolarmente difficoltoso. Di conseguenza, da diversi anni la comunità scientifica ricerca marcatori di malattia di più agile impiego, come biomarcatori liquorali, indici di rischio cardiovascolare e metabolico, caratteristiche demografiche ma anche linguistiche (cfr. ADI, 2021).

Le “spie” presenti nella produzione verbale spontanea dei pazienti, infatti, rappresentano una marca dell'azione di processi patologici per altri aspetti silenti: numerosi studi a livello internazionale (cfr. Boschi *et al.*, 2017) hanno da tempo evidenziato la degradazione delle competenze linguistiche relata alla presenza di una forma di demenza e – d'altro canto – studi di natura longitudinale-retrospettiva hanno messo in luce quanto tale compromissione possa anche fungere da indice prodromico di disfunzioni cognitive (Snowdon, 2003).

Affinché il *profiling* linguistico di una condizione patologica sia corretto, tuttavia, è necessario basare tali studi su *corpora* bilanciati e rappresentativi della popolazione clinica indagata. È fondamentale, inoltre, che le analisi siano condotte su verbalizzazioni

¹ Alma Mater Studiorum - Università di Bologna.

autentiche dei parlanti, senza aprioristicamente basarsi sui risultati ottenuti indagando sistemi linguistici diversi: nonostante aspetti certamente comuni, il profilo comunicativo di una malattia può essere caratterizzato da tratti specificamente relati alla lingua impiegata dai soggetti.

Da questo punto vista, le risorse per l'italiano sono molto limitate poiché, escludendo i singoli *case-studies*, le ricerche condotte al riguardo sono poche e i materiali raccolti non sono confluiti in *corpora* pubblicati e/o accessibili (a differenza di quanto si osserva, ad esempio, per l'inglese). In aggiunta, la notevole variazione diatopica della lingua italiana implica che si raccolgano campioni linguistici appartenenti a soggetti provenienti da zone differenti della Penisola, così da garantire una più robusta generalizzabilità dei risultati ottenuti. Considerando i progetti più rilevanti finora condotti in questo campo, l'eloquio raccolto è riconducibile solamente a tre regioni italiane: Emilia-Romagna (*corpus* OPLON; cfr. Calzà *et al.*, 2021), Campania (*corpus* CIPP-MA; Dovetto, 2021) e Calabria (*corpus* CIPP-mci; Dovetto, 2021). Da menzionare in tale sede, anche se di natura leggermente differente, è il *corpus* Anchise 320 (cfr. Benvenuti *et al.*, 2020), che censisce le produzioni linguistiche spontanee di pazienti affetti da declino cognitivo ricoverati nella RSA italiane (prevalentemente, però, della regione Lombardia).

2. IL PROGETTO DI RICERCA

Al fine di identificare la compromissione semantico-lessicale relata alle condizioni di deterioramento cognitivo per la lingua italiana, il disegno sperimentale ha previsto l'acquisizione di nuovi campioni di linguaggio patologico e normotipico. Nello specifico, allo scopo di arricchire anche in ottica sociolinguistica il materiale già disponibile, abbiamo scelto di coinvolgere soggetti provenienti dalla Basilicata, una regione non precedentemente rappresentata in studi di tale natura.

Il reclutamento è avvenuto previa valutazione neuropsicologica realizzata tramite classici test psicometrici, ovvero MMSE (Magni *et al.*, 1996), MoCA (Conti *et al.*, 2015) e test di fluency verbale, fonemica (Caltagirone *et al.*, 1995; Carlesimo *et al.*, 1995, 1996) e semantica (Spinnler, Tognoni, 1987).

Il campione è composto da 40 individui di origine lucana (24 di sesso femminile, 16 di sesso maschile) distinti in due gruppi, bilanciati per sesso ed età:

- Gruppo dei pazienti (GP): 20 soggetti (12 F, 8 M; età: 81 ± 6.9), con diagnosi conclamata di demenza e in cura presso la Residenza Sanitaria Assistenziale Opera Salute – Don Uva (sede di Potenza, PZ). Nello specifico, il gruppo è più finemente suddivisibile, sulla base della diagnosi ricevuta, in:
 - Malattia di Alzheimer (MA): 9 soggetti (8F, 1 M);
 - Demenza Mista (DM): 2 soggetti (1 F, 1M);
 - Demenza Non Specificata (DNS): 5 soggetti (2 F, 3 M);
 - Demenza Vascolare (DV): 3 soggetti (M);
 - Demenza fronto-temporale (DF-T): 1 soggetto (F);
- Gruppo di controllo (GC): 20 soggetti (12 F e 8 M; età: 81 ± 6.3), con stato cognitivo nella norma.

In Tabella 1 sono riassunti i criteri di reclutamento del campione.

Tabella 1. *Criteri di reclutamento dei partecipanti allo studio*

CRITERI DI RECLUTAMENTO	
GRUPPO P	GRUPPO C
Età > 60 anni	Età > 60 anni
Esposizione linguistica monolingue	Esposizione linguistica monolingue
L1 italiano	L1 italiano
Diagnosi clinica di demenza	Assenza di deficit neurologici / sensoriali
MMSE < 22	MMSE ≥ 22
MoCA ≤ 19,262	MoCA > 19,262
Fluenza fonemica < 17,35	Fluenza fonemica ≥ 17,35
Fluenza semantica < 7,25	Fluenza semantica ≥ 7,25

Dal punto di vista sociolinguistico, è bene precisare che alcuni dei partecipanti reclutati siano stati inevitabilmente esposti anche a sistemi dialettali nel corso della loro vita: tale aspetto, in unione al grado di istruzione raggiunto (spesso non elevato), spiega il frequente riscontro di espressioni linguistiche substandard nell'eloquio raccolto, come sarà esemplificato dagli estratti del corpus (cfr. 5.2)

Il reclutamento dei soggetti e l'acquisizione del materiale linguistico hanno richiesto complessivamente circa quattro mesi (gennaio 2022 - aprile 2022).

La ricerca ha ricevuto l'approvazione del Comitato di Bioetica dell'Alma Mater Studiorum - Università di Bologna (Prot. n. 72032 del 29/01/2022). Per tutti i soggetti sono stati ottenuti i moduli di consenso informato per la partecipazione allo studio e di consenso per il trattamento dei dati personali: per i pazienti, l'autorizzazione è stata concessa da parte dei relativi *caregiver*/tutori legali; i soggetti di controllo, invece, hanno personalmente espresso il loro assenso.

I dati anagrafici, clinici e linguistici sono stati trattati nel rispetto della normativa vigente (Regolamento UE 2016/679 - Regolamento generale sulla protezione dei dati personali; D.Lgs. 196/2003 - Codice in materia di protezione dei dati personali): le informazioni richieste dal progetto e i risultati delle indagini sono stati annotati nella Scheda Raccolta Dati (CRF) e utilizzati in forma pseudo-anonimizzata attraverso l'attribuzione di un codice progressivo univoco ad ogni partecipante.

3. IL CORPUS

3.1. *Acquisizione*

L'eloquio dei partecipanti è stato acquisito mediante la somministrazione di tre task linguistici: due compiti narrativi, per i quali è stato richiesto – rispettivamente – il racconto di un viaggio compiuto nella propria vita e quello delle tradizioni legate alla giornata di Natale; un compito descrittivo, per il quale è stato impiegato come stimolo visivo una figura inclusa nel test neuropsicologico “Esame del Linguaggio II” (Ciurli *et al.*, 1996). Rispettivamente, le frasi stimolo impiegate per i task sono:

1. “Vuoi parlarmi di un viaggio o di una gita che hai fatto?”;
2. “Come passi di solito la giornata di Natale?”;
3. “Potresti descrivermi questo disegno?”.

Le sessioni di registrazione, durate in media circa 30 minuti², sono state svolte per pazienti e gruppo di controllo rispettivamente nella RSA ospitante e all'interno delle proprie abitazioni.

L'eloquio raccolto è pari a circa 8 h e 50 min di sonoro, di cui 7 h e 50 min acquisite per i gruppi reclutati e 1 h attribuibile alle produzioni dell'intervistatrice e di coloro che sono stati coinvolti nelle conversazioni anche se non direttamente intervistati (es. compagni di stanza dei pazienti sottoposti ad intervista nella RSA; familiari presenti durante le interviste dei soggetti di controllo).

3.2. Annotazione

I campioni di linguaggio acquisiti sono stati trascritti a livello ortografico, seguendo le convenzioni dell'Italiano Standard e il formato di trascrizione L-Act (Cresti *et al.*, 2018; *software*: ELAN, 2021).

Contestualmente alla trascrizione ortografica, è stata portata a termine la segmentazione dell'eloquio in enunciati, scelti come unità di riferimento dell'analisi discorsiva. Definibile come “*the linguistic counterpart to the speech act*” (cfr. Austin, 1962), dunque come la minima entità linguistica pragmaticamente interpretabile, ogni enunciato è stato identificato tramite la percezione dei *break prosodici*, ovvero indici prosodici percettivamente salienti poiché correlati con una serie di fenomeni acustici come il *reset* della F0, allungamento finale, presenza di pausa e *rush* iniziale nell'unità prosodica seguente (cfr. Cresti, Moneglia, 2018).

I diacritici del *framework* teorico L-Act impiegati per la segmentazione prosodica del parlato acquisito e per l'annotazione di alcuni fenomeni della lingua parlata sono sinteticamente illustrati in Tabella 2.

Tabella 2. *Diacritici (Framework teorico L-Act)*

Diacritico	Spiegazione
//	break prosodico terminale (conclusione di una sequenza discorsiva)
?	break prosodico terminale (conclusione di una sequenza discorsiva con profilo prosodico interrogativo)
/	break prosodico non terminale (conclusione di un'unità prosodica, ma non di un enunciato)
[/]	break prosodico corrispondente ad una falsa partenza o un <i>retracting</i> con ripetizione, con lo <i>slash</i> eventualmente seguito da una cifra numerica corrispondente al numero di parole ripetute
*	break prosodico coincidente con la terminazione non intenzionale dell'unità discorsiva progettata dal parlante, la cui comprensione risulta dunque potenzialmente compressa
#	esitazione temporanea silente o interruzione del flusso discorsivo (limite minimo: 250 ms)

² L'eloquio è stato acquisito mediante un registratore Zoom H1n (formato .wav, 96 kHz/24-bit).

&	parole incomplete e/o frammenti fonetici
< >	<i>overlapping</i> tra gli interattanti
xxx	parole incomprensibili
hhh	elementi paralinguistici / extralinguistici (es. risata, pianto)
yyy	parole non trascrivibili per privacy

Tutti gli altri livelli linguistici sono stati invece annotati in modo automatico: tramite la *pipeline* computazionale progettata da Gagliardi e Tamburini (2022) sono state portate a termine tokenizzazione, lemmatizzazione, *PoS-tagging* e *parsing* sintattico a dipendenze, annotazioni realizzate sfruttando un *toolkit* di NLP *open source* (STANZA, v. 1.3.0; *Qi et al.*, 2020)³.

3.3. Digital Linguistic Biomarkers

Obiettivo ultimo delle fasi di annotazione descritte è stato l'estrazione automatica dei Biomarker Linguistici Digitali (*Digital Linguistic Biomarkers, DLBs*): *features* linguistiche direttamente derivabili dalla produzione linguistica dei soggetti, la cui analisi consente di descrivere oggettivamente la degradazione delle competenze linguistiche in una data coorte di pazienti (Calzà *et al.*, 2021).

I DLBs estratti sono raggruppabili in cinque tipologie (*features* acustiche, lessicali, semantiche, sintattiche e di leggibilità), a loro volta riconducibili a due gruppi principali a seconda del tipo di input grazie al quale vengono computati:

- *Speech-derived DLBs*: *features* estratte direttamente dai *file* audio, riguardanti aspetti del segnale acustico (quali durata dei segmenti di silenzio/di *speech*, dimensione frattale, centroide spettrale, ecc.).
- *Text-derived DLBs*: *features* estratte dalla trascrizione ortografica e, dunque, dall'input testuale. Riguardano il livello lessicale (LEX), sintattico (SYN), semantico (LWC) e di leggibilità (REA): principalmente descrivono la frequenza di occorrenza delle varie parti del discorso (combinazione di *types*, *token* e *PoS-tag*), la complessità delle costruzioni sintattiche (es. lunghezza media delle relazioni di dipendenza), le categorie semantiche alle quali è possibile ascrivere le parole prodotte dai soggetti (es. processi cognitivi, abitudini quotidiane, socialità), nonché l'intelligibilità (es. lessicale, sintattica) dei testi prodotti.

Ogni tratto linguistico è stato valutato tra le due coorti sia in relazione ai singoli task sia in relazione alla produzione linguistica complessiva (computando, dunque, la media

³ Prima di proseguire con la processazione automatica, gli input sono stati manualmente ripuliti: la frequenza di campionamento dei file è stata ridotta da 96 KHz/24 bit a 44KHz/16 bit, previa eliminazione di tutte le sezioni non riconducibili all'eloquio dei soggetti reclutati; le trascrizioni (in formato .txt) sono state ripulite da tutti i diacritici inseriti in precedenza, fatta eccezione per i *break* prosodici terminali (“//”, sostituiti con il punto “.”) e per le interruzioni improvvise degli enunciati (“+”, sostituiti con i tre puntini sospensivi “...”). Un nome proprio (sempre lo stesso, scelto casualmente), invece, ha sostituito tutte le espressioni inizialmente oscurate per motivi di privacy (= “yyy”), così da continuare a tutelare le informazioni personali dei partecipanti senza compromettere la processazione dei testi orali.

dei valori ottenuti dai partecipanti in ognuno dei compiti somministrati). La significatività⁴ delle divergenze osservate è stata valutata mediante l'applicazione di test di inferenza statistica parametrici e/o non parametrici⁵ (es. t-test di Student, test di Wilcoxon-Mann-Whitney) in R (R Core Team, 2014).

4. ANALISI SEMANTICO-LESSICALE

Sebbene l'estrazione della ricca serie di DLBs abbia consentito di indagare la degradazione di molteplici livelli linguistici, in questa sede proponiamo i risultati dell'analisi semantico-lessicale condotta sui testi orali raccolti: la valutazione delle molteplici *features* estratte, arricchita da riflessioni di tipo qualitativo, consente infatti di caratterizzare con maggiore precisione i deficit pertinenti tale piano linguistico.

Delle *features* semantico-lessicali oggetto di analisi, 34 pertengono al livello lessicale (LEX), quantificando tratti come frequenza di occorrenza delle varie parti del discorso, indici di ricchezza lessicale e frequenza d'uso media delle parole, e 84 riguardano il piano strettamente semantico⁶. Queste ultime sono state computate sulla base del *Linguistic Inquiry and Word Count* (LIWC), un programma ideato per indagare le categorie semantiche (es. processi cognitivi, abitudini quotidiane) verbalizzate da un soggetto durante uno scambio dialogico (cfr. Chung, Pennebaker, 2007). Nello specifico, questa tipologia di analisi è stata automaticamente condotta impiegando la versione italiana del dizionario LIWC (Agosti, Rellini, 2007): esclusa la sottosezione di *features* pertinenti la dimensione linguistica, tutti gli altri DLBs esplorano il riferimento alla sfera emotiva, alla socialità, ai propri stili di vita, e così via.

Le Tabelle 3 e 4 descrivono rispettivamente le *features* lessicali e LWC estratte e analizzate in questo lavoro.

Tabella 3. *Features lessicali (n°: 34)*

Feature	Descrizione
Content Density	Proporzione tra le parole di classe aperta (LEX_OCW) e le parole di classe chiusa (LEX_CCW): (LEX_ContDens; cfr. Roark <i>et al.</i> , 2011).
Part-of-Speech rate	Proporzione di occorrenza di ogni parte del discorso (LEX_PoS_*): nomi, aggettivi, avverbi, verbi, determinanti, articoli, pronomi, congiunzioni coordinanti e subordinanti, numerali, preposizioni semplici e articolate, interiezioni (cfr. Bucks <i>et al.</i> , 2000).
Reference Rate To Reality	Rapporto tra il numero totale dei nomi e il numero totale dei verbi (LEX_RefRRReal; cfr. Vigorelli, 2004).

⁴ La soglia di significatività per tutti i test statistici impiegati è stata fissata a *p-value* < 0.05. Nello specifico: *p-value* < 0.05 (significativo “*”); *p-value* < 0.01 (molto significativo, “**”); *p-value* < 0.001 (altamente significativo, “***”).

⁵ L'applicazione del t-test di Student o del test di Wilcoxon-Mann-Whitney è dipesa dalla tipologia di distribuzione dei dati dei campioni analizzati per ogni DLBs. Nel caso di una distribuzione non dissimile da quella normale o gaussiana, è stato applicato il t-test: nelle tabelle sono indicate media e deviazione standard dei campioni indagati. Quando la distribuzione dei dati è risultata significativamente dissimile da quella normale, è stato applicato il test di Wilcoxon-Mann-Whitney (U-Test): in tal caso, nelle tabelle sono indicati – e segnalati in corsivo – mediana e IQR dei campioni indagati.

⁶ È bene sottolineare che alcuni aspetti (es. uso di persone, tempi e modi verbali, forme ausiliarie, ecc.) vengono in questa sede presentati come parametri lessicali esclusivamente in virtù della modalità con la quale vengono computati, sebbene sia chiaro che siano caratteristiche prettamente grammaticali.

Deissi	Proporzione di espressioni deittiche di natura temporale, spaziale e personale (LEX_TDEIXIS, LEX_SDEIXIS, LEXPDEIXIS; cfr. March <i>et al.</i> , 2006; Cantos-Gòmez, 2009).
Pronomi relativi e avverbi negativi	Proporzione di pronomi relativi (es. chi, cui, ecc; LEX_RPRO) e di avverbi negativi (es. no, neanche, ecc; LEX_NEGADV).
Ricchezza Lessicale	Ricchezza del vocabolario/diversità lessicale, quantificata come: 1) TTR (Types-Tokes Ratio), ovvero il rapporto tra il numero di parole diverse rispetto alla lunghezza complessiva del testo. È, tuttavia, sensibile a quest'ultima: il valore risultante è tanto più alto quanto più il testo è breve (cfr. Holmes and Singh, 1996); 2) Indice di Brunet (LEX_BrunetW), calcolato mediante la formula $W = NV^{(-1.65)}$, dove N=lunghezza totale del testo e V= il vocabolario usato dal parlante. Il range è tipicamente tra 10 e 20: più piccolo è il valore, più il lessico è ricco e diversificato (cfr. Brunet, 1978); 3) Indice di Honoré (LEX_HonoreR), calcolato mediante la formula $R = 100 \log N / (1 - V1/V)$, dove V1 = hapax, V = vocabolario totale, N =lunghezza totale del testo. Indica il numero di espressioni usate solo una volta (<i>hapax legomena</i>) in proporzione al totale delle parole del testo. Più il valore dell'indice è alto, più il lessico del soggetto è ricco (= numero maggiore di <i>hapax</i> impiegati; cfr. Honoré, 1979).
Verbi d'azione	Proporzione dei verbi di azione (LEX_ACTVRB; cfr. Gagliardi, 2014).
Frequenza d'uso delle parole	Peso medio della frequenza d'uso delle parole (LEX_DM_F; cfr. De Mauro, 2000).
Propositional Idea Density	Rapporto tra il numero di proposizioni e il numero di parole (LEX_IDEAD): quantifica la realizzazione di asserzioni o domande (piuttosto che di meri riferimenti ad entità). Marca di "proposizione" è la presenza di verbi / aggettivi / avverbi / preposizioni / congiunzioni, ma non di nomi, considerati parte di proposizioni realizzate mediante l'intera struttura argomentale di un verbo (cfr. Snowdon <i>et al.</i> , 1996; Roark <i>et al.</i> , 2011).
Lunghezza media dell'enunciato	Numero medio di parole per enunciato (LEX_NW).

Tabella 4. *Features semantiche (n°: 84)*

Feature Set	Descrizione
Dimensione linguistica	Occorrenza testuale di molteplici categorie morfosintattiche, tra cui pronomi, articoli, preposizioni, tempi e modi verbali, forme ausiliarie e persone verbali (es. LWC_VOI_Verb).
Emotività	Riferimento ad emozioni/stati d'animo, sia a livello generale (es. LWC_Emo_Neg=emozioni negative) che più specifico (es. LWC_Tristezza).
Cognizione	Riferimento a pensieri personali (es. LWC_Intros) e processi cognitivi (es. percezione: LWC_Vista).

Socialità	Riferimento alla comunicazione interpersonale e, più in generale, ai comportamenti sociali (es. LWC_amici).
Fisicità	Riferimento alla dimensione corporea e a processi ad essa connessi (es. LWC_Corpo, LWC_Mangiare).
Stili di vita	Riferimento alle attività/abitudini quotidiane, comprese quelle del tempo libero (es. LWC_Lavoro, LWC_Tv_it, LWC_Svago).
Spazio/Tempo	Riferimenti di natura temporale / spaziale (es. LWC_Tempo, LWC_Passato).
Conversazione	Presenza di disfluenze (LWC_Nonfluen), parolacce (LWC_parolac) e riempitivi (LWC_riempiti).

5. LA COMPROMISSIONE LESSICALE: RISULTATI

5.1. *Dati quantitativi*

Sulla base delle analisi condotte, la produzione linguistica dei soggetti affetti da demenza risulta essere caratterizzata da 25 DLBs morfosintattico-lessicali significativamente divergenti rispetto alla produzione normotipica (di cui 15 evidenziati tramite l'analisi lessicale e 10 tramite l'analisi LIWC).

Tabella 5: *Features lessicali (analisi – produzione complessiva)*

FEATURE	GRUPPO C		GRUPPO P		T-test / U-Test	
	Media Mediana	SD IQR	Media Mediana	SD IQR	p-value	Signific.
LEX_PoS_DET	0.139	0.0230	0.087	0.031	6,80E-04	***
LEX_ContDens	0.616	0.054	0.471	0.085	3,11E-04	***
LEX_PoS_PUNCT	<i>0.064</i>	<i>0.209</i>	<i>0.209</i>	<i>0.064</i>	2,64E-04	***
LEX_PoS_ADJ	0.036	0.009	0.021	0.008	5,92E-02	***
LEX_PoS_ADP	0.097	0.021	0.064	0.025	5,92E-02	***
LEX_PoS__	<i>0.021</i>	<i>0.011</i>	<i>0.005</i>	<i>0.008</i>	1.1e-05	***
LEX_RPRO	<i>0.019</i>	<i>0.006</i>	<i>0.008</i>	<i>0.005</i>	1,14E-02	***
LEX_OCW	<i>0.379</i>	<i>0.021</i>	<i>0.320</i>	<i>0.045</i>	0.0002117	***
LEX_CCW	<i>0.621</i>	<i>0.021</i>	<i>0.681</i>	<i>0.045</i>	0.0002117	***
LEX_PoS_INTJ	0.029	0.017	0.053	0.025	0.00129	**
LEX_TDEIXIS	<i>0.016</i>	<i>0.009</i>	<i>0.011</i>	<i>0.008</i>	0.001291	**
LEX_PoS_AUX	0.059	0.013	0.044	0.015	0.002091	**
LEX_PoS_X	<i>0.003</i>	<i>0.016</i>	<i>0.037</i>	<i>0.099</i>	0.002227	**
LEX_ACTVRB	0.024	0.008	0.016	0.010	0.007714	**
LEX_PoS_ADV	<i>0.083</i>	<i>0.036</i>	<i>0.066</i>	<i>0.022</i>	0.02447	*

FEATURE	GRUPPO C		GRUPPO P		T-test / U-Test	
	Media <i>Meediana</i>	SD <i>IQR</i>	Media <i>Mediana</i>	SD <i>IQR</i>	<i>p-value</i>	Signific.
LWC_Articol	0.163	0.046	0.106	0.071	3,36E-02	***
LWC_Passato	0.024	0.014	0.010	0.011	0.0001192	***
LWC_Prepos	0.094	0.024	0.059	0.028	0.0001554	***
LWC_Tempo	0.027	0.009	0.015	0.010	0.0001854	***
LWC_Causa	0.008	0.004	0.003	0.005	0.0004983	***
LWC_NoI_Verb	0.006	0.008	0.001	0.004	0.002149	**
LWC_P_pass	0.012	0.011	0.008	0.008	0.009773	**
LWC_Tu_Verbo	0.002	0.002	0.005	0.005	0.01069	*
LWC_NoI	0.004	0.005	0.002	0.002	0.02474	*
LWC_Condizio	0.001	0.003	0	0.001	0.02594	*

L'osservazione dei DLBs risultati significativamente divergenti solo in alcuni task, ma non nella produzione linguistica complessiva, permette di evidenziare ulteriori *features* caratterizzanti l'eloquio patologico, presentate nella tabella 6 e meritevoli di approfondimenti futuri.

Tabella 6: *Features lessicali (analisi – singoli task)*

FEATURE	TASK	GRUPPO C		GRUPPO P		T-test / U-Test	
		Media <i>Mediana</i>	SD <i>IQR</i>	Media <i>Mediana</i>	SD <i>IQR</i>	<i>p-value</i>	Signific.
LEX_PoS_PROPN	1	0.035	0.017	0.023	0.014	0.01812	*
	3	0.000	0	0.002	0.011	0.02961	*
LEX_DM_F	1	2.499	0.660	1.990	0.843	0.04056	*
	2	2.376	0.627	1.997	0.802	0.006715	**
LEX_NW	2	496.0	342.50	250.5	256.25	0.02447	*
LEX_PoS_SCONJ	3	0.015	0.013	0.008	0.017	0.03456	*
LEX_PoS_NOUN	3	0.209	0.053	0.176	0.061	0.03851	*
LWC_Essere	1	0.030	0.017	0.019	0.013	0.03966	*
LWC_Present	3	0.126	0.029	0.105	0.033	0.04491	*
LWC_Transiti	2	0.006	0.007	0.004	0.005	0.04681	*

Sintetizzando le evidenze riscontrate, la produzione linguistica dei soggetti affetti da demenza risulta essere complessivamente caratterizzata da una minore percentuale di

parole di classe aperta (nomi, verbi, aggettivi e avverbi; cfr. LEX_OCW), riconducibile – nello specifico – al ridotto impiego di aggettivi e avverbi (LEX_ADJ; LEX_ADV).

La speculare maggior percentuale di parole di classe chiusa è giustificata dal maggiore ricorso alle interiezioni (LEX_PoS_INTJ), poiché si evidenzia una minore occorrenza di preposizioni semplici (LEX_Prep; LWC_Prepos) e articolate (LEX_PoS__), di determinanti (LEX_Det; LWC_Articol), ma anche di avverbi interrogativi/congiunzioni subordinanti come “perché” (LWC_Causa).

Conseguenza naturale dei dati appena esposti è la registrazione di una minore densità di contenuto (LEX_ContDens) per l'eloquio dei pazienti.

Limitato è anche il ricorso alle espressioni deittiche di natura temporale (LEX_TDEIXIS), confermato da una generale riduzione delle espressioni temporali, soprattutto riferite/formulate al passato (LWC_Tempo, LWC_Passato).

La più alta percentuale di punteggiatura (LEX_PoS_PUNCT) che si osserva è invece interpretabile come maggiore numero di enunciati (significativamente più brevi rispetto a quelli prodotti dai soggetti di controllo) e più frequenti interruzioni improvvise delle pianificazioni discorsive. Questa *feature*, infatti, non è da intendersi come marca dell'uso dei segni di interpunzione (si tratta di testi orali), ma è da ricondursi alle modifiche testuali avvenute negli step di preprocessazione, durante i quali i break prosodici terminali (“//”) e le interruzioni improvvise degli enunciati (“+”) sono stati rispettivamente sostituiti dai segni di punteggiatura “.” e “...”.

Sebbene nel complesso non emergano differenze significative rispetto all'uso dei verbi, l'eloquio dementigeno risulta essere comunque caratterizzato da una percentuale minore di verbi d'azione (LEX_ActVrb) e di verbi coniugati alla prima persona plurale (LWC_Noì_Verb), al netto di una maggiore percentuale di verbi coniugati alla seconda persona singolare (LWC_Tu_Verb). Minori risultano anche gli usi del modo condizionale (LWC_Condizio), di forme ausiliarie (LEX_PoS_AUX) e del participio passato (LWC_P_Pass).

Lo stesso dicasi per l'uso dei pronomi: complessivamente non emergono differenze, anche se si osserva un minore utilizzo di forme pronominali relative (LEX_RPRO) e di forme pronominali di prima persona plurale (LWC_Noì).

Considerando le *features* risultate significativamente differenti solo in alcuni task, se ne deduce che il profilo linguistico della demenza potrebbe caratterizzarsi anche per un minore uso di congiunzioni subordinanti (LEX_PoS_SCONJ), un minore peso medio della frequenza d'uso delle parole (LEX_DM_F), una minore lunghezza media dell'enunciato (LEX_NW), un minore impiego del verbo essere (LWC_Essere), dei verbi transitivi (LWC_Trans) e, infine, dei nomi (LEX_PoS_NOUN).

Peculiare è invece l'uso dei nomi propri, da ridefinire in futuro poiché i dati emersi sono contrastanti (LEX_PROPN; nel task 1 è significativamente inferiore, nel task 3 è significativamente maggiore).

5.2. *Discussione*

I risultati quantitativi ottenuti e finora esposti sono coerenti con quanto riportato in letteratura circa la caratteristica povertà di densità contenutistica dell'eloquio dementigeno. In prima battuta, ciò è spiegabile con il minore impiego di aggettivi e avverbi, prevalentemente responsabili di aggiunta di informazione circostanziale al nucleo del messaggio che il soggetto intende comunicare, principalmente veicolato da nomi e verbi. Per queste ultime due classi, infatti, non si registrano differenze statisticamente significative con le produzioni linguistiche dei soggetti di controllo, un dato che potrebbe anche essere interpretato come il mantenimento di una minima competenza discorsiva,

rappresentata dalla capacità di compiere almeno atti referenziali (nomi) e predicativi (verbi).

Per le stesse ragioni sopra menzionate, la semplificazione lessicale si esplica nella riduzione delle occorrenze di quasi tutte le parole appartenenti a parti del discorso chiuse (es. preposizioni, determinanti, ausiliari), poiché principalmente portatrici di significato funzionale/grammaticale piuttosto che contenutistico/lessicale.

L'uso di pronomi e congiunzioni, in questo studio, non emerge come statisticamente differente da quello dei soggetti cognitivamente integri: uno sguardo alle effettive produzioni linguistiche permette però di sottolineare importanti differenze, relative perlopiù alla funzione con la quale vengono effettivamente impiegate queste parti del discorso.

Per quanto riguarda i pronomi, la letteratura di riferimento ne segnala un consistente impiego da parte dei soggetti affetti da demenza: la motivazione è plausibilmente da ricercarsi nella maggiore accessibilità e nel minore dispendio articolatorio delle forme pronominali, tratti grazie ai quali i pazienti possono compiere atti referenziali altrimenti di difficile realizzazione a causa della severa anomia. Ciononostante, la pronominalizzazione nell'eloquio dementigeno non sempre riveste la sua funzione canonica: nonostante occorrenze quantitativamente comparabili con i soggetti sani, i pazienti frequentemente impiegano forme pronominali senza introdurre un referente nel contesto linguistico antecedente/successivo e/o senza indicarne uno nel contesto extralinguistico. Errori di coesione referenziale sono ampiamente ravvisabili, con evidenti conseguenze sulla struttura pragmatico-discorsiva degli enunciati (cfr. *infra*): la comprensione dei riferimenti pronominali, anche di natura deittica, risulta spesso impossibile, come si evince dai seguenti esempi, emblematici anche dell'alta genericità del lessico impiegato (“ho fatto questa”, “stavo facendo qualcosa”, “cose buone”):

SOGG_02 – Task 1

I: sei andata in vacanza qualche volta?

P: sì

I: e dove?

P: non è stata mai &una [/1] una [/1] &u [/1] # un bel &laggio // # si mettono sempre con # tutte quelle schifezze / # &po mi vengono a prendere / # e che se no # quello che &li devono fare //

SOGG_03

Task 2a:

I: A Natale cosa cucini tu?

P (22): che adesso / # questo qui // po' qua / agg' fatt n'ata + questa qui & # ho fatto questa / e stavo facendo &quaccosa //

Task 2b:

I: stavamo dicendo / a Natale cosa cucini?

P (1): eh / sto # che con il &camo # di # chi faceva # delle &co che [/3] delle cose buone # di là / allora # questo qui # mi viene # a &vede [ride] // (4): e allora [/1] # allora # perché # lui che # ha # il xxx delle + vedi vedi / # non c'è più / linea / con xxx / viene così / xxx che viene / qua / qua //

Il SOGG_02, una paziente affetta da morbo di Alzheimer, impiega sì forme pronominali, ma per nessuna di esse è stato specificato/indicato un referente che possa rendere pienamente comprensibile il messaggio formulato.

Lo stesso discorso vale per gli enunciati della seconda paziente (SOGG_03), ugualmente affetta da Alzheimer: nell'enunciato 22, i pronomi dimostrativi impiegati sono

di difficile interpretazione, poiché un referente non è stato indicato; negli enunciati 1 e 4, il pronome dimostrativo “questo” e il pronome di terza persona singolare maschile “lui” di poco successivo è molto probabile che abbiano lo stesso referente, che però non è stato esplicitato/indicato né in precedenza né successivamente.

Ulteriori riflessioni riguardano l'impiego dei pronomi personali di prima e seconda persona singolare. In letteratura non è raro trovare contributi sulla tendenza dei soggetti con demenza, soprattutto con morbo di Alzheimer, a usare raramente le seconde persone verbali (cfr. Macrì, 2016), aspetto interpretato come riflesso di un quasi totale disinteresse nei confronti dello scambio conversazionale e delle parole degli interattanti. In questo lavoro, invece, emerge un maggiore impiego delle forme verbali coniugate alla seconda persona singolare rispetto ai soggetti di controllo: piuttosto che un atteggiamento di chiusura nei confronti del mondo esterno, gli enunciati qui riportati esemplificano l'intenzione di coinvolgere l'interlocutore nello scambio dialogico e, dunque, il mantenimento di una minima competenza conversazionale.

SOGG_04 – Task 1

80. La vostra (vita) com'è? # &se [/1] sei signorina / perché non hai l'anello //

SOGG_20 – Task 2

4. e mo # dico [/2] mo ti dico proprio la storia //

6. (...) che l'avrai pure sentita (...) //

9. sai quel palazzo (...) //

SOGG_18 – Task 2

23. tu ti dici la storia tua / io mi dico la mia / # e siamo a pari a pari //

Nonostante ciò, non è da escludersi che in alcuni casi la formulazione di domande o il rivolgersi all'interlocutore possano essere una strategia compensatoria attuata quando il paziente non desidera parlare e/o sa di non essere in grado di esprimersi. Ad esempio, la paziente “01” (affetta da Alzheimer) produce un eloquio già drasticamente ridotto, soprattutto confrontato con altri pazienti affetti dalla stessa malattia ma molto più anziani; durante la somministrazione del task 1 (racconto di un viaggio), coinvolge l'intervistatrice nella conversazione, sottintendendo che sia già a conoscenza di aspetti della sua vita: “...ma tu lo sai” appare un'espressione impiegata per evitare il prosieguo della conversazione.

SOGG_01 – Task 1

21. (...) cu u can' / # ma tu lo sai // [= (...) con il cane / # ma tu lo sai //

Un'ulteriore osservazione, in questa sede, riguarda il riferimento alla propria identità, realizzato mediante pronomi e aggettivi (personali e possessivi) e, naturalmente, tramite forme verbali coniugate alla prima persona: in letteratura è stata segnalata “una bassa percentuale di predicati afferenti al soggetto [...] e all'io” e che tale aspetto possa suggerire “uno sganciamento del parlante dalla propria identità e l'incapacità di riconoscersi come soggetto di quanto predicato” (Macrì, 2016: 385). In realtà, l'opinione al riguardo non è unanime, in virtù del riscontro di tendenze contrastanti, soprattutto se analizzate alla luce delle peculiarità morfosintattiche dell'italiano: essendo una lingua *non pro-drop*, infatti, l'esplicitazione del soggetto non è obbligatoria, salvo motivazioni pragmatiche. L'eloquio

dei pazienti è invece spesso denotato dall'espressione dell'“io” soggetto⁷, oltre ad essere spesso costellato da aggettivi e pronomi possessivi, a testimonianza di un mantenimento piuttosto saldo della consapevolezza circa la propria vita, la propria identità e i propri affetti: piuttosto, è la contestualizzazione degli avvenimenti personali a essere deficitaria, coerentemente con i fenomeni di disorientamento spazio-temporale e misidentificazione tipici della demenza. Gli esempi al riguardo sono tanti, eccone alcuni:

SOGG_04 – Task 2

2. io la giornata di Natale / sai com' m' & pass' [= sai come me la passo] ?
3. prima cosa / è morta mia madre / # è morto mio padre / # e sono sola / la [/1] la [/1] diciamo [/1] la più grande delle cinque persone / # sono io / la più grande dei cinque fratelli //
9. non sono infelice io & sono [/1] # sono normale / ecco / sono normale //
10. mh mi piace leggere / mi piace scrivere pure # qualunque cosa / # ahm / discutere pure con le signore # che vengono & perché io tengo il &de[/1] negozio / l'ho detto prima //
59. infatti io sono venuta per una visita //

SOGG_07 – Task 1

10. ij ancora nun l'agg ritt [=io ancora non l'ho detto] & # a mia mamma eccedere / # per il bambino (...) //
25. ij già l'avia ritt ra [/1] [= io già l'avevo detto da] # da mio padre (...) //
45. ij mo ho [/1] ho [/1] aggij vist'qua (...) [= io ora ho visto qua] //

SOGG_11 – Task 1

27. quel là le ha fatte mio padre / # &quan [/1] prima di morire / e a partire pure xxx xxx //
51. ij [= io] quasi quasi xxx (...) //
52. ij veng [=io vengo] <hhh xxx> //

SOGG_12 – Task 2

4. erm [=eravamo] io mia moglie e i miei figli //
18. gemm a truà [=andammo a fare visita] a mia suocera / # ca mio suocero è morto / (...) mia suocera ancora [/1] # tiene ottantacinque anni / # ancora si cucina da sola //
29. io vado a [/1] # mi piace # a giorni # di andare a funghi //
41. ij so' stato fidanzato da piccolino &co mia moglie //
47. io a venticinque anni mi so' spostato //

La riduzione delle forme pronominali di prima persona plurale potrebbe invece ricondursi a un'altra tendenza riscontrabile nell'eloquio dementigeno, ampiamente riconosciuta in letteratura e confermata anche in questo studio⁸: la maggiore prevalenza di costruzioni verbali al presente (0.082, IQR = 0.036), piuttosto che al passato (0.010, IQR = 0.011) e al futuro (0.0001, IQR = 0.0014).

⁷ La prima persona singolare è infatti significativamente maggiore rispetto alla seconda singolare (p -value < 0.001), alla prima plurale (p -value < 0.001), alla seconda plurale (p -value < 0.001) e alla terza plurale (p -value < 0.001). È invece significativamente inferiore rispetto alla terza persona singolare (p -value < 0.01).

⁸ Passato vs Futuro: (KS) p -value < 0.001; Presente vs Futuro: (KS) p -value < 0.001; Presente vs Passato: (KS) p -value < 0.001.

La sovraestensione del presente potrebbe trovare giustificazione nel disorientamento spazio-temporale che spesso caratterizza il quadro sindromico delle patologie dementigene e in un fenomeno di livellamento atemporale, per il quale tale tempo verbale ingloba tutti gli eventi lontani dal momento in cui avviene lo scambio conversazionale: il paziente è indotto a ridurre il riferimento non solo ad avvenimenti passati, ma anche futuri che, di per sé, presentano già un carattere di incertezza. In aggiunta, la marcatezza che contraddistingue questi due tempi verbali rispetto al presente rende l'impiego di quest'ultimo cognitivamente meno dispendioso durante uno scambio conversazionale. Non a caso, lo stesso *pattern* si riscontra osservando le occorrenze dei modi verbali: l'impiego all'interno di costruzioni sintattiche complesse e la non fattualità degli eventi espressi, unitamente alla marcatezza rispetto a un modo come l'indicativo, potrebbero spiegare l'uso pressoché nullo del condizionale.

Dunque, la maggiore focalizzazione dei soggetti con demenza sul tempo presente potrebbe chiarire il ridotto impiego dei pronomi di prima persona plurale e, conseguentemente, delle forme verbali alla prima persona plurale: è indubbio che, a causa della lungodegenza presso la RSA, i pazienti vivano un oggettivo allontanamento dalle usuali situazioni di condivisione (es. con familiari/amici); queste ultime riguardano per lo più il passato e/o il futuro che, come appena sottolineato, sono riferimenti temporali che tendono a non essere verbalizzati (a conferma di ciò, si riscontra anche una generale riduzione dei riferimenti alla socialità, all'inclusione, agli affetti e all'ambiente domestico; cfr. *infra*).

Le riflessioni esplicitate in merito all'impiego dei pronomi possono estendersi anche a quello delle congiunzioni. L'analisi qualitativa del *corpus* sostanzialmente conferma quanto evidenziato anche da altri studi: l'uso delle congiunzioni nei pazienti con demenza risulta spesso "atipico", poiché paiono nel complesso svolgere la funzione di collegamento discorsivo "blando" (cfr. Hernandez-Dominguez *et al.*, 2018).

Sebbene non si evincano occorrenze numericamente dissimili da quelle osservabili nella produzione linguistica normotipica⁹, è però necessario primariamente operare un distinguo, poiché è più plausibile che i pazienti producano costruzioni paratattiche grammaticalmente ben formate piuttosto che strutture ipotattiche ugualmente ben costruite, in ragione della minore dispendiosità cognitiva delle prime (confermata infatti da un maggiore impiego delle congiunzioni di tipo coordinativo¹⁰).

Ciò detto, le evidenze riscontrate sono principalmente riconducibili a due casistiche. Per alcuni pazienti, si osserva un uso "grammaticalmente" corretto delle congiunzioni: le costruzioni paratattiche/ipotattiche sono sintatticamente ben formate, ma deficitarie da un punto di vista pragmatico discorsivo, in quanto caratterizzate da tangenzialità, prolissità e/o deragliamento ideativo. *Deficit* piuttosto frequenti, infatti, concernono la capacità di rispettare coesione e coerenza (a livello locale e/o globale) nei testi prodotti, pena l'inefficacia comunicativa delle produzioni linguistiche stesse:

SOGG_08 – Task 1

24. e quindi insomma / ij naturalmente come agg vist ca qui &affi [/1] facern
quir cos crai / ij agg ritt sì // [e quindi insomma / io naturalmente come ho
visto che quelli &affi [/1] facevano quelle cose domani / io ho detto di sì]

⁹ Una differenza significativa nell'occorrenza delle congiunzioni subordinanti è però emersa per il task 3 (cfr. tab. 6)

¹⁰ Per il GP (task: ALL): LEX_PoS_SCONJ (0.014, IQR: 0.014) *vs* LEX_PoS_CCONJ (0.036, IQR: 0.015), (KS) *p-value* < 0.001. Si riscontra significatività anche per i soggetti di controllo, a conferma del maggiore carico cognitivo implicato dalla subordinazione sintattica rispetto alla coordinazione: LEX_PoS_SCONJ (0.017, IQR: 0.016) *vs* LEX_PoS_CCONJ (0.031, IQR: 0.014), *p-value* < 0.001.

La subordinata temporale e la completiva oggettiva sono correttamente realizzate da un punto di vista sintattico; tuttavia, sono ravvisabili errori coesivi di natura referenziale (per i pronomi dimostrativi non sono stati menzionati referenti nel cotesto linguistico, né ne sono stati indicati alcuni nel contesto extralinguistico), così come di coerenza (l'enunciato è tangenziale rispetto al *topic* principale, ovvero un viaggio compiuto nella propria vita).

SOGG_10 – Task 3

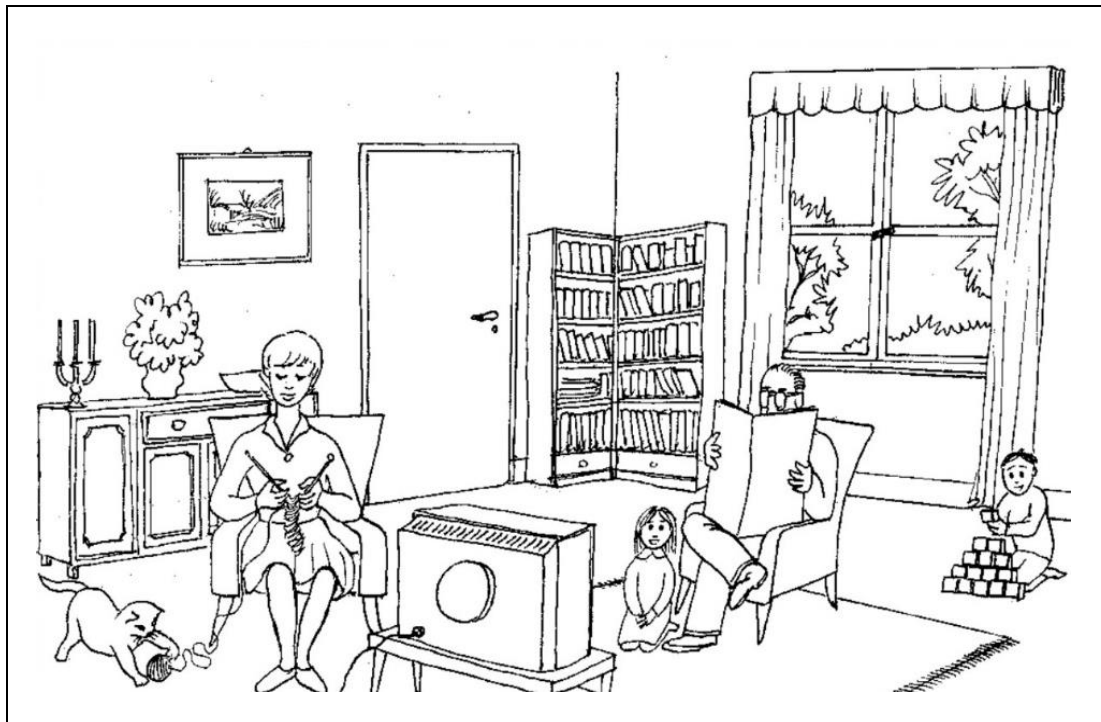
10. questo è un attrezzo # del medioevo / # che veniva usato per stirare [/1] # tirare la luna / # una volta ca [/1] fatto il cardinamento della lana / # a per si [/1] si fanno [/2] delle [/1] delle a forma di palla / # dei rotoli a forma di palla / # e allora xxx più tesa è la [/1] corda e più vengono bene le palle a forma di [/3] a forma di palla //

SOGG_10 – Task 2

14. eh / una volta lì / non c'erano # medici # non c'erano i cosiddetti medici di famiglia / # allora la &buonanim di mio padre che / da questo io ho compreso che poi # era molto affezionato ai figli / # e difatti ne ha avuto nove / due sono morti piccolini / # e all'epoca non c'erano le cure che ci sono oggi / # e allora si moriva facilmente / quando si era &ba [/1] bimbi //

15. a me erano uscite delle [/1] # &de [/1] delle [/1] # delle figure / # una a forma di foglia / l'altra a forma di carrarmato / # e chiese ad un veterinario / dice / ma io sono medico [/1] per gli animali / no / che ti posso dare / dice //

Il SOGG_10, durante la somministrazione del task 3, si pronuncia nel modo sopra riportato durante la descrizione dei ferri per il lavoro a maglia, rappresentati nella figura usata per l'elicitazione linguistica riportata di seguito (cfr. “Esame del Linguaggio II”; Ciurli *et al.*, 1996.)



Per quanto non si ravvisino agrammaticalità sintattiche, l'enunciazione non può dirsi corretta dal punto di vista semantico-lessicale e pragmatico-discorsivo. Sebbene sia stata espressa la funzione dell'oggetto da identificare (i ferri), nel discorso figura un'espressione certamente non pertinente (probabilmente una forma di parafasia verbale = "cardinamento"), così come appare non del tutto appropriata la definizione dell'oggetto come "attrezzo del medioevo". Tali aspetti sono esemplificativi di quanto segnalato in letteratura circa la compromessa coerenza della produzione linguistica dementigena, caratterizzata dall'irrelevanza e dall'implausibilità, nonché dalla ridondanza¹¹ di molti dettagli forniti dai pazienti (cfr. Boschi *et al.*, 2017). Anche gli enunciati prodotti durante il task 2 ben esemplificano quanto appena detto, nonostante la correttezza grammaticale e la complessità delle costruzioni sintattiche realizzate.

In altri casi, invece, l'uso delle congiunzioni è primariamente agrammaticale: anche se ne manifesta la progettazione, infatti, non è certo che il paziente completi la realizzazione della coordinazione/subordinazione. Sono diversi gli esempi di questo tipo, ovvero enunciati nei quali il soggetto enuclea una costruzione ipotattica, salvo poi lasciarla in sospeso (il simbolo "+", si ricordi, è inserito per segnalare l'interruzione improvvisa del discorso):

SOGG_03 – Task 2°

1.sono loro che # stanno # &tr [/1] trovando delle [/1] delle avvolte / nelle cose così / perché +

SOGG_07 – Task 2

4.se devo andare a # tanto per dire / # io vengo da te # prima / # dopo dieci giorni nemmeno +

SOGG_14 – Task 2

22. mh / volendo # reagire / volendo ehm / non si trattava d'altro che # creare +

6. LA COMPROMISSIONE SEMANTICA: RISULTATI

Da un punto di vista prettamente semantico, le analisi condotte consentono di individuare 19 DLBs significativamente divergenti tra la produzione linguistica dei soggetti affetti da demenza e quella normotipica, come evidenziato dalla successiva tabella.

Tabella 7. *Features semantiche (analisi – produzione complessiva)*

FEATURE	GRUPPO C		GRUPPO P		T-test / U-Test	
	Media Mediana	SD IQR	Media Mediana	SD IQR	p-value	Signific.
LWC_Ottimis	0.004	0.003	0.0002	0.001	0.0004921	***
LWC_Sen_Pos	0.021	0.009	0.011	0.007	0.0007168	***
LWC_Affett	0.026	0.011	0.015	0.009	0.001085	**
LWC_Fisico	0.006	0.004	0.002	0.004	0.001609	**

¹¹ L'espressione "palla a forma di palla" ne è un esempio: non solo è di per sé ridondante, ma viene anche ripetuta tre volte, di cui una con sostituzione sinonimica e l'altra con declinazione al plurale.

FEATURE	GRUPPO C		GRUPPO P		T-test / U-Test	
	Media Mediana	SD IQR	Media Mediana	SD IQR	p-value	Signific.
LWC_Sopra	0.002	0.003	0.0001	0.001	0.004391	**
LWC_Corpo	0.003	0.002	0.001	0.002	0.004105	**
LWC_Scuola	0.003	0.002	0	0.002	0.004782	**
LWC_Proc_Sen	0.019	0.007	0.013	0.007	0.007572	**
LWC_Consen	0.001	0.001	0	0.0001	0.01018	*
LWC_Svago	0.011	0.006	0.006	0.005	0.01142	*
LWC_Occupaz	0.007	0.003	0.002	0.007	0.02282	*
LWC_Mangiare	0.002	0.004	0.001	0.002	0.02506	*
LWC_Movimen	0.015	0.007	0.008	0.007	0.02633	*
LWC_riempiti	0.005	0.007	0.002	0.003	0.0263	*
LWC_Sport	0.002	0.004	0	0.002	0.02854	*
LWC_Raggiun	0.002	0.003	0.001	0.002	0.03484	*
LWC_Rif_gen	0.003	0.004	0.002	0.003	0.03596	*
LWC_Casa	0.006	0.005	0.004	0.005	0.03848	*
LWC_Social	0.0378	0.012	0.030	0.012	0.04381	*

Anche in questo caso, focalizzare l'attenzione sui DLBs significativamente divergenti solo in alcuni task, ma non nella produzione linguistica complessiva, evidenzia ulteriori *features* caratterizzanti l'eloquio patologico, presentate nella successiva tabella.

Tabella 8. *Features semantiche (analisi – singoli task)*

FEATURE	TASK	GRUPPO C		GRUPPO P		T-test / U-Test	
		Media Mediana	SD IQR	Media Mediana	SD IQR	p-value	Signific.
LWC_Intros	1	0.016	0.013	0.016	0.012	0.003691	**
	2	0.009	0.013	0.013	0.011	0.000579	***
LWC_Vista	3	0.019	0.019	0.005	0.015	0.004999	**
LWC_Emo_Neg	2	0.005	0.005	0	0.003	0.006195	**
LWC_Famigl	3	0.010	0.021	0	0.004	0.01558	*
LWC_Rabbia	2	0	0.002	0	0	0.01982	*
LWC_possib	3	0.012	0.017	0.005	0.015	0.02506	*
LWC_inclusi	2	0.022	0.012	0.009	0.018	0.02734	*
LWC_Inibiz	2	0	0.003	0	0	0.03903	*
LWC_Mec_Cog	2	0.042	0.015	0.032	0.015	0.04281	*

L'osservazione dei dati ottenuti evidenzia tratti psicopatologici tipicamente associati alla presenza di una sindrome demenziale.

Nello specifico, è evidenziabile un complessivo appiattimento emotivo, testimoniato da minori riferimenti alla sfera affettiva (LWC_Affett), alle sensazioni positive (LWC_Sen_Pos), agli atteggiamenti propositivi (LWC_Ottimis) e consensuali (LWC_Consen).

Ben ravvisabile è la ridotta attività cognitiva, rappresentata – ad esempio – dalla minore verbalizzazione di processi cognitivi di natura sensoriale (LWC_Proc_Sen).

Il classico allontanamento dalla realtà circostante è invece testimoniato da un generale decremento della referenzialità (LWC_Rif_gen), evidente soprattutto per la sfera della socialità, del movimento (LWC_Movimen; WC_Raggiun; LWC_Sopra), così come per buona parte delle tematiche connesse alle attività quotidiane, tra cui l'occupazione, l'istruzione, lo svago, l'ambiente domestico e lo sport. Minore è anche la tendenza a discutere relativamente alla fisicità in generale (LWC_Fisic) e, nello specifico, al corpo (LWC_Corpo), nonché ad attività strettamente connesse con la cura di sé, come il nutrirsi (LWC_Mangiare).

Lo stile conversazionale dei pazienti, infine, si caratterizza per una più bassa occorrenza di riempitivi (LWC_riempiti).

Anche in questo caso, analizzare singolarmente i task somministrati consente di evidenziare ulteriori caratteristiche significative, oscure nella computazione complessiva finale ma meritevoli di futuri approfondimenti.

L'eloquio dementigeno talvolta pare infatti essere caratterizzato da una minore espressione dei processi cognitivi in generale (LWC_Mec_Cog), di processi introspettivi (LWC_Intros), della possibilità (LWC_possib) e di emozioni negative (LWC_Emo_Neg; specialmente della rabbia, cfr. LWC_Rabbia). Ridotto risulta anche il riferimento agli affetti familiari (LWC_Famigl) e alle situazioni inclusive (LWC_Inclusi), mentre è risultata maggiore l'espressione di inibizione (LWC_Inibiz).

7. CONCLUSIONE

L'analisi quantitativa e qualitativa dei testi orali raccolti permette di individuare numerose caratteristiche linguistiche significativamente differenti tra eloquio normotipico e patologico, consentendo di delineare con maggiore precisione la degradazione semantico-lessicale indotta dall'insorgenza di patologie neurodegenerative in soggetti italofofoni.

Dal punto di vista lessicale, la registrazione della ridotta occorrenza di aggettivi e avverbi, di preposizioni semplici e articolate, determinanti e articoli, al netto di una maggiore produzione di interiezioni, concorre a spiegare la povertà di densità contenutistica dell'eloquio dei pazienti: la semplificazione lessicale si esplica nel ridotto impiego delle parole appartenenti a parti del discorso aperte ma responsabili dell'aggiunta di informazione circostanziale al nucleo del messaggio che il soggetto intende veicolare, così come di parole appartenenti alla maggior parte delle classi chiuse, poiché principalmente portatrici di significato funzionale/grammaticale piuttosto che contenutistico/lessicale.

Non si osservano differenze nell'impiego di verbi e pronomi nel complesso, anche se per specifiche sottocategorie si registrano occorrenze minori: ne sono esempio i verbi d'azione, gli ausiliari, i participi passati e i pronomi relativi.

A tal proposito, un aspetto segnalato in letteratura e confermato da questo studio riguarda la maggiore verbalizzazione di eventi al tempo presente, piuttosto che al passato e (ancor meno) al futuro: la “nube di incertezza” comportata dalla malattia in cui vengono inglobati tutti gli eventi lontani dal momento in cui avviene lo scambio conversazionale plausibilmente spiega il ridotto riferimento ad avvenimenti passati e futuri; in aggiunta, la

marcatezza che contraddistingue questi due tempi verbali rispetto al presente rende l'impiego di quest'ultimo cognitivamente meno dispendioso durante uno scambio dialogico. A conferma di ciò, il medesimo *pattern* si riscontra osservando le occorrenze dei modi verbali: l'uso drasticamente ridotto del condizionale plausibilmente può ricondursi al suo impiego all'interno di costruzioni sintattiche complesse, alla non fattualità degli eventi espressi, nonché alla marcatezza da cui è contraddistinto rispetto al modo indicativo.

Ancora, si osserva una maggiore produzione di verbi coniugati alla seconda persona singolare, interpretabile talvolta come il mantenimento di un atteggiamento di apertura nei confronti del mondo esterno e di una minima competenza conversazionale (coinvolgimento dell'interlocutore, formulazione di domande rivolte a quest'ultimo, ecc.), talvolta come strategia compensatoria attuata qualora non si voglia (o ci si renda conto di non essere in grado di) continuare l'interazione. Ridotta è, invece, la produzione di verbi coniugati alla prima persona plurale, riconducibile a un minore riferimento alle usuali situazioni di condivisione e socialità, dalle quali i pazienti sono inevitabilmente più allontanati a causa della lungodegenza nella RSA e del conseguente distacco dagli affetti.

L'analisi qualitativa delle occorrenze pronominali, così come delle congiunzioni (coordinanti e subordinanti) mette in luce *deficit* a livello pragmatico-discorsivo non desumibili esclusivamente dai dati quantitativi computati: nonostante occorrenze non dissimili tra le due coorti, la pronominalizzazione e la coordinazione/subordinazione tra frasi riflettono i tipici errori di coesione referenziale e di coerenza locale/globale frequentemente segnalati in letteratura, emblema della compromissione delle competenze macrolinguistiche nei soggetti affetti da declino cognitivo.

L'analisi semantica degli enunciati prodotti dai pazienti, d'altro canto, evidenzia aspetti psicopatologici classicamente associati alle sindromi demenziali: il generale appiattimento emotivo, testimoniato da un decremento dell'espressione di sensazioni ed emozioni (sia positive che negative), ma anche di atteggiamenti propositivi e consensuali; la ridotta attività cognitiva, evidenziata da una minore esplicitazione di processi cognitivi consapevoli (es. la vista, l'introspezione); la chiusura in sé stessi e il conseguente allontanamento dalla realtà circostante, sottolineati da una minore realizzazione, nel complesso, di atti referenziali (che coinvolge soprattutto lo spazio circostante, le azioni quotidiane, la fisicità, nonché le attività strettamente connesse con la cura di sé, come il nutrirsi).

Ciò viene confermato dall'analisi della leggibilità delle produzioni linguistiche raccolte. L'indice di leggibilità lessicale è risultato – per i pazienti – significativamente minore (C: 78.219, IQR = 26.000; P: 99.690, IQR = 14.659), aspetto imputabile al fatto che i testi prodotti siano maggiormente imperniati su lemmi appartenenti al lessico fondamentale, ad alta frequenza e ad alta disponibilità (maggiore ambiguità lessicale dovuta alla bassa specificità del lessico impiegato), ma anche alla presenza di false partenze, frammenti fonetici e parafrasie di diversa tipologia.

In conclusione, vale la pena notare che tutte le *features* caratterizzanti le condizioni di *Mild Cognitive Impairment* in soggetti italofofoni (cfr. Calzà *et al.*, 2021) risultano altrettanto discriminanti in questo lavoro, ad eccezione proprio di due indici di ricchezza lessicale (ITR e BrunetW). Tale aspetto potrebbe trovare giustificazione in errori di annotazione (e successiva computazione) automatica, o più semplicemente nelle differenti fasce d'età coinvolte nei due studi: la ricchezza lessicale decresce anche in condizioni di normalità cognitiva con l'avanzare dell'età e, dunque, per i soggetti più anziani tali aspetti semantico-lessicali potrebbero non essere più discriminanti.

8. PROSPETTIVE FUTURE

La criticità più evidente dello studio presentato riguarda le modalità di annotazione del *corpus* e di estrazione dei DLBs, poiché – trattandosi di procedure automatiche – la percentuale di errore che può registrarsi è maggiore rispetto a operazioni manuali.

Tuttavia, la questione più complessa è di natura sociolinguistica, poiché i termini dialettali (a livello soprattutto lessicale) sono tutt'altro che rari: non solo non possono escludersi errori di annotazione linguistica automatica dovuti alla mancata individuazione delle espressioni dialettali usate dai soggetti, ma alcuni tratti eventualmente riscontrabili nell'eloquio dei pazienti reclutati potrebbero anche essere solo riconducibili alla varietà di italiano impiegata (italiano popolare o, più in generale, italiano *sub-standard*). In alcuni casi, infatti, gli anziani reclutati sono dialettofoni, con un livello di istruzione non particolarmente elevato, prima ancora di essere soggetti affetti da demenza: tale aspetto mette in luce la necessità di vagliare differenti variabili sociolinguistiche e di valutare l'effettività/significatività di eventuali divergenze anche nei casi in cui i pazienti affetti da demenza godano di un maggiore livello di istruzione e non abbiano vissuto una considerevole esposizione a sistemi linguistici dialettali.

I risultati ottenuti, nonostante tali problematiche, sono in ogni caso promettenti, dal momento che l'annotazione automatica ha comunque evidenziato un notevole numero di *features* significativamente divergenti tra i due gruppi reclutati. Ripetere le indagini condotte dopo aver revisionato manualmente l'annotazione automatica del *corpus* si rivelerà certamente prezioso sia per individuare eventuali divergenze rispetto ai DLBs risultati significativi in questo lavoro (e, dunque, per incrementare / diminuire i correlati linguistici del declino cognitivo per la lingua italiana), sia per stimare la bontà dei risultati ottenibili esclusivamente con una totale automatizzazione delle procedure di annotazione ed estrazione di tratti linguistici.

Integrare nuove analisi quantitative (incrementando numero e tipologia di *Digital Linguistic Biomarker* indagati; cfr. Fuente Garcia *et al.*, 2020; Petti *et al.*, 2020; Vigo *et al.*, 2022) con considerazioni qualitative sarà indubbiamente funzionale alla delineazione di un profilo linguistico-comunicativo del deterioramento cognitivo quanto più completo e preciso.

RIFERIMENTI BIBLIOGRAFICI

- Alzheimer's Disease International, Wimo A., Ali G.C., Guerchet M., Prince M., Prina M., Wu Y.T. (2015), *World Alzheimer Report 2015 – The Global Impact of Dementia: An analysis of prevalence, incidence, cost and trends*, Alzheimer's Disease International, London.
- Alzheimer's Disease International, Patterson C. (2018), *The World Alzheimer Report 2018, 'The state of the art of dementia research: new frontiers*, Alzheimer's Disease International, London.
- Alzheimer's Disease International, Gauthier S., Rosa-Neto P., Morais J. A., Webster C. (2021), *World Alzheimer Report 2021: Journey through the diagnosis of dementia*, Alzheimer's Disease International, London.
- American Psychiatric Association (2013), *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.), American Psychiatric Publishing, Washington (DC).
- Austin J. L. (1962), *How to do things with words*, Clarendon Press, Oxford.

- Becker J. T., Boiler F., Lopez O. L., Saxton J., McGonigle K. L. (1994), "The natural history of Alzheimer's disease: description of study cohort and accuracy of diagnosis", in *Archives of Neurology*, 51, 6, pp. 585-594.
- Benvenuti N. Bolioli A., Mazzei A., Vigorelli P., Bosca A. (2020), "The "Corpus Anchise 320" and the analysis of conversations between healthcare workers and people with dementia", in Dell'Orletta F., Monti J., Tamburini, F. (a cura di.), *Proceedings of the Seventh Italian Conference on Computational Linguistics CLiC-it 2020*, Accademia University Press, Torino.
- Berruto G. (2021), *Sociolinguistica dell'italiano contemporaneo* (Nuova edizione), Carocci Editore, Roma.
- Boschi V., Catricalà E., Consonni M., Chesi C., Moro A., Cappa S. F. (2017), "Connected speech in neurodegenerative language disorders: a review", in *Frontier in Psychology*, 8: <https://www.frontiersin.org/articles/10.3389/fpsyg.2017.00269/full>.
- Caltagirone C., Gainotti G., Carlesimo G., Parnetti L., Fadda L., Gallassi R., Lorusso S., Marfia G., Marra C., Nocentini U. (1995), "Batteria per la valutazione del deterioramento mentale: I. Descrizione di uno strumento di diagnosi neuropsicologica [The Mental Deterioration Battery: I. Description of a neuropsychological diagnostic instrument]", in *Archivio di Psicologia, Neurologia e Psichiatria*, 56, 4, pp. 461-470.
- Calzà L., Gagliardi G., Rossini Favretti R., Tamburini F. (2021), "Linguistic features and automatic classifiers for identifying Mild Cognitive Impairment and dementia", in *Computer Speech & Language*, 65, pp. 101-113.
- Carlesimo G. A., Caltagirone C., Gainotti G., Nocentini U. (1995), "Batteria per la valutazione del deterioramento mentale: standardizzazione ed affidabilità diagnostica nell'identificazione di pazienti affetti da sindrome demenziale", in *Archivio di Psicologia Neurologia e Psichiatria*, 56, pp. 471-488.
- Ciurli P., Marangolo P., Basso A. (1996), *Esame del Linguaggio II. Manuale e materiale d'esame*. Giunti, Firenze.
- Conti S., Bonazzi S., Laiacona M., Masina M., Vanelli Coralli M. (2015), "Montreal Cognitive Assessment (MoCA) - italian version: regression based norms and equivalent scores" in *Neurological Sciences*, 26, pp. 209-214.
- Costa A., Sinforiani E. (2020), *Le demenze. Manuale di diagnosi e trattamento*, NEMS, Milano.
- Cresti E., Moneglia M. (2018), "Chapter 13. The illocutionary basis of Information Structure: The Language into Act Theory (L-Act)", in Adamou E., Haude K., Vanhove M. (eds.), *Information Structure in Lesser-described Languages: Studies in prosody and syntax*, John Benjamins Publishing Company, Amsterdam, pp. 360-402.
- Crystal D. (1981), *Clinical Linguistics*, Springer Verlag, Wien.
- de la Fuente Garcia S., Ritchie C. W., Luz S. (2020), "Artificial Intelligence, Speech, and Language Processing Approaches to Monitoring Alzheimer's Disease: A Systematic Review", in *Journal of Alzheimer's disease: JAD*, 78, 4, pp. 1547-1574.
- De Mauro T. (2000), *Il dizionario della lingua italiana*, Paravia, Torino.
- Dell'Orletta F., Montemagni S., Venturi G. (2011), "READ-IT: Assessing readability of Italian texts with a view to text simplification", in Alm N. (ed.) *Proceedings of the Second Workshop on Speech and Language Processing for Assistive Technologies*, Association for Computational Linguistics, Edinburgh, pp. 73-83.
- Dovetto F. M., Guida A., Pagliaro A. C., Guarasci R., Raggio L, Sorrentino A., Trillocco S. (2021), "Corpora di Italiano Parlato Patologico dell'età adulta e senile", in Cresti E., Moneglia (a cura di), *Corpora e Studi Linguistici*. Atti del LIV Congresso della Società di Linguistica Italiana (Online, 8-10 settembre 2021) Officina, Milano, pp. 165-177:

- https://www.societadilinguisticaitaliana.net/wp-content/uploads/2022/11/010_DovettoEtAlia Atti LIV Congresso SLI.pdf.
- ELAN (Version 6.4) [Computer software] (2022), *Nijmegen: Max Planck Institute for Psycholinguistics*, The Language Archive: <https://archive.mpi.nl/tla/elan>.
- Gagliardi G., Tamburini F. (2021), “Linguistic biomarkers for the detection of Mild Cognitive Impairment”, in *Lingue e linguaggio*, 1, pp. 3-31.
- Gagliardi G. (2019), *Linguistica per le professioni sanitarie*, Pàtron, Bologna.
- Hernández-Domínguez L., Ratté S., Sierra-Martínez G., Roche-Bergua A. (2018), “Computer-based evaluation of Alzheimer’s disease and mild cognitive impairment patients during a picture description task”, in *Alzheimer’s & dementia (Amsterdam, Netherlands)*, 10, pp. 260-268.
- Honore A. (1979), “Some simple measures of richness of vocabulary”, in *Association of Literary and Linguistic Computing Bulletin*, 7, pp. 172-177.
- Macri A. (2016), “La lingua della demenza di Alzheimer. Analisi linguistica del parlato spontaneo”, in Scarpa R. (a cura di), *Le lingue della Malattia. Psicosi, spettro acustico, Alzheimer*, Mimesis Edizioni, Milano.
- Magni E., Binetti G., Bianchetti A., Rozzini R., Trabucchi M. (1996), “Mini-Mental State Examination: a normative study in Italian elderly population”, in *European journal of neurology*, 3, 3, pp. 198-202.
- Pagano C., Bolognini N. (a cura di) (2020), *Neuropsicologia delle demenze*, il Mulino, Bologna.
- Petersen R. C., Smith G. E., Waring S. C., Ivnik R. J., Tangalos E. G., Kokmen E. (1999), “Mild cognitive impairment: clinical characterization and outcome”, in *Archives of neurology*, 56, 3, pp. 303-308.
- Petti U., Baker S., Korhonen A. (2020), “A systematic literature review of automatic Alzheimer’s disease detection from speech and language”, in *Journal of the American Medical Informatics Association: JAMIA*, 27, 11, pp. 1784-1797.
- Qi P., Dozat T., Zhang Y., Manning C. D. (2018), “Universal Dependency Parsing from Scratch”, in *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pp. 160-170.
- R Core Team (2014), *R: A language and environment for statistical computing*, R Foundation for Statistical, Vienna.
- Roark B., Mitchell M., Hollingshead K. (2007), “Syntactic complexity measures for detecting Mild Cognitive Impairment”, in *Biological, translational, and clinical language processing*, Association for Computational Linguistic, Prague, pp. 1-8.
- Roark B., Mitchell M., Hosom J.-P., Hollingshead K., Kaye J. A. (2011), “Spoken Language Derived Measures for Detecting Mild Cognitive Impairment”, in *IEEE Transactions on Audio Speech, and Language Processing*, 19, 7, pp. 2081-2090.
- Satt A., Sorin A., Toledo-Ronen O., Barkan O., Kompatsiaris I., Kokonozi A., Tsolaki M. (2013), “Evaluation of Speech-Based Protocol for Detection of Early-Stage Dementia”, in *Proceedings of Interspeech 2013*, ISCA, Grenoble, pp. 1692-1696.
- Spinnler H., Tognoni G. (1987). *Standardizzazione e taratura italiana di test neuropsicologici: gruppo italiano per lo studio neuropsicologico dell'invecchiamento*, Masson Italia Periodici.
- World Health Organization (2012), *Dementia: a public health priority*, World Health Organization: <https://apps.who.int/iris/handle/10665/75263>.

