FOXP2

Un modello linguistico per lo studio del deterioramento cognitivo

ANTONELLO FABIO CATERINO

ARISTODEMICA EDIZIONI

ARISTODEMICA EDIZIONI

Fuori collana

ISBN 9791281550216

Indice

Introduzione generale

Questo manuale ha per oggetto un modello linguistico, non le sue applicazioni. Chi si accosta al linguaggio spontaneo come indicatore di stato cognitivo trova di solito, in letteratura, il percorso opposto: si parte dal quadro clinico, e gli strumenti linguistici arrivano dopo, in funzione di quello, spesso con criteri di selezione impliciti.[1] Qui il modello, denominato FOXP2, viene prima: la sua architettura, i criteri con cui le componenti sono state incluse ed escluse, la logica che le tiene insieme. Gli ambiti applicativi vengono dopo, come conseguenza, non come giustificazione.

Chi scrive proviene dalla filologia, disciplina che ricostruisce un testo con criteri espliciti prima di chiedersi a cosa serva. Applicata al parlato spontaneo, questa disciplina pretende che le dimensioni linguistiche del modello siano motivate in sé, non ritagliate sulle esigenze di un singolo ambito clinico: un modello costruito per un solo dominio applicativo tende a includere parametri ad hoc, difficili da trasferire altrove. Uno costruito su criteri generali, e poi messo alla prova in più ambiti, regge meglio. Il nome FOXP2 richiama il gene coinvolto nello sviluppo del linguaggio, ma il rapporto è metaforico, non tecnico: il modello non lo implementa, lo evoca. Il Capitolo 2 spiega perché.

Il modello nasce da una domanda di brevetto per invenzione, numero IT202600001294, che definisce cinque indicatori linguistici e una scala di valutazione clinica a dieci punti. Il manuale sviluppa quella base in due direzioni: automatizza il calcolo dei punteggi con formule proprie di ciascuna dimensione, e propone, dove pertinente, estensioni teoriche che il testo depositato non prevede. Le due cose restano distinte capitolo per capitolo.

Stato del brevetto. La domanda IT202600001294 è in esame, non concessa. Ogni riferimento in queste pagine va inteso in questi termini.

Stato della validazione. Il confronto del modello con gli strumenti diagnostici già in uso è un lavoro in corso, non un risultato qui riportato.

Il manuale si articola in due parti. La prima ricostruisce la genesi del modello: perché cinque dimensioni, che rapporto c'è con il gene omonimo, dove si colloca rispetto alla

1. Per una rassegna delle caratteristiche linguistiche impiegate in questo modo nella letteratura sulla malattia di Alzheimer, cfr. K. C. Fraser, J. A. Meltzer, F. Rudzicz, Linguistic features identify Alzheimer's disease in narrative speech, «Journal of Alzheimer's Disease», 49, 2, 2016, pp. 407-422, che classificano 370 variabili linguistiche e acustiche estratte dal medesimo compito di descrizione di immagine.

linguistica clinica computazionale esistente. La seconda riprende ciascuna dimensione, densità lessicale, complessità sintattica, coesione discorsiva, pertinenza e coerenza testuale, fluidità articolatoria, chiude con un caso interamente costruito che le applica insieme a un unico campione fittizio, e si spinge infine oltre il perimetro del brevetto: se la filologia attributiva riconosce un autore dalla stabilità del suo idioletto, la stessa logica, applicata alle campionature ripetute di uno stesso parlante, può rendere osservabile la loro progressiva perdita di coerenza reciproca.

Il nome delle cinque dimensioni adottato qui non coincide sempre con quello degli indicatori del brevetto. La Tabella 1 riporta la corrispondenza.

Tabella 1. Corrispondenza tra le dimensioni del manuale e gli indicatori della domanda di brevetto.

Tabella 1. Corrispondenza tra le dimensioni del manuale e gli indicatori della domanda di brevetto.

Dimensione nel manuale	**Indicatore nella domanda di brevetto**	**Capitolo**
Densità lessicale	Vocabolario	Capitolo 4
Complessità sintattica	Sintassi	Capitolo 5
Coesione discorsiva	Coesione e connessione testuale	Capitolo 6
Pertinenza e coerenza testuale	Pertinenza e coerenza testuale	Capitolo 7
Fluidità articolatoria	Fluidità	Capitolo 8

Un'ultima nota sul metodo di citazione: ogni affermazione empirica è accompagnata da una fonte verificabile; dove la fonte manca, il testo lo dice, e tratta l'affermazione come ipotesi di chi scrive, non come dato acquisito.

Parte I. Genesi e statuto teorico del modello

Capitolo 1 Perché un modello a cinque dimensioni

Criteri di selezione e di esclusione dei parametri linguistici

1.1 Il problema della sovrabbondanza di parametri

Chi costruisce oggi un modello computazionale del linguaggio spontaneo non parte da una scarsità di parametri disponibili, ma dal problema opposto. La letteratura sulla rilevazione automatica del decadimento cognitivo a partire dal parlato spontaneo ha prodotto, nel corso di un decennio, insiemi di variabili linguistiche e acustiche di dimensione crescente: lo studio già citato di Fraser, Meltzer e Rudzicz ne classifica 370, distribuite su otto sottogruppi che vanno dalle parti del discorso alla ricchezza lessicale, dalla complessità grammaticale ai tratti acustici.[2] Un lavoro successivo di ricognizione sistematica su cinquantuno studi dedicati al medesimo compito di descrizione di immagine riporta un insieme di 87 caratteristiche, ripartite in tratti ritmici, acustici, lessicali, morfosintattici e sintattici.[3]

Un insieme così ampio di parametri pone un problema che la sola potenza di calcolo non risolve. Buona parte di queste variabili misura, con formule diverse, uno stesso fenomeno sottostante: la ricchezza del vocabolario impiegato, ad esempio, può essere calcolata attraverso indici distinti quali il rapporto type-token, la misura MTLD, l'indice HDD o l'indice di Honoré, che tendono a covariare fortemente tra loro pur avendo proprietà statistiche differenti.[4] Un modello che includesse l'insieme di queste variabili senza un criterio di selezione rischierebbe di attribuire un peso ridondante a un unico fenomeno linguistico, mascherato da più etichette tecniche, a scapito di altri fenomeni altrettanto rilevanti ma rappresentati da una sola variabile. La costruzione di FOXP2 nasce dall'esigenza di ridurre questo insieme sovrabbondante a un numero di dimensioni

2. Fraser, Meltzer, Rudzicz, Linguistic features identify Alzheimer's disease, cit., p. 410.

3. Il riferimento è alla rassegna citata in E. Zozuk et al., Relationship between language features extracted through NLP and clinically diagnosed Alzheimer's disease and mild cognitive impairment in Slovak, «Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring», 17, 3, 2025, articolo e70122, che riprende la classificazione di E. Eyigoz et al. in 87 variabili linguistiche.

4. Sulla covarianza tra indici di ricchezza lessicale applicati al parlato patologico, cfr. Zozuk et al., Relationship between language features, cit., che riporta la significatività di più indici (GTTR, UBER, SICHEL, MTLD, HDD) nella distinzione tra controlli sani, deterioramento cognitivo lieve e malattia di Alzheimer su un campione di lingua slovacca.

gestibile, sufficiente a rappresentare i principali livelli di organizzazione del linguaggio spontaneo, e al tempo stesso stabile rispetto alla ridondanza interna appena descritta.

La riduzione non è stata condotta per sottrazione arbitraria, ma applicando in sequenza tre criteri distinti, di cui si tratta nei paragrafi che seguono. Il primo riguarda la rilevanza diagnostica del parametro, il secondo la sua indipendenza rispetto agli altri parametri selezionati, il terzo la sua calcolabilità automatica. Un parametro linguistico è stato incluso tra le cinque dimensioni del modello soltanto se ha superato tutti e tre i filtri; è stato escluso se non ha superato anche uno solo di essi, indipendentemente dal grado in cui soddisfaceva gli altri due.

1.2 Primo criterio: rilevanza diagnostica accertata in letteratura indipendente

Il primo criterio richiede che il parametro candidato disponga di un corpo di evidenza pubblicato da gruppi di ricerca indipendenti da chi scrive, relativo alla sua capacità di discriminare tra condizioni linguistiche differenti. Non è richiesto che l'evidenza sia univoca: la letteratura sul linguaggio spontaneo, come si vedrà, contiene risultati non sempre convergenti anche per i parametri meglio studiati. È richiesto che l'evidenza esista, sia stata pubblicata con metodo verificabile, e riguardi un fenomeno linguistico definibile in modo autonomo dal quadro clinico specifico in cui è stata osservata.

La densità lessicale e la ricchezza del vocabolario soddisfano questo criterio in modo ampio. Oltre ai lavori già citati, un contributo che ha analizzato 87 variabili linguistiche in pazienti con malattia di Alzheimer riporta che i risultati di classificazione più significativi sono stati ottenuti impiegando un piccolo sottoinsieme di predittori lessicali, tra cui la lunghezza media delle parole e la frequenza lessicale media.[5]

La complessità sintattica dispone di un corpo di letteratura altrettanto solido, sebbene meno univoco. Diversi studi riportano una riduzione della complessità sintattica nel parlato di soggetti con malattia di Alzheimer rispetto a controlli sani, misurata ad esempio attraverso il punteggio di Yngve sulla struttura ad albero delle frasi.[6] Altri lavori non trovano la stessa associazione nelle fasi più precoci del decadimento cognitivo, il che

5. Zozuk et al., Relationship between language features, cit., che richiama i risultati di Eyigoz et al. su un tasso di classificazione del 76% ottenuto con i dieci predittori linguistici più rilevanti.

segnala un limite reale del parametro piuttosto che la sua irrilevanza.[7] Questo secondo elemento è stato tenuto distinto, nella costruzione del modello, dalla decisione di includere comunque la dimensione: un parametro con evidenza mista resta un candidato valido, purché la sua incertezza sia dichiarata esplicitamente e non presentata come certezza diagnostica.

La coesione discorsiva, intesa come la capacità di mantenere il filo tematico e i legami referenziali all'interno di un tratto di discorso, è documentata in letteratura sia a livello locale, tra frasi contigue, sia a livello globale, tra parti distanti del discorso. Un lavoro dedicato al confronto tra questi due livelli in soggetti con malattia di Alzheimer riporta una riduzione della coerenza semantica globale, in linea con osservazioni precedenti condotte con metodi non automatizzati.[8] Un lavoro sul deterioramento cognitivo lieve giunge a conclusioni analoghe, impiegando misure discorsive distinte da quelle lessicali e sintattiche.[9]

La pertinenza e coerenza testuale, intesa come la capacità di mantenere il discorso aderente al compito richiesto e privo di contraddizioni logiche interne, dispone di un corpo di letteratura autonomo rispetto a quello della coesione discorsiva appena discussa: mentre la coesione riguarda i legami linguistici locali e globali tra le parti del discorso, questa dimensione riguarda la tenuta tematica del discorso rispetto alla consegna e la sua consistenza logica complessiva. Il costrutto di riferimento in letteratura è quello della verbosità fuori tema, introdotto da Arbuckle e Gold per descrivere una tendenza alla tangenzialità del discorso, con inclusione di informazioni eccessive o non

6. Sull'impiego del punteggio di Yngve come misura di complessità sintattica in soggetti con deterioramento cognitivo lieve, cfr. il lavoro di B. Roark e colleghi richiamato in K. Lundholm Fors, Automated Speech and Language Analysis Techniques for the Detection and Assessment of Language and Speech Impairment, letteratura ripresa anche in T. Hakala et al., Analysis of Language for Dementia Diagnosis with Focus on Automatic Speech Recognition, University of Washington, 2017.

7. Cfr. la discussione in K. C. Fraser et al., Language Impairment in Alzheimer's Disease. Robust and Explainable Evidence for AD-Related Deterioration of Spontaneous Speech Through Multilingual Machine Learning, «Frontiers in Aging Neuroscience», 2021, che riporta risultati contrastanti in letteratura sull'associazione tra complessità sintattica e patologia cognitiva nelle fasi iniziali.

8. Cfr. lo studio riportato in Comparing global and local semantic coherence of spontaneous speech in persons with Alzheimer's disease and healthy controls, «Applied Corpus Linguistics», 3, 3, 2023, articolo 100064, che analizza trascrizioni di 81 soggetti con probabile malattia di Alzheimer e 61 controlli sani tratte dal corpus DementiaBank.

9. B. S. Kim, Y. B. Kim, H. Kim, Discourse measures to differentiate between mild cognitive impairment and healthy aging, «Frontiers in Aging Neuroscience»; cfr. anche A. Bertola et al., Exploring the Use of Cohesive Devices in Dementia, in Atti CLiC-it 2024.

pertinenti e perdita del filo tematico.[10] Studi successivi hanno confermato che la tangenzialità del discorso aumenta con l'età ed è associata a un deficit dei processi inibitori, con una replicazione longitudinale dell'effetto a distanza di alcuni anni dalla prima osservazione.[11] Un filone di ricerca più recente collega la scarsa coerenza del discorso negli anziani a un deficit dei processi esecutivi e semantici, distinto da quello che governa la coesione locale tra frasi contigue, un'evidenza che sostiene la scelta del modello di trattare pertinenza e coerenza testuale come dimensione autonoma rispetto alla coesione discorsiva.[12]

La fluidità articolatoria, infine, è il parametro più direttamente legato alla componente motoria della produzione linguistica piuttosto che a quella cognitiva in senso proprio. La letteratura sui disturbi del movimento riporta con concordanza che la disartria, tipica della malattia di Parkinson e di altre condizioni neurodegenerative, altera in modo misurabile la velocità di articolazione, la regolarità del ritmo e la produzione di sillabe in compiti di ripetizione rapida.[13] Un lavoro dedicato all'effetto della disartria sui compiti di fluenza verbale segnala inoltre che la disartria stessa può alterare la misurazione di parametri cognitivi come la fluenza semantica, se i due livelli, motorio e cognitivo, non vengono tenuti distinti nell'analisi.[14] Questa osservazione ha avuto un peso diretto nella costruzione del modello, discusso nel paragrafo seguente: la fluidità articolatoria è stata tenuta come dimensione distinta proprio per evitare che un deficit di natura motoria venga attribuito, in fase di interpretazione dei punteggi, a un deficit di natura cognitiva o discorsiva.

10. T. Y. Arbuckle, D. P. Gold, Aging, inhibition, and verbosity, «Journal of Gerontology», 48, 5, 1993, pp. P225-P232.

11. D. P. Gold, T. Y. Arbuckle, A Longitudinal Study of Off-Target Verbosity, «The Journals of Gerontology: Series B», 50B, 6, 1995, pp. P307-P315.

12. Sul deficit dei processi esecutivi e semantici come spiegazione della scarsa coerenza del discorso negli anziani, distinto dai meccanismi della coesione locale, cfr. Poor coherence in older people's speech is explained by impaired semantic and executive processes, riportato su PubMed, PMID 30179156.

13. Cfr. la rassegna dei test raccomandati in Speech and language biomarkers for Parkinson's disease prediction, early diagnosis and progression, «npj Parkinson's Disease», 2025, che descrive tra i test di riferimento la ripetizione rapida di sillabe per la valutazione di articolazione, velocità e coordinazione.

14. Cfr. Y. Li, J. Yang, K. Evans, J. B.-W. Wong, N. N. Dissanayaka, A. P. Vogel, Optimising verbal fluency analysis in neurological patients with dysarthria: Examples from Parkinson's disease and hereditary ataxia, «Journal of Clinical and Experimental Neuropsychology», 45, 5, 2023, pp. 452-463.

1.3 Secondo criterio: indipendenza reciproca delle dimensioni

Il secondo criterio ha richiesto che ciascuna dimensione candidata rappresentasse un livello di organizzazione del linguaggio distinto da quello delle altre dimensioni già selezionate, e non un'ulteriore variante di misurazione di un fenomeno già rappresentato. Questo criterio ha operato per lo più in senso escludente, riducendo il numero di parametri lessicali e sintattici distinti che la letteratura offre, non introducendo nuove dimensioni.

Il caso più chiaro riguarda proprio gli indici di ricchezza lessicale richiamati nel paragrafo precedente. La covarianza riportata tra indici come il rapporto type-token, la misura MTLD e l'indice di Honoré indica che questi parametri, per quanto calcolati con formule differenti, tendono a rappresentare uno stesso fenomeno sottostante, la varietà del vocabolario impiegato rispetto alla lunghezza del testo prodotto.[15] Il modello ne assume uno solo come rappresentativo della dimensione, rinviando alla Parte II la descrizione della formula adottata e la discussione delle ragioni della scelta rispetto alle alternative disponibili.

Un secondo caso riguarda il rapporto tra complessità sintattica e coesione discorsiva. I due fenomeni sono distinti sul piano teorico, la sintassi riguarda l'organizzazione interna della frase, il discorso riguarda il rapporto tra frasi successive, ma la letteratura registra un'interazione tra i due livelli: un lavoro sul discorso in soggetti con disturbo dello spettro schizofrenico associa la ridotta complessità sintattica e la scarsa verbosità a un deficit dei processi esecutivi di pianificazione, che si riflette anche sulla coerenza discorsiva.[16] Il modello mantiene le due dimensioni distinte, poiché lo stesso studio riporta che la coerenza discorsiva è associata soprattutto alla memoria di lavoro, mentre la verbosità e la complessità sintattica sono associate soprattutto alla fluenza e alla pianificazione, un'evidenza a favore della loro parziale indipendenza funzionale piuttosto che della loro sovrapposizione.

Un terzo caso, già anticipato, riguarda il rapporto tra fluidità articolatoria e le altre quattro dimensioni, tutte di natura più propriamente linguistico-cognitiva. Il modello tiene ferma questa distinzione per una ragione di cautela interpretativa oltre che teorica:

15. Zozuk et al., Relationship between language features, cit.

16. Cfr. lo studio riportato in An Exploration of Cohesion and Coherence Skills in Neuropsychiatric Disorder, «Schizophrenia Journal», 2022, che collega la ridotta verbosità e complessità sintattica a un deficit dei processi di pianificazione esecutiva, distinto dal deficit di coerenza discorsiva, quest'ultimo più associato alla memoria di lavoro.

un punteggio complessivo che confondesse il livello motorio con il livello cognitivo produrrebbe un risultato ambiguo, difficile da attribuire a una causa specifica in sede di lettura clinica o forense del profilo prodotto dal modello.

1.4 Terzo criterio: calcolabilità automatica

Il terzo criterio richiede che il parametro sia calcolabile in modo automatico a partire da una trascrizione del parlato spontaneo, prodotta a sua volta da un sistema di riconoscimento vocale, senza intervento di codifica manuale da parte di un valutatore. Questo criterio ha escluso dal modello, allo stato attuale, alcuni parametri che la letteratura clinica tradizionale considera rilevanti ma che richiedono ancora una valutazione condotta da un esaminatore qualificato.

Un caso rilevante è quello dei fenomeni pragmatici di ordine superiore, come la capacità di inferire lo stato mentale dell'interlocutore in un compito narrativo, la cosiddetta teoria della mente. La letteratura clinica riconosce a questi fenomeni un ruolo nel discorso patologico di alcune condizioni,[17] ma la loro operazionalizzazione richiede tuttora, nella maggior parte degli studi disponibili, una codifica manuale del testo secondo griglie qualitative, non una misura estraibile in modo automatico dalla sola trascrizione. Il modello non include, per ora, una dimensione dedicata a questi fenomeni. Si tratta di un'esclusione dichiaratamente provvisoria, legata allo stato dell'arte degli strumenti di annotazione automatica, non di un giudizio sulla rilevanza clinica del fenomeno.

Un secondo caso riguarda i tratti acustici in senso stretto, quali la variabilità dell'intonazione, il tremore vocale o la qualità timbrica della voce. Questi tratti sono automaticamente calcolabili, e la letteratura ne riconosce l'utilità in alcune condizioni, in particolare nei disturbi del movimento.[18] Sono stati tuttavia esclusi dal modello per una ragione di coerenza dell'oggetto, non di calcolabilità: FOXP2 è un modello del linguaggio, non del segnale vocale in quanto tale, e i tratti puramente acustici appartengono a un livello di analisi distinto, quello della fonetica strumentale, che il

17. Cfr. la discussione sul deficit di teoria della mente nel discorso di soggetti con disturbi dello spettro schizofrenico in An Exploration of Cohesion and Coherence Skills in Neuropsychiatric Disorder, cit.

18. Fraser, Meltzer, Rudzicz, Linguistic features identify Alzheimer's disease, cit., include un sottogruppo di tratti acustici accanto ai sottogruppi linguistici propriamente detti.

modello tiene separato dalla pipeline linguistica propriamente detta pur riconoscendone la complementarità.

Un terzo caso, di segno opposto, riguarda la fluidità articolatoria, che pure ha una componente motoria evidente. Questa dimensione è stata mantenuta, a differenza dei tratti acustici in senso stretto, perché la sua misurazione nel modello non si basa su parametri acustici del segnale ma su indicatori derivabili dalla trascrizione temporizzata, quali la velocità di produzione delle sillabe, la frequenza e la durata delle pause, la regolarità del ritmo articolatorio. Si tratta di una misura linguistica della fluidità, non di una misura acustica della voce, ed è per questo motivo, oltre che per la rilevanza diagnostica già discussa nel paragrafo 1.2, che rientra tra le cinque dimensioni del modello.

1.5 Le cinque dimensioni come risultato del filtro combinato

Le cinque dimensioni che compongono FOXP2, densità lessicale, complessità sintattica, coesione discorsiva, pertinenza e coerenza testuale, fluidità articolatoria, non sono state stabilite in anticipo e poi giustificate a posteriori. Sono il risultato dell'applicazione sequenziale dei tre criteri appena descritti a un insieme di partenza più ampio, dell'ordine delle decine di parametri candidati desunti dalla letteratura richiamata in questo capitolo.

1.6 Dalla misura continua alla scala clinica: l'architettura del modello

Va reso esplicito, a chiusura di questo capitolo, il rapporto tra due livelli del modello che il lettore incontrerà separatamente nei capitoli successivi e che è necessario tenere fermi come un'unica architettura, non come due sistemi alternativi. Il primo livello è quello delle misure linguistiche continue descritte nella Parte II, densità lessicale, distanza di dipendenza sintattica e simili, calcolate in modo automatico dalla pipeline computazionale del modello. Il secondo livello è quello della scala di valutazione clinica a dieci punti, con ancoraggi descrittivi qualitativi per ciascun grado, che la domanda di brevetto definisce per ciascuno dei cinque indicatori come esito dell'analisi quantitativa condotta con gli strumenti software indicati nella domanda stessa.

Il modello FOXP2 non tratta questi due livelli come alternativi, ma come lo stesso processo osservato a due gradi diversi di risoluzione. La misura continua è il dato che il

sistema calcola direttamente dalla trascrizione e dal segnale audio; il grado sulla scala a dieci punti è la traduzione di quella misura, normalizzata rispetto al database normativo del modello, nella forma ordinale con cui la domanda di brevetto definisce ciascun indicatore, e che rende il punteggio leggibile in sede clinica secondo gli ancoraggi qualitativi già fissati nella domanda stessa. Questa architettura a due livelli va dichiarata esplicitamente per una ragione di onestà espositiva: il testo della domanda di brevetto, così come depositato, descrive la scala di valutazione e gli strumenti software di supporto, non una pipeline di riconoscimento vocale interamente automatizzata né una procedura di normalizzazione statistica contro un campione normativo. Le formule computazionali dettagliate nei capitoli della Parte II costituiscono lo sviluppo con cui chi scrive ha inteso automatizzare e rendere riproducibile il calcolo dei punteggi previsti dalla domanda di brevetto, non una componente già rivendicata nel testo depositato. Il manuale segnala questa distinzione ogni volta che sia rilevante, per evitare che al brevetto in esame vengano attribuite capacità che il testo depositato non descrive.

La domanda di brevetto specifica inoltre il protocollo di raccolta dei dati su cui l'intero modello opera: una registrazione audio della durata compresa preferibilmente tra uno e due minuti, in cui il paziente descrive verbalmente le attività svolte nel fine settimana precedente, effettuata in ambiente controllato con un registratore digitale. La prima registrazione costituisce la campionatura di base, indicata come T0; una seconda registrazione, indicata come T1, viene effettuata dopo un intervallo non inferiore a tre mesi, con la possibilità di ulteriori campionature successive, T2, T3 e così via, per un monitoraggio a lungo termine. I punteggi delle diverse campionature vengono confrontati su un grafico radar a cinque assi, uno per ciascun indicatore linguistico, la cui contrazione progressiva segnala il deterioramento delle competenze linguistiche monitorate: la domanda di brevetto indica che un peggioramento significativo si verifica quando il poligono generato dalla campionatura più recente risulta inscritto, per oltre il 50% della propria area, all'interno del poligono della campionatura precedente. Questo protocollo è comune a tutte e cinque le dimensioni.

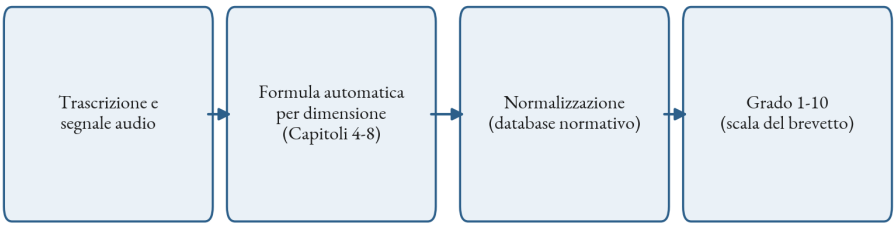

Figura 1.1. I due livelli dell'architettura del modello: dalla trascrizione alla formula automatica di ciascuna dimensione, dalla normalizzazione contro il database normativo al grado sulla scala a dieci punti della domanda di brevetto.

La domanda di brevetto prevede inoltre, come parte integrante della sua documentazione, una scheda per la raccolta dei dati anagrafici e clinici del paziente, allegata come prima figura della domanda. La Tabella 2 riassume, a titolo illustrativo, la struttura tipica di una scheda di questo genere, senza riprodurne il layout grafico originale, che esula dagli scopi di un manuale teorico come il presente.

Tabella 2. Struttura illustrativa della scheda di raccolta dei dati anagrafici e clinici prevista dalla domanda di brevetto, come Figura 1 della documentazione allegata.

Tabella 2. Struttura illustrativa della scheda di raccolta dei dati anagrafici e clinici prevista dalla domanda di brevetto, come Figura 1 della documentazione allegata.

Campo	**Contenuto**
Dati anagrafici	Nome, data di nascita, sesso, livello di scolarizzazione, lingua madre
Dati clinici	Diagnosi o sospetto diagnostico, comorbilità rilevanti, terapie farmacologiche in corso
Dati della registrazione	Data e ora, identificativo della campionatura (T0, T1, T2, ...), durata effettiva, condizioni ambientali
Consegna somministrata	Testo della traccia assegnata al paziente, secondo il protocollo standardizzato di cui al paragrafo 5.1 della domanda di brevetto

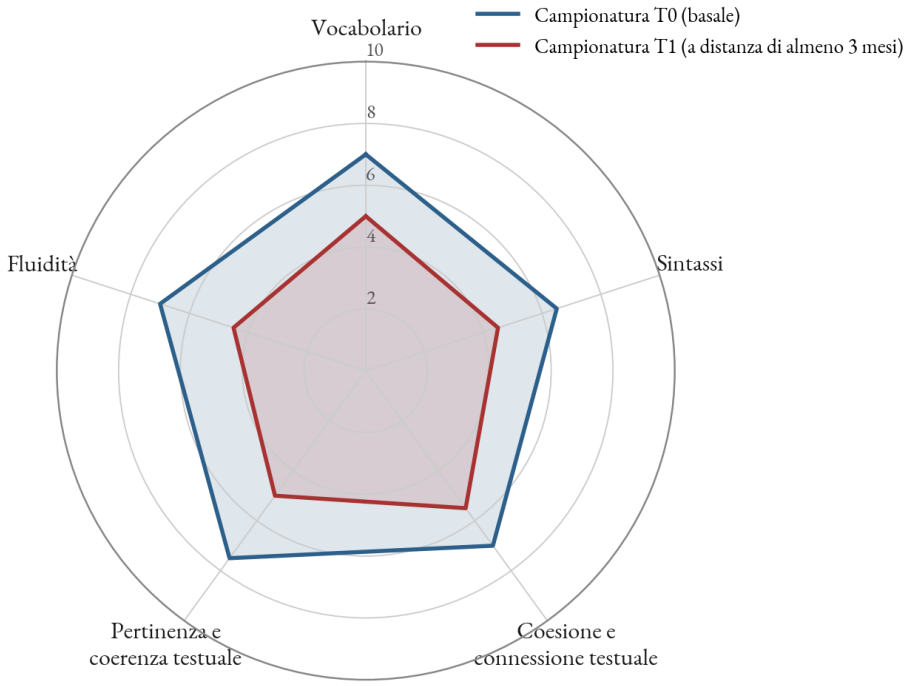

Figura 1.2. Esempio illustrativo, con valori ipotetici, del confronto tra la campionatura di base T0 e una campionatura successiva T1 sul grafico radar previsto dalla domanda di brevetto. I valori riportati non provengono da alcuna somministrazione reale e servono unicamente a illustrare la logica di lettura del grafico: la contrazione del poligono più recente rispetto al precedente, quando inscritto per oltre la metà della propria area in quest'ultimo, segnala secondo la domanda di brevetto un peggioramento significativo delle competenze linguistiche monitorate.

Il capitolo successivo tratta il rapporto tra il nome scelto per il modello e il gene FOXP2 da cui il nome deriva per via metaforica. I capitoli della Parte II riprendono ciascuna delle cinque dimensioni qui introdotte, descrivendone la formula di calcolo, la fonte nella letteratura di riferimento, il rapporto con la scala di valutazione della domanda di brevetto e i limiti specifici che ne accompagnano l'uso.

Riferimenti bibliografici del Capitolo 1

Arbuckle T. Y., Gold D. P., Aging, inhibition, and verbosity, «Journal of Gerontology», 48, 5, 1993, pp. P225-P232.

Bertola A. et al., Exploring the Use of Cohesive Devices in Dementia, in Atti della decima conferenza italiana di linguistica computazionale (CLiC-it 2024).

Fraser K. C., Meltzer J. A., Rudzicz F., Linguistic features identify Alzheimer's disease in narrative speech, «Journal of Alzheimer's Disease», 49, 2, 2016, pp. 407-422.

Fraser K. C. et al., Language Impairment in Alzheimer's Disease. Robust and Explainable Evidence for AD-Related Deterioration of Spontaneous Speech Through Multilingual Machine Learning, «Frontiers in Aging Neuroscience», 2021.

Gold D. P., Arbuckle T. Y., A Longitudinal Study of Off-Target Verbosity, «The Journals of Gerontology: Series B», 50B, 6, 1995, pp. P307-P315.

Hakala T. et al., Analysis of Language for Dementia Diagnosis with Focus on Automatic Speech Recognition, University of Washington, 2017.

Kim B. S., Kim Y. B., Kim H., Discourse measures to differentiate between mild cognitive impairment and healthy aging, «Frontiers in Aging Neuroscience».

Lai C. S. L., Fisher S. E., Hurst J. A., Vargha-Khadem F., Monaco A. P., A forkhead-domain gene is mutated in a severe speech and language disorder, «Nature», 413, 2001, pp. 519-523.

Lundholm Fors K., Automated Speech and Language Analysis Techniques for the Detection and Assessment of Language and Speech Impairment.

Comparing global and local semantic coherence of spontaneous speech in persons with Alzheimer's disease and healthy controls, «Applied Corpus Linguistics», 3, 3, 2023, articolo 100064.

Poor coherence in older people's speech is explained by impaired semantic and executive processes, PubMed, PMID 30179156.

Speech and language biomarkers for Parkinson's disease prediction, early diagnosis and progression, «npj Parkinson's Disease», 2025.

Li Y., Yang J., Evans K., Wong J. B.-W., Dissanayaka N. N., Vogel A. P., Optimising verbal fluency analysis in neurological patients with dysarthria: Examples from Parkinson's disease and hereditary ataxia, «Journal of Clinical and Experimental Neuropsychology», 45, 5, 2023, pp. 452-463.

An Exploration of Cohesion and Coherence Skills in Neuropsychiatric Disorder: Speech Language Pathologist Perspective, «Schizophrenia Journal», 2022.

Zozuk E. et al., Relationship between language features extracted through NLP and clinically diagnosed Alzheimer's disease and mild cognitive impairment in Slovak, «Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring», 17, 3, 2025, articolo e70122.

Capitolo 2 FOXP2 come costrutto

Dal gene alla metafora del nome

2.1 Il gene FOXP2: sintesi essenziale della scoperta

Resta in sospeso, dal capitolo precedente, il nome del modello. FOXP2 richiama direttamente un gene: la selezione delle cinque dimensioni non basta a chiarire il rapporto tra il costrutto teorico e il locus genetico da cui prende il nome. Bisogna prima ricostruire cosa la ricerca genetica abbia accertato, per separare il piano biologico da quello teorico che ne eredita la denominazione.

FOXP2 è stato il primo gene a essere associato in modo diretto a un disturbo ereditario della comunicazione. La scoperta nasce dallo studio di un ampio nucleo familiare britannico, indicato in letteratura con la sigla KE, in cui circa metà dei membri, distribuiti su tre generazioni, presentava un disturbo denominato disprassia verbale evolutiva, caratterizzato da difficoltà nel sequenziare i movimenti muscolari necessari all'articolazione del parlato, accompagnato da difficoltà più ampie nell'elaborazione linguistica e grammaticale.[19] L'analisi genetica identificò nei membri affetti una mutazione puntiforme a carico di un unico gene sul cromosoma 7, che alterava un amminoacido invariante nel dominio forkhead della proteina corrispondente, da cui il nome FOXP2, acronimo di forkhead box P2.[20]

Sul piano molecolare, FOXP2 codifica un fattore di trascrizione: una proteina che si lega a regioni regolatrici di altri geni e ne modula l'espressione, non un gene che produce direttamente una funzione linguistica.[21] Ricerche successive hanno confermato il

19. C. S. L. Lai, S. E. Fisher, J. A. Hurst, F. Vargha-Khadem, A. P. Monaco, A forkhead-domain gene is mutated in a severe speech and language disorder, «Nature», 413, 2001, pp. 519-523. Il pedigree della famiglia KE era stato descritto per la prima volta in J. Hurst, M. Baraitser, E. Auger, F. Graham, S. Norell, An extended family with a dominantly inherited speech disorder, «Developmental Medicine and Child Neurology», 32, 1990, pp. 347-355.

20. Lai et al., A forkhead-domain gene, cit. Sulla struttura del gene e sull'estensione della regione genomica coinvolta, superiore a 267 kilobasi, cfr. la medesima fonte.

21. Sulla natura di FOXP2 come fattore di trascrizione e sui suoi bersagli genici diretti, identificati mediante analisi ad alto rendimento dell'occupazione dei promotori, cfr. S. C. Vernes et al., High-throughput analysis of promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in speech and language disorders, richiamato in Assessing the effects of common variation in the FOXP2 gene on human brain structure, disponibile su PubMed Central, PMC4076884.

quadro clinico osservato nella famiglia KE anche in casi indipendenti, tra cui un soggetto non imparentato portatore di un riarrangiamento cromosomico che interrompeva lo stesso locus, e un piccolo gruppo di soggetti con disprassia verbale portatori di varianti troncanti della proteina.[22] Studi di neuroimmagine sui membri affetti della famiglia KE hanno inoltre documentato anomalie bilaterali in strutture cerebrali coinvolte nel controllo motorio, in particolare nei nuclei della base e nelle aree corticali motorie, coerenti con un disturbo primariamente motorio-articolatorio più che con un deficit linguistico-cognitivo generale.[23]

2.2 Il problema del «gene del linguaggio»

La scoperta del 2001 ebbe una risonanza pubblica sproporzionata rispetto alla cautela con cui era stata formulata dagli stessi autori. La stampa generalista adottò rapidamente l'etichetta di «gene del linguaggio», presentando FOXP2 come la base genetica unica della facoltà linguistica umana.[24] Simon Fisher, uno degli autori dello studio originale del 2001, ha ripetutamente contestato questa lettura nel corso degli anni successivi, sottolineando come un fattore di trascrizione agisca sempre in modo indiretto, modulando l'espressione di altri geni e non producendo da solo alcuna funzione cognitiva complessa.[25]

Le analisi della sequenza codificante di FOXP2 condotte su ampi campioni di soggetti con disturbo specifico del linguaggio, autismo o dislessia non hanno individuato varianti eziologiche paragonabili a quella della famiglia KE.[26] Il gene, in altre parole, spiega un fenotipo clinico circoscritto, il disturbo osservato nella famiglia KE e in pochi

22. K. D. MacDermot et al., Identification of FOXP2 Truncation as a Novel Cause of Developmental Speech and Language Deficits, «American Journal of Human Genetics», 76, 6, 2005, pp. 1074-1080.

23. Cfr. la rassegna delle evidenze di neuroimmagine in FOXP2 and the neuroanatomy of speech and language, «Nature Reviews Neuroscience», che riporta come resti tuttora dibattuto se gli effetti della mutazione derivino da un deficit motorio primario o da deficit cognitivi ulteriori.

24. Sull'origine e la fortuna mediatica dell'etichetta «language gene», cfr. la ricostruzione in A Tentative Role for FOXP2 in the Evolution of Dual Processing Modes and Generative Abilities, che attribuisce a D. Bickerton, 2007, una delle prime critiche pubblicate a questa semplificazione.

25. Sulla presa di posizione di Fisher rispetto alla semplificazione mediatica, cfr. la ricostruzione divulgativa in Understanding the FOXP2 Gene: More Than Just a Language Gene, che richiama le dichiarazioni dello stesso Fisher sul carattere non riducibile a un singolo gene della facoltà linguistica.

26. MacDermot et al., Identification of FOXP2 Truncation, cit., che riporta l'assenza di varianti eziologiche di FOXP2 in coorti con disturbo specifico del linguaggio, autismo e dislessia.

casi analoghi, non l'insieme delle competenze linguistiche della specie. Chi adotta il nome FOXP2 per un modello linguistico non biologico eredita un rischio preciso: ripetere, su un piano diverso, la stessa semplificazione già segnalata dalla comunità genetica.

2.3 Cosa il modello eredita dal gene, e cosa no

Il modello FOXP2 non implementa alcun meccanismo genetico, né riproduce in forma computazionale la funzione biologica del fattore di trascrizione omonimo. Non esiste corrispondenza tecnica tra le cinque dimensioni linguistiche del modello e i bersagli genici regolati da FOXP2 in ambito neurale: leggervi una base genetica diretta travisa l'impostazione del lavoro.

Ciò che il modello eredita dal gene è un'analogia strutturale, circoscritta e dichiarata come tale. Il fenotipo clinico osservato nella famiglia KE combina un deficit primariamente motorio-articolatorio con difficoltà più ampie di elaborazione linguistica e grammaticale, un quadro che gli studi di neuroimmagine collegano al coinvolgimento di strutture implicate sia nel controllo motorio sia in circuiti cognitivi più generali.[27] Questa duplicità, componente motoria e componente cognitivo-linguistica insieme ma concettualmente distinte, offre un'immagine utile per l'architettura del modello: una dimensione dedicata alla fluidità articolatoria accanto a quattro dimensioni lessicali, sintattiche, discorsive e tematiche, discusse nel Capitolo 1. Resta un'analogia espositiva, non una derivazione tecnica. Il Capitolo 1 ha già mostrato che i criteri di selezione delle cinque dimensioni vengono dalla letteratura sulla linguistica clinica computazionale, non da considerazioni genetiche.

Il nome è stato scelto per il suo valore euristico: richiama il rapporto tra una base biologica del linguaggio e la sua articolazione osservabile, senza pretendere di rappresentarne un'implementazione. Espone però il modello allo stesso rischio di lettura riduzionistica già visto per il gene. Per questo la distinzione va dichiarata qui, non lasciata al lettore.

27. Sulla compresenza di deficit motorio-articolatorio e difficoltà linguistico-grammaticali più ampie nella famiglia KE, cfr. Lai et al., A forkhead-domain gene, cit.; sul dibattito relativo al rapporto tra i due livelli, cfr. FOXP2 and the neuroanatomy of speech and language, cit.

2.4 Una precisazione terminologica

D'ora in avanti, «il modello FOXP2» indica sempre e solo il costrutto teorico descritto in queste pagine; dove serva riferirsi al gene, il testo scrive «il gene FOXP2» per esteso, per non lasciare che una comoda ellissi riapra l'ambiguità appena dissolta. Il capitolo seguente colloca il modello così definito nel campo più ampio della linguistica clinica computazionale, di cui condivide gli strumenti ma non le premesse biologiche.

Riferimenti bibliografici del Capitolo 2

Assessing the effects of common variation in the FOXP2 gene on human brain structure, PubMed Central, PMC4076884.

FOXP2 and the neuroanatomy of speech and language, «Nature Reviews Neuroscience».

Hurst J., Baraitser M., Auger E., Graham F., Norell S., An extended family with a dominantly inherited speech disorder, «Developmental Medicine and Child Neurology», 32, 1990, pp. 347-355.

Lai C. S. L., Fisher S. E., Hurst J. A., Vargha-Khadem F., Monaco A. P., A forkhead-domain gene is mutated in a severe speech and language disorder, «Nature», 413, 2001, pp. 519-523.

MacDermot K. D. et al., Identification of FOXP2 Truncation as a Novel Cause of Developmental Speech and Language Deficits, «American Journal of Human Genetics», 76, 6, 2005, pp. 1074-1080.

A Tentative Role for FOXP2 in the Evolution of Dual Processing Modes and Generative Abilities, disponibile su arXiv.

Understanding the FOXP2 Gene: More Than Just a Language Gene, 2025.

Vernes S. C. et al., High-throughput analysis of promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in speech and language disorders.

Capitolo 3 Posizionamento rispetto alla linguistica clinica computazionale

3.1 Due tradizioni distinte

Chiarita la genesi delle cinque dimensioni e il rapporto puramente metaforico tra il modello e il gene omonimo, resta da collocare FOXP2 rispetto al campo di studi in cui si inserisce. Occorre a tal fine distinguere due tradizioni che la letteratura recente tende talvolta a sovrapporre, pur avendo origini e metodi differenti. La prima è la linguistica clinica, disciplina che applica gli strumenti della linguistica teorica e descrittiva allo studio, alla diagnosi e al trattamento dei disturbi del linguaggio. Una definizione ormai classica del campo la descrive come l'applicazione della linguistica teorica e descrittiva alla patologia del linguaggio e alla sua riabilitazione.[28] Il linguista clinico, in questa tradizione, mette a disposizione del clinico una descrizione rigorosa dei fenomeni linguistici osservati, senza che questo comporti necessariamente l'uso di strumenti computazionali.

La seconda tradizione è quella dell'elaborazione automatica del linguaggio applicata a scopi clinici, comunemente indicata con l'acronimo NLP, dall'inglese natural language processing. Si tratta di un insieme di tecniche computazionali che permettono di analizzare e rappresentare il testo a diversi livelli linguistici in modo automatico.[29] Quando questi strumenti vengono applicati alla produzione linguistica di un soggetto per ricavarne un indicatore clinicamente informativo, la letteratura recente parla di marcatore NLP, definito come una misura quantificabile della produzione linguistica umana, acquisita in forma digitale e derivata computazionalmente, che riflette lo stato dei processi biologici, neurocognitivi e sociali che la determinano.[30]

Il modello FOXP2 si colloca all'incrocio di queste due tradizioni, ma non le confonde. Le cinque dimensioni descritte nel Capitolo 1 sono state selezionate con un

28. M. R. Perkins, S. J. Howard, Case Studies in Clinical Linguistics, Whurr, London, 1995, p. 11, citato in Clinical Linguistics: Analysis of Mapping Knowledge Domains in Past, Present and Future, disponibile su PubMed Central, PMC9406678.

29. Sulla definizione di NLP come insieme di tecniche computazionali per l'analisi del testo a più livelli linguistici, cfr. la rassegna in Artificial intelligence and natural language processing in modern clinical neuropsychology, «Journal of Clinical and Experimental Neuropsychology», 2025, che richiama la definizione proposta in Crema et al., 2022.

30. La definizione di marcatore NLP qui riportata è quella proposta in Corona Hernández et al., 2023, richiamata nella rassegna citata alla nota precedente.

criterio di rilevanza diagnostica che deriva dalla linguistica clinica, e sono rese misurabili attraverso una pipeline computazionale che appartiene alla tradizione del NLP applicato. Il modello non è quindi riducibile all'una o all'altra tradizione, ma ne rappresenta un tentativo di integrazione, i cui termini vanno resi espliciti per evitare l'ambiguità terminologica che accompagna spesso l'uso disinvolto di entrambe le etichette.

3.2 Gli strumenti già esistenti: la logica del insieme ampio di caratteristiche

Il campo dei marcatori linguistici digitali applicati alla diagnosi non è privo di strumenti già sviluppati, alcuni dei quali commercializzati. Il caso più rappresentativo è quello della piattaforma sviluppata da Winterlight Labs, che elabora le registrazioni del parlato spontaneo per generare centinaia di marcatori linguistici e acustici, combinati successivamente in un punteggio composito.[31] Una rassegna sistematica pubblicata nel 2026 su ventisette strumenti digitali di screening cognitivo riporta coefficienti di correlazione, rispetto a strumenti clinici di riferimento come il Montreal Cognitive Assessment, il Mini-Mental State Examination e la batteria cognitiva composita preclinica per la malattia di Alzheimer, compresi tra 0,39 e 0,84 a seconda del metodo statistico impiegato.[32]

Questi sistemi condividono un'impostazione metodologica di fondo: partono da un numero molto ampio di caratteristiche estratte in modo sistematico dal segnale linguistico e acustico, spesso senza una selezione teorica a priori, e affidano a un modello statistico o di apprendimento automatico il compito di individuare, tra queste centinaia di variabili, quelle più predittive rispetto a un determinato esito clinico. Si tratta di una logica per lo più induttiva: le caratteristiche rilevanti emergono dai dati, non da un impianto teorico definito prima dell'analisi.

31. Sull'estrazione di oltre cinquecento marcatori acustici e linguistici e sulla loro combinazione in un punteggio composito, cfr. il comunicato relativo allo studio condotto in collaborazione con Genentech sullo sperimentazione Tauriel, Winterlight Labs, 2022. Il sito istituzionale della società dichiara accuratezze comprese tra l'82% e il 100% nella rilevazione di più condizioni cliniche a partire dagli stessi marcatori, riportate in una serie di studi peer reviewed elencati nella sezione dedicata alla ricerca clinica.

32. S. Hong, C. Kim, S. H. Lee, Digital tools and biomarkers for cognitive screening: A systematic review, «Digital Health», 2026.

3.3 La scelta di FOXP2: un numero ridotto di dimensioni motivate

Il modello descritto in questo manuale adotta una logica diversa, deduttiva più che induttiva, i cui criteri sono stati esposti nel Capitolo 1. Le cinque dimensioni non emergono da un processo statistico applicato a un insieme ampio di variabili disponibili, ma sono state selezionate a monte sulla base di una rilevanza diagnostica documentata in letteratura indipendente, di un requisito di indipendenza reciproca tra i fenomeni misurati, e della loro calcolabilità automatica. Questa scelta comporta un compromesso che va dichiarato con onestà, non nascosto dietro la sistematicità dell'esposizione teorica.

Un insieme ridotto di dimensioni teoricamente motivate offre un vantaggio di interpretabilità: ogni punteggio prodotto dal modello è riconducibile a un fenomeno linguistico definito e a una letteratura che ne discute il significato, un requisito rilevante in particolare per l'uso in ambito forense, dove la tracciabilità della singola misura rispetto alla sua fonte teorica è parte integrante della sua difendibilità in sede di contraddittorio. Un insieme ampio di centinaia di caratteristiche, per contro, può in linea di principio raggiungere un potere discriminativo superiore, come indicano le accuratezze dichiarate per gli strumenti commerciali già richiamati, ma a costo di una minore trasparenza interpretativa: un punteggio composito derivato da centinaia di variabili correlate tra loro in modo non sempre esplicitato è più difficile da motivare punto per punto rispetto a un punteggio costruito su cinque dimensioni dichiarate.

Quale delle due logiche sia preferibile dipende dallo scopo, non ammette una risposta assoluta. Il confronto empirico tra le due, in termini di potere discriminativo effettivo, resta materia da accertare: qui interessa solo dichiarare quale logica il modello adotta.

Un ultimo aspetto distingue il modello FOXP2 dai sistemi a punteggio composito opaco richiamati nel paragrafo 3.2, in un modo diverso dal solo numero di variabili impiegate. Winterlight e strumenti analoghi restituiscono un punteggio composito unico, ottenuto combinando centinaia di variabili attraverso un modello statistico o di apprendimento automatico i cui pesi relativi non sono, di norma, resi interpretabili all'utente clinico. Il modello FOXP2, per come definito nella domanda di brevetto, restituisce invece cinque punteggi distinti, uno per dimensione, ciascuno collocato su una scala a dieci punti con ancoraggi descrittivi qualitativi espliciti. Questa architettura, discussa nel dettaglio nel Capitolo 1 e ripresa in ciascun capitolo della Parte II, non è quindi soltanto una scelta di parsimonia nel numero delle dimensioni, discussa nel paragrafo 3.3, ma anche una scelta di trasparenza nella forma dell'output: un punteggio

per dimensione, leggibile secondo un ancoraggio qualitativo dichiarato, contro un punteggio unico la cui composizione interna resta in gran parte opaca all'utente finale.

Tabella 3.1. Confronto architetturale tra il modello FOXP2 e i sistemi a punteggio composito del tipo descritto nel paragrafo 3.2.

Tabella 3.1. Confronto architetturale tra il modello FOXP2 e i sistemi a punteggio composito del tipo descritto nel paragrafo 3.2.

Caratteristica	**Sistemi a punteggio composito (es. Winterlight)**	**Modello FOXP2**
Numero di variabili in ingresso	Centinaia, selezionate per via induttiva	Cinque dimensioni, selezionate per via deduttiva secondo i tre criteri del Capitolo 1
Output restituito	Un punteggio composito unico	Cinque punteggi distinti, uno per dimensione
Interpretabilità dei pesi	Non resa nota, di norma, all'utente clinico	Ogni dimensione tracciabile a una formula dichiarata (Parte II)
Forma di lettura clinica	Punteggio numerico o probabilità	Grado su scala a dieci punti con ancoraggio qualitativo esplicito, definito dalla domanda di brevetto
Fonte della selezione delle variabili	Ottimizzazione statistica su dati di addestramento	Letteratura indipendente e definizione della domanda di brevetto

3.4 Il confronto con i test da tavolo: MMSE e MoCA

Un confronto tra FOXP2 e gli strumenti oggi impiegati dai neurologi per la valutazione del decadimento cognitivo va condotto con una cautela preliminare. I due test cognitivo-linguistici da tavolo più diffusi nella pratica clinica, il Mini-Mental State Examination e il Montreal Cognitive Assessment, dispongono di decenni di validazione, di soglie diagnostiche consolidate e di adattamenti in decine di lingue; il modello FOXP2 non dispone ancora di nulla di tutto questo. Il confronto che segue riguarda dunque la struttura e il metodo dei tre strumenti, non la loro accuratezza diagnostica comparata: un confronto di questo secondo tipo richiederebbe dati di validazione di FOXP2 che allo stato non esistono, e presentarlo comunque significherebbe suggerire una parità che il modello non ha ancora guadagnato.

Il Mini-Mental State Examination fu introdotto da Folstein, Folstein e McHugh nel 1975 come strumento di screening rapido dello stato cognitivo, articolato in undici prove che coprono orientamento, memoria, attenzione, linguaggio e prassia costruttiva, con un punteggio complessivo su trenta punti e un tempo di somministrazione di circa dieci minuti.[33] Il Montreal Cognitive Assessment, sviluppato da Nasreddine e colleghi nel 2005 per colmare la scarsa sensibilità del MMSE al deterioramento cognitivo lieve, copre un numero maggiore di domini, in particolare le funzioni esecutive e l'astrazione, anch'esso su una scala a trenta punti e con un tempo di somministrazione paragonabile. [34] Il MoCA riporta una sensibilità al deterioramento cognitivo lieve superiore a quella del MMSE, indicativamente 90% contro 81%, a fronte di una specificità lievemente inferiore per le sindromi non amnesiche e non riconducibili alla malattia di Alzheimer.[35]

Entrambi gli strumenti condividono tre limiti strutturali rilevanti per il confronto con FOXP2. Il primo riguarda la mediazione linguistica degli item: prove pensate per misurare l'orientamento, la memoria o le funzioni esecutive richiedono in realtà una competenza linguistica intatta per essere superate, un fatto che rende ambiguo se un punteggio basso rifletta un deficit cognitivo generale o una difficoltà specificamente linguistica.[36] FOXP2 non elimina questo problema, ma lo capovolge: tratta il linguaggio come oggetto diretto della misura, non come canale neutro attraverso cui misurare altro.

Il secondo limite riguarda la sensibilità del punteggio alla scolarizzazione. Una quota rilevante della varianza nei punteggi di MMSE e MoCA è spiegata dagli anni di istruzione del soggetto, al punto che una soglia numerica fissa produce tassi di falso positivo sistematicamente più alti nei soggetti con minore scolarizzazione.[37] In una po-

33. M. F. Folstein, S. E. Folstein, P. R. McHugh, "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician, «Journal of Psychiatric Research», 12, 3, 1975, pp. 189-198.

34. Z. S. Nasreddine et al., The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment, «Journal of the American Geriatrics Society», 53, 4, 2005, pp. 695-699.

35. Sul confronto di sensibilità tra MoCA e MMSE nella rilevazione del deterioramento cognitivo lieve, cfr. la sintesi in Montreal Cognitive Assessment (MoCA), voce di raccolta bibliografica consultata nel 2025.

36. Sulla difficoltà di distinguere il contributo linguistico da quello cognitivo generale negli item del MMSE, e sulle modifiche necessarie in fase di traduzione per costrutti privi di equivalente diretto in altre lingue, cfr. Cross-cultural adaptation and psychometric properties of the MMSE and MoCA questionnaires in Tanzanian Swahili for a traumatic brain injury population, «BMC Neurology», 2019.

37. Sull'effetto della scolarizzazione sui punteggi di MMSE e MoCA e sul rischio di falsi positivi in soggetti con minore istruzione, cfr. Y. Wu, Y. Zhang, X. Yuan, J. Guo, X. Gao, Influence of education level on MMSE and MoCA scores of elderly inpatients, «Applied Neuropsychology: Adult», 30, 2023, pp. 414-418; I. Pellicer-Espinosa, U. Díaz-Orueta, Cognitive screening instruments for older adults with low educational and literacy levels: a systematic review, «Journal of Applied Gerontology», 41, 2022, pp. 1222-1231.

polazione statunitense, questo effetto si è tradotto in un tasso di falso positivo del 42% tra i soggetti afroamericani contro il 6% tra i soggetti bianchi sottoposti al medesimo test.[38] FOXP2 non è esente da un problema analogo: le sue cinque dimensioni richiedono a loro volta un database normativo stratificato per scolarizzazione, discusso nei Capitoli 4 e 8, la cui costruzione resta un compito aperto. La differenza, allo stato, è che il problema viene dichiarato come tale fin dalla progettazione, non scoperto dopo decenni di uso su una soglia unica.

Il terzo limite, il più direttamente rilevante per il confronto con il protocollo descritto nel Capitolo 1, riguarda l'effetto di apprendimento nelle somministrazioni ripetute. Poiché MMSE e MoCA impiegano item identici a ogni somministrazione, un soggetto tende a migliorare il proprio punteggio per la sola familiarità con il test, indipendentemente da un reale cambiamento cognitivo: in media un soggetto guadagna circa un punto al MMSE se ritestato a un anno di distanza, e l'effetto è stato documentato anche a un intervallo di tre mesi, lo stesso intervallo minimo previsto dal protocollo di FOXP2 per le campionature successive.[39] Un effetto di questo tipo maschera il declino reale, perché la componente di apprendimento e la componente di deterioramento agiscono in direzioni opposte sullo stesso punteggio. Il protocollo di FOXP2, fondato su un compito di produzione libera anziché su una batteria di item fissi, non è strutturalmente esposto allo stesso meccanismo: il contenuto prodotto dal parlante cambia a ogni somministrazione, anche a fronte di una consegna stabile. Resta però una domanda aperta se la sola familiarità con la consegna produca un effetto di apprendimento residuo, distinto da quello legato agli item: è materia di validazione, non un vantaggio da dare per acquisito.

3.5 Una collocazione dichiaratamente ibrida

Il modello FOXP2 attinge dunque alla linguistica clinica per la selezione delle dimensioni, al NLP applicato per la loro misurazione automatica, alla filologia

38. Sul divario nei tassi di falso positivo tra gruppi demografici nella somministrazione del MMSE, cfr. Cognitive decline assessment in speakers of understudied languages, «Alzheimer's & Dementia: Translational Research & Clinical Interventions», 2023.

computazionale per l'impianto metodologico. Non è indecisione disciplinare: un modello del linguaggio va costruito con criteri propri, anche a costo di attraversare più tradizioni di ricerca. La Parte II riprende ciascuna delle cinque dimensioni così collocate, entrando nel dettaglio della loro formula di calcolo e della letteratura specifica che le riguarda.

Riferimenti bibliografici del Capitolo 3

Cognitive decline assessment in speakers of understudied languages, «Alzheimer's & Dementia: Translational Research & Clinical Interventions», 2023.

Cross-cultural adaptation and psychometric properties of the MMSE and MoCA questionnaires in Tanzanian Swahili for a traumatic brain injury population, «BMC Neurology», 2019.

Folstein M. F., Folstein S. E., McHugh P. R., "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician, «Journal of Psychiatric Research», 12, 3, 1975, pp. 189-198.

Montreal Cognitive Assessment (MoCA), voce di raccolta bibliografica consultata nel 2025.

Nasreddine Z. S. et al., The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment, «Journal of the American Geriatrics Society», 53, 4, 2005, pp. 695-699.

Pellicer-Espinosa I., Díaz-Orueta U., Cognitive screening instruments for older adults with low educational and literacy levels: a systematic review, «Journal of Applied Gerontology», 41, 2022, pp. 1222-1231.

Test-Retest Reliability of a Serious Game for Delirium Screening in the Emergency Department, «Frontiers in Aging Neuroscience», 2016.

39. Sull'effetto di apprendimento del MMSE a un anno di distanza e sulla sua documentazione a un intervallo di tre mesi, cfr. Helkala et al., 2002, richiamato in Validity and reliability of two alternate versions of the Montreal Cognitive Assessment (Hong Kong version) for screening of Mild Neurocognitive Disorder, PubMed Central, PMC5965815; sull'effetto di apprendimento del MoCA in campioni longitudinali di anziani sani, cfr. Cooley et al., 2015, citato in Test-Retest Reliability of a Serious Game for Delirium Screening in the Emergency Department, «Frontiers in Aging Neuroscience», 2016.

Validity and reliability of two alternate versions of the Montreal Cognitive Assessment (Hong Kong version) for screening of Mild Neurocognitive Disorder, PubMed Central, PMC5965815.

Wu Y., Zhang Y., Yuan X., Guo J., Gao X., Influence of education level on MMSE and MoCA scores of elderly inpatients, «Applied Neuropsychology: Adult», 30, 2023, pp. 414-418.

Artificial intelligence and natural language processing in modern clinical neuropsychology: A narrative review, «Journal of Clinical and Experimental Neuropsychology», 2025.

Clinical Linguistics: Analysis of Mapping Knowledge Domains in Past, Present and Future, PubMed Central, PMC9406678.

Hong S., Kim C., Lee S. H., Digital tools and biomarkers for cognitive screening: A systematic review, «Digital Health», 2026.

Perkins M. R., Howard S. J., Case Studies in Clinical Linguistics, Whurr, London, 1995.

Winterlight Labs, Winterlight Labs announces data demonstrating the sensitivity of speech-based digital measures in detecting longitudinal change in Alzheimer's disease in collaboration with Genentech, 2022.

Winterlight Labs, Clinical Research, pagina istituzionale, consultata nel 2026.

Parte II. Le cinque dimensioni

Capitolo 4 La densità lessicale

4.1 Una nozione del 1971

Il concetto di densità lessicale, la prima delle cinque dimensioni del modello e la più antica sul piano della tradizione di studi da cui proviene, fu introdotto da Jean Ure nel 1971 per descrivere la proporzione di parole lessicali, o parole di contenuto, rispetto al totale delle parole di un testo, nell'ambito di uno studio sulla differenziazione dei registri linguistici tra parlato e scritto.[40] La formula originaria di Ure calcola la densità lessicale come il rapporto percentuale tra il numero di elementi lessicali e il numero totale di parole del testo.[41] Michael Halliday, nel 1985, rivide il denominatore della formula, proponendo di calcolare la densità lessicale come rapporto tra il numero di elementi lessicali e il numero di proposizioni, non di parole, e introducendo una distinzione più netta tra elementi lessicali, che possono includere più di una parola quando formano un'unica unità di significato, ed elementi grammaticali.[42]

Il criterio distintivo tra parole lessicali e parole grammaticali riguarda la classe aperta o chiusa a cui la parola appartiene. Sono considerate parole lessicali i nomi, i verbi non ausiliari, gli aggettivi e la maggior parte degli avverbi, classi a cui la lingua può aggiungere nuovi membri con facilità. Sono considerate parole grammaticali gli articoli, le preposizioni, le congiunzioni, i pronomi e i verbi ausiliari o modali, classi chiuse a cui nuovi membri si aggiungono raramente.[43] Ure osservò inoltre che la densità lessicale tende a essere strutturalmente più bassa nel parlato spontaneo che nello scritto, un dato empirico rilevante per un modello come FOXP2, costruito specificamente sull'analisi del parlato spontaneo: un valore di densità lessicale che apparirebbe basso se giudicato con

40. J. Ure, Lexical density and register differentiation, in G. Perren, J. L. M. Trim (a cura di), Applications of Linguistics, Cambridge University Press, London, 1971, pp. 443-452.

41. Sulla formula originaria e sulla sua revisione, cfr. la voce Lexical density, in HandWiki, che riporta la formula di Ure, Ld = numero di elementi lessicali / numero totale di parole × 100, e la successiva revisione proposta da Halliday.

42. Sulla revisione hallidayana e sulla distinzione tra elementi lessicali ed elementi grammaticali, cfr. Lexical density, cit.; per la discussione originaria, M. A. K. Halliday, Spoken and Written Language, Deakin University Press, Geelong, 1985.

43. Sulla distinzione tra classi aperte, o parole di contenuto, e classi chiuse, o parole grammaticali, cfr. la ricostruzione in Lexical diversity and lexical density in speech and writing, «Lund Working Papers in Linguistics», che riprende la classificazione originaria di Ure.

gli standard dello scritto può essere del tutto atteso nel parlato, e la sua interpretazione deve tenere conto di questa differenza strutturale di registro.[44]

4.2 Il problema della lunghezza del campione

Il Capitolo 1 ha segnalato che gli indici di ricchezza lessicale disponibili in letteratura sono numerosi e tendono a covariare fortemente tra loro. Un secondo problema, distinto da quello della covarianza ma altrettanto rilevante per un modello applicato al parlato spontaneo, riguarda la sensibilità di molti di questi indici alla lunghezza del campione analizzato. L'indice più semplice, il rapporto type-token, cioè il rapporto tra il numero di parole diverse e il numero totale di parole di un testo, è stato definito da Covington e McFall una misura insoddisfacente della variazione lessicale proprio per questa ragione: il suo valore tende a diminuire in modo sistematico all'aumentare della lunghezza del testo, indipendentemente da un'effettiva variazione della ricchezza del vocabolario impiegato.[45]

Il problema è particolarmente rilevante per un modello costruito su campioni di parlato spontaneo, che sono tipicamente più brevi dei testi scritti impiegati nella maggior parte degli studi di linguistica dei corpora, spesso dell'ordine di poche centinaia di parole o meno. Un confronto sistematico tra più indici di diversità lessicale, condotto su indici quali TTR, Maas, MTLD, D e HD-D, ha mostrato che il TTR grezzo presenta una correlazione forte con la lunghezza del testo, pari a 0,81, mentre MTLD presenta una correlazione trascurabile, pari a -0,02.[46] Uno studio dedicato specificamente a testi brevi, tra 50 e 200 parole, prodotti da parlanti non nativi, ha confermato che MTLD resta relativamente stabile anche al di sotto della soglia delle 100 parole, a differenza del TTR e di indici correlati.[47] Una rassegna successiva, condotta su un ampio corpus di saggi in

44. Ure riporta che la maggioranza dei testi parlati del suo corpus presenta una densità lessicale inferiore al 40%, mentre la maggioranza dei testi scritti la presenta pari o superiore al 40%; cfr. Lexical density and register differentiation, cit., discussione ripresa in Lexical diversity and lexical density in speech and writing, cit.

45. M. A. Covington, J. D. McFall, Cutting the Gordian Knot: The Moving-Average Type-Token Ratio (MATTR), «Journal of Quantitative Linguistics», 17, 2, 2010, pp. 94-100, citato in A refined and concise model of indices for quantitatively measuring lexical richness, «English for Specific Purposes» o rivista analoga, che riporta la definizione di TTR come misura insoddisfacente della variazione lessicale a causa della sua sensibilità alla lunghezza del testo.

46. P. M. McCarthy, S. Jarvis, MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment, «Behavior Research Methods», 42, 2, 2010, pp. 381-392. Sui coefficienti di correlazione con la lunghezza del testo riportati per le diverse misure, cfr. la sintesi in Effects of text length on lexical diversity measures, «Assessing Writing», 2012.

seconda lingua, ha confermato che MATTR e MTLD restano gli indici più stabili al variare della lunghezza del testo tra quelli comunemente impiegati in letteratura.[48]

4.3 La formula adottata dal modello

Sulla base di queste evidenze, il modello FOXP2 misura la dimensione della densità lessicale attraverso due componenti distinte, poi combinate in un unico punteggio normalizzato. La prima componente riprende direttamente la formula di Ure, calcolando la proporzione di parole lessicali sul totale delle parole della trascrizione:

$$Ld = N_{lex} / N_{tot} \times 100$$

dove N_{lex} è il numero di parole lessicali, individuate secondo il criterio della classe aperta descritto nel paragrafo 4.1, e N_{tot} è il numero totale di parole della trascrizione.[49]

Un esempio minimo chiarisce il calcolo. Nell'enunciato «Sabato sono andato al mercato e ho comprato del pane fresco», di undici parole, sei sono lessicali e cinque grammaticali, come mostra la Figura 4.1: Ld = 6/11 × 100 ≈ 54,5. Il valore isolato non ha significato clinico al di fuori del confronto con il database normativo del modello; l'esempio serve solo a mostrare il funzionamento della formula.

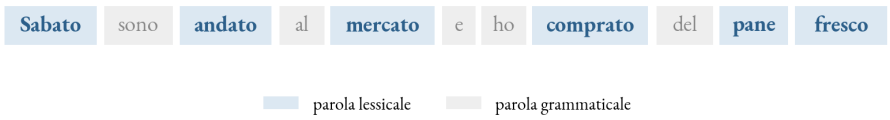

Figura 4.1. Classificazione lessicale/grammaticale dell'enunciato d'esempio. Le parole in blu sono lessicali (Ure 1971), quelle in grigio grammaticali.

47. Sulla stabilità di MTLD anche per testi compresi tra 50 e 200 parole, cfr. Koizumi, 2012, richiamato in Investigating minimum text lengths for lexical diversity indices, «Journal of Second Language Writing», 49, 2020, che conferma MATTR e le due versioni di MTLD come gli indici più stabili rispetto alla lunghezza del testo su un corpus di 4542 saggi argomentativi.

48. Investigating minimum text lengths for lexical diversity indices, cit.

49. La formula qui adottata corrisponde alla definizione originaria di Ure con denominatore sul totale delle parole, non sulla revisione hallidayana basata sulle proposizioni; cfr. la discussione delle due varianti in Exploring Automated Essay Scoring for Nonnative English Speakers, che riporta la stessa formula nella forma $N_{lex}/N \times 100$.

La seconda componente misura la diversità del vocabolario impiegato attraverso MTLD, calcolata come lunghezza media delle sequenze di parole necessarie perché il rapporto type-token, calcolato in modo progressivo lungo il testo, raggiunga la soglia di 0,72, con il calcolo condotto sia in avanti sia all'indietro rispetto al testo e i due valori mediati.[50] Le due componenti, la densità in senso stretto e la diversità lessicale, sono concettualmente distinte, la prima riguarda la proporzione tra parole di contenuto e parole grammaticali, la seconda la varietà del vocabolario impiegato all'interno delle sole parole di contenuto, ma vengono trattate come sotto-componenti di un'unica dimensione del modello, coerentemente con la loro collocazione comune nel medesimo sottogruppo di caratteristiche lessicali nella letteratura sulla rilevazione automatica del decadimento cognitivo già richiamata nel Capitolo 1.[51] Il punteggio finale della dimensione è ottenuto normalizzando entrambe le componenti rispetto al database normativo del modello e mediandone i valori normalizzati.

4.4 L'identificazione automatica delle parole lessicali in italiano

L'applicazione della formula richiede che ogni parola della trascrizione venga classificata in modo automatico come lessicale o grammaticale. Questa classificazione si basa sull'annotazione grammaticale automatica della trascrizione, un passaggio della pipeline che qui va richiamato solo nei suoi aspetti rilevanti per la dimensione in esame. L'italiano pone alcuni problemi specifici che un sistema tarato su altre lingue non risolverebbe in modo corretto. I pronomi clitici, che in italiano si legano graficamente al verbo in costruzioni come «dirglielo» o «guardandola», vanno scomposti nei loro costituenti prima della classificazione, per evitare che un'unica forma grafica venga contata come una sola parola lessicale quando contiene in realtà un verbo e uno o più pronomi.[52] I verbi ausiliari e modali, «avere», «essere», «potere», «dovere», «volere» nelle loro funzioni di ausiliari o modali, sono esclusi dal conteggio delle parole lessicali anche quando la stessa forma può funzionare, in altri contesti sintattici, come verbo

50. Sulla definizione operativa di MTLD come lunghezza media delle sequenze di parole necessarie a raggiungere la soglia di TTR pari a 0,72, calcolata in entrambe le direzioni del testo, cfr. P. M. McCarthy, An Assessment of the Range and Usefulness of Lexical Diversity Measures and the Potential of the Measure of Textual Lexical Diversity (MTLD), tesi di dottorato, University of Memphis, 2005; McCarthy, Jarvis, MTLD, vocd-D, and HD-D, cit., p. 384.

51. Sulla collocazione di densità e diversità lessicale nel medesimo sottogruppo di caratteristiche, cfr. K. C. Fraser, J. A. Meltzer, F. Rudzicz, Linguistic features identify Alzheimer's disease in narrative speech, «Journal of Alzheimer's Disease», 49, 2, 2016, pp. 407-422.

pieno dotato di significato lessicale autonomo: la classificazione richiede quindi un'analisi contestuale della funzione sintattica, non la sola forma della parola.

4.5 Perché non un solo indice tra i molti disponibili

Il Capitolo 1 ha ricordato che indici quali il rapporto type-token, MTLD, HD-D e l'indice di Honoré tendono a covariare fortemente tra loro, un dato confermato anche su un campione di lingua slovacca applicato a parlato patologico.[53] La scelta di MTLD tra le alternative disponibili non è quindi arbitraria: McCarthy e Jarvis, nel loro studio di validazione del 2010, hanno mostrato che MTLD è l'unico tra gli indici esaminati a non variare in funzione della lunghezza del testo, un requisito decisivo per un modello che deve trattare in modo comparabile campioni di parlato spontaneo di lunghezza disomogenea.[54] Gli stessi autori raccomandano comunque l'uso congiunto di più indici complementari nella ricerca in generale; il modello se ne discosta per la ragione già esposta nel Capitolo 1, la necessità di ridurre la ridondanza tra indici fortemente correlati a un'unica misura rappresentativa per dimensione, a costo di rinunciare alla convergenza di più misure che uno studio non vincolato da questo criterio potrebbe permettersi.

La domanda di brevetto indica il rapporto type-token come esempio di misura per la ricchezza lessicale dell'indicatore Vocabolario, senza vincolare il metodo a questo specifico indice. Il modello FOXP2 aggiorna questa esemplificazione con MTLD per le ragioni di robustezza rispetto alla lunghezza del campione appena discusse.

4.6 Dalla misura al grado della scala «Vocabolario»

Il modello FOXP2 traduce ogni misura continua nel grado corrispondente di una scala di valutazione clinica a dieci punti, definita per questa dimensione dalla domanda

52. Il trattamento dei pronomi clitici come unità morfologicamente distinte dal verbo cui si legano graficamente è un problema standard nell'annotazione grammaticale automatica dell'italiano, affrontato dai principali sistemi di analisi morfosintattica per questa lingua; il modello risolve questi casi in fase di annotazione grammaticale automatica.

53. E. Zozuk et al., Relationship between language features extracted through NLP and clinically diagnosed Alzheimer's disease and mild cognitive impairment in Slovak, «Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring», 17, 3, 2025, articolo e70122.

54. McCarthy, Jarvis, MTLD, vocd-D, and HD-D, cit., che riportano MTLD come l'unico indice, tra quelli esaminati, a non presentare una variazione significativa in funzione della lunghezza del testo.

di brevetto sotto la voce «Vocabolario». Il punteggio finale normalizzato descritto nel paragrafo 4.3 viene collocato, rispetto al database normativo del modello, in una delle dieci fasce corrispondenti agli ancoraggi qualitativi della domanda di brevetto, in modo che il valore quantitativo prodotto dalla formula resti sempre accompagnato da una descrizione clinicamente leggibile, non da un numero isolato privo di interpretazione. La domanda di brevetto specifica inoltre che l'analisi stilometrica di supporto a questo indicatore viene condotta con Voyant Tools e con il pacchetto Stylo, strumenti che il modello impiega in fase di verifica accanto alla pipeline di annotazione grammaticale automatica qui presentata.

Tabella 4.1. Scala di valutazione dell'indicatore «Vocabolario».

Tabella 4.1. Scala di valutazione dell'indicatore «Vocabolario».

Grado	**Descrizione**
1	Uso estremamente limitato del vocabolario, difficoltà nell'articolare parole; presenza di gravi errori semantici.
2	Molte difficoltà nella scelta delle parole; vocabolario molto ridotto e numerosi errori di significato.
3	Vocabolario scarso, scarsa precisione lessicale; utilizzo di parole non appropriate al contesto.
4	Vocabolario basilare, presenza di errori e imprecisioni; qualche difficoltà nella scelta delle parole giuste.
5	Vocabolario sufficiente, errori semantici moderati; utilizzo accettabile di termini generici.
6	Buon controllo del vocabolario di base, con pochi errori; capacità di esprimere concetti semplici.
7	Vocabolario abbastanza ricco e preciso; rare imprecisioni semantiche.
8	Buona padronanza del vocabolario, scelta accurata delle parole; quasi nessun errore di significato.
9	Ottima padronanza lessicale, vocabolario variegato e preciso; nessun errore evidente.
10	Eccellente padronanza del vocabolario, uso perfetto e preciso di termini complessi e appropriati al contesto.

4.7 Rinvio alla validazione e ai limiti della dimensione

La rilevanza diagnostica della densità e della diversità lessicale nel parlato spontaneo è stata discussa nel Capitolo 1 con riferimento alla letteratura sulla malattia di Alzheimer, che riporta risultati di classificazione significativi ottenuti a partire da un piccolo insieme di predittori lessicali.[55] I limiti propri della dimensione comprendono la sensibilità residua di entrambe le componenti a fattori non patologici quali il livello di scolarizzazione e la familiarità del parlante con il compito linguistico richiesto, insieme ai limiti delle altre quattro dimensioni del modello. Il capitolo seguente tratta la seconda dimensione, la complessità sintattica.

Riferimenti bibliografici del Capitolo 4

A refined and concise model of indices for quantitatively measuring lexical richness, disponibile su ERIC, EJ1437591.

Covington M. A., McFall J. D., Cutting the Gordian Knot: The Moving-Average Type-Token Ratio (MATTR), «Journal of Quantitative Linguistics», 17, 2, 2010, pp. 94-100.

Effects of text length on lexical diversity measures: Using short texts with less than 200 tokens, «Assessing Writing», 2012.

Exploring Automated Essay Scoring for Nonnative English Speakers, arXiv: 1706.03335.

Fraser K. C., Meltzer J. A., Rudzicz F., Linguistic features identify Alzheimer's disease in narrative speech, «Journal of Alzheimer's Disease», 49, 2, 2016, pp. 407-422.

Halliday M. A. K., Spoken and Written Language, Deakin University Press, Geelong, 1985.

Investigating minimum text lengths for lexical diversity indices, «Journal of Second Language Writing», 49, 2020.

Lexical density, in HandWiki.

55. Zozuk et al., Relationship between language features, cit., che richiama i risultati di Eyigoz et al. su un tasso di classificazione del 76% ottenuto con i dieci predittori linguistici più rilevanti, tra cui predittori lessicali.

Lexical diversity and lexical density in speech and writing, «Lund Working Papers in Linguistics».

McCarthy P. M., An Assessment of the Range and Usefulness of Lexical Diversity Measures and the Potential of the Measure of Textual Lexical Diversity (MTLD), tesi di dottorato, University of Memphis, 2005.

McCarthy P. M., Jarvis S., MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment, «Behavior Research Methods», 42, 2, 2010, pp. 381-392.

Ure J., Lexical density and register differentiation, in G. Perren, J. L. M. Trim (a cura di), Applications of Linguistics, Cambridge University Press, London, 1971, pp. 443-452.

Zozuk E. et al., Relationship between language features extracted through NLP and clinically diagnosed Alzheimer's disease and mild cognitive impairment in Slovak, «Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring», 17, 3, 2025, articolo e70122.

Capitolo 5 La complessità sintattica

5.1 Il costo cognitivo della subordinazione

Includere una proposizione dentro un'altra ha un costo. La subordinazione è riconosciuta in letteratura come uno dei fenomeni più onerosi dal punto di vista dell'elaborazione cognitiva, sia negli studi sperimentali sulla comprensione del linguaggio sia negli studi clinici sulla produzione linguistica di parlanti anziani e afasici.[56] È a questo fenomeno che deve la propria definizione la seconda dimensione del modello, la complessità sintattica, distinta dalla ricchezza del vocabolario discussa nel capitolo precedente per il fatto di riguardare la struttura della frase, non le parole che la compongono. Diversi studi hanno impiegato il numero di proposizioni per frase come indice sintetico di questo fenomeno, sulla base dell'osservazione che i parlanti anziani tendono a produrre un numero minore di proposizioni subordinate per frase rispetto ai parlanti più giovani.[57]

La tradizione di studi computazionali su questo fenomeno risale al lavoro di Victor Yngve, che nel 1960 propose un modello della struttura sintattica basato sulla rappresentazione della frase come albero gerarchico a costituenti, e un punteggio, noto come profondità di Yngve, che assegna a ciascuna parola un valore proporzionale al numero di rami sinistri non ancora completati nell'albero al momento in cui la parola viene raggiunta in una visita top-down da sinistra a destra.[58] Yngve propose questo punteggio come approssimazione del carico di memoria a breve termine richiesto per elaborare la frase, un'ipotesi nota come ipotesi della profondità.

56. Sulla subordinazione come fenomeno cognitivamente oneroso, corroborato da studi sperimentali sulla comprensione e da studi clinici sulla produzione in soggetti anziani e afasici agrammatici, cfr. la rassegna in Automated Measures of Syntactic Complexity in Natural Speech Production: Older and Younger Adults as a Case Study, «Journal of Speech, Language, and Hearing Research», 2023, disponibile su PubMed Central, PMC12510242.

57. Sull'impiego del numero di proposizioni per frase come indice di complessità sintattica e sulla sua riduzione nei parlanti anziani, cfr. la stessa fonte, che richiama tra gli altri Cheung, Kemper, 1992.

58. V. H. Yngve, A model and an hypothesis for language structure, «Proceedings of the American Philosophical Society», 104, 1960, pp. 444-466.

5.2 Un'ipotesi psicologica contestata, una misura empiricamente utile

La letteratura successiva ha reso necessaria una distinzione, che il modello FOXP2 recepisce integralmente. L'ipotesi della profondità, nella sua versione originaria di modello del carico di memoria a breve termine, non ha trovato conferma sperimentale solida: due esperimenti sulla ritenzione di frasi condotti negli anni immediatamente successivi non hanno trovato la relazione attesa tra profondità sintattica e capacità di richiamo mnestico, in un caso osservando perfino una relazione di segno opposto a quella prevista.[59] Un terzo studio, che ha confrontato la profondità sintattica media con la densità lessicale come predittori della ritenzione di frasi, ha trovato che la prima non è associata alla ritenzione, mentre la seconda lo è, ponendo ulteriori dubbi sul valore generale dell'ipotesi di Yngve come modello psicologico del carico di memoria.[60]

Questo esito non ha tuttavia impedito che il punteggio di Yngve, calcolato in modo automatico su ampi corpora, si dimostrasse empiricamente utile come indicatore correlato all'età e allo stato cognitivo del parlante, indipendentemente dalla validità del meccanismo psicologico che Yngve aveva originariamente proposto per spiegarlo. Il punteggio di Yngve calcolato automaticamente risulta ridotto nei parlanti anziani rispetto ai parlanti più giovani e nei soggetti con demenza rispetto ai controlli sani, un risultato replicato da più gruppi di ricerca indipendenti.[61] Il modello FOXP2 tratta quindi il punteggio di Yngve, e le misure imparentate discusse nel paragrafo seguente, come indicatori empiricamente validati di un fenomeno linguistico osservabile, non come conferma di una specifica teoria del carico di memoria a breve termine. Questa distinzione, tra validità di un costrutto come modello psicologico e validità della sua misura come indicatore diagnostico, va tenuta ferma anche nella lettura dei risultati del modello. È coerente con il criterio del Capitolo 1: conta la rilevanza diagnostica accertata di ogni dimensione, non la correttezza della teoria che la spiega.

59. Cfr. i due esperimenti sulla ritenzione di frasi riportati in Sentence retention and the depth hypothesis, «Journal of Verbal Learning and Verbal Behavior», 1969, che non trovano supporto per l'ipotesi della profondità e osservano in un caso una relazione di segno opposto a quella prevista, con le frasi più profonde che mostrano un richiamo mnestico superiore.

60. Lexical density and phrase structure depth as variables in sentence retention, «Journal of Verbal Learning and Verbal Behavior», 1969, che riporta l'assenza di associazione tra profondità sintattica media e ritenzione, a fronte di un'associazione inversa significativa con la densità lessicale.

61. Sulla riduzione del punteggio di Yngve nei parlanti anziani e nei soggetti con demenza, cfr. la rassegna delle fonti in Automated Measures of Syntactic Complexity, cit., che richiama Cheung, Kemper, 1992; Kemper, Rash, 1988; Kemper et al., 2001 per l'invecchiamento, e Fraser et al., 2015; Pakhomov et al., 2011; Roark et al., 2011 per la demenza.

5.3 La formula adottata dal modello

Il modello FOXP2 non calcola il punteggio di Yngve nella sua formulazione originaria, che richiede un albero sintattico a costituenti prodotto da un parser calibrato sulla grammatica a struttura di frase di una lingua specifica, una risorsa meno disponibile e meno robusta per l'italiano rispetto agli strumenti di analisi a dipendenze.[62] Il modello adotta invece la distanza di dipendenza media, una misura ricavabile direttamente da un albero sintattico a dipendenze, oggi lo standard di riferimento per l'annotazione automatica dell'italiano attraverso gli schemi di annotazione universali.[63]

La distanza di dipendenza tra due parole legate da una relazione sintattica è definita come il numero di parole intermedie tra la testa della relazione e il suo dipendente, calcolato sulla base della posizione lineare delle parole nella frase.[64] La distanza di dipendenza media di una frase si ottiene sommando le distanze di tutte le relazioni di dipendenza della frase e dividendo per il numero di relazioni:

$$\mathrm{MDD} = (1 / (n - 1)) \times \Sigma \left|\mathrm{posizione}(\mathrm{testa}_i) - \mathrm{posizione}(\mathrm{dipendente}_i)\right|$$

dove n è il numero di parole della frase e la sommatoria è estesa a tutte le $n - 1$ relazioni di dipendenza dell'albero.[65]

Un esempio minimo illustra il calcolo. Nella frase «Il gatto nero dorme sul divano», di sei parole, l'analisi a dipendenze individua cinque relazioni, mostrate nella Figura 5.1: la somma delle distanze è 7, e MDD = 7/5 = 1,4. Anche qui il valore isolato non ha

62. Sulla distinzione tra parsing a costituenti, che richiede categorie sintagmatiche quali sintagma nominale e sintagma verbale, e parsing a dipendenze, in cui ogni nodo dell'albero è trattato come nodo terminale legato agli altri da relazioni di dipendenza, cfr. A Review of Automated Speech and Language Features for Dementia Detection, arXiv:1906.01157.

63. Sulla distanza di dipendenza come misura ricavabile da un albero a dipendenze e sul suo impiego come indicatore di complessità sintattica, cfr. R. Ferrer-i-Cancho, discusso in The optimality of syntactic dependency distances, arXiv:2007.15342; per l'introduzione del concetto di distanza di dipendenza, R. Hudson, Word Meaning, Routledge, London, 1995, richiamato in Dependency Distance as a Metric of Language Comprehension Difficulty.

64. Sulla definizione operativa della distanza di dipendenza come distanza tra le posizioni lineari di testa e dipendente in una relazione sintattica, cfr. la rassegna in The distribution of syntactic dependency distances, arXiv:2211.14620.

65. La formula corrisponde alla definizione standard di distanza di dipendenza media adottata nella letteratura sulla linguistica computazionale delle dipendenze, cfr. The distribution of syntactic dependency distances, cit.

significato clinico al di fuori della normalizzazione rispetto al database normativo del modello.

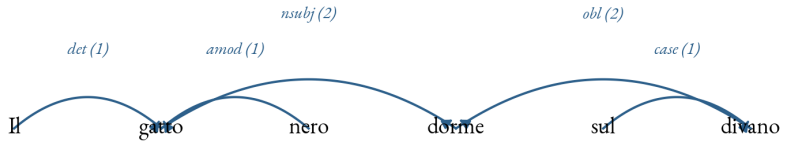

Figura 5.1. Albero delle dipendenze della frase d'esempio, con la distanza di ciascuna relazione tra parentesi.

Una distanza di dipendenza media più alta indica una struttura sintattica in cui le parole sintatticamente collegate tendono a essere separate da un maggior numero di parole intermedie, una configurazione che la letteratura associa a un maggiore carico di elaborazione, coerentemente con il principio di minimizzazione della distanza di dipendenza osservato in modo pressoché universale nelle lingue naturali.[66] A questa prima componente il modello affianca un indice di subordinazione, calcolato come proporzione delle proposizioni subordinate sul totale delle proposizioni identificate nella trascrizione, in continuità diretta con la tradizione di studi sulla subordinazione richiamata nel paragrafo 5.1.

5.4 L'estrazione automatica per l'italiano

Il calcolo della distanza di dipendenza media e dell'indice di subordinazione richiede un'analisi sintattica a dipendenze automatica della trascrizione, un ulteriore passaggio della pipeline computazionale del modello. L'applicazione al parlato spontaneo dell'italiano pone alcuni problemi che un'applicazione a testo scritto non incontrerebbe nella stessa misura. Le disfluenze tipiche del parlato spontaneo, quali le false partenze, le autocorrezioni e le ripetizioni, possono produrre strutture sintattiche incomplete o anomale che un parser addestrato su testo scritto rischia di analizzare in modo errato, con effetti diretti sul calcolo sia della distanza di dipendenza sia dell'identificazione dei confini di proposizione.[67] L'identificazione dei confini di proposizione nel parlato spontaneo,

66. Sul principio di minimizzazione della distanza di dipendenza come tendenza pressoché universale delle lingue naturali, cfr. The distribution of syntactic dependency distances, cit., che richiama tra gli altri Ferrer-i-Cancho, 2004; Liu, 2007; Futrell et al., 2015.

necessaria per il calcolo dell'indice di subordinazione, richiede inoltre una segmentazione della trascrizione in unità enunciative che non sempre coincidono con le frasi ortografiche della lingua scritta, un problema affrontato attraverso l'annotazione prosodica delle pause prodotta nella fase di trascrizione automatica.

5.5 Perché non le sole misure a costituenti

La letteratura sulla rilevazione automatica del decadimento cognitivo riporta che il punteggio di Yngve e il punteggio correlato proposto da Frazier hanno funzionato bene nel discriminare soggetti con lieve compromissione della memoria da controlli sani.[68] Uno studio che ha confrontato otto misure automatiche di complessità sintattica, incluse il punteggio di Yngve, il punteggio di Frazier e la distanza di dipendenza media, per discriminare parlanti anziani da parlanti giovani a partire da un compito di descrizione di immagine, ha inoltre mostrato un elevato grado di accordo tra un sistema computazionale basato sul parser sintattico di Stanford e la valutazione manuale di un linguista esperto, con una differenza media minima proprio per la misura della distanza di dipendenza sintattica.[69] La scelta del modello di adottare la distanza di dipendenza media anziché le misure a costituenti non nasce quindi da un giudizio negativo sulla loro capacità discriminativa, empiricamente buona, ma dalla maggiore robustezza e disponibilità degli strumenti di analisi a dipendenze per l'italiano rispetto ai parser a costituenti calibrati su questa lingua, in coerenza con il criterio di calcolabilità automatica discusso nel Capitolo 1.

67. La gestione delle disfluenze del parlato spontaneo come fattore di rischio per l'accuratezza dell'analisi sintattica automatica è un problema riconosciuto nella letteratura sull'elaborazione del linguaggio parlato; questo modello affronta il problema in una fase dedicata di normalizzazione della trascrizione, precedente all'analisi sintattica.

68. Cfr. Roark et al., 2007, richiamato in The Fewer Splits are Better: Deconstructing Readability in Sentence Splitting, arXiv:2302.00937, che riporta buoni risultati discriminativi sia per il punteggio di Yngve sia per il punteggio di Frazier in soggetti con lieve compromissione della memoria.

69. Sull'elevato accordo tra valutazione automatica e manuale, con una differenza media particolarmente contenuta per la distanza di dipendenza sintattica, pari a 0,02 con deviazione standard 0,17, cfr. Computerized assessment of syntactic complexity in Alzheimer's disease: a case study of Iris Murdoch's writing, «Behavior Research Methods», 2010.

5.6 Dalla misura al grado della scala «Sintassi»

La distanza di dipendenza media e l'indice di subordinazione descritti nel paragrafo 5.3 vengono tradotti nel grado corrispondente della scala di valutazione clinica a dieci punti che la domanda di brevetto definisce per questa dimensione sotto la voce «Sintassi», secondo l'architettura del Capitolo 1. Anche in questo caso il valore numerico prodotto dalla formula del paragrafo 5.3 non sostituisce la scala della domanda di brevetto ma la alimenta, fornendo il dato oggettivo su cui si basa la collocazione del campione in una delle dieci fasce descrittive, secondo la stessa logica di normalizzazione rispetto al database normativo descritta nel Capitolo 1.

Tabella 5.1. Scala di valutazione dell'indicatore «Sintassi».

Tabella 5.1. Scala di valutazione dell'indicatore «Sintassi».

Grado	Descrizione
1	Sintassi praticamente assente, frasi frammentarie e incoerenti.
2	Costruzione delle frasi molto povera, struttura sintattica disordinata o incompleta.
3	Frasi molto semplici e ripetitive, con numerosi errori strutturali.
4	Struttura delle frasi limitata, errori frequenti nella costruzione di frasi complesse.
5	Sintassi sufficiente, frasi complete ma semplici; errori occasionali nelle subordinate.
6	Buona gestione della sintassi di base, con poche difficoltà nelle frasi complesse.
7	Buona varietà nelle costruzioni sintattiche; errori minimi.
8	Ottima padronanza della sintassi complessa, con una buona gestione delle subordinate.
9	Sintassi quasi perfetta, frasi ben costruite con pochi errori.
10	Eccellente padronanza della sintassi, costruzione complessa e articolata senza errori.

5.7 Rinvio alla validazione e ai limiti della dimensione

La rilevanza diagnostica della complessità sintattica nel parlato spontaneo è stata discussa nel Capitolo 1 con riferimento alla letteratura sulla malattia di Alzheimer, che riporta risultati non sempre univoci a seconda dello stadio della patologia considerato. I limiti propri della dimensione comprendono la sensibilità della distanza di dipendenza

media alla lunghezza della frase e l'effetto delle disfluenze del parlato spontaneo sull'accuratezza del parser automatico. Il capitolo seguente tratta la terza dimensione, la coesione discorsiva.

Riferimenti bibliografici del Capitolo 5

A Review of Automated Speech and Language Features for Dementia Detection, arXiv:1906.01157.

Automated Measures of Syntactic Complexity in Natural Speech Production: Older and Younger Adults as a Case Study, «Journal of Speech, Language, and Hearing Research», 2023, PubMed Central, PMC12510242.

Computerized assessment of syntactic complexity in Alzheimer's disease: a case study of Iris Murdoch's writing, «Behavior Research Methods», 2010.

Dependency Distance as a Metric of Language Comprehension Difficulty.

Lexical density and phrase structure depth as variables in sentence retention, «Journal of Verbal Learning and Verbal Behavior», 1969.

Hudson R., Word Meaning, Routledge, London, 1995.

Sentence retention and the depth hypothesis, «Journal of Verbal Learning and Verbal Behavior», 1969.

The distribution of syntactic dependency distances, arXiv:2211.14620.

The Fewer Splits are Better: Deconstructing Readability in Sentence Splitting, arXiv: 2302.00937.

The optimality of syntactic dependency distances, arXiv:2007.15342.

Yngve V. H., A model and an hypothesis for language structure, «Proceedings of the American Philosophical Society», 104, 1960, pp. 444-466.

Capitolo 6 La coesione discorsiva

6.1 L'unità del testo secondo Halliday e Hasan

Un testo si presenta come un insieme unitario, non come una sequenza di frasi indipendenti, grazie a meccanismi linguistici espliciti che Michael Halliday e Ruqaiya Hasan catalogarono nel 1976 in cinque categorie: il riferimento, la sostituzione, l'ellissi, la connessione e la coesione lessicale.[70] È a questo repertorio che deve la propria definizione la coesione discorsiva, terza dimensione del modello, corrispondente all'indicatore che la domanda di brevetto definisce come «Coesione e connessione testuale». I primi quattro meccanismi vengono classificati come coesione grammaticale, che riguarda la struttura sintattica con cui gli elementi del testo si richiamano a vicenda, in particolare attraverso pronomi e altri elementi anaforici; il quinto meccanismo, la coesione lessicale, riguarda invece la relazione semantica tra le parole, realizzata attraverso la ripetizione, la sinonimia o termini semanticamente affini.[71]

La coesione discorsiva va tenuta distinta dalla pertinenza e coerenza testuale, trattata nel Capitolo 7, come già anticipato nel Capitolo 1. La coesione riguarda i legami linguistici espliciti tra le parti del testo, osservabili anche isolando singole coppie di frasi contigue; la pertinenza e la coerenza riguardano invece la tenuta tematica del discorso nel suo complesso rispetto al compito assegnato. Un discorso può essere ben coeso, con pronomi correttamente risolti e connettivi appropriati, e tuttavia poco pertinente rispetto alla consegna, così come un discorso pertinente può presentare legami coesivi deboli tra le frasi che lo compongono. La letteratura clinica distingue inoltre la coesione locale, tra frasi contigue, dalla coesione globale, tra parti distanti del discorso, una distinzione già richiamata nel Capitolo 1 a proposito degli studi sulla malattia di Alzheimer e sul deterioramento cognitivo lieve, che riportano una riduzione della coerenza semantica globale più marcata di quella locale.[72]

70. M. A. K. Halliday, R. Hasan, Cohesion in English, Longman, London, 1976. Sui cinque meccanismi di coesione, riferimento, sostituzione, ellissi, connessione e coesione lessicale, cfr. la sintesi in Discourse Cohesion Evaluation for Document-Level Neural Machine Translation, arXiv:2208.09118.

71. Sulla distinzione tra coesione grammaticale, che comprende riferimento, sostituzione, ellissi e connessione, e coesione lessicale, cfr. Quantitative Discourse Cohesion Analysis of Scientific Scholarly Texts using Multilayer Networks, arXiv:2205.07532.

6.2 Dalla teoria qualitativa alla misura automatica

Il modello di Halliday e Hasan nasce come strumento di analisi qualitativa, applicato manualmente da un analista che identifica e classifica ciascuna occorrenza dei cinque meccanismi coesivi in un testo. La sua automatizzazione ha richiesto lo sviluppo di strumenti computazionali distinti, il più influente dei quali è Coh-Metrix, sviluppato da Arthur Graesser, Danielle McNamara e colleghi, che analizza il testo su oltre duecento misure di coesione e linguaggio, utilizzando tra l'altro l'analisi semantica latente per stimare la sovrapposizione di significato tra frasi adiacenti.[73] L'analisi semantica latente, introdotta da Landauer, Foltz e Laham, rappresenta ogni parola e ogni frase come un vettore in uno spazio semantico ad alta dimensionalità, costruito a partire da un ampio corpus di addestramento, e stima la coesione locale tra due frasi come similarità coseno tra i rispettivi vettori.[74] Uno strumento successivo, TAACO, ha reso disponibile gratuitamente un'analisi automatica della coesione locale, globale e testuale basata su un insieme più ampio di indici lessicali e semantici, applicabile senza le limitazioni di accesso di Coh-Metrix.[75]

6.3 La formula adottata dal modello

Il modello FOXP2 misura la coesione discorsiva attraverso due componenti, in continuità con l'impostazione a due componenti già adottata per la densità lessicale nel Capitolo 4. La prima componente misura la coesione semantica locale, calcolata come similarità coseno media tra le rappresentazioni vettoriali di frasi consecutive della trascrizione:

72. Cfr. Comparing global and local semantic coherence of spontaneous speech in persons with Alzheimer's disease and healthy controls, «Applied Corpus Linguistics», 3, 3, 2023, articolo 100064; B. S. Kim, Y. B. Kim, H. Kim, Discourse measures to differentiate between mild cognitive impairment and healthy aging, «Frontiers in Aging Neuroscience»; cfr. anche A. Bertola et al., Exploring the Use of Cohesive Devices in Dementia, in Atti CLiC-it 2024.

73. A. C. Graesser, D. S. McNamara, M. M. Louwerse, Z. Cai, Coh-Metrix: Analysis of text on cohesion and language, «Behavior Research Methods, Instruments, & Computers», 36, 2004, pp. 193-202.

74. Sull'impiego dell'analisi semantica latente per calcolare la coesione semantica tra frasi adiacenti, cfr. T. K. Landauer, P. W. Foltz, D. Laham, An Introduction to Latent Semantic Analysis, «Discourse Processes», 25, 2-3, 1998, pp. 259-284, tecnica ripresa e descritta anche in Analyzing the Cohesion of English Text and Discourse with Coh-Metrix, che ne illustra l'uso per computerizzare la coesione semantica tra frasi adiacenti.

75. S. A. Crossley, K. Kyle, D. S. McNamara, The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion, «Behavior Research Methods», 48, 2016, pp. 1227-1237.

$$CL = (1 / (k - 1)) \times \Sigma \cos(v_i, v_{i+1})$$

dove k è il numero di unità enunciative identificate nella trascrizione, v_i è la rappresentazione vettoriale semantica dell'unità i-esima, e la sommatoria è estesa a tutte le $k - 1$ coppie di unità consecutive.[76] A differenza dell'implementazione originaria di Coh-Metrix, che impiegava l'analisi semantica latente, il modello utilizza rappresentazioni vettoriali del significato addestrate su corpora dell'italiano, un aggiornamento tecnico che non altera la logica di fondo della misura, la stima della sovrapposizione semantica tra unità di discorso contigue, ma ne migliora l'accuratezza per questa lingua.

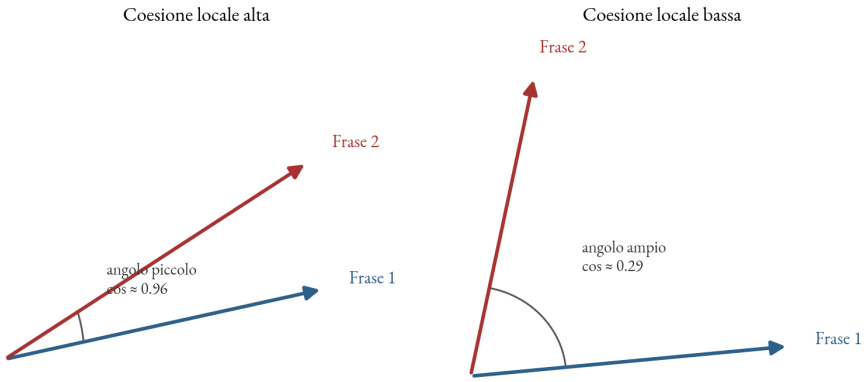

Figura 6.1. Rappresentazione schematica della coesione semantica locale come similarità coseno tra i vettori di due unità enunciative consecutive: un angolo piccolo tra i vettori corrisponde a una coesione alta, un angolo ampio a una coesione bassa.

La seconda componente misura la densità referenziale, calcolata come proporzione degli elementi anaforici della trascrizione, principalmente pronomi e forme di ellissi del soggetto, per cui il sistema individua in modo automatico un antecedente esplicito entro i tre enunciati precedenti, sul totale degli elementi anaforici individuati:

$$DR = N_{ref.risolti} / N_{ref.totali}$$

Questa seconda componente opera un'operazionalizzazione ristretta del meccanismo di riferimento descritto da Halliday e Hasan, limitata agli elementi per cui la risoluzione automatica è tecnicamente affidabile, secondo il criterio di calcolabilità automatica

76. La formula riprende l'impostazione classica della coesione semantica locale via similarità coseno tra rappresentazioni vettoriali di frasi adiacenti, descritta in Landauer, Foltz, Laham, cit., aggiornata nella sua implementazione computazionale con modelli di rappresentazione vettoriale del significato successivi all'analisi semantica latente.

discusso nel Capitolo 1; non include, allo stato attuale, i meccanismi di sostituzione ed ellissi diversi da quella del soggetto, né la coesione lessicale in senso stretto, la cui rilevazione automatica affidabile per l'italiano parlato spontaneo resta un problema aperto.

6.4 L'estrazione automatica per l'italiano

L'italiano pone per questa dimensione un problema strutturale che una pipeline tarata su altre lingue non affronterebbe nello stesso modo. È infatti una lingua a soggetto nullo, in cui il pronome soggetto viene comunemente omesso quando recuperabile dalla flessione verbale o dal contesto, una costruzione linguisticamente corretta e frequentissima nel parlato spontaneo che tuttavia non deve essere confusa, in fase di calcolo della densità referenziale, con un'ellissi anomala o con un'interruzione del filo referenziale.[77] Il sistema deve quindi riconoscere il soggetto nullo come un caso regolare di coesione referenziale realizzata morfologicamente, non lessicalmente, e trattarlo di conseguenza nel computo della densità referenziale, includendolo tra le forme anaforiche risolte quando il referente è identificabile dal contesto verbale precedente.

6.5 Dalla misura al grado della scala «Coesione e connessione testuale»

Le due componenti descritte nel paragrafo 6.3 vengono tradotte nel grado corrispondente della scala di valutazione clinica a dieci punti che la domanda di brevetto definisce per questa dimensione, secondo l'architettura del Capitolo 1. Il valore composito ottenuto dalla coesione semantica locale e dalla densità referenziale, normalizzato rispetto al database normativo del modello, colloca il campione in una delle dieci fasce descrittive della domanda di brevetto, secondo la stessa logica già illustrata per le dimensioni trattate nei Capitoli 4 e 5.

Tabella 6.1. Scala di valutazione dell'indicatore «Coesione e connessione testuale».

Tabella 6.1. Scala di valutazione dell'indicatore «Coesione e connessione testuale».

77. Il trattamento del soggetto nullo come fenomeno grammaticalmente regolare dell'italiano, da non computare come anomalia coesiva, richiede una fase dedicata di identificazione morfosintattica del soggetto implicito a partire dalla flessione verbale.

Grado	Descrizione
1	Assenza di coesione, frasi completamente slegate e prive di connessioni logiche.
2	Connessioni minime tra le frasi, uso scarso o inappropriato di congiunzioni.
3	Coesione debole, frasi poco legate tra loro; connettivi usati in modo limitato o errato.
4	Frasi collegabili ma con connessioni deboli; uso minimo dei connettivi.
5	Coesione sufficiente, uso basico dei connettivi; testo comprensibile ma con qualche salto logico.
6	Connessione testuale generalmente buona, uso adeguato di connettivi ma con qualche incoerenza.
7	Buona coesione, frasi ben collegate; connessioni logiche presenti e ben gestite.
8	Connessioni solide tra frasi, con uso fluido e vario di congiunzioni e altri connettivi.
9	Testo molto coeso, frasi perfettamente connesse; connessioni logiche efficaci e fluide.
10	Eccellente coesione e connessione, uso magistrale dei connettivi; nessuna incoerenza logica.

6.6 Rinvio alla validazione e ai limiti della dimensione

La rilevanza diagnostica della coesione discorsiva nel parlato spontaneo è stata discussa nel Capitolo 1 con riferimento alla letteratura sulla malattia di Alzheimer e sul deterioramento cognitivo lieve. I limiti propri della dimensione comprendono la copertura ancora parziale dei meccanismi coesivi previsti dal modello teorico di Halliday e Hasan e la dipendenza della coesione semantica locale dalla qualità delle rappresentazioni vettoriali impiegate. Il capitolo seguente tratta la quarta dimensione, la pertinenza e coerenza testuale, di cui il paragrafo 6.1 ha già anticipato il rapporto con la dimensione qui trattata.

Riferimenti bibliografici del Capitolo 6

Analyzing the Cohesion of English Text and Discourse with Coh-Metrix.

Bertola A. et al., Exploring the Use of Cohesive Devices in Dementia, in Atti della decima conferenza italiana di linguistica computazionale (CLiC-it 2024).

Comparing global and local semantic coherence of spontaneous speech in persons with Alzheimer's disease and healthy controls, «Applied Corpus Linguistics», 3, 3, 2023, articolo 100064.

Crossley S. A., Kyle K., McNamara D. S., The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion, «Behavior Research Methods», 48, 2016, pp. 1227-1237.

Discourse Cohesion Evaluation for Document-Level Neural Machine Translation, arXiv:2208.09118.

Graesser A. C., McNamara D. S., Louwerse M. M., Cai Z., Coh-Metrix: Analysis of text on cohesion and language, «Behavior Research Methods, Instruments, & Computers», 36, 2004, pp. 193-202.

Halliday M. A. K., Hasan R., Cohesion in English, Longman, London, 1976.

Landauer T. K., Foltz P. W., Laham D., An Introduction to Latent Semantic Analysis, «Discourse Processes», 25, 2-3, 1998, pp. 259-284.

Quantitative Discourse Cohesion Analysis of Scientific Scholarly Texts using Multilayer Networks, arXiv:2205.07532.

Capitolo 7 La pertinenza e coerenza testuale

7.1 Quando un discorso coeso non è pertinente

Un discorso può essere perfettamente coeso, con ogni pronome correttamente risolto, e tuttavia fuori tema, oppure logicamente contraddittorio: è questo lo spazio proprio della quarta dimensione del modello, la pertinenza e coerenza testuale, corrispondente all'indicatore omonimo della domanda di brevetto. La dimensione riprende un costrutto già introdotto nel Capitolo 1, la verbosità fuori tema descritta da Arbuckle e Gold, distinguendolo dalla coesione discorsiva trattata nel Capitolo 6: la coesione riguarda i legami linguistici tra le parti del testo, la pertinenza e coerenza riguardano la tenuta tematica del discorso rispetto al compito assegnato e la sua consistenza logica interna.[78] Questa dimensione, più delle altre quattro, riunisce sotto un'unica etichetta due fenomeni linguistici distinguibili sul piano teorico. Il primo è la pertinenza in senso stretto, l'aderenza del contenuto del discorso al compito richiesto, il cui opposto è la tangenzialità descritta nel paragrafo 1.2. Il secondo è la coerenza logica, l'assenza di contraddizioni interne tra le affermazioni successive del parlante, un fenomeno distinto dalla coesione linguistica perché riguarda il contenuto proposizionale del discorso, non la sua forma superficiale: un discorso può essere coeso, con pronomi correttamente risolti, e tuttavia logicamente incoerente, se le proposizioni che lo compongono si contraddicono a vicenda.

7.2 Dalla teoria qualitativa alla misura automatica

La rilevazione automatica della pertinenza tematica ha una tradizione consolidata nella valutazione automatica di elaborati scritti, dove il problema si presenta come rilevazione del fuori tema rispetto a una consegna data. Louis e Higgins hanno proposto un metodo non supervisionato che stima l'aderenza al tema attraverso la similarità semantica tra il testo prodotto e la consegna, senza richiedere elaborati già annotati come esempio di addestramento.[79] Un problema analogo, applicato specificamente al parlato spontaneo anziché al testo scritto, è stato affrontato da Malinin e colleghi nel contesto

78. T. Y. Arbuckle, D. P. Gold, Aging, inhibition, and verbosity, «Journal of Gerontology», 48, 5, 1993, pp. P225-P232.

79. A. Louis, D. Higgins, Off-topic essay detection using short prompt texts, in Proceedings of the NAACL HLT 2010 Fifth Workshop on Innovative Use of NLP for Building Educational Applications, 2010, pp. 92-95.

della valutazione automatica dell'inglese parlato, con un sistema di rilevazione delle risposte fuori tema calibrato su produzioni orali spontanee piuttosto che su elaborati scritti.[80]

Per la coerenza logica, la tradizione computazionale di riferimento è il modello a griglia di entità proposto da Barzilay e Lapata, che rappresenta il discorso come una sequenza di transizioni tra i ruoli grammaticali ricoperti da ciascuna entità menzionata nel testo, da frase a frase, distinguendo transizioni regolari, in cui un'entità mantiene un ruolo di rilievo sintattico costante, da transizioni brusche, in cui il ruolo dell'entità cambia in modo marcato o l'entità scompare senza continuità apparente con il discorso successivo.[81] Il modello, originariamente sviluppato per la valutazione automatica di elaborati scritti e per compiti di ordinamento delle frasi, è stato ripreso in numerosi lavori successivi come misura generale di coerenza applicabile a diversi generi testuali.

7.3 La formula adottata dal modello

Il modello FOXP2 misura la pertinenza e coerenza testuale attraverso due componenti, corrispondenti ai due fenomeni distinti nel paragrafo 7.1. La prima componente misura la pertinenza tematica come similarità coseno tra la rappresentazione vettoriale complessiva della trascrizione e la rappresentazione vettoriale della consegna assegnata al parlante:

$$PT = \cos(v_{trascrizione}, v_{consegna})$$

dove $v_{trascrizione}$ è ottenuto aggregando le rappresentazioni vettoriali delle singole unità enunciative della trascrizione, secondo la stessa infrastruttura di rappresentazione semantica impiegata per la coesione locale nel Capitolo 6, e $v_{consegna}$ è la rappresentazione vettoriale del compito assegnato, fisso per tutte le somministrazioni poiché la domanda di brevetto specifica un'unica consegna standardizzata.[82] La seconda compo-

80. A. Malinin, R. Van Dalen, K. Knill, Y. Wang, M. Gales, Off-topic response detection for spontaneous spoken English assessment, in Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016, pp. 1075-1084.

81. R. Barzilay, M. Lapata, Modeling Local Coherence: An Entity-Based Approach, «Computational Linguistics», 34, 1, 2008, pp. 1-34.

82. Sul metodo di calcolo della similarità tra testo prodotto e consegna, cfr. Louis, Higgins, Off-topic essay detection, cit.

nente misura la coerenza logica come proporzione di transizioni regolari sul totale delle transizioni osservate nella griglia di entità della trascrizione:

$$CO = N_{\text{transizioni regolari}} / N_{\text{transizioni totali}}$$

dove le transizioni sono classificate secondo la tipologia a quattro categorie proposta da Barzilay e Lapata, in base al ruolo grammaticale, soggetto, oggetto, altro ruolo obliquo, o assenza, ricoperto da ciascuna entità salente in frasi consecutive.[83]

Un esempio minimo illustra la costruzione della griglia. Si considerino le tre unità enunciative «Il medico ha esaminato il paziente», «Ha rilevato una lieve difficoltà articolatoria» e «La difficoltà era comparsa da pochi mesi». La Tabella 7.2 riporta il ruolo di ciascuna entità in ciascuna unità.

Tabella 7.2. Griglia di entità dell'esempio: S = soggetto, O = oggetto, – = assente.

Tabella 7.2. Griglia di entità dell'esempio: S = soggetto, O = oggetto, – = assente.

Unità	**medico**	**paziente**	**difficoltà**
1	S	O	–
2	S	–	O
3	–	–	S

Delle cinque transizioni tra unità consecutive che coinvolgono almeno un'occorrenza dell'entità, medico(S,S) tra le unità 1 e 2 è l'unica regolare; paziente(O,–), difficoltà(–,O) tra le unità 1 e 2, e medico(S,–), difficoltà(O,S) tra le unità 2 e 3, sono tutte transizioni brusche, poiché nessuna delle due mantiene lo stesso ruolo. La coerenza logica dell'esempio risulta quindi CO = 1/5 = 0,2, un valore basso che riflette il cambio di soggetto grammaticale a ogni unità, coerentemente con l'intuizione che un testo in cui il soggetto della frase cambia di continuo, senza che nessuna entità venga ripresa con ruolo costante, risulti meno coerente di un testo che mantiene lo stesso soggetto lungo più frasi.

83. Sulla classificazione delle transizioni in quattro categorie in base al ruolo grammaticale dell'entità, cfr. Barzilay, Lapata, Modeling Local Coherence, cit.

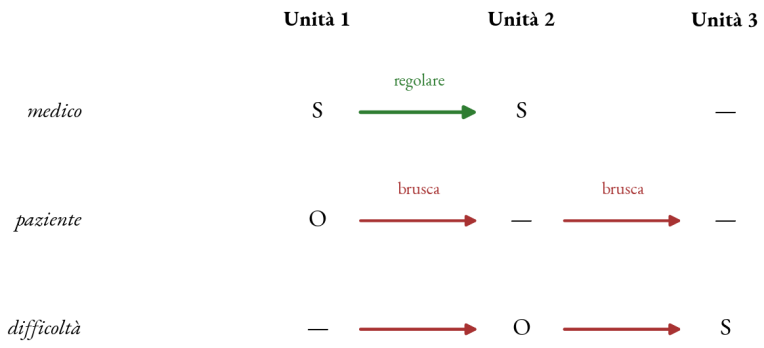

Figura 7.1. Le transizioni della griglia di entità della Tabella 7.2: in verde l'unica transizione regolare (medico, S-S), in rosso le quattro transizioni brusche.

Le due componenti, pertinenza tematica e coerenza logica, restano concettualmente distinte, ma vengono trattate come sotto-componenti di un'unica dimensione in coerenza con la definizione dell'indicatore corrispondente nella domanda di brevetto, secondo la stessa logica compositiva già adottata per la densità lessicale nel Capitolo 4.

7.4 L'estrazione automatica per l'italiano

Il calcolo della coerenza logica attraverso il modello a griglia di entità richiede l'identificazione automatica delle entità menzionate nella trascrizione e la risoluzione dei riferimenti anaforici che le collegano da una frase all'altra, un compito che condivide l'infrastruttura di risoluzione referenziale già descritta nel Capitolo 6 a proposito della densità referenziale, incluso il trattamento del soggetto nullo tipico dell'italiano. A questa infrastruttura si aggiunge, per questa dimensione, l'attribuzione del ruolo grammaticale di ciascuna occorrenza dell'entità, che il modello ricava dall'analisi sintattica a dipendenze già descritta nel Capitolo 5.[84] La costruzione della griglia di entità per il parlato spontaneo dell'italiano condivide quindi i medesimi problemi di segmentazione in unità enunciative già discussi nel Capitolo 5 a proposito delle disfluenze, che il modello affronta nella medesima fase di normalizzazione della trascrizione già richiamata per le altre dimensioni.

84. Il modello riutilizza per questa dimensione l'annotazione grammaticale prodotta dalla pipeline di analisi sintattica a dipendenze descritta nel Capitolo 5, evitando una duplicazione dell'infrastruttura di analisi tra dimensioni diverse del modello.

7.5 Dalla misura al grado della scala «Pertinenza e coerenza testuale»

Le due componenti descritte nel paragrafo 7.3 vengono tradotte nel grado corrispondente della scala di valutazione clinica a dieci punti che la domanda di brevetto definisce per questa dimensione, secondo l'architettura del Capitolo 1. Il valore composito ottenuto dalla pertinenza tematica e dalla coerenza logica, normalizzato rispetto al database normativo del modello, colloca il campione in una delle dieci fasce descrittive della domanda di brevetto, secondo la stessa logica già illustrata per le dimensioni trattate nei capitoli precedenti.

Tabella 7.1. Scala di valutazione dell'indicatore «Pertinenza e coerenza testuale».

Tabella 7.1. Scala di valutazione dell'indicatore «Pertinenza e coerenza testuale».

Grado	**Descrizione**
1	Discorso completamente fuori tema, illogico e incoerente rispetto alla richiesta.
2	Informazioni largamente irrilevanti o incoerenti, con evidenti contraddizioni logiche.
3	Discorso poco pertinente, con molte incongruenze e difficoltà nel mantenere il filo logico.
4	Alcune informazioni pertinenti, ma con deviazioni significative e incoerenze logiche.
5	Testo generalmente pertinente, con qualche incoerenza o disconnessione logica.
6	Discorso pertinente e abbastanza coerente, con occasionali deviazioni dal tema.
7	Buona coerenza e pertinenza; poche contraddizioni e buona gestione delle informazioni.
8	Discorso molto pertinente, logico e coerente, con solo lievi deviazioni o incongruenze.
9	Discorso estremamente coerente e pertinente, logica impeccabile e nessuna contraddizione.
10	Massima pertinenza e coerenza; il testo è perfettamente logico, senza alcuna incoerenza o contraddizione.

7.6 Rinvio alla validazione e ai limiti della dimensione

La rilevanza diagnostica della tangenzialità e della scarsa coerenza discorsiva è stata discussa nel Capitolo 1 con riferimento alla letteratura sulla verbosità fuori tema, che ne segnala un aumento con l'età e un'associazione con un deficit dei processi inibitori. I

limiti propri della dimensione comprendono la dipendenza della pertinenza tematica da una consegna standardizzata e la sensibilità del modello a griglia di entità alla qualità della risoluzione anaforica automatica. Il capitolo seguente tratta la quinta e ultima dimensione, la fluidità articolatoria.

Riferimenti bibliografici del Capitolo 7

Arbuckle T. Y., Gold D. P., Aging, inhibition, and verbosity, «Journal of Gerontology», 48, 5, 1993, pp. P225-P232.

Barzilay R., Lapata M., Modeling Local Coherence: An Entity-Based Approach, «Computational Linguistics», 34, 1, 2008, pp. 1-34.

Louis A., Higgins D., Off-topic essay detection using short prompt texts, in Proceedings of the NAACL HLT 2010 Fifth Workshop on Innovative Use of NLP for Building Educational Applications, 2010, pp. 92-95.

Malinin A., Van Dalen R., Knill K., Wang Y., Gales M., Off-topic response detection for spontaneous spoken English assessment, in Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016, pp. 1075-1084.

Capitolo 8 La fluidità articolatoria

8.1 Un fatto meccanico, non cognitivo

Parlare è anche un fatto meccanico: la velocità con cui i suoni vengono prodotti in sequenza, la distribuzione delle pause, la regolarità del ritmo articolatorio. È a questa componente, primariamente motoria piuttosto che cognitivo-linguistica, che è dedicata la quinta e ultima dimensione del modello, la fluidità articolatoria, corrispondente all'indicatore «Fluidità» della domanda di brevetto, già anticipata nel Capitolo 1 e ripresa nel Capitolo 2 a proposito del fenotipo motorio-articolatorio osservato nella famiglia KE: riguarda l'organizzazione motoria del parlato, non l'organizzazione lessicale, sintattica o discorsiva del contenuto linguistico prodotto.[85] Questa distinzione non è solo teorica ma ha una conseguenza diretta sulla lettura clinica dei punteggi prodotti dal modello: un deficit isolato di questa dimensione, in presenza di punteggi regolari sulle altre quattro, orienta l'interpretazione verso una compromissione del controllo motorio del parlato piuttosto che verso un deficit cognitivo-linguistico, una distinzione clinicamente rilevante che il Capitolo 1 ha indicato come una delle ragioni per cui la dimensione è stata mantenuta separata dalle altre.

La letteratura clinica di riferimento, già richiamata nel Capitolo 1, riguarda i disturbi del movimento, in primo luogo la malattia di Parkinson, in cui la disartria altera in modo misurabile la velocità di articolazione, la regolarità del ritmo e la produzione di sillabe in compiti di ripetizione rapida.[86] Un aspetto già segnalato nel Capitolo 1 riguarda l'interazione tra il livello motorio e il livello cognitivo: una disartria di grado severo può alterare la misurazione di parametri cognitivi come la fluenza verbale semantica, se i due livelli non vengono tenuti distinti in fase di analisi, il che rafforza ulteriormente la scelta di trattare la fluidità articolatoria come dimensione autonoma piuttosto che come componente delle altre quattro.[87]

85. Sul carattere primariamente motorio della fluidità articolatoria, distinto dalle componenti cognitivo-linguistiche delle altre quattro dimensioni, cfr. Capitolo 1, paragrafo 1.3, e Capitolo 2, paragrafo 2.3.

86. Cfr. Speech and language biomarkers for Parkinson's disease prediction, early diagnosis and progression, «npj Parkinson's Disease», 2025.

87. Cfr. Y. Li, J. Yang, K. Evans, J. B.-W. Wong, N. N. Dissanayaka, A. P. Vogel, Optimising verbal fluency analysis in neurological patients with dysarthria: Examples from Parkinson's disease and hereditary ataxia, «Journal of Clinical and Experimental Neuropsychology», 45, 5, 2023, pp. 452-463.

8.2 Dalla teoria clinica alla misura automatica

La foniatria e la logopedia dispongono di una tradizione consolidata di misure della fluidità articolatoria, imperniata sulla distinzione tra velocità di eloquio, che include il tempo occupato dalle pause, e velocità di articolazione in senso stretto, che lo esclude.[88] La velocità di articolazione si ottiene dividendo il numero di sillabe prodotte per il tempo di articolazione effettivo, al netto delle pause, ed è la misura meno influenzata da fattori non motori quali le esitazioni di pianificazione del discorso.[89] L'identificazione automatica delle sillabe a partire dal solo segnale acustico, senza necessità di una trascrizione ortografica preliminare, è resa possibile da algoritmi che rilevano i nuclei sillabici come picchi di intensità preceduti e seguiti da minimi, un metodo validato che consente il calcolo automatico sia della velocità di eloquio sia della velocità di articolazione.[90] La domanda di brevetto prevede a sua volta, come terza figura della documentazione allegata, un esempio di tracciato spettrografico impiegato per l'analisi della fluidità di eloquio. La Figura 2 di questo manuale ne offre un esempio illustrativo, generato a partire da un segnale sintetico e non da una registrazione reale, utile a mostrare visivamente cosa un tracciato di questo tipo renda osservabile: le bande orizzontali corrispondono alle armoniche del segnale vocalico, la loro estensione verticale ai formanti che caratterizzano i suoni prodotti, e gli intervalli bianchi alle pause tra un nucleo sillabico e il successivo.

88. Sulla distinzione tra velocità di eloquio, comprensiva del tempo di pausa, e velocità di articolazione, che lo esclude, cfr. G. S. Turner, G. Weismer, Characteristics of speaking rate in the dysarthria associated with amyotrophic lateral sclerosis, «Journal of Speech and Hearing Research», 36, 6, 1993, pp. 1134-1144, distinzione ripresa anche in Comparing manual and automated methods for calculating speaking rate in Parkinson's disease, PubMed Central, PMC11905113, 2025.

89. Sulla formula della velocità di articolazione come rapporto tra numero di sillabe e tempo di articolazione al netto delle pause, cfr. Speech and pause characteristics associated with voluntary rate reduction in Parkinson's disease and Multiple Sclerosis, «Journal of Communication Disorders», 2011.

90. N. H. de Jong, T. Wempe, Praat script to detect syllable nuclei and measure speech rate automatically, «Behavior Research Methods», 41, 2, 2009, pp. 385-390.

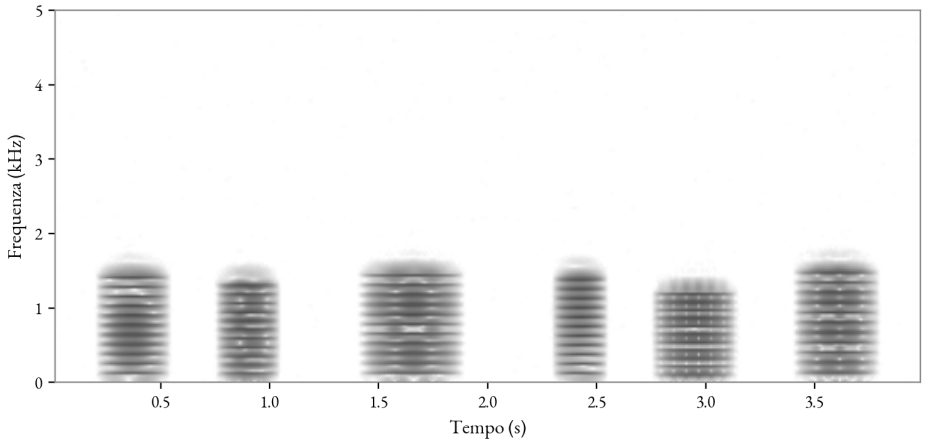

Figura 2. Esempio illustrativo di tracciato spettrografico, generato da un segnale sintetico a scopo puramente dimostrativo e non da una registrazione reale, corrispondente al tipo di analisi indicato come Figura 3 nella documentazione allegata alla domanda di brevetto. Le bande scure verticali corrispondono ai nuclei sillabici, separati da intervalli di pausa; l'analisi automatica della loro durata e frequenza costituisce la base delle componenti di velocità di articolazione e indice di pausa descritte nel paragrafo 8.3.

Per la regolarità del ritmo articolatorio, la misura di riferimento in fonetica sperimentale è l'indice di variabilità a coppie normalizzato, proposto da Esther Grabe ed Ee Ling Low per la classificazione ritmica delle lingue, che quantifica la differenza di durata tra intervalli vocalici successivi, normalizzata per la durata media della coppia.[91] Come nel caso della profondità di Yngve discusso nel Capitolo 5, l'indice fu concepito per confrontare classi ritmiche tra lingue tipologicamente distinte, non come misura clinica di patologia del linguaggio: il suo impiego nel modello FOXP2 riprende lo strumento di misura, non l'ipotesi tipologica originaria per cui era stato proposto.

8.3 La formula adottata dal modello

Il modello FOXP2 misura la fluidità articolatoria attraverso tre componenti, corrispondenti ai tre indicatori derivabili dalla trascrizione temporizzata già annunciati nel Capitolo 1: la velocità di produzione delle sillabe, la distribuzione delle pause e la regolarità del ritmo articolatorio. La prima componente è la velocità di articolazione:

91. E. Grabe, E. L. Low, Durational variability in speech and the rhythm class hypothesis, in C. Gussenhoven, N. Warner (a cura di), Papers in Laboratory Phonology 7, Cambridge University Press, Cambridge, 2002, pp. 515-546.

$$VA = N_{sillabe} / T_{articolazione}$$

dove $N_{sillabe}$ è il numero di nuclei sillabici rilevati nella registrazione e $T_{articolazione}$ è il tempo totale di produzione effettiva del parlato, al netto delle pause.[92] La seconda componente è l'indice di pausa, calcolato come proporzione del tempo totale della registrazione occupato da intervalli di silenzio superiori a una soglia minima:

$$IP = T_{pause} / T_{totale}$$

con una soglia minima di durata della pausa fissata, in linea con la prassi corrente in questo tipo di analisi, a poche centinaia di millisecondi, per escludere le micropause fisiologiche dell'articolazione dal conteggio delle pause comunicativamente rilevanti.[93] La terza componente è la regolarità ritmica, calcolata come indice di variabilità a coppie normalizzato applicato alle durate degli intervalli vocalici successivi della registrazione:

$$RR = (100 / (m - 1)) \times \Sigma |d_k - d_{k+1}| / ((d_k + d_{k+1}) / 2)$$

dove m è il numero di intervalli vocalici identificati e d_k è la durata del k-esimo intervallo.[94] Valori più alti dell'indice segnalano una maggiore irregolarità del ritmo articolatorio, coerentemente con l'osservazione clinica di un'articolazione meno regolare nei disturbi del movimento discussi nel paragrafo 8.1.

8.4 L'estrazione automatica per l'italiano

La regolarità ritmica misurata dall'indice di variabilità a coppie richiede una cautela interpretativa specifica per l'italiano, che la letteratura di fonetica sperimentale classifica tra le lingue a ritmo sillabico, insieme al francese e allo spagnolo, distinte dalle lingue a ritmo accentuale come l'inglese e il tedesco, che presentano valori dell'indice strutturalmente più alti indipendentemente da qualunque condizione patologica.[95] Questo dato impone che la normalizzazione del punteggio di regolarità ritmica avvenga rispetto a un database normativo costruito su parlanti italiani, secondo quanto già

92. La formula riprende la definizione operativa di velocità di articolazione descritta in de Jong, Wempe, cit.

93. Sulla soglia di durata minima impiegata per distinguere le pause comunicativamente rilevanti dalle micropause fisiologiche dell'articolazione, cfr. la prassi descritta in Cross-lingual Dysarthria Severity Classification for English, Korean, and Tamil, arXiv:2209.12942, che adotta una soglia di 0,1 secondi.

94. La formula corrisponde alla definizione dell'indice di variabilità a coppie normalizzato proposta in Grabe, Low, cit.

stabilito come principio generale nel Capitolo 1: un valore dell'indice che segnalerebbe un'anomalia ritmica se calcolato su un parlante di una lingua a ritmo accentuale può rientrare nella piena norma per un parlante italiano, e l'interpretazione clinica del punteggio deve tenere conto di questa differenza strutturale tra lingue, esattamente come il Capitolo 4 ha segnalato per la densità lessicale a proposito della differenza tra parlato e scritto.

8.5 Dalla misura al grado della scala «Fluidità»

Le tre componenti descritte nel paragrafo 8.3 vengono tradotte nel grado corrispondente della scala di valutazione clinica a dieci punti che la domanda di brevetto definisce per questa dimensione, secondo l'architettura del Capitolo 1. Il valore composito ottenuto dalla velocità di articolazione, dall'indice di pausa e dalla regolarità ritmica, normalizzato rispetto al database normativo italiano secondo la cautela indicata nel paragrafo 8.4, colloca il campione in una delle dieci fasce descrittive della domanda di brevetto. La domanda di brevetto specifica inoltre che l'analisi di supporto a questo indicatore viene condotta attraverso il software Praat e il programma Spek per la generazione dello spettrogramma, strumenti che il modello impiega in fase di verifica accanto all'algoritmo di rilevazione automatica dei nuclei sillabici qui presentato.

Tabella 8.1. Scala di valutazione dell'indicatore «Fluidità».

Tabella 8.1. Scala di valutazione dell'indicatore «Fluidità».

Grado	Descrizione
1	Fonoarticolazione molto scarsa; discorso frammentato e difficile da comprendere.
2	Gravi difficoltà nella fluidità, con frequenti esitazioni e interruzioni.
3	Fluidità limitata, con numerose pause e difficoltà nell'articolare le parole.
4	Fluidità ridotta, ma comprensibile; discorso spezzato con pause frequenti.
5	

95. Sulla classificazione dell'italiano, del francese e dello spagnolo come lingue a ritmo sillabico, con valori dell'indice di variabilità a coppie strutturalmente più bassi rispetto alle lingue a ritmo accentuale come l'inglese e il tedesco, cfr. Grabe, Low, cit., e la discussione ripresa in LFSAG, Laboratorio di Fonetica Sperimentale «Arturo Genre», sezione dedicata al PVI.

Grado	Descrizione
	Fluidità sufficiente, con qualche esitazione e occasionali pause, ma discorso complessivamente fluido.
6	Buona fluidità, con solo lievi esitazioni o pause; comprensibilità adeguata.
7	Fluidità notevole, discorso scorrevole con rari momenti di esitazione.
8	Ottima fluidità, discorso fluido e ben articolato, con rare pause.
9	Fluidità quasi perfetta, discorso estremamente fluido e scorrevole, senza esitazioni.
10	Eccellente fluidità, fonoarticolazione impeccabile; discorso perfettamente scorrevole senza alcuna pausa o difficoltà.

8.6 Rinvio alla validazione e ai limiti della dimensione, e chiusura della Parte II

La rilevanza diagnostica della fluidità articolatoria nei disturbi del movimento è stata discussa nel Capitolo 1 con riferimento alla letteratura sulla malattia di Parkinson. I limiti propri della dimensione comprendono la sensibilità della rilevazione automatica dei nuclei sillabici alla qualità della registrazione audio e la necessità di un database normativo costruito specificamente su parlanti italiani per l'interpretazione della regolarità ritmica.

Con questo capitolo si chiude la Parte II del manuale. I cinque capitoli che la compongono hanno seguito ciascuna dimensione dalla teoria alla formula, fino al grado corrispondente sulla scala della domanda di brevetto. L'architettura computazionale che integra le cinque dimensioni in un'unica pipeline, dalla trascrizione automatica del parlato al database normativo utilizzato per la normalizzazione dei punteggi qui descritti, resta l'infrastruttura tecnica presupposta da ciascun capitolo di questa parte.

Riferimenti bibliografici del Capitolo 8

Comparing manual and automated methods for calculating speaking rate in Parkinson's disease, PubMed Central, PMC11905113, 2025.

Cross-lingual Dysarthria Severity Classification for English, Korean, and Tamil, arXiv:2209.12942.

de Jong N. H., Wempe T., Praat script to detect syllable nuclei and measure speech rate automatically, «Behavior Research Methods», 41, 2, 2009, pp. 385-390.

Grabe E., Low E. L., Durational variability in speech and the rhythm class hypothesis, in C. Gussenhoven, N. Warner (a cura di), Papers in Laboratory Phonology 7, Cambridge University Press, Cambridge, 2002, pp. 515-546.

LFSAG, Laboratorio di Fonetica Sperimentale «Arturo Genre», sezione dedicata al PVI.

Speech and language biomarkers for Parkinson's disease prediction, early diagnosis and progression, «npj Parkinson's Disease», 2025.

Speech and pause characteristics associated with voluntary rate reduction in Parkinson's disease and Multiple Sclerosis, «Journal of Communication Disorders», 2011.

Li Y., Yang J., Evans K., Wong J. B.-W., Dissanayaka N. N., Vogel A. P., Optimising verbal fluency analysis in neurological patients with dysarthria: Examples from Parkinson's disease and hereditary ataxia, «Journal of Clinical and Experimental Neuropsychology», 45, 5, 2023, pp. 452-463.

Turner G. S., Weismer G., Characteristics of speaking rate in the dysarthria associated with amyotrophic lateral sclerosis, «Journal of Speech and Hearing Research», 36, 6, 1993, pp. 1134-1144.

Capitolo 9 Un caso illustrativo

::: avviso Il caso presentato in questo capitolo è interamente costruito a scopo didattico. Nessuna delle due trascrizioni proviene da una registrazione reale, nessun dato clinico è coinvolto, e i punteggi qui calcolati non hanno alcun valore diagnostico. Lo scopo è mostrare, su un esempio concreto e verificabile riga per riga, come le cinque dimensioni descritte nei capitoli precedenti operino insieme sullo stesso campione, cosa che i singoli esempi dei Capitoli 4-7 non potevano mostrare, ciascuno limitato a una frase isolata.

9.1 Il campione T0

Un uomo di settantatré anni, secondo il protocollo descritto nel Capitolo 1, viene registrato mentre risponde alla consegna standardizzata sulle attività del fine settimana. La trascrizione, normalizzata secondo la procedura di pulizia del testo adottata dal modello, è la seguente.

«Allora, sabato mattina sono andato al mercato con mia moglie, abbiamo comprato della frutta e del pesce fresco, poi siamo tornati a casa e ho preparato il pranzo per tutta la famiglia. Nel pomeriggio è venuto mio nipote, abbiamo giocato a carte e guardato una partita di calcio insieme. La domenica invece siamo stati più tranquilli: siamo andati a messa la mattina e poi abbiamo fatto una passeggiata nel parco vicino casa, faceva bel tempo e c'era molta gente.»

Il campione conta settantanove parole, una lunghezza modesta ma sufficiente, secondo quanto discusso nel Capitolo 4, per un calcolo attendibile degli indici meno sensibili alla brevità del testo.

9.2 Le cinque dimensioni applicate al campione T0

Della densità lessicale descritta nel Capitolo 4, la prima componente si calcola direttamente: delle settantanove parole del campione, quarantatré sono lessicali, per una densità Ld = 43/79 × 100 ≈ 54,4, un valore in linea con quanto atteso per il parlato spontaneo secondo Ure. La seconda componente, la diversità del vocabolario, mostra un tratto rilevante anche senza il calcolo completo di MTLD: delle quarantatré parole lessicali, quaranta sono forme diverse tra loro, con un rapporto tipo-token limitato alle

sole parole di contenuto pari a 0,93, segno di un vocabolario ricco e non ripetitivo. Collocato sulla Tabella 4.1, il profilo corrisponde a un grado 7, «vocabolario abbastanza ricco e preciso, rare imprecisioni semantiche».

Per la complessità sintattica, si consideri la frase «Siamo andati a messa la mattina»: sei parole, cinque relazioni di dipendenza, di cui «Siamo» dipende da «andati» a distanza 1, «a» da «messa» a distanza 1, «messa» da «andati» a distanza 2, «la» da «mattina» a distanza 1, «mattina» da «andati» a distanza 2. La somma è 7, e MDD = 7/5 = 1,4. Il valore, isolato, non distingue ancora molto; ciò che caratterizza l'intero campione è piuttosto la densità di coordinazione tra le proposizioni, «abbiamo comprato... poi siamo tornati... e ho preparato», più tipica di un eloquio ben organizzato che di uno frammentato. Il grado corrispondente sulla Tabella 5.1 è un 7, «buona varietà nelle costruzioni sintattiche, errori minimi».

La coesione discorsiva risulta solida: ogni soggetto sottinteso delle forme verbali, «abbiamo comprato», «ho preparato», «abbiamo giocato», trova un antecedente immediato nella prima persona plurale o singolare già introdotta, senza ambiguità referenziale. Il grado sulla Tabella 6.1 è un 7, «buona coesione, frasi ben collegate». La pertinenza e coerenza testuale è la dimensione più alta del profilo: il racconto risponde punto per punto alla consegna, segue l'ordine cronologico sabato-domenica senza salti né contraddizioni, e si colloca a un grado 8 sulla Tabella 7.1, «discorso molto pertinente, logico e coerente». La fluidità articolatoria, infine, per come la trascrizione registra l'assenza di pause prolungate o esitazioni, corrisponde a un grado 7 sulla Tabella 8.1, «fluidità notevole, discorso scorrevole con rari momenti di esitazione».

9.3 Il campione T1, tre mesi dopo

Trascorsi tre mesi, come previsto dal protocollo, lo stesso uomo viene sottoposto a una seconda registrazione con la medesima consegna.

«Allora... sabato... sono stato a casa. Poi, non so, ho fatto... delle cose. Mia moglie è andata al mercato. Io sono stato a casa. Domenica... anche domenica sono stato a casa. Faceva freddo. Non mi ricordo bene cosa ho fatto.»

Il campione conta quaranta parole, la metà circa del precedente, un primo segnale, seppur grezzo, di riduzione della produttività verbale.

9.4 Le cinque dimensioni applicate al campione T1

Il calcolo della densità lessicale riserva una sorpresa istruttiva: delle quaranta parole del campione, ventuno sono lessicali, per Ld = 21/40 × 100 ≈ 52,5, un valore vicino a quello di T0 e che, preso da solo, non segnalerebbe alcun mutamento. È la seconda componente a rivelare il deterioramento: delle ventuno parole lessicali, solo quindici sono forme diverse, con «stato», «casa», «domenica» e «fatto» ciascuna ripetuta più volte, per un rapporto tipo-token limitato alle parole di contenuto pari a 0,71, sensibilmente più basso dello 0,93 di T0. È esattamente il caso che il Capitolo 4 aveva anticipato in astratto: la densità lessicale da sola, insensibile alla ripetizione, può restare stabile mentre la diversità del vocabolario, la componente MTLD del modello, già segnala il restringimento. Il profilo composito colloca il campione a un grado 4 sulla Tabella 4.1, «vocabolario basilare, presenza di errori e imprecisioni».

Anche la complessità sintattica illustra lo stesso principio. La frase «Ho fatto delle cose» dà MDD = 4/3 ≈ 1,3, un valore isolato non lontano da quello calcolato per T0: la differenza reale non è nella distanza di dipendenza di una singola frase breve, ma nella quasi totale assenza di coordinazione tra le proposizioni dell'intero campione, ridotte a una sequenza di enunciati brevi e giustapposti, un solo caso di subordinazione, «cosa ho fatto», per il resto paratassi minima. Il grado sulla Tabella 5.1 scende a 4, «struttura delle frasi limitata, errori frequenti nella costruzione di frasi complesse».

La coesione discorsiva si indebolisce: «delle cose» è un riferimento vago, privo di un antecedente lessicale specifico, e il passaggio da un enunciato all'altro procede per giustapposizione più che per connessione esplicita. Il grado sulla Tabella 6.1 è un 3, «coesione debole, frasi poco legate tra loro». La pertinenza e coerenza testuale, coerentemente con quanto discusso nel Capitolo 7 a proposito della verbosità fuori tema, mostra il calo più marcato: «non so», «delle cose» e «non mi ricordo bene cosa ho fatto» sono formulazioni che eludono il contenuto richiesto dalla consegna invece di soddisfarlo, un pattern testuale, non una semplice impressione soggettiva di chi legge. Il grado sulla Tabella 7.1 è un 3, «discorso poco pertinente, con molte incongruenze e difficoltà nel mantenere il filo logico». La fluidità articolatoria, infine, registra le pause segnalate nella trascrizione a inizio di enunciato, «Allora...», «sabato...», «Domenica...», corrispondenti a un grado 4 sulla Tabella 8.1, «fluidità ridotta, ma comprensibile, discorso spezzato con pause frequenti».

9.5 Il confronto sul grafico radar

La Tabella 9.1 riepiloga i dieci gradi ottenuti, e la Figura 9.1 li riporta sul grafico radar previsto dalla domanda di brevetto.

Tabella 9.1. Gradi delle cinque dimensioni per le campionature T0 e T1 del caso illustrativo.

Tabella 9.1. Gradi delle cinque dimensioni per le campionature T0 e T1 del caso illustrativo.

Dimensione	T0	T1
Vocabolario	7	4
Sintassi	7	4
Coesione e connessione testuale	7	3
Pertinenza e coerenza testuale	8	3
Fluidità	7	4

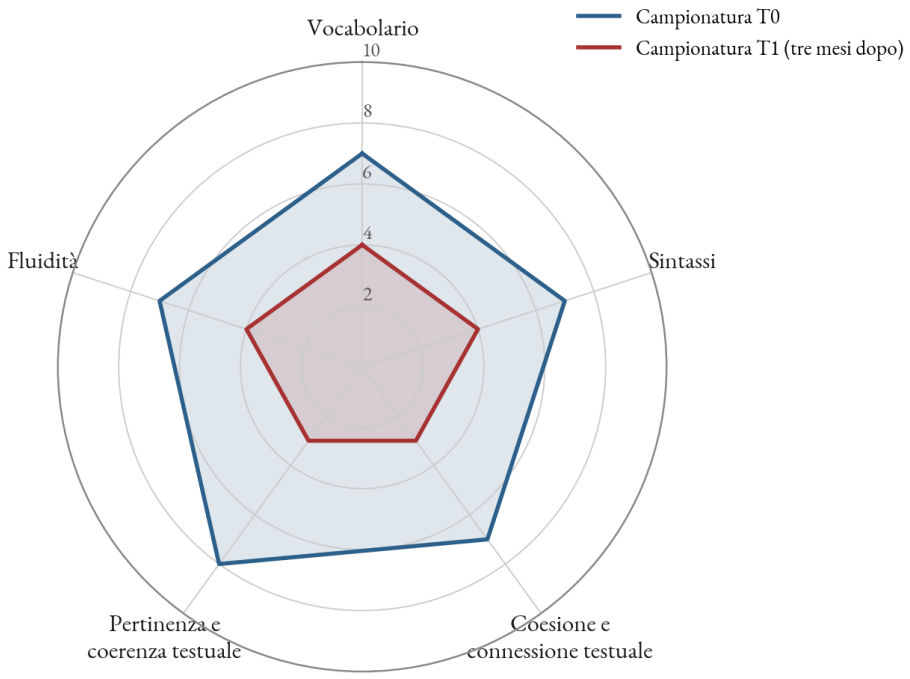

Figura 9.1. Confronto tra le due campionature del caso illustrativo sul grafico radar. Il poligono di T1 risulta inscritto per una porzione ben superiore alla metà della propria area in quello di T0 su tutti e cinque gli assi, la condizione che la domanda di brevetto associa a un peggioramento significativo.

La riduzione non è uniforme: la pertinenza e coerenza testuale, la dimensione più alta in T0, è anche quella che si contrae di più, un pattern coerente con l'ipotesi, discussa nel Capitolo 7, che il calo dei processi inibitori ed esecutivi incida sulla capacità di restare aderenti al tema prima ancora che sulle altre componenti linguistiche. Si tratta di un'osservazione plausibile su questo singolo caso costruito, non di una generalizzazione: un solo campione fittizio non prova alcuna regolarità clinica, e la sua unica funzione è mostrare come una lettura differenziata, dimensione per dimensione, aggiunga informazione rispetto al solo punteggio composito.

9.6 Cosa mostra, e cosa non mostra, questo esempio

L'esempio dimostra la meccanica del modello: come le formule dei Capitoli 4-8 si applichino a un campione reale di parlato, come misure apparentemente stabili, la densità lessicale grezza, la distanza di dipendenza di una singola frase, possano nascondere un deterioramento che altre componenti della stessa dimensione, la diversità lessicale, la densità di coordinazione sull'intero campione, rendono visibile. Non dimostra, e non potrebbe dimostrare, che il modello funzioni: la validazione è un processo in corso, di natura ben diversa da quanto un singolo caso costruito a tavolino possa illustrare. L'infrastruttura computazionale che rende automatico, su campioni reali, il calcolo qui condotto a mano resta l'oggetto proprio di uno sviluppo tecnico separato.

Capitolo 10 Filologia attributiva e perdita di autorialità

10.1 Il testo come prodotto della mente: la filologia cognitiva

La filologia attributiva discussa in questo capitolo non è un caso isolato. Appartiene a un principio più ampio, che una disciplina distinta ha reso esplicito fin dal nome: la filologia cognitiva, che tratta i testi scritti e orali come prodotto dei processi mentali di chi li compone, li trasmette o li riceve, e confronta l'evidenza testuale con i risultati della ricerca sperimentale in psicologia cognitiva, neuroscienze e intelligenza artificiale.[96] Il termine fu coniato da Paolo Canettieri, professore di filologia romanza alla Sapienza, in una recensione a Cesare Segre del 1999; dal 2003 è anche il nome di una rivista propria, che la stessa università pubblica in accesso aperto.[97]

Il programma della disciplina non si ferma a osservare che i testi riflettono la mente di chi li produce, un'osservazione su cui poca filologia dissentirebbe. Lo rende operativo: applica alla trasmissione dei testi strumenti quantitativi mutuati dalla teoria dell'informazione, trattando la variazione testuale, l'errore di copia, la struttura ritmica come dati misurabili, non solo interpretabili.[98] È lo stesso spostamento, dall'osservazione qualitativa alla misura, che la filologia attributiva compie quando sostituisce il giudizio sulla compatibilità stilistica con una distanza calcolabile, discussa nel paragrafo seguente, e che il modello FOXP2 compie a sua volta quando traduce la valutazione clinica del linguaggio in punteggi su cinque dimensioni. Le tre discipline condividono la premessa, non gli strumenti: la filologia cognitiva la applica al funzionamento della mente in generale, la filologia attributiva alla stabilità dell'identità individuale nel tempo, il modello descritto in questo manuale al suo deterioramento.

96. Sulla definizione della filologia cognitiva come scienza che studia i testi scritti e orali quali prodotti dei processi mentali umani, confrontando l'evidenza testuale con la ricerca sperimentale in psicologia, neuroscienze e intelligenza artificiale, cfr. About the Journal, «Cognitive Philology», Sapienza Università di Roma.

97. Sulla coniazione del termine da parte di Canettieri nel 1999 e sulla fondazione della rivista, cfr. S. Argurio, A proposito di «Filologia cognitiva». Silvia Argurio dialoga con Paolo Canettieri, «Insula europea», 2018.

98. P. Canettieri, V. Loreto, M. Rovetta, G. Santini, Ecdotics and Information Theory, «Rivista di Filologia Cognitiva», 2005.

10.2 La filologia attributiva come strumento cardine

La filologia attributiva assegna un testo di paternità ignota o contestata a un autore noto, sulla base della premessa che il comportamento linguistico di un individuo presenti regolarità stabili e in gran parte inconsapevoli, tali da produrre una coerenza stilistica riconoscibile attraverso i testi che quello stesso individuo produce.[99] La misura più diffusa per operazionalizzare questa premessa è la Delta di Burrows, che quantifica la distanza stilistica tra testi a partire dalle frequenze relative delle parole più comuni, tipicamente parole grammaticali, sotto l'assunto che la loro distribuzione sfugga al controllo consapevole dello scrivente e per questo tradisca l'abitudine linguistica individuale meglio del lessico tematico.[100] Lavori recenti hanno raffinato questa impostazione con l'ipotesi dei profili chiave, secondo cui un autore si riconosce non dalla deviazione di singoli tratti ma dalla configurazione complessiva delle sue scelte linguistiche entro uno spazio multidimensionale.[101]

Il presupposto comune a ogni impiego della filologia attributiva, dal caso classico della paternità dei testi antichi al caso forense della email anonima, è che l'idioletto individuale sia sufficientemente stabile da fungere da termine di paragone: si confronta il testo controverso con un corpus di riferimento di paternità certa, e la distanza stilistica misurata autorizza l'attribuzione o l'esclusione.

Questo presupposto non nasce con la stilometria computazionale. La filologia lo utilizza da secoli sotto il nome di usus scribendi, l'insieme delle abitudini linguistiche ricorrenti e in gran parte involontarie con cui un copista o un autore compone i propri testi, abitudini che la critica testuale impiega per attribuire un manoscritto anonimo, per distinguere la mano di uno scriba da quella di un altro in un codice composito, o per stabilire se un'opera tramandata sotto un certo nome sia effettivamente compatibile con lo stile abituale di quell'autore.[102] La Delta di Burrows non introduce quindi un principio nuovo: gli affianca una misura numerica, sostituendo il giudizio filologico sulla

99. Sulla premessa di stabilità idiolettale come fondamento teorico della stilometria attributiva, cfr. J. Burrows, 'Delta': A Measure of Stylistic Difference and a Guide to Likely Authorship, «Literary and Linguistic Computing», 17, 3, 2002, pp. 267-287.

100. Sull'uso delle parole a più alta frequenza, in prevalenza grammaticali, come tratto distintivo indipendente dal contenuto, cfr. la sintesi in Towards a better understanding of Burrows's Delta in literary authorship attribution, Association for Computational Linguistics, 2015.

101. Sull'ipotesi dei profili chiave e sulla configurazione complessiva delle scelte linguistiche come firma stilometrica, cfr. Nini, 2023, richiamato in Understanding and explaining Delta measures for authorship attribution, «Digital Scholarship in the Humanities», 32, suppl. 2, 2017, pp. ii4-ii16.

compatibilità stilistica con una distanza calcolabile, ma lascia intatto l'assunto di fondo, che le scelte linguistiche più automatiche e meno sorvegliate di un individuo compongano una firma riconoscibile nel tempo. È questa continuità, tra il criterio filologico tradizionale e la sua versione statistica novecentesca, a rendere legittimo il passo ulteriore proposto nel paragrafo seguente.

10.3 L'inversione: quando l'autore si allontana da sé stesso

Chi scrive propone qui un'estensione di questo schema, presentata esplicitamente come ipotesi di lavoro e non come risultato acquisito. Il ragionamento si lascia scomporre in passaggi distinti, perché la conclusione a cui arriva è meno intuitiva delle premesse da cui muove.

Il primo passaggio riguarda cosa renda possibile un'attribuzione. La filologia attributiva, sia nella forma tradizionale dell'usus scribendi sia nella forma computazionale della Delta di Burrows, funziona perché confronta un testo di paternità incerta con un corpus di riferimento della stessa persona, di cui si assume la paternità. Se la distanza stilistica tra i due è piccola, l'attribuzione è confermata; se è grande, l'attribuzione è messa in dubbio o esclusa. Il metodo, va sottolineato, non sa nulla della biografia dell'autore: giudica solo sulla base della distanza numerica tra due insiemi di testi.

Il secondo passaggio consiste nel cambiare cosa venga confrontato con cosa, lasciando invariato il metodo. Anziché confrontare un testo controverso di autore ignoto con il corpus noto di un autore diverso, si può confrontare una campionatura tarda di un parlante con le campionature precoci dello stesso parlante, raccolte quando le sue competenze linguistiche non erano ancora compromesse. Il metodo non cambia: resta una misura di distanza stilistica calcolata sulle stesse categorie di parole. Cambia solo il fatto che, in questo caso, sappiamo con certezza biografica che l'autore è il medesimo in entrambe le campionature.

Il terzo passaggio è quello che dà al fenomeno il nome di deautorializzazione, e richiede una precisazione per non essere frainteso. Non si sostiene che una persona

smetta, in un senso qualsiasi, di essere l'autore delle proprie parole: l'identità biografica del parlante non è in discussione, e nessuna misura stilometrica potrebbe metterla in discussione. Quello che può accadere è più circoscritto e più tecnico: se la demenza altera in modo sufficientemente ampio le abitudini linguistiche automatiche di un individuo, quelle stesse su cui si fonda l'usus scribendi, la distanza stilistica tra la campionatura tarda e il corpus di riferimento precoce dello stesso parlante può diventare grande quanto quella che, in un caso attributivo ordinario, indurrebbe a escludere la paternità. In altri termini, se si sottoponesse la campionatura tarda alla stessa procedura con cui la filologia attributiva verifica la paternità di un testo, usando come termine di paragone i testi precoci dello stesso individuo, la procedura potrebbe restituire un esito negativo. Non perché l'autore sia cambiato, ma perché la sua firma linguistica si è allontanata da sé stessa, al punto che il metodo, cieco alla biografia per costruzione, non la riconosce più come propria. È questo scarto, tra un dato biografico che non cambia e un dato stilometrico che può cambiare fino a contraddirlo, a costituire il fenomeno che questo capitolo chiama deautorializzazione. Lo distingue da un generico mutamento linguistico legato all'età un tratto preciso: la grandezza del cambiamento è la stessa che la filologia attributiva tratta come segno di paternità diversa.

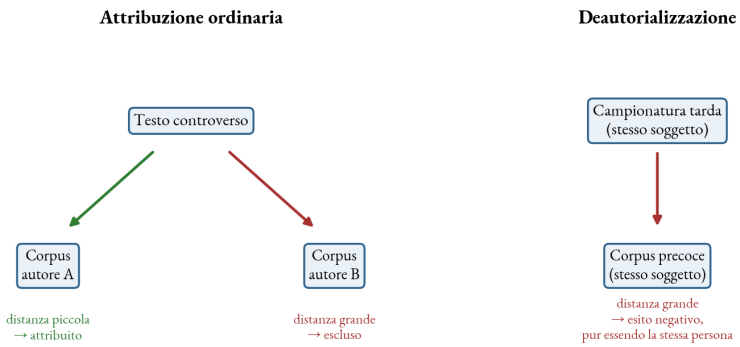

Figura 10.1. A sinistra, la logica dell'attribuzione ordinaria: un testo controverso confrontato con corpora di due autori diversi, con attribuzione al più vicino. A destra, la deautorializzazione: la campionatura tarda confrontata con il corpus precoce dello stesso soggetto può restituire una distanza altrettanto grande, pur trattandosi della stessa persona.

102. Sull'uso dell'usus scribendi come criterio di attribuzione nella critica testuale, applicato tra l'altro alla questione dell'autenticità delle lettere pseudoepigrafe nel corpus paolino, cfr. P. N. Harrison, The Problem of the Pastoral Epistles, Oxford University Press, London, 1921, che fonda l'esclusione delle lettere pastorali dal corpus autentico di Paolo su uno scarto sistematico nel vocabolario e nelle costruzioni sintattiche ricorrenti rispetto alle lettere di paternità non contestata.

Il precedente empirico più diretto per questa ipotesi, sebbene condotto con finalità diverse e non nel quadro teorico qui proposto, è lo studio sulla scrittura tarda di Iris Murdoch, già richiamato nel Capitolo 5, che applica misure di complessità sintattica automatica al confronto tra i romanzi scritti dall'autrice prima e durante l'insorgenza della malattia di Alzheimer, riscontrando un deterioramento misurabile della complessità sintattica riconoscibile a livello del singolo autore nel tempo.[103] Un caso analogo, condotto su base lessicale anziché sintattica, riguarda i romanzi tardi di Agatha Christie, in cui un'analisi quantitativa del vocabolario ha rilevato una riduzione della ricchezza lessicale e un aumento delle espressioni vaghe interpretati dagli autori dello studio come possibile indizio di un declino cognitivo mai diagnosticato in vita.[104] In entrambi i casi la letteratura esistente misura il cambiamento linguistico intra-autore nel tempo, non lo formalizza esplicitamente come una perdita di attribuibilità dell'opera tarda a sé stessa nei termini in cui la filologia attributiva definisce l'attribuzione; è questo secondo passaggio, il trattamento del declino come caso particolare, e per certi versi speculare, del problema attributivo, a costituire il contributo teorico proposto in questo capitolo.

10.4 Implicazioni per il modello FOXP2

Il modello descritto nei Capitoli 4-8 già possiede, nella dimensione del vocabolario, una componente sensibile a un fenomeno imparentato: la diversità lessicale misurata da MTLD, discussa nel Capitolo 4 e applicata al caso illustrativo del Capitolo 9, registra la contrazione del repertorio di un individuo rispetto a sé stesso nel tempo, una volta normalizzata contro il proprio andamento pregresso più che contro la sola popolazione di riferimento. L'ipotesi qui avanzata suggerisce una direzione di sviluppo ulteriore, non ancora implementata né prevista dalla domanda di brevetto: affiancare alle cinque dimensioni un indice di distanza stilometrica intra-soggetto, calcolato secondo il metodo di Burrows o una sua variante tra le parole a più alta frequenza delle campionature successive dello stesso parlante, trattando ogni nuova campionatura come un testo la cui paternità rispetto al proprio autore pregresso va, in un certo senso, sempre riverificata.

103. Computerized assessment of syntactic complexity in Alzheimer's disease: a case study of Iris Murdoch's writing, «Behavior Research Methods», 2010.

104. I. Lancashire, G. Hirst, Vocabulary changes in Agatha Christie's mysteries as an indication of dementia: A case study, 19th Annual Rotman Research Institute Conference, Cognitive Aging: Research and Practice, 2009.

Una simile misura non sostituirebbe le cinque dimensioni già descritte, ma ne offrirebbe una lettura di sfondo, complementare: mentre densità lessicale, complessità sintattica, coesione, pertinenza e fluidità misurano proprietà del singolo campione secondo criteri linguistici generali, la distanza stilometrica intra-soggetto misurerebbe la tenuta dell'identità linguistica individuale attraverso le campionature, un livello di analisi ulteriore rispetto a quelli già discussi.

10.5 Cautele

Vanno segnalati con chiarezza i limiti di questa proposta. Primo, l'estensione della filologia attributiva qui descritta non è stata validata: gli studi citati sul singolo caso Murdoch e sul singolo caso Christie non costituiscono una dimostrazione generalizzabile, e chi scrive non ne presenta il contenuto come tale. Secondo, la distanza stilometrica intra-soggetto richiede una quantità di testo per campionatura superiore a quella prevista dal protocollo attuale, di uno o due minuti a somministrazione, un problema di dimensione del campione già segnalato nel Capitolo 4 a proposito della sensibilità degli indici lessicali alla lunghezza del testo, e qui potenzialmente più severo data la minore quantità di dati per confronto stilometrico rispetto ai corpora letterari su cui la Delta di Burrows è stata tradizionalmente impiegata. Terzo, la stessa nozione di stabilità idiolettale su cui la filologia attributiva si fonda è stata di recente messa in discussione dalla disponibilità di modelli linguistici di grandi dimensioni capaci di imitare lo stile altrui, un problema di natura diversa da quello qui trattato ma che invita alla stessa cautela metodologica: prima di trattare una distanza stilometrica come segno di declino cognitivo, occorre escludere spiegazioni alternative del mutamento osservato.[105] Questa direzione di sviluppo, per tutte queste ragioni, resta un'estensione teorica non ancora consolidata del modello.

Riferimenti bibliografici del Capitolo 10

About the Journal, «Cognitive Philology», Sapienza Università di Roma.

105. Sulla messa in discussione delle assunzioni fondative sull'idioletto a fronte della capacità dei modelli linguistici di grandi dimensioni di imitarne lo stile, cfr. G. K. Mikros, Large Language Models and Forensic Linguistics: Navigating Opportunities and Threats in the Age of Generative AI.

Argurio S., A proposito di «Filologia cognitiva». Silvia Argurio dialoga con Paolo Canettieri, «Insula europea», 2018.

Burrows J., 'Delta': A Measure of Stylistic Difference and a Guide to Likely Authorship, «Literary and Linguistic Computing», 17, 3, 2002, pp. 267-287.

Canettieri P., Loreto V., Rovetta M., Santini G., Ecdotics and Information Theory, «Rivista di Filologia Cognitiva», 2005.

Computerized assessment of syntactic complexity in Alzheimer's disease: a case study of Iris Murdoch's writing, «Behavior Research Methods», 2010.

Harrison P. N., The Problem of the Pastoral Epistles, Oxford University Press, London, 1921.

Lancashire I., Hirst G., Vocabulary changes in Agatha Christie's mysteries as an indication of dementia: A case study, 19th Annual Rotman Research Institute Conference, Cognitive Aging: Research and Practice, 2009.

Mikros G. K., Large Language Models and Forensic Linguistics: Navigating Opportunities and Threats in the Age of Generative AI.

Towards a better understanding of Burrows's Delta in literary authorship attribution, Association for Computational Linguistics, 2015.

Understanding and explaining Delta measures for authorship attribution, «Digital Scholarship in the Humanities», 32, suppl. 2, 2017, pp. ii4-ii16.

Conclusione

Un brevetto è un testo tecnico-giuridico che rivendica un metodo. Un modello è un'altra cosa: una spiegazione di perché quel metodo dovrebbe funzionare, costruita in modo che chiunque possa verificarla indipendentemente da chi l'ha scritta. Questo manuale ha provato a ricavare il secondo dal primo con gli strumenti della filologia più che con quelli dell'ingegneria: non prendendo la domanda di brevetto per buona alla lettera, ma ricostruendone i presupposti, e segnando ogni volta dove finisce il documento depositato e dove comincia l'interpretazione di chi scrive.

Il risultato è un oggetto ibrido, e lo è di proposito. Le cinque dimensioni linguistiche vengono dal brevetto, le formule che le calcolano no. La scala clinica a dieci punti viene dal brevetto, il database normativo che la renderebbe leggibile no. Il capitolo sull'attribuzione d'autore non viene dal brevetto affatto: è un'estensione della sua logica oltre il perimetro legale che lo definisce.

Le cinque dimensioni non misurano una malattia. Misurano il linguaggio: densità lessicale, complessità sintattica, coesione, pertinenza, fluidità. Un profilo linguistico stabile in questi termini, e la sua variazione nel tempo, sono un dato che parla a qualunque disciplina si occupi di ricostruire, da tracce testuali o vocali, lo stato o l'identità di chi parla. Da questa genericità derivano campi di applicazione distinti, ciascuno con un proprio grado di maturità e un proprio orizzonte di sviluppo.

In linguistica clinica il campo d'uso è quello per cui il modello è stato pensato: il monitoraggio longitudinale del parlato spontaneo in soggetti a rischio di decadimento cognitivo, dove il delta tra due somministrazioni pesa più del valore assoluto in una singola seduta. È l'applicazione più vicina alla maturità operativa, perché la letteratura di riferimento è ampia e le cinque dimensioni vi trovano una rilevanza diagnostica già documentata; resta da costruire il database normativo su parlanti italiani, condizione perché ogni punteggio trovi una soglia interpretabile.

In linguistica forense il modello apre due strade distinte. La prima è la valutazione della capacità di intendere e di volere in soggetti con sospetto deterioramento cognitivo, dove il profilo linguistico integra la valutazione clinica tradizionale con un dato oggettivo e riproducibile, prodotto secondo una formula dichiarata anziché affidato all'impressione del singolo consulente. La seconda è l'attribuzione d'autore in senso

proprio, discussa nel capitolo dedicato alla filologia deautoriale: qui il modello non misura una patologia ma la tenuta di uno stile, e la sua caduta, offrendo alla consulenza tecnica un metodo tracciabile in ogni suo passaggio, requisito decisivo per reggere il vaglio del contraddittorio.

Un terzo campo, più distante dal perimetro clinico del brevetto ma coerente con la sua logica, riguarda l'analisi del testo come indicatore comportamentale in ambito di prevenzione della radicalizzazione, contrasto al terrorismo e intelligence. Se un profilo linguistico stabile nel tempo permette di intercettare un deterioramento cognitivo prima che diventi manifesto, lo stesso principio, applicato a un corpus di produzioni testuali o discorsive di un soggetto o di una comunità online, permette di intercettare un percorso di radicalizzazione prima che si traduca in azione. Le cinque dimensioni offrono in questo campo un vantaggio specifico rispetto ai sistemi a punteggio composito discussi nel Capitolo 3: ogni segnale resta tracciabile alla propria formula, requisito decisivo quando l'esito dell'analisi deve reggere il vaglio di un'autorità giudiziaria o alimentare una segnalazione che un analista dovrà motivare.

Il campo di applicazione più immediato è quello dell'early warning testuale su corpora aperti: forum, canali social, produzioni scritte sequestrate o intercettate. L'irrigidimento sintattico, la riduzione della coesione argomentativa, la chiusura lessicale attorno a un nucleo semantico ristretto e la crescente tangenzialità rispetto a un contesto dato sono pattern già documentati in letteratura sulla radicalizzazione linguistica; il modello FOXP2 offre uno strumento per quantificarli con la stessa architettura a cinque punteggi elaborata per il decadimento cognitivo, riproposta qui su un oggetto diverso ma strutturalmente affine: il cambiamento nel tempo dello stile di un soggetto o di un gruppo.

Un secondo campo è quello della profilazione integrata in sede di analisi preventiva, dove l'indicatore linguistico si combina con la profilazione legale e cognitiva già in uso presso i centri specializzati: il punteggio FOXP2 diventa una componente aggiuntiva di un dossier più ampio, utile a corroborare un quadro costruito su altre fonti, dal comportamento online ai precedenti, alle segnalazioni di rete. Il modello arricchisce così il lavoro dell'analista con un dato ulteriore, misurabile e riproducibile.

Un terzo campo è quello della formazione degli operatori: forze dell'ordine, servizi di intelligence, professionisti dei centri di prevenzione possono utilizzare l'output del modello come materiale didattico, per riconoscere pattern linguistici ricorrenti nei casi

già chiusi e costruire una sensibilità diagnostica trasferibile a casi nuovi, in modo analogo a come la Parte II di questo manuale ha trattato ogni dimensione come strumento leggibile anche da chi non ha una formazione linguistica specialistica.

Un quarto campo riguarda l'integrazione del modello in piattaforme consortili di ricerca e prevenzione a livello europeo, dove la componente linguistica si affianca a moduli di screening cognitivo-vocale costruiti sulla stessa architettura a cinque dimensioni, e dove la scala a dieci punti, già discussa nei capitoli precedenti per l'uso clinico, offre un linguaggio comune tra partner di paesi e discipline diverse: giuristi, informatici, linguisti, operatori della sicurezza, chiamati a leggere lo stesso punteggio secondo gli stessi ancoraggi qualitativi.

Un quinto campo è quello dell'uso probatorio in sede giudiziaria: un profilo linguistico costruito con criteri dichiarati e formule verificabili può accompagnare la ricostruzione di un percorso di radicalizzazione con lo stesso statuto di tracciabilità che il Capitolo 3 ha rivendicato per l'uso clinico-forense del modello, offrendo alla pubblica accusa e alla difesa un dato tecnico discutibile in contraddittorio, non un punteggio opaco prodotto da un sistema non ispezionabile.

Questi cinque campi, insieme alla linguistica clinica e alla linguistica forense discusse in apertura, condividono lo stesso oggetto tecnico, le cinque dimensioni linguistiche descritte nella Parte II, e la stessa condizione di sviluppo: ciascuno richiede un database normativo costruito sul proprio dominio specifico, parlanti sani e patologici per la clinica, testi di radicalizzazione conclamata e testi di controllo per la prevenzione. È il lavoro ancora da fare perché ogni possibilità elencata diventi strumento maturo, non una premessa che le manchi.

Un modello si giudica dalla coerenza dei propri criteri, non dalla propria completezza. Se questo manuale ha un merito, è di aver reso quei criteri abbastanza espliciti da poter essere, a loro volta, controllati, in qualunque campo li si voglia applicare.