



UNIVERSITA' DEGLI STUDI DI PAVIA

DIPARTIMENTO DI SCIENZE DEL SISTEMA NERVOSO E DEL
COMPORAMENTO

CORSO DI LAUREA MAGISTRALE IN PSICOLOGIA

**Linguistic Inquiry and Word Count (LIWC) per l'analisi di
Categorie Linguistiche e Psicologiche
in una Psicoterapia conclusa**

RELATORE:

PROF. DAVIDE LICCIONE

CORRELATORE:

Dott.ssa Luisa Siciliano

Tesi di Laurea di

Gessica Valente

507367

Anno Accademico 2023/2024

*“Tu, superando le ore avvolte dal dolore,
fai brillare la nobiltà del tuo cuore.
Niente lo potrà fermare!”*

- *La Mappa della Felicità. Daisaku Ikeda.*

INDICE

CAPITOLO 1 Process-Outcome Research.....	6
1.1 Premessa.....	6
1.2 Outcome Research e Process Research.....	8
1.3 Indagine qualitativa e quantitativa.....	21
1.4 Vantaggi e limiti della Process-outcome research.....	24
CAPITOLO 2. L'intelligenza artificiale e l'indagine linguistica.....	28
2.1 Cos'è l'Intelligenza artificiale.....	28
2.1.1 Approcci scientifici all'Intelligenza artificiale.....	33
2.2 Natural Language Processing (NLP).....	35
2.2.1 Approcci a vocabolario.....	36
2.2.2 Approcci a vocabolario chiuso.....	37
2.2.3 Approcci a vocabolario aperto.....	39
2.3 LIWC: Linguistic inquiry and word count.....	41
2.3.1 Il significato sociale e psicologico delle parole alla base di LIWC.....	43
CAPITOLO 3. Tratti di personalità e Text Mining.....	48
3.1 Tratti e stili di personalità.....	48
3.2 Principali sistemi di valutazione di tratti e disturbi di personalità.....	49
3.2.1 Il modello di Eysenck.....	50
3.2.2 Il modello di Cloninger.....	51
3.2.3 Il modello di Zuckerman.....	52
3.2.4 Il modello dei Big Five Factors.....	52
3.3 La Text Mining applicata allo studio dei tratti di personalità.....	54
3.4 La Text Mining applicata allo studio di utenti Facebook.....	59
3.5 Tratti di personalità: Utilizzo di LIWC.....	60
3.6 Conclusione.....	69

CAPITOLO 4. La Ricerca.....	71
4.1 Considerazioni preliminari.....	71
4.1.1 Analisi testuale ed ermeneutica.....	73
4.1.2 Il Topic come unità di significato.....	76
4.2 Il Caso di I.....	80
4.3 Ipotesi e obiettivi della ricerca.....	81
4.4 Il metodo.....	83
4.4.1 Creazione del Dataset.....	83
4.4.2 L’assegnazione dei Topic.....	87
4.4.3 Il software: LIWC.....	88
4.4.4 Le analisi con Rstudio.....	93
4.5 Analisi dei colloqui: Un approccio integrato	94
4.5.1 Analisi dei Colloqui dal 1° al 5°: Risultati e Interpretazioni.....	94
4.5.2 Analisi dei Colloqui dal 6° al 10°: Risultati e Interpretazioni.....	103
4.5.3 Analisi dei Colloqui dall’11° al 15°: Risultati e Interpretazioni.....	110
4.5.4 Analisi dei Colloqui dal 16° al 20°: Risultati e Interpretazioni.....	114
4.5.5 Analisi dei Colloqui dal 21° al 25°: Risultati e Interpretazioni.....	117
4.5.6 Analisi dell’andamento della terapia attraverso i punteggi LIWC: Approccio longitudinale.....	120
4.6 Conclusioni: Limiti e prospettive future.....	127
BIBLIOGRAFIA.....	130

CAPITOLO 1

PROCESS-OUTCOME RESEARCH

1.1 Premessa

La psicoterapia è stata definita da Brent e Kolko (1998) come “una modalità di trattamento dove il terapeuta e il paziente lavorano insieme per migliorare le condizioni psicopatologiche e la compromissione funzionale attraverso la focalizzazione su: a) relazione terapeutica; b) attitudini del paziente, pensieri, affetti e comportamenti; c) contesto sociale e lo sviluppo.

Discutendone ampiamente, Liccione (2019) la definisce:

Un'esperienza clinica relazionale, attiva e responsabilizzante, che ha l'obiettivo di ripercuotersi nell'esistenza del paziente per orientarlo verso una forma migliore [...]. È necessario distinguere l'intervento psicoterapeutico dalle esperienze relazionali non cliniche come lo possono essere il dialogo amicale o familiare; dalle attività cliniche simili come i colloqui psicologici di sostegno e la riabilitazione neuropsicologica; e da terapie come quella farmacologica, di magnetica transcranica ecc. [...]. La psicoterapia ha il compito di riposizionare il paziente rispetto ai suoi modi di essere disfunzionali e l'obiettivo è quello di curarlo, guidandolo verso nuove aperture di mondo. Il paziente deve necessariamente essere attivo nel processo di cura e guarigione e quindi pronto a prendersi le sue responsabilità e impegnarsi nel seguire le indicazioni terapeutiche. Il processo terapeutico poi, non si risolve con la mera spiegazione di un'esistenza ma deve necessariamente, portare a galla delle nuove progettualità e modi di vedersi nel mondo, conducendo il paziente a tendere verso la sua forma migliore. (Liccione, 2019, pp. 260-262).

Friederick van Eeden (1889), un noto psichiatra e scrittore olandese, è stato il primo ad introdurre il termine psicoterapia per indicare un metodo di cura delle funzioni psichiche di un individuo, mediante processi psicologici.

A partire dagli anni '60, dopo aver risposto alla domanda “Che cos'è la psicoterapia?”, il quesito centrale nella comunità dei ricercatori è diventato “Qual è la psicoterapia più

efficace?” (Bergin e Garfield, 2021). Più nello specifico, Gordon Paul, nel 1969 scrive una delle domande più citate sull’obiettivo della scienza evidence-based, “Quale trattamento, da parte di chi, è più efficace per uno specifico paziente con uno specifico problema, e in quale specifico contesto? (Paul, 1969, p.44)”. Durante il periodo iniziale di sviluppo della terapia comportamentale, Gordon Paul, allora a pochi anni dal suo dottorato di ricerca, pose una delle domande più citate sull’obiettivo corretto di una scienza basata sull’evidenza. Questo incitò un nuovo approccio scientifico: interventi testati per specifiche aree problematiche che si adattano ai bisogni degli individui, sulla base di processi noti di cambiamento. Questo inizio promettente non si estese abbastanza perché i primordi della terapia comportamentale si basavano su principi e teorie dell’apprendimento che erano in gran parte tratti dal laboratorio animale. I primi terapeuti del comportamento presumevano che i laboratori potessero essere affidabili per tracciare i principi di cambiamento necessari per la scienza (Franks & Wilson, 1967).

A partire dalla seconda metà del secolo scorso, dunque, si è acceso il dibattito sul tema della ricerca in psicoterapia. Nel rapporto tra psicoterapia e scienza, ci sono due estremi che bisogna evitare: lo scientismo e lo sciamanesimo pseudoscientifico. Nel primo estremo troviamo coloro che pur di adottare il metodo delle scienze naturali eliminano l’essenza dell’uomo stravolgendola in un dominio di impersonali variabili computabili (Liccione, 2019). Dall’altra parte ci sono coloro che rigettano qualsiasi tentativo di valutare oggettivamente il processo psicoterapeutico e il suo outcome e rifiutano qualunque forma di falsificazione empiricamente orientata/verifica empirica dei propri/adottati modelli teorici. Si rischia così di ridurre la psicoterapia a un dominio di giochi relazionali, frequentemente basati su strutture interpersonali suggestive, come se la psicoterapia fosse ancora una pratica sciamanica. (Liccione, 2019). Vi è quindi la necessità di mettere a punto degli strumenti di valutazione del processo psicoterapeutico che trattino l’esperienza umana secondo modi scientifici (quindi astratti e categoriali), senza svuotare di senso l’oggetto (soggetto) di studio, ossia l’esperienza della persona (Liccione, 2019).

Il campo si è ora sviluppato a sufficienza per tornare a una forma ampliata della visione originale di Paul, per cui la psicoterapia moderna e la scienza si concentrano su una nuova domanda fondamentale: “Quali processi biopsicosociali di base dovrebbero essere mirati con questo paziente, dato questo obiettivo in questa situazione, e come possono essere cambiati nel modo più efficiente ed effettivo?”. Questa area di studio non riguarda solo prove di verifica e valutazione dell’efficacia o meno della psicoterapia (psychotherapy

outcome research), ma anche come e perché avviene il cambiamento in psicoterapia (psychotherapy process research).

La ricerca sperimentale, che verrà descritta nel corso dei successivi capitoli, si pone l'obiettivo di contribuire al filone della Process-Outcome Research per cui, per una comprensione ottimale, è utile introdurre che cos'è, come si è evoluta nel corso del tempo e evidenziarne principalmente obiettivi e vantaggi, non tralasciando eventuali potenziali limiti. Se si vuole aumentare l'efficacia della psicoterapia, dobbiamo capire esattamente quali processi portano a risultati migliori per i clienti. Questa è la sfida principale e il potenziale contributo della process-outcome research in psicologia clinica.

1.2 Outcome Research e Process Research

I termini “process” e “outcome” sono stati definiti nella ricerca psicoterapeutica in molti modi. Il risultato, generalmente, significa un cambiamento nel comportamento, nelle esperienze o nelle caratteristiche del cliente dopo un intervento terapeutico. Il termine processo originariamente si riferiva agli eventi che accadono durante le sessioni psicoterapeutiche che conducono al cambiamento del cliente (Greenberg, 1986). Con il concetto di processo ci riferiamo allo studio dell'interazione tra il cliente e il terapeuta; l'interazione riguarda sia ciò che avviene nella seduta, sia ciò che si sviluppa fuori (il pensare al terapeuta/cliente, l'eseguire compiti assegnati, etc.), sia gli incontri vis a vis in condizioni sperimentali, virtuali, etc. Con interazione stiamo contemplando tanto il processo quanto il suo esito immediato, a medio e a lungo termine. Fin dalla fine degli anni '60 a guidare la ricerca in psicoterapia vi era l'interrogativo su quale trattamento fosse più efficace per un tipo di individuo con un certo disturbo specifico (Paul, 1967), e infatti molti filoni di indagini si focalizzarono sul “cosa funziona per chi” in un'ottica però molto causalistica. In ciò però viene a mancare l'attenzione al “come”, quindi all'osservazione di tutto ciò che accade verso il risultato. Ma “prima di poter spiegare, forse dobbiamo descrivere cosa succede” (Kazdin, pag. 541).

In effetti, sulla base di questo fattore centrale, la ricerca in psicoterapia si è sviluppata lungo tre ambiti: la ricerca sui risultati (Outcome Research), sul processo (Process Research) e sul processo in relazione ai risultati ottenuti (Process-Outcome Research).

Dato il numero di opzioni di trattamento che esistono per problemi di salute mentale e comportamentale (ad esempio, farmacoterapia, auto-aiuto, supporto tra pari, nessun

trattamento), forse la prima domanda che dovrebbe essere posta è se la psicoterapia funziona ovvero produce cambiamenti. Dato che il cambiamento può essere definito in molti modi diversi e i gruppi e le impostazioni di confronto possono differire, questa domanda apparentemente semplice è in realtà piuttosto complessa. A questo proposito, in psicoterapia l'Outcome Research si concentra sul valutare se la terapia funziona, osservando il cambiamento del paziente in relazione al trattamento a cui si è sottoposto (Gelo, Auletta & Braakmann, 2010). Tale cambiamento può essere misurato, ad esempio, tramite l'utilizzo di strumenti standardizzati pre- e post- terapia. Il cambiamento nella psicoterapia, per citarne alcuni, può essere rappresentato dal livello di soddisfazione del cliente, da una riduzione dei sintomi, da un aumento del funzionamento, dal miglioramento dell'alleanza terapeutica, della speranza per la vita.

Inoltre, il cambiamento può essere valutato da più prospettive, tra cui misurazioni comportamentali e / o fisiologiche da parte del cliente, del terapeuta e / o di osservatori esterni. La maggior parte concorda sul fatto che una combinazione di prospettive e domini di esito dovrebbero essere inclusi nella ricerca sui risultati della psicoterapia al fine di valutare con precisione se si è verificato o meno un cambiamento. Inoltre, l'outcome research deve dimostrare cambiamenti clinicamente significativi nei risultati misurati, piuttosto che solo differenze statistiche. La maggior parte concorda anche sul fatto che l'efficacia della psicoterapia deve essere dimostrata sia in studi altamente controllati che in contesti più naturalistici. Un certo numero di metodi è stato utilizzato per testare gli effetti della psicoterapia e degli interventi individuali, compresi i disegni pre-post, gli studi controllati con placebo e i confronti tra i vari trattamenti.

Il primo contributo dell'outcome research può essere rintracciato agli arbori della psicologia, quando Freud negli ultimi anni dell'Ottocento descrisse in che modo l'ipnosi e la tecnica interpretativa fossero efficaci nella riduzione sintomatologica della sua paziente Anna O. (Breuer & Freud, 1892-95). Nel corso poi degli anni '20 si hanno i primi tentativi di dimostrare l'efficacia delle psicoterapie psicoanalitiche ottenendo però risultati imprecisi, i quali successivamente furono criticati da Eysenck nella sua rassegna del 1952.

Costruire un disegno di ricerca (o piano) vuol dire mettere a punto un insieme di elementi e di procedure che consentono di dire, con la più alta sicurezza, che gli effetti riscontrati nella variabile dipendente sono causati dalla manipolazione di quella indipendente, e non dall'influenza di variabili non controllate. Il disegno di ricerca costituisce l'ossatura stessa di un'indagine empirica e fonda l'affidabilità dei risultati.

Nello studio dell'outcome vengono adottati principalmente disegni di ricerca sperimentali e semi-sperimentali, sviluppatasi a partire dai primi anni '50. Entrambi si pongono come obiettivo quello di individuare i rapporti di causa-effetto tra una variabile indipendente, ovvero i fattori che lo sperimentatore manipola come il trattamento, e una variabile dipendente (ad esempio, la sintomatologia dei pazienti), misurata al *passare del tempo* (ad esempio, tra l'inizio e la fine della terapia). Inoltre, entrambi gli approcci implicano il confronto di almeno due gruppi, il cui numero dipende dai valori (detti anche livelli o condizioni) che la variabile indipendente può assumere. Nella ricerca sull'esito, tali valori sono due (0 = assenza di trattamento, 1 = trattamento); nel caso della ricerca comparata sull'esito – che intende confrontare l'efficacia di più trattamenti tra loro – sono invece possibili n livelli in base al numero di trattamenti che si vogliono confrontare (ad esempio: 0 = assenza di trattamento, 1 = terapia psicoanalitica, 2 = terapia cognitivo-comportamentale, ecc.). I disegni di ricerca sperimentali consentono allo studioso di avere un controllo completo su tutte le variabili della ricerca: *sul chi, su cosa, sul quando, sul dove e sul come*. Il gruppo che riceve il trattamento viene denominato *gruppo sperimentale*; il gruppo non sottoposto a trattamento viene invece chiamato *gruppo di controllo*.¹

Il controllo sul “chi” comporta che si possano assegnare in maniera casuale i soggetti alle varie condizioni sperimentali come, ad esempio, chi farà una terapia e chi no oppure a quale tipologia di terapia saranno sottoposti. Il controllo sul che cosa, quando, dove, come permette di padroneggiare tutte le modalità di esecuzione dell'esperimento, cioè il momento dell'esecuzione, il luogo, la strumentazione, ecc. per questo alto livello di controllo, il loro grado di costrizione è molto elevato. Negli studi *sperimentali* (che nella ricerca clinica e psicoterapeutica vengono anche chiamati Randomized Controlled Trials – RCT) è possibile la randomizzazione, per cui è di conseguenza possibile “formare” tanti gruppi quanti sono i livelli della variabile indipendente. Ad esempio, è possibile assegnare casualmente i pazienti che si presentano in un ambulatorio psicoterapeutico alle varie condizioni sperimentali, cioè stabilire casualmente chi farà una terapia e chi no.

I quasi- esperimenti (chiamati anche Preferential Trials), invece, non controllano tutte le condizioni, perché i soggetti possono essere sottoposti alle varie condizioni sperimentali

¹ Nella ricerca in psicoterapia, il gruppo di controllo dovrebbe essere idealmente sottoposto agli effetti di una qualche attività il più possibile simile al trattamento psicoterapeutico di cui si vuole studiare l'efficacia, ma che non contiene nessun “ingrediente” attivo nel trattamento in esame – in analogia con la condizione placebo della ricerca medica. La strategia tipicamente utilizzata per costituire un gruppo di controllo nella ricerca in psicoterapia è quella dell'utilizzo delle liste d'attesa: ai soggetti viene comunicato che saranno sottoposti ad un trattamento psicoterapeutico, che però non avrà inizio fino alla fine dello studio.

solo in ragione di raggruppamenti già costituiti che si sono formati in base a criteri non controllati dal ricercatore (ad esempio, nello studiare la terapia psicoanalitica e cognitivo-comportamentale, sono il consiglio del medico di base, le inclinazioni personali del paziente, l'esito delle prime interviste, ecc. a stabilire chi farà una terapia piuttosto che un'altra). Il ricercatore cioè non può manipolare a suo piacimento i livelli della variabile indipendente e proprio perché sono basati su gruppi già esistenti, i quasi esperimenti talvolta vengono chiamati anche *ex post facto*.

L'analisi dei dati generalmente viene condotta utilizzando l'analisi della varianza (ANOVA) che valuta gli effetti su una variabile di interesse (variabile dipendente-risposta) di uno o più fattori di controllo (variabili indipendenti categoriali con due o più modalità) con l'obiettivo di valutare se esiste una relazione significativa tra i cambiamenti della variabile indipendente, la quale può essere rappresentata dalla sintomatologia esperita dai pazienti, misurata entro soggetti, andando a confrontare i dati raccolti pre-trattamento con quelli registrati alla conclusione dello stesso, e tra soggetti, confrontando quindi i risultati ottenuti dal gruppo sperimentale e da quello di controllo (Gelo et al., 2010). L'outcome research in psicoterapia non deve solo dimostrare differenze statistiche ma anche e soprattutto cambiamenti clinicamente significativi nei risultati misurati. Inoltre, la maggior parte concorda anche sul fatto che l'efficacia della psicoterapia deve essere dimostrata sia in studi altamente controllati che in contesti più naturalistici. Quando si parla di osservazione condotta in ambiente naturale non significa che il ricercatore sia svincolato da qualsiasi regola o che il suo lavoro sia asistemico e privo di ipotesi. Anzi, questo metodo comporta una selezione attenta delle osservazioni; infatti, una delle caratteristiche fondamentali dell'osservazione naturalistica è la sistematicità. Il ricercatore, cioè, da un'ampia quantità di interazioni comportamentali, sceglie solo determinati aspetti, quelli che suppone siano più strettamente legati all'ipotesi da verificare. Inoltre, tra tutti i metodi descrittivi, l'osservazione naturalistica è la tecnica che maggiormente rispetta il fluire naturale dei comportamenti ma non è utilizzabile quando il ricercatore vuole individuare le cause di un comportamento; dato che ogni comportamento può essere il prodotto di più variabili che operano indipendentemente o in combinazione, essa non ha i mezzi per isolarle e comporta tempi molto lunghi (cfr. A. Pedon & A. Gnisci 2016).

La ricerca sull'esito misura l'efficacia in riferimento a diverse variabili dipendenti, che solitamente includono diversi aspetti della sintomatologia e della personalità del paziente così come la qualità delle sue relazioni interpersonali o il suo grado di funzionamento. I

primi strumenti utilizzati a tal fine sono questionari *auto-somministrati*. Tra questi ricordiamo lo Shapiro Personal Questionnaire (Shapiro, 1961) e i più noti Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, e Erbaugh, 1961; nel 1996 la BDI-II), e la Symptom Check List 90-Revised (SCL-90-R; Derogatis, 1977). A questi seguiranno, in tempi recenti, l'Inventory of Interpersonal Problems (IIP; Horowitz, Rosenberg, Baer, Ureño, e Villaseñor, 1988), il Clinical Outcome Routine Evaluation-Outcome Measure (CORE-OM; Evans, et al., 2000) e l'Outcome Questionnaire-45 (OQ-45; Lambert et al., 2004).

Nelle prime fasi della ricerca sull'esito, tali strumenti vengono applicati all'*inizio* del trattamento (*pre-treatment*), al *termine* di esso (*post-treatment*), ed eventualmente in un periodo *successivo al termine* del trattamento (*follow up*). Dunque, nella tradizionale ricerca sull'efficacia della psicoterapia, l'esito viene inteso esclusivamente nei termini del miglioramento clinico del paziente osservabile al *termine* del trattamento (detto anche esito *finale*² o "*big-O*"; Greenberg, 1986)². Si possono impiegare anche altri strumenti osservazionali, i quali solitamente sono utilizzati da terapeuti che hanno ricevuto un training specifico. Tra questi strumenti quello maggiormente utilizzato è la Hamilton Rating Scale for Depression (Hamilton, 1960).

I disegni di ricerca utilizzati nell'outcome research rischiano di "appiattire" le differenze individuali ed è di tipo estensivo³, ed ha il merito di garantire la validità interna dei risultati e la loro generalizzabilità alla popolazione da cui il gruppo sperimentale è tratto – a condizione, ovviamente, che il campione sperimentale sia rappresentativo. Tuttavia, è importante sottolineare che i dati analizzati all'interno di un disegno di ricerca del genere rappresentano la media di una "performance" del gruppo di soggetti studiati (cioè, sottoposti a psicoterapia) e non la performance di un dato soggetto. Dunque, il fatto che i soggetti studiati presentano *mediamente* dei miglioramenti significativi in seguito ad un

² Greenberg (1986) propone una distinzione tra esito finale (osservato al termine del trattamento), intermedio (osservato durante il trattamento alla fine delle sedute), e immediato (osservato durante le sedute).

³ La distinzione tra ricerca *estensiva* ed *intensiva* rispecchia, in gran parte, la distinzione tra approccio *nomotetico* (teso alla spiegazione di fenomeni attraverso le formulazioni di leggi o modelli di tipo *generale*) e approccio *idiografico* (teso alla comprensione il significato di fenomeni *singoli, individuali* ed, eventualmente, *casuali*). Nel primo, i fenomeni oggetto di indagine vengono studiati prediligendo l'*ampiezza* delle osservazioni – condotte quindi su un campione che deve essere necessariamente numeroso al fine di essere rappresentativo della popolazione di riferimento, ma in cui i soggetti vengono studiati in modo più "superficiale". Nel secondo caso, si predilige la *profondità* – ottenuta ad esempio utilizzando campioni meno numerosi ma lasciando più "spazio" ai soggetti studiati di esprimere i propri vissuti ed esperienze. Questi due approcci, solitamente considerati come due categorie distinte, rappresentano in realtà gli estremi di un continuum su cui uno specifico disegno di ricerca si può variamente collocare in base a numerosi fattori. Senza entrare nel dettaglio (si veda Gelo, Braakmann, e Benetka, 2008 per un approfondimento e discussione)

trattamento non è garanzia del fatto che *ciascun* soggetto del gruppo ha effettivamente tratto giovamento da esso. In altri termini, un disegno estensivo di tipo sperimentale può effettivamente dimostrare che un trattamento è efficace per un soggetto *medio*, ma non fornisce alcuna informazione sull'efficacia che quella terapia potrebbe avere per uno *specifico* individuo (per una discussione sulla necessità di riportare l'individuo al centro della ricerca in psicoterapia e più generalmente dell'indagine sui processi psicologici si veda Del Rio e Molina, 2008; Molenaar, 2004; Molenaar e Valsiner, 2008; Salvatore e Valsiner, 2008).

In contrasto con la semplice domanda se la psicoterapia funzioni o meno (*outcome research*), la *process research* chiede come e perché si verificano cambiamenti nella terapia. Pertanto, mentre la ricerca sui risultati della psicoterapia porta alla conoscenza di quali trattamenti dovrebbero essere utilizzati, la ricerca sul processo aiuta a identificare possibili modi per migliorare la psicoterapia al fine di aumentarne l'efficacia.

Quindi, la process research ha come obiettivo quello di studiare “cosa avviene in terapia”, osservando le dinamiche che si instaurano tra il paziente e il terapeuta (Gelo et al., 2010). Se non viene accompagnata dai risultati, la process research risulta ben poco utile (Migone, 1996). Il processo terapeutico viene tradizionalmente definito come quell'insieme di fenomeni, avvenimenti e/o attività ascrivibili al paziente, al terapeuta e alla loro relazione, osservabili nel corso di una psicoterapia (Kiesler, 1973). Lo studio del processo consiste, dunque: (i) nel selezionare un livello d'analisi che permetta di osservare ciò che accade durante il trattamento, (ii) nell'identificare quali variabili siano più adeguate per descrivere i fenomeni oggetto d'interesse al livello d'analisi selezionato, (iii) nel descrivere la frequenza o la percentuale di occorrenza con cui questi fenomeni si presentano, e (iv) nel valutare relazione tra i costrutti selezionati così come l'eventuale diversa distribuzione di questi fenomeni tra diversi gruppi o nel tempo.

Anche in questo caso, una prima forma di ricerca sui processi fu introdotta da Freud tramite la descrizione degli episodi salienti sviluppatisi durante le sedute con i pazienti, partendo dagli appunti presi durante i colloqui. Ne costituisce un esempio il caso de “L'uomo dei topi” (1909).

La moderna ricerca sul processo terapeutico nasce dal lavoro di Rogers (1942), il quale iniziò ad utilizzare in maniera sistematica i trascritti delle sedute sviluppando in questo modo l'analisi del contenuto dei trascritti terapeutici e, contemporaneamente, vennero messi a punto numerosi sistemi di codifica del comportamento verbale del paziente e/o del terapeuta dunque della loro relazione. La Terapia Centrata sulla Persona, infatti, ha

caratteristiche proprie, che sono state elencate da Rogers e Sandford (1985) tra cui una elaborazione teorica che il processo terapeutico è caratterizzato da un cambiamento nel modo di essere del cliente e nell'aumento della capacità di vivere più pienamente nel presente, un interesse maggiore per il processo di cambiamento piuttosto che per la struttura della personalità e sostiene la grande importanza della ricerca nell'aumentare le conoscenze sulla psicoterapia e l'importanza di costruire tutte le formulazioni teoriche, a partire da una base di esperienza, piuttosto che deformare l'esperienza per adattarla ad una teoria preformata.

A differenza dell'*outcome research*, la *process research* utilizza disegni di ricerca non sperimentali che prevedono l'analisi della relazione tra due o più variabili in termini di *associazioni* o *covariazioni*⁴, creando così un approccio denominato *correlazionale*. Gli studi condotti secondo un approccio correlazionale vengono anche definiti *naturalistici*, in riferimento al fatto che la rilevazione delle variabili avviene in setting naturali, cioè in cui il ricercatore può esercitare un controllo estremamente ridotto. Il termine "correlazionale" viene a volte criticato per il fatto di riferirsi ad una specifica modalità di analisi dei dati (l'indice di correlazione) che, benché molto usata in questo approccio, non rappresenta l'unico modo in cui i dati possono essere analizzati (è ad esempio possibile confrontare medie, frequenze, ecc.). Si ha così un'osservazione delle variabili nel corso delle sedute, definita come osservazione *within-session*. Ciò è dovuto al fatto che l'ipotetica variabile indipendente (che in questo caso corrisponde ad un aspetto del processo terapeutico) non viene *manipolata* – per impossibilità o per scelta – ma semplicemente *osservata*.

L'*osservazione within-session* consiste nell'analisi dei trascritti delle sedute di psicoterapia da parte di terze persone e permette di misurare certi aspetti del processo terapeutico mentre è in corso. Per l'analisi dei dati vengono utilizzati solitamente indici di correlazione, t-test per campioni indipendenti o appaiati, chi-quadro o ANOVA per campioni indipendenti o misure ripetute o, ancora, analisi della regressione (Gelo et al., 2010). Le analisi vengono condotte utilizzando dati *aggregati*⁵. Nei disegni di ricerca

⁴ Esistono una serie di procedure di analisi statistica (ad esempio, analisi della regressione e la path analysis) in cui la misura della covarianza di due o più variabili può permettere di fare inferenze relative a *possibili* rapporti di causa-effetto tra le variabili studiate. I risultati così ottenuti sono tuttavia da interpretare con cautela.

⁵ Il processo di raccolta di dati implica la rilevazione di determinate variabili in singoli soggetti (disegni trasversali), in singoli momenti (disegni longitudinali) o in entrambi (disegni misti). L'analisi di dati aggregati consiste nell'analizzare il rapporto tra due o più variabili facendo riferimento non ai dati raccolti nei singoli soggetti e/o momenti, bensì aggregando le singole rilevazioni con riferimento ai gruppi e/o ai momenti che si vogliono confrontare.

trasversali, i dati vengono solitamente aggregati in riferimento ai gruppi che si vogliono confrontare, come ad esempio il comportamento (a) di pazienti vs. terapeuti, (b) di un certo tipo di pazienti con un altro tipo di pazienti, e/o (c) di un certo tipo di terapeuti con un altro tipo di terapeuti. Nei disegni longitudinali, l'aggregazione dei dati viene fatta con riferimento ai diversi momenti che si vogliono confrontare, come ad esempio (a) le fasi della seduta (parte iniziale, intermedia e finale)⁶, (b) ciascuna seduta, e/o (c) le fasi della terapia (parte iniziale, intermedia e finale). Infine, nei disegni misti, l'aggregazione dei dati viene fatta in riferimento sia ai gruppi che ai momenti temporali.

Tali metodi aggregativi sono stati criticati in prima istanza da Kiesler (1966, 1973) in quanto fondati sul "mito dell'uniformità" [Elliott (1983); Rice & Greenberg (1984); Greenberg (1986); Greenberg & Pincus (1986); Hilliard (1993); si veda anche Gelo e Auletta (2010) in questo numero], e in successive, ricerche [Gurman, 1973; Gottman & Markman (1978) cit. in Rice & Greenberg (1984)] che ne hanno dimostrato la debolezza metodologica, spingendo diversi studiosi all'esplorazione di nuovi paradigmi. Soluzioni in questa direzione sono rappresentate dal paradigma del "change process research" (Rice & Greenberg, 1984; Greenberg, 1991) o dal "treatment process" (Stiles, Shapiro, & Elliott, 1986; Stricker & Gold, 1993; Orlinsky et al., 2004). Grazie all'introduzione di diversi strumenti (ad esempio, il Therapy Session Report di Orlinsky & Howard o il Barret-Lennard's Relationship Inventory) e al successo ottenuto nel predire l'outcome, si è rilevato che i self-report, se paragonati alle misure osservative, sono maggiormente predittivi dell'outcome.

La *process research* di questo periodo fa uso di disegni di ricerca prevalentemente *trasversali e longitudinali*. Nel primo caso, aspetti specifici del processo terapeutico vengono misurati all'interno e/o dopo le sedute considerate e poi messi a confronto in riferimento a diversi *gruppi*. Nel caso di disegni longitudinali, al contrario, specifici aspetti del processo terapeutico vengono misurati nel corso delle sedute e/o dei trattamenti, con l'obiettivo di studiare il rapporto tra le variabili oggetto di interesse in diversi gruppi *al passare del tempo*. Infine, anche se raramente, la ricerca di processo di questo periodo fa uso di disegni di ricerca *misti*, che permettono ad esempio di analizzare se il comportamento verbale di pazienti e terapeuti (variabile tra i soggetti) viene utilizzato in maniera diversa all'interno di diverse fasi del trattamento (variabile entro i soggetti).

⁶ È chiaro come, in relazione a quanto detto, l'aggregazione con riferimento a parti della seduta è possibile solo se i dati sono stati raccolti attraverso l'analisi dei trascritti.

La process research in psicoterapia affronta tre ampi domini: il corso del cambiamento, i moderatori del cambiamento e i mediatori del cambiamento. In primo luogo, studiare il corso del cambiamento ha una serie di potenziali benefici. Ad esempio, nonostante la presenza di due approcci terapeutici molto simili, si potrebbe presentare la situazione per cui uno dei due comporta potenziali benefici dopo sole poche sessioni, mentre l'altro approccio può richiedere anni. Inoltre, studiando il corso del cambiamento, i ricercatori hanno la possibilità di individuare con più efficacia quali variabili terapeutiche possono produrre miglioramenti nel cliente. Ad esempio, se il cambiamento di un cliente si verifica all'inizio del processo terapeutico, prima di qualsiasi tecnica specifica, i risultati suggerirebbero che le aspettative del cliente e l'alleanza terapeutica hanno svolto un ruolo importante nei risultati osservati. Lo studio del corso del cambiamento può anche fornire suggerimenti per interventi su misura. Il modello a fasi sostiene che un cliente, prima di sperimentare una diminuzione dei sintomi, sperimenta miglioramenti nella speranza e nel benessere, che a sua volta è seguita da un aumento del funzionamento (Howard, Lueger, Mal- ing, & Martinovich, 1993). Questo modello, empiricamente supportato, suggerisce agli psicoterapeuti di utilizzare tecniche che affrontano il livello di speranza di un cliente prima di qualsiasi intervento finalizzato al benessere. Un esempio è l'area dieci della Mental Health Recovery Star (MHRS) che fa riferimento alla percezione di fiducia personale e alle aspettative positive per il futuro prendendo in considerazione la fiducia in sé stessi, la fiducia negli altri e l'aspettativa di trovare delle possibili soluzioni al malessere.

I moderatori del cambiamento, noti anche come variabili di input, si riferiscono a variabili che predicono quali tipi di individui in quali condizioni trarranno beneficio dalla psicoterapia; quindi, precedono il trattamento e indicano per chi e sotto quali circostanze il trattamento funziona (Kraemer, Wilson, Fairburn, & Agras, 2002). Di conseguenza, essi aiutano i ricercatori a stabilire più accurati criteri di inclusione ed esclusione dei pazienti per massimizzare la potenza degli studi randomizzati e controllati (RCT). Dunque, suggeriscono ai clinici quali pazienti hanno più probabilità di rispondere al trattamento e a quali invece consigliare una cura diversa.

Esempi di variabili del moderatore includono variabili del cliente (ad esempio, età, sesso, grado di funzionamento e motivazione per il trattamento), variabili del terapeuta (ad esempio, esperienza del terapeuta), variabili di impostazione (ad esempio, ospedaliero rispetto a ambulatoriale) e variabili di trattamento (ad esempio, individuale contro gruppo, 8 sessioni contro 16 sessioni).

I mediatori del cambiamento, noti anche come variabili di processo, sono le variabili causali che stabiliscono come e perché un determinato trattamento funziona (Kraemer et al., 2002). I mediatori identificano possibili “meccanismi” attraverso cui un trattamento può raggiungere i suoi effetti. Queste variabili di processo possono riguardare il comportamento del cliente, il comportamento del terapeuta o un'interazione tra il cliente e il terapeuta. Dimostrare la causalità è molto più difficile di stabilire lo stato di mediazione. Per questo motivo è più conveniente inizialmente focalizzare la ricerca sull'identificazione dei mediatori. Successivamente solo un RCT, in cui un trattamento potenziato nelle componenti associate con il mediatore identificato mostra di essere più efficace del trattamento originale non potenziato, può stabilire se il mediatore è anche un meccanismo. Comprendere i meccanismi attraverso cui un trattamento funziona può favorire lo sviluppo di trattamenti innovativi più efficaci o meno costosi. Questa conoscenza può anche aiutare a potenziare e a rifinire le componenti attive del trattamento e a eliminare le componenti inattive o ridondanti. Infine, può aiutare anche a migliorare la nostra conoscenza sulla natura del disturbo clinico trattato. Per esempio, c'è una buona evidenza che la terapia cognitivo comportamentale (CBT) per il disturbo di panico ha successo perché elimina le cognizioni catastrofiche riguardanti i cambiamenti corporei: un'osservazione che ha fornito un buon supporto alla teoria cognitiva del disturbo di panico (Clark, 1986). In psicoterapia un esempio di mediatore è la speranza. Secondo Frank e Frank (1991), i clienti iniziano la terapia in uno stato di demoralizzazione. Il trattamento fornisce la convinzione che il percorso psicoterapeutico, tramite una serie di tecniche e procedure scientificamente valide, possa aiutarli a riprendersi dal loro disagio emotivo. Questa convinzione si traduce quindi in cambiamenti effettivi o risultati migliori. Un altro esempio di variabile mediatrice riguarda l'alleanza terapeutica e il modo attraverso cui potrebbe migliorare la speranza nel paziente durante un percorso psicoterapeutico. La psicoterapia fornisce un ambiente in cui un terapeuta non giudicante è di supporto al cliente. Se il paziente percepisce il supporto del terapeuta, sarà più semplice instaurare l'alleanza terapeutica e rafforzare la speranza. La percezione che il paziente ha del terapeuta è determinante per l'esito della terapia, infatti, può rivelarsi determinante ai fini di buon esito se il terapeuta viene percepito come interessato al suo benessere, come competente ed affidabile.

Queste caratteristiche svilupperanno una relazione terapeutica positiva che produrrà sicuramente dei benefici per cui il paziente avvertirà la possibilità di “esistere altrimenti” (Liccione, 2016) al di là di quello che sarà stato il tipo di trattamento utilizzato.

Dunque, nella process research, il processo terapeutico viene considerato un fenomeno omogeneo ovvero presenta minime variazioni nel corso delle sedute. Tale assunto, che rappresenta un altro aspetto del mito dell'uniformità nella ricerca in psicoterapia denunciato da Kiesler (1966, 1983), rivela la propria infondatezza innanzitutto da una prospettiva clinica. Nel corso di una psicoterapia, infatti, ci troviamo spesso di fronte a improvvise o inattese variazioni nel modo del paziente di riportare la propria esperienza emotiva, o nella sua abilità a riflettere su di essa. Tali variazioni non devono essere "appiattite" dall'utilizzo di valori medi, ma devono essere adeguatamente misurate nel corso dell'intero trattamento. Inoltre, da un punto di vista metodologico, tale assunto sostiene implicitamente la convinzione che le variabili di processo misurate avrebbero un significato (cioè una funzione psicologica e/o relazionale all'interno della psicoterapia che viene studiata) costante nel tempo e indipendente dal contesto.

Si impiegano procedure di analisi statistica che fanno uso di valori medi aggregati tra le osservazioni compiute. Ciò determina una serie di conseguenze negative fondamentali: si corre il rischio di sovrastimare o sottostimare l'entità dei comportamenti osservati e/o della loro relazione. Ad esempio, l'utilizzo di valori medi può impedire l'osservazione di valori eventualmente estremi, mentre la pratica psicoterapeutica, al contrario, ci insegna che variazioni "atipiche" nel comportamento del paziente possono essere particolarmente significative da un punto di vista clinico. Infine, con riferimento agli indici di correlazione, è possibile che una correlazione positiva (o negativa) tra i valori aggregati (tra i gruppi o entro i soggetti) di due determinate variabili possa oscurare una serie di correlazioni negative (o positive) che le due variabili potrebbero presentare all'interno di ciascun gruppo o di ciascun soggetto.

Riguardo il rapporto *tra outcome e process*, è sempre stata proposta una visione dicotomica dell'esito e del processo terapeutico. Il primo viene infatti identificato con il cambiamento del paziente misurato quasi esclusivamente "al di fuori del trattamento" (prospettiva *extra-therapy*) in relazione ad un qualche criterio clinico-diagnostico, mentre il processo terapeutico riguarda ciò che avviene "all'interno delle sedute" (prospettiva *within-therapy*). In realtà, processo ed esito rappresentano due costrutti euristici che possono essere utili per individuare e descrivere alcune delle dimensioni essenziali della psicoterapia, ma che non devono per questo essere reificati. Inoltre, la pratica clinica ci insegna che, ciò che da una determinata prospettiva potrebbe essere considerato un aspetto del processo terapeutico, da un'altra prospettiva potrebbe rappresentare un aspetto dell'esito. Ciò è tanto più vero quanto più la scala d'osservazione si riduce lungo un

continuum che va dal livello *macro* (la terapia e le fasi che la costituiscono) al livello *micro* (le sedute e i diversi momenti all'interno di esse), che porta quasi a far scomparire una netta distinzione tra ciò che è processo e ciò che è esito. Ad esempio, l'insight terapeutico osservato in specifici momenti di una o più sedute può essere considerato sia un aspetto del processo (ciò che succede nel corso della seduta) che un aspetto dell'esito (ciò che viene prodotto dagli avvenimenti che hanno luogo nel corso della seduta).

Il fatto che l'esito e il processo terapeutico vengano considerati come due aspetti indipendenti della psicoterapia giustifica, di fatto, una rigida e netta separazione tra la ricerca sull'esito e la ricerca sul processo. È come se ciò che accade durante la terapia non avesse a che fare con l'esito osservato al termine di essa. Il tipo di ricerca empirica associata a questo modo di concepire l'esito e il processo terapeutico conduce così a dei risultati abbastanza sterili. La ricerca sull'esito, infatti, pur consentendoci di affermare che una terapia è efficace, nulla ci dice su meccanismi responsabili di tale efficacia. Al contrario, la ricerca sul processo ci permette di valutare cosa accade nel corso delle sedute, ma non è in grado di fornire informazioni relative alla significatività clinica di questi fenomeni.

Una scienza della psicoterapia dovrebbe prefiggersi l'obiettivo di spiegare in che modo ciò che accade durante le sedute può o meno produrre cambiamenti clinicamente significativi. Per far ciò, è necessario sviluppare strategie di ricerca che permettano di investigare in modo sistematico il rapporto tra processo ed esito terapeutico. In effetti, si è iniziato ad osservare un crescente interesse per il rapporto tra il processo e l'esito terapeutico, interesse che sarebbe poi esploso a partire dai primi anni '80. Ciò è avvenuto perché alcuni clinici e ricercatori hanno iniziato ad interessarsi e a porsi domande sul "come e perché la terapia funziona". Ciò ha dato origine alla *process-outcome research*, che rappresenta un primo tentativo di collegare il processo all'esito terapeutico.

Quando il processo viene studiato prendendo in considerazione anche l'esito, si parla di *Process-Outcome Research*.

I due campi di ricerca sono arrivati a polarizzarsi tanto che il processo viene generalmente ignorato negli studi sui risultati e spesso negli studi sui processi si suppone che il risultato sia irrilevante: questo complica la già articolata e difficile indagine sul cambiamento in psicoterapia. Volendo indagare però i fattori comuni al cambiamento, ci muoviamo entro un'integrazione tra questi due filoni. Nel quadro attuale sul cambiamento in psicoterapia si possono identificare più componenti, almeno tre principali: i processi di cambiamento, i meccanismi di cambiamento e poi i risultati finali (Doss, 2004).

La process-outcome research può anche essere definita come l'indagine scientifica sulla relazione tra ciò che avviene durante la psicoterapia e i suoi effetti clinici (Orlinsky et al. 2004; Elliott 2010); quindi, si concentra sulla relazione tra il processo terapeutico (considerato come processo di trattamento o processo di cambiamento) e l'esito terapeutico.

La nascita del POR, negli anni '50, è stata motivata dalla crescente consapevolezza dei ricercatori che la psicoterapia dovrebbe mirare a identificare quali aspetti del processo psicoterapeutico possono essere considerati responsabili dell'esito del trattamento dunque comprendere il perché la psicoterapia funziona. Ciò ha portato ad un'integrazione *dell'outcome research e della process research* (Lambert 2013).

L'obiettivo si è spostato progressivamente verso lo studio dei "micro-processi" terapeutici considerando ad esempio come unità singoli enunciati o turni conversazionali paziente-terapeuta, con l'intenzione di comprendere ciò che effettivamente avviene a livello di processo terapeutico, inteso come una seduta o come intero percorso (Migone, 1996; Llewelyn et al., 2016).

Nella process-outcome research possono essere studiati e analizzati innumerevoli variabili del processo tra cui: (a) i processi del terapeuta (ad esempio, la terapia specifica e le eventuali tecniche terapeutiche) che facilitano il cambiamento del cliente; (b) i processi del cliente (ad esempio, le tipologie di azioni del cliente, gli argomenti trattati in seduta, o il livello di motivazione) che facilitano l'esito positivo della terapia; c) i processi interpersonali tra il terapeuta e il cliente che facilitano il cambiamento (ad esempio, empatia e alleanza terapeutica); e (d) il contesto (ad esempio, sociale e storico o i contesti sociali, culturali e politici).

Oltre ad esplorare questi diversi elementi di processo, i ricercatori possono studiare gli effetti (cambiamenti immediati o successivi) di una qualsiasi variabile di processo all'interno o tra le sessioni. Ognuna di esse può essere esaminata dal punto di vista del cliente, del terapeuta o di un terzo osservatore e così via.

Per uno studio adeguato del processo terapeutico e dell'outcome, le variabili di interesse devono essere misurate il più spesso possibile nel corso del trattamento. Inoltre, Si fa spesso uso di un'arbitraria suddivisione della terapia e/o delle sedute. Quando si vuole studiare l'andamento delle variabili di interesse al passare del tempo, si è soliti suddividere arbitrariamente la terapia (quando le variabili vengono misurate dopo le sedute) o le sedute (quando le variabili vengono misurate all'interno dei trascritti delle sedute) in fasi, solitamente due o tre. Questo atteggiamento è problematico se la

suddivisione è imposta dal ricercatore. Le fasi temporali in cui è possibile suddividere una seduta e/o un trattamento dovrebbero essere determinate attraverso una strategia di tipo bottom-up, in cui il ricercatore, dopo un'attenta esplorazione dei dati, si lascia "informare" da questi nello stabilire le fasi in cui è possibile suddividere una seduta e/o un trattamento secondo un approccio più ermeneutico (Gelo, Salvatore, e Mergenthaler, in prep.). In altri termini, il tempo è uno parametro d'osservazione. Il processo terapeutico è per sua natura un fenomeno dinamico (cioè, tempo-dipendente), nella misura in cui si dispiega al passare del tempo. Si predilige lo studio del rapporto tra variabili isolate (ad esempio, il rapporto tra la frequenza degli interventi supportivi del terapeuta e i livelli di insight del paziente). Molte delle analisi tradizionali (ad esempio indici di correlazione o di regressione) presuppongono l'adesione implicita ad un modello causale lineare del tipo: "se x, allora y". Come già affermato, il processo terapeutico è infatti un fenomeno che si modifica, varia nel tempo, e diversi processi assumono un significato diverso in diversi momenti della seduta e della terapia (contestualità). Di conseguenza, sono le diverse occorrenze di specifiche sequenze e pattern di variabili, e non le occorrenze di singole variabili, a poter indicare la significatività clinica di ciò che accade nel corso della terapia. Infine, nella maggior parte dei casi si fa uso di un approccio esclusivamente quantitativo. Tuttavia, ad oggi, si propende sempre di più verso una metodologia "Theory free" per la Process-Outcome Research che unisce un approccio quantitativo ad un approccio qualitativo. Come ogni tipologia di ricerca in campo scientifico, anche la Process-Outcome Research presenta numerosi vantaggi, ma anche alcuni limiti che verranno meglio delineati nel paragrafo 1.4

1.3. Indagine qualitativa e quantitativa

Le principali tipologie di approccio all'analisi empirica del processo terapeutico sono l'analisi quantitativa e quella qualitativa (Hill e Lambert, 2004). La pratica metodologica empirico-quantitativa ha le sue radici nel paradigma del positivismo e del post-positivismo, con cui si identifica la visione della cosiddetta "received view" (visione ricevuta) della scienza (Suppe, 1977). Questa pratica metodologica attualmente è la più utilizzata nella ricerca in psicoterapia (Braakmann, 2015; Orlinsky & Russell 1994). In accordo con il paradigma scientifico di matrice post-positivista (dagli anni '50 fino ai primi anni '80), lo studio del processo e dell'esito vengono condotti all'interno di un approccio esclusivamente quantitativo. Ciò implica l'utilizzo di procedure per cui i dati

vengono raccolti attraverso scale di misurazione e/o questionari e poi analizzati attraverso statistiche. Più specificamente, i fenomeni rilevanti della psicoterapia, come, ad esempio, la sintomatologia del paziente, le strutture di personalità, i comportamenti verbali ed interpersonali visibili tra paziente e terapeuta, e i loro relativi processi sottostanti che possono essere nascosti, quali i meccanismi di difesa del paziente, i pattern relazionali, le strategie terapeutiche, possono essere indagati attraverso l'uso di questionari, o scale di punteggio (Gelo, 2012). L'espressione "ricerca quantitativa" compare regolarmente nei manuali e nei testi in cui si cerca di presentare tutto il ventaglio delle metodologie di ricerca, e viene spesso usata in contrapposizione alla ricerca qualitativa, nel formato a due colonne esaminato pocanzi.

Qualitative paradigm	Quantitative paradigm
Advocates the use of qualitative methods	Advocates the use of quantitative methods
Phenomenologism and verstehen: 'concerned with understanding human behavior from the actor's own frame of reference'	Logical-positivism: 'seeks the facts or causes of social phenomena with little regard for the subjective states of individuals'
Naturalistic and uncontrolled observation	Obstrusive and controlled measurement
Subjective	Objective
Close to the data; the 'insider' perspective	Removed from the data; the 'outsider' perspective
Grounded, discovery-oriented, explanatory, expansionist, descriptive, and inductive	Ungrounded, verification-oriented, confirmatory, reductionist, inferential, and hypothetico-deductive
Process-oriented	Outcome-oriented
Valid; 'real', 'rich', and 'deep' data	Reliable; 'hard' and replicable data
Ungeneralizable; single case studies	Generalizable; multiple case studies
Holistic	Particularistic
Assumes a dynamic reality	Assumes a stable reality

Tabella 1. «Attributes of the qualitative and quantitative paradigms» (Larsen-Freeman, Long 1991, p. 12)

A partire dagli anni '90, tuttavia, si svilupperà una riflessione critica sulla necessità di utilizzare (anche) approcci di tipo qualitativo nella ricerca in psicoterapia. All'interno dell'approccio qualitativo – che da una prospettiva epistemologica fa riferimento alla fenomenologia, all'ermeneutica e all'interazionismo simbolico (Corbetta, 2009; Gelo, Braakmann, e Benetka, 2008) – i dati vengono solitamente raccolti attraverso interviste aperte o semi-strutturate, offrendo il vantaggio di lasciar maggior "spazio" ai soggetti studiati (paziente e/o terapeuta) di esprimere il proprio personale vissuto soggettivo relativamente a quanto avvenuto nel corso di una o più sedute. Il materiale testuale così raccolto viene poi analizzato attraverso specifiche analisi qualitative del testo, che permettono di identificare sia il contenuto sia la struttura di ciò che è stato espresso e analizzati attraverso specifiche procedure di analisi qualitativa del testo (Bernard, 2009).

È stato già sottolineato nel paragrafo precedente che la process-outcome research può essere progettata per misurare processi specifici di un modello, oppure può concentrarsi su processi che si ipotizza promuovano risultati positivi tra i diversi orientamenti teorici. I processi studiati possono assumere una forma qualitativa o quantitativa, oppure una combinazione di entrambi gli approcci. Gli approcci quantitativi forniscono risultati numerici o punteggi che possono essere sottoposti ad analisi statistica e in genere coinvolgono un gran numero di clienti e partecipanti terapeuti. Sebbene storicamente la maggior parte delle ricerche di tipo process-outcome sono state correlazionali (ad esempio, correlare l'empatia dei terapeuti con le valutazioni del miglioramento del cliente), recentemente sono state sviluppate tecniche analitiche avanzate e più sofisticate (McLeod et al., 2013). La Multilevel (hierarchical) modeling, ad esempio, è ideale per analizzare i dati della process-outcome research per una serie di motivi. Innanzitutto, esso è una generalizzazione del linear modeling e, in primo luogo, consente ai ricercatori di esplorare l'impatto che ogni singolo terapeuta ha all'interno dello studio oltre ad analizzare i risultati a livello di cliente e viceversa. Può gestire tante variabili dei clienti in varie sessioni, e tante variabili dei clienti per terapeuta. Inoltre, può essere utilizzata per valutare la relazione temporale tra le variabili di processo e le variabili di esito coinvolgendo misurazioni in più punti nel corso del trattamento. Latent growth curve modeling statistics è una tecnica statistica utilizzata nel quadro della Structural equation modeling (SEM) per stimare le traiettorie di crescita. È ampiamente utilizzata nel campo della psicologia, delle scienze comportamentali, dell'educazione e delle scienze sociali. Gli studi quantitativi hanno dominato il campo della process-outcome research, ma la ricerca mista quantitativa e qualitativa svolge un ruolo importante nello sviluppo della comprensione delle relazioni tra processi e risultati. Gli approcci qualitativi offrono l'opportunità di raccogliere informazioni approfondite dai pazienti e ascoltare le loro prospettive e la loro esperienza in terapia. Gli approcci qualitativi possono coinvolgere interviste faccia a faccia, interviste telefoniche o risposte scritte (includendo un questionario), alla fine della terapia o a metà strada. Esempi di approcci qualitativi nella process-outcome research sono la Change Interview (Elliott, Slatick, & Urman, 2001) e the Helpful Aspects of Therapy Form (Llewelyn, 1988). Ambedue questi approcci incoraggiano i clienti a descrivere le cose più utili o importanti che sono accadute in una sessione appena completata e a indicare ciò che ha contribuito ad aumentare la loro disponibilità in terapia. I resoconti qualitativi dei processi di cambiamento possono essere analizzati utilizzando metodi analitici qualitativi sistematici, come la Grounded theory

(Rennie, Phillips, & Quartaro, 1988) che è un metodo di ricerca che nasce nell'ambito della ricerca sociologica ispirata al cosiddetto paradigma interpretativo, allo scopo di interpretare i processi sottesi da un determinato fenomeno. Secondo la Grounded Theory, osservazione ed elaborazione teorica procedono di pari passo, in un'interazione continua. Il ricercatore scopre la teoria nel corso della ricerca empirica, e preferibilmente dovrebbe ignorare la preesistente letteratura sull'argomento, per non esserne condizionato. L'accento in questa tecnica viene quindi posto sui dati (si dice che «lascia parlare i dati»), piuttosto che sulle teorie, le quali derivano direttamente dall'analisi dei dati, che sono locali e contestuali. L'approccio dovrebbe quindi essere il più possibile libero da pre-strutturazioni teoriche. Un altro metodo qualitativo è l'analisi fenomenologica interpretativa (Smith, Flowers, & Larkin, 2009) o la Consensual qualitative research (CQR) (Hill et al., 2005).

1.4 Vantaggi e limiti della Process-Outcome Research

Si ritiene importante illustrare i vantaggi e i limiti che derivano dall'utilizzo di questa tipologia di ricerca. Sicuramente il vantaggio più evidente è che essa fornisce una visione più ampia della terapia e dei suoi esiti rispetto alla mera Outcome Research o alla sola Process Research. Questa è una diretta conseguenza dell'unione delle due tipologie di ricerca in un'unica modalità che potremmo pertanto definire “mista”; considerare il processo terapeutico nella sua interezza consente, infatti, di fare un'analisi più attenta e dettagliata rispetto alla semplice considerazione finale circa l'esito della terapia per n. soggetti slegata dalle caratteristiche individuali che rendono ogni paziente unico. Questa osservazione porta alla luce, dunque, un secondo vantaggio, ovvero che la process-outcome research permette di tenere conto dell'individualità dei soggetti. Un ultimo, ma non meno importante, beneficio è che lo sviluppo crescente delle tecnologie ha portato alla nascita di tipologie di intelligenza artificiale che possono essere usate per svolgere queste ricerche. La Process-Outcome Research oggi si avvale di strumenti di Intelligenza Artificiale che possono fornire supporto alle analisi, permettendo anche l'individuazione di nuove prospettive di ricerca. Con l'avvento dell'Intelligenza Artificiale, sono state perfezionate tecniche di Machine Learning che hanno dato un impulso nuovo alla ricerca in diversi campi, tra cui quello clinico/medico. A seguito della scelta di usare questi nuovi strumenti si è “obbligati” a delineare uno studio che

rientri nell'area della Process–Outcome Research, in quanto i sistemi di Machine Learning si basano su input legati al processo per prevedere in termini di probabilità i possibili outcome. Questo è un enorme progresso nella ricerca perché permette di velocizzare gli studi, di analizzare aspetti che prima non potevano essere indagati e di dare un sostegno tecnologico anche negli ambiti che prima erano di dominio unicamente umano. (per ulteriori approfondimenti si rimanda il lettore al capitolo 2 di questa ricerca).

Tuttavia, la strategia di correlare un processo con l'esito del trattamento è sì alla base della process-outcome research, ma questo tipo di ricerca comporta diversi problemi metodologici (Crits- Christoph et al., 2013). Il primo è che la correlazione non implica necessariamente causalità. Infatti, se ad esempio si evidenzia una correlazione tra x e y, potrebbe certamente essere possibile che x causi y, ma allo stesso modo vi è la possibilità che invece sia y a causare x, o ancora che un terzo fattore z causi sia x che y. Questa mancanza di causalità ha portato ad una limitazione nel considerare i risultati ottenuti con la Process Outcome Research utili a livello clinico, e ciò ha contribuito all'esclusione della Process dalle reviews di psicoterapia. Molti ricercatori concordano poi sul fatto che la cosiddetta “drug metaphore” (metafora della droga), sia sottostante a molti studi di POR. Secondo questa, se una componente del processo terapeutico, come ad esempio l'empatia o i compiti terapeutici, contribuisce ad una terapia ben riuscita e quindi ad outcome positivi, allora livelli più alti di questa componente, dovrebbe produrre un outcome conseguentemente più positivo (Stiles & Shapiro, 1994). Si pensa quindi che una variabile di processo influenzi l'outcome non rispettando però il contesto nel quale essa emerge. Infatti, nella realtà dei fatti, un terapeuta modifica continuamente l'intensità degli interventi in risposta al paziente. Un altro limite nell'utilizzo delle correlazioni nella POR è il rischio che vi sia una causalità inversa. Infatti, ad esempio la correlazione tra l'alleanza terapeutica e gli outcome di una terapia è stata messa in discussione su un piano metodologico ed empirico (Llewelyn et al. 2016). Infatti, l'alleanza terapeutica è incline ad aumentare man mano che il paziente si sente meglio, e questo potrebbe implicare il fatto che degli outcome migliori portino ad un più alto livello di alleanza terapeutica (DeRubeis et al., 2005). Il rapporto tra alleanza e outcome non è così lineare (buona alleanza = buon outcome e viceversa) come potrebbe sembrare. A renderlo più complesso intervengono infatti numerose variabili: la patologia del paziente, il tipo di interventi del terapeuta, lo strumento utilizzato per la misurazione dell'alleanza, il momento in cui avviene la misurazione, ecc.

Come emerge dalle due metanalisi effettuate da Horvath (Horvath, Symonds 1991; Horvath, Bedi, 2002), il momento in cui viene misurata l'alleanza sembra avere una grande influenza sulla sua capacità predittiva rispetto all'outcome: le valutazioni compiute nelle fasi iniziali (prima-quinta seduta) e finali della terapia hanno una maggiore capacità predittiva rispetto alle valutazioni intermedie (Luborsky, 1990; Horvath, Symonds, 1991). Una possibile spiegazione di queste differenze risiede nel fatto che nelle fasi intermedie di una terapia vengono affrontate le problematiche più strutturali, causa di possibili stress alla relazione tra paziente e terapeuta e di conseguenza all'alleanza terapeutica. Anche la diversa prospettiva che può essere assunta nella POR, ovvero quella del terapeuta, quella del paziente, o quella di un osservatore esterno, può essere un limite stesso, poiché conduce a risultati differenti. Tutto ciò va purtroppo a minare la validità dell'assunto secondo il quale i fattori del processo operano allo stesso modo nonostante il punto di vista.

Inoltre, non è possibile una totale manualizzazione della ricerca poiché essa si deve adattare alla terapia, e come ben sappiamo, ogni terapia è a sé stante. Ciascuna psicoterapia dovrebbe infatti essere “fatta su misura” per il soggetto che svolge il percorso terapeutico e, per questo motivo, standardizzare la terapia e la ricerca su di essa, non permetterebbe di cogliere l'essere umano nella sua unicità, irripetibilità e nei suoi modi di essere. “La psicologia, e in particolare la psicoterapia, non possono fare a meno di considerare l'essere umano nella sua essenza [...] perché perderebbero l'oggetto (soggetto) di studio” (Liccione, 2019). Dall'altro lato, è anche vero che non bisogna rendere la ricerca troppo individualizzata, altrimenti si otterrebbero dei risultati troppo specifici e non replicabili. Queste problematiche derivano dal fatto che la metodologia scientifica si basa su categorie, che vengono stabilite affinché si possa garantire l'oggettivizzazione dei risultati e la loro universalità. Tuttavia, per natura, la tematica della psicologia sembra sfuggire al metodo scientifico, e ogni tentativo di applicare alla psicologia il metodo scientifico dissolve la tematica (Galimberti, 2019). Infatti, l'applicazione tout-court del metodo delle scienze naturali alla psicologia porta alla creazione di protocolli terapeutici che tengono in considerazione solo le caratteristiche della malattia, indipendentemente dall'individuo. Si rende evidente quindi che “il motivo principale per cui è necessario il pluralismo scientifico è che il nostro oggetto di indagine è estremamente complesso e diversificato, così come le domande di ricerca che gli studiosi di psicoterapia possono porre sono molto complesse e diverse” (Stiles, 2015). Un ulteriore limite, che si rende evidente, riguarda la strutturazione dei gruppi di

controllo: questi permetterebbero il confronto con un campione normativo. Di fatto questa è la modalità base con cui vengono svolte le ricerche scientifiche, ma nel caso dello studio dei processi terapeutici ciò non è possibile, poiché non è possibile determinare che cosa è normalizzato o meno. In particolare, non è possibile identificare una condizione “non terapeutica” e l’applicazione di un protocollo standardizzato porterebbe ad una categorizzazione del soggetto. Da queste osservazioni, emerge che questo tipo di approccio all’indagine si adatta bene alla ricerca in altri ambiti, ma non in ambito psicoterapeutico. In più, la POR è dispendiosa: sia in termini di tempo perché ci si focalizza sull’esito e sul processo intero e questo implica un lavoro impegnativo, sia a livello economico. Infatti, potremmo anche considerare come svantaggio il fatto che la ricerca in psicoterapia richieda costosi finanziamenti e purtroppo in molti casi i ricercatori si debbano accontentare di esigue pubbliche sovvenzioni. Questo comporta il dover adattare il progetto di ricerca iniziale alle risorse realmente disponibili. Si spera che in futuro la ricerca in questo ambito possa beneficiare di adeguati finanziamenti, aprendo così alla possibilità di nuove scoperte e nuovi campi di indagine, con la possibilità di una diminuzione dei tempi necessari per ottenere dei risultati. Lo studio che è oggetto di questa tesi, si propone tramite l’utilizzo di strumenti quantitativi, l’obiettivo di ottenere poi delle considerazioni qualitative, e di creare uno strumento che possa guidare al meglio i terapeuti nel loro lavoro clinico futuro.

CAPITOLO 2

L'INTELLIGENZA ARTIFICIALE E L'INDAGINE LINGUISTICA

2.1 Che cos'è l'Intelligenza artificiale

L'IA è una moderna disciplina che negli anni recenti ha fornito un importante contributo al progresso dell'intera informatica. Per questo motivo riveste un indubbio interesse per coloro che siano interessati alla valutazione delle potenzialità dell'informatica e degli effetti che il progresso di questa disciplina ha nei confronti dell'uomo. Inoltre, l'IA è stata influenzata da numerose discipline fra le quali, oltre ovviamente all'informatica, la filosofia, la matematica, l'economia, le neuroscienze, la psicologia, la cibernetica, le scienze cognitive e la linguistica. L'IA è quella disciplina, appartenente all'informatica, che studia i fondamenti teorici, le metodologie e le tecniche che permettono di progettare sistemi hardware e sistemi di programmi software capaci di fornire all'elaboratore elettronico delle prestazioni che, a un osservatore comune, sembrerebbero essere di pertinenza esclusiva dell'intelligenza umana [Somalvico 1987]. Volendo utilizzare una serie di metafore scientifiche della fisica (più efficaci a descrivere algoritmi di rappresentazione dei fenomeni naturali) possiamo, secondo Somalvico (1987), delineare alcuni principi generali nella strutturazione dei processi utilizzati per l'organizzazione di un'AI: quello che in fisica si definisce come "fenomeno", per l'AI può indicarsi come *problema intuitivo*, che corrisponde alla percezione del problema da risolvere e alla formalizzazione del problema stesso; In Intelligenza Artificiale esiste una differenza tra un problema intuitivo e la sua formalizzazione che viene denominata *problema rappresentato*. Inoltre, il fisico utilizza il modello di un fenomeno stesso al fine di derivarne, mediante un'attività inferenziale, una legge che consenta di caratterizzare le proprietà del fenomeno e di prevederne di nuove. Allo stesso modo, in Intelligenza Artificiale il risolutore automatico di problemi è in grado di operare un'elaborazione sul problema rappresentato che permetta di costruire *il problema risolto*, espresso sia dall'algoritmo risolvete il problema rappresentato sia dalla sua esecuzione, che produce effettivamente la soluzione del problema. Pertanto, in fisica ai tre concetti di fenomeno, modello e legge corrispondono, nell'intelligenza artificiale, quelli di *problema intuitivo*, *problema rappresentato* e *problema risolto*.

È interessante a questo proposito analizzare con attenzione il ruolo dell'uomo e della macchina algoritmica (l'elaboratore) nel complesso processo della risoluzione di un problema. Mentre nel caso della fisica le attività sono compiute solamente dall'uomo, nel caso dell'IA le attività sono compiute tanto dall'uomo quanto dalla macchina algoritmica. In particolare, l'IA si pone l'obiettivo di fornire alla macchina la capacità di effettuare la risoluzione automatica dei problemi, ossia il passaggio dal problema rappresentato al problema risolto. Vengono invece mantenuti come esclusivi dell'attività umana tanto il passaggio dal problema intuitivo al problema rappresentato, ossia la capacità di "formalizzazione" di un problema, quanto il passaggio dal problema risolto al problema intuitivo, ossia la capacità di "confronto" della soluzione ottenuta con l'esigenza da soddisfare, che ha rappresentato l'inizio del procedimento di risoluzione (approccio qualitativo). Considerando da un punto di vista gnoseologico tanto il caso della fisica quanto il caso dell'IA, appare chiaro che le attività in gioco possono essere articolate nel modo seguente:

1. Passaggio dal mondo della realtà (fenomeno, problema intuitivo) al mondo della conoscenza della realtà (modello, problema rappresentato). Questo passaggio, oltre che formalizzazione (in fisica) e rappresentazione (in IA), in filosofia può essere chiamato "abduzione".
2. Passaggio all'interno della conoscenza della realtà (dal modello alla legge, dal problema rappresentato al problema risolto). Questo passaggio, oltre che derivazione e formulazione (in fisica) e risoluzione ed esecuzione (in IA), in filosofia può essere chiamato "inferenza".
3. Passaggio dal mondo della conoscenza della realtà (legge, problema risolto) al mondo della realtà (fenomeno, problema intuitivo). Questo passaggio, oltre che aspettazione e previsione (in fisica) e confronto e soddisfacimento (in IA), in filosofia può essere chiamato "adduzione".

Le attività dell'abduzione e dell'adduzione possono essere chiamate con il termine complessivo di afferenza, che si contrappone all'inferenza che opera all'interno della conoscenza della realtà.

Il tema centrale dell'AI resta la sua capacità di autorealizzarsi, di auto-strutturarsi e quindi di concepire risposte sempre più articolate e complesse a partire dall'analisi di elevate quantità di dati; accesso a banche dati da cui attingere rapidamente informazioni, per poi scrivere e riscrivere algoritmi sempre più rispondenti alla realtà e alle esigenze di

risoluzione. Il percorso storico alla base dell'evoluzione dell'AI si può descrivere guardando allo schema sotto riportato (Rainsberger, 2022):

Period	Development step AI	Functional principle
<1960	Native algorithms	I repeat
<2010	Machine learning	I imitate
<2018	Deep learning	I learn
2020	Deep reinforcement learning	I learn to learn
Future	Swarm deep reinforcement learning	I contribute

Tabella 1. Evoluzione dell'AI secondo lo schema di Rainseberger (2022: 2)

È evidente che il percorso evolutivo è quello di una crescente capacità di auto-strutturazione dei sistemi d'Intelligenza Artificiale a cui seguono crescenti capacità di discriminare e selezionare le risposte più adeguate a partire da una più approfondita capacità di analisi di una elevata mole di dati. Si può così affermare che l'AI sia un procedimento automatizzato di selezione di dati per pervenire a una risposta quanto più rispondente alla realtà o al problema posto.

Un sistema di AI deve essere in grado di acquisire, rappresentare ed elaborare conoscenza (e non semplicemente dati simbolici separati dal loro significato) relativa al compito da eseguire e di applicarla in meccanismi di elaborazione intelligenti, al fine di fornire le prestazioni richieste. I sistemi di intelligenza artificiale sono pertanto caratterizzati dalla capacità di gestire, elaborare e scambiare conoscenza mediante meccanismi tipici dell'intelligenza umana, quali (ma non solamente): l'inferenza, la deduzione, il ragionamento con incertezza, il ragionamento analogico, la generalizzazione, la particolarizzazione, la valutazione di ipotesi, l'apprendimento.

Quando parliamo di Artificial Intelligence – AI ci viene automatico pensare che si tratti di una disciplina relativamente recente, in realtà essa ha ormai più di 60 anni. Fra i primi lavori in questo ambito di ricerca spicca quello di Warren McCulloch e Walter Pitts (1943), che proposero un modello di rete neurale ispirato al funzionamento del cervello umano. Mostrarono come un semplice sistema di neuroni artificiali potesse essere in grado di eseguire delle funzioni logiche basilari. Almeno in teoria, questo sistema poteva imparare nello stesso modo in cui impara l'uomo: usando l'esperienza ed eseguendo quei tentativi ed errori che rafforzano o indeboliscono le connessioni tra neuroni. Ogni neurone, interconnesso con gli altri da sinapsi, poteva rappresentare uno stato binario: "on" o "off". Donald Hebb nel 1949 ideò una regola di modifica dei pesi

delle sinapsi che rendeva possibile l'apprendimento. “La regola di Hebb” rappresenta un modello “ponte” tra apprendimento biologico e artificiale. Essa riesce a descrivere la capacità adattiva delle unità neuronali in contesti di apprendimento associativo.

Importante fu anche il primo computer a rete neurale, sviluppato da Marvin Minsky nel 1950, detto SNARC, che simulava una rete di 40 neuroni.

Tuttavia, la nascita dell'AI risale al 1956 in occasione del convegno di Dartmouth. Alan Turing, considerato uno dei padri fondatori dell'informatica moderna, pose le basi per la computabilità, calcolabilità e di quello, che in seguito, verrà definito il “Test di Turing”. Secondo il test, una macchina poteva essere considerata intelligente se il suo comportamento, osservato da un essere umano, fosse stato considerato indistinguibile da quello di una persona (Osservatori Digital Innovation, 2019). È riduttivo e per alcuni aspetti anche fuorviante incentrare sul Test di Turing (TT), l'analisi dei contributi di Alan Turing alla nascita e allo sviluppo iniziale dell'Intelligenza Artificiale (IA). Anzitutto, la funzione intesa del TT sembra essere stata fin dall'inizio puramente divulgativa: Turing aveva l'obiettivo di raggiungere un ampio pubblico di persone colte, alle quali trasmettere e illustrare la possibilità tecnologica di sviluppare macchine capaci di elaborare strutture simboliche e di manifestare comportamenti intelligenti.

L'espressione “intelligenza artificiale” viene solitamente attribuita a McCarthy e compare nel documento di proposta della Conferenza di Dartmouth del 1955. McCarthy voleva un nuovo termine neutrale che potesse raccogliere e organizzare i vari sforzi di ricerca in un unico campo, incentrato sullo sviluppo di macchine in grado di simulare in tutto e per tutto l'intelligenza umana. Venne presentato un documento di 17 pagine chiamato “proposta di Dartmouth” in cui, per la prima volta, si utilizzò il termine “Intelligenza artificiale”.

Il documento discute di alcuni temi che gli organizzatori ritenevano fondamentali per il campo di ricerca, come le reti neurali, la teoria della computabilità, la creatività e l'elaborazione del linguaggio naturale, motivando la necessità della conferenza con l'obiettivo di costruire delle “macchine intelligenti”, capaci di simulare ogni aspetto dell'intelligenza umana.

A Dartmouth i ricercatori Herbert Simon e Allen Newell presentarono quello che si può considerare il primo programma di IA: *Logic Theorist*, ideato dai due già nel 1955, che fu il primo programma euristico della storia. Per sviluppare il programma si sono avvalsi dell'aiuto del programmatore di computer *Principia Mathematica's*.

I primi anni dell'IA (1952-1969) furono segnati da una serie di grandi successi; i programmi come Logic Theorist risultavano infatti sorprendenti, dimostrando che i calcolatori non erano semplici manipolatori di espressioni matematiche ma potevano fare di più. Newell e Simon sviluppano nel 1957 il General Problem Solver (GPS), in cui implementano un processo inferenziale ispirato al modo di ragionare della mente umana. In America si sviluppò la scienza cognitiva, esattamente due mesi dopo la nascita dell'IA stessa nel 1956, durante un workshop tenutosi al MIT (Massachusetts Institute of Technology), durante il quale vennero presentati alcuni paper che mostrarono come i modelli informatici potessero essere utili in alcuni ambiti quali la psicologia della memoria, del linguaggio e del pensiero prospettico. Anche la linguistica ha contribuito allo sviluppo dell'intelligenza artificiale, mostrando che l'utilizzo del linguaggio è una prova del fatto che gli esseri umani possono essere considerati come macchine che processano informazioni, partendo dalla domanda: "In che modo il linguaggio è collegato al pensiero?". Inoltre, la linguistica condivide molto con l'IA anche oggi, poiché questi due ambiti di studio sono "nati" all'incirca nello stesso periodo, tanto che si sono sviluppati insieme, fino a dare luce ad una loro intersezione, quella della linguistica computazionale, o processamento del linguaggio naturale (Natural Language Processing, NLP). Il termine NLP (Natural Language Processing) si riferisce al settore dell'informatica, e più specificamente al settore dell'AI, che si occupa di dare ai computer la capacità di comprendere il testo e le parole nello stesso modo in cui lo fanno gli esseri umani. L'NLP combina la linguistica computazionale, la modellazione basata su regole del linguaggio umano, con modelli statistici, di machine learning e di deep learning. Insieme, queste tecnologie consentono ai computer di elaborare il linguaggio umano sotto forma di testo o dati vocali e di "capire" il suo pieno significato, compreso l'intento e il sentiment di chi parla o scrive.

Nel 1959, Oliver Selfridge al Lincoln Laboratory propose il Pandemonium, un primo programma per il "pattern recognition", modello di percezione delle lettere basato sul riconoscimento delle lettere tramite le loro caratteristiche.

Nei primi anni del 2000 la disponibilità di vastissimi database ha portato un grande contributo a molti ambiti dell'IA. Infatti, non è più necessario, per realizzare sistemi basati sulla conoscenza, inserire manualmente i dati necessari al funzionamento del programma, ma affidandosi agli algoritmi di apprendimento è sufficiente solo un'ampia base di dati. In conclusione, negli ultimi anni l'utilizzo dell'Intelligenza Artificiale (IA) è diventato massivo nell'ambito della ricerca in quasi tutti i settori del mercato, della

medicina e, in generale, della scienza sia naturale sia umanistica. Restano, tuttavia, ancora alcuni campi, come quello psicoterapeutico, in cui non si è ancora sviluppata una riflessione circa i possibili usi e abusi dell'implementazione di algoritmi di IA.

In ambito psichiatrico e psicologico tecniche di IA vengono ad oggi implementate per la farmacopea degli antidepressivi o degli ansiolitici (Horn & Weitz, 2020), ma anche per la previsione del comportamento aggressivo in base all'analisi di materiale testuale valutativo, o di scale standardizzate per la valutazione della rabbia. In psicoterapia, invece, entrano in gioco aspetti cognitivi, quali le caratteristiche della personalità del paziente, gli aspetti emotivi e comportamentali, così come le caratteristiche del terapeuta. Essendo l'oggetto di studio dell'IA un soggetto, un Esser-ci (Heidegger, 1927; Liccione 2011, 2019), per gli algoritmi diventa un compito arduo quantificare gli aspetti personologici sopracitati. L'IA serve a migliorare un prodotto o uno strumento che si rivolge alle persone, intese come esistenze, e allo stesso tempo è da esse progettato. Nello specifico di questa tesi, è stato dimostrato come L'IA possa essere un utile strumento di aiuto in ambito psicoterapeutico non dimenticando il ruolo della psicoterapia in sé come strumento di cura e di dialogo tra uno specialista e i modi-di-essere e di fare esperienza del paziente. (Liccione D., 2019). In ambito psicoterapeutico la ricerca dovrà, tuttavia, essere sempre integrata dalla valutazione qualitativa dello specialista, in quanto i dati risultanti dovranno essere interpretati alla luce sia della storia di vita dei pazienti sia dell'esperienza del terapeuta (Gelo et al., 2008; Gelo et al., 2020).

2.1.1 Approcci scientifici all'Intelligenza artificiale

Quando si parla di Intelligenza Artificiale si parla del campo di ricerca che si occupa della progettazione di sistemi mirati a dotare le macchine di una o più caratteristiche che sono considerate prettamente umane: 1. l'inferenza, 2. la deduzione, 3. il ragionamento con incertezza, 4. il ragionamento analogico, 5. la generalizzazione, 6. la particolarizzazione, 7. la generazione e la valutazione di ipotesi, 8. l'apprendimento.

Le applicazioni dei sistemi di IA sono innumerevoli e una loro completa catalogazione risulta molto difficile a causa della quantità ed eterogeneità dei settori interessati. La concezione attualmente dominante nello sviluppo dei sistemi di IA è quella dell'“agire razionalmente” che individua una visione più ampia e più complessa di intelligenza legata non solo alle capacità intellettuali, ma anche all'abilità di saper interagire in generale con il mondo. L'ambito di applicazione di nostro interesse è *la comprensione del linguaggio*:

il riconoscimento e la sintesi del linguaggio vocale dell'uomo richiedono l'elaborazione di informazioni complesse di natura fonetica, morfologica, sintattica, semantica e pragmatica, e rappresentano un obiettivo molto ambizioso.

All'interno di questo ampio scenario possiamo trovare due diverse tipologie di Intelligenza Artificiale: IA debole e IA forte.

In particolare, l'IA forte si riferisce a sistemi capaci di comportarsi in maniera del tutto autonoma a prescindere dal contesto e dal compito che gli viene assegnato, quindi è incentrata sull'agire umanamente e pensare razionalmente. La tecnica di riferimento per l'IA forte è il cosiddetto Deep Learning (apprendimento profondo), che è una sotto forma di Machine Learning poiché è basato sull'apprendimento non supervisionato, ed è un sistema che utilizza degli input per arrivare ad emulare il comportamento del cervello umano ed è costituito da strutture di algoritmi multilivello per la creazione di "reti neurali artificiali" sulla base del funzionamento biologico delle reti neurali cerebrali.

L'IA debole si riferisce, invece, a sistemi progettati per risolvere problemi specifici, simula alcune capacità della cognizione senza avere la pretesa di comprendere e replicare il funzionamento di tutto ciò che il cervello umano è in grado di fare, e la tecnica alla base è il cosiddetto Machine Learning, ovvero l'apprendimento automatico. La differenza tra *Machine Learning* e *Deep Learning* consiste nel fatto che il secondo, l'apprendimento "profondo", analizza i dati applicando continuamente logiche euristiche di natura ontologica con algoritmi a più livelli che funzionano in parallelo (Rainsberger, 2022). Maggiore è il numero di algoritmi che lavorano contemporaneamente più "profondo" è il grado di capacità di analisi e decisione che il sistema è in grado di prendere. In poche parole, il Deep Learning è una tecnica di apprendimento in cui, le reti neurali artificiali, vengono esposte a una vasta quantità di dati; queste sono, poi, in grado di imparare da sole a svolgere determinati compiti senza la necessità di un pre-processamento dei dati. Invece, il Machine Learning è l'apprendimento automatico e si riferisce, quindi, a sistemi in grado di apprendere dall'esperienza, con un meccanismo simile a ciò che un essere umano fa sin dalla nascita. Grazie all'addestramento del modello, la macchina apprende come prendere una decisione. Quando si sente parlare di "apprendimento automatico", si intende esattamente questo. Più aggiungiamo features per aumentare le osservazioni e quindi più aumentiamo l'affidabilità della decisione del modello, e più la sua rappresentazione (dal punto di vista matematico) diventa complessa, rendendo necessarie funzioni matematiche strutturate.

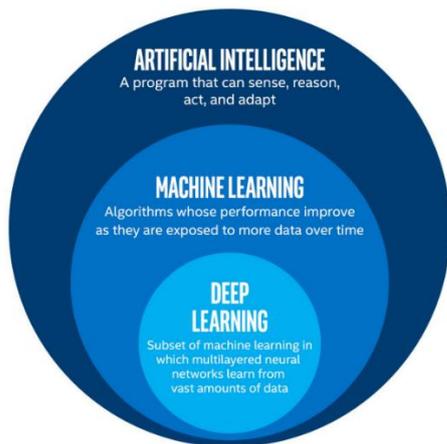


Figura 1. Classificazione concettuale delle sottocategorie dell'Intelligenza Artificiale. (Fonte: Lindekin.com)

2.2 Natural Language Processing (NLP)

Il termine NLP (Natural Language Processing) si riferisce al settore dell'informatica - e più specificamente al settore dell' AI (Artificial Intelligence)- che si occupa di dare ai computer la capacità di comprendere il testo e le parole con la stessa modalità degli esseri umani. L'NLP combina la linguistica computazionale con modelli statistici, di machine learning e di deep learning per estrarre, classificare ed etichettare automaticamente elementi di testo e dati vocali, poi assegnare una probabilità statistica ad ogni possibile significato di quegli elementi e per consentire ai computer di comprendere l'intento e il sentiment di chi parla o scrive. L'NLP è alla base di programmi informatici che traducono il testo da una lingua all'altra, rispondono a comandi vocali e riassumono grandi volumi di testo rapidamente, anche in tempo reale. Oggi l'analisi del linguaggio con l'impiego della tecnologia permette di comprendere in modo efficiente e su larga scala come le persone utilizzano e combinano le parole in modo da rilevarne i pensieri, i comportamenti e le emozioni. Un modo per svolgere tale analisi è mediante testo scritto. Negli ultimi anni il Machine Learning (ML) ha assunto un ruolo sempre più importante. Infatti, L'uso di modelli ML per prevedere i tratti della personalità ha guadagnato una notevole popolarità nel campo della *Affective Computing*, con diversi studi che si sono già impegnati a concepire modelli ML per valutazione della personalità, come vedremo nel capitolo 4. In questa sezione specifica verranno descritti i vari approcci all'elaborazione

del linguaggio naturale e lo strumento utilizzato in questa ricerca come contributo alla *Process outcome research*.

2.2.1 Approcci a vocabolario

L'analisi di un testo scritto può essere effettuata con un approccio a vocabolario:

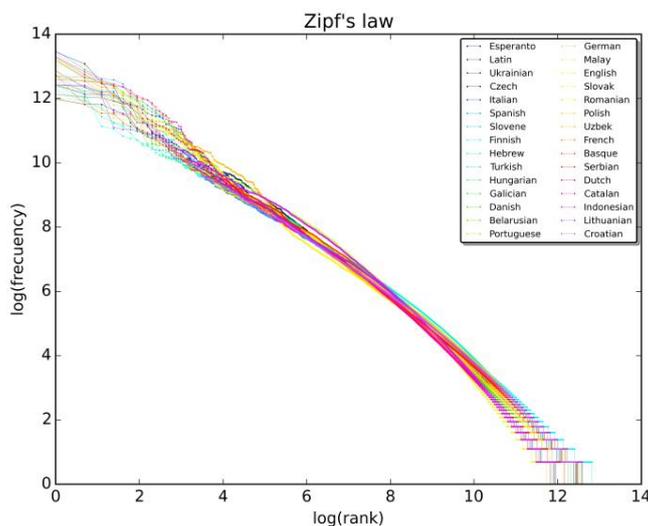
* Chiuso: dizionari o elenchi di parole associati, dai ricercatori, a categorie psico-socialmente rilevanti (come processi sociali e biologici, lavoro ed emozioni). Introdotti per la prima volta nel 1960 e incorporati in programmi informatici che consentono di scansionare automaticamente un testo, di contare la frequenza delle parole di ciascun dizionario e di produrre le frequenze relative; misure utilizzabili anche come variabili in successive analisi statistiche. Alcuni esempi di dizionari, incorporati all'interno di programmi con approccio a vocabolario chiuso, sono General Inquirer, DICTION e Linguistic Inquiry and Word Count (LIWC).

* Aperto: metodi che si basano sull'informatica, sviluppati soprattutto dal mondo accademico, ma non solo, negli ultimi due decenni, come Word Embeddings. In una definizione più rigorosa il Word Embedding è un termine complessivo che indica, nell'elaborazione del linguaggio naturale, un insieme di tecniche di modellazione in cui parole o frasi di un vocabolario vengono mappate in vettori di numeri reali. Concettualmente consiste in un'operazione matematica di "immersione" in conseguenza della quale uno spazio costituito da una dimensione per parola viene trasformato in uno spazio vettoriale continuo di dimensione molto inferiore. Queste tecniche trovano applicazione nello studio della semantica distribuzionale del discorso. Il Word Embedding (tradotto letteralmente "immersione di parole"), anche conosciuto come "rappresentazione distribuita delle parole", permette di memorizzare le informazioni sia semantiche che sintattiche delle parole partendo da un corpus non annotato e costruendo uno spazio vettoriale in cui i vettori delle parole sono più vicini se le parole occorrono negli stessi contesti linguistici, cioè se sono riconosciute come semanticamente più simili (secondo l'ipotesi della semantica distribuzionale). Questi sono caratterizzati da algoritmi, che identificano da un insieme di dati linguistici, gruppi di parole semanticamente correlate in grado di far ottenere informazioni sui campioni e formulare nuove ipotesi sui modelli. Alcuni esempi di approcci a vocabolario aperto sono Latent Semantic Analysis (LSA) e Latent Dirichlet Allocation (LDA). Studi recenti definiscono gli approcci a vocabolario chiuso più adatti per verificare in che misura sono presenti

precise categorie all'interno di un testo, fornendo così una rappresentazione compatta, oggettiva e stabile dei testi, evidenziando i concetti chiave discussi. Gli approcci a vocabolario aperto, invece, sono un modo efficace per estrarre i temi presenti in uno scritto e il significato delle parole, per chiarire cosa viene discusso nel testo e in che contesto le parole vengono usate. Su questa linea di pensiero si stanno sviluppando nuovi approcci che combinano entrambe le strategie, con l'obiettivo di analizzare i processi psicologici e come si verificano nella vita quotidiana.

2.2.2 Approcci a vocabolario chiuso

La metrica principale utilizzata negli approcci a vocabolario chiuso è la frequenza di occorrenza delle parole, soggetta, trattandosi di linguaggio naturale, alla Zipf's Law. La distribuzione di frequenza delle parole è stata un oggetto chiave di studio nella linguistica statistica negli ultimi 70 anni. Questa distribuzione segue approssimativamente una semplice forma matematica conosciuta come legge di Zipf. Zipf notò che contando le parole di un testo osservò che la parola più presente nel testo era presente almeno il doppio delle volte della parola successiva. Zipf per capirlo annotò la parola più usata e il numero di volte in cui era presente nel testo, così scorrendo con le altre parole meno ripetute, fino ad ottenere una vera e propria classifica di parole presenti nel testo (che nel suo caso era in inglese). Il fatto che la prima parola in classifica si ripeteva il doppio delle volte della seconda e così scorrendo, portò ad approfondire il discorso e mettere a confronto questa regola con altre lingue che non siano inglese. Il risultato affascinante, è stato che in tutte le lingue del mondo la legge di Zipf è applicabile. Ecco un grafico che dimostra quanto detto:



Tutte le lingue inserite nel grafico seguono la legge di Zipf in maniera molto simile (Alessandro Lenci, Università di Pisa)

Nel linguaggio naturale le parole si presentano secondo una distribuzione di frequenza notoriamente sistematica, tale per cui ci sono poche parole ad altissima frequenza che rappresentano la maggior parte dei token nel testo (ad esempio, “a”, “il”, “io”, ecc.) e molte parole a bassa frequenza (ad esempio, “fisarmonica”, “catamarano”, “ravioli”). Ciò che colpisce è che la distribuzione è matematicamente semplice e obbedisce approssimativamente a una legge di potenza nota come legge di Zipf: la r- esima parola più frequente ha una frequenza $f(r)$ che si adatta a:

$$F(r) \propto 1/R^\alpha$$

per $\alpha \approx 1$ (Zipf, 1936 , 1949). In questa equazione, r è chiamato il rango di frequenza di una parola e $f(r)$ è la sua frequenza in un corpus naturale. Poiché la frequenza effettiva osservata dipenderà dalla dimensione del corpus esaminato, questa legge stabilisce le frequenze in modo proporzionale: la parola più frequente ($r = 1$) ha una frequenza proporzionale a 1 , la seconda parola più frequente ($r = 2$) ha una frequenza proporzionale a $1/2^\alpha$, la terza parola più frequente ha una frequenza proporzionale a $1/3^\alpha$, e così via. Mandelbrot ha proposto e derivato una generalizzazione di questa legge che si adatta più da vicino alla distribuzione di frequenza nel linguaggio “spostando” il rango di una quantità β (Mandelbrot, 1953 , 1962):

$$F(r) \propto 1/(r + \beta)^\alpha$$

per $\alpha \approx 1$ e $\beta \approx 2,7$ (Mandelbrot, 1953 , 1962 ; Zipf, 1936 , 1949). Adesso però Sander Lestrade, linguista dell'Università Riboud (Paesi Bassi), propone una spiegazione: la legge di Zipf può essere spiegata dall'interazione in un testo tra la struttura delle frasi (sintassi) e il significato delle parole (semantica). Utilizzando simulazioni di computer, Lestrade è stato in grado di mostrare che né la sintassi né la semantica riescono da sole ad indurre una distribuzione zipfiana, ma la loro combinazione ha successo. «Dal momento che si usano gli articoli prima di quasi tutti i nomi, i singoli articoli appaiono più spesso dei nomi». Ma non è sufficiente: «Nell'ambito dei nomi ci sono anche grandi differenze: la parola "cosa", ad esempio, è molto più comune di "sottomarino" e può essere utilizzata più frequentemente. Ma per essere effettivamente frequente, una parola non deve neppure essere troppo generica» (Lestrade, 2017). Tale metrica permette agli approcci a vocabolario chiuso di identificare due gruppi di parole; le functions words

(parole di stile come articoli, pronomi e preposizioni) ovvero quelle parole con frequenza maggiore in quanto più comuni in un linguaggio e che occorrono maggiormente, usate in un discorso per definire la struttura sociale e le relazioni, e le contest words (parole di contesto, nomi, verbi regolari, aggettivi e avverbi) che sono invece quelle parole con frequenza inferiore e che per questo occorrono solo in specifiche parti del discorso, descrivendo le emozioni e le caratteristiche di un individuo. Su entrambi i set di parole sono stati sviluppati una serie di dizionari specifici:

* **General Inquirer**: sviluppato il 1960 dalla Harvard University. L'ultima versione include 182 dizionari con 8'281 parole uniche, divise in 3 principali sets: 63 Lasswell dictionaries (potere e società), 107 Harvard Psychosociological dictionaries (psicologia e teorie sociali) e 12 Stanford Political dictionaries (interazioni politiche).

* **DICTION**: sviluppato il 1980 con lo scopo di analizzare il tono verbale in 500 discorsi presidenziali. DICTION assume che nei testi politici ci siano 5 variabili master (attività, certezza, comunanza, ottimismo e realismo) e si compone di 31 dizionari non-overlapping con 4 variabili, che decodificano la lunghezza delle parole (complessità), il rapporto tra aggettivi e verbi, la frequenza relativa delle parole (insistenza) e il rapporto tra parole uniche e totali (varietà). DICTION è stato sviluppato soprattutto nelle politiche commerciali.

* **Linguistic Inquiry and Word Count**: sviluppato nel 1990 con il fine di analizzare saggi scritti durante gli interventi di scrittura espressiva. LIWC è organizzato gerarchicamente, con dizionari che si suddividono in altri sottostanti (esempio, affective processes dictionary si suddivide in positive emotion e negative emotion dictionaries, quest'ultimo che comprende sadness, anxiety e anger dictionaries); di conseguenza quando un risultato è correlato a un dizionario di livello inferiore, anche i dizionari ai livelli più alti sono correlati con il medesimo risultato.

2.2.3 Approcci a vocabolario aperto

Gli approcci a vocabolario aperto sono un'alternativa agli approcci a vocabolario chiuso. Tra questi particolare interesse stanno acquisendo gli approcci di clustering (task di Data Mining). Si immagina di avere un insieme di oggetti da analizzare in cui, a differenza della classificazione, non è nota l'etichetta di classe di ciascun oggetto. Il processo di raggruppamento di un insieme di oggetti fisici o astratti in classi di oggetti simili è denominato clustering. Un cluster è una collezione di oggetti che sono simili l'un l'altro

e sono dissimili dagli oggetti di altri cluster. Sebbene la classificazione è un mezzo efficace per distinguere gruppi o classi di oggetti essa richiede una costruzione e un'etichettatura del training set che risultano spesso costose. Spesso può essere desiderabile procedere in senso inverso partizionando prima i dati in gruppi sulla base della loro similarità (ovvero, utilizzando il clustering) e, successivamente, assegnando le etichette al numero relativamente piccolo di gruppi così ottenuto. Ulteriori vantaggi di tale processo riguardano il fatto che esso è adattabile ai cambiamenti e consente di scegliere quali sono le caratteristiche di interesse per distinguere i vari gruppi. Inoltre, in letteratura psicologica hanno ricevuto molta attenzione Latent Semantic Analysis (LSA) e Latent Dirichlet Allocation (LDA):

* Latent Semantic Analysis: sviluppata nel 1980 con lo scopo di determinare la somiglianza tra due corpi di testo. LSA è un approccio molto simile all'analisi fattoriale e permette a dei documenti di essere presentati come una combinazione di punteggi fattoriali, i quali clusterizzano le parole che vi appartengono. Di conseguenza le parole che si trovano vicine hanno uno spazio che tende a coincidere e a co-occorrere con le stesse parole nei documenti, questo significa che tali parole sono tendenzialmente correlate. Tale rappresentazione dimensionale permette inoltre di quantificare la distanza semantica tra due parole. Tuttavia, anche se LSA è un metodo robusto in grado di quantificare le differenze semantiche tra i documenti, l'interpretazione delle sue dimensioni è limitata (può capitare che parole caricate sullo stesso fattore siano semanticamente non coerenti); questo perché la tecnica ignora totalmente la molteplicità di senso delle parole. Per questo motivo utilizzare LSA per l'individuazione su testo scritto di processi psicologici e sociali è limitante.

* Latent Dirichlet Allocation: LDA è un approccio Generative Probabilistic Clustering che raggruppa le parole in topics o argomenti coerenti di un corpus di testo. Gli argomenti possono essere visti come micro-dizionari se paragonati con l'approccio a vocabolario chiuso, tuttavia in questo caso i micro-dizionari vengono generati automaticamente dai dati. Come LSA anche LDA è un'analisi fattoriale tecnica, però qui l'algoritmo presuppone che ogni parola che occorre possa essere attribuita a uno o più argomenti generati dal corpus. Il numero di topics viene assegnato a priori; le parole vengono assegnate a un topic in base alla co-occorrenza con altre parole presenti nel corpus; processo reiterato fino a quando non si raggiunge un equilibrio ottimale, ovvero quando tutte le parole del documento sono assegnate a un insieme di topics con altre parole semanticamente simili. Questi topics rappresentano dunque clusters di parole, in cui alle

parole vengono assegnati dei pesi in base al contributo nel topic. LDA non presenta gli stessi problemi di LSA, in quanto usa una rappresentazione più strutturata che permette di separare le parole in base al contesto. Importante è distinguere la fase di topic modeling (costruzione delle variabili) da quella di topic extraction (studio delle caratteristiche individuali), che possono essere fatte anche su due datasets distinti. Durante la prima fase avviene la generazione di topics da un corpus (auspicabile ottenere topics di alta qualità e con coerenza semantica); nella seconda c'è l'etichettatura, sui topics modellati nella prima fase, che può anche essere svolta su un set di dati di dimensioni inferiori al corpus.

2.3 LIWC: Linguistic inquiry and Word Count

LIWC è un programma di text Analysis trasparente che conta le parole in categorie psicologicamente significative ovvero quantifica il testo utilizzando la frequenza e le categorie di parole combinando metodi tradizionali di apprendimento automatico (Support Vector Machine e Naive Bayes). Support Vector Machine (SVM) e il Naive Bayes sono due tecniche di Machine Learning usate principalmente per la classificazione e, in alcuni casi, per la regressione. I punteggi forniti dal LIWC (Linguistic Inquiry and Word Count) sono percentuali che rappresentano la frequenza di determinate categorie di parole nel testo analizzato. In termini statistici, questi punteggi sono descrittori quantitativi delle caratteristiche linguistiche di un testo. I punteggi sono generalmente espressi come percentuali del numero totale di parole nel testo. Ad esempio, se una categoria di parole ("positive emotion") ha un punteggio di 5, significa che il 5% delle parole nel testo appartiene a questa categoria. Supponiamo di avere i seguenti punteggi medi per alcune categorie LIWC da un gruppo di testi:

- Emozioni Positive (positive emotion): 4.5%
- Emozioni Negative (negative emotion): 2.0%
- Analiticità (analytic): 15%
- Autenticità (authentic): 10%

In questo caso:

- 4.5% delle parole nei testi analizzati appartengono alla categoria "emozioni positive".
- 2.0% delle parole sono classificate come "emozioni negative".
- 15% indicano parole associate a pensiero analitico.
- 10% delle parole suggeriscono autenticità nel linguaggio.

I risultati empirici LIWC dimostrano la sua capacità di rilevare il significato in un'ampia varietà di contesti sperimentali, tra cui l'attenzione, l'emotività, le relazioni sociali, gli stili di pensiero e le differenze individuali (Pennebaker, Francis & Booth, 2001). Il giudizio iniziale del LIWC si è svolto tra il 1992 e il 1994. Una significativa revisione è stata intrapresa nel 1997 e di nuovo nel 2007 per snellire il programma e i dizionari originali. Sono stati analizzati i file di testo di diverse decine di studi, per un totale di oltre cento milioni di parole. Sono state eliminate alcune categorie di parole e ne sono state aggiunte altre. Per i dettagli del processo e i risultati specifici, si veda Pennebaker, Chung, Ireland, Gonzales e Booth (2007).

Attualmente usare LIWC-22 per analizzare il linguaggio può aiutare a comprendere i pensieri, sentimenti, personalità e il modo con cui ci si relaziona con gli altri. L'indagine linguistica e il conteggio delle parole (LIWC) è lo standard di riferimento nei software per l'analisi dell'uso delle parole. Può essere utilizzato per studiare un singolo individuo, gruppi di persone nel tempo o tutti i social media. La nuova suite LIWC include moduli integrati che consentono di sviluppare e personalizzare i dizionari, nonché valuta la coerenza interna dei dizionari, genera nuvole di parole in un attimo, converte i set di dati in un formato di modellazione degli argomenti, basato sul metodo di estrazione del significato, calcola il Language Style Matching (LSM) per confrontare i testi tra loro, modella la valutazione dell'Arco narrativo (AON) per misurare e comprendere come le narrazioni "si svolgono" e analizza i testi con le categorie del dizionario LIWC aggiornate, migliorate e ampliate (Dimensioni linguistiche, Processi psicologici, Relatività, Preoccupazioni personali, Dimensioni sperimentali) per identificare se ciascuna parola specifica del set di dati di origine corrisponde a una qualsiasi delle parole o frammenti rinvenuti nella biblioteca LIWC. Per quanto riguarda l'emozione, le parole vengono confrontate con ciascuna delle tre categorie di emozione (ad esempio, espressione emotiva, emozione positiva ed emozione negativa). Per ogni parola stabilita come corrispondente con una parola o un frammento in un dizionario delle emozioni LIWC, il programma ripete un conteggio di tutte le parole delle emozioni identificate in quel particolare dizionario delle emozioni (ad esempio, emozione positiva). Dunque, possiamo rappresentare le proprietà degli enunciati rilevando la presenza di alcuni segnali o modelli osservabili rilevanti, come astrarre i rapporti tra le caratteristiche linguistiche (parole, n-grammi , tempi verbali, persone verbali, ecc.), astrarre sostantivi tematici da lessici emotivi (ad esempio, Mairesse et al., 2007 ; Pennebaker, Francis, & Booth, 2001), segnali sublessicali (punteggiatura, ortografia, maiuscole, numero di lettere, sillabe), o

anche parole di classe chiuse come indicatori di personalità, ad esempio, pronomi personali per indicare la presenza di espansione sociale (Holzmann ; Campbell & Pennebaker, 2003) o analizzare il tempo dei verbi per rilevare il focus temporale dell'attenzione (Gunsch, 2000). Lo strumento LIWC (Pennebaker et al., 2001 ; Pennebaker & King, 1999) è l'esempio più famoso di questo tipo di paradigma e ha dimostrato la sua utilità nel rilevare stili di coping, personalità e cambiamenti di salute (Campbell & Pennebaker, 2003 ; Mitra, Conti e Pennebaker, 2016 ; Tausczik e Pennebaker, 2010).

LIWC-22 funziona come un'applicazione desktop all-inclusive che ora si integra con altri linguaggi di programmazione come Python e R sfruttando comunque il motore di elaborazione LIWC.

2.3.1. Il significato sociale e psicologico delle parole alla base di LIWC

Le parole che usiamo nella vita quotidiana riflettono ciò a cui prestiamo attenzione, ciò a cui pensiamo, ciò che cerchiamo di evitare, come ci sentiamo e come organizziamo e analizziamo il nostro mondo. Le 80 categorie linguistiche del LIWC sono state collegate in centinaia di studi a interessanti processi psicologici. Per esempio, la categoria "attenzione" può rivelare non solo a chi si sta assistendo, ma anche come si sta elaborando la situazione. Gli studenti che hanno scritto delle loro esperienze di prese in giro hanno usato pronomi diversi a seconda che stessero prendendo in giro gli altri o fossero presi in giro da altri (Kowalski, 2000). I partecipanti hanno usato più pronomi in prima persona singolare e meno pronomi di terza persona (per esempio, "lui", "lei") quando hanno descritto un evento in cui erano stati presi in giro rispetto a quando descrivevano un evento in cui stavano prendendo in giro qualcun altro. In entrambi i casi, l'attenzione si concentra sulla vittima dell'evento. Mentre i pronomi personali forniscono informazioni sul soggetto dell'attenzione, l'analisi del tempo dei verbi comuni può dirci quale sia il focus temporale dell'attenzione e indicare una maggiore distanza psicologica e una maggiore risoluzione per gli eventi rilevati rispetto a quelli non rilevati (Pasupathi, 2007). Nello stesso studio sugli annunci politici, gli autori hanno riscontrato che gli annunci positivi utilizzano maggiormente i verbi al presente e al futuro, mentre gli annunci negativi utilizzavano più verbi al passato (Gunsch et al., 2000). Lo studio dell'attenzione, dunque, ci permette anche di capire meglio come le persone elaborano una situazione o un evento. Ai partecipanti è stato chiesto di ricordare un evento di cui avevano discusso

con un'altra persona o un evento non rivelato; sono state riscontrate differenze significative nel tempo verbale utilizzato nelle due condizioni (Pasupathi, 2007). I partecipanti hanno utilizzato maggiormente il passato quando parlavano di un evento rivelato e un tempo presente quando parlavano di un evento non rivelato. Il grado in cui le persone esprimono emozioni e la valenza di tali emozioni può dirci come le persone stanno vivendo il mondo. Le persone reagiscono in modo radicalmente diverso a eventi traumatici o a eventi importanti; il modo in cui reagiscono può dire molto sul modo in cui affrontano l'evento e in che misura l'evento giocherà un ruolo nel futuro. La ricerca suggerisce che LIWC identifica con precisione le emozioni nell'uso del linguaggio. Ad esempio, le parole che delineano un'emozione positiva (ad esempio, amore, bello, dolce) sono usate per scrivere di un evento positivo, mentre le parole di emozione più negativa (ad esempio, ferito, brutto, cattivo) sono usate per scrivere di un evento negativo (ad esempio, hurt, ugly, nasty) (Kahn, Tobin, Massey, & Anderson, 2007). Le valutazioni LIWC delle parole di emozione positiva e negativa corrispondono alle valutazioni umane degli estratti di scrittura (Alpers, Tobin, Masse, & Anderson, 2007). L'uso di parole emozionali è stato utilizzato anche come misura del grado di coinvolgimento in una situazione (Holmes et al., 2007). Estrapolando casualmente le parole emozionali da circa 2.800 testi, si rilevò che esse erano correlate negativamente con gli articoli (r 0.33), le preposizioni (r 0.38) e le parole di relatività (r 0.40). Queste caratteristiche linguistiche, come discuteremo in seguito, possono essere importanti per la complessità cognitiva e gli stili di pensiero. Le parole di "emozione" erano correlate positivamente con l'uso dei pronomi (r 0.29), l'uso dei verbi ausiliari (r 0.29) e l'uso della negazione (r 0.32). Tutte le correlazioni sono altamente significative, p 0.001. La natura di queste correlazioni suggerisce una più profonda importanza dell'espressione delle emozioni, degli stili di pensiero e della consapevolezza sociale.

La funzione più elementare del linguaggio è quella di comunicare. Le parole forniscono informazioni anche sui processi sociali. La scelta delle parole fornisce anche informazioni sulla percezione, su come le persone elaborano le informazioni e le interpretano per dare senso al loro ambiente (Semin & Fiedler, 1988). Alcuni indizi linguistici rivelano le relazioni. I pronomi rivelano il modo in cui un individuo si riferisce a chi è presente nell'interazione e a chi ne è estraneo. Il conteggio delle parole spiega chi sta dominando la conversazione e quanto è impegnato in essa, in particolare le parole di assenso e di emozione positiva misurano i livelli di accordo. Anche il pensiero può variare in profondità e complessità, e questo si riflette nelle parole che le persone usano per

collegare i pensieri. Il linguaggio cambia quando le persone stanno attivamente rivalutando un evento passato. Può anche differire a seconda della misura in cui un evento è già stato valutato. La complessità cognitiva può essere considerata come la ricchezza di due componenti del ragionamento: la misura in cui una persona distingue tra più soluzioni in competizione e la misura in cui si integrano le soluzioni (Tetlock, 1981). Questi due processi sono racchiusi in due categorie LIWC: le parole di esclusione e le congiunzioni. Le parole di esclusione (ad esempio, ma, senza, escludere) sono utili per fare distinzioni. In effetti, le persone usano le parole di esclusione quando cercano di fare una distinzione tra ciò che è in una categoria e ciò che non ne fa parte. Le parole di esclusione sono più usate tra le persone che dicono la verità (Newman et al., 2003). Le congiunzioni (e, anche, sebbene) uniscono più pensieri insieme e sono importanti per creare una narrazione coerente (Graesser, McNamara, Louwerse, & Cai, 2004). Le preposizioni (ad esempio, to, with, above), i meccanismi cognitivi (ad esempio, cause, know, ought), le parole causali e di comprensione, le parole con più di sei lettere sono tutte indicative di un linguaggio più complesso (Pennekaber, Mayne e Francis, 2001). In una rianalisi di sei studi sulla scrittura espressiva, Pennebaker, Mayne e Francis (1997) hanno riscontrato che l'aumento dell'uso delle parole causali e di intuizione ha portato a miglioramenti nella salute. Boals & Klein nel 2005 sostennero che le parole causali vengono utilizzate nelle parti più traumatiche, la rottura e il post-rottura, perché vengono utilizzate per creare spiegazioni causali e per organizzare i pensieri dei partecipanti. Il linguaggio che le persone usano per discutere di un evento può rivelare qualcosa sulla misura in cui una storia può essere stata stabilita o si sta ancora formando. Quando le persone sono incerte o insicure sul loro argomento, usano un linguaggio incerto (ad esempio, forse, credo) e parole più riempitive (ad esempio, bla, cioè, sai). Forse l'uso maggiore di parole provvisorie suggerisce che un partecipante non ha ancora elaborato un evento e non lo ha trasformato in una storia. Analogamente, Beaudreau, Storandt e Strube (2006) hanno riscontrato che, nel raccontare una storia personale, i partecipanti più giovani usavano più parole riempitive rispetto a quelli più anziani.

La disponibilità dell'uso del linguaggio naturale e le nostre risorse computazionali stanno trasformando l'analisi del linguaggio e le moderne scienze sociali. LIWC rappresenta un programma di analisi testuale di transizione dall'analisi linguistica tradizionale alla nuova era dell'analisi del linguaggio. Le nuove analisi testuali saranno in grado di analizzare strutture linguistiche più complesse, pur mantenendo la trasparenza di LIWC. I metodi di text analysis dovrebbero anche aumentare la loro flessibilità, consentendo al ricercatore

di esaminare categorie linguistiche specifiche per il suo programma di ricerca. La ricerca che si avvale di questi nuovi metodi di analisi del testo sarà ampliata anche per cogliere le differenze culturali che si riflettono nell'uso del linguaggio. Le informazioni sociali rilevanti possono variare notevolmente tra lingua e cultura (Maass, Karasawa, Politi, & Suga, 2006). In effetti, alcune delle differenze culturali più evidenti nel linguaggio sono i marcatori di cortesia, formalità e vicinanza sociale.

Category	Examples	Words in Category	Psychological Correlates
<i>Linguistic processes</i>			
Word count			Talkativeness, verbal fluency
Words/sentence			Verbal fluency, cognitive complexity
Dictionary words	(Percentage of all words captured by the program)		Informal, nontechnical language
Words >6 letters	(Percentage of all words longer than 6 letters)		Education, social class
Total function words		464	
Total pronouns	I, them, itself	116	Informal, personal
Personal pronouns	I, them, her	70	Personal, social
First-person singular	I, me, mine	12	Honest, depressed, low status, personal, emotional, informal
First-person plural	We, us, our	12	Detached, high status, socially connected to group (sometimes)
Second person	You, your, thou	20	Social, elevated status
Third-person singular	She, her, him	17	Social interests, social support
Third-person plural	They, their, they'd	10	Social interests, out-group awareness (sometimes)

Figure 1 - Summary Table Linking LIWC Word Categories (Fonte: Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and social psicologia*)

Category	Examples	Words in Category	Psychological Correlates
Indefinite pronouns	It, it's, those	46	
Articles	A, an, the	3	Use of concrete nouns, interest in objects and things
Common verbs	Walk, went, see	383	
Auxiliary verbs	Am, will, have	144	Informal, passive voice
Past tense	Went, ran, had	145	Focus on the past
Present tense	Is, does, hear	169	Living in the here and now
Future tense	Will, gonna	48	Future and goal oriented
Adverbs	Very, really, quickly	69	
Prepositions	To, with, above	60	Education, concern with precision
Conjunctions	And, but, whereas	28	
Negations	No, not, never	57	Inhibition
Quantifiers	Few, many, much	89	
Numbers	Second, thousand	34	
Swear words	Damn, piss, fuck	53	Informal, aggression,
<i>Psychological processes</i>			
Social processes	Mate, talk, they, child	455	Social concerns, social support
Family	Daughter, husband	64	
Friends	Buddy, friend, neighbor	37	
Humans	Adult, baby, boy	61	
Affective processes	Happy, cried, abandon	915	Emotionality

Figure 2- Summary Table Linking LIWC Word Categories (Fonte: Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and social psychology*)

Category	Examples	Words in Category	Psychological Correlates
Positive emotion	Love, nice, sweet	406	
Negative emotion	Hurt, ugly, nasty	499	
Anxiety	Worried, nervous	91	
Anger	Hate, kill, annoyed	184	
Sadness	Crying, grief, sad	101	
Cognitive processes	Cause, know, ought	730	
Insight	Think, know, consider	195	
Causation	Because, effect, hence	108	
Discrepancy	Should, would, could	76	
Tentative	Maybe, perhaps, guess	155	
Certainty	Always, never	83	Social/verbal skills, emotional stability
Inhibition	Block, constrain, stop	111	
Inclusive	And, with, include	18	
Exclusive	But, without, exclude	17	Cognitive complexity, honesty

Figure 3- Summary Table Linking LIWC Word Categories (Fonte: Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. Journal of Language and social psychology)

Category	Examples	Words in Category	Psychological Correlates
Perceptual processes	Observing, heard, feeling	273	
See	View, saw, seen	72	
Hear	Listen, hearing	51	
Feel	Feels, touch	75	
Biological processes	Eat, blood, pain	567	
Body	Cheek, hands, spit	180	
Health	Clinic, flu, pill	236	
Sexual	Horny, love, incest	96	
Ingestion	Dish, eat, pizza	111	
Relativity	Area, bend, go	638	
Motion	Arrive, car, go	168	
Space	Down, in, thin	220	
Time	End, until, season	239	
<i>Personal concerns</i>			
Work	Job, majors, xerox	327	
Achievement	Earn, hero, win	186	
Leisure	Cook, chat, movie	229	
Home	Apartment, kitchen, family	93	
Money	Audit, cash, owe	173	
Religion	Altar, church, mosque	159	
Death	Bury, coffin, kill	62	
<i>Spoken categories</i>			
Assent	Agree, OK, yes	30	Agreement, passivity
Nonfluencies	Er, hm, umm	8	
Fillers	Blah, lmean, yaknow	9	Informal, Unprepared speech

Figure 4- Summary Table Linking LIWC Word Categories (Fonte: Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. Journal of Language and social psychology)

CAPITOLO 3

TRATTI DI PERSONALITÀ E TEXT MINING

3.1 Tratti e stili di personalità

Dai modelli di Allport (1965) e Eysenck (1952) fino alle più recenti teorie fattoriali (Costa, McCrae, 1992), i tratti sono considerati le dimensioni di base della personalità in grado di rendere conto sia delle caratteristiche comuni alle diverse persone sia delle differenze individuali, se letti in chiave quantitativa. I tratti forniscono dunque uno schema generale per rappresentare la personalità, un insieme di caratteristiche riconoscibili che l'individuo tende a esprimere in un'ampia gamma di situazioni per un arco di tempo relativamente lungo. Tratto: modalità costante di percepire, rapportarsi e pensare nei confronti dell'ambiente e di sé stessi, che si manifesta in un'ampia varietà di situazioni sociali e personali. (Caprara, Gennaro 1999); Tutti i modelli basati sul riconoscimento di "tratti" condividono però la convinzione che la personalità sia costituita da tendenze comportamentali organizzate gerarchicamente: i tratti sovraordinati organizzano le tendenze che si collocano a un livello più basso, le quali, a loro volta, controllano le abitudini comportamentali, situate a un livello ancora inferiore. Quando invece di indagare le dimensioni di base della personalità, studiamo i pattern più generali di comportamento, emozione, pensiero e motivazione si sta indagando lo stile di personalità. Ci si riferisce cioè a una configurazione più o meno stabile di tratti di personalità che definisce una struttura psicologica più generale. È possibile così indagare i modi in cui i diversi tratti tendono a organizzarsi determinando stili di personalità o tipi. Il costrutto di stile cerca di fornire una descrizione più globale del funzionamento individuale. Quando queste modalità di agire, percepire e rapportarsi e pensare diventano rigide o disadattive e interferiscono con il funzionamento relazionale e lavorativo di un individuo e con il suo benessere allora è possibile parlare di "disturbo della personalità". La rigidità si riferisce alla presenza di un ristretto repertorio di risposte che vengono ripetute anche quando la soluzione richiederebbe soluzioni diverse, o di fronte alla chiara evidenza che tale comportamento è inadeguato e disfunzionale.

Un ampio studio di Widiger e Costa (2002) ha concluso che i disturbi di personalità rappresentano una variante disadattiva di tratti presenti nella popolazione normale.

Studi longitudinali hanno infine dimostrato la stabilità dei tratti nel tempo dall'infanzia all'età adulta (Caspi, Harrington, Milne et al., 2003; Roberts, Delvecchio, 2000).

La componente genetica ereditaria dei tratti di personalità influenza il comportamento del 30/50 % (Cloninger et al 1993; Livesley 2006). Tratti come l'aggressività o la timidezza sono attribuibili in una certa misura a una predisposizione biologica, ma è pur vero che questa predisposizione influenza lo sviluppo di una psicopatologia di personalità connessa ai relativi tratti. Inoltre, i tratti di personalità sono moderatamente ereditabili e ugualmente influenzati da fattori genetici e ambientali.

3.2 Principali sistemi di valutazione di tratti e disturbi di personalità

La costruzione di un sistema di classificazione dei tratti di personalità implica delle scelte teoriche e metodologiche che influenzano non solo il tipo di classificazione prodotta, anche il tipo di approccio al trattamento della personalità. Lo scopo principale di un sistema di classificazione dovrebbe essere infatti quello di guidare l'osservazione clinica e il trattamento, processi che possono essere facilitati dalla risoluzione di alcune questioni teoriche e metodologiche di fondo.

L'approccio categoriale alla classificazione dei disturbi di personalità vede una discontinuità e una differenza significativa tra ciò che è "normale" e ciò che è "disturbato". Viceversa, l'approccio dimensionale presuppone un continuum tra questi due concetti. Nonostante esista in letteratura la tendenza a far coincidere l'approccio dimensionale allo studio dei tratti di personalità e l'approccio categoriale alla descrizione degli stili o tipologie di personalità (Westen, 2006), la soluzione clinicamente più utile consiste in un sistema misto, in grado cioè di coniugare la valenza clinico-metodologica dell'approccio dimensionale con l'efficienza e semplicità d'uso delle categorie diagnostiche. La diagnosi classica di stampo psichiatrico si fonda sulla presenza o meno di specifici descrittori che definiscono il disturbo. Questo tipo di diagnosi possiede diverse caratteristiche categoriali e sostiene l'esistenza di un chiaro confine tra normalità e patologia. Le diagnosi psichiatriche possono essere monotetiche, laddove il disturbo sia definito dalla presenza di un intero pool di criteri, o politetiche, laddove per fare diagnosi è sufficiente la presenza di un sottogruppo di criteri presi da un pool più ampio.

Un approccio prototipico, invece, valuta la presenza di un certo stile o di un certo disturbo della personalità in base al livello di somiglianza tra la presentazione clinica di un paziente e dei prototipi dei diversi stili o disturbi di personalità. In questo modo gli

individui possono somigliare a prototipi diversi di personalità ma differire per quanto si somigliano. I dati di ricerca depongono a favore di un approccio politetico-prototipico perché alcuni sintomi e disturbi hanno un maggior peso diagnostico rispetto ad altri e inoltre una procedura in cui si chiede al clinico di delineare il peso e la funzione dei singoli criteri diagnostici all'interno della personalità globale facilita il processo diagnostico perché è vicina al modo in cui i clinici effettuano diagnosi nella vita reale.

Storicamente, le teorie fattoriali della personalità hanno utilizzato, come variabili, liste di aggettivi tratte dal linguaggio quotidiano, ricorrendo a popolazioni normali per estrarre i fattori di personalità. Questo tipo di approccio parte dal presupposto che la personalità è un oggetto di osservazione "naturale" e per questo le diverse lingue ha in sé parole che ne descrivono le caratteristiche fondamentali. Diversi autori (per esempio, Pervin 1994; Westen, 1996) hanno criticato l'utilizzo di aggettivi come variabili di studio, sostenendo che il linguaggio quotidiano non è in grado di raggiungere la complessità necessaria a descrivere i fenomeni clinici importanti. In questo caso sono più utili frasi o interi periodi (Block, 1995).

3.2.1 Il modello di Eysenck

Il modello proposto da Hans Eysenck (1987) è precursore degli approcci che mettono i tratti di personalità in relazione alla biologia e si contraddistinguono sia per il ristretto numero di dimensioni considerate sia per la forte importanza data alle componenti biogenetiche della personalità. Il suo modello è centrato su tre grandi fattori: estroversione (E), nevroticismo (N), psicoticismo (P). L'estroversione corrisponderebbe a caratteristiche di socievolezza, vivacità, attività ecc. Il nevroticismo include tratti come la variabilità del tono dell'umore e la predisposizione all'ansia ed è considerato perciò una misura della stabilità emotiva. Nel fattore psicoticismo confluiscono invece aspetti come l'impulsività, l'aggressività, la ricerca di sensazioni e la scarsa sensibilità sociale. Nel sistema di Eysenck, inoltre, ciascuno di questi tratti è correlato a specifiche attività neurochimiche del cervello.

3.2.2 Il modello di Cloninger

L'autore che meglio è riuscito a sviluppare un modello dei tratti normali e patologici basato su correlati biologici è Robert Cloninger (1998). Cloninger è arrivato a dividere la personalità in due domini di base: temperamento (formato da risposte automatiche a stimoli emotivi che determinano abitudini e abilità) e carattere (composto da concezioni consapevoli che influenzano volontariamente le intenzioni e gli atteggiamenti). Secondo questo modello, tutti i disturbi di personalità si caratterizzerebbero per bassi punteggi nelle dimensioni dell'auto direttività e della cooperatività. Quello che differenzia i diversi disturbi sarebbe invece il modo in cui queste due dimensioni si legano agli altri tratti: i pazienti del cluster A presenterebbero una bassa dipendenza dalla ricompensa, i pazienti del cluster B avrebbero una forte ricerca di novità, mentre i pazienti del cluster C si caratterizzano per un marcato evitamento del danno. Le sette dimensioni di personalità sono: *ricerca di novità*, *evitamento del danno*, *dipendenza dalla ricompensa*, *persistenza*, *auto direzionalità*, *cooperatività* e *autotrascendenza*. La *ricerca di novità* è un tratto caratterizzato dalla presenza di una forte tendenza all'esplorazione e da aspetti di stravaganza e impulsività, *l'evitamento del danno* è una dimensione caratterizzata dalla presenza di pessimismo, paura ed eccessiva timidezza, *la dipendenza dalla ricompensa* è un tratto caratterizzato da sentimentalismo, una forte tendenza a rispondere prontamente ai segnali di approvazione sociale e apertura all'esperienza, *la persistenza* è un tratto definito alla presenza di determinazione e ambizione, *l'auto direzionalità* è una dimensione connessa ad aspetti di responsabilità, accettazione di sé e responsabilità, *la cooperatività* è il tratto che descrive la presenza di una forte empatia, compassione e disponibilità verso l'altro e infine *l'auto trascendenza* aspetto indicante la sensazione di sentirsi parte integrante di un sistema universale, definito da spiritualità e idealismo.

I profili dei singoli disturbi di personalità possono essere descritti in maniera più precisa attraverso il self-report Temperament and Character Inventory (TCI; Cloninger, Svrakic, 1994), un questionario auto somministrato di 240 item volti a valutare e quantificare i sette fattori del modello. Il TCI è stato utilizzato anche per indagare le dimensioni della personalità nei disturbi psichiatrici. Peirson e Heuchert (2001) hanno indagato il rapporto tra dimensioni del TCI e depressione, riscontrando come la personalità di pazienti con una diagnosi di depressione si caratterizzi per la presenza di un alto evitamento del danno e un basso livello di auto direttività.

3.2.3 Il modello di Zuckerman

Per Myron Zuckerman, al centro della personalità vi sarebbe un tratto di base, la *sensation seeking*, descrivibile come un continuo bisogno di esperienze e sensazioni nuove, che determina la propensione ad assumere rischi fisici e sociali. Le differenze individuali nell'espressione di questo tratto corrispondono all'espressione di una dimensione fondamentale della personalità. I cinque fattori determinanti alla base di questo costrutto sono: *nevroticismo-ansia*, *attività*, *socialità*, *ricerca di sensazioni impulsiva*, *aggressione-ostilità*. Secondo Zuckerman (2002) sarebbe una combinazione di ricerca di sensazioni impulsiva e bassa socialità a caratterizzare i disturbi gravi di personalità (come psicopatie, condotte antisociali e forme di ostilità paranoidee). Ognuna di queste dimensioni può essere valutata attraverso lo Zuckerman- Kuhlman Personality Questionnaire (ZKPQ; Zuckerman, 2002), un questionario self report a 99 item a risposta dicotomica.

3.2.4 Il modello dei Big Five Factors

A differenza dei modelli biologici dei tratti, che cercano di esplorare le origini genetiche della personalità, le teorie fattoriali utilizzano l'approccio lessicale per studiare l'architettura generale della personalità. Il Five Factors Model (FFM; McCrae, Costa, 1997; Widiger, Costa, 1994) è, tra le teorie fattoriali della personalità, quella che ha raccolto maggior credito. Si tratta di un modello che, pur avendo come finalità esplicita l'individuazione di un certo numero di fattori, tenta di coniugare l'operatività dei sistemi nati in ambito clinico con la complessità di quelli di derivazione fattoriale e psicolinguistica. Attraverso lo studio di numerosi campioni non clinici, hanno ridotto la vasta gamma di aggettivi possibili a cinque costrutti sovraordinati con i loro relativi poli positivi e negativi (Tabella 1).

Tratto	Alto	basso
Amichevolezza	Sensibile, elogiativo, altruista, gentile, amichevole, cooperativo, piacevole, indulgente, disponibile, fiducioso, caloroso, generoso, cortese, comprensivo	Critico, freddo, ostile, litigioso, scortese, irrispettoso
Estroversione	Loquace, assertivo, attivo, energico, estroverso, schietto, dominante, entusiasta, esibizionista, socievole, vivace, avventuroso, rumoroso, autoritario	Tranquillo, riservato, timido, silenzioso, chiuso, ritirato
Coscienziosità	Organizzato scrupoloso, pianificatore, efficiente, responsabile, affidabile, coscienzioso, preciso, pratico, ponderato, attento	Sbadato, disordinato, frivolo, irresponsabile, approssimativo, inaffidabile, smemorato
Nevroticismo	Teso, ansioso, nervoso, lunatico, preoccupato, permaloso, pauroso, autocommiserativo, irascibile, instabile, autopunitivo, emotivo	Stabile, calmo, soddisfatto, non emotivo
Apertura	Fantasiioso, intelligente, originale, perspicace, curioso, acuto, ingegnoso	Interessi ristretti, superficiale

- Tabella 1. Tratti di personalità del Big Five Factors Model

La definizione dei cinque fattori di base è stata raggiunta attraverso due approcci, quello *lessicale*, dove le categorie descrittive sono state estratte setacciando l'intero patrimonio lessicale alla ricerca delle parole più rappresentative dei tratti, e quello *strutturale*, che giunge alla definizione dei tratti attraverso l'analisi fattoriale e una successiva riflessione teorica sui dati emersi. Secondo gli autori, i cinque fattori del FFM rappresenterebbero la "struttura universale" della personalità umana. Studi di ricerca transculturali hanno evidenziato la rilevanza transculturale del FFM che si è dimostrato utile anche nello studio della costanza e del cambiamento della personalità nel tempo (McCrae, Costa, 1990) ed è stato applicato con successo allo studio dell'ereditarietà dei tratti di personalità (Jang, McCrae, Angleitner et al., 1998). Più recentemente, il FFM è stato utilizzato nello studio della psicopatologia di personalità.

Il FFM si basa su una visione dimensionale dei tratti e considera i disturbi di personalità un'amplificazione patologica dei normali tratti della personalità. Poiché, tuttavia, l'analisi a livello dei soli cinque fattori non possiede una specificità tale da catturare la complessità dei disturbi di personalità, ogni fattore è stato successivamente scomposto in sei sotto-fattori di ordine inferiore (Widiger et al., 1994).

Il sistema dei FFM consente di “misurare” i disturbi di personalità in due modi diversi. Il primo consiste nell’identificare la patologia della personalità con punteggi estremi nei cinque fattori. In questo caso, per esempio, una patologia di personalità si può identificare con un alto punteggio sul fattore nevroticismo. Un altro modo di procedere consiste invece nel tradurre le categorie del DSM nel linguaggio dei cinque fattori (Coker et al., 2002; Lynam, Widiger, 2001). Widiger e collaboratori (2002), per esempio, hanno ridefinito il disturbo antisociale di personalità come una combinazione di punteggi bassi su fattori gradevolezza e coscienziosità. I pazienti con disturbo borderline di personalità si differenzerebbero dai pazienti con disturbo dipendente per un alto punteggio su tutti e sei i fattori del nevroticismo (Widiger et al., 2002). La valutazione dei cinque fattori è affidata a un questionario autosomministrato, il NEO PI-R (Neuroticism, Extroversion, Openness to experience Personality Inventory- Revised; Costa, McCrae, 1992), formato da 240 item che valutano sia i cinque fattori di ordine superiore e sia i sei tratti inferiori. Le principali tecniche di raccolta dei dati per la valutazione dei tratti e dei disturbi di personalità presentano problematiche specifiche che spesso impongono l’utilizzo di tecniche e fonti di informazioni multiple. I self-report sono la tecnica di valutazione della personalità più utilizzata in ambito di ricerca; gli informant-report sono, invece, degli strumenti che ricorrono a informatori esterni come fonte di dati sul paziente e infine il clinician-report. Nella sezione seguente sono state messe in rassegna alcune ricerche che hanno utilizzato l’intelligenza artificiale attraverso sistemi di text analysis per prevedere i tratti di personalità.

3.3. La Text Mining applicata allo studio dei tratti di personalità

Per predire i tratti di personalità vengono comunemente utilizzati dei questionari specializzati. Tuttavia, già da qualche anno, anche l’Intelligenza Artificiale sta emergendo come strumento adatto a determinare i profili di personalità. Quest’ultima, difatti, si presta bene a decodificarne la digital footprint che permette di associare l’individuo a tratti di personalità specifici. Le categorie utilizzate nell’approccio sono state le categorie LIWC.

Le parole che le persone usano nella loro vita quotidiana possono rivelare aspetti importanti del loro mondo sociale e psicologico. Con i progressi della tecnologia informatica, l’analisi del testo consente ai ricercatori di valutare, in modo affidabile e rapido, le caratteristiche dei loro stili linguistici. Dopo una breve rassegna nel capitolo 2

di diversi programmi di analisi del testo, riassumiamo alcune delle prove che collegano l'uso naturale delle parole alla personalità, alle fluttuazioni sociali e situazionali e agli interventi psicologici. La personalità è l'insieme duraturo di tratti e stili che un individuo esibisce, cioè quelle caratteristiche che rappresentano le sue disposizioni, vale a dire tendenze naturali o inclinazioni personali. Essere consapevoli delle caratteristiche della personalità delle persone le aiuterà a migliorare le loro capacità di gestione delle relazioni e anche a migliorare le loro comunicazioni interpersonali. Certo, avere informazioni sulla personalità con cui comunichiamo sarebbe molto vantaggioso. Può essere effettuato analizzando i testi scambiati (detta anche lingua scritta), ad esempio, tra gli utenti social. Di conseguenza, la previsione automatica della personalità umana attraverso approcci computazionali è chiamata Previsione Automatica della Personalità. L'elaborazione della personalità o *personality computing* è un campo di ricerca correlato all'intelligenza artificiale e alla psicologia della personalità che, mediante tecniche computazionali, estrapolano i dati dal testo, dai multimedia e dai social network. L'elaborazione della personalità affronta tre problemi principali: il riconoscimento automatico, la percezione e la sintesi (Vinciarelli,2014). Il riconoscimento automatico della personalità è la deduzione del tipo di personalità degli individui selezionati attraverso l'analisi di dati da loro prodotti o da comportamenti osservati. La percezione automatica della personalità è l'inferenza della personalità attribuita da un osservatore a un individuo sulla base di un comportamento che sia osservabile. Infine, la sintesi automatica della personalità è la generazione artificiale di stile o comportamento di una personalità in Avatar e agenti virtuali. I test di personalità, autovalutati o compilati da degli osservatori esperti, sono sempre utilizzati come banco di prova per testare e validare le prestazioni degli algoritmi di intelligenza artificiale per la previsione automatica dei tipi di personalità (Celli, 2013). Esiste un'ampia varietà di test di personalità, come l'indicatore del tipo Myers Briggs (MBTI) o MMPI, ma i più utilizzati sono test basati sul modello a cinque fattori, come l'inventario della personalità NEO PI-R (Costa, 1992). La disciplina dell'elaborazione della personalità ha avuto origine intorno al 2005 con alcuni lavori di ricerca pionieristici per il riconoscimento della personalità. Queste ricerche mostrarono che i tratti della personalità potrebbero essere dedotti con ragionevole accuratezza da testi quali: blog, auto-presentazioni (Argamon,2005; Oberlander,2006; Mairesse,2007), e indirizzi e-mail (Back,2008). Pochi anni dopo sono iniziate le ricerche sul riconoscimento e la percezione della personalità attraverso l'analisi di segnali multimodali e sociali, come video e chiamate vocali. Nel 2010 la ricerca si è concentrata

principalmente sul riconoscimento e la percezione della personalità delle persone presenti sui social media, in particolare su Facebook, Twitter e Instagram . Negli stessi anni la sintesi automatica della personalità ha contribuito a migliorare il comportamento simulato negli agenti virtuali (Faur, Caroline, et al. 2013). Ricerche scientifiche hanno dimostrato la validità del calcolo della personalità partendo da alcuni dati comportamentali, in particolare dalle preferenze degli utenti come i "Mi piace" delle pagine di Facebook (Kosinski,2013) e hanno dimostrato che gli algoritmi sono in grado di riconoscere le personalità meglio degli umani. Le tecniche di elaborazione della personalità hanno applicazioni nel Social media marketing, dove possono aiutare a ridurre i costi delle campagne pubblicitarie attraverso l'identificazione psicologica (Matz,2017 ; Celli,2017). In generale, i campioni vocali contengono due potenziali fonti di informazione sulla personalità di chi parla: caratteristiche linguistiche e vocali. Qui delineiamo sei potenziali meccanismi per spiegare come i campioni vocali potrebbero essere associati alla personalità di chi parla. I primi due meccanismi derivano dal fatto che la personalità è legata alle attività, agli interessi, agli stati emotivi e ai luoghi degli individui; per esempio, gli estroversi hanno maggiori probabilità degli introversi di trascorrere del tempo socializzando, di provare emozioni positive e di frequentare luoghi con alti livelli di attività sociale. Quindi, il primo meccanismo cattura l'idea che la personalità guiderà ciò di cui le persone parlano. Nella misura in cui gli individui parlano delle loro attività, interessi, stati emotivi e luoghi, persone con personalità diverse parleranno di cose diverse. La prova di questo meccanismo è stata fornita in un ampio studio sui blog, che ha mostrato che le personalità dei blogger erano associate alla frequenza delle parole rilevanti per i tratti della personalità che usavano; per esempio, gli estroversi sono più propensi degli introversi a usare parole associate ad interessi e attività estroverse, come andare al bar e socializzare (ad esempio, "bar", "drink", "amici", "incontrato"; Yarkoni, 2010). Più in generale, esiste un ampio corpus di ricerche che collegano il linguaggio alla personalità (Pennebaker et al., 2003); questa ricerca ha identificato solide connessioni tra i tratti della personalità e un'ampia gamma di campioni linguistici, incluso il linguaggio trovato nei libri, nella poesia, nelle canzoni, nei discorsi, nella scrittura del flusso di coscienza e nei campioni vocali.

Il secondo modo in cui si rilevano le differenze di personalità nelle attività, negli interessi, negli stati emotivi e nei luoghi deriva dal modo in cui le persone parlano. Ad esempio, un estroverso che socializza abitualmente in gruppo può imparare a parlare ad alta voce per essere ascoltato in luoghi rumorosi come i bar.

Il terzo e il quarto meccanismo che collega potenzialmente i campioni vocali e la personalità rispecchiano i primi due meccanismi descritti sopra ma invertono la direzione causale. Potrebbe cioè darsi che ciò di cui le persone parlano (terzo meccanismo) e il modo in cui parlano (quarto meccanismo) abbiano un impatto sui tratti della loro personalità. Ad esempio, il contenuto del discorso di una persona (ad esempio, parlare di filosofia o di sport) e il modo in cui lo pronuncia (ad esempio, ad alta voce o a voce bassa) potrebbero avere un impatto sul modo in cui vengono trattati dagli altri (ad esempio, quali compiti gli vengono assegnati sul lavoro o se sono percepiti come intelligenti o dominanti), il che potrebbe, a sua volta, influenzare la loro personalità.

Il quinto e il sesto meccanismo che collega potenzialmente i campioni vocali e la personalità si basa sull'idea che una terza variabile potrebbe influenzare la personalità, nonché ciò di cui parla (quinto meccanismo) e come parla (sesto meccanismo), ad esempio, l'istruzione, l'educazione o la cultura. Inoltre, i fattori biologici potrebbero anche avere un impatto sulla personalità, sulla lingua e sulle caratteristiche vocali.

Negli ultimi due decenni, prove empiriche hanno dimostrato che i tratti della personalità potrebbero essere correlati alle caratteristiche della lingua scritta. Moreno e colleghi nel 2021 (*Can personality traits be measured analyzing written language? A meta-analytic study on computational methods*) descrivono una metanalisi che sintetizza 23 stime indipendenti delle correlazioni tra i cinque principali tratti della personalità e alcuni indicatori ottenuti computazionalmente dalla lingua scritta. I risultati mostrano stime combinate significative delle correlazioni, sebbene da piccole a moderate secondo le convenzioni di Cohen per interpretare le dimensioni dell'effetto, per i cinque tratti (tra $r = 0,26$ per gradevolezza e nevroticismo e $0,30$ per apertura). Queste stime sono moderate dal tipo di informazioni contenute nei testi, dall'uso di meccanismi di previsione e dalla fonte di pubblicazione degli studi primari. Generalmente, per i cinque tratti operano gli stessi moderatori efficaci. Si conclude che la lingua scritta analizzata attraverso metodi computazionali potrebbe essere utilizzata per estrarre informazioni rilevanti sulla personalità. (Moreno, 2021). Ma sono ancora necessarie ulteriori ricerche per considerarlo uno strumento predittivo o esplicativo delle differenze individuali. Hanno verificato che la stragrande maggioranza degli studi ha misurato i tratti della personalità utilizzando il modello Big Five Personality (Big Five; McCrae & Costa Jr., 2008): Apertura all'esperienza (O, relativa alla curiosità intellettuale e all'apertura a provare cose nuove), Coscienziosità (C, relativa alla autodisciplina e comportarsi secondo il dovere), Estroversione (E, generalmente caratterizzata da un pronunciato

impegno con il mondo esterno), Piacevolezza (A , gli individui con punteggi elevati in questo tratto tendono ad essere premurosi, gentili, generosi e disponibili con gli altri) e Nevroticismo (N , per lo più legato alla tendenza a provare emozioni negative come rabbia e ansia). Il modello Big Five Personality è stato utilizzato in molte ricerche accumulando un solido corpus di conoscenze sulla stabilità e l'universalità dei tratti della personalità proposti al suo interno, e mostrando anche elevati coefficienti di validità (ad esempio, Grucza e Goldberg, 2007), che riflettono la fedeltà nel prevedere ciascun tipo di comportamento all'interno del suo intervallo.

Tuttavia, la loro scoperta principale è stata che la lingua scritta mostra relazioni significative con le cinque dimensioni fondamentali della personalità, tanto da poter essere utilizzata come predittore del profilo della personalità dell'individuo. Questi risultati delle dimensioni degli effetti combinati sono anche coerenti e congruenti con la letteratura precedente (ad esempio, Farnadi et al., 2013 ; Farnadi et al., 2016 ; Gao et al., 2013 ; Golbeck, 2016 ; Golbeck et al., 2011 ; Hawkins et al., 2017 ; Kwantes et al., 2016 ; Mairesse et al., 2007 ; Park et al., 2015 ; Schwartz et al., 2013 ; Skowron et al., 2016 ; Thilakaratne et al., 2016 ; Yarkoni, 2010). Questi risultati rafforzano la rilevanza delle relazioni tra personalità e linguaggio (Boyd e Pennebaker, 2015 , Boyd e Pennebaker, 2017).

Un piccolo gruppo di ricercatori con orientamento psicoanalitico si è interessato al modo in cui i clienti utilizzano il linguaggio nelle sessioni di terapia. Bucci (ad esempio, 1995) e Mergenthaler (1996) hanno identificato separatamente e insieme (Mergenthaler & Bucci 1999) modelli di parole che predicono risultati terapeutici positivi. Gli autori hanno identificato tre categorie di parole che possono essere facilmente catturate dalle analisi computerizzate: tono emotivo, astrazione e attività referenziale. Utilizzando questo sistema di codifica, gli autori sostengono che una terapia di successo richiede che i clienti passino da un'attività referenziale altamente specifica e da un elevato tono emotivo a livelli elevati di astrazione. In effetti, l'analisi di sedute selezionate di terapia psicoanalitica (Mergenthaler 1996) così come i saggi scritti di divulgazione (Bucci 1995) supportano queste previsioni. Questi modelli sono notevolmente coerenti con le analisi LIWC di Pennebaker et al. (1997).

3.4 La Text Mining applicata allo studio di utenti Facebook

I lavori di Maarten Sap et al. (“Developing Age and Gender Predictive Lexica over Social Media”,2014), H. Andrew Schwartz et al. (“Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach”,2013) e Margaret L. Kern et al. (“The Online Social Self: An Open Vocabulary Approach to Personality”,2014) sono molto simili negli obiettivi. In ciascuno è stata svolta l’analisi del linguaggio naturale su post di Facebook di utenti che avevano fatto uso dell’applicazione myPersonality; su questi sono stati studiati il genere, l’età e la personalità. Gli approcci testati sono stati a vocabolario chiuso, e a vocabolario aperto come verifica finale soltanto da Margaret L. Kern et al, e la creazione di un lessico pesato da Maarten Sap et al.

Il lavoro di Schwartz et al. 2013 ha testato sui post un approccio a vocabolario chiuso. Tale tipo di approccio associa il linguaggio usato da un individuo a una categoria, per questo è stato indicato dagli autori anche come word-count approach, e lo hanno definito con la formula:

$$p(category|subject) = \frac{\sum_{word \in category} freq(word, subject)}{\sum_{word \in vocab(subject)} freq(word, subject)},$$

dove $freq(word, subject)$ è il numero di volte in cui un soggetto utilizza la parola $word$ e, al denominatore, $vocab(subject)$ indica l’insieme di tutte le parole utilizzate dal soggetto. Le categorie utilizzate nell’approccio di Schwartz et al. sono state le categorie LIWC e il $subject$ gli utenti di myPersonality. Dopodiché per collegare le categorie con i tratti di personalità, genere ed età i ricercatori hanno utilizzato la regressione dei minimi quadrati ordinari. Il coefficiente della variabile esplicativa, in categoria LIWC, è stata considerata la forza della relazione. Il lavoro di Eleonora Signor e Nicolò Navarin (“Intelligenza Artificiale e Psicografia: come personalità e genere influiscono sulla scelta occupazionale”, 2023) risponde alla domanda se il genere di appartenenza e la personalità possono influire sulla scelta occupazionale utilizzando modelli di Machine Learning per l’analisi del linguaggio naturale su post pubblici di Twitter, per cogliere le relazioni esistenti tra genere, tratti di personalità Big Five e occupazione. Per la predizione dei tratti di personalità Big Five si è deciso di istanziare il modello Word and phrase correlations di Shwartz, il quale si compone di cinque sotto modelli, uno per ciascun tratto di personalità:

* A: agreeableness (amichevolezza);

- * C: conscientiousness (coscienziosità);
- * E: extraversion (estroversione);
- * N: neuroticism/ emotional stability (nevroticismo/stabilità emotiva);
- * O: openness to experience (apertura alle esperienze).

Ognuno dei sotto modelli è costituito dalle prime 100 frasi, composte al massimo da 3-grammi, che Shwartz ha individuato, con l'utilizzo della tecnica di Differential Language, come appartenenti a ogni polarità dei tratti di personalità. Ciascun sotto modello si compone dei parametri frase/parola, correlation e p-value; con correlazione positiva quando la frase/parola appartiene a un alto tratto invece negativa quando a un basso tratto. Inoltre, per selezionare le prime 100 frasi, più descrittive per ciascun tratto, gli autori hanno scelto di valutare le correlazioni con maggior valore assoluto e p-value inferiore a 4×10^{-9} . Per il modello Word and phrase correlations le frasi/parole sono state estratte da Facebook e da utenti che hanno utilizzato l'applicazione myPersonality. Anche per predire la personalità degli utenti di Twitter Occupation Dataset è stato fatto riferimento all'approccio a vocabolario chiuso di Schwartz et al. 2013; Essendo che il loro scopo era solo quello di individuare quali erano i tratti di personalità Big Five degli utenti, hanno riutilizzato la formula di word-count approach, con categorie i tratti di personalità e sostituendo la frequenza con la correlazione delle frasi/parole espresse in Word and phrase correlations quando presenti anche nella Bag of words di un utente, tralasciando l'utilizzo della regressione ai minimi quadrati ordinari.

3.5 Tratti di personalità: Utilizzo di LIWC

I metodi basati sul lessico cercano principalmente di prevedere la personalità dello scrittore assegnando le sue parole a categorie predeterminate. L'indagine linguistica e il conteggio delle parole (LIWC) è uno degli strumenti più comuni che conta le parole in categorie psicologicamente significative. I punteggi forniti dal LIWC (Linguistic Inquiry and Word Count) sono percentuali che rappresentano la frequenza di determinate categorie di parole nel testo analizzato. In termini statistici, questi punteggi sono descrittori quantitativi delle caratteristiche linguistiche di un testo. I punteggi LIWC sono generalmente espressi come percentuali del numero totale di parole nel testo. Ad esempio, se una categoria di parole ("positive emotion") ha un punteggio di 5, significa che il 5% delle parole nel testo appartiene a questa categoria.

L'idea principale alla base di LIWC è che l'uso delle parole nel linguaggio quotidiano rivela i pensieri, la personalità e i sentimenti degli individui. Nella loro analisi di APP, dalle parole che le persone usano, Yuan et al. hanno indagato la personalità dei personaggi dei romanzi in volgare. Hanno creato un vettore per ogni dialogo utilizzando le funzionalità LIWC, che riflette la psicologia dei personaggi. Infine, i vettori sono stati mappati sui tratti della personalità dei Big Five, per prevedere le etichette finali della personalità.

In particolare, Mairesse et al. Nel 2007 hanno anche studiato una varietà eterogenea di caratteristiche basate sul lessico al fine di prevedere i tratti della personalità dei Big Five dal testo scritto e dalla conversazione parlata.

Nonostante alcuni limiti noti (Eysenck, 1997; Paunonen & Jackson, 2000), negli ultimi 50 anni il modello Big Five è diventato uno standard in psicologia e gli esperimenti che utilizzano i Big Five hanno dimostrato che i tratti della personalità influenzano molti aspetti del comportamento individuale correlato ai compiti. Ad esempio, il successo della maggior parte dei compiti interpersonali dipende dalla personalità dei partecipanti e i tratti influenzano la capacità di leadership (Hogan, Curphy e Hogan, 1994), la prestazione lavorativa generale (Furnham, Jackson e Miller, 1999), capacità di vendita (Furnham et al., 1999), capacità accademica e motivazione (Furnham & Mitchell, 1991; Komarraju & Karau, 2005). Tuttavia, fino ad oggi c'è stato poco lavoro sul riconoscimento automatico dei tratti della personalità (Argamon, Dhawle, Koppel e Pennebaker, 2005; Oberlander & Nowson, 2006). Un recente lavoro sull'intelligenza artificiale esplora metodi per il rilevamento automatico di altri tipi di variazione pragmatica nel testo e nella conversazione, come l'emozione (Oudeyer, 2002; Liscombe, Ven-ditti e Hirschberg, 2003), l'inganno (Newman, Pennebaker, Berry, & Richards, 2003; Graciarena, Shriberg, Stolcke, Enos, Hirschberg e Kajarekar, 2006; Hirschberg, Benus, Brenier, Enos, Friedman, Gilman, Girand, Graciarena, Kathol, Michaelis, Pellom, Shriberg e Stolcke, 2005), carisma dell'oratore (Rosenberg & Hirschberg, 2005), umore (Mishne, 2005), dominanza nelle riunioni (Rienks & Heylen, 2006), punto di vista o soggettività (Somasundaran, Ruppenhofer e Wiebe, 2007) e il sentimento o opinione (Breck, Choi, & Cardie, 2007). In contrasto con questi fenomeni pragmatici, che possono essere relativamente contestualizzati o di breve durata, la personalità è solitamente considerata essere un aspetto degli individui a lungo termine e più stabile (Scherer, 2003). Tuttavia, ci sono prove che la personalità interagisce e influenza altri aspetti della produzione linguistica. Ad esempio, ci sono forti relazioni tra i tratti di estroversione e coscienziosità e gli affetti

positivi, e tra nevroticismo, antipatia e vari affetti negativi (Watson & Clark, 1992). Mentire porta a incoerenze visivo vs. acustico del tratto della personalità gradevolezza, e queste incoerenze sono usate come segnali per il rilevamento dell'inganno da parte dei giudici (Heinrich & Borkenau, 1998). Le persone estroverse ed energiche hanno più successo nell'inganno, mentre gli individui apprensivi (cioè nevrotici) non hanno lo stesso successo (Riggio, Salinas e Tucker, 1988) e gli individui che ottengono punteggi elevati in termini di gradevolezza e apertura all'esperienza sono anche più bravi a individuare gli inganni (Enoset al., 2006). Le caratteristiche utilizzate per riconoscere automaticamente l'introversione e l'estroversione sono importanti anche per identificare automaticamente l'inganno (Newman et al., 2003). È stato dimostrato che il carisma dell'oratore è fortemente correlato all'estroversione (Bono & Judge, 2004) e gli individui che dominano le riunioni hanno caratteristiche simili agli estroversi, come la verbosità (Rienks & Heylen, 2006). Mairesse ha esplorato se sia possibile acquisire modelli di personalità per i tratti della personalità dei Big Five mediante l'osservazione di output linguistici individuali nel testo e nella conversazione. Ad oggi, conosciamo solo due studi sul riconoscimento automatico della personalità dell'utente (Argamon et al., 2005; Mairesse & Walker, 2006a, 2006b; Oberlander & Nowson, 2006), oltre quello di Yla R. Tausczik and James W. Pennebaker del 2010.

I risultati di Mairesse del 2006 sono i primi a esaminare il riconoscimento della personalità nel dialogo (Mairesse & Walker, 2006a, 2006b) e ad applicare modelli di regressione e classificazione che consentono di modellare il riconoscimento della personalità utilizzando le scale continue tradizionali della psicologia.

Inoltre, gli psicologi hanno documentato l'esistenza di correlazioni tra una serie di variabili linguistiche e tratti della personalità, attraverso un'ampia gamma di livelli linguistici, inclusi parametri acustici (Smith, Brown, Strong e Rencher, 1975; Scherer, 1979), categorie lessicali (Pennebaker & King, 1999; Pennebaker, Mehl, & Niederhoffer, 2003; Mehl, Gosling, & Pennebaker, 2006; Fast & Funder, 2007), n-grammi (Oberlander & Gill, 2006) e tipo di atto linguistico (Vogel & Vogel, 1986).

Di tutti i tratti dei Big Five, l'estroversione ha ricevuto la maggiore attenzione da parte dei ricercatori. Tuttavia, gli studi che si concentrano sistematicamente su tutti i tratti dei Big Five stanno diventando sempre più comuni.

Riassumiamo i vari risultati che collegano estroversione e segnali linguistici nella Table 1, per diversi livelli di produzione linguistica come discorso, sintassi e selezione del contenuto. Una revisione di Furnham (1990) descrive le caratteristiche linguistiche legate

all'estroversione e altri tratti, e Dewaele e Furnham (1999) esaminano studi concentrandosi sul legame tra estroversione e produzione del linguaggio. I risultati includono che esiste una correlazione più elevata tra estroversione e linguaggio orale, soprattutto quando lo studio comporta un compito complesso. Gli estroversi parlano di più, più forte e in modo più ripetitivo, con meno pause ed esitazioni, hanno ritmi di parola più elevati, silenzi più brevi, una produzione verbale più elevata, un rapporto tipo/token più basso e un linguaggio meno formale, mentre gli introversi utilizzano un vocabolario più ampio (Scherer, 1979; Furnham, 1990; Gill & Oberlander, 2002). Gli estroversi usano anche parole che esprimono emozioni più positive e mostrano più accordi e complimenti rispetto agli introversi (Pennebaker & King, 1999). Gli studenti estroversi che imparano il francese come seconda lingua producono più canali secondari, hanno uno stile più implicito e una ricchezza lessicale inferiore nelle situazioni formali. Sembra che più il compito è complesso e maggiore è il livello di ansia, più facile sarà distinguere tra introversi ed estroversi (Dewaele & Furnham, 1999). Heylighen e Dewaele (2002) notano anche che l'estroversione è significativamente correlata alla contestualità, in contrapposizione alla formalità. La contestualità può essere vista come un'elevata dipendenza dalla conoscenza condivisa tra i partner di conversazione, che porta all'uso di molte espressioni deittiche come pronomi, verbi, avverbi e interiezioni, mentre il linguaggio formale è meno ambiguo e presuppone una conoscenza meno comune. Per misurare questa variazione, Heylighen e Dewaele suggeriscono l'uso di una metrica chiamata formalità, definita come:

$$\text{Formalità} = (\text{sostantivo freq} + \text{aggettivo freq} + \text{preposizione freq} + \text{articolo freq} - \text{pronomi freq} - \text{verbfreq} - \text{avverbio freq} - \text{interiezione freq} + 100) / 2$$

Sostengono che questa misura è la dimensione più importante della variazione tra le espressioni linguistiche, come mostrato nell'analisi fattoriale di Biber di vari generi (Biber, 1988). Oltre all'introversione, gli autori scoprono anche che la formalità è correlata positivamente con il livello di istruzione e la femminilità di chi parla. Le variabili situazionali legate all'uso del linguaggio formale sono la dimensione del pubblico, l'intervallo di tempo tra i dialoghi, l'indisponibilità di feedback, la differenza di background e la posizione spaziale tra i parlanti, nonché la quantità di conversazione precedente.

Level	Introvert	Extravert
Conversational behaviour	Listen Less back-channel behaviour	Initiate conversation More back-channel behaviour
Topic selection	Self-focused Problem talk, dissatisfaction Strict selection Single topic Few semantic errors Few self-references	Not self-focused* Pleasure talk, agreement, compliment Think out loud* Many topics Many semantic errors Many self-references
Style	Formal Many hedges (tentative words)	Informal Few hedges (tentative words)
Syntax	Many nouns, adjectives, prepositions (explicit) Elaborated constructions Many words per sentence Many articles Many negations	Many verbs, adverbs, pronouns (implicit) Simple constructions* Few words per sentence Few articles Few negations
Lexicon	Correct Rich High diversity Many exclusive and inclusive words Few social words Few positive emotion words Many negative emotion words	Loose* Poor Low diversity Few exclusive and inclusive words Many social words Many positive emotion words Few negative emotion words
Speech	Received accent Slow speech rate Few disfluencies Many unfilled pauses Long response latency Quiet Low voice quality Non-nasal voice Low frequency variability	Local accent* High speech rate Many disfluencies* Few unfilled pauses Short response latency Loud High voice quality Nasal voice High frequency variability

Table 1: Summary of identified language cues for extraversion and various production levels, based on previous studies by Scherer (1979), Furnham (1990), Pennebaker and King (1999), Dewaele and Furnham (1999), Gill (2003), Mehl et al. (2006). Asterisks indicate that the cue is only based on a hypothesis, as opposed to study results.

Pennebaker e King (1999) identificano molte caratteristiche linguistiche associate a ciascuno dei tratti della personalità dei Big Five. Usano lo strumento Linguistic Inquiry and Word Count (LIWC) per contare le categorie di parole dei saggi scritti da studenti la cui personalità è stata valutata utilizzando un questionario. Gli autori trovano piccole ma significative correlazioni tra le loro dimensioni linguistiche e i tratti della personalità. I nevrotici usano più pronomi di prima persona singolare, più parole emotive negative e meno parole emotive positive. D'altro canto, le persone gradevoli esprimono più emozioni positive e meno emozioni negative. Usano anche meno articoli. Le persone coscienti evitano le negazioni, le parole che esprimono emozioni negative e le parole che riflettono discrepanze (ad esempio, dovrei e vorrei). Infine, l'apertura all'esperienza è caratterizzata dalla preferenza di parole più lunghe e parole che esprimono tentazione (ad esempio, forse e forse), così come dall'evitamento dei pronomi della prima persona singolare e delle forme del tempo presente. Per quanto riguarda la prosodia, Smith et al. (1975) mostrano anche che la velocità di parola è correlata positivamente con la competenza percepita (coscientosità) e che la velocità di parola ha una relazione invertita con la benevolenza (gradevolezza), suggerendo la necessità di modelli non lineari.

Alcuni tratti hanno prodotto più risultati di altri. Una ragione potrebbe essere che alcuni si riflettono maggiormente attraverso il linguaggio, come l'estroversione. Tuttavia, è possibile che questa attenzione sia una conseguenza della correlazione dell'estroversione

con segnali linguistici che possono essere analizzati più facilmente (ad esempio, verbosità).

Pennebaker e colleghi hanno estratto una serie di caratteristiche linguistiche da ogni trascrizione del saggio e della conversazione, iniziando con il conteggio della frequenza di 88 categorie di parole del Linguistic Inquiry and WordCount (LIWC) (Pennebaker et al., 2001). Queste caratteristiche includono informazioni sia sintattiche (ad esempio, rapporto tra i pronomi) che semantiche (ad esempio, parole con emozioni positive), che sono state convalidate da giudici esperti.

Alcune caratteristiche del LIWC sono illustrate nella Tabella 5. Pennebaker e King (1999) avevano precedentemente trovato correlazioni significative tra queste caratteristiche e ciascuno dei tratti della personalità dei Big Five. Le categorie di parole rilevanti per l'estroversione includono parole sociali, parole emotive, pronomi in prima persona e verbi al presente. Poiché gli introversi impiegano più tempo a riflettere sulle proprie espressioni, Heylighen e Dewaele (2002) suggeriscono che il loro vocabolario è più ricco e più preciso, il che implica una minore frequenza di utilizzo come abbiamo già sottolineato. Il set di funzionalità MRC è stato precedentemente utilizzato da Gill e Oberlander (2004), che hanno dimostrato che l'estroversione è negativamente correlata alla concretezza. La concretezza indica anche nevroticismo (Gill & Oberlander, 2003). La tabella 5 mostra esempi di scale MRC. Ciascuna caratteristica MRC viene calcolata facendo la media del valore della caratteristica di tutte le parole nel saggio o nell'estratto della conversazione.

Feature	Type	Example
Anger words	LIWC	hate, kill, pissed
Metaphysical issues	LIWC	God, heaven, coffin
Physical state/function	LIWC	ache, breast, sleep
Inclusive words	LIWC	with, and, include
Social processes	LIWC	talk, us, friend
Family members	LIWC	mom, brother, cousin
Past tense verbs	LIWC	walked, were, had
References to friends	LIWC	pal, buddy, coworker
Imagery of words	MRC	Low: future, peace - High: table, car
Syllables per word	MRC	Low: a - High: uncompromisingly
Concreteness	MRC	Low: patience, candor - High: ship
Frequency of use	MRC	Low: duly, nudity - High: he, the

Table 5: Examples of LIWC word categories and MRC psycholinguistic features (Pennebaker et al., 2001; Coltheart, 1981). MRC features associate each word to a numerical value.

<p>LIWC FEATURES (Pennebaker et al., 2001):</p> <ul style="list-style-type: none"> - Standard counts: <ul style="list-style-type: none"> - Word count (WC), words per sentence (WPS), type/token ratio (Unique), words captured (Dic), words longer than 6 letters (Sixltr), negations (Negate), assents (Assent), articles (Article), prepositions (Preps), numbers (Number) - Pronouns (Pronoun): 1st person singular (I), 1st person plural (We), total 1st person (Self), total 2nd person (You), total 3rd person (Other) - Psychological processes: <ul style="list-style-type: none"> - Affective or emotional processes (Affect): positive emotions (Posemo), positive feelings (Posfeel), optimism and energy (Optim), negative emotions (Negemo), anxiety or fear (Anx), anger (Anger), sadness (Sad) - Cognitive Processes (Cogmech): causation (Cause), insight (Insight), discrepancy (Discrep), inhibition (Inhib), tentative (Tentat), certainty (Certain) - Sensory and perceptual processes (Senses): seeing (See), hearing (Hear), feeling (Feel) - Social processes (Social): communication (Comm), other references to people (Othref), friends (Friends), family (Family), humans (Humans) - Relativity: <ul style="list-style-type: none"> - Time (Time), past tense verb (Past), present tense verb (Present), future tense verb (Future) - Space (Space): up (Up), down (Down), inclusive (Incl), exclusive (Excl) - Motion (Motion) - Personal concerns: <ul style="list-style-type: none"> - Occupation (Occup): school (School), work and job (Job), achievement (Achieve) - Leisure activity (Leisure): home (Home), sports (Sports), television and movies (TV), music (Music) - Money and financial issues (Money) - Metaphysical issues (Metaph): religion (Relig), death (Death), physical states and functions (Physical), body states and symptoms (Body), sexuality (Sexual), eating and drinking (Eating), sleeping (Sleep), Grooming (Groom) - Other dimensions: <ul style="list-style-type: none"> - Punctuation (Allpct): period (Period), comma (Comma), colon (Colon), semi-colon (Semic), question (Qmark), exclamation (Exclam), dash (Dash), quote (Quote), apostrophe (Apostro), parenthesis (Parenth), other (Otherp) - Swear words (Swear), nonfluencies (Nonfl), fillers (Fillers)
<p>MRC FEATURES (Coltheart, 1981):</p> <p>Number of letters (Nlet), phonemes (Nphon), syllables (Nsyl), Kucera-Francis written frequency (K-F-freq), Kucera-Francis number of categories (K-F-ncats), Kucera-Francis number of samples (K-F-nsamp), Thorndike-Lorge written frequency (T-L-freq), Brown verbal frequency (Brown-freq), familiarity rating (Fam), concreteness rating (Conc), imageability rating (Imag), meaningfulness Colorado Norms (Meanc), meaningfulness Paivio Norms (Meamp), age of acquisition (AOA)</p>
<p>UTTERANCE TYPE FEATURES:</p> <p>Ratio of commands (Command), prompts or back-channels (Prompt), questions (Question), assertions (Assertion)</p>
<p>PROSODIC FEATURES:</p> <p>Average, minimum, maximum and standard deviation of the voice's pitch in Hz (Pitch-mean, Pitch-min, Pitch-max, Pitch-stddev) and intensity in dB (Int-mean, Int-min, Int-max, Int-stddev), voiced time (Voiced) and speech rate (Word-per-sec)</p>

Table 6: Description of all features, with feature labels in brackets.

Al fine di valutare quali caratteristiche individuali sono importanti per modellare la personalità indipendentemente dal modello utilizzato, ho riportato precedenti studi di correlazione per le caratteristiche LIWC sugli stessi dati così come analisi del nuovo MRC, tipo di enunciato e caratteristiche prosodiche. Le caratteristiche del LIWC erano già state analizzate da Mehl et al. (2006) per il set di dati EAR, e da Pennebaker e King (1999) per i saggi. Sorprendentemente, possiamo vedere che le funzionalità LIWC superano le funzionalità MRC per ogni caratteristica, e le funzionalità LIWC da sole funzionano sempre leggermente meglio del set completo di funzionalità. Ciò suggerisce chiaramente che le funzionalità MRC non sono utili quanto le funzionalità LIWC per classificare la personalità dal testo scritto. Per quanto riguarda gli algoritmi, troviamo che AdaboostM1 offre le prestazioni migliori per estroversione (classificazioni corrette del 56,3%), mentre SMO produce i modelli migliori per tutti gli altri tratti. Ciò suggerisce che il supporto delle macchine vettoriali è promettente per la modellazione della

personalità in generale. La caratteristica più semplice da modellare è ancora l'apertura all'esperienza, con una precisione del 62,5% utilizzando solo le funzionalità LIWC.

Successivamente, in uno studio che è stato proposto per prevedere i tratti della personalità dai microblog dei social network, Han et al. (2020) hanno scoperto che la conoscenza delle parole basata sul contesto può essere vantaggiosa per la previsione della personalità. Ritengono che, poiché i tradizionali lessici psicologici (come LIWC) sono appropriati per i testi formali, non potrebbero essere applicati in modo efficiente nei testi informali dei social network. Pertanto, hanno proposto un approccio per estrarre automaticamente un lessico della personalità dai social network, utilizzando tecniche di estrazione di parole chiave e quindi raggruppando semanticamente le parole chiave estratte. Alla fine, hanno semplicemente combinato il lessico estratto (come fonte di conoscenza precedente) con i vettori di incorporamento delle parole e li hanno inseriti in un modello di classificazione, per prevedere le etichette dei tratti della personalità dei Big Five. Hanno parzialmente migliorato la precisione della previsione, anche se hanno semplicemente sfruttato la conoscenza lessicale delle parole.

Tuttavia, un ampio e crescente corpo di letteratura ha studiato la combinazione di metodi basati sul lessico con metodi più ricchi di conoscenza che hanno migliorato proporzionalmente l'accuratezza delle previsioni. La progettazione di una rete neurale di convoluzione (CNN), che utilizza le caratteristiche Mairesse a livello di documento (estratte dal testo di input) in uno strato interno, ha costituito il focus centrale di uno studio di Majumder et al. Nella loro ricerca del 2017 hanno presentato un metodo per estrarre i tratti della personalità da saggi sul flusso di coscienza utilizzando una rete neurale convoluzionale (CNN). Hanno addestrato cinque reti diverse, tutte con la stessa architettura, per i cinque tratti della personalità. Ogni rete era un classificatore binario che prevedeva che il tratto corrispondente fosse positivo o negativo. A tal fine, hanno sviluppato una nuova tecnica di modellazione dei documenti basata su un estrattore di caratteristiche della CNN. Vale a dire, che hanno alimentato le frasi dei saggi con filtri di convoluzione per ottenere il modello della frase sotto forma di vettori di caratteristiche n-grammi. Hanno rappresentato ogni singolo saggio aggregando i vettori delle sue frasi. Hanno concatenato i vettori ottenuti con le caratteristiche Mairesse, che sono state estratte dai testi direttamente in fase di preelaborazione; questo ha migliorato le prestazioni del metodo. Scartare frasi di input emotivamente neutre dai saggi ha ulteriormente migliorato i risultati. Per la classificazione finale, hanno inserito questo vettore di documenti in una rete neurale completamente connessa con un livello

nascosto. Hanno ipotizzato che una frase rilevante contenga almeno una parola carica di emozioni così hanno scartato tutte le frasi che non avevano parole cariche di carica emotiva. È stato utilizzato il NRC Emotion Lexicon per ottenere parole cariche di emozioni. L'NRC Emotion Lexicon è un elenco di parole inglesi associate con otto emozioni di base (rabbia, paura, anticipazione, fiducia, sorpresa, tristezza, gioia e disgusto) e due sentimenti (negativo e positivo). Questo lessico contiene 14.182 parole. Hanno considerato se una parola carica emotivamente avesse almeno uno di questi attributi; ci sono 6.468 parole di questo tipo nel lessico (la maggior parte delle parole in questo lessico non hanno attributi). Quindi, se una frase non conteneva nessuna delle 6.468 parole, la rimuovevano prima di estrarre il fileword2vec caratteristiche del testo. Quindi nel loro set di dati, tutti i saggi contenevano almeno una parola emotivamente carica. La rimozione delle frasi emotivamente neutre, come descritto in precedenza, ha ulteriormente migliorato i risultati, producendo i migliori risultati per tutti e cinque i tratti. Anche il filtraggio ha migliorato il tempo di formazione del 33,3%. Le annotazioni sono state eseguite manualmente tramite crowdsourcing.

Yuan et al. hanno condotto uno studio per prevedere la personalità degli utenti dai contenuti del loro stato su Facebook. In realtà, hanno combinato le funzionalità del LIWC con funzionalità più profonde che sono state estratte attraverso un modello di deep learning. Innanzitutto, hanno estratto le caratteristiche linguistiche tramite lo strumento LIWC e poi, utilizzando una CNN, hanno estratto automaticamente le caratteristiche dai contenuti testuali. Successivamente, le due caratteristiche estratte sono state combinate per predire le etichette di personalità.

In uno studio che è stato proposto per rilevare la personalità sulla base dell'analisi del contenuto del testo, Ren et al. (2021) hanno studiato un nuovo modello di apprendimento per la previsione della personalità multi-etichetta, che combina caratteristiche emotive e semantiche. In particolare, hanno sfruttato una rappresentazione di codificatore bidirezionale di Transformers (BERT), per generare incorporamenti a livello di frase per estrarre caratteristiche semantiche dal testo, nonché un dizionario del sentiment per scopi di analisi del sentiment del testo. I codificatori sono progettati principalmente per ottenere una rappresentazione completa del testo di input. Nel loro studio hanno utilizzato il Myers-Briggs Type Indicator (MBTI) e i modelli dei tratti della personalità Big Five. I 100 marcatori unipolari di Goldberg rimangono uno dei metodi di misurazione più popolari dei tratti della personalità, in particolare i Big Five. Un'importante riduzione fu poi compiuta da Saucier, utilizzando un sottoinsieme di 40 marcatori. Entrambe le

valutazioni vengono eseguite presentando una serie di marcatori, o aggettivi, al soggetto, chiedendogli di quantificare ciascun marcatore utilizzando una scala di valutazione a 9 punti. Il Gradient Boosting fa uso di regressori per fornire un punteggio esatto dei Big Five ed è in grado di quantificare i Big Five con un errore approssimativo di 5 unità di misura.

3.6 Conclusione

Le tradizionali tecniche di machine learning e deep learning possono essere efficaci nel rilevare i tratti della personalità degli utenti e questi metodi dimostrano che i modelli basati sul linguaggio possono identificarli in modo accurato ed efficiente. Tuttavia, ci sono ancora alcune sfide: 1) Per raggiungere una comprensione profonda della semantica del testo, la maggior parte dei modelli esistenti per la codifica si basano su rappresentazioni statiche di vettori di parole, ad esempio Word2Vec (Mikolov, Chen, Corrado & Dean, 2013) e GloVe (Pennington, Socher & Manning, 2014), che non possono risolvere problemi come la polisemia di una parola. Boxman-Shabtai et al. Hanno dimostrato che la polisemia di una parola in un testo potrebbe portare ad alcuni tratti ambigui della personalità (Boxman-Shabtai & Shifman, 2014). 2) Un problema impegnativo che si pone in questo ambito è che gli attuali metodi di rilevamento della personalità sono lungi dall'essere ottimali.

Seguendo la tassonomia precedente, in primo luogo, possiamo rappresentare le proprietà degli enunciati rilevando la presenza di alcuni segnali o modelli osservabili rilevanti, come l'estrazione di rapporti tra le varie caratteristiche linguistiche (parole, n-grammi, tempi verbali, persone verbali, ecc.), sostantivi tematici da lessici emotivi (ad esempio, Mairesse et al., 2007 ; Pennebaker, Francis, & Booth, 2001), segnali sublessicali (punteggiatura, ortografia, maiuscole, numero di lettere, sillabe), o anche parole di classe chiuse come indicatori di personalità (ad esempio, pronomi personali per indicare la presenza di espansione sociale in Holtzman et al., 2019 ; Campbell & Pennebaker, 2003). Lo strumento LIWC (Pennebaker et al., 2001 ; Pennebaker & King, 1999) è l'esempio più famoso di questo tipo di paradigma e ha dimostrato la sua utilità nel rilevare stili di coping, personalità e cambiamenti di salute (Campbell & Pennebaker, 2003 ; Mitra, Conti e Pennebaker, 2016 ; Tausczik e Pennebaker, 2010). Altre opzioni per l'individuazione di segnali osservabili sono prese dal lessico NRC (Mohammad, Zhu, & Martin, 2014) o dal database psicolinguistico MRC (ad esempio, Gill, Oberlander, &

Austin, 2006). Esistono diversi modelli computazionali in cui le occorrenze delle parole sono vettorizzate algebricamente come LSA, word2vec o BEAGLE. Tutti coincidono nel senso che rappresentano il lessico in uno spazio vettoriale a dimensionalità ridotta. Il prodotto finale è che i significati latenti di questi vettori rappresentano possibili argomenti rilevanti che possono essere una descrizione estensionale del dominio della personalità (ad esempio, relazioni sociali, malattia, morte, famiglia, ecc.; vedere ad esempio, Kwantes et al. ., 2016).

Nel nostro studio ci limiteremo ad utilizzare un approccio a vocabolario chiuso (LIWC) per analizzare il linguaggio naturale dell'intera terapia e infine ipotizzare tratti e stili di personalità della paziente come contributo alla process-outcome research.

CAPITOLO 4

LA RICERCA

4.1 Considerazioni preliminari

Dopo aver delineato l'approccio usato nel corso della psicoterapia e spiegato in cosa consiste la Process–Outcome Research con il supporto dell'Intelligenza Artificiale, passiamo ora alla vera e propria parte sperimentale di questo progetto.

La Content analysis (analisi del contenuto) è una metodologia di ricerca utilizzata per analizzare e interpretare il contenuto di testi, immagini, video e altre forme di comunicazione. Questo tipo di analisi è ampiamente utilizzato nelle scienze sociali, nella psicologia, nella comunicazione, negli studi sui media e in molte altre discipline per identificare modelli, temi e significati all'interno dei dati qualitativi.

Nella pratica attuale, ci sono essenzialmente due approcci alla Content Analysis: l'approccio “top down” e l'approccio “bottom up” (Mehl e Gill 2008). L'approccio “top down” utilizza un dizionario di parole chiave e quindi, identifica e conta la presenza di parole chiave nel testo. L'approccio “bottom up” codifica tutti i concetti presenti nel testo, calcola le differenze nella frequenza con cui tali concetti emergono nel testo e rinomina i cambiamenti di significato, nel tempo o tra i testi, statisticamente significativi (Rayson, 2009). I metodi top-down (come LIWC), più diffusi, spaziano dagli approcci basati su dizionari ai più sofisticati algoritmi di apprendimento supervisionato nel mondo dell'elaborazione del linguaggio naturale. Nell'analisi che seguirà si utilizza un approccio “Top down” o approccio basato sul dizionario.

LIWC è un programma di text analysis trasparente che conta le parole in categorie psicologicamente significative ovvero quantifica il testo utilizzando la frequenza e le categorie di parole combinando metodi tradizionali di apprendimento automatico (Support Vector Machine e Naive Bayes). I punteggi forniti dal LIWC (Linguistic Inquiry and Word Count) sono percentili (basati sull'area sotto una curva normale), compresi tra 1 e 99, che rappresentano la frequenza di determinate categorie di parole nel testo analizzato. In termini statistici, questi punteggi sono descrittori quantitativi delle caratteristiche linguistiche di un testo. I punteggi sono generalmente espressi come percentuali del numero totale di parole nel testo. Ad esempio, se una categoria di parole ("positive emotion") ha un punteggio di 5, significa che il 5% delle parole nel testo

appartiene a questa categoria. Supponiamo di avere i seguenti punteggi medi per alcune categorie LIWC da un gruppo di testi:

- Emozioni Positive (positive emotion): 4.5%
- Emozioni Negative (negative emotion): 2.0%
- Analiticità (analytic): 15%
- Autenticità (authentic): 10%

In questo caso:

- 4.5% delle parole nei testi analizzati appartengono alla categoria "emozioni positive".
- 2.0% delle parole sono classificate come "emozioni negative".
- 15% indicano parole associate a pensiero analitico.
- 10% delle parole suggeriscono autenticità nel linguaggio.

I risultati empirici LIWC dimostrano la sua capacità di rilevare il significato in un'ampia varietà di contesti sperimentali, tra cui l'attenzione, l'emotività, le relazioni sociali, gli stili di pensiero e le differenze individuali (Pennebaker, Francis & Booth, 2001).

Gli algoritmi principali utilizzati da LIWC includono:

1 Tokenizzazione: il testo viene suddiviso in unità minime, generalmente parole o frasi.

- **Algoritmi:** Spaziatura e punteggiatura vengono utilizzati per determinare i confini delle parole.

2 Confronto con il Dizionario: Ogni parola tokenizzata viene confrontata con un dizionario predefinito di parole categorizzate.

- **Algoritmi:** Lookup Table, dove ogni parola trovata nel testo viene cercata in una tabella di hash o in un'altra struttura dati efficiente per il confronto.

3 Conteggio delle Parole: Le parole identificate vengono contate e classificate nelle rispettive categorie.

- **Algoritmi:** Word Count, spesso implementate come semplici cicli che incrementano i contatori per le categorie corrispondenti.

4 Calcolo delle Percentuali: Il numero di parole in ogni categoria viene trasformato in una percentuale del totale delle parole.

- **Algoritmi:** (conteggio totale parole/conteggio parole categoria) ×100

5 Analisi delle Frasi e del Contesto: In alcuni casi, LIWC può analizzare le frasi per contesto e significato più ampio, anche se questo è limitato rispetto a strumenti di elaborazione del linguaggio naturale più avanzati.

- **Algoritmi:** Algoritmi di pattern matching e regole grammaticali di base per identificare frasi e contesto.

6 Categorization Using Predefined Categories: Le parole vengono categorizzate secondo le dimensioni psicologiche, emotive e tematiche definite dal dizionario LIWC.

- **Algoritmi:** Algoritmi di categorizzazione

7 Funzioni di Post-Processing: Dopo il conteggio e la categorizzazione, LIWC può eseguire ulteriori analisi per fornire output strutturati e report.

- **Algoritmi:** Generazione di report e visualizzazione dei dati usando statistiche descrittive e altre funzioni di sintesi dei risultati.

Questi algoritmi sono implementati in modo efficiente per permettere l'analisi rapida di grandi quantità di testo. Anche se relativamente semplici, combinati insieme forniscono una potente strumentazione per l'analisi linguistica e psicologica.

In questo capitolo verranno spiegate la metodologia seguita, le analisi attuate con RStudio ed i risultati ottenuti; verranno poi analizzati qualitativamente i dati osservati. Infine, si descriveranno i limiti dello studio ed eventuali prospettive future che reputiamo interessanti.

4.1.1. Analisi testuale ed ermeneutica

Se in passato gli studi per monitorare i progressi in psicoterapia erano focalizzati sui questionari self-report, al giorno d'oggi lo sviluppo di tecniche di analisi del testo ha permesso di concentrarsi sul vero e proprio trascritto della sessione terapeutica, andando a considerare il testo nella sua interezza. I dati linguistici che vengono raccolti con la trascrizione dei colloqui possono fornire importanti informazioni circa il benessere e i

progressi del paziente. Tipicamente le sessioni psicoterapeutiche prevedono conversazioni tra paziente e terapeuta che costituiscono, quindi, un grande corpus di dati linguistici. La ricerca qualitativa è di fatto volta alla comprensione dei significati e delle esperienze a partire dal testo che il paziente porta in terapia. In questo ambito una figura molto importante fu Carl Rogers che per primo utilizzò, nella ricerca, le verbalizzazioni dei trattamenti. Già nel 1942 egli videoregistrava e trascriveva i colloqui e li utilizzava per la verifica scientifica delle proprie ipotesi. Per questo motivo, nel suo libro “Counseling and Psychotherapy” (1942), affermò di voler adottare una nuova prospettiva, che mettesse “la persona al centro”. Conseguenza di questo suo pensiero fu la scelta di rifiutare la definizione di paziente, sostituendola con quella di cliente, delineando così un rapporto più paritario tra le figure, in quanto il soggetto è il vero esperto della propria sofferenza. Con questo passaggio importante, egli fu il primo ad allontanarsi dal classico approccio meccanicistico-riduzionista per adottare un paradigma olistico/sistemico (Rogers, 1989). In seguito, nel corso della storia, altri studiosi hanno preso in considerazione il testo dei colloqui come elemento da indagare nelle ricerche. Per questo motivo, oggi è una delle metodologie più proficue per fare ricerca in ambito psicoterapeutico. La psicoterapia è una scienza ermeneutica nella misura in cui il discorso del paziente è strutturato come un testo; quindi, parte della metodologia psicoterapeutica ha lo stesso tipo di procedura dell’interpretazione dei testi (Liccione & Liccione, 2016). Tuttavia, è bene ricordare che l’applicazione di un approccio ermeneutico alla psicopatologia non si deve limitare alla comprensione dell’oggettiva struttura narrativa della storia di vita del paziente. La parola “ermeneutica”, infatti, deriva dal greco ἐρμηνευτική che significa «spiegare/tradurre/interpretare», di fatto, si tratta dell’arte di interpretare. Quotidianamente usiamo l’ermeneutica in molteplici occasioni, anche senza esserne consapevoli. Basti pensare a quando cerchiamo di capire le intenzioni sottostanti un’azione, il discorso di un soggetto che ci parla, i segnali stradali o, in generale, il linguaggio non verbale. Di fatto, l’interpretazione che noi diamo varia in base al contesto nel quale avviene l’azione. In un contesto non specifico, il gesto di mettere un pezzo di carta in una scatola potrebbe sembrare insignificante, in realtà se la stessa azione avvenisse nel contesto delle elezioni avrebbe un preciso significato: esprimere il voto inserendo la scheda elettorale nell’apposita urna. Avendo definito la psicoterapia una disciplina ermeneutica, occorre ricordare che il testo è caratterizzato da tre proprietà semantiche fondamentali: continuità, unitarietà, progressione. Per continuità si intende il lavoro della configurazione, che secondo Paul Ricoeur fa in modo che eventi diversi tra

loro vengano integrati in un unico racconto. Con il termine unitarietà è invece indicata l'identitarietà del racconto stesso. Infine, parlando di progressione ci si riferisce al concetto della temporalità, che permette di definire l'intreccio, chiamato anche narrative plot (Liccione & Liccione, 2016). Inoltre, ciò che contribuisce nel rendere un testo comprensibile è la strutturazione di esso secondo delle regole fondamentali di ordine semantico. La proprietà semantica più importante è la coerenza, essa permette che vengano svolte due operazioni importanti per ricondurre il contenuto ad una particolare architettura semantica (Ferrari, 2014). Le due operazioni citate dall'autore vengono svolte dal terapeuta e sono: la decodifica e l'inferenza. La decodifica consiste nella comprensione delle strutture linguistiche del testo, mentre l'inferenza è la capacità di formulare delle deduzioni che rendono il testo più comprensibile, partendo dai significati contestuali. L'idea di fondo è che l'analisi del testo del paziente possa rivelare cambiamenti e processi nel corso della terapia. L'utilizzo del trascritto della sessione terapeutica per intero permette di approfondire più ambiti di conversazione, senza concentrarsi solo su risposte a domande standardizzate, come invece accade con l'utilizzo degli strumenti di self-report. Consente, inoltre, di seguire il flusso del discorso del paziente, lasciando che sia lui a guidarci verso quegli aspetti che può essere utile approfondire e che risultano più importanti per il soggetto stesso. Ad esempio, il paziente può introdurre il tema della sofferenza raccontando un episodio ed è, in seguito, compito dello psicoterapeuta cercare di comprendere meglio quali temi sono motivo di sofferenza guidando il paziente nel colloquio. Inoltre, l'utilizzo delle videoregistrazioni ha aperto alla possibilità di tenere in considerazione anche il linguaggio non-verbale, come le pause, le espressioni facciali e i cambi nel tono della voce: tutti elementi che possono contribuire a rendere l'indagine più completa. Occorre, infatti, considerare che la semplice analisi delle parole dette dal paziente potrebbe non rendere conto della reale emotività sottostante che caratterizza la personalità, ad esempio nei casi in cui il paziente usa l'ironia per nascondere l'imbarazzo o la tristezza, che invece possono essere evidenti considerando il comportamento non verbale. Dopo che il racconto del paziente è stato trascritto, è possibile suddividerlo in argomenti, ovvero unità di analisi che potranno poi essere usate per le fasi successive della ricerca che sfruttano gli strumenti di Intelligenza Artificiale. Nel nostro lavoro questi argomenti vengono definiti "Topics". Di fatto, quello che si propone è una cooperazione tra tecnologia e uomo, dalla quale entrambi possono trarre benefici, pur mantenendo delle funzioni distinte (Smink, Sools, Van der Zwaan, Wiegersma, Veldkamp & Westerhof, 2019). A livello pratico, la tecnologia può fornire un

supporto per l'analisi degli elementi soggettivi che consentono di non categorizzare l'uomo considerando solo ciò che lo accomuna agli altri, ma di andare a osservare la sua individualità, i suoi modi di essere-nel-mondo e quindi a prendere in considerazione il "Chi" come soggetto di ricerca. Attraverso questi strumenti, inoltre, è possibile considerare quali fattori abbiano influito positivamente, negativamente o non hanno avuto alcun peso all'interno del processo terapeutico. L'intenzione è quella di verificare, attraverso metodi quantitativi ed oggettivi di analisi dei dati testuali, l'andamento e la scelta dei processi che hanno portato ad esiti più o meno ottimali. Nel caso di una terapia possiamo considerare come esito positivo la chiusura delle sedute psicoterapeutiche in quanto il paziente ha risolto, eliminando o gestendo, la sintomatologia negativa e ha rfigurato alcuni aspetti e stili tipici della sua personalità in modo maggiormente identitario.

4.1.2 Il Topic come unità di significato

Il topic è un movimento cooperativo che orienta il destinatario nell'individuazione delle isotopie come proprietà semantiche di un testo" (Traini, 2008). All'interno di una psicoterapia, il topic è una scelta pragmatica grazie alla quale si può stabilire di che cosa si sta parlando, quindi è un macro-tema (Traini, 2008) Il fatto che sia una scelta pragmatica sta a significare che si è in grado di scegliere all'interno di un testo, tra tutti i temi che compaiono, quelli che sono più interessanti e che possono rappresentare dei nuclei di significato. Ad esempio, leggendo *I Promessi sposi*, i topic scelti potrebbero essere il perdono, la punizione divina al tempo della peste, la trasformazione di sé, la divina provvidenza etc. Nel caso di una psicoterapia, i topic riguardano l'esistenza del paziente, ed egli li porta secondo magnificazioni e narcotizzazioni. Un topic è magnificato quando il paziente ne parla in maniera ricorrente, mentre è narcotizzato nel caso in cui è presente nella vita del paziente, ma sottoforma di esperienza non configurata. Sarà allora il terapeuta a dover porre delle domande che siano in grado di far emergere quanto è narcotizzato nel testo (Liccione, 2019). Solitamente all'inizio di una terapia il topic principale riguarda la problematica per la quale il paziente ha deciso di intraprendere il percorso psicoterapeutico, quindi con la sintomatologia nel caso in cui essa sia presente. Successivamente emergono nuovi topic, alcuni dei quali potenzialmente critici e quindi da rfigurare, negoziati di volta in volta in base agli accadimenti che intercorrono nella vita del paziente tra una seduta e l'altra (Liccione, 2019).

Alla luce di questa definizione di topic, risulta allora comprensibile l'individuazione di alcuni topic da parte di un gruppo di psicoterapeuti esperti con formazione nel campo della psicoterapia cognitiva neuropsicologica, che hanno preso parte a questo gruppo di ricerca, e che dopo l'attenta lettura di diverse psicoterapie concluse, grazie alla loro esperienza e seguendo un approccio ermeneutico fenomenologico, sono riusciti a sviluppare una tabella di riferimento contenente un certo numero di topic. Inizialmente, la tabella, sviluppata dai precedenti gruppi di ricerca, contava 14 topic definiti nucleari, poiché comuni a tutte le terapie, accompagnati ognuno da una sigla, per semplificare il lavoro di etichettatura di cui si discuterà successivamente.

N°	IDENTIFICATIVO	TEMA	DESCRIZIONE
1	FDO	Famiglia Di Origine	Informazioni e significati relativi alla famiglia di origine del paziente (padre, madre, fratelli, cugini e così via), comprese le relazioni che intrattiene con essi. Da definire in base a MCE.
2	FA	Famiglia Acquisita	Informazioni e significati relativi alla famiglia acquisita – nel corso della vita – dal paziente (marito, figli, cognati, suoceri e così via), comprese le relazioni che intrattiene con essi. Vd. MCE per capire se il topic si attiva o resta inattivo.
3	FDF	Famiglia di Fatto	Informazioni e significati relativi alla famiglia di fatto del paziente (compagno, figli e così via), comprese le relazioni che intrattiene con essi. Vd. MCE per capire se il topic si attiva o resta inattivo.
4	SV	Sistema Valoriale	Insieme di assunti, regole, norme di natura etica e morale, di natura astratta e impersonale, che regolano e guidano il comportamento della persona.
5	RA	Rapporti Affettivi	Informazioni e significati relativi ai rapporti affettivi passati, attuali.
6	SES	Sessualità	Informazioni riguardanti l'orientamento, le preferenze e il comportamento sessuale del paziente. Significati relativi all'attrazione fisica.
7	SPS	Sofferenza/Patologia/Sintomi	Informazioni e significati riguardanti la sofferenza, la sintomatologia, la psicopatologia, il disagio esistenziale passato e/o attuale del paziente.

8	OS	Occupazione/Studio	Informazioni riguardanti l'occupazione lavorativa, il percorso scolastico e/o accademico del paziente. Comprende anche le esperienze lavorative e scolastiche/ accademiche del paziente con i colleghi.
9	RS	Rapporti Sociali	Informazioni e significati relativi alle amicizie e alla rete sociale del paziente.

Tabella 1. TOPIC NUCLEARI: temi toccati in tutte le terapie benché con diversi gradi di intensità e significatività

N°	IDENTIFICATIVO	TEMA	DESCRIZIONE
10	ALT	Altro	Informazioni e significati non specificati (tendenzialmente convenevoli, organizzazione della seduta successiva e saluti). Temi altrimenti non categorizzabili e/o non particolarmente rilevanti ai fini del percorso terapeutico.
11	MCE	Macro Contesto Esistenziale	Informazioni e significati esistenziali fondamentali del paziente (anni, famiglia, occupazione, luogo di vita, relazione affettiva attuale). Raccolta ai fini della comprensione di quello che sarà poi raccontato. Comprende anche valori e cultura del paziente e della sua famiglia (religione, usanze).

Tabella 2. TOPIC NUCLEARI: Altro e Macro Contesto Esistenziale

N°	IDENTIFICATIVO	TEMA	DESCRIZIONE
12	AUT	Autostima	Informazioni e significati riguardanti le valutazioni che il paziente esprime su se stesso. Comprende giudizi, capacità, abilità (livello fisico e/o intellettuale).
13	MDE	Modi di essere	Informazioni e significati dei modi di essere emotivamente incarnati (embodied) e situati (embedded) del paziente. Emozione incarnata che emerge nell'interazione con l'alterità.
14	PRG	Progettualità	Informazioni e significativi relativi ai progetti, desideri futuri e gli orizzonti di attesa del paziente.

15	COR	Corporeità	Significati espressi dal paziente relativi al suo essere corpo vissuto, che vive e patisce, in prima persona (può inglobare più aspetti relativi alla corporeità, come aspetti motorio-percettivi, emotivi, ecc.).
----	------------	------------	--

Tabella 3. TOPIC NUCLEARI: legati ai modi d'essere, i progetti, l'autostima, la corporeità del/della paziente

N°	IDENTIFICATIVO	TEMA	DESCRIZIONE
16	RET	Relazione Terapeutica	Informazioni e significati riguardanti il percorso terapeutico in corso del paziente.
17	CT	Compito Terapeutico	Informazioni riguardanti le attività esperienziali, le tecniche e/o gli strumenti terapeutici indicati alla paziente su suggerimento del terapeuta.
18	RT	Rifigurazione Terapeutica	Rifigurazioni di alcuni dati riportati dal paziente secondo direzioni di senso identitarie o maggiormente identitarie. Le rifigurazioni sono un'analisi dei modi non sempre riconosciuti di essere-nel-mondo del paziente e disvelano nuove possibilità di esistere e nuovi orizzonti di attesa.

Tabella4. TOPIC NUCLEARI: legato alle strategie psicoterapeutiche e alla relazione terapeutica

N°	IDENTIFICATIVO	TEMA	DESCRIZIONE
	IDG	Identità di Genere	Un senso di appartenenza di una persona a un sesso o un genere con cui essa si identifica (binaria, non binaria).
	UDS	Uso di Sostanze	Il riferimento e il racconto rispetto a un uso patologico di diversi generi di sostanze (dal cibo alle bevande alle droghe alle dipendenze comportamentali) che portano a dipendenza.
	SAF	Sport / Attività fisica	Informazioni inerenti qualunque attività fisica e sportiva svolta dalla persona.
	HP	Hobbies / Passioni	Informazioni e significati rispetto alle attività svolte dal paziente durante il proprio tempo libero.
	REL	Religione / Spiritualità	Informazioni rispetto ai valori religiosi e spirituali di una persona e della sua rete sociale di riferimento (amici, famiglia primaria, marito, ecc.).
	GEN	Genitorialità	Modalità e attitudini dei genitori di stare con i figli e saper comprendere/soddisfare i loro bisogni evolutivi.
	AP	Anamnesi psicologica	Informazioni rispetto a psicoterapie pregresse, psicopatologie passate e in corso, diagnosi di precedenti specialisti, visite specialistiche passate o parallele, ecc.

Tabella 5. TOPIC SATELLITI: essenziali che non sono attrattori e possono essere specifici di una psicoterapia

Come precedentemente affermato, questa tabella è stata la base utilizzata dai precedenti gruppi di ricerca, ma già questi gruppi, con la loro ricerca, hanno lavorato per migliorarla, grazie all'utilizzo dell'algoritmo LDA. Grazie ad esso è infatti stato possibile modificare la tabella ove necessario, e trovare nuovi topic, che sono poi stati etichettati dagli psicoterapeuti esperti del presente gruppo di ricerca. In particolare, sono stati individuati dei topic comuni a tutte le terapie e per questo definiti nucleari, come in precedenza, divisi però per macroaree. Una prima novità riguarda il topic FAM, famiglia, che è stato attualmente modificato e reso maggiormente specifico; infatti, si distingue adesso tra i topic FDO, famiglia d'origine, FA, famiglia acquisita, e FDF, famiglia di fatto. Nella stessa macroarea è stato aggiunto il topic SV, Sistema Valoriale. Quale di questi topic si attiveranno sarà da definire in base al MCE, ovvero il Macro-Contesto Esistenziale. Nella macroarea relativa ai modi di essere, all'autostima, ai progetti e alla corporeità del paziente è stato aggiunto appunto il topic COR, corporeità, a sottolineare il fatto che l'essere umano è un corpo vivo che patisce. Infine, la più grande novità riguarda l'aggiunta dei topic cosiddetti satelliti, dei topic che non sono comuni a tutte le terapie, ma potrebbero essere la parte principale e caratterizzante di alcune di esse; infatti, anche se spesso non si attivano, quando invece ciò accade, diventano dei topic preponderanti. È importante sottolineare come i topic in generale presentino una componente di flessibilità, perché seppur sia possibile pensare ad una loro caratteristica rigidità, essendo alcuni nucleari e quindi relativi ad ogni terapia, vi è sempre la possibilità di aggiungerne di nuovi, toglierne altri o modificare le etichette, sia grazie all'ausilio di algoritmi come LDA, sia grazie ad una sempre crescente esperienza degli specialisti in formazione e degli psicologi e psicoterapeuti esperti che fanno parte dei gruppi di ricerca.

4.2 Il caso di I.

Il caso clinico preso in considerazione per portare avanti questa ricerca è quello di I., e consiste in una psicoterapia di 26 colloqui, tutti audio e videoregistrati previa firma del consenso informato da parte della paziente. La psicoterapia è stata svolta presso uno studio privato, da un terapeuta esperto secondo un approccio cognitivo-neuropsicologico ed ha avuto inizio nel Marzo 2021. La paziente si presenta al primo incontro puntuale, molto curata nell'aspetto, vestita in maniera semplice ma consona al contesto, e senza accompagnatori. Riferisce di avere il desiderio di cominciare un percorso di questo tipo da lungo tempo, ma non ne ha mai intrapreso uno. La motivazione che l'ha spinta a chiedere un primo colloquio riguarda varie problematiche che la paziente riscontra in

diversi ambiti della sua vita quali la famiglia, il posto di lavoro e la relazione affettiva in corso al momento della richiesta, argomento preponderante per la maggior parte delle sedute. Si può affermare che la paziente non presenta nessun tipo di patologia conclamata, o sintomatologia; quindi, il suo è un disagio di tipo esistenziale. Già dal primo colloquio il terapeuta si avvale di uno strumento terapeutico, le assegna infatti il diario emotivo, nel quale la paziente deve, durante il giorno, prendere nota degli episodi che sono per lei emotivamente significativi, sia in positivo che in negativo, e questo strumento è stato molto importante per il percorso terapeutico in quanto sono emersi degli aspetti particolarmente critici che la paziente non riesce ad affrontare in maniera identitaria. Successivamente le sono stati anche assegnati diversi compiti esperienziali. Questa psicoterapia è stata inizialmente caratterizzata da incontri a cadenza settimanale, con una distanza tra gli incontri progressivamente maggiore, da due settimane di distanza nella parte centrale della terapia, fino ad un mese di distanza nel periodo conclusivo della stessa. Attualmente la terapia è conclusa e la paziente si è riappropriata di quei modi di essere-nel-mondo non identitari che le procuravano sofferenza, grazie anche allo sviluppo di nuovi progetti e orizzonti di sé nel futuro.

4.3 Ipotesi e obiettivi di ricerca

Questa ricerca si inserisce nell'ambito di studio della Process-Outcome Research e, in particolare, nasce dalla necessità di comprendere se e in quale modo l'utilizzo di strumenti di Intelligenza Artificiale possa portare a miglioramenti nella pratica clinica, consentendo un uso combinato di analisi di tipo quantitativo e qualitativo. Negli ultimi anni, vi sono infatti sempre più tentativi di applicare il Machine Learning all'ambito della salute mentale e, in particolare, della psicoterapia.

L'obiettivo principale di questa ricerca è comprendere se e come strumenti di Intelligenza Artificiale possano supportare la pratica psicoterapeutica e possano contribuire a definire i tratti di personalità della paziente utilizzando il Five Factor Model (FFM). Questo sarà realizzato attraverso l'analisi del linguaggio della paziente con il motore di elaborazione LIWC-22, integrato con i linguaggi di programmazione R e RStudio. L'obiettivo è di delineare i tratti di personalità caratteristici della paziente basandosi sul linguaggio utilizzato nelle sessioni di terapia e focalizzandosi sulla frequenza e sul conteggio delle parole utilizzate. Il Linguistic Inquiry and Word Count (LIWC) è uno strumento di analisi del testo sviluppato per esaminare le componenti emotive, cognitive e strutturali del

linguaggio scritto o parlato. LIWC utilizza un dizionario predefinito di parole categorizzate che permette di quantificare diversi aspetti del linguaggio, come le emozioni, i processi cognitivi, gli stili linguistici e i contenuti tematici. Attraverso la tokenizzazione del testo e il confronto con questo dizionario, LIWC è in grado di fornire un'analisi dettagliata delle parole usate, identificando pattern e frequenze specifiche che possono essere correlati a vari tratti di personalità e stati psicologici. L'idea principale alla base di LIWC è che l'uso delle parole nel linguaggio quotidiano rivela i pensieri, la personalità e i sentimenti degli individui. Gli obiettivi principali di questa ricerca sono:

1. Comprendere se e come strumenti di Intelligenza Artificiale possano supportare la pratica psicoterapeutica.
2. Definire i tratti di personalità della paziente utilizzando il Five Factor Model (FFM). Questo sarà realizzato attraverso l'analisi del linguaggio della paziente con il motore di elaborazione LIWC-22, integrato con i linguaggi di programmazione R e RStudio. L'obiettivo è di ipotizzare i tratti di personalità caratteristici della paziente basandosi sullo stile linguaggio utilizzato nelle sessioni di terapia e sulla letteratura di riferimento. (vd. Paragrafo 2.3.1. e 4.5)
3. Integrare metodi di analisi quantitativa, tipici dell'uso di strumenti di Intelligenza Artificiale come LIWC, con l'analisi qualitativa, per ottenere una comprensione più completa e approfondita del paziente e del processo terapeutico

La speranza è che l'utilizzo di strumenti di Intelligenza Artificiale come LIWC possa non solo supportare i terapeuti nel loro lavoro quotidiano, ma anche contribuire in modo significativo alla Process-Outcome Research, migliorando la qualità e l'efficacia degli interventi psicoterapeutici. Integrando le analisi qualitative e quantitative, questa ricerca mira a fornire un quadro completo e integrato che possa essere utile per i terapeuti nella pratica clinica, offrendo nuove prospettive e strumenti per comprendere meglio i loro pazienti e per monitorare i progressi terapeutici in modo più accurato.

4.4 IL METODO

4.4.1 Creazione del dataset

Il testo è la base della presente ricerca, ed a partire da esso è stato creato il dataset. In particolare, il testo e quindi il dataset sono composti dai 25 colloqui video e audio registrati, previa firma del consenso informato, della psicoterapia di I., che sono stati trascritti secondo gli standard di trascrizione delle sessioni di psicoterapia delineati da Mergenthaler e Stinson (1992) nel loro lavoro “Psychotherapy Transcription Standards”. Ciò che gli autori hanno tentato di standardizzare riguarda in primis ciò che è necessario trascrivere, già a partire da tutte quelle espressioni para verbali quali espressioni di consenso, di dubbio, di esclamazione, di esitazione, che vengono trascritte a livello ortografico in modo preciso, ad esempio ah, mm, eh, ah ah. Le espressioni non verbali come, ad esempio, uno starnuto o un colpo di tosse vengono invece scritte come commenti tra parentesi. Si sottolinea anche l'importanza delle pause, che vengono indicate con un trattino se sono brevi, ma se sono più lunghe di cinque secondi vengono riportate con un commento indicante la loro durata temporale. Quando non viene compreso qualcosa del dialogo magari a causa di qualche problema di audio, le incomprensioni vengono segnate attraverso uno slash /, con annesso commento della possibile parola corretta tra parentesi. Se chi parla in un determinato momento cita una frase detta da qualcun altro, questa viene riportata all'interno delle virgolette singole, mentre se cambia modo di parlare usando una voce diversa dal solito, queste parole vengono inserite all'interno delle virgolette doppie, nelle quali possono essere riportate anche frasi e modi di dire del linguaggio comune. Per quanto riguarda la punteggiatura, viene usata anche se solitamente il linguaggio parlato non è così bene ordinato, ma il suo utilizzo si rende necessario per agevolare la lettura. A livello strutturale, ogni nuovo turno del discorso inizia in una nuova riga, e viene sempre preceduto da una lettera che indica chi sta parlando, in particolare, trattandosi di una psicoterapia i codici utilizzati saranno T per indicare i turni conversazionali in cui parla il terapeuta, e P per indicare quelli in cui a parlare è il paziente. Sono anche inserite le parole composte separate da un trattino, e i neologismi o gli errori che possono presentarsi nel discorso sono riportati così come sono stati pronunciati. Infine, una questione importante riguarda i nomi. Trattandosi infatti di una reale psicoterapia di una paziente che ha firmato il consenso informato, è necessario mantenere inviolata la privacy. Per questa ragione tutti i nomi propri di persona, luoghi o istituzioni sono stati modificati, per evitare qualsiasi riconoscibilità

della paziente. Dopo aver terminato le trascrizioni di tutte le sedute, esse sono state ricontrollate da un altro ricercatore partecipante al gruppo di ricerca, e ciò, unito al fatto di aver seguito questi standard (Mergenthaler & Stinson, 1992), ha fatto in modo che fossero disponibili delle trascrizioni più fedeli possibili al testo originale. Le trascrizioni sono state fatte su un documento di Microsoft Word, e successivamente il testo del colloquio trascritto è stato riportato su un foglio di Microsoft Excel contenente diverse colonne, che indicano ognuna un aspetto specifico del colloquio. È possibile osservarne un esempio nella tabella 5.

COLLOQUIO	BLOCCO CONVERSAZIONALE	ATTORE	TURNO	FRASE	PHRASE	TEMA1	TEMA2	TEMA3	TEMA [...]
1	1	T	1	Buongiorno.	Good Morning.	ALT			
1	1	P	2	Buongiorno.	Good Morning.	ALT			
1	1	T	3	Prego.si accomodi pure.	Please, come in.	ALT			
1	1	P	4	Grazie.	Thank you.	ALT			
1	1	T	5	Allora, lei si chiama?	So, what's your name?	MCE			

Tabella5. Esempio del foglio Excel contenente i 25 colloqui di I. Viene qui riportato uno stralcio esclusivamente del primo colloquio perché qui non sono presenti dati sensibili della paziente. Nella prima colonna è riportato il numero del colloquio, nella seconda il Blocco Conversazionale, mentre nella terza colonna è indicato l'attore. La quarta colonna è riservata al numero del turno conversazionale, nella quinta e nella sesta sono riportate le frasi del colloquio rispettivamente in italiano e in inglese, e in tutte le colonne successive possono essere inseriti i topic assegnati ad ogni turno conversazionale.

Dopo aver creato un foglio Excel per ogni colloquio come da esempio, essi sono stati uniti tutti in un unico documento. La prima colonna indica il numero del colloquio, per un totale di 25 colloqui. La colonna “Blocco Conversazionale” indica invece il macroargomento del quale si sta parlando ed ha una corrispondenza col concetto di “Macro-Thematic Unit” (MTU), sviluppato da Stinson e colleghi nel 1994. In particolare, essi fanno riferimento al concetto di Thematic Units (TU), ovvero dei segmenti che contengono un singolo concetto, nei quali è possibile dividere il testo. Una volta determinate le TU, è possibile creare le MTU, dei segmenti più grandi contenenti diverse idee raggruppabili sotto un unico tema (Stinson, Milbrath, Reidbord & Bucci, 1994). In

sostanza, il blocco conversazionale riguarda il tema che si sta affrontando in uno specifico momento del colloquio, quindi è possibile trovarne diversi all'interno se ad esempio la paziente e il terapeuta si spostano dall'affrontare l'argomento della relazione affettiva a quello del rapporto con i familiari e via dicendo. Nella terza colonna viene indicato l'attore, ovvero la persona che ha pronunciato una determinata frase, dove T identifica il terapeuta e P identifica la paziente. Nella quarta colonna vengono indicati in ordine crescente, a partire dal numero 1, i turni conversazionali. Essi possono essere rappresentati da una singola espressione, da una parola o da un'intera frase, addirittura da un'intera pagina di testo. La divisione in turni conversazionali, oltre ad essere definita in base agli scambi tra paziente e terapeuta nel corso della conversazione, corrisponde alla divisione nelle Thematic Unit citate in precedenza. Nella quinta e sesta colonna è inserito il testo del colloquio rispettivamente in italiano e in inglese. La traduzione, effettuata da un esperto madrelingua inglese, si è resa necessaria per sfruttare al meglio il software LIWC utilizzato per le analisi. Il Linguistic Inquiry and Word Count (LIWC) utilizza un dizionario predefinito per analizzare il linguaggio. Originariamente sviluppato negli Stati Uniti, il dizionario di LIWC è basato principalmente sull'inglese americano. Tuttavia, nel corso degli anni, sono state create versioni adattate del dizionario LIWC per diverse lingue e contesti culturali. Nella scheda **“Dizionario”** vi è un elenco dei dizionari interni di LIWC e di eventuali dizionari esterni. È possibile selezionare il dizionario che si desidera (Tabella 6) utilizzare per analizzare il set di dati. In alternativa, si può caricare un file di dizionario esterno dal computer, che può essere nel "vecchio" formato dizionario LIWC (.dic) o nel nuovo formato dizionario LIWC-22 (.dicx).

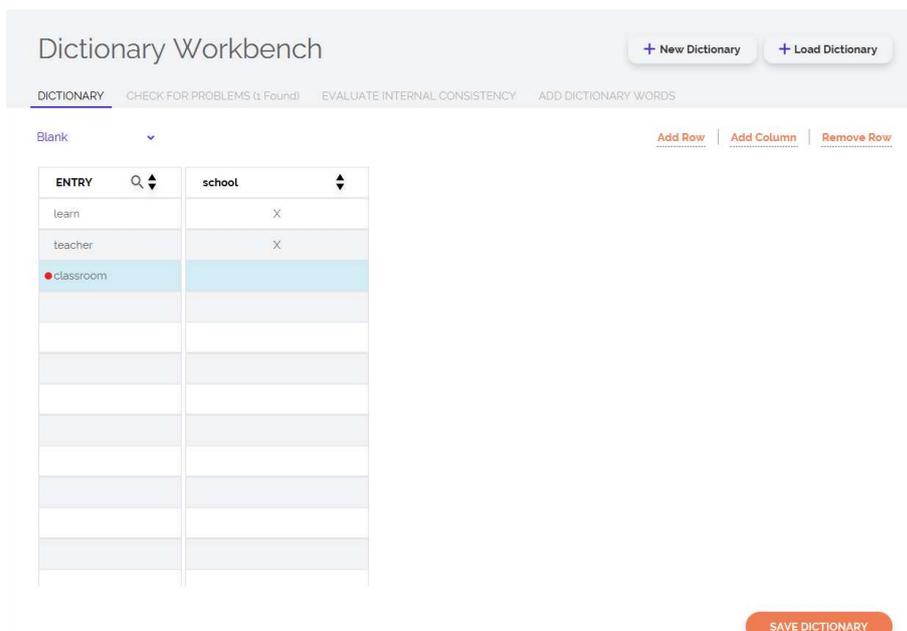


Tabella 6. Dictionary Workbench di LIWC

Per verificare, poi, la correttezza della traduzione, è stata eseguita la cosiddetta reverse translation, ovvero la traduzione dall'inglese alla lingua originale, l'italiano, ad opera di un traduttore indipendente, che non era a conoscenza del contenuto dei colloqui. Infine, nelle ultime colonne sono inseriti i topic assegnati ad ogni turno conversazionale. A differenza del metodo utilizzato dai precedenti gruppi di ricerca, che procedevano assegnando un massimo di tre topic per ogni turno conversazionale, nel presente gruppo di ricerca è stato deciso di non porre una limitazione al numero dei possibili topic assegnabili. Ciò significa che potenzialmente per un turno conversazionale sarebbe possibile assegnare fino a 5, 10, 20 topic e via dicendo, anche se vi è una grande probabilità che non si arrivi ad un tale numero, in quanto è possibile che in molti turni conversazionali sia presente un solo topic.

Come precedentemente affermato il dataset è formato dai 26 colloqui della psicoterapia di I., che sono stati divisi in cinque blocchi per consentire una maggiore diversificazione delle analisi possibili e per scandire maggiormente l'andamento della psicoterapia.

4.4.2 L'assegnazione dei topic

L'assegnazione dei topic ai singoli turni conversazionali dei colloqui psicoterapeutici è stata effettuata sulla base del metodo fenomenologico di Giorgi (1985), da tre psicoterapeuti esperti che hanno preso parte al presente gruppo di ricerca, i quali prima di procedere con l'assegnazione avevano letto il singolo colloquio per intero. Come precedentemente affermato, a differenza dei precedenti gruppi di ricerca, non vi è stato un limite al numero dei topic assegnabili ad ogni turno conversazionale. Trattandosi di un'operazione che non è immune alla soggettività, è prevista un'analisi cosiddetta Inter-Rater Reliability basata sull'indice K di Cohen (1960), utile per calcolare il grado di accordo fra gli sperimentatori. Tuttavia, in questa ricerca sono stati considerati solo i topic maggiormente frequenti dell'intera terapia che hanno contribuito a rilevare i tratti di personalità più significativi del paziente, eliminando le righe di topic poco rilevanti al fine di questa ricerca. (Tabella 7).

IDENTIFICATIVO	TEMA
MCE	Macro Contesto Esistenziale
RET	Relazione Terapeutica
SPS	Sofferenza/Patologia/Sintomi
OS	Occupazione/Studio
FDO	Famiglia d'Origine
RA	Rapporti Affettivi
AUT	Autostima
SV	Sistema Valoriale
RS	Rapporti Sociali
MDE	Modi di essere
PRG	Progettualità
FAM	Famiglia

Tabella7. Esempio di TOPIC considerati nei 25 colloqui di I. Sono considerate le frasi del colloquio, rispettivamente in inglese, con i seguenti topic assegnati ad ogni turno conversazionale.

Successivamente sono state eliminate su Excel le righe che fanno riferimenti ai blocchi conversazionali del terapeuta perché lo scopo della nostra ricerca è considerare lo stile linguistico e di personalità solo del paziente (Tabella 8). Successivamente sono state eliminate, all'interno di ciascun file, le colonne "FRASE", perché ai fini di una migliore

prestazione di LIWC-22 è bene considerare solo i blocchi conversazionali in inglese. I colloqui sono stati salvati come fogli di lavoro di Microsoft Excel (.xlsx).

COLLOQUIO	BLOCCO CONVERSAZIONALE	ATTORE	TURNO	PHRASE	TEMA1	TEMA2	TEMA3	TEMA [...]
1	1	P	1	Good Morning.	ALT			
1	1	P	3	Thank you.	ALT			

Tabella8. Esempio di foglio Excel contenente solo i blocchi conversazionali di I in inglese.

4.4.3 Il software: LIWC

LIWC è un programma di text analysis trasparente che conta le parole in categorie psicologicamente significative ovvero quantifica il testo utilizzando la frequenza e le categorie di parole combinando metodi tradizionali di apprendimento automatico (Support Vector Machine e Naive Bayes). (vd. Paragrafo 2.3). L'analisi "tradizionale" eseguita da LIWC può essere trovata tramite l'icona "Home" etichettata "Analisi LIWC". LIWC-22 è in grado di leggere testi contenuti in file .txt, MS Word o PDF separati e di analizzarli per categorie psicologiche del linguaggio. Per iniziare, LIWC-22 chiede di importare il set di dati in uno dei formati sopra descritti (Tabella 9)

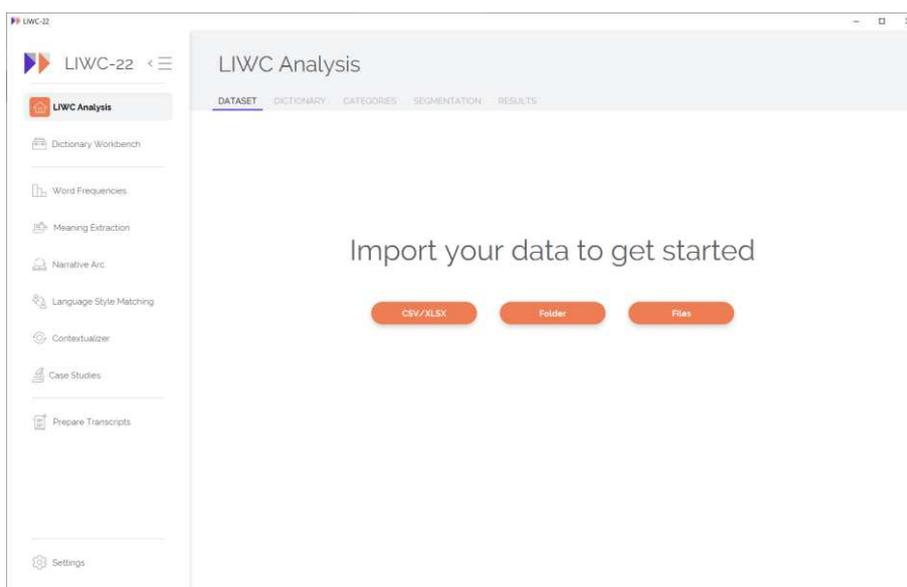


Tabella9. Anteprima "Home" etichettata "Analisi LIWC".

Successivamente LIWC-22 mostra un'anteprima del set di dati e chiede di indicare quali colonne si desidera analizzare come testo. (Tabella 9) Si può scegliere di analizzare ciascuna colonna di testo separatamente selezionando l'opzione "Tratta ogni cella come testo separato". Altrimenti, LIWC aggregherà le colonne di testo all'interno di ciascuna riga per l'analisi. Per impostazione predefinita, la prima riga è considerata una riga di "intestazione" che contiene nomi di variabili, ecc.

The screenshot shows the 'LIWC Analysis' interface. At the top, there are navigation tabs: DATASET (selected), DICTIONARY, CATEGORIES, SEGMENTATION, and RESULTS. On the right, there are links for 'Clear Dataset' and 'Replace Dataset'. Below the tabs, there are two checkboxes: 'Treat Each Cell as a Separate Text' (unchecked) and 'Use first row as a header' (checked). The main area displays a table with the following columns: TEXT (checked), GPA, GENDER, AGE, ETHNICITY, and INCOME. The table contains 14 rows of data. At the bottom, there are 'CSV options', 'DICTIONARY: LIWC-22 Dictionary (English)', 'SEGMENTATION: NONE', 'ALL CATEGORIES: YES', and a 'RUN LIWC' button. A note at the bottom right says 'First 100 rows are previewed'.

TEXT	GPA	GENDER	AGE	ETHNICITY	INCOME
All of my classes were online, so my rela...	3.9	2	27	1	
Although reluctant at first, I now love an...	3.67	1	20	1	
Back when I went to school there the ca...	3.2	1	44	1	
College life is literally what you make it ...	3	1	23	1	
College was a wake-up call for me. I fee...	3	2	39	1	
During my years at the University I had a ...	3.49	1	28	2	
First, I felt different from the school. It's a...	2.25	1	23	1	
Going to college was never my idea of fu...	3.5	2	27	3	
I am a good student in my college. I give ...	3.1	1	21	2	
I am happy with the school I am going to...	3.2	1	21	2	
I am not sure I have one. It's where I studi...	3.9	2	27	1	
I enjoy my school. It is a small school but...	3.4	2	29	1	
I enjoyed attending the college when I w...	3.75	2	16	7	

Tabella10. Anteprima LIWC Analysis dove si possono selezionare le colonne che si desidera analizzare

Nella scheda "Dizionario" è presente un elenco dei dizionari interni di LIWC-22 e di eventuali dizionari esterni utilizzati in precedenza. Puoi selezionare il dizionario che desideri utilizzare per analizzare il set di dati. In alternativa, è possibile caricare un file di dizionario esterno dal computer, che può essere nel "vecchio" formato dizionario LIWC (.dic) o nel nuovo formato dizionario LIWC-22 (.dicx). Nella nostra ricerca abbiamo utilizzato il dizionario standard: LIWC-22 Dictionary (English). La schermata "Categorie" del software LIWC è tipicamente strutturata in modo intuitivo e organizzato, per consentire agli utenti di selezionare e visualizzare le diverse categorie linguistiche e psicologiche che il software può analizzare. Elenca tutte le categorie contenute nel dizionario selezionato. In questa scheda si può scegliere di abilitare/disabilitare varie categorie di interesse. Tuttavia, non tutte le dimensioni LIWC-22 sono disponibili quando si utilizza un dizionario personalizzato; ad esempio, il calcolo della misura di riepilogo

"Pensiero analitico" si basa sui dizionari interni di LIWC-22 e non può essere calcolato quando si utilizza un file di dizionario personalizzato. Le Categorie LIWC selezionate ai fini di questa ricerca sono così organizzate:

Categorie Linguistiche:

- **Funzioni Linguistiche:**
 - Analytic: Misura del pensiero analitico e sistematico.
 - Clout: Percezione di sicurezza e autorità.
 - Authentic: Grado di genuinità e sincerità.
 - Tone: Tono emotivo generale del discorso.
 - Linguistic function: Utilizzo di funzioni linguistiche generali.
- **Pronomi:**
 - Pronoun: Uso di pronomi personali.
 - Ppron: Uso di pronomi personali (totale).
 - I: Uso del pronome "io".
 - We: Uso del pronome "noi".
 - You: Uso del pronome "tu".
 - Shehe: Uso dei pronomi "lei/lui".
 - They: Uso del pronome "essi".
 - Ipron: Uso di pronomi indefiniti.
- **Altre Particelle Grammaticali:**
 - Det: Uso di determinanti.
 - Article: Uso di articoli.
 - Prep: Uso di preposizioni.
 - Auxverb: Uso di verbi ausiliari.

Categorie Psicologiche:

- **Processi Motivazionali:**
 - Drives: Motivazioni e impulsi.
 - Affiliation: Bisogno di affiliazione e relazioni sociali.
 - Achieve: Orientamento al successo e raggiungimento degli obiettivi.
 - Power: Orientamento al potere e influenza.
- **Processi Cognitivi:**
 - Cognition: Utilizzo di parole legate ai processi cognitivi generali.
 - Cogproc: Utilizzo di parole che indicano processi cognitivi specifici.
 - Insight: Uso di parole che indicano insight e comprensione.
 - Cause: Uso di parole che indicano causalità.
 - Discrep: Uso di parole che indicano discrepanze.
- **Processi Emotivi:**
 - Affect: Espressione di emozioni generali.
 - Tone_pos: Espressione di emozioni positive.
 - Tone_neg: Espressione di emozioni negative.
 - Emotion: Espressione emotiva generale.
 - Emo_pos: Espressione di emozioni positive.
 - Emo_neg: Espressione di emozioni negative.
 - Emo_anx: Espressione di ansia.

- Emo_anger: Espressione di rabbia.
- Emo_sad: Espressione di tristezza.

Categorie Sociali:

- **Comportamenti Sociali:**
 - Social: Interazioni sociali e comportamenti.
 - Socbehav: Comportamenti sociali specifici.
 - Prosocial: Comportamenti prosociali.
 - Allure: Capacità di affascinare e coinvolgere.

Categorie Perceptive e di Attenzione:

- Perception: Riferimenti alla percezione sensoriale.
- Attention: Uso di parole che indicano attenzione.
- Space: Uso di parole che indicano spazio e relazioni spaziali.
- Focuspast: Riferimenti al passato.
- Focuspresent: Riferimenti al presente.
- Focusfuture: Riferimenti al futuro.
- Certitude: Grado di certezza nelle affermazioni.

Queste categorie sono utilizzate dal software LIWC per analizzare i testi, fornendo un'analisi dettagliata basata sulle parole e le frasi utilizzate. Le categorie selezionate aiutano a identificare e quantificare specifici aspetti linguistici, emotivi, cognitivi e sociali presenti nel trascritto terapeutico analizzato.

La scheda “Segmentazione” consente di specificare se si desidera che i testi vengano segmentati o suddivisi in blocchi prima di essere analizzati. Sono disponibili numerose opzioni di segmentazione: scegliere di dividere tutti i testi in un numero predefinito di segmenti, dividere tutti i testi in segmenti di circa N numero di parole o segmentare in base a caratteri speciali o numero di ritorni a capo. Dunque, facendo clic sul pulsante Esegui LIWC sono stati analizzati i trascritti di ciascun colloquio uno alla volta per un totale di 25 analisi; al termine dell'elaborazione, automaticamente sei indirizzato a una scheda contenente i risultati. Nella pagina Risultati abbiamo salvato il set specifico di risultati. Se si analizzano i dati in un foglio di calcolo CSV o Excel, LIWC-22 offre la possibilità di selezionare quale colonna del set di dati originale si desidera conservare/riportare nel file di output LIWC. Per impostazione predefinita, LIWC salva tutte le colonne. I punteggi forniti dal LIWC (Linguistic Inquiry and Word Count) sono percentili (basati sull'area sotto una curva normale), compresi tra 1 e 99, che rappresentano la frequenza di determinate categorie di parole nel testo analizzato. In termini statistici, questi punteggi sono descrittori quantitativi delle caratteristiche

linguistiche di un testo. I punteggi sono generalmente espressi come percentuali del numero totale di parole nel testo. Ad esempio, se una categoria di parole ("positive emotion") ha un punteggio di 5, significa che il 5% delle parole nel testo appartiene a questa categoria. La maggior parte delle variabili LIWC-22 sono ragionevolmente semplici. Ad esempio, i pronomi personali si riferiscono semplicemente alla percentuale di parole in un dato testo che sono pronomi personali. Il significato psicologico dell'uso dei pronomi personali è invece meno chiaro. Tuttavia, le parole più semplici e apparentemente banali come pronomi, articoli, verbi ausiliari, congiunzioni, preposizioni o negazioni rivestivano un ruolo fondamentale nei saggi dei partecipanti (Pennebaker, 2014). Nello specifico gli indicatori da tenere in considerazione sono soprattutto i pronomi: la grande differenza sta nell'uso del pronome personale "io" (Pennebaker, 2014).

LIWC-22 contiene quattro misure riassuntive: pensiero analitico, influenza, autenticità e tono emotivo. Ciascuna delle misure riassuntive sono algoritmi derivati da varie variabili LIWC basate su precedenti ricerche empiriche: Pensiero analitico (Pennebaker et al., 2014), Peso (Kacewicz et al., 2014), autenticità (ML Newman et al., 2003) e il tono emotivo (Cohn et al., 2004).

Le misure dipendono dalla presenza di una certa quantità di informazioni all'interno di un testo per produrre risultati significativi. Di conseguenza, diverse misure riassuntive non sono state calcolate per un dato testo poiché mancavano le informazioni necessarie. Ad esempio, considerando un testo contenente solo la parola "Stop". Dati così minimi non contengono contenuti sufficienti per determinarne il livello di pensiero analitico o di autenticità. In questi casi, LIWC-22 considera questi dati come mancanti piuttosto che fornire un punteggio impreciso che è essenzialmente privo di significato. Ciò si verifica in genere con testi molto brevi: quanto più estesi sono i dati linguistici disponibili per l'analisi, tanto maggiore è la probabilità che tutte e quattro le misure riassuntive possano essere calcolate accuratamente. La quantità di dati è il fondamento sia dell'affidabilità che dell'accuratezza nella quantificazione dei processi psicologici, che può essere incredibilmente difficile da valutare utilizzando qualsiasi metodo di ricerca. Lo stesso vale quando si usa il linguaggio per estrarre informazioni psicologiche.

La variabile del pensiero analitico è una dimensione derivata dal fattore analitico basata su diverse categorie di parole funzionali. Pubblicato originariamente come Indice Categorico-Dinamico, o CDI, il Pensiero Analitico cattura il grado in cui le persone usano parole che suggeriscono modelli di pensiero formali, logici e gerarchici. Le persone con

un basso livello di pensiero analitico tendono a scrivere e pensare utilizzando un linguaggio più intuitivo e personale. Un punteggio linguistico elevato nel pensiero analitico tende ad essere premiato in contesti accademici ed è correlato a fattori come i voti e le capacità di ragionamento. Il linguaggio con un punteggio basso nel Pensiero Analitico tende ad essere visto come meno freddo e rigido e più amichevole e gradevole. Il peso si riferisce al relativo status sociale, fiducia o leadership che le persone mostrano attraverso la loro scrittura o la loro conversazione. Il peso è una misura che trasmette la dominanza sociale. L'algoritmo Clout è stato sviluppato sulla base dei risultati di una serie di studi in cui le persone interagivano tra loro (ad esempio, Kacewicz et al., 2013). Si noti che Clout è diverso dal concetto di "Potenza" (inclusa la variabile "potenza" LIWC-22). Il potere o, più precisamente, il *bisogno* di potere, riflette l'attenzione o la consapevolezza delle persone rispetto allo status relativo in un contesto sociale. Puoi avere un leader fiducioso che non ha alcun interesse per la posizione degli altri nella gerarchia sociale. Quando le persone si rivelano in modo "autentico" o onesto, tendono a parlare in modo più spontaneo e non si autoregolamentano né filtrano ciò che dicono. L'algoritmo per l'Autenticità è stato originariamente derivato da una serie di studi in cui le persone venivano indotte a essere oneste o ingannevoli (Newman et al., 2003), nonché da una sintesi di studi sull'inganno pubblicati negli anni successivi (Pennebaker, 2014). Tuttavia, nel corso degli anni abbiamo capito che la misura dell'Autenticità ha meno a che fare con l'"inganno" in senso tradizionale ed è, invece, più un riflesso del grado in cui una persona si auto-monitora. Esempi di testi con un punteggio basso in Autenticità includono testi preparati (cioè discorsi scritti in anticipo) e testi in cui una persona si comporta. Sebbene LIWC-22 includa sia la dimensione del tono positivo che quella del tono negativo, la variabile Tono inserisce le due dimensioni in un'unica variabile di riepilogo. L'algoritmo è costruito in modo tale che maggiore è il numero, più positivo è il tono. I numeri inferiori a 50 suggeriscono un tono emotivo più negativo.

4.4.4 Le analisi con RStudio

LIWC può essere utilizzato efficacemente con RStudio. Tuttavia, è importante notare che LIWC è un software proprietario, quindi l'analisi del testo deve essere eseguita utilizzando il software LIWC stesso o utilizzando pacchetti alternativi in R che possono fornire funzionalità simili. Dopo aver ottenuto i risultati dall'analisi di LIWC, abbiamo importato e analizzato ulteriormente questi dati in RStudio in un file CSV. Di Excel. Ogni

riga rappresenta un blocco conversazionale e ogni colonna una metrica di LIWC (come “Authentic”).

Per calcolare la media dei punteggi di ciascuna categoria di LIWC tramite R Studio, abbiamo importato i dati di 5 file per volta e li abbiamo uniti in un unico data frame, per poi calcolare le medie punteggi delle categorie desiderate. I passaggi:

1. Installare i pacchetti necessari

```
install. Packages("dplyr")
```

```
install. Packages("readr")
```

2. Leggere i dati da un file CSV `liwc_data` ← `read_csv("liwc_data.csv")`
3. Importare i dati dei 5 file di colloqui per volta (dal 1° trascritto di colloquio al 5°, dal 6° al 10°, dall'11° al 15°, dal 16° al 20° e infine dal 21° al 25°)
4. Unire i dati di 5 gruppi per volta in un unico data frame. → `combined_data = pd.concat([colloquio1, colloquio2, colloquio3, colloquio4, colloquio5])`
5. _Calculate the mean for each LIWC category

```
liwc_means = combined_data.mean()
```

Lo stesso procedimento è stato effettuato per gli altri gruppi di colloqui.

4.5 ANALISI DEI COLLOQUI: UN APPROCCIO INTEGRATO

4.5.1 Analisi dei Colloqui dal 1° al 5° : Risultati e Interpretazioni

In questa sezione vengono presentati i risultati percentili delle analisi dei colloqui dal primo al 5° utilizzando il software LIWC (Tabella10). L'analisi si focalizza sulle categorie linguistiche e psicologiche emergenti dai discorsi della paziente, evidenziando i principali tratti di personalità e le emozioni predominanti. La sintesi dei colloqui fornisce una visione integrata del percorso terapeutico e dei cambiamenti osservati nel tempo.

Category	Mean		
		insight	2.5
Analytic	23.75	cause	1.33
Clout	35.95	discrep	1.37
Authentic	60.86	Affect	3.17
Tone	42.18	tone_pos	1.81
Linguistic function	67.26	tone_neg	1.11
pronoun	54.09	emotion	1.3
ppron	18.3	emo_pos	0.44
i	12.71	emo_neg	0.64
we	6.8	emo_anx	0.13
you	1.3	emo_anger	0.12
shehe	1.19	emo_sad	0.13
they	2.61	Social	11.2
ipron	5.59	socbehav	3.11
det	12.16	prosocial	0.3
article	4.82	allure	7.46
prep	9.74	Perception	6.53
auxverb	9.01	attention	0.3
Drives	4.49	space	4.28
affiliation	2.92	focuspast	4.24
achieve	0.92	focuspresent	5.21
power	0.71	focusfuture	1.11
Cognition	14.06	Certitude	0.67
cogproc	11.05		

Tabella10. Risultati percentili del valore medio di ciascuna categoria LIWC dei primi cinque trascritti terapeutici.

Il grafico mostrato nella Figura 1 è un **grafico a barre orizzontali** (barplot) in cui le categorie sono visualizzate sull'asse delle ordinate (y) e i punteggi medi su quello delle ascisse (x). Questo tipo di grafico è utile per confrontare i punteggi medi di diverse categorie in modo visivamente chiaro. Usando su Rstudio, questa è la funzione utilizzata:

```
barplot(mean_scores, names.arg = categories, horiz = TRUE, las = 2, cex.names = 0.7,
col = "steelblue",
main = "Mean Scores for Each Category 1-5", xlab = "Mean Score", ylab =
"Category")
grid (nx = NA, ny = NULL)
```

categories: Un vettore contenente i nomi delle categorie.
mean_scores: Un vettore contenente i punteggi medi per ciascuna categoria.
barplot: La funzione che crea il grafico a barre. L'argomento horiz = TRUE rende le barre orizzontali.
names.arg: Specifica i nomi delle categorie.
las = 2: Ruota le etichette delle categorie per renderle leggibili.
cex. names = 0.7: Riduce la dimensione del testo delle etichette delle categorie.
col = "steelblue": Colora le barre di blu acciaio.
main, xlab, ylab: Titoli e etichette degli assi del grafico.
grid (nx = NA, ny = NULL): Aggiunge una griglia al grafico

Tabella11. Spiegazione funzione barplot di Rstudio utilizzata per creare il grafico a barre orizzontali. Lo stesso procedimento è stato applicato a tutti i 5 gruppi di colloqui

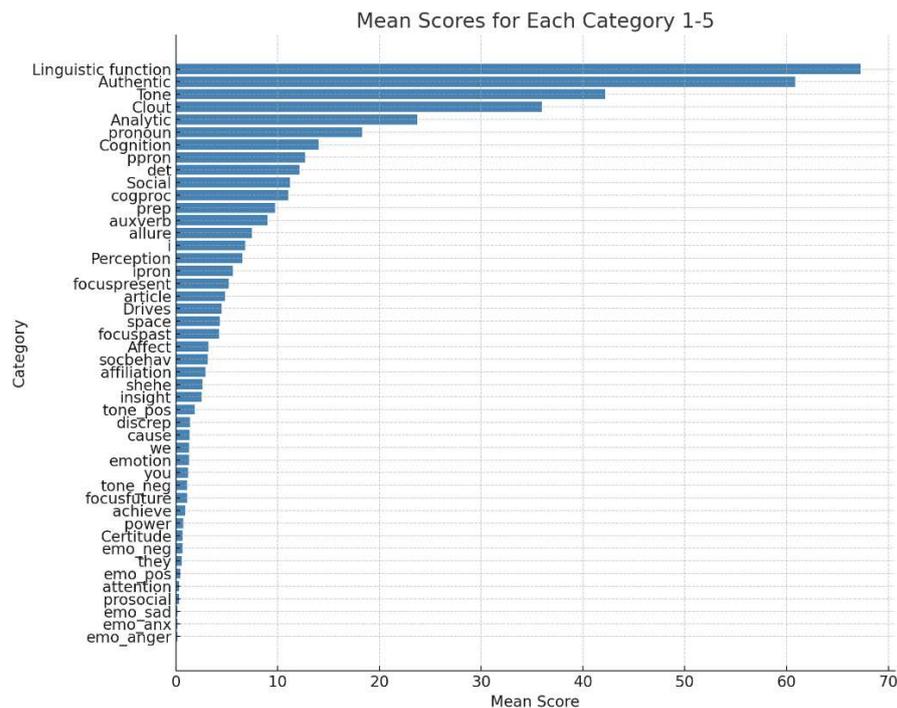


Figura 1. Punteggi Medi per Categoria nei Colloqui dal 1° al 5°.

Questo grafico mostra i punteggi medi per ciascuna categoria LIWC (Linguistic Inquiry and Word Count) calcolati attraverso i colloqui dal 1° al 5°.

- **Asse delle X:** Rappresenta i punteggi medi delle categorie, che indicano la frequenza relativa delle parole appartenenti a ciascuna categoria nel testo analizzato. I valori variano da 0 a 70.
- **Asse delle Y:** Elenca le categorie LIWC, che includono dimensioni linguistiche, psicologiche e sociali.

Le categorie sono ordinate in base ai punteggi medi, con le categorie "Linguistic function", "Authentic" e "Tone" che mostrano i valori più alti, suggerendo che queste categorie hanno avuto la maggiore rilevanza nei discorsi analizzati. Al contrario, categorie come "emo_anx" (Espressione di ansia) e "emo_anger" (Espressione di rabbia) hanno punteggi medi più bassi, indicando una minore frequenza di queste espressioni nei colloqui. Per fornire un'analisi accurata dei punteggi LIWC riportati nelle tabelle, dobbiamo esaminare i valori delle diverse categorie e ciò che possono indicare riguardo alla personalità della paziente. Ecco una valutazione di alcune categorie chiave:

Analytic (23.75): Un punteggio relativamente moderato-basso in questa categoria può suggerire che la paziente tende a un pensiero meno analitico e più intuitivo, basato sull'esperienza piuttosto che su una rigorosa analisi logica. Secondo uno studio di Pennebaker e colleghi (2014), i punteggi medi "Analytic" per scritti di studenti universitari erano intorno ai 40-50 punti, riflettendo un pensiero relativamente strutturato e logico. In contesti terapeutici, i punteggi "Analytic" possono essere più bassi, riflettendo il tono più narrativo e personale delle conversazioni terapeutiche. Un punteggio di 23.75 nella categoria "Analytic" suggerisce un livello basso-moderato di pensiero analitico nel testo. Questo indica che, sebbene ci sia una certa struttura logica e formale, il discorso non è altamente analitico. Anche se il punteggio analitico è relativamente basso, un uso frequente dei determinanti (**12.16**) può indicare che la paziente tende a fornire dettagli specifici, il che può contribuire a una maggiore chiarezza nel discorso e la comprensione delle sue espressioni.

Cause (1.33): Questi punteggi indicano un moderato livello di riflessione e di analisi delle cause nelle sue espressioni, un'organizzazione non molto elevata del pensiero e un linguaggio non molto complesso. Secondo Tausczik e Pennebaker (2010), in un'analisi di testi diversi, i punteggi per la categoria "Cause" sono solitamente intorno al 1-2% nei discorsi che coinvolgono spiegazioni e riflessioni causali. In contesti terapeutici, un punteggio di 1.33% può essere considerato relativamente tipico, indicando una presenza moderata di discussioni causali.

L'alta autenticità (60.86) riflette il grado di genuinità e sincerità nel modo di esprimersi, senza cercare di nascondere o alterare la verità per apparire in un certo modo. I punteggi delle categorie **Clout** e **Authentic** possono fornire informazioni significative sul modo in cui la paziente si percepisce e si presenta agli altri. Il punteggio di **Clout (35.95)** misura il grado di sicurezza, influenza e autorità percepita nel linguaggio. Nei contesti terapeutici, i punteggi di "Clout" possono essere più bassi, poiché i pazienti spesso esprimono incertezze e vulnerabilità. Un punteggio più alto in questa categoria generalmente indica che la persona si esprime in modo più assertivo e dominante, mentre un punteggio più basso può suggerire una comunicazione più umile o meno sicura. Secondo uno studio di Pennebaker et al. (2014), i discorsi presidenziali hanno punteggi di "Clout" relativamente alti, spesso superiori a 50, riflettendo l'autorità e la sicurezza tipiche di tali discorsi.

Nei blog personali o nelle conversazioni informali, i punteggi di "Clout" tendono ad essere più bassi, intorno a 30-40, a causa del tono più personale e meno formale. **35.95** è un punteggio intermedio-basso. Questo suggerisce che la paziente potrebbe non sentirsi particolarmente dominante o influente nelle sue comunicazioni. Potrebbe esprimersi in modo più riservato o umile e non cercare attivamente di influenzare o controllare gli altri attraverso il suo linguaggio. Combinando i punteggi di Clout e Authentic, la paziente sembra essere autentica e genuina, preferendo un approccio comunicativo sincero piuttosto che uno manipolativo. Tuttavia, il suo basso punteggio di Clout suggerisce che non si sente particolarmente influente o assertiva. Potrebbe essere una persona che esprime i propri pensieri e sentimenti in modo sincero ma con una certa modestia. Questo equilibrio tra autenticità alta e Clout basso potrebbe rendere la paziente più affidabile nelle relazioni interpersonali. Potrebbe essere percepita come una persona onesta e di cui ci si può fidare, anche se tende a non essere molto assertiva nel difendere la sua posizione. Questo potrebbe portarla a cercare soluzioni collaborative piuttosto che imporsi sugli altri. Nonostante l'alto punteggio nella categoria **Linguistic Function (67.26)**, il basso punteggio **Certitude (0.67)** in "certezza" può indicare che la paziente tende a esprimere dubbi o incertezze nelle sue affermazioni.

Social (11.2): Secondo Tausczik e Pennebaker (2010), i punteggi medi per la categoria "Social" nei testi di blog e social media possono variare da 10 a 15, a seconda della natura del contenuto e del contesto delle interazioni. In contesti terapeutici, i punteggi di "Social" possono variare in base alla dinamica della sessione e al focus della discussione. Discussioni che coinvolgono il supporto sociale, relazioni e interazioni con altre persone tendono ad avere punteggi più alti. Un punteggio di 11.2 nella categoria "Social" è considerato moderato-alto. Questo suggerisce che nel testo c'è una presenza significativa di parole che indicano interazioni sociali e relazioni. Nel contesto di trascritti terapeutici, un punteggio di 11.2 può indicare che durante la sessione ci sono state discussioni rilevanti sulle relazioni sociali del paziente, come la famiglia, gli amici, o le dinamiche sociali che influenzano il loro benessere emotivo. Questo può essere un segno positivo di esplorazione dei supporti sociali e delle relazioni interpersonali, che sono spesso elementi cruciali nel lavoro terapeutico.

Tone (42.18): La categoria "Tone" di LIWC misura il tono emotivo generale del testo, dove punteggi più alti indicano un tono più positivo, mentre punteggi più bassi indicano un tono più negativo. I punteggi di "Tone" sono normalizzati su una scala da 0 a 100.

Punteggi sopra 50 indicano un tono generalmente positivo. Ad esempio, testi come blog ottimistici o discorsi di celebrazione possono avere punteggi alti. Punteggi sotto 50 indicano un tono più negativo. Testi che discutono argomenti tristi, stressanti o conflittuali tendono ad avere punteggi bassi. **Pennebaker et al. (2015)**: Lo studio ha evidenziato che in contesti di social media, come Twitter, i punteggi di "Tone" variano ampiamente, ma tendono ad essere più positivi rispetto a contesti clinici o terapeutici. Un punteggio di 42.18 nella categoria "Tone" indica un tono moderatamente negativo. Questo suggerisce che il testo contiene più emozioni negative che positive, ma non in modo estremo. Nel contesto dei trascritti terapeutici, un punteggio di 42.18 può indicare che le discussioni durante la sessione sono state orientate verso emozioni negative o preoccupazioni, il che è comune nelle sessioni terapeutiche dove i pazienti esprimono difficoltà, stress o problemi personali. In conclusione, questo punteggio indica un tono emotivo misto, con una prevalenza di sentimenti negativi. Un tono misto può indicare che la paziente non esprime in modo predominante né emozioni positive né negative, il che può essere indicativo di un approccio bilanciato e riflessivo, tipico degli introversi.

Affect (3.17): Studi di Tausczik e Pennebaker (2010) hanno trovato che i punteggi medi di "Affect" nei blog e nei diari personali possono essere intorno al 4-6%, riflettendo l'uso più frequente di parole emozionali in questi contesti. Nei trascritti terapeutici, i punteggi di "Affect" possono variare ampiamente. Tuttavia, poiché le sessioni terapeutiche spesso coinvolgono l'espressione di emozioni, è comune vedere punteggi superiori al 3%. Un punteggio di 3.17 nella categoria "Affect" è considerato moderato. Questo suggerisce che il testo contiene una quantità ragionevole di parole emozionali, ma non è dominato da esse. Nel contesto dei trascritti terapeutici, un punteggio di 3.17 potrebbe indicare che c'è una discussione bilanciata delle emozioni, con una presenza significativa ma non predominante di parole che esprimono affetto ed emozioni. Questo può riflettere sessioni in cui il paziente esplora le proprie emozioni in modo equilibrato. Un moderato livello di affettività indica che la paziente esprime le emozioni in modo equilibrato, con un punteggio più alto nella categoria **Negative Emotion (0.64)**.

Achievement (0.92) e Power (0.71): La categoria "Achievement" misura la presenza di parole associate ai temi del successo, del raggiungimento di obiettivi e dell'ambizione, come "vincere", "successo", "realizzare", ecc.

In testi motivazionali o di autostima, i punteggi di "Achievement" tendono ad essere più alti, spesso superiori al 2-3%, a causa della frequente menzione di successi e obiettivi. Nei testi che non sono orientati al successo o all'ambizione, come articoli informativi neutrali, i punteggi di "Achievement" tendono ad essere più bassi, spesso inferiori all'1%. Nel contesto dei trascritti terapeutici, un punteggio "Achievement" di 0.92 è considerato basso-moderato e indica che c'è una menzione limitata di temi legati al successo e alla realizzazione. Questo suggerisce che le discussioni terapeutiche includono qualche riferimento agli obiettivi personali, ma non in modo predominante. Nel contesto dei trascritti terapeutici, un punteggio "Power" di 0.71 è considerato basso e indica una presenza limitata di parole associate a potere e controllo. Questo suggerisce che le discussioni terapeutiche non sono fortemente orientate verso temi di dominio e autorità.

La categoria **Pronouns (54.09)** misura l'uso totale di pronomi personali nel discorso. Nel contesto dei trascritti terapeutici, un punteggio "Pronouns" di 54.09 è considerato alto e riflette un'alta focalizzazione su esperienze personali e relazioni interpersonali. Questo è coerente con la natura delle sessioni terapeutiche, dove l'attenzione è spesso sulle esperienze soggettive dei pazienti. **Ppron (18.3)**: Indica l'uso di pronomi personali, inclusi "io", "noi", "tu", "lui/lei", "essi".

I (12.71): Un punteggio elevato in "I" indica una focalizzazione su sé stessi, suggerendo introspezione e un'attenzione ai propri pensieri e sentimenti. **We (6.8)**: Un punteggio moderato in "We" indica una certa focalizzazione sulle relazioni di gruppo e sulla collaborazione. **You (1.3)**: Un punteggio basso in "You" indica che la persona non si rivolge frequentemente direttamente agli altri nel discorso, suggerendo un approccio comunicativo meno interattivo e meno focalizzato sull'interlocutore. **Shehe (1.19)**: Un punteggio basso in "Shehe" suggerisce che la persona fa riferimenti occasionali ad altre persone specifiche, ma non in modo predominante. **They (2.61)**: Un punteggio moderato in "They" indica una focalizzazione su gruppi o terze parti non specificate, suggerendo un riferimento a entità collettive o generalizzate. L'analisi dei pronomi indica che la persona ha una forte focalizzazione su sé stessa (alto uso del pronome "io") e una moderata focalizzazione sulle relazioni di gruppo (uso del pronome "noi"). Tuttavia, tende a utilizzare meno i pronomi che si rivolgono direttamente agli altri (basso uso del pronome "tu"), suggerendo una comunicazione meno interattiva.

La teoria della personalità dei Big Five (o Cinque Grandi Fattori) descrive la personalità in termini di cinque dimensioni fondamentali: Estroversione/ Introversione, Gradevolezza (o Amicalità), Coscienziosità, Stabilità Emotiva (o Nevroticismo) e Apertura all'Esperienza. Utilizzando i punteggi delle categorie LIWC, possiamo fare alcune ipotesi. L'estroversione si riferisce a quanto una persona è socievole, energica e assertiva. Una bassa assertività (Clout) combinata a un'alta frequenza di uso di pronomi personali, soprattutto "I" (12.71), può indicare un focus su sé stessi piuttosto che sugli altri, suggerendo una bassa estroversione e una maggiore riflessione interiore. Un punteggio moderato in riferimenti sociali indica che le interazioni sociali sono presenti, ma non dominanti. **Il Focus presente (5.21)**, Pennebaker et al. (2015) hanno osservato che i testi che riflettono preoccupazioni immediate, come i blog e i diari, tendono ad avere punteggi di "Focus Present" più alti rispetto ai testi più riflessivi o analitici. Nei testi terapeutici, i pazienti possono alternare tra parlare di eventi passati, presenti e futuri, a seconda delle loro preoccupazioni attuali e degli obiettivi terapeutici. Un punteggio di 5.21 nella categoria "Focus Present" è considerato moderato. Questo suggerisce che il testo contiene una quantità significativa di parole che indicano un'attenzione al momento presente, ma non è dominato da esse. Nel contesto dei trascritti terapeutici, un punteggio di 5.21 potrebbe indicare che il paziente discute eventi e preoccupazioni attuali in modo bilanciato, senza trascurare del tutto il passato o il futuro. Una persona con un punteggio moderato in "Focus Present" potrebbe essere curiosa e interessata agli eventi attuali, mostrando una certa apertura all'esperienza e desiderio di esplorare il presente. Le interazioni sociali non sono al centro dei suoi discorsi e questo potrebbe essere un segno di introversione, dato che gli introversi spesso preferiscono interazioni meno frequenti ma più significative (Cain, S. 2013). Un punteggio alto in autenticità suggerisce che la paziente è sincera e genuina nel suo modo di esprimersi. Gli introversi spesso preferiscono conversazioni autentiche e significative rispetto a interazioni superficiali. (Cain, S. 2013). Un punteggio relativamente basso in emozioni espresse (sia positive che negative) può indicare che la paziente tende a non esternare molto le sue emozioni, preferendo un approccio più riservato, riflessivo e bilanciato. Le persone con punteggi più bassi di estroversione tendono a riportare meno emozioni positive e una minore frequenza di esperienze di piacere, il che può riflettersi in un tono generale più negativo del testo.

Conscienziosità: Un punteggio moderatamente negativo nel tono non è fortemente correlato con la conscienziosità, ma può riflettere una predisposizione a essere critici nei confronti di sé stessi e degli altri, il che è comune in persone altamente conscienziose. **La Gradevolezza** fa riferimento a quanto una persona è affabile, cooperativa e socievole. Un alto punteggio in autenticità suggerisce sincerità e trasparenza, caratteristiche tipiche di alta amicalità. La conscienziosità riguarda quanto una persona è organizzata, diligente e responsabile. **Studi di Tausczik e Pennebaker (2010)** suggeriscono che persone con elevata conscienziosità tendono ad avere punteggi più alti nella categoria "Analytic" di LIWC. Questo perché individui conscienziosi sono generalmente più meticolosi e organizzati nei loro pensieri e comunicazioni. Un punteggio di 23.75 nella categoria "Analytic" di LIWC è moderato e suggerisce un pensiero analitico presente ma non predominante. Dal punto di vista della conscienziosità, questo punteggio indica una personalità moderatamente conscienziosa, caratterizzata da un equilibrio tra organizzazione e flessibilità nel pensiero e nella comunicazione. Un punteggio moderato basso nella categoria **Analytic (23.75)** potrebbe indicare un approccio meno sistematico e più intuitivo con un basso punteggio **Achievement (0.92)** in riferimento a comportamenti orientati verso il successo o il raggiungimento degli obiettivi. Tuttavia un alto punteggio nella categoria Cognitive fa riferimento a una maggiore riflessione e analisi del pensiero. Per quanto riguarda la **Stabilità Emotiva** (Nevroticismo), il bilancio leggermente più alto di emozioni negative suggerisce una certa tendenza verso stati emotivi negativi. **Anxiety (0.13), Anger (0.12), Sadness (0.13):** questi bassi punteggi specifici indicano che, anche se presenti, i sentimenti di ansia, rabbia e tristezza non sono predominanti. La paziente potrebbe avere un livello moderato di nevroticismo, con una leggera tendenza verso emozioni negative, ma non in modo marcato. L'apertura all'esperienza riguarda quanto una persona è creativa, curiosa e aperta a nuove idee. Nel complesso, i punteggi moderati in "Affect" e alti in "Social" suggeriscono che la paziente ha un buon livello di apertura all'esperienza ed è attenta alle emozioni e alle relazioni sociali, indicando una propensione a esplorare nuovi stimoli e a essere curiosa riguardo alle esperienze emotive e sociali.

Riassumendo, sulla base dei punteggi LIWC e della teoria dei Big Five:

- Estroversione: Bassa
- Introversione: Moderata-Alta
- Gradevolezza: Moderata-Alta

- Coscienziosità: Moderata
- Stabilità Emotiva (Nevroticismo): Moderata- Bassa
- Apertura all'Esperienza: Moderata- Alta

4.5.2 Analisi dei Colloqui dal 6° al 10°: Risultati e Interpretazioni

In questa sezione vengono presentati i risultati percentili delle analisi dei colloqui dal 6° al 10° utilizzando il software LIWC (Tabella12). L'analisi si focalizza sulle categorie linguistiche e psicologiche emergenti dai discorsi della paziente, evidenziando i principali tratti di personalità e le emozioni predominanti. La sintesi dei colloqui fornisce una visione integrata del percorso terapeutico e dei cambiamenti osservati nel tempo.

Category	Mean		
Analytic	23.05	insight	2.17
Clout	35.48	cause	1.51
Authentic	57.57	discrep	1.11
Tone	37.43	Affect	3.44
Linguistic function	66.49	tone_pos	1.76
pronoun	18.13	tone_neg	1.52
ppron	12.31	emotion	1.5
i	6.26	emo_pos	0.57
we	1.41	emo_neg	0.73
you	0.74	emo_anx	0.12
shehe	2.94	emo_anger	0.22
they	0.61	emo_sad	0.15
ipron	5.82	Social	11.07
det	11.86	socbehav	3.17
article	4.43	prosocial	0.29
prep	9.59	allure	7.47
auxverb	8.42	Perception	7.01
Drives	3.73	attention	0.15
affiliation	2.42	space	4.5
achieve	0.55	focuspast	4.47
power	0.82	focuspresent	5.48
Cognition	13.6	focusfuture	1.21
cogproc	10.8	Certitude	0.88

Tabella12. Risultati percentili del valore medio di ciascuna categoria LIWC dei trascritti terapeutici dal sesto al decimo colloquio.

Il grafico mostrato nella Figura 2 è un **grafico a barre orizzontali** (barplot) in cui le categorie sono visualizzate sull'asse delle ordinate (y) e i punteggi medi su quello delle ascisse (x). Questo tipo di grafico è utile per confrontare i punteggi medi di diverse categorie in modo visivamente chiaro. Usando su Rstudio la funzione barplot con

l'argomento `horiz = TRUE` per disegnare le barre orizzontali è stato possibile creare il grafico. Questa è la funzione utilizzata:

```
barplot(mean scores, names.arg = categories, horiz = TRUE, las = 2, cex.names = 0.7,
col = "steelblue",
main = "Mean Scores for Each Category 6-10", xlab = "Mean Score", ylab =
"Category")
grid(nx = NA, ny = NULL)
```

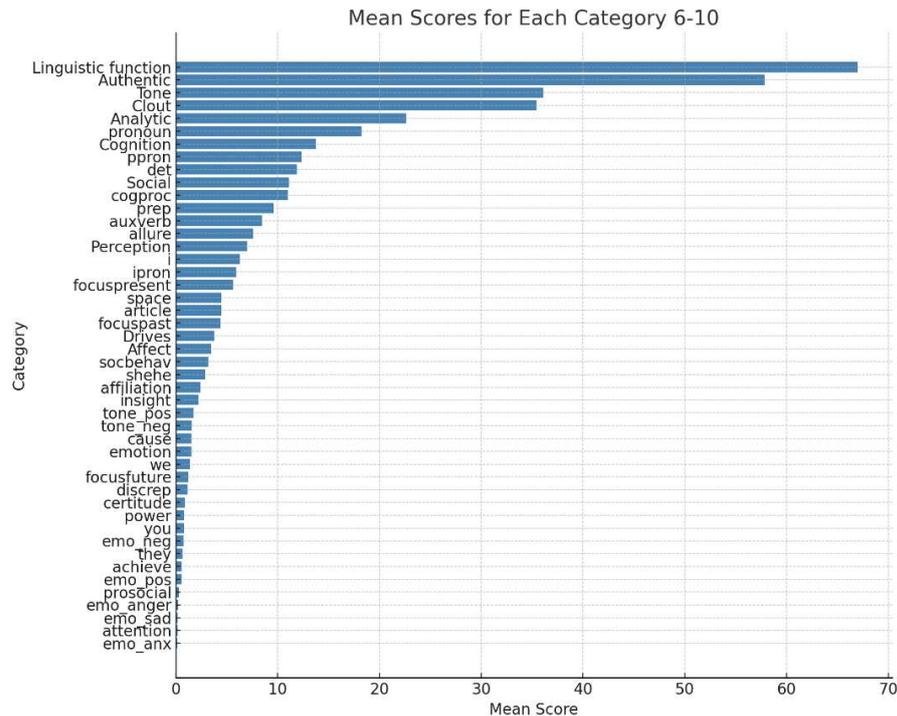


Figura2. Punteggi Medi per Categoria nei Colloqui dal 6° al 10°.

Questo grafico mostra i punteggi medi per ciascuna categoria LIWC (Linguistic Inquiry and Word Count) calcolati attraverso i colloqui dal 6° al 10°.

- **Asse delle X:** Rappresenta i punteggi medi delle categorie, che indicano la frequenza relativa delle parole appartenenti a ciascuna categoria nel testo analizzato. I valori variano da 0 a 70.
- **Asse delle Y:** Elenca le categorie LIWC, che includono dimensioni linguistiche, psicologiche e sociali.

Per fornire un'analisi accurata dei punteggi LIWC riportati nelle tabelle, dobbiamo esaminare i valori delle diverse categorie e ciò che possono indicare riguardo alla personalità della paziente. Ecco una valutazione di alcune categorie chiave:

Punteggio **Authentic: 57.57**: Un punteggio elevato in autenticità suggerisce che la paziente tende a essere onesta e trasparente nelle sue comunicazioni. Questo implica che esprime i suoi veri sentimenti e pensieri senza cercare di nasconderli o mascherarli. Le persone con punteggi elevati in autenticità spesso hanno una buona consapevolezza di sé, sono in grado di riconoscere e articolare i propri stati emotivi e pensieri in modo chiaro. La genuinità nel discorso è una qualità che può facilitare relazioni interpersonali autentiche e fiduciose. La paziente è probabilmente percepita come una persona affidabile e sincera, il che può promuovere relazioni di fiducia. Sebbene la paziente sia genuina, il suo punteggio intermedio-basso in **Clout (35.48)** indica che non tende a dominare o a essere particolarmente assertiva nelle sue comunicazioni. Questo suggerisce che la sua autenticità è accompagnata da umiltà e modestia. La paziente potrebbe essere più incline a seguire piuttosto che a guidare le interazioni sociali.

Un alto punteggio nella categoria Pronoun (18.13) suggerisce che la paziente utilizza frequentemente pronomi nel suo discorso. Questo può indicare una forte focalizzazione sulle relazioni personali e sulla descrizione delle esperienze soggettive. **Pronoun Usage (18.13)**: L'uso significativo dei pronomi, specialmente "I" (6.26), indica una focalizzazione su sé stessa. Questo può essere coerente con un'alta autenticità, in quanto la paziente parla apertamente dei propri sentimenti e esperienze personali. **We (1.41)**: Un uso relativamente basso del pronome "noi" suggerisce che la paziente potrebbe non focalizzarsi molto su esperienze condivise o su un senso di appartenenza a un gruppo. Questo può indicare una tendenza a vedere le esperienze in modo più individuale piuttosto che collettivo. **You (0.74)**: L'uso raro del pronome "tu" può suggerire che la paziente non si rivolge direttamente agli altri molto spesso nel suo discorso. Questo può indicare una comunicazione meno diretta o meno focalizzata sugli altri. **Shehe (2.94)**: Un uso moderato dei pronomi "lui" o "lei" indica che la paziente fa riferimento ad altre persone specifiche in modo significativo. **They (0.61)**: L'uso raro del pronome "loro" suggerisce che la paziente parla raramente di gruppi di persone. Questo può indicare una focalizzazione più su individui specifici piuttosto che su gruppi. **Ipron (5.82)**: L'uso moderato dei pronomi impersonali (come "it" in inglese, che potrebbe corrispondere a "esso" o "essa" in italiano) indica che la paziente utilizza anche riferimenti non personali nel suo discorso. Questo può riflettere una certa distanza emotiva o un'attenzione a oggetti e concetti oltre che a persone.

Tone (37.43) misura il tono emotivo complessivo del linguaggio utilizzato. Un punteggio alto indica un tono più positivo, mentre un punteggio basso indica un tono più negativo. Dal punto di vista dei tratti di personalità, un punteggio "Tone" di 37.43 riflette un tono emotivo generalmente negativo. Questo può indicare una predisposizione verso emozioni negative, spesso associata a un alto neuroticismo. Tuttavia, in un contesto terapeutico, questo punteggio può anche riflettere la natura del lavoro terapeutico che coinvolge l'esplorazione e l'elaborazione delle emozioni negative. La sua alta autenticità suggerisce che queste espressioni emotive sono probabilmente sincere e non filtrate. Un punteggio basso in Clout è spesso associato a una minore estroversione. La paziente potrebbe essere più introversa, preferendo ruoli di supporto piuttosto che ruoli di leadership nelle interazioni sociali e potrebbe mostrarsi più introversa, cooperativa e riflessiva, con una tendenza a evitare di imporre le proprie opinioni sugli altri. L'umiltà e la mancanza di assertività che caratterizzano un basso punteggio di Clout possono contribuire a una maggiore piacevolezza. La paziente potrebbe essere percepita come più cooperativa e meno competitiva.

La categoria **Drives** della paziente ha un punteggio medio di 3.73 e, nel LIWC include diverse sottocategorie che misurano le motivazioni e i bisogni dell'individuo, come affiliazione, realizzazione, potere, e così via. Un punteggio di 3.73 su una scala che generalmente va da 0 a 100 indica che le motivazioni della paziente, in generale, non sono molto forti o predominanti nel suo linguaggio. Questo può suggerire che i suoi discorsi non sono fortemente orientati verso bisogni specifici come affiliazione, potere o realizzazione. **Affiliation (2.42)**: Un punteggio basso in affiliazione indica che la paziente non esprime forti bisogni di appartenenza o di connessione sociale. Potrebbe non enfatizzare molto le relazioni sociali nel suo linguaggio. **Achievement (0.55)**: Un punteggio molto basso in realizzazione suggerisce che la paziente non parla frequentemente di obiettivi, traguardi o successi personali. **Power (0.82)**: Un punteggio basso in potere indica che non cerca di dominare o controllare gli altri. Non è motivata dal desiderio di avere autorità o influenza. Il basso punteggio di Clout è coerente con i punteggi bassi nelle sottocategorie di Drives; infatti, la paziente non cerca di affermare la sua autorità o di influenzare gli altri, riflettendo una mancanza di motivazione verso il potere. **Social (11.07)**: Un punteggio moderato in social indica che le interazioni sociali sono presenti, ma non centrali. Questo è coerente con il basso punteggio in affiliazione, suggerendo che le relazioni sociali sono importanti ma non sono la principale fonte di motivazione. I bassi punteggi nelle sottocategorie di Drives (specialmente affiliazione) suggeriscono una bassa estroversione. La paziente potrebbe essere più introversa e meno orientata verso l'interazione sociale e il successo esterno. Analizzando più a fondo le categorie **FocusPresent (5.48)**, **FocusFuture (1.21)** e **FocusPast(4.47)**, per comprendere come la paziente orienta il suo pensiero temporale nel linguaggio ,possiamo affermare che un punteggio relativamente alto in FocusPresent suggerisce che la paziente è incline a parlare del presente, riflettendo una concentrazione sulle esperienze e gli eventi attuali. Questo indica che la sua attenzione è principalmente rivolta a ciò che sta accadendo ora, piuttosto che a eventi passati o futuri. Un punteggio basso in FocusFuture indica che la paziente parla raramente del futuro. Questo suggerisce che non è fortemente orientata verso la pianificazione o la previsione di eventi futuri. La mancanza di focus sul futuro può riflettere una minore preoccupazione per i piani a lungo termine o per le aspettative future. Un punteggio moderato in FocusPast indica che la paziente fa riferimento agli eventi passati, ma non in modo predominante. Questo suggerisce che, sebbene il passato sia rilevante, non è il principale focus del suo discorso. La mancanza di focus sul futuro potrebbe suggerire una minore coscienziosità, in quanto la paziente potrebbe non essere

orientata verso la pianificazione e il raggiungimento di obiettivi a lungo termine. **Focus Present (5.48)**: Un'attenzione moderata al presente può indicare una persona che è consapevole e attenta alle circostanze attuali. Questo può essere correlato a tratti come la **coscienziosità**. **Focus Future (1.21)**: Un basso punteggio in focus futuro può suggerire una minore preoccupazione o pianificazione per il futuro. Questo potrebbe indicare una **bassa proattività** o una minore enfasi sulla **pianificazione a lungo termine** dunque una bassa coscienziosità. **Focus Past (4.47)**: Una moderata attenzione al passato può indicare una certa misura di riflessione e ruminazione. Questo può essere correlato con tratti come **neuroticismo** (se associato a ruminazione negativa). La paziente potrebbe tendere a riflettere sugli eventi passati, che possono influenzare le sue emozioni e comportamenti attuali.

Perception (7.01) riflette l'attenzione della persona ai dettagli sensoriali nel mondo esterno. Un punteggio di 7.01 è moderatamente alto, suggerendo che la paziente spesso fa riferimento a esperienze sensoriali nel suo linguaggio. Questo indica una consapevolezza e un'attenzione ai dettagli del mondo esterno, che può essere collegata a una maggiore attenzione e consapevolezza del contesto circostante e maggiore apertura all'esperienza. Un buon livello di Perception potrebbe suggerire che la paziente è attenta e osservatrice, capace di descrivere le sue esperienze in modo vivido e dettagliato. Questi punteggi indicano che, nonostante alcune tendenze verso l'introversione e una bassa assertività (Clout), la paziente possiede tratti comunicativi positivi che possono migliorare le sue interazioni sociali e la comprensione degli altri.

La categoria **Cognition (13.6)** misura l'uso di parole che indicano processi cognitivi generali, come pensare, sapere, imparare, ecc. Indica il livello di attività mentale e riflessione della persona. **Tausczik e Pennebaker (2010)**: Hanno trovato che i punteggi di "Cognition" sono più alti nei contesti in cui si discute di problemi complessi o si cerca di comprendere situazioni difficili. I blog, i diari personali e i testi terapeutici possono avere punteggi variabili ma spesso sopra la media nei processi cognitivi. Un punteggio di 13.6 è relativamente alto, suggerendo che la paziente utilizza frequentemente parole che riflettono il pensiero e la riflessione. Questo può indicare che è una persona riflessiva, che tende a elaborare e analizzare le informazioni in modo attivo e può essere associato a una maggiore consapevolezza e comprensione di sé e del mondo circostante. Un punteggio alto in "Cognition" è spesso correlato **con l'apertura all'esperienza**. Le persone con alta apertura tendono ad essere curiose, riflessive e propense all'analisi critica delle informazioni.

Coscienziosità: Questo tratto, che include attenzione ai dettagli e meticolosità, può anche essere correlato con un punteggio alto in "Cognition". Le persone coscienziose tendono a riflettere e analizzare attentamente le informazioni. La categoria Cognitive Processes (10.8) misura l'uso di parole specifiche che indicano processi cognitivi come causalità, insight, analisi, ragionamento, ecc. Fornisce una visione più dettagliata delle modalità specifiche di pensiero e ragionamento. Un punteggio di 10.8 indica un buon livello di utilizzo di parole che riflettono processi cognitivi complessi. Questo suggerisce che tende a ragionare e analizzare le situazioni, mostrando un buon livello di consapevolezza e comprensione dei nessi causali e delle inferenze. Questo può essere un indicatore positivo della sua capacità di elaborare e affrontare le sfide emotive e cognitive durante la psicoterapia.

LIWC fornisce punteggi percentuali basati sulla frequenza relativa delle parole in ciascuna categoria. Un punteggio del 13.6% significa che il 13.6% delle parole utilizzate nel testo rientrano nella categoria "Cognition", il che è una percentuale considerevole per i processi cognitivi, suggerendo una riflessione e un'analisi costanti.

In conclusione, la paziente mostra tratti di **introversione**, evidenziati dal basso punteggio di Clout e dall'alto uso del pronome "io", indicando una focalizzazione su sé stessa piuttosto che su interazioni sociali estese. Inoltre, il basso uso dei pronomi "tu" e "noi" suggerisce una minore inclinazione verso le interazioni sociali e l'appartenenza a gruppi. La **gradevolezza** è moderata: sebbene la paziente non mostri un forte bisogno di affiliazione, il suo alto punteggio in autenticità indica che è genuina e sincera, caratteristiche che possono favorire relazioni positive e di fiducia. La **Coscienziosità**: moderata, con capacità di pensiero strutturato e analitico. La **Stabilità Emotiva (Nevroticismo)** è moderata- bassa e L'**Apertura all'Esperienza** è moderata- alta. La paziente è riflessiva e autentica, suggerendo una certa apertura mentale e una volontà di esplorare i propri pensieri e sentimenti. Tuttavia, la mancanza di focus sul futuro può indicare una minore inclinazione verso la pianificazione e l'esplorazione di nuove idee a lungo termine.

4.5.3 Analisi dei Colloqui dall'11° al 15°: Risultati e Interpretazioni

In questa sezione vengono presentati i risultati percentili delle analisi dei colloqui dal 11° al 15° utilizzando il software LIWC (Tabella13). L'analisi si focalizza sulle categorie linguistiche e psicologiche emergenti dai discorsi della paziente, evidenziando i principali tratti di personalità e le emozioni predominanti. La sintesi dei colloqui fornisce una visione integrata del percorso terapeutico e dei cambiamenti osservati nel tempo.

Category	Mean		
Analytic	22.11	insight	2.66
Clout	32.73	cause	1.92
Authentic	60.67	discrep	1.51
Tone	35.91	Affect	3.49
Linguistic function	68.51	tone_pos	1.71
pronoun	19.53	tone_neg	1.57
ppron	13.42	emotion	1.46
i	7.00	emo_pos	0.48
we	0.75	emo_neg	0.88
you	1.94	emo_anx	0.20
shehe	2.91	emo_anger	0.29
they	0.66	emo_sad	0.07
ipron	6.10	Social	11.21
det	12.15	socbehav	3.30
article	4.52	prosocial	0.22
prep	10.30	allure	7.38
auxverb	9.04	Perception	7.22
Drives	2.76	attention	0.27
affiliation	1.79	space	4.66
achieve	0.55	focuspast	3.85
power	0.46	focuspresent	5.99
Cognition	13.83	focusfuture	1.83
cogproc	11.56	certitude	0.88

Tabella13. Risultati percentili del valore medio di ciascuna categoria LIWC dei trascritti terapeutici dall'undicesimo al quindicesimo colloquio.

Il grafico mostrato nella Figura 3 è un **grafico a barre orizzontali** (barplot) in cui le categorie sono visualizzate sull'asse delle ordinate (y) e i punteggi medi su quello delle ascisse (x). Questo tipo di grafico è utile per confrontare i punteggi medi di diverse categorie in modo visivamente chiaro. Usando su Rstudio la funzione barplot con l'argomento `horiz = TRUE` per disegnare le barre orizzontali è stato possibile creare il grafico. Questa è la funzione utilizzata:

```
barplot(mean scores, names.arg = categories, horiz = TRUE, las = 2, cex.names = 0.7,
col = "steelblue",
main = "Mean Scores for Each Category 11-15", xlab = "Mean Score", ylab =
"Category")
```

grid (nx = NA, ny = NULL)

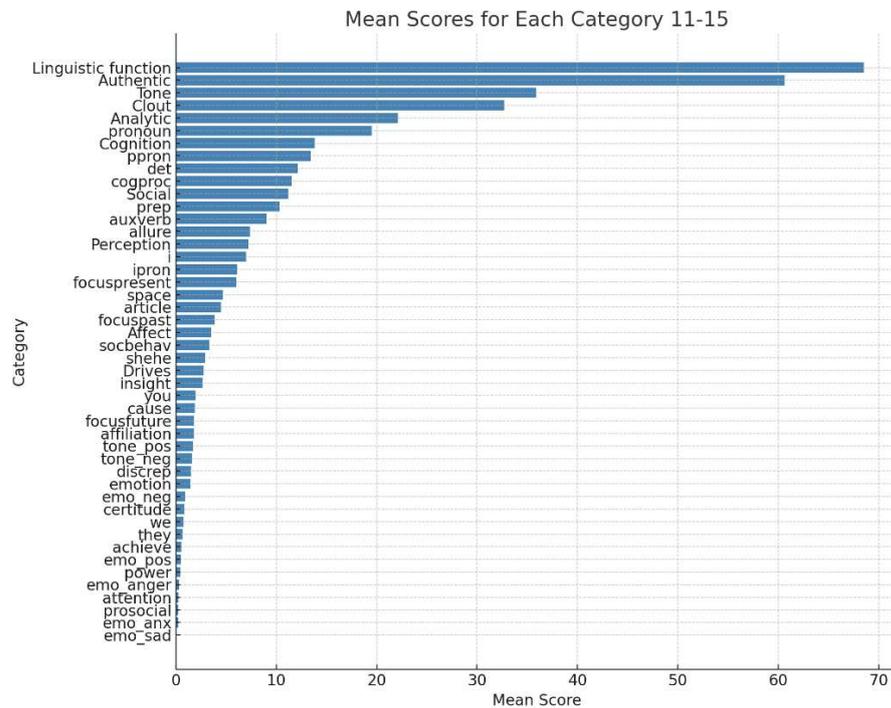


Figura3. Punteggi Medi per Categoria nei Colloqui dal 11° al 15°.

Per fornire un'analisi accurata dei punteggi LIWC riportati nelle tabelle, dobbiamo esaminare i valori delle diverse categorie e ciò che possono indicare riguardo alla personalità della paziente. Ecco una valutazione di alcune categorie chiave:

Punteggio **Authentic (60.67)** su una scala che generalmente va da 0 a 100 indica che la paziente tende a esprimersi in modo molto genuino e sincero, non teme di esprimere le sue emozioni reali, anche se queste possono includere vulnerabilità, ansie o insicurezze. Questo è un segno di buona autoconsapevolezza e di accettazione di sé stessa. Questa autenticità può avere un effetto stabilizzante sulle sue emozioni e migliorare la sua apertura mentale, pur mantenendo una comunicazione umile e non assertiva (**Clout 32.73**). La paziente potrebbe essere più incline a seguire piuttosto che a guidare le interazioni sociali. **Pronoun (19.53)** e **I (7.00)**: L'alto uso del pronome "io" è coerente con l'alta autenticità, indicando che la paziente parla spesso delle proprie esperienze e sentimenti personali in modo genuino e mostra un certo livello di consapevolezza sociale.

Affect (3.49) e **Emotion (1.46)**: Un punteggio **moderato** in affettività e emozioni indica che la paziente esprime un'ampia gamma di emozioni in modo genuino. Non esita a mostrare sia emozioni positive che negative, che è coerente con un alto punteggio di autenticità. **Tone (35.91)**: tendenza a esprimere più sentimenti negativi rispetto a quelli positivi. **Interazioni Sociali**: un punteggio di 11.21 su una scala che generalmente va da 0 a 100 indica che le interazioni sociali sono presenti nel discorso della paziente, ma non sono centrali, però è un punteggio più alto rispetto ai primi 10 colloqui. La paziente riconosce le dinamiche sociali ma potrebbe non cercare attivamente interazioni sociali frequenti o intense. La consapevolezza delle dinamiche sociali può essere associata a una maggiore apertura all'esperienza, poiché la paziente riconosce e discute una varietà di situazioni sociali. Un punteggio moderato in Social suggerisce che le interazioni sociali sono importanti per la paziente, ma il suo basso Clout indica che potrebbe preferire ruoli meno centrali nelle dinamiche di gruppo.

La categoria **Drives**, con un punteggio medio di **2.76** su una scala che generalmente va da 0 a 100 indica che le motivazioni della paziente, in generale, non sono molto forti o predominanti nel suo linguaggio. Questo suggerisce che i suoi discorsi non sono fortemente orientati verso bisogni specifici come affiliazione, potere o realizzazione. **Affiliation (1.79)**: un punteggio basso in affiliazione indica che la paziente non esprime forti bisogni di appartenenza o di connessione sociale. Potrebbe non enfatizzare molto le relazioni sociali nel suo linguaggio. Questo è coerente con un punteggio moderato in Social (11.21). **Achievement (0.55)**: Un punteggio molto basso in realizzazione suggerisce che la paziente non è fortemente motivata da obiettivi di successo o di raggiungimento e questo può riflettersi in una bassa coscienza. **Power (0.46)**: Un punteggio basso in potere indica che la paziente non cerca di dominare o controllare gli altri. Non è motivata dal desiderio di avere autorità o influenza. Questo è coerente con un punteggio basso di Clout (32.73), che indica una comunicazione più umile e meno assertiva. I bassi punteggi nelle sottocategorie di Drives suggeriscono una bassa estroversione. La paziente potrebbe essere più introversa, concentrandosi sulle proprie esperienze piuttosto che cercando il successo sociale o l'influenza.

I punteggi nelle categorie temporali ovvero FocusPresent (5.99), FocusPast(3.85) e FocusFuture(1.83), utili per comprendere come la paziente orienta il suo pensiero temporale nel linguaggio, suggeriscono che la paziente è incline a parlare del presente con un moderato riferimento al passato e una bassa preoccupazione per il futuro. Molto

probabilmente in terapia si stanno affrontando problematiche legate al momento presente con un moderato aumento del “focus future” rispetto ai colloqui precedenti. La minore enfasi sul futuro può riflettere una minore preoccupazione per la pianificazione a lungo termine e una maggiore stabilità emotiva, che non è invece presente nel presente. L'utilizzo di **You** (1.94) è aumentato rispetto ai primi 10 colloqui, il che porta a ipotizzare che si sta costruendo fiducia e alleanza con il terapeuta. L'uso moderato del pronome "tu" suggerisce che la paziente si rivolge direttamente agli altri o al terapeuta nel suo discorso, indicando una certa attenzione verso l'interazione. **Shehe (2.91)**: Un uso moderato dei pronomi "lui" o "lei" indica che la paziente fa riferimento ad altre persone specifiche in modo significativo. Questo può riflettere l'importanza delle altre persone nel suo discorso. Il punteggio nella categoria **Ipron (6.10)** in diminuzione suggerisce che la paziente inizia a focalizzare i suoi discorsi su situazioni o persone specifiche e fa meno riferimenti non personali nel suo discorso (Macro contesto esistenziale meno rilevante). Tuttavia, l'uso frequente di pronomi, in particolare "io", può indicare una forte espressione emotiva e una focalizzazione sui sentimenti personali.

La paziente mostra tratti di introversione, evidenziati dal basso punteggio di Clout e dall'alto uso del pronome "io", indicando una focalizzazione su sé stessa piuttosto che su interazioni sociali estese. Inoltre, il basso uso dei pronomi "noi" e "tu" suggerisce una minore inclinazione verso le interazioni sociali e l'appartenenza a gruppi. La gradevolezza è moderata. Sebbene la paziente non mostri un forte bisogno di affiliazione, il suo alto punteggio in autenticità indica che è genuina e sincera, caratteristiche che possono favorire relazioni positive e di fiducia. L'uso moderato di pronomi come "lui" e "lei" indica che le relazioni individuali sono comunque importanti. La coscienziosità è bassa: La paziente non è orientata verso il raggiungimento di obiettivi, come suggerito dal basso punteggio in achievement e dalla mancanza di focus sul futuro. Questo suggerisce una minore pianificazione e orientamento al successo personale. La paziente mostra una certa vulnerabilità emotiva, come indicato dalla prevalenza di emozioni miste e dal leggero sbilanciamento verso sentimenti negativi. Tuttavia, la sua alta autenticità e focalizzazione sul presente suggeriscono una capacità di esprimere e gestire le proprie emozioni in modo genuino, riducendo potenzialmente l'ansia associata alla repressione emotiva. La paziente è riflessiva e autentica, suggerendo una certa apertura mentale e una volontà di esplorare i propri pensieri e sentimenti. Tuttavia, la mancanza di focus sul futuro può indicare una minore inclinazione verso la pianificazione e l'esplorazione di nuove idee a lungo termine

però è in aumento. In sintesi, la paziente mostra una personalità introversa, genuina e riflessiva, con una buona capacità di esprimere le proprie emozioni in modo genuino e una minore focalizzazione su obiettivi e successi futuri.

4.5.4 Analisi dei Colloqui dal 16° al 20° : Risultati e Interpretazioni

In questa sezione vengono presentati i risultati percentili delle analisi dei colloqui dal quindicesimo al 20° utilizzando il software LIWC (Tabella14). L'analisi si focalizza sulle categorie linguistiche e psicologiche emergenti dai discorsi della paziente, evidenziando i principali tratti di personalità e le emozioni predominanti. La sintesi dei colloqui fornisce una visione integrata del percorso terapeutico e dei cambiamenti osservati nel tempo.

Category	Mean		
Analytic	22.24	insight	2.12
Clout	27.53	cause	1.64
Authentic	60.06	discrep	1.40
Tone	31.56	Affect	3.71
Linguistic function	66.87	tone_pos	1.70
pronoun	18.07	tone_neg	1.79
ppron	12.19	emotion	1.84
i	6.94	emo_pos	0.56
we	0.78	emo_neg	1.13
you	1.11	emo_anx	0.36
shehe	2.41	emo_anger	0.08
they	0.79	emo_sad	0.39
ipron	5.88	Social	9.66
det	12.33	socbehav	2.98
article	4.50	prosocial	0.22
prep	9.42	allure	8.28
auxverb	8.88	Perception	7.00
Drives	2.84	attention	0.34
affiliation	1.91	space	4.32
achieve	0.53	focuspast	3.80
power	0.42	focuspresent	6.05
Cognition	13.79	focusfuture	1.58
cogproc	10.71	certitude	0.73

Tabella14. Risultati percentili del valore medio di ciascuna categoria LIWC dei trascritti terapeutici dal sedicesimo al ventesimo colloquio.

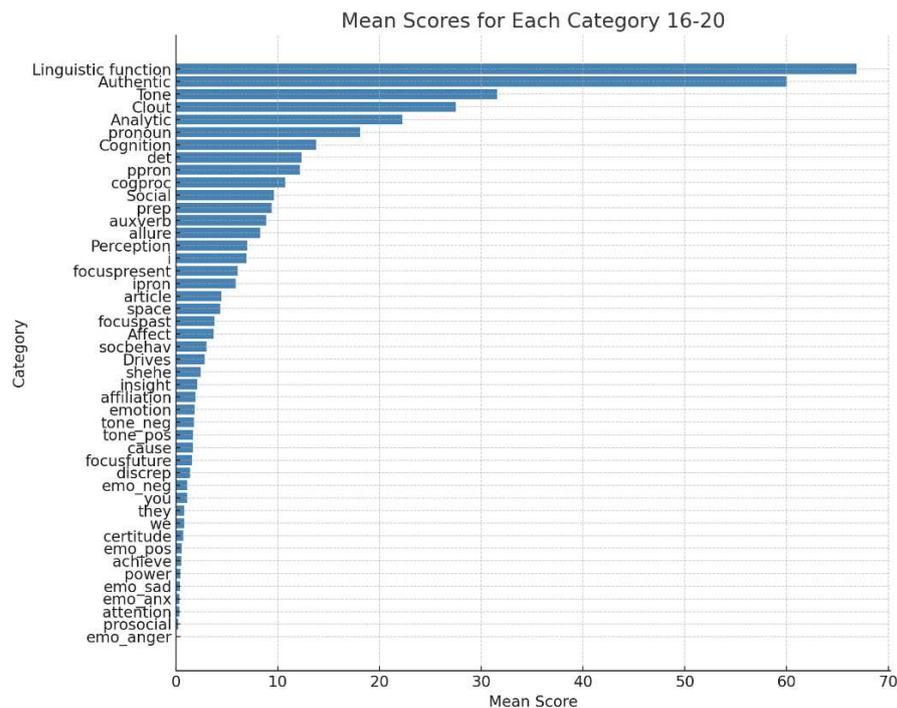


Figura4. Punteggi Medi per Categoria nei Colloqui dal 16° al 20°.

Questo grafico a barre orizzontali mostra i punteggi medi per ciascuna categoria LIWC (Linguistic Inquiry and Word Count) calcolati attraverso i colloqui dal 16° al 20°.

- **Asse delle X:** Rappresenta i punteggi medi delle categorie, che indicano la frequenza relativa delle parole appartenenti a ciascuna categoria nel testo analizzato. I valori variano da 0 a 70.
- **Asse delle Y:** Elenca le categorie LIWC, che includono dimensioni linguistiche, psicologiche e sociali.

Per fornire un'analisi accurata dei punteggi LIWC riportati nelle tabelle, dobbiamo esaminare i valori delle diverse categorie e ciò che possono indicare riguardo alla personalità della paziente. Ecco una valutazione di alcune categorie chiave:

Analytic (22.24): La paziente mostra un basso livello di pensiero analitico, suggerendo uno stile più intuitivo e meno basato sulla logica rigorosa. Un punteggio di 22.24 nella categoria "Analytic" di LIWC suggerisce un livello moderato di pensiero analitico. Dal punto di vista della personalità, questo potrebbe indicare una moderata coscienza e un bilanciamento tra pensiero analitico e narrazione personale. In un contesto terapeutico, un tale punteggio riflette un discorso che include elementi di riflessione e struttura logica,

senza essere eccessivamente formale. **Clout (27.53)**: un punteggio di 27.53 può riflettere un discorso caratterizzato da umiltà e apertura, dove il paziente può esprimere vulnerabilità e incertezze. Questo è comune e spesso positivo in un contesto terapeutico, dove l'onestà e l'apertura sono cruciali per il progresso terapeutico. Dal punto di vista dei tratti di personalità, un punteggio di 27.53 nella categoria "Clout" di LIWC riflette una bassa autorità e dominanza. Questo può indicare un livello più alto di umiltà e empatia. In un contesto terapeutico, un tale punteggio è coerente con l'espressione di vulnerabilità e la ricerca di supporto, elementi fondamentali per un'efficace terapia.

Authentic (60.06): Un punteggio alto in autenticità indica che la paziente si esprime in modo genuino e sincero. Tuttavia, il punteggio nella categoria "Tone" (31.56) **diminuisce** rispetto ai punteggi dei colloqui precedenti. **Linguistic Function (66.87)**: Un uso ampio di funzioni linguistiche, suggerendo un linguaggio complesso e ben strutturato con un livello moderato di riflessione e analisi cognitiva (Cognition 13.79). **Pronoun (18.07)** e **I (6.94)**: L'alto uso del pronome "io" indica una focalizzazione su sé stessa. Tuttavia, il basso punteggio di Clout suggerisce che questa focalizzazione non è usata per dominare le conversazioni, ma piuttosto per esprimere i propri sentimenti e pensieri. **Social (9.66)** Un punteggio basso-moderato rispetto ai colloqui precedenti. Il Punteggio **Drives: 2.84** su una scala che generalmente va da 0 a 100 indica che le motivazioni della paziente, in generale, non sono molto forti o predominanti nel suo linguaggio. Questo suggerisce che i suoi discorsi non sono fortemente orientati verso bisogni specifici come affiliazione, potere o realizzazione. Tuttavia, il punteggio nella categoria "Affiliation" è in aumento, il che suggerisce un iniziale bisogno di connessione sociale e appartenenza ad un gruppo. **FocusPresent (6.05)** indica che la paziente è incline a parlare del presente, riflettendo una concentrazione sulle esperienze e gli eventi attuali. Questo suggerisce che la sua attenzione è principalmente rivolta a ciò che sta accadendo ora, piuttosto che a eventi passati o futuri. La focalizzazione sul presente può essere trovata sia negli estroversi che negli introversi. Tuttavia, dato il basso punteggio di Clout e l'alta autenticità, è probabile che la paziente sia più introversa, concentrandosi sulle sue esperienze immediate. Tuttavia, la concentrazione sul presente e sul passato può suggerire una minore apertura all'esperienza, poiché potrebbe essere meno incline a esplorare nuove idee o a pianificare per il futuro. Tuttavia, l'autenticità e la consapevolezza emotiva indicano una certa apertura a esperienze immediate. **Shehe (2.41)**: Un uso moderato dei pronomi "lui" o "lei" indica che la paziente fa riferimento ad altre persone specifiche in

modo significativo. Questo può riflettere l'importanza delle altre persone nel suo discorso. In particolare, il pronome “Loro” (0.79) è in leggero aumento il che indica un iniziale riferimento a gruppi di persone e non solo singole persone.

Introversione: Alta.

Coscienziosità: Moderata- Bassa. La paziente sembra avere un livello di coscienziosità non molto alto, (**Analytic 22.24**) con una minore focalizzazione sugli obiettivi e il successo. **Achievement (0.53) e Power (0.42):** Punteggi bassi in queste categorie indicano che la paziente potrebbe non essere molto orientata verso il successo o il raggiungimento degli obiettivi, suggerendo una bassa coscienziosità.

Gradevolezza (Amicalità): Moderata. La paziente è sincera e trasparente, ma mostra un livello medio di comportamento cooperativo e prosociale.

Apertura all'Esperienza: Alta. La paziente mostra curiosità intellettuale, riflessione e attenzione ai dettagli sensoriali.

Stabilità Emotiva (Nevroticismo): Moderata- bassa

4.5.5 Analisi dei Colloqui dal 20° al 25°: Risultati e Interpretazioni

In questa sezione vengono presentati i risultati percentili delle analisi dei colloqui dal ventesimo al 25° utilizzando il software LIWC (Tabella15).

Category	Mean		
Analytic	20.88	insight	2.56
Clout	29.07	cause	1.64
Authentic	68.83	discrep	1.51
Tone	38.22	Affect	3.97
Linguistic function	67.80	tone_pos	1.90
pronoun	18.94	tone_neg	1.77
ppron	12.75	emotion	1.56
i	7.51	emo_pos	0.57
we	0.70	emo_neg	0.86
you	1.92	emo_anx	0.15
shehe	1.82	emo_anger	0.19
they	0.53	emo_sad	0.14
ipron	6.19	Social	9.51
det	11.98	socbehav	3.15
article	4.43	prosocial	0.30
prep	8.56	allure	8.44
auxverb	9.27	Perception	6.91
Drives	2.50	attention	0.41
affiliation	1.36	space	4.36
achieve	0.60	focuspast	4.10
power	0.60	focuspresent	6.03
Cognition	14.52	focusfuture	1.61
cogproc	11.58	certitude	0.85

Tabella15. Risultati percentili del valore medio di ciascuna categoria LIWC dei trascritti terapeutici dal ventesimo al venticinquesimo colloquio.

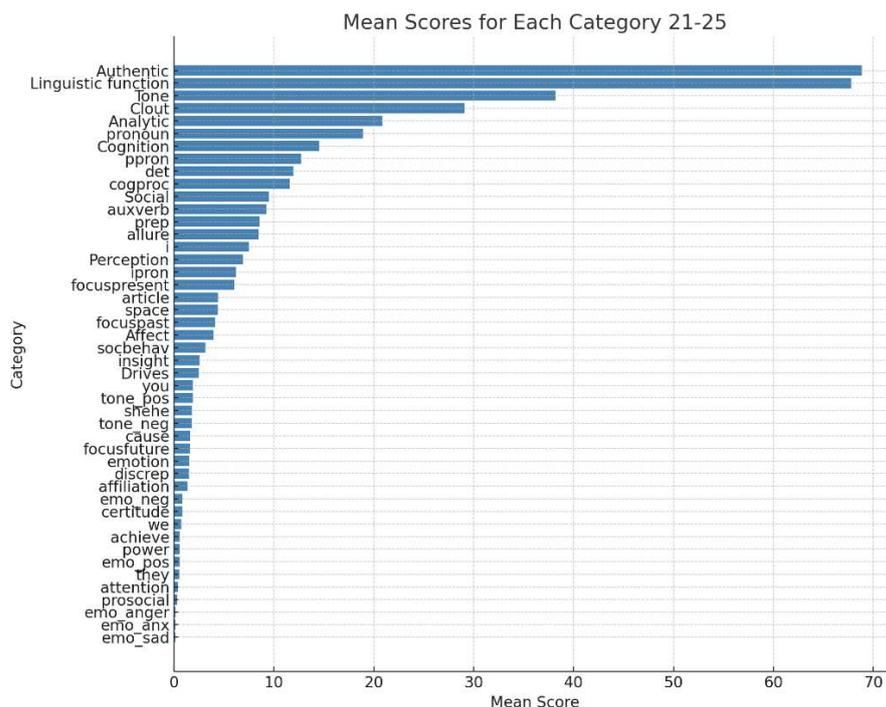


Figura5. Punteggi Medi per Categoria nei Colloqui dal 20° al 25°.

Questo grafico a barre orizzontali mostra i punteggi medi per ciascuna categoria LIWC (Linguistic Inquiry and Word Count) calcolati attraverso i colloqui dal 16° al 20°.

- **Asse delle X:** Rappresenta i punteggi medi delle categorie, che indicano la frequenza relativa delle parole appartenenti a ciascuna categoria nel testo analizzato. I valori variano da 0 a 70.
- **Asse delle Y:** Elenca le categorie LIWC, che includono dimensioni linguistiche, psicologiche e sociali.

Per fornire un'analisi accurata dei punteggi LIWC riportati nelle tabelle, dobbiamo esaminare i valori delle diverse categorie e ciò che possono indicare riguardo alla personalità della paziente. Ecco una valutazione di alcune categorie chiave:

Analytic (20.88): La paziente mostra un basso livello di pensiero analitico, suggerendo uno stile più intuitivo e meno basato sulla logica rigorosa. **Clout (29.07):** Un livello di autorità e fiducia percepita moderatamente, indicando una comunicazione più umile e meno assertiva ma in aumento. **Authentic (68.83):** Un punteggio molto alto in autenticità indica che la paziente si esprime in modo estremamente genuino e sincero verso la fine della terapia e non teme di esprimere le sue emozioni reali segno di consapevolezza e accettazione di sé stessa. **Linguistic Function (67.80):** Un uso ampio di funzioni linguistiche, suggerendo un linguaggio complesso e ben strutturato. (aumento di verbi ausiliari) ma con un livello moderato di riflessione e analisi cognitiva (**Cognition 14.52**). **Affect (3.97):** esprime le emozioni in modo equilibrato, senza una forte enfasi su emozioni positive o negative. **Socbehav (3.15):** Un punteggio moderatamente basso in comportamento sociale indica che la paziente fa riferimento a comportamenti sociali, ma non in modo frequente. Questo può suggerire che, sebbene la paziente riconosca e discuta le dinamiche sociali, non le enfatizza fortemente. **Pronoun (18.94) e Ppron (12.75):** L'uso significativo di pronomi personali, inclusi "io" (7.51) e "lui/lei" (1.82), suggerisce che la paziente parla spesso di sé stessa e degli altri, indicando un certo livello di consapevolezza sociale, con un aumento di percentuale nell'utilizzo di parole che fanno riferimento a comportamenti prosociali. Un punteggio basso in Clout (29.07), combinato con un punteggio moderato in Social, suggerisce che la paziente partecipa alle interazioni sociali in modo umile e non assertivo. **Affect (3.97) e Emotion (1.56):** Un punteggio moderato in affettività e emozioni indica che la paziente esprime un'ampia gamma di emozioni e non esita a mostrare sia emozioni positive che negative, il che è coerente con

un alto punteggio di autenticità. La combinazione con il punteggio di Tone (38.22) suggerisce che le emozioni espresse sono equilibrate. **Tone_pos (1.90)** e **Tone_neg (1.77)**: Questi punteggi specifici confermano che ci sono sia emozioni positive che negative presenti, con un aumento del tono positivo generale. La forte focalizzazione sul presente (6.03) e il basso punteggio in Certitude (0.85) suggeriscono che la paziente vive principalmente nel momento attuale, ma senza una forte convinzione o certezza nelle sue affermazioni. Questo può indicare una persona che è attenta e consapevole delle sue esperienze attuali, ma che mantiene una mente aperta e non rigida riguardo alle interpretazioni e alle possibilità. La bassa certezza nelle affermazioni riflette una comunicazione aperta e non dogmatica, suggerendo una persona che vive nel momento attuale ma che mantiene una mente aperta riguardo alle possibilità e alle interpretazioni delle sue esperienze. Questa combinazione di focalizzazione temporale e apertura mentale può influenzare vari aspetti della sua personalità, rendendola una persona riflessiva, autentica e flessibile. La paziente mostra una moderata stabilità emotiva. Sebbene ci sia una leggera prevalenza di emozioni negative, queste non sono estremamente intense. La capacità di esprimere emozioni in modo genuino e la focalizzazione sul presente contribuiscono a ridurre l'ansia e lo stress emotivo.

4.5.6 Analisi dell'Andamento della Terapia Attraverso i Punteggi LIWC: Un Approccio Longitudinale

I dati sono stati successivamente convertiti in formato lungo (long format) per essere utilizzati con ggplot2 su RStudio. Ggplot2 viene utilizzato per creare un grafico a linee che mostra l'andamento dei punteggi delle varie categorie LIWC nel tempo. I presenti grafici si propongono di analizzare l'evoluzione dei tratti di personalità di una paziente durante il corso di una terapia, utilizzando il Linguistic Inquiry and Word Count (LIWC). Questo strumento è stato utilizzato per rilevare diverse categorie linguistiche e psicologiche dai colloqui trascritti, al fine di ottenere una visione complessiva dei cambiamenti nel tempo. Il grafico è personalizzato con un tema minimal, etichette angolate per migliorare la leggibilità, e una legenda per le categorie. Il grafico è stato poi salvato come file PNG. (Figura 6). Per l'analisi, sono stati presi in considerazione i punteggi medi di 25 colloqui, suddivisi in cinque gruppi (Tabella 16).

Category	Mean 1-5	Mean 6-10	Mean 11-15	Mean 16-20	Mean 21-25
Analytic	23.75	23.05	22.11	22.24	20.88
Clear	35.05	35.48	32.73	27.53	29.07
Authentic	60.86	57.57	60.67	60.06	68.83
Tone	42.18	37.43	35.91	31.56	38.22
Linguistic function	67.26	66.49	68.51	66.87	67.80
pronoun	54.09	18.13	19.53	18.07	18.94
ppron	18.30	12.31	13.42	12.19	12.75
i	12.71	6.26	7.00	6.94	7.51
we	6.8	1.41	0.75	0.78	0.70
you	1.3	0.74	1.11	1.11	1.92
she/he	1.19	2.94	2.91	2.41	1.82
they	2.61	0.61	0.66	0.79	0.53
ipron	5.59	5.82	6.10	5.88	6.19
det	12.16	11.86	12.15	12.33	11.98
article	4.82	4.43	4.52	4.50	4.43
prep	9.74	9.59	10.30	9.42	8.56
auxverb	9.01	8.42	9.04	8.88	9.27
Drives	4.49	3.73	2.76	2.84	2.50
affiliation	2.92	2.42	1.79	1.91	1.36
achieve	0.92	0.55	0.55	0.53	0.60
power	0.71	0.82	0.46	0.42	0.60
Cognition	14.06	13.60	13.83	13.79	14.52
cogproc	11.05	10.80	11.56	10.71	11.58
insight	2.50	2.17	2.66	2.12	2.56
emtie	1.33	1.51	1.92	1.64	1.64
discrep	1.37	1.11	1.51	1.40	1.51
Affect	3.17	3.44	3.49	3.71	3.97
toae_pos	1.81	1.76	1.71	1.70	1.90
toae_neg	1.11	1.52	1.57	1.79	1.77
emotion	1.30	1.50	1.46	1.84	1.56
emo_pos	0.44	0.57	0.48	0.56	0.57
emo_neg	0.64	0.73	0.88	1.13	0.86
emo_anx	0.13	0.12	0.20	0.36	0.15
emo_anger	0.12	0.22	0.29	0.08	0.19
emo_sad	0.13	0.15	0.07	0.39	0.14
Social	11.20	11.07	11.21	9.66	9.51
soctbehav	3.11	3.17	3.30	2.98	3.15
prosocial	0.30	0.29	0.22	0.22	0.30
allure	7.46	7.47	7.38	8.28	8.44
Perception	6.53	7.01	7.22	7.00	6.91
attention	0.30	0.15	0.27	0.34	0.41
space	4.28	4.50	4.66	4.32	4.36
focuspast	4.24	4.47	3.85	3.80	4.10
focuspresent	5.21	5.48	5.99	6.05	6.03
focusfuture	1.11	1.21	1.83	1.58	1.61
Centrude	0.67	0.88	0.88	0.73	0.85

Tabella 16. Percentuali medie di frequenza di parole all'interno di ciascuna categoria LIWC di riferimento nei 5 gruppi di colloqui analizzati (1-5; 6-10; 11-15; 16-20; 21-25)

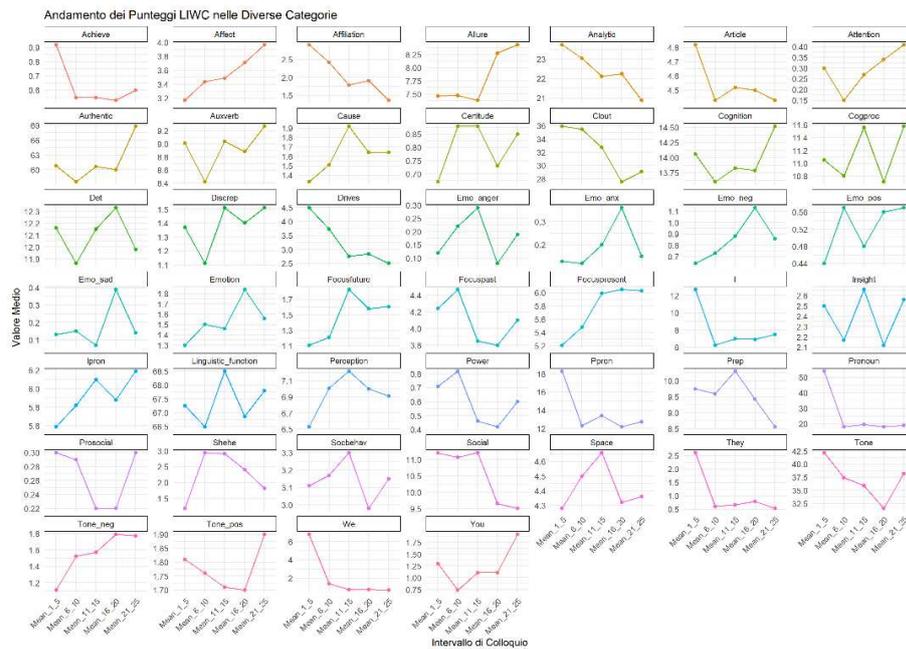


Figura 6. Questi grafici mostrano l'evoluzione dei punteggi medi LIWC in varie categorie linguistiche e psicologiche nel corso dei 25 colloqui della paziente. Ogni grafico rappresenta una specifica categoria LIWC. L'Asse X: Intervallo di Colloquio (Mean 1-5, Mean 6-10, Mean 11-15, Mean 16-20, Mean 21-25), l'Asse Y: Valore Medio del Punteggio LIWC per ciascuna categoria. I grafici mostrano come le diverse dimensioni psicologiche e linguistiche cambiano durante la terapia, fornendo una visione dettagliata del progresso e delle dinamiche della paziente nel corso del trattamento.

Per fornire un'analisi accurata dei punteggi LIWC riportati nei grafici, abbiamo esaminato i valori delle diverse categorie e ciò che possono indicare riguardo alla personalità della paziente. Il punteggio "Analytic" mostra una leggera diminuzione nel tempo, passando da 23.75 a 20.88. Questo calo indica una possibile riduzione nell'uso di un linguaggio analitico durante la terapia il che potrebbe suggerire che la paziente abbia usato un linguaggio meno strutturato e più spontaneo nel tempo. Questo potrebbe essere un segno di un aumento del comfort e della naturalezza durante le sessioni di terapia. Una riduzione nel linguaggio analitico può indicare anche una diminuzione del bisogno di controllo e precisione, suggerendo che la paziente potrebbe sentirsi più sicura e meno bisognosa di analizzare ogni dettaglio e potrebbe anche indicare una maggiore espressività emotiva ovvero si concentra più sulle sue emozioni e meno sull'analisi logica dei suoi pensieri e comportamenti. Tuttavia, è necessario considerare il contesto delle sessioni di terapia e qualsiasi evento significativo che potrebbe aver influenzato il linguaggio della paziente. Il punteggio "Tone" mostra una diminuzione generale nel tempo, passando da 42.18 a 31.56 prima di risalire a 38.22. Questo andamento indica una fluttuazione emotiva durante la terapia con periodi di maggiore e minore positività. Un calo iniziale del tono potrebbe indicare un periodo di maggiore negatività o stress. La risalita successiva potrebbe suggerire un miglioramento nel tono emotivo generale durante le ultime sessioni. Le variazioni nel tono potrebbero riflettere il processo di adattamento della paziente alla terapia e la sua risposta ai diversi temi trattati durante le sessioni. I punteggi nelle emozioni positive oscillano leggermente ma tendono a crescere nel tempo. I punteggi delle emozioni negative mostrano un leggero aumento da 0.64 a 0.86 con un picco di 0.88 al centro della terapia. L'aumento delle emozioni negative potrebbe indicare che la paziente ha affrontato e riconosciuto sentimenti negativi durante la terapia, il che può essere un segno positivo di consapevolezza emotiva. La drastica riduzione nell'uso dei pronomi suggerisce un cambiamento significativo di prospettiva della paziente. L'uso stabile del pronome "io" indica una continua focalizzazione su sé stessa e la terapia potrebbe aver incoraggiato la paziente a esplorare più profondamente i propri sentimenti e pensieri individuali, piuttosto che focalizzarsi sulle dinamiche di gruppo; mentre la riduzione nell'uso del pronome "noi" potrebbe indicare una diminuzione nel senso di connessione con gli altri, o un cambiamento nel modo di vedere le sue relazioni sociali. Potrebbero riflettere un cambiamento nel contesto sociale della paziente, come un allontanamento o cambiamento nelle dinamiche familiari o amicali. Questi cambiamenti qualitativi offrono uno spunto su come la paziente sta evolvendo nel corso della terapia,

sia a livello emotivo che nel modo in cui si percepisce e interagisce con gli altri. L'utilizzo del pronome "You" suggerisce un andamento fluttuante, con un aumento nell'ultimo intervallo, il che potrebbe suggerire una maggiore attenzione verso gli altri o una comunicazione più interattiva. Il punteggio stabile del pronome "io" indica che la paziente continua a focalizzarsi su se stessa, riflettendo costantemente sulle proprie esperienze e sentimenti. La riduzione marcata dell'uso del pronome "noi" potrebbe indicare che la percezione del gruppo è cambiata e potrebbe anche riflettere un cambiamento nella natura delle discussioni durante la terapia, passando da questioni collettive a questioni più personali. Questi cambiamenti nei pronomi possono fornire preziose informazioni sulle dinamiche interne della paziente e sulla sua interazione con il mondo esterno. Monitorare e discutere questi cambiamenti durante la terapia può aiutare a capire meglio i progressi e le aree che potrebbero richiedere ulteriore attenzione. Aiutare la paziente a trovare un equilibrio tra la sua identità individuale e le sue relazioni sociali può essere un obiettivo importante. Potrebbe essere utile lavorare sulla costruzione di relazioni sane che rispettino sia i bisogni individuali che quelli collettivi. Monitorare come e perché il linguaggio della paziente cambia nel tempo può offrire importanti indicazioni sul suo progresso terapeutico. Discutere questi cambiamenti apertamente può favorire una maggiore consapevolezza e una comprensione più profonda delle proprie dinamiche interne. Un graduale aumento dei punteggi nella categoria "Affect" (da 3.17 a 3.97) indica che nel corso del tempo c'è stata una maggiore espressione delle emozioni generali. Questo potrebbe suggerire che la paziente sta diventando più aperta nell'esprimere le proprie emozioni, il che può essere un segnale positivo di fiducia nel contesto terapeutico. La categoria "Tone_pos" rappresenta l'espressione delle emozioni positive. I punteggi relativamente stabili, con un leggero aumento finale, suggeriscono che l'espressione delle emozioni positive è rimasta costante con un piccolo miglioramento verso la fine del periodo osservato. Questo può indicare un miglioramento graduale del tono emotivo generale della paziente. La categoria "Tone_neg" rappresenta l'espressione delle emozioni negative. I punteggi mostrano un aumento graduale fino al quarto intervallo, con una leggera diminuzione nell'ultimo intervallo. Questo potrebbe indicare che, nonostante un aumento iniziale nell'espressione delle emozioni negative, c'è stato un tentativo di gestione o riduzione verso la fine del periodo osservato.

I punteggi nella categoria "Emo_anx" mostrano un aumento costante fino al quarto intervallo, seguito da una riduzione. Questo suggerisce un aumento iniziale dei livelli di

ansia espressi, seguito da un tentativo di gestione o riduzione dell'ansia. I punteggi nella categoria "Emo_sad" mostrano una fluttuazione, con una riduzione fino al terzo intervallo, un aumento nel quarto e una successiva diminuzione. Questo suggerisce variazioni nell'espressione della tristezza, con un picco a metà del periodo osservato, seguito da una gestione o riduzione. I cambiamenti nei punteggi delle categorie emotive suggeriscono che la paziente ha mostrato una maggiore apertura nell'esprimere le proprie emozioni nel corso del tempo. In particolare: un aumento dell'espressione emotiva generale (Affect) indica una maggiore apertura emotiva. Questi cambiamenti possono riflettere progressi nella terapia, con la paziente che diventa più capace di esprimere e gestire le proprie emozioni.

La categoria "Drives" riflette le motivazioni e le forze che spingono le azioni e i comportamenti della paziente. Questi includono sottocategorie come affiliazione, potere, realizzazione e altri motivi psicologici. I punteggi della categoria "Drives" mostrano una diminuzione costante da 4.49 a 2.50 nel tempo. Questo potrebbe riflettere una fase di introspezione e ritiro emotivo, dove la paziente potrebbe essere meno motivata a perseguire obiettivi esterni e più concentrata sulla risoluzione di conflitti interni. Questa riduzione potrebbe essere dovuta a diversi fattori, come il confronto con difficoltà emotive, cambiamenti nelle priorità personali, o una fase di riflessione durante il processo terapeutico. Le variazioni nei punteggi "Certitude" suggeriscono che la sicurezza della paziente nelle proprie affermazioni può fluttuare. Questo potrebbe riflettere momenti di incertezza seguiti da periodi di maggiore sicurezza.

La stabilità nei punteggi "Space" indica che la paziente continua a utilizzare parole spaziali in modo coerente nel tempo. Questo suggerisce una consistenza nella sua capacità di descrivere contesti e situazioni in modo dettagliato e vivace.

L'aumento della categoria "Authentic" nel LIWC può offrire importanti indicazioni sulla sincerità, la trasparenza e l'autenticità della paziente nel suo linguaggio. Questa categoria riflette quanto le persone siano aperte e genuine nelle loro comunicazioni. Un aumento dei punteggi "Authentic" può suggerire vari cambiamenti positivi nella terapia. Un aumento dei punteggi "Authentic" suggerisce che la paziente sta diventando più aperta e sincera nelle sue comunicazioni. Questo può riflettere una crescente fiducia nel terapeuta e un comfort crescente nel condividere i suoi pensieri e sentimenti più profondi. L'aumento dell'autenticità può indicare una riduzione dei meccanismi di difesa. La

paziente potrebbe sentirsi meno necessitata a nascondere o minimizzare i suoi veri sentimenti, mostrando una maggiore vulnerabilità. L'aumento dell'autenticità potrebbe riflettere anche una maggiore profondità e complessità nelle comunicazioni della paziente. L'autenticità nelle comunicazioni può riflettersi anche nelle relazioni interpersonali al di fuori della terapia. La paziente potrebbe iniziare a comunicare in modo più genuino e aperto con gli altri, migliorando la qualità delle sue relazioni.

L'uso dei verbi ausiliari (auxiliary verbs) è spesso correlato con l'espressione di emozioni poiché questi verbi sono fondamentali per costruire tempi verbali complessi che esprimono stati emotivi, desideri, possibilità e obblighi. I punteggi della categoria "Auxiliary Verbs" mostrano piccole fluttuazioni, rimanendo relativamente stabili intorno a 8.42-9.27. La stabilità nell'uso dei verbi ausiliari può indicare che la paziente mantiene un modo costante di esprimere i suoi stati emotivi. I verbi ausiliari sono utilizzati per costruire frasi che descrivono stati interiori, desideri e obblighi (ad esempio, "sono triste", "vorrei", "devo").

L'analisi dei tempi verbali, specificamente le categorie "Focus Present", "Focus Past" e "Focus Future" nel LIWC, può fornire informazioni importanti su come la paziente si concentra temporalmente nelle sue narrazioni. Ogni categoria riflette un diverso orientamento temporale che può avere implicazioni significative per la terapia. I punteggi mostrano un aumento da 5.21 a circa 6.05, suggerendo una crescente focalizzazione sul presente. I punteggi nella categoria "Focus Future" mostrano un aumento significativo da 1.11 a 1.83. Un aumento nell'orientamento al futuro può indicare che la paziente sta diventando più ottimista e pianificatrice, pensando ai suoi obiettivi e aspettative future.

L'analisi dei punteggi LIWC e le relative categorie può fornirci indicazioni sui tratti di personalità secondo il modello Big Five (Cinque Grandi Fattori).

Apertura all'Esperienza: L'uso di parole associate a cognizione (Cogproc) e percezione (Perception) può suggerire un'alta apertura all'esperienza. L'uso costante di parole spaziali (Space) e dettagliate può indicare una descrizione vivida e ricca del mondo, tipica di chi è aperto a nuove esperienze. L'uso costante e dettagliato di parole nella categoria "Space" (4.28 - 4.66) e il mantenimento di punteggi stabili in "Cogproc" indicano una buona capacità di riflessione e una mente aperta.

Coscienziosità: L'uso di verbi ausiliari (Auxverb) e articoli (Article) può essere collegato alla struttura nel linguaggio, indicando coscienziosità. L'uso delle parole di realizzazione (Achieve) e pianificazione (Focus Future) può anche essere indicativo. La stabilità nei punteggi dei verbi ausiliari (8.42 - 9.27) e l'aumento nel focus sul futuro (1.11 - 1.83) suggeriscono un certo livello di organizzazione e pianificazione, sebbene la riduzione nei punteggi "Drives" (4.49 - 2.50) potrebbe indicare una flessione temporanea nella motivazione.

Estroversione: La riduzione nei punteggi "We" (6.80 - 0.70) e "Social" (11.20 - 9.51) può suggerire una diminuzione temporanea dell'interazione sociale, che potrebbe essere un segno di introversione. Tuttavia, l'aumento nei punteggi "You" (1.30 - 1.92) potrebbe indicare una maggiore interazione diretta. Ma comunque il tratto prevalente è l'introversione.

Gradevolezza: L'uso di parole di affiliazione (Affiliation), comportamenti prosociali (Prosocial) e un tono positivo (Tone) potrebbero suggerire amicalità. L'aumento nell'uso di parole positive suggerisce una disposizione amichevole e positiva. La diminuzione nell'uso del pronome "We" potrebbe indicare una temporanea distanza dalle relazioni di gruppo, ma non necessariamente una riduzione nel tratto gradevolezza.

Stabilità Emotiva: L'aumento delle emozioni negative potrebbe indicare una presenza instabilità emotiva, ma anche una maggiore consapevolezza e espressione delle emozioni negative.

In conclusione, sarebbe necessario discutere con la paziente i cambiamenti nei suoi tratti di personalità e comprendere come si sente a riguardo. Questo può offrire spunti importanti per il lavoro terapeutico. Continuare a lavorare sulla gestione delle emozioni negative e sviluppare strategie di coping per migliorare la stabilità emotiva. Incoraggiare la paziente a mantenere o aumentare le interazioni sociali, lavorando su strategie per migliorare la connessione con gli altri, e infine promuovere ulteriormente l'apertura all'esperienza e la pianificazione del futuro, utilizzando tecniche terapeutiche che facilitino la consapevolezza e l'espressione autentica. Tuttavia, potremmo ipotizzare che La paziente ha mostrato progressi in termini di stabilità emotiva, autenticità e apertura verso gli altri. I tratti di introversione sono rimasti predominanti, mentre si osserva una tendenza alla maggiore riflessione e consapevolezza verso la fine della terapia. Questo

può coincidere con i progressivi cambiamenti dei modi di essere emotivamente situati della paziente, associati alla conseguente significativa diminuzione della sintomatologia (Liccione, 2019). Infatti, l'esperienza relazionale che si instaura tra paziente e terapeuta ha lo specifico obiettivo di riposizionare il paziente rispetto ai suoi modi di essere disfunzionali e/o francamente psicopatologici (Liccione, 2019).

4.6 Conclusioni: Limiti e prospettive future

L'analisi della terapia attraverso l'uso del Linguistic Inquiry and Word Count (LIWC) ha offerto preziose intuizioni sui cambiamenti linguistici e psicologici della paziente. Tuttavia, questo approccio presenta alcuni limiti significativi che meritano una riflessione approfondita. LIWC utilizza un dizionario predefinito che non copre tutte le parole e le espressioni possibili. Parole nuove, gerghi o linguaggi specifici di nicchia potrebbero non essere riconosciuti o categorizzati correttamente. Il dizionario di LIWC potrebbe non essere aggiornato frequentemente, limitando la sua capacità di analizzare linguaggi in evoluzione. Inoltre, LIWC si concentra sulle singole parole piuttosto che sull'intero contesto in cui queste parole appaiono. Questo può portare a interpretazioni errate del significato di frasi complesse o ambigue. Associa le parole a categorie emozionali predefinite, ma non tiene conto delle sfumature emozionali che possono essere presenti nel testo. Ad esempio, l'ironia e il sarcasmo possono essere difficili da rilevare. Il dizionario di LIWC potrebbe riflettere bias culturali e linguistici, essendo sviluppato principalmente per l'inglese e basato su un contesto culturale specifico e sebbene esistano versioni per altre lingue, le traduzioni potrebbero non catturare tutte le sfumature culturali e linguistiche. Inoltre, gli algoritmi di LIWC si basano su regole rigide e semplici per categorizzare le parole senza usare modelli più complessi di elaborazione del linguaggio naturale (NLP) che potrebbero fornire una comprensione più profonda e contestuale del testo. Infine, è importante sottolineare che questo software si basa esclusivamente sul linguaggio verbale trascurando il linguaggio non verbale che è fondamentale in un contesto terapeutico. Espressioni facciali, gesti, toni di voce e altre forme di comunicazione non verbale forniscono indicazioni essenziali sullo stato emotivo e sui progressi della paziente, che non possono essere catturate attraverso questo strumento.

Un altro limite riguarda l'interpretazione dei risultati. È cruciale discutere i punteggi LIWC direttamente con la paziente per esplorare il significato personale della sua esperienza. Questo processo di riflessione congiunta non è stato possibile nella presente

ricerca, limitando così la profondità dell'analisi. Inoltre, la presente analisi è stata condotta su un singolo caso. Questo approccio limita la generalizzabilità dei risultati. Per confermare e ampliare i risultati ottenuti, sarebbe opportuno replicare la ricerca su un campione più ampio di pazienti, includendo diverse tipologie di interventi terapeutici e contesti clinici. Guardando al futuro, si auspica di replicare questa ricerca con un numero maggiore di casi e di sperimentare l'integrazione del LIWC con altre tecniche psicoterapeutiche. L'obiettivo è di esplorare come l'uso combinato di diversi approcci possa arricchire la pratica terapeutica. Infine, un'altra interessante direzione di ricerca potrebbe essere l'impiego di strumenti di intelligenza artificiale come supporto alla pratica psicoterapeutica. In sintesi, mentre il LIWC rappresenta uno strumento potente e innovativo per l'analisi del linguaggio terapeutico, il suo utilizzo deve essere integrato con altre metodologie e considerato all'interno di un quadro più ampio di pratiche terapeutiche e di ricerca.

Bibliografia

- Allport, W. G. (1965). *Psicologia della personalità*. Tr. It. LAS, Roma 1977.
- Alpers, G. W., Winzelberg, A. J., Classen, C., Roberts, H., Dev, P., Koopman, C., et al. (2005). Evaluation of computerized text analysis in an Internet breast cancer support group. *Computers in Human Behavior*, 21, 361-376.
- Argamon, S., Dhawle, S., Koppel, M., & Pennebaker, J. W. (2005, June). Lexical predictors of personality type. In *Proceedings of the 2005 joint annual meeting of the interface and the classification society of North America* (pp. 1-16).
- Argamon, Shlomo, et al. "Lexical predictors of personality type." (2005).
- Back, Mitja D., Stefan C. Schmukle, and Boris Egloff. "How extraverted is honey.bunny77@hotmail.de? Inferring personality from e-mail addresses." *Journal of Research in Personality*, 42.4 (2008): 1116-1122.
- Beaudreau, S. A., Storandt, M., & Strube, M. J. (2006). A comparison of narratives told by younger and older adults. *Experimental Aging Research*, 32, 105-117.
- Beck, A.T., Steer, R.A., & Brown, G.K. (1996). *Manual for the Beck Depression Inventory-II*. San Antonio, TX: Psychological Corporation.
- Bernard, H. R. (2009). *Analyzing Qualitative Data: Systematic Approaches*. Sage Publications.
- Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin*, 117(2), 187.
- Boals, A., & Klein, K. (2005). Word use in emotional narratives about failed romantic relationships and subsequent mental health. *Journal of Language and Social Psychology*, 24, 252-268.
- Bono, J. E., & Judge, T. A. (2004). Personality and transformational and transactional leadership: a meta-analysis. *Journal of Applied Psychology*, 89(5), 901.

- Bowen, N. K., & Guo, S. (2011). *Structural Equation Modeling*. Oxford University Press.
- Boxman-Shabtai, L., & Shifman, L. (2014). Evasive targets: Deciphering polysemy in mediated humor. *Journal of Communication*, 64(5), 977-998.
- Boyd, R. L., & Pennebaker, J. W. (2015). Did Shakespeare write *Double Falsehood*? Identifying individuals by creating psychological signatures with text analysis. *Psychological Science*, 26(5), 570-582.
- Boyd, R. L., & Pennebaker, J. W. (2017). Language-based personality: A new approach to personality in a digital world. *Current Opinion in Behavioral Sciences*, 18, 63-68.
- Braakmann, D. (2015). Historical paths in psychotherapy research. In O. C. G. Gelo, A. Pritz, & B. Rieken (Eds.), *Psychotherapy Research: Foundations, Process, and Outcome* (pp. 39-65). Vienna: Springer.
- Breck, E., Choi, Y., & Cardie, C. (2007, January). Identifying expressions of opinion in context. In *IJCAI* (Vol. 7, pp. 2683-2688).
- Brown, B. L., Strong, W. J., & Rencher, A. C. (1975). Acoustic determinants of perceptions of personality from speech.
- Cain, S. (2013). *Quiet: The Power of Introverts in a World That Can't Stop Talking*. Crown.
- Caprara, G. V., & Gennaro, A. (1999). *Psicologia della Personalità*. Seconda Edizione. *Personality Psychology*.
- Caspi, A., Harrington, H., Milne, B., Amell, J. W., Theodore, R. F., & Moffitt, T. E. (2003). Children's behavioral styles at age 3 are linked to their adult personality traits at age 26. *Journal of Personality*, 71(4), 495-514.
- Celli, Fabio, et al. "Workshop on computational personality recognition (shared task)." *Proceedings of the Workshop on Computational Personality Recognition*. 2013.
- Celli, Fabio, Pietro Zani Massani, and Bruno Lepri. "Profilio: Psychometric Profiling to Boost Social Media Advertising." *Proceedings of the 2017 ACM on Multimedia Conference*. ACM, 2017.

- Chessick, R. D. (1990). Hermeneutics for Psychotherapists. *American Journal of Psychotherapy*, 44(2), 256–273.
- Cloninger, C. R., & Svrakic, D. M. (1994). Differentiating normal and deviant personality by the seven-factor personality model.
- Cloninger, C. R., Svrakic, D. M., & Przybeck, T. R. (1993). A psychobiological model of temperament and character. *Archives of General Psychiatry*, 50(12), 975-990.
- Coker, A. L., Davis, K. E., Arias, I., Desai, S., Sanderson, M., Brandt, H. M., & Smith, P. H. (2002). Physical and mental health effects of intimate partner violence for men and women. *American Journal of Preventive Medicine*, 23(4), 260-268.
- Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A*, 33(4), 497-505.
- Corbetta, P. (2003). *La Ricerca Sociale: Metodologia e Tecniche. I. I Paradigmi di Riferimento*. Bologna: Il Mulino.
- Costa Jr, P. T., & McCrae, R. R. (1992). Four ways five factors are basic. *Personality and Individual Differences*, 13(6), 653-665.
- Costa, P. T., & McCrae, R. R. (1992). *Neo Personality Inventory-Revised (NEO PI-R)*. Odessa, FL: Psychological Assessment Resources.
- Costa, V. (2015). *Heidegger*. Editrice La Scuola, Milano.
- Costa, V. (2016). *Husserl*. Carocci Editore, Roma.
- Crits-Christoph, P., Gibbons, M. B. C., & Mukherjee, D. (2013). Psychotherapy process–outcome research. In M. J. Lambert (Ed.), *Bergin and Garfield’s Handbook of Psychotherapy and Behavior Change* (6th ed., pp. 298–340).
- D’Alessandris, F. (2019). The Person a ‘Material’ Story: The Construction of Ipseity and the Writing and Reading of the Shared Space. *Critical Hermeneutics*, 3(2), 75-102.

Dazzi, N., Lingiardi, V., & Gazzillo, F. (2009). La diagnosi in psicologia clinica. Personalità e psicopatologia. In *La diagnosi in psicologia clinica. Personalità e psicopatologia*. Raffaello Cortina Editore.

De la Rie, S. M., Duijsens, I. J., & Cloninger, C. R. (1998). Temperament, character, and personality disorders. *Journal of Personality Disorders*, 12(4), 362-372. degli Studi sul Risultato e sul Processo. *Rivista Sperimentale di Freniatria*, CXX, 2, 182-238.

Derogatis, L. R. (1977). *SCL-90-R: Administration, scoring, and procedures manual-I for the R(evised) version*. Baltimore: The Johns Hopkins University Press.

DeRubeis, R. J., Brotman, M. A., & Gibbons, C. J. (2005). A conceptual and methodological analysis of the nonspecifics argument. *Clinical Psychology: Science and Practice*, 12, 174–183. <http://dx.doi.org/10.1093/clipsy.bpi022>

DeRubeis, R. J., Brotman, M. A., & Gibbons, C. J. (2005). A conceptual and methodological analysis of the nonspecifics argument. *Clinical Psychology: Science and Practice*, 12, 174–183.

Dewaele, J. M., & Furnham, A. (1999). Extraversion: The unloved variable in applied linguistic research. *Language Learning*, 49(3), 509-544.

Dewaele, J. M., & Furnham, A. (2000). Personality and speech production: A pilot study of second language learners. *Personality and Individual Differences*, 28(2), 355-365.

Elliott, R. (1989). Comprehensive process analysis: understanding the change process in significant therapy events. In M.J. Packer & R. B. Addison (Eds.), *Entering the Circle: Hermeneutic Investigation in Psychology* (pp.165–184). Albany: State University of New York Press.

Elliott, R., Slatick, E., & Urman, M. (2001). *Qualitative change process research on psychotherapy: Alternative strategies*

Ernandes, M. (2010). *Un ponte tra apprendimento biologico e artificiale. La “Regola di Hebb” e le evidenze sperimentali nella Long-Term Potentiation* (Doctoral dissertation, Tesi per il dottorato in Neuroscienze. Università degli studi di Siena).

Evans, C., Mellor-Clark, J., Margison, F., Barkham, M., McGrath, G., Connell, J., e Audin, K. (2000). Clinical Outcomes in Routine Evaluation: The CORE-OM. *Journal of Mental Health, 9*, 247-255.

Eysenck, H. J. (1952). The effects of psychotherapy: An evaluation. (*Psychology Revivals*) (pp. 319-324).

Eysenck, H. J. (1952). The effects of psychotherapy: an evaluation. *Journal of consulting psychology, 16*(5), 319.

Eysenck, H. J. (1987). The definition of personality disorders and the criteria appropriate for their description. *Journal of Personality Disorders, 1*(3), 211-219.

Farnadi, G., Sitaraman, G., Sushmita, S., Celli, F., Kosinski, M., Stillwell, D., ... & De Cock, M. (2016). Computational personality recognition in social media. *User modeling and user-adapted interaction, 26*, 109-142.

Farnadi, G., Zoghbi, S., Moens, M. F., & De Cock, M. (2013). Recognising personality traits using facebook status updates. In *Proceedings of the international AAAI conference on web and social media* (Vol. 7, No. 2, pp. 14-18).

Fasanella, A. (2004). *Disegni sperimentali e quasi-sperimentali per la ricerca; tr. it. di DT Campbell, JC Stanley, Experimental and Quasi-experimental Designs for Research* (pp. 1-193). Eucos.

Faur, Caroline, et al. "PERSEED: a self-based model of personality for virtual agents inspired by socio-cognitive theories." *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on. IEEE, 2013.*

Fernandes, B., González-Briones, A., Novais, P., Calafate, M., Analide, C., & Neves, J. (2020). An adjective selection personality assessment method using gradient boosting machine learning. *Processes, 8*(5), 618.

Frank, J. D., & Frank, J. B. (1993). *Persuasion and healing: A comparative study of psychotherapy*. JHU Press.

Freud, S. (1892–95). Studi sull'isteria. OSF, vol. I.

Furnham, A., & Mitchell, J. (1991). Personality, needs, social skills and academic achievement: A longitudinal study. *Personality and Individual Differences*, 12(10), 1067-1073.

Furnham, A., Jackson, C. J., & Miller, T. (1999). Personality, learning style and work performance. *Personality and individual differences*, 27(6), 1113-1122.

Furnham, A., Petrides, K. V., Jackson, C. J., & Cotter, T. (2002). Do personality factors predict job satisfaction?. *Personality and individual differences*, 33(8), 1325-1342.

Galimberti U (2016). *Psichiatria e fenomenologia*. Feltrinelli Editore, Milano.

Gao, R., Hao, B., Bai, S., Li, L., Li, A., & Zhu, T. (2013, October). Improving user profile with personality traits predicted from social media content. In *Proceedings of the 7th ACM conference on recommender systems* (pp. 355-358).

Gelman, A. (2006). Multilevel (hierarchical) modeling: what it can and cannot do. *Technometrics*, 48(3), 432-435.

Gelo O., Braakmann, D., e Benetka, G. (2008). Quantitative and Qualitative Research: Beyond the Debate. *Integrative Psychological and Behavioural Science*, 42, 3, 266-90.

Gelo O., Pritz A. & Rieken B. (2015). *Psychotherapy Research: Foundations, process and outcome*.

Gelo, O. C. G (2012). On research methods and their philosophical assumptions: “Raising the consciousness of researchers” again. *Psychoterapie & Sozialwissenschaft*, 14(2), 111-130.

Gelo, O. C. G., & Manzo, S. (2014). Quantitative approaches to treatment process, change process, and process-outcome research. In *Psychotherapy research: Foundations, process, and outcome* (pp. 247-277). Vienna: Springer Vienna.

Gelo, O. C. G., Lagetto, G., Dinoi, C., Belfiore, E., Lombi, E., Blasi, S., ... & Ciavolino, E. (2020). Which methodological practice (s) for psychotherapy science? A systematic review and a proposal. *Integrative Psychological and Behavioral Science*, 54, 215-248.

Gelo, O. C. G., Salcuni, S., & Colli, A. (2012). Text analysis within quantitative and qualitative psychotherapy process research: introduction to special issue. *Research in Psychotherapy: Psychopathology, Process and Outcome*, 15(2), 45-53.

Gill, A. J., & Oberlander, J. (2003). Perception of e-mail personality at zero-acquaintance: Extraversion takes care of itself; neuroticism is a worry. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 25, No. 25).

Gill, A. J., Oberlander, J., & Austin, E. (2006). Rating e-mail personality at zero acquaintance. *Personality and Individual Differences*, 40(3), 497-507.

Gnisci, A., & Pedon, A. (2021). *Manuale di psicodiagnostica: aspetti teorici e applicativi dei test*.

Golbeck, J. A. (2016). Predicting personality from social media text. *AIS Transactions on Replication Research*, 2(1), 2.

Golbeck, J., Robles, C., Edmondson, M., & Turner, K. (2011, October). Predicting personality from twitter. In *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing* (pp. 149-156). IEEE.

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American psychologist*, 48(1), 26.

Graciarena, M., Shriberg, E., Stolcke, A., Enos, F., Hirschberg, J., & Kajarekar, S. (2006, May). Combining prosodic lexical and cepstral systems for deceptive speech detection. In *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings* (Vol. 1, pp. I-I). IEEE. *personality and Social Psychology Bulletin*, 29, 665-675

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-metrix: analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, 36, 193-202.

Grainger, J., Rey, A., & Dufau, S. (2008). Letter perception: From pixels to pandemonium. *Trends in cognitive sciences*, 12(10), 381-387.

Greenberg, L.S. (1986). Change process research. *Journal of Consulting and Clinical Psychology*, 54, 1, 4-9.

Greenberg, L.S., & Pinsof, W. M. (1986). *Process Research: Current Trends and Future Perspectives*. New York: Guilford Press.

Greenberg, L.S., & Pinsof, W. M. (1986). *The Psychotherapeutic Process: A Research Handbook*. New York: Guilford Press.

Grucza, R. A., & Goldberg, L. R. (2007). The comparative validity of 11 modern personality inventories: Predictions of behavioral acts, informant reports, and clinical indicators. *Journal of personality assessment*, 89(2), 167-187.

Gunsch, M. A., Brownlow, S., Haynes, S. E., & Mabe, Z. (2000). Differential linguistic content of various forms of political advertising. *Journal of Broadcasting & Electronic Media*, 44, 27-42.

H. Andrew Schwartz et al. «Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach». In: *PLOS ONE* 8 (set. 2013), pp. 1–16. doi: 10.1371/journal.pone.0073791. url: <https://doi.org/10.1371/journal.pone.0073791>

Han, S., Huang, H., & Tang, Y. (2020). Knowledge of words: An interpretable approach for personality recognition from social media. *Knowledge-Based Systems, 194*, 105550.

Hawkins II, R. C., & Boyd, R. L. (2017). Such stuff as dreams are made on: Dream language, LIWC norms, and personality correlates. *Dreaming, 27*(2), 102.

Heidegger M. (1927), *Essere e Tempo*, Trad. It. (2015), Arnoldo Mondadori Editore, Milano.

Heidegger M. (1975), *I problemi fondamentali della fenomenologia*, Trad. It (1999) Il Melangolo, Genova.

Heinrich, C. U., & Borkenau, P. (1998). Deception and deception detection: The role of cross-modal inconsistency. *Journal of Personality, 66*(5), 687-712.

Heylighen, F., & Dewaele, J. M. (2002). Variation in the contextuality of language: An empirical measure. *Foundations of science, 7*, 293-340.

Hill, C. E. (1994) Special section: qualitative research in counseling process and outcome. *JCouns Psychol 41*(4).

Hill, C. E., Knox, S., Thompson, B. J., Williams, E. N., Hess, S. A., & Ladany, N. (2005). Consensual qualitative research: An update. *Journal of Counseling Psychology, 52*, 196–205.

Hirschberg, J. B., & Rosenberg, A. (2005). Acoustic/prosodic and lexical correlates of charismatic speech.

Hirschberg, J. B., Enos, F., Benus, S., Cautin, R. L., Graciarena, M., & Shriberg, E. (2006). Personality factors in human deception detection: Comparing human to machine performance.

Hofmann, S. G., & Hayes, S. C. (2019). The future of intervention science: Process-based therapy. *Clinical Psychological Science*, 7(1), 37-50. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6350520/>.

Hogan, R., Curphy, G. J., & Hogan, J. (1994). What we know about leadership: Effectiveness and personality. *American psychologist*, 49(6), 493.

Holmes, D., Alpers, G. W., Ismailji, T., Classen, C., Wales, T., Cheasty, V., et al. (2007). Cognitive and emotional processing in narratives of women abused by intimate partners. *Violence Against Women*, 13, 1192-1205

Horn, R. L., & Weisz, J. R. (2020). *Can Artificial Intelligence Improve Psychotherapy Research and Practice?* *Administration and Policy in Mental Health*, 47(5), 852–855 [<https://doi.org/10.1007/s10488-020-01056-9>].

Horowitz, L.M., Rosenberg, S.E., Baer, B.A., Ureño, G., e Villaseñor, V.S. (1988). Inventory of interpersonal problems: psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology*, 56, 885-892.

Horvath, A. O., & Symonds, B. D. (1991). Relation between working alliance and outcome in psychotherapy: A meta-analysis. *Journal of counseling psychology*, 38(2), 139.

Howard, K. I., Lueger, R. J., Maling, M. S., & Martinovich, Z. (1993). A phase model of psychotherapy outcome: Causal mediation of change. *Journal of Consulting and Clinical Psychology*, 61, 678–685.

Husserl E. (1931). *Cartesianische Meditationen und Pariser Vorträge*. (Tr. it. *Meditazioni cartesiane*. Roma: Armando, 1997).

Isabel Briggs Myers and Peter B Myers. 2010. *Giftsdiffering: Understanding personality type*. Davies-Black Publishing.

Jang, K. L., McCrae, R. R., Angleitner, A., Riemann, R., & Livesley, W. J. (1998). Heritability of facet-level traits in a cross-cultural twin sample: support for a hierarchical model of personality. *Journal of personality and social psychology*, 74(6), 1556.

Jordan, K. N., Sterling, J., Pennebaker, J. W., & Boyd, R. L. (2019). Examining long-term trends in politics and culture through language of political leaders and cultural institutions. *Proceedings of the National Academy of Sciences*, 116(9), 3476-3481.

Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). Pronoun use reflects standings in social hierarchies. *Journal of Language and Social Psychology*, 33(2), 125-143.

Kahn, J. H., Tobin, R. M., Massey, A. E., & Anderson, J. A. (2007). Measuring emotional expression with the Linguistic Inquiry and Word Count. *American Journal of Psychology*, 120, 263-286.

Kiesler, D.J. (1973). *The process of psychotherapy: Empirical foundations and systems of analysis*. Chicago: Aldine

Komarraju, M., & Karau, S. J. (2005). The relationship between the big five personality traits and academic motivation. *Personality and individual differences*, 39(3), 557-567.

Kosinski, Michal, David Stillwell, and Thore Graepel. "Private traits and attributes are predictable from digital records of human behavior." *Proceedings of the National Academy of Sciences* (2013): 201218772.

Kowalski, R. M. (2000). "I was Only Kidding!" Victims' and perpetrators' perceptions of teasing. *Personality and Social Psychology Bulletin*, 26, 231-241.

Kraemer, H. C., Wilson, G. T., Fairburn, C. G., & Agras, W. S. (2002). Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry*, 59, 877 – 883.

Kwantes, P. J., Derbentseva, N., Lam, Q., Vartanian, O., & Marmurek, H. H. (2016). Assessing the Big Five personality traits with latent semantic analysis. *Personality and Individual Differences, 102*, 229-233.

Lambert MJ (2013) Bergin and Garfield's handbook of psychotherapy and behavior change, 6th edn. Wiley, New York

Lambert, M.J., Morton, J., Hatfield, D., Harmon, C., Hamilton, S., Reid, R., Shimokawa, K., Christopherson, C.D., e Burlingame, G. (2004). *Administration and scoring manual for the Outcome Questionnaire (OQ-45.2)*, 3rd Edition. Wilmington, DE: American Professional Credentialing Services LLC.

Larsen-Freeman, D., & Long, M. H. (2014). *An introduction to second language acquisition research*. Routledge.

Lestrade S. (2017). Unzipping Zipf's law. *PloS one, 12*(8), e0181987. <https://doi.org/10.1371/journal.pone.0181987>

Liccione D, Liccione D (2016). (A cura di). Il primo colloquio in psicoterapia. Libreria Universitaria Editore: Padova

Liccione D. (2011), *Psicoterapia Cognitivo Neuropsicologica*, Bollati Boringhieri, Torino.

Liccione D. (2013) Verso una neuro psicopatologia ermeneutica. *Rivista internazionale di filosofia e psicologia*, Vol.4, n.3, pp. 305-324
Liccione D. (2019). *Psicoterapia Cognitiva Neuropsicologica*. II Edizione ampliata e riveduta. Bollati Boringhieri, Torino.

Liscombe, J., Venditti, J., & Hirschberg, J. B. (2003). Classifying subject ratings of emotional speech using acoustic features.

Llewelyn, S. P. (1988). Psychological therapy as viewed by clients and therapists. *British Journal of Clinical Psychology, 27*, 223–237.

Llewelyn, S., Macdonald, J., & Aafjes-van Doorn, K. (2016). Process–outcome studies. In J. C. Norcross, G. R. VandenBos, D. K. Freedheim, & B. O. Olatunji (Eds.), *APA handbook of clinical psychology: Theory and research*, 451–463. American Psychological Association. <https://doi.org/10.1037/14773-020>

Luborsky, L. (1990). Therapeutic alliance measures as predictors of future benefits of psychotherapy. In *annual meeting of the Society for Psychotherapy Research, Wintergreen, VA*.

Lynam, D. R., & Widiger, T. A. (2001). Using the five-factor model to represent the DSM-IV personality disorders: an expert consensus approach. *Journal of abnormal psychology*, 110(3), 401.

Maarten Sap et al. «Developing Age and Gender Predictive Lexica over Social Media». In: (gen. 2014), pp. 1146–1151. doi: 10.3115/v1/D14-1121

Maass, A., Karasawa, M., Politi, F., & Suga, S. (2006). Do verbs and adjectives play different roles in different cultures? A cross-linguistic analysis of person representation. *Journal of Personality and Social Psychology*, 90, 734-750

Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of artificial intelligence research*, 30, 457-500.

Mairesse, François, et al. "Using linguistic cues for the automatic recognition of personality in conversation and text." *Journal of artificial intelligence research* 30 (2007): 457-500.

Majumder, N., Poria, S., Gelbukh, A., & Cambria, E. (2017). Deep learning-based document modeling for personality detection from text. *IEEE Intelligent Systems*, 32(2), 74-79.

Mandelbrot, B. (1953). An informational theory of the statistical structure of language. *Communication theory*, 84, 486-502.

Manzo, S. (2010). La ricerca di processo nel decennio 1998-2007: contributo a una recensione critica. *Ricerca in psicoterapia: psicopatologia, processo e risultato*, 13 (1), 92-119.

Margaret L. Kern et al. «The Online Social Self: An Open Vocabulary Approach to Personality». In: *Assessment* 21.2 (2014). PMID: 24322010, pp. 158–169. doi: 10.1177/1073191113514104. eprint: <https://doi.org/10.1177/1073191113514104>. url: <https://doi.org/10.1177/1073191113514104>

Marmar, C.R. (1990). Psychotherapy Process Research: Progress, Dilemmas, and Future Direction. *Journal Consulting and Clinical Psychology*, 58, 265–272.

Matz, S. C., et al. "Psychological targeting as an effective approach to digital mass persuasion." *Proceedings of the National Academy of Sciences* (2017): 201710966.

McCarthy, John, *A proposal for the Dartmouth summer research project on artificial intelligence*, 1955.

McCrae, R. R., & Costa Jr, P. T. (1997). Personality trait structure as a human universal. *American psychologist*, 52(5), 509.

McCrae, R. R., Costa Jr, P. T., Ostendorf, F., Angleitner, A., Hřebíčková, M., Avia, M. D., ... & Smith, P. B. (2000). Nature over nurture: temperament, personality, and life span development. *Journal of personality and social psychology*, 78(1), 173.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5, 115-133.

McLeod, B. D., Islam, N. Y., & Wheat, E. (2013). Designing, conducting, and evaluating therapy process research. In J. S. Comer & P. C. Kendall (Eds.), *The Oxford handbook of research strategies for clinical psychology* (pp. 142–164).

Mehl, M. R., & Gill, A. J. (2010). Automatic text analysis.

Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology*, 90, 862–877.

Mehl, M., Pennebaker, J., Crow, M., Dabbs, J., & Price, J. (2001). The Electronically Activated Recorder (EAR): A device for sampling naturalistic daily activities and conversations. *Behavior Research Methods, Instruments, and Computers*, 33, 517–523.

Mergenthaler E. & Stinson C. H. (1992). Psychotherapy transcription standards. *Psychotherapy Research*, 2(2), 125–142. <https://doi.org/10.1080/1050330921233133290>.

Migone P. (1996). La ricerca in psicoterapia: storia, principali gruppi di lavoro, stato attuale degli studi sul risultato e sul processo. *Rivista Sperimentale di Freniatria*, CXX, 2: 182-238.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26.

Milbrath C., Bauknight R., Horowitz M., Amaro R. & Sugahara C. (1995). *Sequential Analysis of Topics in Psychotherapy Discourse: Single-Case Study*. *Psychotherapy Research*, 5:3, 199-217, DOI: 10.1080/10503309512331331316.

Mishne, G., Carmel, D., & Lempel, R. (2005, May). Blocking Blog Spam with Language Model Disagreement. In *AIRWeb* (Vol. 5, pp. 1-6).

Mohammad, S. M., Zhu, X., Kiritchenko, S., & Martin, J. (2015). Sentiment, emotion, purpose, and style in electoral tweets. *Information Processing & Management*, 51(4), 480-499.

Monzani, D., Vergani, L., Pizzoli, S. F. M., Marton, G., & Pravettoni, G. (2021). Emotional tone, analytical thinking, and somatosensory processes of a sample of Italian tweets during the first phases of the COVID-19 pandemic: observational study. *Journal of Medical Internet Research*, 23(10), e29820.

Moreno, J. D., Martinez-Huertas, J. A., Olmos, R., Jorge-Botana, G., & Botella, J. (2021). Can personality traits be measured analyzing written language? A meta-analytic study on computational methods. *Personality and Individual Differences*, 177, 110818.

Murphy R, Cooper Z, Hollon SD, Fairburn CG. How do psychological treatments work? Investigating mediators of change. *Behav Res Ther*. 2009; 47:1-5.

Naugle, D., & Nussbaum, C. (1993). WILHELM DILTHEY'S DOCTRINE OF WORLD VIEWS AND ITS RELATIONSHIP TO HERMENEUTICS. *Philosophy*, 5329(501).

Newell, A., & Simon, H. (1956). The logic theory machine--A complex information processing system. *IRE Transactions on information theory*, 2(3), 61-79.

Newell, A., & Simon, H. A. (1961). GPS, a program that simulates human thought.

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles.

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, 29, 665-675.

Norcross, J. C., Krebs, P. M., & Prochaska, J. O. (2011). *Stages of change*. *Journal of Clinical Psychology*, 67(2), 143–154.

Oberlander, J., & Gill, A. (2004, April). Language generation and personality: Two dimensions, two stages, two hemispheres. In *Proceedings from the AAAI Spring Symposium on Architectures for Modeling Emotion: Cross-Disciplinary Foundations* (pp. 104-111).

Oberlander, J., & Gill, A. J. (2006). Language with character: A stratified corpus comparison of individual differences in e-mail communication. *Discourse Processes*, 42, 239-270

Oberlander, J., & Nowson, S. (2006, July). Whose thumb is it anyway? Classifying author personality from weblog text. In *Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions* (pp. 627-634).

Orlinsky DE, Rønnestad MH, Willutzki U (2004) Fifty years of psychotherapy process-outcome research: continuity and change. In: Lambert M (ed) *Bergin and Garfield's handbook of psychotherapy and behavior change*, 5th edn. Wiley, New York, pp 307–389

Orlinsky, D. E., & Russell, R. L. (1994). Tradition and change in psychotherapy research: Notes on the fourth generation.

Oudeyer, P. Y. (2002). Novel useful features and algorithms for the recognition of emotions in human speech. In *Speech Prosody 2002, International Conference*.

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., ... & Seligman, M. E. (2015). Automatic personality assessment through social media language. *Journal of personality and social psychology*, 108(6), 934.

Pasupathi, M. (2007). Telling and the remembered self: Linguistic differences in memories for previously disclosed and previously undisclosed events. *Memory*, 15, 258-270.

Paul T Costa and Robert R McCrae. 2008. The re-vi-sed neo personality inventory (neo-pi-r). In G.J.Boyle, G Matthews and D. Saklofske (Eds.). *TheSAGE handbook of personality theory and assessment*2:179–198

Paul, G. L. (1967). Strategy of outcome research in psychotherapy. *Journal of consulting psychology*, 31(2), 109.

Paunonen, S. V., & Jackson, D. N. (2000). What is beyond the big five? Plenty! *Journal of personality*, 68(5), 821-835.

Peirson, A. R., & Heuchert, J. W. (2001). The relationship between personality and mood: comparison of the BDI and the TCI. *Personality and Individual differences*, 30(3), 391-399.

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, 77, 1296–1312.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Inquiry and Word Count: LIWC2001*. Lawrence Erlbaum, Mahwah, NJ.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: our words, our selves. *Annual review of psychology*, 54, 547–577. <https://doi.org/10.1146/annurev.psych.54.101601.145041>

Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014). When small words foretell academic success: The case of college admissions essays. *PloS one*, 9(12), e115844.

Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532-1543).

Pervin, L. A. (1994). A critical analysis of current trait theory. *Psychological Inquiry*, 5(2), 103-113.

Picci, L., & Pacini, B. (2001). Capitolo 9-Test sulla varianza e l'analisi della varianza. *Capitolo 9-Test sulla varianza e l'analisi della varianza*, 1000-1021.

Piacentino, A., Lucchi, F., Scarsato, G., & Fazzari, G. (2017). Mental Health Recovery Star: features and validation study of the Italian version. *Rivista di psichiatria*, 52(6), 247-254.

Pritz, & B. Rieken (Eds.), *Psychotherapy research: Foundations, process, and outcome*, 39–65. Vienna: Springer. https://doi.org/10.1007/978-3-7091-1382-0_3.

Quercia, Daniele, et al. "The personality of popular Facebook users." *Proceedings of the ACM 2012 conference on computer supported cooperative work*. ACM, 2012.

Rayson, M. D., Jones, N. L., & Ivey, G. N. (2019). Observations of large-amplitude mode-2 nonlinear internal waves on the Australian Northwest shelf. *Journal of Physical Oceanography*, 49(1), 309-328.

Rainsberger, L. (2022). Explanation: What Is New and Different About AI? In L. Rainsberger (A c. Di), *AI - The new intelligence in sales: Tools, applications and potentials of Artificial Intelligence* (pp. 1–16).

Ramezani, M., Feizi-Derakhshi, M. R., & Balafar, M. A. (2022). Knowledge Graph-Enabled Text-Based Automatic Personality Prediction. *Computational intelligence and neuroscience*, 2022, 3732351. <https://doi.org/10.1155/2022/3732351>

Ren Z., Shen Q., Diao X., Xu H. A sentiment-aware deep learning approach for personality detection from text. *Information Processing & Management*. 2021;58(3) doi: 10.1016/j.ipm.2021.102532.102532

Rennie, D. L., Phillips, J. R., & Quartaro, G. K. (1988). Grounded theory: A promising approach to conceptualization in psychology? *Canadian Psychology*, 29, 139–150.

Rice, L.N., & Greenberg, L.S. (1984). *Patterns of Change: Intensive Analysis of Psychotherapy Process*. New York: Guilford

Ricoeur P. (1982). *Il conflitto delle interpretazioni*. Editoriale Jaca Book, p. 26. Prima pubblicazione: 1969.

Ricoeur P. (1986). *Dal testo all'azione. Saggi di ermeneutica*. Trad. it di G. Grampa, Jaca Book, Milano 2003.

Ricoeur, P. (1990). *Sé come un altro*. (Tr. It. a cura di D. Iannotta, Milano: Jaca Book, 2005).

Riggio, R. E., Salinas, C., & Tucker, J. (1988). Personality and deception ability. *Personality and Individual Differences*, 9(1), 189-191.

Rogers, C. (2000). Terapia centrata sulla persona. *Sei approcci chiave alla consulenza e alla terapia*, 1, 98-105.

Rosenzweig, S. (1936). Some implicit common factors in diverse methods of psychotherapy. *American journal of Orthopsychiatry*, 6(3), 412.

Ruppenhofer, J., Somasundaran, S., & Wiebe, J. (2008, May). Finding the sources and targets of subjective expressions. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)* (pp. 278)

Salerno, D. (2018). Introduzione. Big data e social networks. *E| C*, (23), 33-35.

Sap, M., Park, G., Eichstaedt, J., Kern, M., Stillwell, D., Kosinski, M., ... & Schwartz, H. A. (2014, October). Developing age and gender predictive lexica over social media. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1146-1151).

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., ... & Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS one*, 8(9), e73791.

Semin, G. R., & Fiedler, K. (1988). The cognitive functions of linguistic categories in describing persons: Social cognition and language. *Journal of Personality and Social Psychology*, 54, 558-568.

Shapiro, M.B., (1961). A method of measuring changes specific to the individual psychiatric patient. *British Journal of Medical Psychology*, 34, 151–155.

SIGNOR, E. Intelligenza Artificiale e Psicografia: come personalità e genere influiscono sulla scelta occupazionale. (2023) <https://hdl.handle.net/20.500.12608/50233>

Skowron, M., Tkalčič, M., Ferwerda, B., & Schedl, M. (2016, April). Fusing social media cues: personality prediction from twitter and Instagram. In *Proceedings of the 25th international conference companion on world wide web* (pp. 107-108).

Smink, W., Sools, A. M., van der Zwaan, J. M., Wiegersma, S., Veldkamp, B. P., & Westerhof, G. J. (2019). Towards text mining therapeutic change: A systematic review of text-based methods for Therapeutic Change Process Research. *PloS one*, 14(12), e0225703.

Smith, B. L., Brown, B. L., Strong, W. J., & Rencher, A. C. (1975). Effects of speech rate on personality perception. *Language and Speech*, 18, 145–152.

Smith, J. A., Flowers, P., & Larkin, M. (2009). *Interpretative phenomenological analysis: Theory, method and research*. London, England: Sage.

Somalvico, M. (1987). *Intelligenza artificiale*. Scienza & vita nuova.

Spinosi, N. (2005). Capitolo II. La ricerca in psicologia sociale. *Capitolo II. La ricerca in psicologia sociale*, 1000-1015.

Stiles, W. B. (2015). Theory building, enriching, and fact gathering: Alternative purposes of psychotherapy research. In O. C. G. Gelo, A. Pritz, & B. Rieken (Eds.), *Psychotherapy research: Foundations, process, and outcome* (pp. 159–179). Vienna: Springer

Stiles, W. B., & Shapiro, D. A. (1994). Disabuse of the drug metaphor: Psychotherapy process–outcome correlations. *Journal of Consulting and Clinical Psychology*, 62, 942–948. <http://dx.doi.org/10.1037/0022-006X.62.5.942>

Stiles, W.B., & Shapiro, D.A., & Elliott, R. (1986). Are all psychotherapies equivalent? *American Psychologist*, 41, 165–180.

Stinson C. H., Milbrath C., Reidbord S. P. & Bucci W., (1994). Thematic segmentation of psychotherapy transcripts for convergent analyses. *Psychotherapy: Theory, Research, Practice, Training*, 31, 36-48. doi: 10.1037/0033-3204.31.1.36

Stricker, G., & Gold, J.R. (1993) *Comprehensive handbook of psychotherapy integration*. New York: Plenum.

Suppe, F. (1977). *The structure of scientific theories* (2nd ed.). Chicago, IL. University of Illinois Press.

Tarozzi, M. (2008). *Cos'è la Grounded Theory*. Carocci, Roma.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology*, 29(1), 24-54.

Tetlock, P. E. (1981). Pre- to post-election shifts in presidential rhetoric: Impression management or cognitive adjustment. *Journal of Personality and Social Psychology*, 41, 207-212

Thilakaratne, M., Weerasinghe, R., & Perera, S. (2016, October). Knowledge-driven approach to predict personality traits by leveraging social media data. In *2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI)* (pp. 288-295). IEEE.

Thorne, B., & Sanders, P. (2012). *Carl rogers*. Sage.

Tompkins, K. A., & Swift, J. K. (2014). Psychotherapy process and outcome research. *The enciclopedia of clinical psychology*, 1-7.

Traini S., (2013). *Le basi della semeiotica*. Bompiani, Milano.

Troop, N. A., Chilcot, J., Hutchings, L., & Varnaite, G. (2013). Expressive writing, self-criticism, and self-reassurance. *Psychology and psychotherapy*, 86(4), 374–386. <https://doi.org/10.1111/j.2044-8341.2012.02065.x>

Turing, A.M. (1950), Computing machinery and intelligence, in «Mind», 59, pp. 433-460, trad. it. Macchine calcolatrici e intelligenza, in V. Somenzi e R. Cordeschi (a cura di), La filosofia degli automi. Origini dell'Intelligenza Artificiale, Bollati Boringhieri, Torino, 1994, pp. 166- 193.

Turing, A.M. (1951), Can digital computers think, in Copeland B.J. (2004), trad. it. I calcolatori digitali possono pensare? in Sistemi Intelligenti, 10, 1998, pp. 21-26.

Verbree, D., Rienks, R., & Heylen, D. (2006, December). Dialogue-act tagging using smart feature selection; results on multiple corpora. In *2006 IEEE Spoken Language Technology Workshop* (pp. 70-73). IEEE.

Vinciarelli, Alessandro, and Gelareh Mohammadi. "A survey of personality computing." *IEEE Transactions on Affective Computing* 5.3 (2014): 273-291.

Vogel, K., & Vogel, S. (1986). L'interlangue et la personnalité de l'apprenant. *International Journal of Applied Linguistics*, 24(1), 48–68

Watson, D., Clark, L. A., McIntyre, C. W., & Hamaker, S. (1992). Affect, personality, and social activity. *Journal of personality and social psychology*, 63(6), 1011.

Westen, D. (1996). A model and a method for uncovering the nomothetic from the idiographic: An alternative to the five-factor model?. *Journal of Research in Personality*, 30(3), 400-413.

Westen, D. (2006). Commentary on Trull: Drizzling on the 5±3 factor parade.

Westen, D., & Shedler, J. (2000). A prototype matching approach to diagnosing personality disorders: Toward DSM-V. *Journal of Personality Disorders*, 14(2), 109-126.

Widiger, T. A., Trull, T. J., Clarkin, J. F., Sanderson, C., & Costa Jr, P. T. (1994). A description of the DSM-III-R and DSM-IV personality disorders with the five-factor model of personality.

Widiger, T. A., Trull, T. J., Clarkin, J. F., Sanderson, C., & Costa Jr, P. T. (2002). A description of the DSM-IV personality disorders with the five-factor model of personality.

Yamagata, S., Suzuki, A., Ando, J., Ono, Y., Kijima, N., Yoshimura, K., ... & Jang, K. L. (2006). Is the genetic structure of human personality universal? A cross-cultural twin study from North America, Europe, and Asia. *Journal of personality and social psychology*, 90(6), 987.

Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of research in personality*, 44(3), 363-373.

Yla R. Tausczik e James W. Pennebaker. «The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods». In: *Journal of Language and Social Psychology* 29.1 (2010), pp. 24–54. doi: 10.1177/0261927X09351676.

Zipf, G. K. (2016). *Human behavior and the principle of least effort: An introduction to human ecology*. Ravenio Books.

Zuckerman, M. (2002). Zuckerman-Kuhlman Personality Questionnaire (ZKPQ): an alternative five-factorial model.

Zuckerman, N. (2015). Heidegger and the Essence of Dasein. *The Southern Journal of Philosophy*, 53(4), 493-516.