



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Università degli Studi di Padova

Dipartimento di Studi Linguistici e Letterari

Corso di Laurea Magistrale in Linguistica
Classe LM-39

Tesi di Laurea

Testi scritti con la Comunicazione Facilitata: autorialità dei facilitati e influenza di facilitatori.

Relatore

Prof. Michele Cortelazzo

Correlatore

Prof.ssa Arjuna Tuzzi

Laureando

Giovanni Nicoli

n° matr.1189763 / LMLIN

Anno Accademico 2018/ 2019

“Non berrò alla fontana del cavallino”

CAPITOLO 1

1.1 I metodi quantitativi di analisi del testo	7
1.1.1 Sistemi quantitativi	9
1.1.1.1 Tratti utilizzati dai sistemi quantitativi	12
1.1.1.1.1 Tratti lessicali	13
1.1.1.1.2 Tratti grafici/ n-grammi di caratteri	16
1.1.1.1.3 Tratti sintattici	18
1.1.1.1.4 Tratti semantici	19
1.1.1.1.5 Tratti specifici per applicazione	20
1.2 Metodi per l'attribuzione d'autore	21
1.2.1 Metodi per l'attribuzione d'autore: classificazioni multi-classe	21
1.2.1.1 Approcci <i>profile-based</i>	22
1.2.1.2 Approcci <i>instance-based</i>	24
1.2.1.2.1 Modelli basati su vettori spaziali	25
1.2.1.2.1.1 <i>Supported Vector Machine (SVM)</i>	26
1.2.1.2.1.2 Alberi decisionali e <i>random forest</i>	26
1.2.1.2.2 Modelli basati sulla similarità	27
1.2.1.3 Approcci ibridi	27
1.2.2 Metodi di attribuzione d'autore: <i>hierarchical clustering methods</i> .	28

CAPITOLO 2

2.1 La comunicazione facilitata: introduzione	31
2.2 La comunicazione facilitata: lo stato dell'arte	32
2.2.1 Gli studi di message passing	32
2.2.2 Gli studi con il tracciamento oculare	34
2.2.3 Gli studi di analisi linguistica	35
2.2.3.1 La risposta dei detrattori	38
2.3 Il dibattito sulla comunicazione facilitata	38

CAPITOLO 3

3.1 L'analisi dei testi scritti con la CF	42
3.2 IL progetto EASIEST	42
3.3 Indicatori di autorialità e di influenzamento	43

CAPITOLO 4	
4.1 Un'analisi semi-qualitativa	54
4.1.1 Materiali	54
4.1.1.1 Il <i>corpus 2018</i>	54
4.1.1.2 I partecipanti	55
4.1.2 Il corpus: uno sguardo d'insieme	56
4.1.3 Corrispondenze all'interno del corpus	64
4.1.3.1 La lunghezza media dell'enunciato (LME)	64
4.1.3.2 Indicatori lessicali e sintagmatici di influenza e autorialità	66
4.1.3.2.1 Metodologie di indagine	66
4.1.3.2.2 Tipicità legate al facilitatore: indicatori di influenza?	67
4.1.3.2.3 Tipicità legate al facilitato: indicatori di autorialità?	72
CAPITOLO 5	
5.1 Un'analisi quantitativa	92
5.1.1 Materiali e metodi	92
5.1.1.1 <i>Hierarchical clustering analysis</i>	92
5.1.1.2 <i>Supported vector machine e Random Forest</i>	92
5.1.1.3 Il corpus	93
5.1.2 Risultati	94
5.2 Analisi conversazione WhatsApp	103
5.2.1 Il <i>corpus WA</i>	103
5.2.2 Metodologie di indagine	103
5.2.3 Risultati	104
CAPITOLO 6	
6.1 Discussione	106
6.1.1 Analisi quantitativa di testi tratti da WhatsApp	106
6.1.2 Limiti e prospettive future	109
6.2 Conclusione	110
BIBLIOGRAFIA	112

Capitolo 1

1.1 I metodi quantitativi di analisi del testo

Conoscere l'autore che si cela dietro un testo è una pratica di interesse notevole e di non facile soluzione. I motivi che spingono alla ricerca della persona dietro lo scritto sono molteplici come numerosi sono gli usi dell'indagine sulla paternità di uno scritto. Possono essere utili nella critica letteraria, per aggiungere una chiave di lettura al testo sulla base delle informazioni biografiche dell'autore, in sede legale per certificare il cosiddetto diritto d'autore o individuare situazioni di plagio. In ciascuno di questi casi, l'indagine sull'autorialità si fonda sulla ricerca di caratteri di similarità e caratteri di differenziazione tra due o più opere; da qui la necessità metodologica di organizzare ogni affermazione e ogni risultato sulla base di un confronto.

Fino a circa la metà dell'ottocento l'unico metodo di ricerca è stato di tipo qualitativo; il confronto tra le opere verteva su aspetti legati ai contenuti o particolarità stilistiche, ricorrenti in un autore e meno frequenti in un altro. Xenone ed Ellenico, grammatici alessandrini, ad esempio, sulla base di differenze nei contenuti affrontati, affermarono che l'autore dell'Iliade non fosse lo stesso dell'Odissea, negando, per primi, a Omero la paternità di ambedue le opere.

A partire dalla metà dell'ottocento alle soluzioni qualitative si sono affiancati alcuni metodi quantitativi, che da allora si sono sempre più affinati. L'introduzione di queste strategie, pur mantenendo lo stesso oggetto di indagine, il testo, ha spostato il focus dell'indagine su aspetti più discreti e misurabili del testo, tanto meno legati alla complessità del testo quanto più alle sue singole componenti. Alcuni metodi hanno addirittura abbandonato il materiale linguistico come variabile d'indagine affidandosi a misurazioni indirette, come nel caso dei modelli di compressione del testo (Joula, 2006). Anche per questo motivo la ricerca in merito ai metodi di indagine si è allargata dallo specifico ambito linguistico-letterario a un ambito matematico-statistico-informatico. Il presupposto che fonda questo tipo di approccio sta nella convinzione che misurando alcune proprietà del testo sia possibile identificare l'impronta digitale linguistica (linguistic fingerprint) di un autore e, a partire da questa, distinguere testi scritti da autori diversi (Stamatatos, E. 2009)

L'utilizzo di metodi quantitativi non è, tuttavia, da considerarsi come sostitutivo dell'indagine qualitativa, ma complementare. È un approccio che permette di assumere una prospettiva diversa sul testo, un punto di vista da lontano per usare la definizione di Franco Moretti,

«Distant reading», ho chiamato una volta, un po' per scherzo e un po' no, questo modo di lavorare: dove la distanza non è però un ostacolo alla conoscenza, bensì una sua forma specifica. La distanza fa vedere meno dettagli, vero: ma fa capire meglio i rapporti, i pattern, le forme. *(Moretti 2005)*

Questa rinnovata prospettiva va comunque integrata e riferita a una analisi qualitativa. In questo senso i risultati di una lettura da distante del testo possono essere utilizzati in una duplice ottica: un'ottica confermativa e un'ottica esplorativa (Cortelazzo 2013). L'ottica confermativa consente di utilizzare i dati di una ricerca quantitativa a conferma dei risultati di una analisi stilistica qualitativa, per confermare le intuizioni e le informazioni individuate come caratterizzanti lo stile di un autore. L'ottica esplorativa consente invece di ampliare le prospettive di ricerca, di generare nuove ipotesi, di individuare aspetti e relazioni che a causa della limitatezza, quantitativa appunto, della mente umana possono sfuggire o passare inosservati.

A partire da questa duplice prospettiva i vantaggi di un'analisi quantitativa sembrano ben delineati: permettono di affrontare corpora di grandi dimensioni, che non possono essere gestiti nella loro interezza dalla memoria umana; consentono di ristrutturare il testo rendendo disponibili in tempi rapidi informazioni utili per la classificazione dei testi o per procedere con ulteriori indagini qualitative; facilitano la visualizzazione dei rapporti di somiglianza o distacco tra testi o tra unità testuali e testi. Il costo di questi vantaggi sta nel sacrificio di altri aspetti di rilievo; un'analisi con gli n-grammi di caratteri, ad esempio, potrà concentrarsi maggiormente sulla ripetitività delle sequenze di caratteri tralasciando le informazioni riguardo al lessico mentre un'analisi statistica basata sul lessico rischia di non considerare gli aspetti legati alla sintassi o alla retorica. Inoltre, l'utilizzo di questo tipo di analisi restituisce solamente l'aspetto formale del lessico e non quello semantico; non consente, ad esempio, la discriminazione degli

omografi o il riconoscimento dei sinonimi (Cortelazzo 2013). Alcuni di questi problemi potrebbero essere evitati procedendo con una lemmatizzazione del corpus in esame. Tuttavia, anche la lemmatizzazione porta con sé degli ostacoli. È un processo non ancora completamente automatizzato, che richiede un importante investimento di tempo da parte del ricercatore, a maggior ragione in corpora di dimensioni elevate. Inoltre, con la lemmatizzazione si affacciano alcuni problemi teorico-pratici di difficile soluzione (trattamento delle forme composte del verbo, distinzione tra participi presenti e passati con funzione aggettivale o verbale) (Cortelazzo 2013). In questo senso, nonostante le recenti innovazioni e accelerazioni nell'ambito dei software e dell'intelligenza artificiale hanno affinato gli strumenti di analisi testuale quantitativa rimane fondamentale ai fini di una miglior comprensione e specificazione del dato quantitativo l'affiancamento dell'approccio qualitativo all'analisi del testo.

1.1.1 Sistemi quantitativi.

Con attribuzione d'autore si intende il tentativo di identificare l'autore di un testo, per lo più scritto, a partire da un insieme di testi la cui autorialità è certa (Stamatatos 2009, IN Bozkurt, 2007).

In termini generali, l'attribuzione d'autore cerca di affinare le sue strategie per rispondere a tre quesiti principali (Joula, 2006):

- a. Identificare l'autore a partire da un frammento testuale che, per certo, è attribuibile a uno degli autori di un dato insieme
- b. Identificare, se possibile, l'autore a partire da un frammento testuale che, si suppone, è attribuibile a uno degli autori di un dato insieme
- c. Identificare l'autore a partire da un documento testuale.

I fini di queste indagini possono essere molteplici e includono la verifica di autorialità (decidere se un dato testo è stato scritto da un autore oppure no), il riconoscimento di un plagio (falsa auto-attribuzione di un testo o frammento testuale che è invece opera di altri), il riconoscimento di incoerenze stilistiche (come ad esempio nel caso di testi scritti da più autori). (Stamatatos 2009)

La stilometria o la profilazione d'autore, invece, si occupano di determinare le caratteristiche o proprietà dell'autore di un campione testuale (Joula, 2006): sapere se il testo è stato scritto da una sola persona o al contrario è un lavoro a più mani, conoscere il genere dell'autore, la sua provenienza, la sua formazione.

Come abbiamo visto, gli approcci a questo tipo di indagine possono basarsi su considerazioni qualitative o quantitative.

Nel caso delle analisi supportate da analisi quantitative e stilistiche i primi studi di cui si ha traccia risalgono ai primi tentativi di quantificare lo stile di scrittura da parte di Mendenhall (1887) sulle commedie di Shakespeare; a questo fecero seguito gli studi di Zipf (1932) e Yule (1938, 1944).

È con lo studio di Mosteller e Wallace (1964) sul caso dei Federalist Paper, un insieme di saggi pubblicati tra il 1787 e il 1788 da un autore con lo pseudonimo di Publius, che l'attribuzione d'autore basata su informazioni quantitative ha il suo vero inizio. I due ricercatori utilizzarono un metodo basato su una analisi statistica baesiana sulla frequenza di un insieme ristretto di parole comuni raggiungendo ottimi risultati.

Fino al 1990 la ricerca sull'attribuzione d'autore è stata dominata da tentativi di individuare quei tratti che permettessero di misurare quantitativamente e distinguere lo stile autoriale. Per questo motivo sono stati sviluppati in quegli anni numerosi indici di indagine: lunghezza della frase, lunghezza delle parole, frequenza delle parole, frequenza dei caratteri, ricchezza lessicale. Rudman (1998) stima che sono state proposte quasi 1000 misurazioni differenti. Tuttavia, le ricerche erano assistite dal computer più che basate su di esso, dal momento che non si era ancora raggiunta una completa automatizzazione dei sistemi. Inoltre, uno dei maggiori problemi delle ricerche dell'epoca è da ascrivere all'assenza di oggettività dei metodi impiegati: questi venivano utilizzati per analizzare testi di dubbia o addirittura sconosciuta autorialità non rendendo possibile la valutazione dell'accuratezza del metodo; a questo vanno aggiunte delle limitazioni metodologiche: i dati testuali utilizzati erano troppo lunghi, il numero di candidati troppo piccolo, i corpora non erano controllati in funzione dell'argomento, la valutazione del metodo di indagine era soprattutto intuitiva che conseguiva in una difficoltà nel confrontare i metodi diversi.

Dagli anni '90 la situazione è migliorata; la grande quantità di testi depositati in internet spinse ad affinare i metodi di gestione dei dati stessi con ricadute positive nell'ambito del recupero di informazioni, nella machine learning e nel processamento del linguaggio naturale. Nello specifico, nel caso del recupero delle informazioni, si migliorarono le tecniche di classificazione e rappresentazione di testi di ampiezza sempre più elevata, nell'ambito della machine learning si svilupparono nuovi algoritmi in grado di gestire dati multidimensionali facilitando una rappresentazione dei dati più espressiva. Nell'ambito dell'elaborazione del linguaggio naturale vennero individuati nuovi strumenti per analizzare il testo fornendo nuove forme di misura di rappresentazione dello stile. L'affinamento della tecnica ha permesso un utilizzo dei metodi di attribuzione d'autore ad altri settori, dai servizi di intelligence, alla lotta al crimine, al diritto civile, (dispute sul copyright).

L'attribuzione d'autore è diventata dunque una questione multi disciplinare, sia nel momento in cui si conduce l'analisi (in cui è richiesto l'intervento di linguisti, filologi, statistici e matematici) sia nel momento di fruizione dei risultati, dato che possono interessare persone impiegate nell'ambito dell'educazione, della legge, della sicurezza. Ciò che deve guidare le metodologie dell'attribuzione d'autore è la ricerca dell'accuratezza e dell'affidabilità dello strumento di valutazione e accanto a questa un sistema in grado di rendere conto delle caratteristiche o dei tratti del testo che sono risultati decisivi in una specifica attribuzione.

Infine, in un sistema perfetto di attribuzione d'autore è necessario disporre di un corpus regolarmente costruito. (Joula 2006)

Secondo la definizione di Tuzzi (2003):

il corpus si configura come una collezione di testi. Il corpus raccoglie testi coerenti con gli scopi perseguiti dalla ricerca e questa coerenza è valutabile solo discrezionalmente. Nello studio dell'intera opera di un autore i testi costituenti il corpus possono essere, per esempio, le singole opere inedite e/o inedite di cui si conosce l'esistenza; nello studio di un romanzo i singoli capitoli; nell'analisi dei risultati di un'indagine con intervista a domande aperte le trascrizioni dei colloqui, nell'analisi di annate di stampa i quotidiani (o i settimanali o mensili ecc.) pubblicati.

In un'ottica di analisi quantitativa per l'attribuzione d'autore la costruzione del corpus dovrà tener conto di diversi parametri; innanzitutto l'insieme dei testi dovrà essere organizzato in funzione di un ben preciso quesito o scopo di indagine, dovrà considerare testi dello stesso genere letterario e infine dovrà essere equilibrato dal punto di vista della lunghezza dei testi coinvolti.

1.1.1.1 Tratti utilizzati dai sistemi quantitativi

Per analizzare e misurare il linguaggio è necessario creare dei modelli semplificati; si può osservare come un testo è una struttura di eventi separati selezionati da una popolazione di eventi possibili. Gli eventi presi in considerazione possono essere caratteri, sequenze di caratteri, parole, sequenze di parole, frasi o periodi. I diversi eventi stanno tra di loro in un rapporto che non è governato dal caso, ma da regole di più alto ordine. Ci sono diversi modi per trattare questo tipo di regolarità. Uno dei più sofisticati e psicologicamente plausibili (Joula 2006) sono le *context free grammars* e le loro specificazioni. Queste possono essere descritte come un insieme di regole che permettono di riscrivere simboli astratti come stringhe di altre categorie o parole specifiche. Ad esempio, il codice per il sintagma preposizionale (in inglese) dato dal simbolo PP potrebbe essere riscritto come preposizione (PREP) seguito da un sintagma nominale (NP) che a sua volta potrà essere espresso come articolo (ART) più nome e così dicendo. Questo modello, sensibile nella rappresentazione della dipendenza strutturale degli elementi non è altrettanto funzionale alla descrizione di dipendenze semantiche o lessicali. Più spesso il linguaggio è analizzato come un semplice "borsa di parole", un insieme di tutte le parole che compaiono nel testo senza contare l'ordine di occorrenza. Più precisamente una *bag of words* è la raccolta di tutte le parole (*tokens*) che compaiono nel testo.

Un'altra proprietà fondamentale del linguaggio è la distribuzione delle parole e dei sintagmi; questa tende a essere non uniforme e irregolare in funzione del contesto. Alcune parole sono molto più comuni di altre e i contesti diversi originano diverse probabilità di selezione di una parola; la probabilità assoluta di selezione della parola *the* in inglese è pari al 7%, ma dato un contesto in cui la parola precedente a quella da selezionare sia *the* la probabilità che sia di nuovo selezionata *the* scende allo 0%.

La distribuzione delle parole è descritta dalla legge di Zipf o distribuzione di Zipf, che dimostra come la frequenza di una parola sia inversamente proporzionale alla sua posizione nella distribuzione: il legame matematico tra la frequenza di una word type e il rango della parola nel vocabolario ordinato per lunghezza decrescente. Le parole del vocabolario si distribuiscono nei testi in modo che la frequenza e il rango siano inversamente proporzionali, secondo una costante di proporzionalità c (Tuzzi 2003).

Va inoltre notato come alcune proprietà delle parole variano in funzione della frequenza. Le parole più frequenti tendono infatti a essere più corte, meno precise nella denotazione e meno scomponibili in tratti semantici più piccoli. Allo stesso modo le diverse parti grammaticali del discorso tendono a occupare posizioni diverse nello spettro di frequenza. Le parole che occupano le posizioni più alte nella classifica distribuzionale, e che quindi sono utilizzate con più frequenza all'interno di un testo, sono le parole che appartengono alla classe chiusa, le cosiddette parole funzione. Sia le parole funzione che i costrutti marcati aprono a diverse possibilità di espressione di un concetto, di conseguenza gli autori possono scegliere vie diverse, facilitandone il riconoscimento attraverso i modelli di analisi.

The essence of authorship analysis is the formation of a set of features, or metrics, that remain relatively constant for a large number of writings created by the same person. In other words, a set of writings from one author would exhibit greater similarity in terms of these features than a set of writings from different authors (Zheng et al. 2003).

Nei paragrafi successivi si affronteranno singolarmente le diverse proprietà che possono essere utilizzate per quantificare lo stile di un testo, tra quelle maggiormente citate in letteratura.

1.1.1.1.1 Tratti lessicali

Le parole utilizzate all'interno di un testo possono essere utilizzate come forti indicatori dell'identità di un autore. Le scelte lessicali possono indicare il momento storico in cui un testo è stato scritto, identificare la provenienza o l'appartenenza sociale di un autore;

anche gli errori, se ripetuti con costanza possono essere utilizzati come strumento per riconoscere l'identità di un autore; in questo senso sono stati sviluppati degli strumenti in grado di estrarre autonomamente queste proprietà di vocabolario. (Koppel, Schler, 2003)

Vi sono essenzialmente due problemi legati a questo tipo di analisi:

- a. La possibilità che ciascun autore possa dirigere intenzionalmente la scelta delle parole in modo tale da deviare o sviare gli indizi che porterebbero a un suo riconoscimento.
- b. Le parole interessanti in senso diatopico, diafasico, diacronico o errori sistematici di *spelling* potrebbero non essere presenti nel testo, non dando appigli ad alcuna osservazione in un'ottica di attribuzione d'autore.

I successivi metodi che si sono basati sulle proprietà lessicali del corpus sono partiti dalla considerazione secondo la quale ciascuna parola di un documento ha delle proprietà come la lunghezza, il numero di sillabe, la lingua di origine, la parte del discorso. Molti approcci classici hanno provato a utilizzare questo tipo di proprietà in una forma statistica. Kruh ad esempio utilizzò la lunghezza media della frase per dimostrare che i manoscritti di Beale fossero dei falsi. Un'altra misurazione utilizzata è quella della ricchezza del vocabolario, una stima dell'ampiezza del vocabolario dell'autore. (Joula 2006)

In altri casi (Mosteller, Wallace) l'analisi si è concentrata su un sottoinsieme ben definito di parole (parole funzione).

Il vantaggio che l'utilizzo di un metodo di questo tipo può garantire è il poter utilizzare testi scritti in qualsiasi lingua naturale avvalendosi semplicemente di un *tokenizer* (uno strumento in grado di contare le parole di un testo e organizzarle in una lista sulla base della frequenza), secondo il modello delle *bag of words*.

Una volta che tale lista è generata, è possibile misurare la ricchezza di vocabolario: attraverso la *type/token ratio* (TTR, il rapporto tra il numero di *types* [forme] e il numero di *tokens* [il numero di parole di un testo]) e il numero di *hapax* (parole che ricorrono soltanto una volta nel testo) presenti nei testi. Questo tipo di dato è fortemente dipendente dalla lunghezza del testo, nonostante i tentativi di Yule 1944 e Honore 1979

di individuare una stabilità indipendentemente dalla lunghezza del testo (Stamatatos, 2008)

L'approccio più comunemente utilizzato è la rappresentazione del testo attraverso dei vettori di frequenza delle parole, in cui ogni testo rappresenta un vettore a n dimensioni, dove a ciascuna dimensione è associata una parola del testo. Il peso, inteso come frequenza di occorrenza di ciascuna parola, influisce nella specificazione della dimensione, la somma delle dimensioni restituisce una posizione nello spazio al vettore (il testo). Le posizioni nello spazio di questi vettori testuali consentono di stabilire dei rapporti di vicinanza tra i testi e di conseguenza dedurre delle somiglianze o differenze di stile.

Nell'attribuzione d'autore sono, spesso, utilizzate come parametri le sole parole funzione dal momento che sono meno manipolabili direttamente e consciamente dall'autore, oltre a essere indipendenti dall'argomento di cui il testo tratta; per questo motivo sono considerate più affidabili nella discriminazione tra autori. La selezione delle parole funzione da utilizzare è piuttosto arbitraria oltre che lingua-dipendente. Diversi autori hanno proposto un insieme di parole funzione da utilizzare per le analisi della lingua inglese; Abbasi e Chen hanno riportato un insieme di 150 parole, Zhao e Zobel un insieme di 365.

Quando invece si vuole stabilire una classificazione in funzione dell'argomento o del tema di un gruppo di testi è più probabile che vengano utilizzate le parole piene.

Una volta identificate le parole più frequenti è da decidere fino a quale posizione nella classifica di occorrenza utilizzare le parole ai fini dell'attribuzione d'autore. Nei primi studi vennero considerate le prime 100 parole più frequenti (Burrows 1987, 1992); con il miglioramento dei dispositivi di analisi del testo (*machine learning, supported vector machine*) il numero di parole più frequenti è salito fino a 1000 (cfr. Stamatatos 2006a) o in altri casi si è arrivati a considerare tutte le parole che comparissero almeno due volte nel testo. Una analisi che utilizzasse le prime 100 parole più frequenti si baserebbe, sostanzialmente, sulle parole funzione (che di norma occupano buona parte delle prime 100 posizioni).

L'analisi per tratti lessicali può giovare della lemmatizzazione che permette di abbassare il rumore delle alternative morfologiche di ciascuna parola, restituendo una

classificazione più accurata dell'ampiezza lessicale dell'autore e delle sue scelte lessicali. Allo stesso modo gli strumenti che riducono le parole al tema morfologico (*stemmers*), come quelli che sono in grado di distinguere le omografe, consentono una classificazione più accurata delle scelte lessicali di un autore.

Come abbiamo visto l'approccio per *bag of words*, nel restituire una lista delle parole occorrenti nel testo, in un ordine governato dalla frequenza di occorrenza, rompe la struttura lineare del testo e scompone i rapporti tra le parole. Per rendere conto dei rapporti tra le parole si sono costruiti i cosiddetti n-grammi di parole, sequenze di 2 o 3 parole, anche queste ordinate per frequenza, che rendono conto del contesto di utilizzo delle parole, disambiguando in alcuni casi i significati delle parole.

I risultati ottenuti con questi metodi tuttavia non hanno dato risultati migliori di quelli ottenuti attraverso una analisi delle singole parole (Stamatatos 2006). Uno dei problemi maggiori sta soprattutto nella difficoltà di trovare sequenze di parole con una frequenza tale da poter stilare una classificazione (soprattutto al crescere di n).

1.1.1.1.2 Tratti grafici/ n-grammi di caratteri.

Il presupposto dell'analisi per tratti grafici consiste nel considerare ciascun testo come una semplice sequenza di caratteri. Partendo da questo presupposto le analisi che si possono condurre sono molteplici.

Si possono contare tutti i caratteri che compaiono nel testo, compresi gli spazi e i segni di punteggiatura o contare i soli caratteri alfabetici, i caratteri in maiuscolo, quelli in minuscolo, la sola punteggiatura, l'occorrenza di una precisa lettera. Un approccio più elaborato è quello del conteggio delle sequenze di caratteri nel testo (n-grammi). Una analisi per trigrammi di caratteri dell'inizio di questo paragrafo risulterebbe in: [il_], [l_p], [_pr], [pre], [res], [esu], [sup], [...]. Questo tipo di sistema è in grado di catturare diverse sfumature, informazioni lessicali, indizi sui contenuti, usi della punteggiatura, etc. Un ulteriore vantaggio di questo metodo è la resistenza al rumore dei testi che potrebbe essere dato da un uso scorretto della punteggiatura, da alcuni errori, da una cattiva formattazione. La parola "comunque" e la parola "*comunque" presentano infatti un buon numero di trigrammi in comune, mentre in un conteggio delle parole sarebbero classificate come due forme diverse.

All'interno dell'analisi d'autore vengono conteggiate le sequenze di caratteri più frequenti, che si è visto sono degli indicatori piuttosto efficaci dello stilo di un autore. A differenza delle parole, le sequenze di caratteri o n-grammi non sono lingua-dipendenti; inoltre, la divisione in sequenze di caratteri consente di fornire un campione di dati molto più alto rispetto all'approccio basato sulle parole, consentendo di avere una maggior efficacia con testi di dimensioni ridotte.

L'efficacia dell'uso degli n-grammi in compiti di analisi quantitativa e attribuzione d'autore è stato avvalorato da numerose ricerche; alcuni studiosi hanno anche affermato che l'utilizzo dei caratteri sia più funzionale e dia risultati migliori rispetto all'uso delle parole più frequenti. Grieve (2007) dimostra poi, attraverso una analisi di uno stesso corpora con strumenti diversi, che gli n-grammi di caratteri sono la misura più efficace nell'attribuzione d'autore.

Uno degli aspetti da tenere maggiormente in considerazione quando si parla di *character n-grams* è il valore di n. Non c'è un accordo sostanziale riguardo al numero di caratteri da considerare all'interno della sequenza. Certamente ogni scelta porta con sé delle conseguenze. Un valore di n alto è più sensibile nel cogliere aspetti lessicali e aspetti legati al contesto, mentre un valore di n più piccolo restituirebbe informazioni legate alla sfera sub-lessicale o sillabica. Inoltre, la scelta del valore di n- è da considerarsi lingua-dipendente; alcune lingue (tedesco-greco) tendono a utilizzare parole con una lunghezza media superiore a quelle utilizzate, ad esempio, dall'inglese. Per queste lingue, probabilmente, la scelta di un valore alto di n- potrebbe essere più sensibile nel cogliere le particolarità e i tratti della lingua dell'autore.

Un uso particolare dei caratteri è l'approccio basato sulla compressione (*compression-based approach*) [Benedetto, Caglioti, Loreto]. Questi sistemi si basano sull'utilizzo dello schema di compressione, con il quale si è compresso un testo A, per comprimere un testo B. Si è visto che quando i testi sono stati scritti dallo stesso autore la dimensione in bit del file compresso è più bassa rispetto a quando i due testi sono di due autori differenti. Il modello di compressione è solitamente basato sulla ripetizione dei caratteri e, per questo motivo, è annoverato tra i metodi basati sugli n-grammi.

1.1.1.1.3 Tratti sintattici.

L'utilizzo di tratti sintattici come proprietà dalle quali desumere le informazioni riguardo allo stile di un autore assume una importanza ancora maggiore se si considera che è proprio la ricorrenza di schemi sintattici, più che di parole, a costituire una solida impronta digitale dell'autore. Basti considerare l'efficacia dei metodi attribuzione d'autore che si basano unicamente sulle parole funzione, direttamente connesse, poiché le introducono, alle strutture sintattiche. A differenza dell'analisi dei caratteri, quella sintattica è altamente lingua-dipendente e richiede un software specializzato per ciascuna lingua e che riconosca le diverse strutture.

Primi in questo tipo di analisi furono van Halteren e Tweedie che riuscirono in maniera semi automatica a dare per ciascuna frase di un corpus il rispettivo albero sintattico e da questi estrassero le frequenze delle regole. Ciascuna regola esprimeva una parte dell'analisi sintattica. Le regole erano descritte in questo modo:

A: PP → P: PREP + PC: NP

(un sintagma avverbiale preposizionale è costituito da una preposizione seguita da un sintagma nominale come complemento preposizionale). Una struttura di questo genere consente di individuare la classe grammaticale per ciascuna parola e la modalità con cui ciascuna parola è combinata alle altre.

Un altro tentativo di utilizzare le informazioni sintattiche in maniera quantitativa è quello proposto da Stamatatos (2000). La modalità proposta semplifica le operazioni di analisi sintattica, non c'è una analisi dettagliata del sintagma né dei rapporti tra i diversi sintagmi. Per cui una frase come quella che introduce il paragrafo potrebbe essere analizzata nel seguente modo:

NP (un altro tentativo) VP (di utilizzare) NP (le informazioni sintattiche) PP (in maniera quantitativa) VP (è quello proposto da S.)

Da questo tipo di analisi venivano estratti il numero di sintagmi di ciascun tipo (verbale, nominale, preposizionale) e la lunghezza di ciascun sintagma. Allo stesso modo possono

essere utilizzati sistemi di analisi sintattica parziale per estrarre informazioni riguardo alle proprietà sintattiche di un testo. Hirst e Feiguina (2007) trasformarono il risultato di una analisi parziale in una sequenza di etichette sintattiche per cui un sintagma come “a simple example” risultava in:

NX DT JJ NN

che si può tradurre in un sintagma nominale composto da un articolo, un aggettivo e un nome. A partire da questa sequenza di codici misurarono la frequenza dei bigrammi e risultò in uno strumento piuttosto efficace nell’attribuzione di testi di dimensioni ridotte (sotto le 200 parole).

Un altro metodo piuttosto efficace e più semplice rispetto a quello dell’analisi sintattica parziale è considerato quello dell’analisi delle parti del discorso (POS tagger). Esistono dei sistemi automatici che consentono di assegnare a ciascuna parola una marca morfo-sintattica. Questi sistemi permettono di calcolare la frequenza o il rapporto tra le parti del discorso nel testo ma non permettono di descrivere le combinazioni delle parole a formare sintagmi o dei sintagmi a formare frasi.

La rassegna condotta da Stamatatos nel 2009 cita altre forme di analisi sintattica; alcuni ricercatori hanno basato le indagini di autoriali sugli errori sintattici, in maniera piuttosto simile a queglii strumenti che si basavano sugli errori di digitazione delle parole. Altri ricercatori (Karlgrén, Eriksson 2007) descrissero un modello preliminare basato sulle espressioni avverbiali e sull’occorrenza di frasi incassate all’interno di frasi complesse, un tentativo di rappresentare le proprietà distribuzionali delle frasi nel testo.

1.1.1.1.4 Tratti semantici

La computazione automatica di dati relativi alla semantica e alla pragmatica non ha raggiunto i livelli che si sono raggiunti con gli altri livelli della lingua (fonologico, lessicale, sintattico). Diversi sono stati i tentativi di rendere in forma quantitativa il dato semantico. Gamon (2004) creò un grafico di dipendenza semantica dal quale venivano estratti due tipi di informazione: tratti semantici binari e relazioni di modificazione semantica. Il primo tipo restituiva le informazioni riguardo al numero e alla persona del

nome, il tempo e l'aspetto del verbo etc. Il secondo dato invece informava della relazione sintattica e semantica tra un nodo del grafico e il nodo sorella (ad esempio un nodo nominale con un modificatore nominale di tipo locativo). Un metodo interessante nell'ambito dell'analisi semantica è quello proposto da Argamon (2007). Definì un insieme di tratti funzionali che associava ciascuna parola o ciascun sintagma a una informazione semantica. Ad esempio, il sintagma "in altre parole" è classificato come una APPOSIZIONE di una ELABORAZIONE, di una CONGIUNZIONE. Una volta creato un vocabolario di termini classificati in questo modo fu possibile contare e misurare quante volte, ad esempio, una congiunzione fosse espansa in una elaborazione e quante elaborazioni in apposizioni. I risultati dimostrarono una efficacia del metodo quando associata ad analisi basate sulle parole funzione.

Un sistema simile è quello sviluppato all'interno delle strategie di *Sentiment Analysis* che, a partire da un vocabolario di termini a cui è associata un'etichetta a seconda che veicoli un significato positivo, negativo o neutro, sono in grado di calcolare il grado di positività o negatività di un testo, sulla base del numero di occorrenze etichettate in un modo o nell'altro.

1.1.1.1.5 Tratti specifici per applicazione

I tratti fino a ora discussi si definiscono come indipendenti dall'applicazione dal momento che possono essere estratti da qualsiasi tipo di testo. In alcuni casi un ricercatore può servirsi di tratti specifici in grado di cogliere le sfumature di un tipo di testo.

Analisi strutturali per misurare lo stile autoriale: indentazione, l'uso di formule fisse, emoticon, lunghezza dei paragrafi, tipo di font, grandezza del font, colore del font, l'uso della giustificazione, il numero di spazi che seguono la punteggiatura. È chiaro che informazioni di questo tipo potranno essere utilizzate unicamente contestualmente a determinati e ben specifici generi testuali. Sono molto utili nell'analisi di testi di piccole dimensioni anche se possono essere, più che gli aspetti linguistici, influenzati da scelte editoriali o di impaginazione.

Quando i testi che formano il corpus di indagine trattano di uno stesso argomento anche la scelta o la frequenza di occorrenza di termini legati all'argomento possono rivelare delle precise scelte autoriali.

Questi metodi, dipendenti dal testo del quale si vuole indagare l'origine o la natura autoriale, hanno di per sé un'ampiezza di estensione limitata ma al contempo possono determinare un'altissima affidabilità in questioni specifiche; un chiaro esempio è lo strumento sviluppato da Tambouratsiz, che sulla base dell'analisi delle terminazioni verbali è riuscito a mettere a punto uno strumento capace di discriminare tra due varianti diglossiche.

1.2 Metodi per l'attribuzione d'autore

La maggior parte delle indagini di attribuzione d'autore implicano un compito di classificazione multi-classe nel quale un testo di autorialità incerta viene classificato secondo un autore più probabile sulla base di tratti stilistici e utilizzando algoritmi di machine learning con apprendimento supervisionato (Srinivasa Rao O. et al, 2017). In altri casi, in particolare quando le informazioni riguardo l'autorialità di un documento non ci sono o non sono affidabili è preferibile ricorrere a metodi di attribuzione con apprendimento non supervisionato attraverso il *clustering* (o analisi dei gruppi). Mentre nella classificazione sono costruite delle classi di dati predeterminate a cui è assegnata un'etichetta (es. caratteristiche stilistiche di 1, caratteristiche stilistiche di 2) e alle quali si dovrà fare riferimento nella fase di classificazione (es. il testo X di origine sconosciuta è più vicino ad 1 che a 2), nel clustering non sono date delle classi precostituite. L'insieme dei dati, nel caso della stilometria i testi che compongono il corpus, è organizzato in gruppi (*clusters*); gli oggetti che fanno parte di uno stesso gruppo avranno una maggiore similarità e di conseguenza oggetti appartenenti a due gruppi diversi tenderanno a essere molto diversi tra loro.

1.2.1 Metodi per l'attribuzione d'autore: classificazioni multi-classe.

Questo tipo di metodo si serve di un corpus contenente i testi riconosciuti di un gruppo di autori, che chiameremo corpus A, e un corpus di testi di cui non si conosce l'autore ma si ritiene essere compreso all'interno del gruppo di autori considerato, corpus B. Di

conseguenza il corpus contenente i testi di autore conosciuto, suddiviso in tanti sub-corpora quanti gli autori considerati, è utilizzato come modello per riconoscere i testi di paternità incerta. La modalità di approccio ai testi del corpus A può svilupparsi in tre modi differenti. Stamatatos riconosce degli approcci *profile-based*, degli approcci *instance-based* e infine degli approcci ibridi. La distinzione tra i due metodi, l'approccio ibrido unisce alcuni aspetti del primo e altri del secondo, è da considerarsi come fondamentale nell'ambito dell'attribuzione d'autore.

1.2.1.1 Approcci profile-based.

Secondo questo tipo di approccio i testi di ciascun sub-corpus, ovvero tutti i testi raccolti di un determinato autore, sono sommati in un unico file; e proprio a partire da questo file unico verranno estratte le informazioni stilistiche dell'autore da confrontare alle caratteristiche di un testo di paternità sconosciuta, come mostrato dalla figura 1.

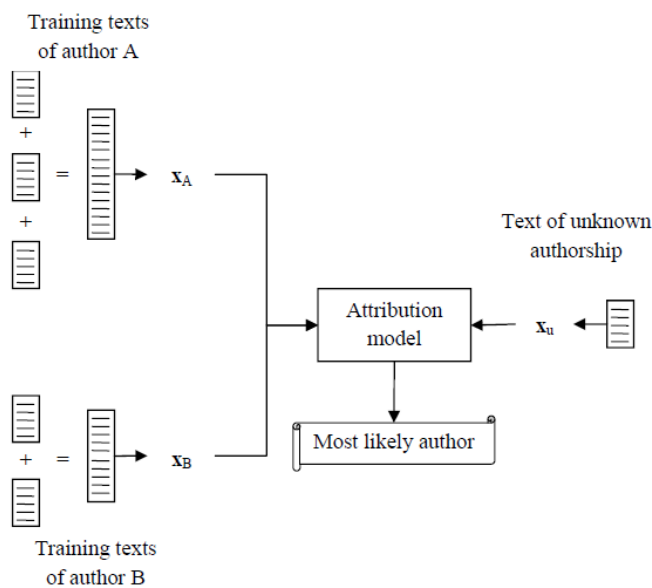


Figura 1: Struttura di un approccio profile-based (Stamatatos E. 2006)

Un approccio di questo genere consente di neutralizzare le differenze interne a ciascun testo del sub-corpus creando un unico grande documento di testo.

Il procedimento di *training* utilizzato da questo modello consiste nell'estrazione di un profilo per ciascun autore candidato; il sistema di attribuzione misurerà poi la distanza tra il profilo emerso dal testo di origine sconosciuta con il profilo dei diversi candidati.

Consideriamo $P(x)$ il profilo emerso di un testo X , profilo che come abbiamo visto potrà basarsi sulle parole più frequenti, oppure sugli n -grammi, oppure su caratteristiche sintattiche; la distanza tra $P(x)$ e $P(y)$ sarà espressa dalla formula $d(P(x), P(y))$. A decretare l'autore più probabile, all'interno del gruppo di candidati, sarà il minore rapporto di distanza tra i due profili.

Le applicazioni di questo metodo possono servirsi di modelli di diversa natura.

Si possono servire di modelli probabilistici con diverse varianti (Mosteller & Wallace 1964, Clement & Sharp 2003, Sebastiani 2002, Peng 2004). Gli approcci *profile based* possono anche servirsi di modelli di compressione. Abbiamo discusso nel paragrafo precedente, in particolare nella descrizione dei sistemi che basavano sulla sequenza di caratteri come indice di autorialità, dei meccanismi di compressione del testo. In questo senso i testi di un intero sub-corpus sono riuniti in un unico file $P(x)$ che a sua volta, attraverso un algoritmo di compressione, è compresso in $C(x_a)$. La stessa operazione è condotta per ciascun sub-corpus. Poi a ciascun sub-corpus è sommato il testo di cui si vuole indagare l'origine autoriale e su questo profilo $P(x+z)$ è svolta una nuova compressione. La differenza in termini di dimensione in bit dei documenti compressi indicherà l'autore più probabile. La differenza calcola la cross-entropia tra i due testi, una misura di quanto due eventi differiscono; una differenza pari a 0 si verifica quando i due testi sono completamente identici, valori di differenza più alti indicano invece una crescente dissimilarità tra i testi (Joula, 2006).

In genere l'utilizzo di metodi basati sull'approccio probabilistico è più rapido rispetto ai metodi basati sulla compressione; inoltre i metodi probabilistici possono applicarsi sia alle frequenze di caratteri sia alle frequenze di parole mentre i metodi di compressione agiscono sulle sequenze di caratteri.

L'approccio CNG (*Common N-Grams*) è infine meritevole di menzione tra quelli *profile based*. Secondo questo approccio il profilo $P(x)$ di un testo è dato dagli engrammi (L) più frequenti in quel testo. La misura della distanza è espressa dalla formula:

$$d(PR(x), PR(y)) = \sum_{g \in P(x) \cup P(y)} \left(\frac{2(f_x(g) - f_y(g))}{f_x(g) + f_y(g)} \right)^2$$

dove g è un n -gramma e $f_x(g)$ e $f_y(g)$ sono rispettivamente la frequenza di occorrenza del n -gramma nel testo x e nel profilo y . La differenza tra due profili è calcolata sulla base differenze relative della frequenza degli n -grammi che condividono. Gli n -grammi non condivisi dai due profili rivestono un ruolo di costante nella misura della distanza. La validità degli studi di attribuzione sviluppati con questo metodo sono state verificate per un numero di n -grammi comuni compreso tra 1000 e 5000 e per una lunghezza della stringa di caratteri compresa tra 3 e 5.

Un problema che spesso si riscontra, quando si vogliono utilizzare sistemi basati sul profilo, può essere dato dal disequilibrio tra i diversi sub-corpora. Non è affatto raro infatti che il sub-corpora di un autore comprenda un maggior numero di testi rispetto a un secondo sub-corpora o che tale disparità sia altresì data da testi di dimensioni estremamente diversi. Allo stesso modo, nei casi di metodi *instance based*, ci si può trovare di fronte a sub-corpora che non possono essere scomposti, a causa della lunghezza, in testi separati. È il problema della classe sbilanciata. L'approccio CNG tende a fallire quando si trova di fronte a trarre le sue conclusioni a partire da un corpus A non bilanciato. Per superare il problema Frantzeskou, Stamatatos, Gritzalis e Katsikas (2006) proposero un nuovo tipo di distanza chiamata (SPI) intersezione di profili semplificata che contava solamente la quantità di n -grammi comuni e tralasciando il resto.

1.2.1.2 Approcci instance-based

Gli approcci *instance-based*, quelli utilizzati nella maggior parte dei moderni studi di attribuzione del messaggio utilizzano i testi di ciascun sub-corpus in maniera individuale. Nella fase di addestramento (*training*) ciascun testo di un dato autore concorre singolarmente al processo di decodifica delle informazioni stilistiche (descritte nella sezione dei tratti) e nella comparazione con il testo di autorialità incerta (come mostrato dalla figura 2).

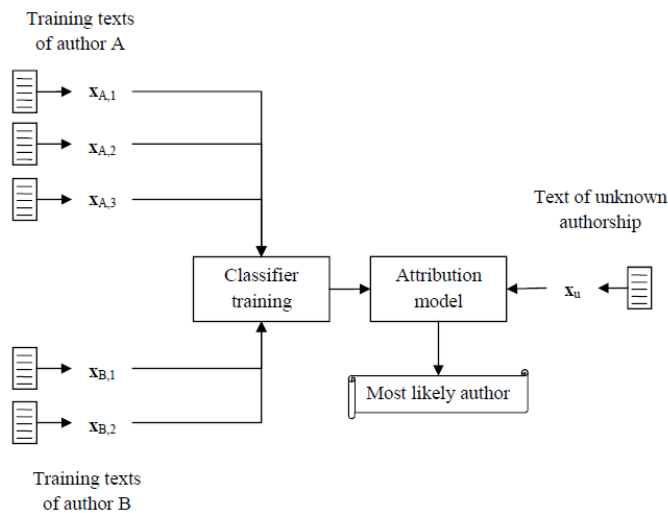


figura 2: struttura di un approccio instance-based (Stamatatos 2006)

Ciascun testo del sub-corpus è definito da un vettore di attributi (x) e un algoritmo di classificazione è addestrato a utilizzare l'insieme dei testi di ciascun sub-corpus per sviluppare un modello di attribuzione. Affinché questi approcci funzionino e siano affidabili, è necessario che ciascun sub-corpus sia rappresentato da più campioni di testo di un autore. Nei casi, dunque, in cui sia presente un solo testo a rappresentare un autore sarà necessario suddividerlo in più parti, possibilmente di uguale lunghezza. Allo stesso modo quando all'interno di un sub-corpus i testi di un autore sono rappresentati da più campioni ma di lunghezza variabile si deve procedere a una normalizzazione della lunghezza testuale. I testi dovranno essere sufficientemente lunghi per poter estrarre delle informazioni adeguate. Sanderson e Guenter nel 2006 hanno proposto una suddivisione dei testi in documenti da 500 caratteri. Altri studiosi hanno proposto una divisione in documenti di 500 parole. Hirst e Feiguina hanno condotto esperimenti con documenti di 200, 500 e 1000 parole, osservando una diminuzione dell'accuratezza nell'attribuzione inversamente proporzionale alla lunghezza del testo. I modelli che possono essere utilizzati con l'approccio *instance based* possono essere basati su vettori spaziali (*vector base model*) o basati sulla similarità (*similarity-based models*).

1.2.1.2.1. Modelli basati su vettori spaziali

In questo tipo di modelli ciascun testo è considerato come un vettore in uno spazio multivariato: di conseguenza la classificazione di questi testi è resa possibile da una serie di algoritmi statistici e di machine learning tra i quali il *Supported Vector Machine* (SVM)

(de Vel et al 2001; Diederich et al. 2003; Teng et al. 2004, Sanderson e Guenter 2006), gli alberi decisionali (Uzuner & Katz 2005; Zhao e Zobel 2005), le reti neurali o algoritmi genetici. Meritano approfondimento in questa sezione i meccanismi che stanno alla base del *supported vector machine* e degli alberi decisionali e del *Random Forest*, dal momento che ritorneranno più avanti nella discussione.

1.2.1.2.1.1. Supported Vector Machine (SVM)

Una macchina di vettore a supporto è un algoritmo che determina il miglior confine decisionale tra vettori che appartengono a un determinato gruppo e vettori che non appartengono a quel gruppo. Un vettore può essere immaginato come una lista di numeri che rappresentano un insieme di coordinate nello spazio. Una macchina di vettore a supporto consente di decidere dove disegnare la miglior linea o il miglior iperpiano che divida lo spazio in due porzioni: una per i vettori che appartengono a una determinata categoria e una per i vettori che non appartengono a quella categoria. Una volta che l'algoritmo ha determinato la posizione del confine (attraverso l'analisi dei testi del *training*) si potranno testare e classificare altri testi sulla base della porzione in cui ricade la sua rappresentazione. (Diederich et al. 2003)

1.2.1.2.1.2 Alberi decisionali e random forest

Il sistema decisionale Random Forest è un insieme di metodi di apprendimento che permette di classificare una serie di dati (testi nel nostro caso) attraverso la costruzione di alberi decisionali. Un albero decisionale può essere definito come una struttura binaria nella quale ciascun nodo rappresenta una domanda relativa a un attributo e alla sua espressione nell'insieme di dati. Ciascun ramo rappresenta una risposta o un risultato della domanda mentre ciascuna foglia rappresenta la classificazione, la decisione presa dopo aver interrogato ciascun attributo. Le proprietà utilizzate per generare le domande all'interno dell'albero decisionale possono essere le parole più frequenti, le sequenze di parole più frequenti (*words bigram*), le sequenze di bigrammi o trigrammi più frequenti, la frequenza delle parole funzione, la ricchezza di Hapax legomena, la lunghezza media delle parole (Khonji M., Iraqui Y., Jones A. 2015). La necessità di una foresta di alberi (*random forest*) permette di compensare e correggere

l'abitudine dell'albero decisionale all'*overfitting*: la tendenza di una analisi a essere troppo specifica nella rappresentazione di un particolare insieme di dati e, di conseguenza, a essere meno precisa e affidabile nelle predizioni future.

1.2.1.2.2 Modelli basati sulla similarità.

Nei modelli basati sulla similarità testuale l'obiettivo consiste in una misurazione a coppie della somiglianza tra un testo di origine sconosciuta e tutti i testi del *training*; la stima dell'autore più probabile sarà basata sul *nearest-neighbor algorithm*. Tra i modelli basati sulla similarità troviamo l'approccio Delta, secondo Burrows, l'approccio di meta-learning tra cui l'*unmasking method* di Koppel e altri (Koppel, Schler & Bonchek-Dokow 2007).

L'approccio Delta calcola la distribuzione z di un gruppo di parole funzione (in origine le prime 150 parole più frequenti). Quindi per ciascun documento viene assegnato un valore z a seconda della deviazione della frequenza della parola rispetto alla media; un valore $z > 0$ è assegnato a parole che compaiono di frequente, un valore $z = 0$ a parole che hanno una frequenza all'interno della media, mentre è assegnato un valore $z < 0$ a parole con una frequenza sotto la norma. La misura Delta che indica la differenza tra un gruppo di testi scritti da uno stesso autore e testi scritti da autori diversi è data dalla differenza in valori assoluti dei punteggi z delle parole del *training set* e il valore z delle corrispondenti parole funzione del testo di autorialità incerta. Più il valore Delta è piccolo maggiore sarà la somiglianza del testo di autorialità incerta con i testi di un determinato autore. Questo metodo è risultato efficace con testi di almeno 1500 parole, ma ha dimostrato di poter essere significativo anche con testi di dimensioni inferiori. Uno studio condotto da Hoover (2004a) dimostrò come l'efficacia del Delta di Burrows fosse maggiore con un numero di parole frequenti superiore alle 500.

1.2.1.3 Approcci ibridi

Questo tipo di approcci, descritto da van Halteren (2007), condensa i due approcci descritti precedentemente. I testi che per ciascun autore compongono i campioni di *training* sono rappresentati separatamente, come nell'approccio instance based. In seguito, i vettori che rappresentano ciascuna istanza testuale sono raggruppati per

autore e unificati in un unico vettore che risulta dalla media dei valori delle proprietà che componevano il vettore di ogni testo, come accade negli approcci *profile-based*. Infine, è misurata la distanza tra il profilo del testo di origine sconosciuta e il profilo dell'autore.

1.2.2 Metodi di attribuzione d'autore: hierarchical clustering methods.

I metodi di *clustering*, come detto, consentono di raggruppare i testi di un determinato corpus in funzione della loro somiglianza o differenza. I testi del corpus concorrono allo stesso modo alla costruzione del dendrogramma; da ciascun testo è estratto un vettore di informazioni, basato sulle parole o sequenze di caratteri più frequenti, che è confrontato con tutti i vettori di dati degli altri testi per generare dei rapporti di somiglianza o distanza di stile (Evert et al 2017).

La misura della distanza testuale si rifà originariamente al lavoro di Burrows e della distanza delta di cui abbiamo parlato nel paragrafo precedente. L'implementazione di questa misura di distanza introdotta da Burrows ha dato il via a una serie di ricerche che hanno dapprima validato l'efficacia del metodo (Hoover 2004a) e la possibilità di utilizzarlo con altre lingue (Eder, Rybicki 2013). In seguito, ha determinato un cambio di paradigma nell'ambito della stilometria. Eder e Rybicki hanno implementato in R (linguaggio di programmazione statistica) la misura di Delta e alcune variazioni di questa misura, dando vita a una interfaccia molto semplice da utilizzare. Il poter utilizzare in maniera massiccia e semplificata questo tipo di misurazione ha consentito di testare l'efficacia di Delta anche con diversi generi testuali o testi di diversi periodi storici. I notevoli risultati e la facilità di implementazione di questo tipo di misura hanno spinto la ricerca a provare a migliorarla e perfezionarla. Tra le misure di distanza introdotte si trova la *Manhattan distance* ($\sum_i |u_i - v_i|$) che somma le distanze assolute tra le frequenze normalizzate di ciascuna parola nei due documenti, la "distanza Euclidea" ($\sqrt{\sum_i (u_i - v_i)^2}$) che calcola la distanza *straight line* tra i vettori, la *cosine Delta* che corrisponde all'angolo ϕ individuato da due vettori. $(\cos \phi = \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\|_2 \|\vec{v}\|_2} = \frac{\sum_i u_i v_i}{\sqrt{\sum_i u_i^2} \sqrt{\sum_i v_i^2}})$

Il vantaggio della distanza Delta coseno (*cosine Delta*) consiste nel fatto che un valore molto alto in un'unica dimensione, che risulta da una altissima frequenza di una parola in un testo rispetto agli altri che compongono il corpus, non ha un impatto decisivo

sull'angolo tra i due vettori; impatto che è invece molto più decisivo nel caso della distanza Manhattan o della distanza Euclidea (cfr. figura 3).

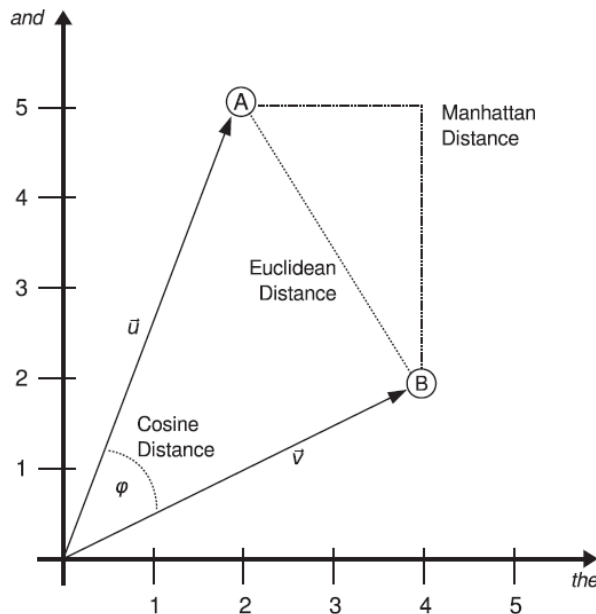


Figura 3: comparazione tra le diverse distanze vettoriali tra due esempi (A, B) descritti in uno spazio bidimensionale. (Evert et al. 2017)

I rapporti di distanza tra i diversi vettori sono poi rappresentati graficamente sulla base della matrice di distanza. Di norma si utilizza un raggruppamento gerarchico e agglomerativo (*clustering*) in cui i due testi che presentano una distanza reciproca minore sono uniti a formare un nodo. Il nodo appena creato si unirà agli altri nodi sulla base di un metodo di *linkage* a sua volta basato sulla vicinanza. Questo procedimento è ripetuto fino a unire tutti i documenti e tutti i nodi intermedi.

I tratti considerati nella misurazione della distanza tra due testi sono solitamente le parole e le sequenze di caratteri (*character n-grams*) più frequenti. La decisione riguardo al numero di parole più frequenti da utilizzare è, invece, dibattuta. I primi studi presentati da Burrows dimostrarono un'alta efficacia considerando le prime 100-150 parole più frequenti (in sostanza le parole funzione); studi più recenti hanno mostrato come un aumento del numero di parole considerate è direttamente proporzionale all'efficacia della misurazione della distanza, con un picco di efficacia intorno alle 1000-1500 parole più frequenti (Evert et al 2017).

L'efficacia di questo tipo di misurazioni è subordinata alla lunghezza dei testi presi in esame. Eder et al. Dimostrano infatti come solamente quei corpora a cui partecipano testi di lunghezza superiore alle 5000 parole garantiscono delle misurazioni veramente affidabili. Al contrario testi di dimensioni ridotte sono più suscettibili di incappare in errori di attribuzione; sotto le 5000 parole la percentuale di corrette attribuzioni si assesta al di sotto del 50%, per la lingua inglese.

Capitolo 2

2.1 Comunicazione facilitata (CF): introduzione.

La comunicazione facilitata (a cui ci si riferirà anche con la sigla CF) è una tecnica utilizzata da persone con un disturbo severo del linguaggio e con diversi tipi di disabilità (autismo, paralisi cerebrale infantile, sindrome di Down, sindrome dell'X fragile); può essere classificata all'interno delle strategie di Comunicazione Aumentativa Alternativa (CAA), seppure non tutti concordino con questa classificazione, (ISAAC).

Il tema della comunicazione facilitata è di notevole interesse dal momento che, a partire dagli anni '90, ha polarizzato le opinioni. Da una parte la maggior parte della comunità scientifica che si è espressa pubblicamente a sfavore di questa tecnica (ASHA, ISAAC, APA) dall'altra parte gli utilizzatori, i famigliari, i professionisti e alcuni ricercatori che ne hanno rivendicato l'efficacia e i vantaggi.

Il nucleo della discussione intorno alla comunicazione facilitata riguarda l'effettiva paternità degli scritti. I detrattori della CF affermano che i testi scritti in CF sono in realtà il prodotto dei facilitatori che intervengono nel processo di scrittura; al contrario, coloro che supportano la tecnica riconoscono negli scritti l'autorialità delle persone con disabilità.

La comunicazione facilitata prevede la presenza di un facilitatore che "facilita" il processo di scrittura della persona con disabilità fornendo un supporto emotivo, comunicativo, curriculare e fisico all'azione di digitazione alla tastiera (Bliken 1990); il supporto fisico può essere a livello del polso (facilitazione bassa) oppure a livello del gomito, dell'avambraccio o della spalla (facilitazione alta). In alcuni casi la facilitazione regredisce a una parte del corpo non direttamente coinvolta nel movimento di digitazione (schiena, gamba). Ciascun percorso di comunicazione facilitata ha come obiettivo il raggiungimento della totale autonomia comunicativa, traguardo che è stato raggiunto in un buon numero di casi (Ashby, C. 2018).

La comunicazione facilitata prende questo nome da Rosemary Crossley che nei primi anni '70 iniziò a utilizzare la tecnica presso il St. Nicholas Hospital di Melbourne. L'interesse da parte del professor Bliken (Syracuse University) al lavoro di Crossley determinò lo sviluppo e il propagarsi della CF negli Stati Uniti negli anni '90 e da lì in

diverse nazioni europee, tra cui la Francia, la Germania, il Regno Unito e l'Italia. È tuttavia interessante notare come tecniche di comunicazione molto simili alla CF si siano sviluppate in maniera spontanea e indipendente in diverse parti del mondo e in epoche diverse. Ci si riferisce alla *talking typewriter* descritta da F.K. Goodwin e R.P. Goodwin alla fine degli anni '60 (Megarry J. 1983), o alla tecnica molto simile alla CF scoperta da Carol Lee Berger all'inizio degli anni '80 (Lee Berger, 1992).

2.2 Comunicazione facilitata: lo stato dell'arte

La ricerca sulla comunicazione facilitata ha un punto d'inizio ben più recente rispetto al momento in cui la tecnica è stata scoperta e si è iniziato a utilizzarla. Negli anni '90, a seguito di alcuni casi giudiziari che vedevano coinvolte testimonianze redatte attraverso la comunicazione facilitata, si è cominciato a indagare circa l'efficacia della tecnica e la validità degli scritti (Wehrenfennig et al. 2008). Gli studi che sono stati condotti hanno seguito tre direttive principali: studi di passaggio del messaggio, studi con tracciamento dei movimenti oculari e studi di analisi dei testi linguistici. Generalmente la comunità scientifica ha individuato nel compito del passaggio del messaggio la prova principe per validare o meno la CF come strumento efficace di comunicazione.

2.2.1 Prove di passaggio del messaggio.

Le prove di passaggio del messaggio si preoccupano di indagare la capacità da parte del facilitato di scrivere (passare) informazioni non conosciute dal facilitatore. In generale la prova può essere strutturata in tre parti che individuano tre diverse condizioni sperimentali: nella prima condizione sperimentale al facilitato è chiesto di passare, attraverso la comunicazione facilitata, una informazione condivisa; ad esempio viene mostrata sia alla coppia facilitato-facilitatore una stessa immagine o suggerita una stessa parola, poi viene chiesto al facilitato di trascriverla. Nella seconda situazione sperimentale il compito rimane il medesimo ma l'informazione viene consegnata unicamente al facilitato mentre il facilitatore rimane non informato. La terza condizione sperimentale, che compare solamente in alcuni studi sul passaggio del messaggio, il facilitatore è malinformato. Viene, infatti, presentata al facilitatore una immagine o una

frase diversa rispetto a quella presentata al facilitato. Il compito del facilitato rimane invariato: trascrive la parola o la figura con la quale è stato "informato".

Frequentemente i risultati di queste prove hanno marcato una difficoltà da parte dei facilitati nel passare correttamente le informazioni non condivise con il facilitatore (situazione sperimentale 2 e situazione sperimentale 3). La facile replicabilità di questi risultati ha portato la comunità scientifica a considerare la comunicazione facilitata come inefficace e non affidabile (Wehrenfenning, Surian 2008, Saloviita, T. 2014). È tuttavia opportuno osservare come, nonostante la quasi totalità di questi studi sperimentalmente controllati abbia condotto a un giudizio negativo riguardo all'affidabilità della Comunicazione Facilitata, i risultati presentati mostrano una percentuale di successo, nelle prove di passaggio del messaggio nelle condizioni sperimentali 2 e 3, che varia dallo 0 al 30%, dimostrando la possibilità di effettivo successo nel passaggio di una informazione non condivisa.

Accanto a questi studi sperimentalmente controllati, si sono poi affiancati alcuni studi naturalistici. Secondo i sostenitori della comunicazione facilitata la causa dell'inefficacia delle prove di passaggio del messaggio è da cercare proprio nel contesto sperimentale, che può creare pressione e ansia nei facilitati, impedendo loro di rispondere adeguatamente al compito. Inoltre, come affermano Douglas Bliken e Donald Cardinal in *Communication Unbound* (1993), il problema potrebbe essere legato all'adeguatezza del test al contesto reale: una prova che identifica unicamente dei casi negativi, non è una prova che ha reale capacità esplicativa. Se voglio, infatti, testare le abilità matematiche di una classe di quinta elementare e il mio strumento di misura è l'assegnazione di un compito di algebra di 3^a superiore, otterrò sicuramente degli esiti negativi per tutti ma avrò fallito nella mia intenzione iniziale. In risposta agli studi controllati sono quindi stati costruiti degli studi naturalistici, in cui la pressione e i motivi di ansia e preoccupazione erano attenuati da allenamenti alla prova, familiarizzazione con il facilitatore e spiegazione degli obiettivi del compito. Inoltre, per questi studi, si sono selezionati utilizzatori esperti della CF. I risultati consegnati da questo tipo di indagine offrono dati significativi e favorevoli all'ipotesi della CF come tecnica efficace, dimostrando come, in una buona percentuale di occasioni, gli utilizzatori della CF erano stati in grado di passare una informazione non conosciuta dal facilitatore. Un altro studio

spesso trascurato nell'ambito del passaggio del messaggio è quello condotto dal professor Steven Laureys (*Twentieth Meeting of the European Neurological Society*, Giugno 2010) in cui si descrive il successo alla prova di passaggio del messaggio in un paziente con un severo trauma cranico.

Allo stesso modo i frequenti casi di passaggio spontaneo del messaggio, in normali situazioni di scambio comunicativo sono stati utilizzati come argomenti a favore della tecnica. Tuttavia, la comunità scientifica ha rigettato la validità di queste prove in quanto aneddotiche e non sperimentalmente controllabili; lo scarso valore scientifico di queste prove si scontra con la loro straordinaria efficacia nel rafforzare la convinzione del funzionamento della tecnica da parte delle persone coinvolte (familiari e professionisti): rappresentano quelle piccole certezze quotidiane che permettono alla CF di sopravvivere e diffondersi in assenza di valore scientifico.

2.2.2. Prove con il tracciamento oculare e la video-analisi.

Accanto alle prove di passaggio del messaggio (unico strumento accettato dalla comunità scientifica al fine di validare la tecnica) sono state utilizzate diverse altre metodologie per verificare la validità della tecnica o quantomeno rendere conto di ciò che succede durante il complesso processo della facilitazione. Alcuni autori hanno condotto una video analisi dettagliata dei movimenti degli utilizzatori della CF nelle diverse sessioni di scrittura, identificando per ciascuno uno schema personale e autonomo nel gesto rivolto alla tastiera: uno schema di movimento indipendente dal facilitatore ma facilitato-specifico (Grayson 2001). Allo stesso modo gli studi sul tracciamento oculare, svolti da Emerson, Grayson e Griffith (2012), precursori di uno studio più recente ma ancora in corso presso la Nottingham Trent University, hanno tentato di svelare i movimenti oculari degli utilizzatori della CF all'interno di più sessioni di scrittura. I risultati ottenuti mostrano come lo sguardo dei facilitati si soffermi con una non casuale frequenza e costanza sulla lettera, della tastiera, che di lì a poco sarà digitata; questi dati sembrano avallare l'ipotesi che effettivamente il facilitato sia autore degli scritti o che almeno sia in grado di prevedere con ottima frequenza l'intenzione del proprio facilitatore e orientare lo sguardo di conseguenza. Alcuni studiosi hanno criticato l'interpretazione che gli autori della ricerca hanno avanzato a

partire dai dati dello studio, affermando che l'orientamento dello sguardo verso la lettera sia una conseguenza del movimento del dito verso la stessa e non viceversa. Grayson e Emerson hanno tuttavia confutato questa spiegazione sostenendo che l'orientamento dello sguardo verso la tastiera fosse ben precedente al momento in cui la direzione del gesto di indicazione fosse inequivocabile.

2.2.3 Gli studi di analisi linguistica

Gli studi di analisi linguistica a partire dai testi scritti in CF lasciano per un istante da parte il complesso processo della facilitazione e del rapporto facilitato-facilitatore e si concentrano sui prodotti di questo processo: i testi e le conversazioni.

Già a prima vista i testi scritti in CF presentano dei caratteri non comuni o inusuali. Ci si accorge di una sintassi corretta ma marcata rispetto alla sintassi utilizzata nel linguaggio corrente. Allo stesso modo si osserva l'utilizzo di parole a bassa frequenza oltre al frequente impiego di neologismi.

Gli studi che sono stati condotti hanno affrontato il tema CF in una duplice prospettiva: a) trovare nei testi prove di autorialità da parte degli utilizzatori della CF e b) descrivere accuratamente le caratteristiche e le particolarità del linguaggio utilizzato durante la CF. Diversi sono stati gli approcci metodologici: alcuni studi (Bliken 1991, Niemi e Karma Lin 2002, Janzen-Wilde, Duchan e Higginbotham 1995, Zanobini 2001) hanno proposto un'analisi dei testi a sé stante, senza confronto con altro materiale testuale. Questi studi hanno identificato come prove della validità della tecnica la presenza nei testi di caratteristiche uniche e proprie del facilitato, costanti nel tempo e indipendenti dal facilitatore (Bliken), nella ricchezza e nella peculiarità del lessico e nell'uso di strutture linguistiche inusuali (Niemi e Karma), nella presenza di contenuti inaspettati, di errori atipici, nel passaggio di informazioni non condivise con il facilitatore, nelle autocorrezioni e nella sillabazione orale della parola contemporaneamente alla sua digitazione (Janzen-Wilde, Duchan, Higginbotham e Zanobini).

Vi è poi un insieme di studi che hanno invece affrontato la questione con modalità diverse: alcuni comparando i testi scritti dai facilitati con i testi scritti dai rispettivi facilitatori (Tuzzi 2009, Bernardi, Tuzzi 2011), altri confrontando le scelte di vocabolario degli utilizzatori della CF quando facilitati da assistenti diversi. Considerata l'importanza

di questi studi in relazione al lavoro di ricerca sviluppato da questa tesi, si proporrà una trattazione a parte nel capitolo 3.

La tabella 1 qui sotto riporta un riassunto dei principali studi che hanno coniugato l'analisi del testo scritto con il tema della comunicazione facilitata.

Studi fatti	Metodologia	Oggetto di indagine	Risultati
Blikem 1991	Senza confronto con altro materiale testuale per dimostrare l'unicità dei testi	<u>Partecipante:</u> 21 persone con diagnosi di autismo.	Caratteristiche uniche dei testi scritti dai facilitati, idiosincrasie in testi scritti con facilitatori diversi. Errori tipografici, errori fonetici e di spelling, stile di scrittura e espressioni inusuali
Niemi e Karna Lin 2002		<u>Partecipante:</u> un giovane adulto finlandese con una paralisi cerebrale di grado severo e danni cerebrali, con diagnosi di disabilità intellettiva <u>Materiale di indagine:</u> 1600 frasi scritte tra il 1992 e il 1996.	Prodotto lessicale altamente innovative Strutture linguistiche inusuali e creatività inaspettata degli scritti. Caratteristiche specifiche della scrittura, errori, ordine inusuale delle parole, elementi lessicali speciali
Janzen-Wilde, Duchan and Higginbotham 1995		<u>Partecipante:</u> bambino di 6 anni con disabilità intellettiva di grado moderato <u>Materiale di indagine:</u> scritti raccolti durante 12 sessioni.	Contenuti inaspettati Spelling atipico Sintagmi unici Evidenze aneddotiche che l'informazione scritta non era conosciuta dal facilitatore Spelling orale simultaneo alla scrittura Autocorrezioni
Tuzzi, Cemin, and Castagna 2004	Testi confrontati con il lessico standard.	<u>Partecipanti:</u> 11 ragazzi con autismo, che utilizzassero la CF da almeno 1 anno e che fossero facilitati a livello del gomito. <u>Materiale:</u> testi dei facilitati (24.449 tokens) testi dei facilitatori 23.386	Le persone con autismo utilizzano più forme (word type) rispetto ai loro facilitatori, sono più inclini all'utilizzo di parole inusuali. Questo supporta l'ipotesi secondo la quale i testi sono il frutto della produzione individuale della persona con autismo, non influenzata dal facilitatore.

Zanobini and Scopesi 2001		<p><u>Partecipanti:</u> bambino di 7 anni con autismo</p> <p><u>Materiali:</u> testi raccolti in 28 sessioni</p>	<p>Se comparati con i testi dei facilitatori, quelli dei facilitati risultavano più brevi e contenevano un maggior numero di parole originali e non comuni, spesso idiosincratiche. Alcune parole risultavano oltre la portata del lessico di un bambino di 7 anni. Il bambino utilizzava inoltre un linguaggio altamente creativo e metaforico.</p>
Scopesi, Zanobini and Cresci 2003		<p>Testi del bambino descritto dallo studio precedente, raccolti in 58 sessioni. Questi studi propongono un confronto statistico tra i testi prodotti attraverso la CF e i testi prodotti dai facilitatori. I ricercatori confrontano infine il lessico utilizzato dal partecipante con la frequenza delle parole generalmente utilizzate dai bambini in età scolare.</p>	
Progetto EASIEST (Berardi 2008, Bernardi Tuzzi 2011a, 2011b)	Testi dei facilitati confrontati con i testi dei facilitatori, e di un gruppo di controllo.	<p><u>Partecipanti gruppo 2:</u> 37 facilitati, 92 facilitatori.</p> <p><u>Partecipanti gruppo controlli:</u> 6 facilitati, 6 controlli senza autismo.</p> <p><u>Materiali:</u> testi dei facilitati e testi dei facilitatori.</p> <p><u>Metodologia:</u> Analisi statistica con <i>clustering method</i> e analisi qualitativa dei testi.</p>	<p>I risultati dimostrano una significativa differenza tra i due gruppi. Nel particolare i testi scritti dai facilitati presentavano un maggior ricchezza lessicale e una maggior complessità in termini di strutture lessicali, morfologiche e sintattiche; inoltre, i testi scritti dagli utilizzatori della CF presentavano una più alta frequenza di parole inusuali mentre i verbi e le parole funzione erano meno frequenti.</p>
Emerson, 2010	Confronto e analisi delle occorrenze lessicali dei facilitati quando assistiti da	<p><u>Partecipanti:</u> 10 facilitati e 10 facilitatori</p> <p><u>Materiali:</u> testi dei facilitati raccolti in modo tale che ciascuno dei 10 facilitati avesse</p>	<p>Presenza di parole tipiche ed esclusive dei facilitati (utilizzate con più di un facilitatore), e parole esclusive e tipiche dei facilitatori (utilizzate dai facilitati, esclusivamente con uno stesso facilitatore).</p>

	facilitatori diversi.	scritto con almeno due dei 10 facilitatori. <u>Metodi:</u> analisi quantitativa e qualitativa.	
--	-----------------------	---	--

Tabella 1: elenco degli studi di analisi linguistica sulla CF tratta da T. Saloviita 2016 e integrata.

2.2.3.1 La risposta dei detrattori della CF

I detrattori della CF non hanno considerato questi risultati come una prova dell'efficacia della CF essenzialmente per due motivi: a) l'unica prova di efficacia della CF può essere data dalle prove di passaggio del messaggio e qualsiasi altro risultato legato alla CF ma non supportato da evidenze di passaggio del messaggio è da considerare come irrilevante e non significativo; b) le conclusioni a cui pervengono gli studi sono l'esito di un ragionamento circolare: come lo stile che emerge dai testi scritti in CF è diverso dai testi dei facilitatori, così i testi scritti in CF mettono in risalto uno stile diverso dai testi scritti dai facilitati ma senza l'utilizzo della CF (T. Saloviita 2014). Inoltre, la presenza di uno stile inusuale potrebbe essere figlio del meccanismo stesso di scrittura, governato dal fenomeno ideomotorio. L'approdo a uno stile che si mostra estremamente diverso rispetto allo stile del facilitatore, ma anche a quello del facilitato scevro delle facilitazione potrebbe essere unicamente il risultato del metodo di scrittura nel quale: a) la prima lettera è scelta autonomamente dal facilitato, casualmente, b) a partire dalla digitazione delle lettere successive il facilitatore potrebbe essere inconsciamente colto dalla paura di arrivare a una parola senza senso e dirigere il movimento verso una sequenza di lettere tale da formare una parola di senso compiuto, c) quante più lettere sono state digitate casualmente minore sarà la probabilità di creare una parola a partire da questa sequenza, da qui si spiegano le parole a bassa frequenza e i neologismi.

2.3 Il dibattito sulla Comunicazione Facilitata.

Il tema della comunicazione facilitata rappresenta tuttora un motivo di acceso dibattito. L'eterogeneità dei risultati (da una parte gli studi in condizioni controllate a negare la validità della tecnica, dall'altra gli studi naturalistici a portare prove a favore della CF) ha

contribuito alla polarizzazione delle opinioni e a un arroccamento su posizioni mutualmente esclusive. Ma, mentre i sostenitori della CF hanno cercato di presentare sempre nuove tipologie di ricerca e sempre nuove argomentazioni, i detrattori della tecnica hanno spostato il confronto dal piano scientifico e dalla discussione dei dati a un piano più prosaico e di attacco, spesso con toni violenti e di scherno alla posizione opposta, togliendo ossigeno e campo a una mediazione verso posizioni intermedie.

La narrazione della CF veicolata dai principali mezzi di comunicazione ha contribuito a identificare gli scettici o detrattori della CF con i difensori di una scienza rigorosa e razionale e con esponenti di posizioni antiscientifiche e pseudoscientifiche coloro che nel corso degli anni hanno cercato di riaprire il dibattito sulla CF. Le due fazioni contrapposte, che si sono costituite, hanno attaccato e difeso la tecnica perdendo di vista il vero quesito scientifico riguardo alla CF. In questo senso hanno cercato di rispondere alla domanda “la CF funziona?” anziché alla domanda “può la CF essere utile per il comunicatore (X)?” (C. Ashby, *Comunico Ergo Sum*, Reggio Emilia, 2018).

L’argomento utilizzato più di frequente dai detrattori della tecnica è quello del fallimento alle prove di passaggio del messaggio, argomento che è spesso presentato come completamente a favore della tesi secondo cui la comunicazione facilitata è inefficace. Abbiamo visto, tuttavia, che questa argomentazione si presta a una considerazione più ampia. Gli studi condotti da Blikem e Cardinal hanno infatti dimostrato un incremento del successo alla prova del passaggio del messaggio quando preceduto da una fase di allenamento e contestualizzazione della prova. Inoltre, gli stessi studi in condizioni sperimentali controllate riportano un percentuale di successo alla prova del passaggio del messaggio fino al 30% in alcuni casi (Wehrenfenning, Surian 2008). Uno sguardo attento a questi dati dovrebbe sicuramente indurre cautela nei confronti della tecnica ma allo stesso tempo dovrebbe favorire un approfondimento maggiore verso quei casi e quelle situazioni in cui effettivamente la CF ha permesso di comunicare a una persona che prima d’allora non aveva potuto farlo, come sottolineato anche dagli stessi autori della revisione più recente riguardo alla letteratura sulla CF (Wehrenfenning, Surian 2008).

Un altro argomento spesso utilizzato dai detrattori della tecnica fa riferimento all’effetto ideomotorio, anche conosciuto come effetto Carpenter. L’effetto ideomotorio consiste

in una reazione inconscia generata dalla mente che determina un movimento sottile del corpo del quale il soggetto non è consapevole e tende ad attribuirlo a una forza esterna. Il movimento ideomotorio si articola dunque in due fasi: nella prima si genera la risposta motoria (*action production*), nella seconda la causa di tale azione è proiettata all'esterno (*action projection*) (Wegner, Sparrow, Fuller 2002). Lo stesso accadrebbe durante la CF. Il facilitatore coinvolto nel processo comunicativo produrrebbe dei micromovimenti involontari che indirizzerebbero il braccio del facilitato alla tastiera. A causa del meccanismo di proiezione dell'azione il risultato di tali movimenti di indicazione sarebbe attribuito a cause al di fuori di sé e di conseguenza come prodotto dei facilitati.

Una spiegazione di questo tipo, per quanto risulti efficace nel rendere conto dei risultati delle prove di passaggio del messaggio, non tiene conto di alcuni aspetti altrettanto fondamentali: primo fra tutti la negazione della partecipazione del facilitato al processo di scrittura: il comunicatore è relegato al ruolo di antenna dei micromovimenti del facilitatore senza alcuna possibilità di autodeterminazione. Inoltre, la percezione dei micromovimenti del facilitatore e la traduzione di questi in un movimento finalizzato alla tastiera richiederebbe un livello di attenzione focalizzata tale, soprattutto nei casi in cui la facilitazione è alta, da implicare delle competenze cognitive elevate nel facilitato, a tal punto da rendere plausibile l'autorialità degli stessi testi.

Duchan (1999) sostiene che il motivo alla base di questa disputa sia da riferire a una diversa concezione teorica sulla comunicazione da parte dei due schieramenti. Nello specifico i detrattori della CF considererebbero la comunicazione dal punto di vista informativo, dell'*information processing*, dando soprattutto peso alle abilità di codifica e decodifica nella trasmissione e ricezione del messaggio. Al contrario, i sostenitori della CF adotterebbero una prospettiva sulla comunicazione più di tipo sociale sottolineando maggiormente la componente collaborativa e di condivisione. L'integrazione di queste prospettive teoriche potrebbe garantire una miglior comprensione e un miglior inquadramento complessivo della tecnica aprendo a nuove interpretazioni circa l'effettiva natura della CF.

Quello che è certo e che è confermato dal dibattito mai sopito sulla Comunicazione Facilitata, è la necessità di sempre nuove proposte di ricerca sul tema insieme all'impegno da parte di entrambi gli schieramenti di abbandonare le proprie posizioni di

difesa e proseguire verso un sentiero comune che possa rendere conto delle diverse sfaccettature della CF.

Capitolo 3

3.1 L'analisi dei testi scritti con la CF

Gli studi che si sono occupati di investigare la comunicazione facilitata attraverso l'analisi dei testi sono numerosi e di diversa natura e, considerato il lavoro presentato in questa tesi, meritano una trattazione a parte e più dettagliata. Come già anticipato nel capitolo 2, gli studi che hanno proposto un'analisi dei testi scritti in CF hanno utilizzato strategie e metodologie diverse, ma hanno servito lo stesso obiettivo: dimostrare l'effettiva paternità degli scritti e in questo modo dare un giudizio circa la validità della tecnica. L'approccio diverso allo studio dei testi ha permesso, di volta in volta, di integrare le affermazioni circa l'autorialità dei testi con spunti di natura qualitativa in grado di far luce sui processi che potrebbero essere alla base del pensiero che si esprime attraverso la comunicazione facilitata. La capacità di poter abbinare alla ricerca sulla validità della tecnica l'indagine sui meccanismi che la sottendono rappresenta uno dei vantaggi maggiori dell'analisi linguistica. Il materiale fornito dai testi consente, infatti, una discussione multidimensionale sulla CF slegata dalla logica binaria del "funziona - non funziona" propria degli studi sperimentalmente controllati (passaggio del messaggio), lasciando spazio a osservazioni di carattere più qualitativo.

Va riconosciuto, tuttavia, che il paradigma scientifico odierno predilige informazioni supportate da dati quantitativi, controllati e verificabili e spesso le affermazioni di natura qualitativa, soprattutto nel caso della CF, vengono etichettate come tangenziali alla discussione scientifica. Per questo motivo, gran parte delle ricerche di analisi degli scritti in CF sono state poco considerate dalla comunità scientifica che, come abbiamo visto, trova nella prova di passaggio del messaggio l'unico strumento di indagine valido per la comunicazione facilitata. Pur non negando l'utilità della prova del passaggio del messaggio, soprattutto in prospettiva dell'autonomia comunicativa del facilitato e considerato che sono numerosi i casi di successo in questa prova, riteniamo che gli studi che riescono ad amalgamare informazioni di natura quantitativa ad altre di tipo qualitativo siano più efficaci nel caso complesso della comunicazione facilitata.

Oggetto di questo capitolo saranno gli studi condotti in seno al progetto EASIEST dell'Università di Padova e i risultati del lavoro di Emerson presso il centro di

Ravenswood; questi saranno confrontati e integrati con i risultati degli studi presentati nel capitolo 2.

3.2 Il progetto EASIEST

Il progetto EASIEST, diretto da Lorenzo Bernardi dell'Università di Padova, si propone, nel panorama degli studi di analisi linguistica sulla Comunicazione Facilitata, come il più completo dal momento che integra dati di natura quantitativa a dati di natura semi-quantitativa. I risultati presentati sono frutto dell'analisi su due diversi corpora:

- Corpus denominato *gruppo 2* (testi di 37 facilitati, con diagnosi di autismo, confrontati con i testi di 92 facilitatori).
- Corpus *casi-controlli* (6 bambini con autismo abbinati per sesso e per età a 6 bambini non autistici)

L'analisi per gruppi condotta su entrambi i corpora dimostrava uno scarto tra il linguaggio degli utilizzatori della CF e quello dei loro facilitatori o del gruppo di controllo. Lo scarto tra i due diversi stili di scrittura è raffigurato dalle figure 4 e 5.

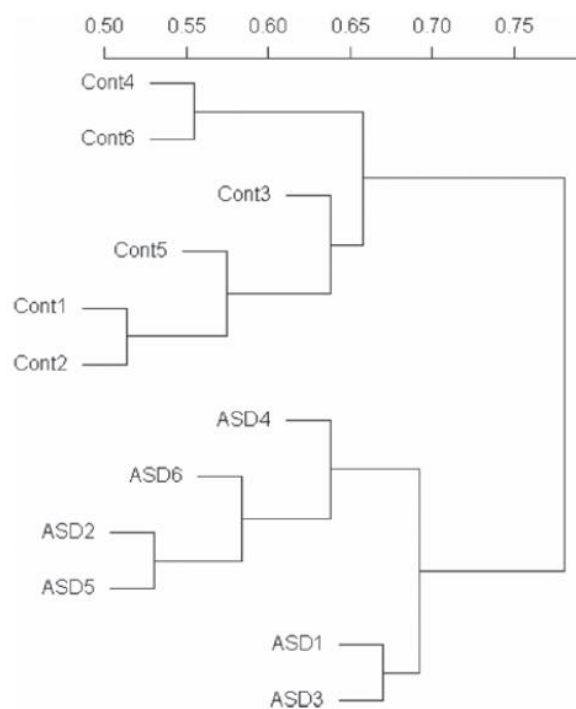


Figura 4: cluster analysis *Gruppo Casi-Controlli*. (Tuzzi 2011 a)

I due dendrogrammi delineano graficamente la distinzione compatta rispettivamente tra il gruppo controllo e il gruppo dei facilitati e tra il gruppo dei facilitatori e il gruppo dei facilitati. I testi scritti dai controlli, codificati con la sigla *cont*, e quelli prodotti dai facilitatori, codificati con la lettera *f*, si raggruppano tra loro, sono simili dal punto di vista stilistico e differiscono dal gruppo dei facilitati. Allo stesso modo i testi scritti dai facilitati presentano delle caratteristiche che li accomunano, sono raggruppati omogeneamente nel dendrogramma e differiscono dai testi dei facilitatori come un unico gruppo. Questo scarto linguistico dimostra la netta differenza sussistente tra i testi dei facilitati e quello dei rispettivi facilitatori o dei controlli. Una differenza che, secondo gli autori, supporta l'ipotesi per cui gli utilizzatori della CF siano effettivamente gli autori di ciò che è prodotto nel processo di scrittura.



Figura 5: cluster analysis Gruppo 2. (Tuzzi 2011 b).

La distanza stilistica è stata inoltre arricchita, sempre a livello quantitativo, dall'indagine sull'occorrenza nei testi di lemmi di uso non comune: i testi dei facilitati mostrano un maggior numero di lemmi a bassa frequenza d'uso rispetto ai testi dei facilitatori; nello specifico i testi dei facilitati presentano una percentuale di parole non comuni pari al 18% nel *gruppo 2* e del 11% nel *gruppo casi-controlli* a fronte del 10% dei facilitatori e del 4,9% dei controlli. L'ulteriore analisi sulle categorie grammaticali dimostra come i testi scritti in CF, se rapportati con i testi dei facilitatori, si caratterizzano per un utilizzo più elevato di aggettivi e avverbi parallelo a un sottoutilizzo di verbi e parole grammaticali. I testi dei facilitati, già a una prima indagine quantitativa, denunciano una maggior ricchezza e densità informativa unita a una minor attenzione verso gli aspetti grammaticali. Lo studio proposto da Tuzzi (2008) si sofferma, inoltre, sulla presenza di "segmenti ripetuti" esclusivamente nei testi delle persone con autismo e altre sequenze che comparivano unicamente nei testi dei facilitatori. L'analisi delle corrispondenze ribadisce come gli utilizzatori della CF abbiano un profilo lessicale tra loro simile e allo stesso tempo diverso dai controlli o dai loro facilitatori, come indicato dalla figura 6, in cui si può osservare come gli scriventi con autismo occupino uno stesso piano dello spazio mentre i controlli si posizionano in un piano diametralmente opposto.

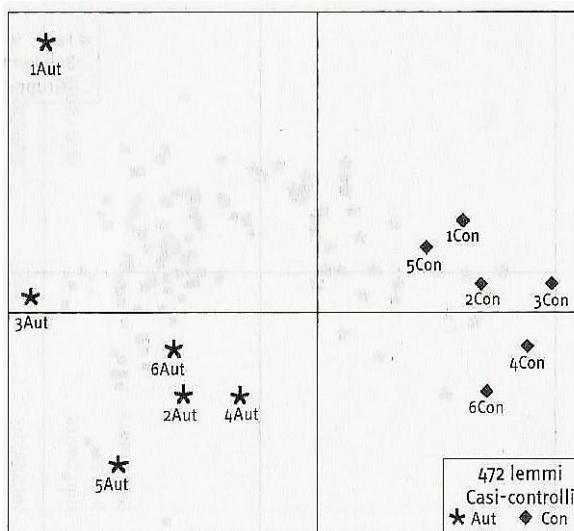


Figura 6: primo piano fattoriale dell'analisi delle corrispondenze della tabella di contingenza lessicale, del corpus casi controlli (Tuzzi, A. 2008 pag. 54)

Le indagini qualitative presentate all'interno dello studio di Bernardi (2008) hanno la capacità di addentrarsi all'interno del linguaggio di chi utilizza la CF specificandone peculiarità e tratti sia dal punto di vista delle scelte lessicali sia dal punto di vista della realizzazione sintattica. Il capitolo 3, a cura di Michele Cortelazzo, è dedicato

interamente alla descrizione della ricchezza lessicale del corpus EASIEST, mentre il capitolo 4, a cura di Chiara Di Benedetto, si occupa di affrontare lo stesso corpus da un punto di vista sintattico, retorico e semantico.

Soffermandosi sugli aspetti lessicali, l'analisi proposta conferma la diversa occorrenza di parole di uso non comune tra i facilitati rispetto ai controlli e ad ai facilitatori: inoltre l'utilizzo di parole a bassa frequenza si dimostra nella maggior parte dei casi adeguato al contesto; tra i lemmi non comuni utilizzati dai facilitati, la categoria grammaticale più rappresentata è quella degli aggettivi, seguita da quella degli avverbi per concludere con sostantivi e predicati. Addentrandosi nella descrizione lessicale del corpus, è sottolineata la presenza tra le parole di uso non comune di termini colti o appartenenti a livelli tecnici della lingua. È osservata l'alta frequenza di forme costruite con il prefisso negativo *in-*, oltre a un utilizzo massiccio degli avverbi in *-mente*. Il corpus *gruppo 2*, più ampio e frutto dello scambio dialogico tra facilitati e facilitatori, permette di notare con maggior frequenza l'occorrenza di parole italiane non esistenti, ma possibili secondo le regole di formazione delle parole che confermano la tendenza da parte degli utilizzatori della CF di creare, oltre agli avverbi in *-mente*, verbi a partire da aggettivi o nomi (*mommare* < *mamma*, *olgare* < *Olga*, *zozzare* < *zozzo*).

L'insieme di queste considerazioni permette di affermare come gli scriventi sappiano gestire e dominare una buona parte del lessico italiano e utilizzare produttivamente le regole di formazione delle parole. Inoltre, la presenza di regolarità (maggiore frequenza di aggettivi e avverbi, frequente ricorso a forme con il prefisso *in-* o ad avverbi in *-mente*) comuni a tutti gli utilizzatori della CF, dimostra come le differenze stilistiche rispetto ai facilitatori o all'uso comune non siano individuali o idiosincratice, ma siano in realtà un tratto comune di coloro che utilizzano la tecnica.

È a partire da queste considerazioni sul lessico che si sviluppano le riflessioni e le indagini contenute nel capitolo 4. La lingua utilizzata dai facilitati del corpus *gruppo 2* si mostra ricca di termini di registro elevato "utilizzati in chiave metaforica" (→ *coltri di molte paure*), di neologismi (→ *dolcificabile*) e di risemantizzazioni (parole adottate in un contesto diverso dal proprio → *puoi pulire l'urlo?*). Gli scriventi si dimostrano infatti estremamente espressivi e creativi come testimoniano:

- il frequente ricorso alla metafora (→ *mostro rumoroso che distrugge la calma dei miei pensieri*, oppure → *ho una giostra pirotecnica*)
- l'utilizzo di enfasi e iperbole (largo utilizzo del verbo *amare* e *odiare*)
- il gran numero di sinestesie (→ *parli teneramente, sonante fame, ritmi di luce*)
- le focalizzazioni semantiche attraverso iperbato (→ *lei mi deve cercare di limiti miei bene spiegare*), anastrofe (→ *molto ho paura del futuro*) e tmesi (→ *felice sono quando coraggio mi dai*).

Dal punto di vista sintattico, oltre alle numerosi dislocazioni e focalizzazioni semantiche, si rilevano strutture fortemente marcate; l'ordine degli elementi è spesso invertito e il verbo è posizionato alla fine; vi sono strutture a forte densità nominale; in alcune circostanze gli elementi sono semplicemente giustapposti senza legami grammaticali. Secondo l'autrice chi scrive con la CF sembra dare precedenza al sintagma nominale, in virtù del suo più alto valore informativo, a discapito, in alcuni casi, del verbo (→ *amicizie non esperienza ho, mi pronta se parli a ascoltare sembra*). Il forte valore informativo affidato dagli scriventi al sintagma nominale è poi confermato dal massiccio utilizzo di aggettivi; aggettivi che, frequentemente, sono disposti in posizioni inusuali. Non è raro, infatti, trovare costruzioni con l'aggettivo in posizione pre-nominale (più marcata rispetto a quella nominale, soprattutto se si tratta di costruzione lunghe → *mio rotto gretto perduto corpo*). Allo stesso modo non è raro il ricorso a participi presenti con funzione aggettivale, con il quale si conferma la tendenza a non considerare importante dal punto di vista informativo il verbo.

Il giudizio finale dell'autore del quarto capitolo è quello di una scrittura "a motore semantico" dalla quale emerge una forte urgenza comunicativa; la modalità con cui sono disposti gli elementi all'interno della frase denota una precedenza semantica che si manifesta con la preferenza degli aggettivi e degli avverbi alle categorie grammaticali e, di conseguenza, con il frequente ricorso alla "giustapposizione di isole di significato". Il capitolo si conclude con un confronto con gli studi internazionali fino ad allora condotti; molte delle riflessioni presentate all'interno di Bernardi (2008) e più generalmente del progetto EASIEST, trovano infatti conferma negli studi di Bliken (1991), Janzen- Wilde et al. (1995) e Niemi et al. (2002); in questi studi si riporta, analogamente, la presenza di

forme creative, il forte ricorso a parole inusuali e neologismi, l'ordine marcato dei costituenti con posposizione del verbo, il largo utilizzo di avverbi, o la rivitalizzazione di *(semi) frozen idioms*.

Da un lato, la conferma di caratteristiche comuni, indipendenti dalla lingua specifica, coopera all'idea che i processi sottesi alla CF siano degli universali. Dall'altro lato questa considerazione può difficilmente aggiungere argomenti a favore dell'autorialità degli scriventi o dell'influenza dei facilitatori; come testimonia l'articolo di T. Saloviita (2016) la presenza di tratti comuni e peculiari potrebbe essere il frutto di una tendenza universale dei facilitatori a influenzare gli scritti in tale direzione, a maggior ragione se l'influenzamento ha origini inconscie. Allo stesso modo l'universalità stilistica denotata dai testi in CF potrebbe dipendere dalle strutture e dagli schemi di pensiero delle persone con disabilità comunicativa uniti al processo di facilitazione. Questi studi si dimostrano quindi non conclusivi del dibattito sulla CF: lo studio del linguaggio degli scriventi in termini generali è utile nella descrizione di ciò che accade durante il processo di facilitazione ma si dimostra meno efficace nella definizione di autorialità individuale dei testi. Come sostiene Saloviita (2016) dimostrare una differenza tra lo stile dei facilitati e quello dei rispettivi facilitatori o dei controlli non equivale a provare l'indipendenza degli scriventi ma semplicemente l'esistenza di un comportamento linguistico differente durante il processo di facilitazione. Inoltre, continua Saloviita, affermare sulla base di queste differenze l'autorialità dei facilitati significa commettere un corto circuito logico, una petizione di principio: infatti, come i testi dei facilitati sono diversi da quelli dei loro facilitatori allo stesso modo sono lontani dai testi o dalle modalità comunicative degli scriventi quando non sono facilitati.

3.3 Indicatori di autorialità e di influenzamento.

Per poter prendere parte alla discussione sull'efficacia della tecnica è dunque necessario che gli studi di analisi del linguaggio si concentrino su quei tratti linguistici in grado di differenziare un facilitato da un altro facilitato, piuttosto che da un facilitatore; dimostrare una coerenza interna allo scrivente, indipendente dal facilitatore, potrebbe rappresentare un forte argomento a sostegno dell'autorialità degli scritti e di conseguenza della funzionalità della CF. All'interno di questa prospettiva di ricerca si

inserisce lo studio di Emerson (2010), condotto all'interno del centro di Ravenswood (UK). La domanda generale che ha guidato l'analisi di Emerson è stata la seguente: *la CF può aiutare alcune persone con una forte disabilità comunicativa a comunicare in un modo che non sarebbe possibile senza supporto, anche se parte della loro comunicazione è soggetta all'influenza del facilitatore?* Per rispondere a questo quesito Emerson propone una analisi del vocabolario di facilitatori e facilitati: la presenza di parole esclusive di un facilitato e che compaiano con più di un facilitatore rappresenta un chiaro indizio di autorialità; viceversa, l'individuazione di parole che occorrono unicamente con lo stesso facilitatore diventa una prova di un possibile influenzamento. A ciascuna occorrenza che risponda a queste caratteristiche è assegnato un livello:

- livello 1: *prova forte di uso individuale di una parola* → parole a bassa frequenza d'uso che occorrono in uno stesso facilitato con diversi facilitatori
- livello 2: *prova moderatamente forte di uso individuale di una parola* → parole a frequenza d'uso più alta che occorrono in uno stesso facilitato con più di un facilitatore
- livello 3: *prova moderata di uso individuale di una parola* → parole ad alta frequenza d'uso che occorrono in uno stesso facilitato con più di un facilitatore
- livello 4: *uso individuale di una parola comune a più partecipanti* → parole che sono utilizzate individualmente da più di un partecipante, ad esempio parole che ricorrono in un facilitato con diversi facilitatori ma che compaiono anche nello stesso facilitatore con utilizzatori della CF diversi.
- livello 5: nessuna evidenza di utilizzo individuale di una parola.

Ciascuna valutazione riguardo al livello a cui attribuire la parola è stata controllata da un esaminatore esterno.

Da questo tipo di analisi, condotta su un campione di dieci facilitati e sette facilitatori, sono emerse sia prove a favore dell'autorialità degli utilizzatori della tecnica sia prove a favore dell'ipotesi dell'influenzamento da parte del facilitatore.

Per ciascun partecipante sono state trovate delle parole utilizzate in maniera individuale, tuttavia per alcuni facilitati e un facilitatore il numero di parole individuato non è risultato sufficiente per affermarne rispettivamente l'autorialità o

l'influenzamento. Cinque utilizzatori della tecnica invece hanno mostrato forti prove di essere autori di almeno parte delle loro comunicazioni; questo risultato è secondo l'autrice sorprendente considerato il grave livello di disabilità comunicativa dei partecipanti e tenuto conto del fatto che gli studi precedenti sulla CF avevano al massimo dimostrato l'efficacia della tecnica per un solo utilizzatore della tecnica, che spesso veniva considerato come una eccezione. Un altro aspetto degno di menzione è quello relativo all'influenzamento da parte del facilitatore; infatti, gli stessi utilizzatori che mostrano tracce di individualità nell'utilizzo del vocabolario sembrano essere al contempo suscettibili di influenzamento da parte dei facilitatori. L'indipendenza maggiore che si verifica in alcuni facilitati non sembra inoltre dipendere né dall'abilità del facilitatore, né dall'esperienza nell'utilizzo della CF né dal tipo di diagnosi ricevuta. Lo studio non si preoccupa di approfondire la questione dell'origine dell'influenzamento, tuttavia avanza alcune ipotesi sul fenomeno all'interno del gruppo oggetto della ricerca; secondo Emerson un facilitatore potrebbe aver influenzato lo scrivente in questi modi:

- *diretta manipolazione della mano del facilitato* → sebbene in molti studi l'influenzamento è descritto come inconscio e non volontario, capita che la lettera o la parola che sarà di lì a poco digitata possa essere intuita in anticipo; alcuni micromovimenti o esitazioni da parte del facilitatore conseguenti all'intuizione della parola potrebbero influenzare lo scrivente nella scelta del tasto da digitare. Alcuni facilitati potrebbero addirittura cercare questi segnali per avere una conferma della correttezza o della bontà di ciò che stanno scrivendo.
- *influenzamento dovuto a una familiarizzazione con un linguaggio tipico* → possibile nel caso in cui i facilitatori abbiano accesso a tutte le scritture dei facilitati; è tuttavia poco probabile che i facilitatori dominino tutte le occorrenze di tutti i facilitati con cui interagiscono, soprattutto quelle meno frequenti.
- *influenza della personalità del facilitatore* → il modo di comunicare di ciascuno è spesso determinato dalla personalità e dalla reazione al contesto in cui si trova. Il facilitatore P descritto da Emerson, ad esempio, si presenta come molto estroverso e maggiormente incline ad affrontare discussioni sulla sessualità.

Questo, unito a un aspetto particolarmente femminile, può aver influenzato i facilitati in molte conversazioni; è pertanto difficoltoso valutare l'origine delle affermazioni dei facilitati, facilitati da P, sui temi riguardanti la sessualità e la relazione tra uomo e donna.

Dal punto di vista della "quantità di influenzamento", il numero di parole utilizzate solo da coloro che scrivono con uno stesso facilitatore è risultato piuttosto basso soprattutto in relazione ai precedenti studi quantitativi (cfr. cap. 2). Secondo Emerson questa tendenza è dovuta alla minor pressione esercitata sui facilitatori dal contesto naturalistico, in cui sono stati raccolti gli scritti, rispetto al "setting" sperimentale; una spiegazione che può essere accomunata a quella proposta da Bliken per giustificare i fallimenti nelle prove di passaggio del messaggio (cfr. cap. 2). L'influenzamento, dunque, potrebbe essere un effetto collaterale della comunicazione facilitata ma non il suo motore. La partecipazione alla scrittura da parte dei facilitatori, come si è visto, è svincolata dal grado di indipendenza del facilitato: per questo influenzamento e autorialità non vanno considerati come due aspetti che si escludono a vicenda ma possono essere compresenti all'interno del processo di facilitazione. La presenza, inoltre, all'interno del gruppo di Ravenswood di un facilitatore che sembra non esercitare alcuna influenza sulla comunicazione delle persone che assiste sembra corroborare questa ipotesi. È infatti possibile, sostiene Emerson, che alcuni facilitatori possano essere più inclini all'influenzamento rispetto ad altri, sulla base di aspetti legati alla personalità o alla relazione costruita con l'utilizzatore della CF.

Il quadro presentato da questi risultati non è in grado, conclude l'autrice, di produrre una conclusione ferma e certa sul grado in cui un facilitato è autore dei suoi testi. A questa affermazione si associano due precise conseguenze: (i) la tecnica va maneggiata con cautela poiché il rischio di influenzamento è reale; i facilitatori devono essere consapevoli di questo rischio e devono essere formati per gestirlo. (ii) La CF sembra essere effettivamente in grado di aumentare le possibilità comunicative di persone che altrimenti non potrebbero dimostrare alcuna intenzione o competenza. Per questo motivo, proibire definitivamente l'utilizzo della CF significherebbe privare alcune persone del proprio diritto alla comunicazione.

La risposta della critica scientifica a questo lavoro è giunta attraverso il lavoro di Bober (2010). In questa revisione bibliografica di tutti gli studi condotti fino ad allora sulla comunicazione facilitata, soffermandosi sullo studio di Emerson l'autrice afferma che utilizzare come indicatore di autorialità semplicemente un parametro del testo, in questo caso l'occorrenza di alcune parole, non è sufficiente, dal momento che è necessario considerare più di un parametro (Kniffka, 1996) e che pertanto, Emerson ha semplicemente finto di fare una analisi linguistica sul corpus del gruppo di Ravenswood. Confermare inoltre l'effettivo influenzamento del facilitatore, continua Bober, equivale ad avvalorare il giudizio di coloro che ritengono la tecnica dannosa e pericolosa e, di conseguenza contribuire al discredito della CF.

Se è vero che un unico parametro linguistico non può essere sufficiente a dirimere la questione dell'autorialità di un gruppo di scritti (Emerson infatti non propone un giudizio conclusivo sul suo corpus, ammesso che sia possibile), è allo stesso tempo utile per avanzare delle ipotesi interpretative sulla comunicazione facilitata. In questo senso, dimostrare l'esistenza di parole utilizzate individualmente da un facilitato può rappresentare un indizio di tracce di indipendenza da parte di chi scrive con la CF e di conseguenza di un aumento delle sue possibilità comunicative. La differenza di interpretazione dei risultati è giustificata da un differente quesito-guida della ricerca: da una parte chi si domanda se la CF sia del tutto efficace, dall'altra chi si domanda se la CF può essere utile al miglioramento della comunicazione e quindi delle relazioni di una persona.

La portata innovativa della ricerca di Emerson si misura dalla seconda osservazione di Bober. Affermare e confermare il possibile influenzamento da parte del facilitatore all'interno di uno studio che porta anche elementi a favore della comunicazione facilitata, significa cambiare sia la percezione sulla CF sia il paradigma di approccio al dibattito sulla stessa (tradizionalmente impostato sullo schema "funziona-non funziona"). Come sostenuto anche da Duchan e da Zanobini e dimostrato dalla stessa Emerson, la comunicazione facilitata è un processo che vede coinvolti attivamente due soggetti; i testi che risultano dalla facilitazione sono frutto di una costruzione condivisa e di una collaborazione che si sviluppano all'interno dell'interazione facilitato-

facilitatore. Di conseguenza, è naturale rinvenire tracce di influenzamento del facilitatore accanto a tratti che indicano l'indipendenza e l'autorialità degli scriventi.

È chiaro che una definizione della CF di questo tipo non possa soddisfare coloro che chiedono a uno strumento di comunicazione aumentativa alternativa (CAA) la massima affidabilità e in termini di autonomia e in termini di espressione effettiva del messaggio che l'emittente vuole comunicare. Se si esce tuttavia dalla prospettiva che intende la comunicazione esclusivamente come scambio di contenuti informativi e si considera invece la comunicazione come un aspetto sociale e di scambio relazionale (Duchan, 1991), anche la comunicazione facilitata, così descritta, assume una differente connotazione.

Capitolo 4

4.1 Un'analisi semi-qualitativa

La tesi qui presentata si propone come un prolungamento dei due principali studi che intorno alla metà degli anni 2000 hanno affrontato l'argomento *Comunicazione Facilitata* attraverso l'analisi linguistica dei testi: *Il delta dei significati* (Bernardi, 2008) e la tesi di dottorato di Anne Emerson (2010) entrambi dettagliatamente descritti nel capitolo precedente (vedi capitolo 3). L'approccio semi-qualitativo ai testi scritti in CF è ciò che accomuna maggiormente le due ricerche, sebbene le analisi condotte si differenzino negli obiettivi. Lo studio diretto da Bernardi (2008), infatti, individua quegli aspetti più caratteristici degli scritti dei soggetti con autismo, sottolineandone lo scarto dallo stile della popolazione generale non autistica. Al contrario, lo studio di Emerson si propone di rintracciare delle prove di autorialità da parte degli utilizzatori della tecnica, attraverso l'analisi delle corrispondenze lessicali.

Il progetto qui presentato intende applicare le strategie di analisi adottate dagli studi precedenti a un corpus completamente nuovo, recuperato presso uno dei centri dell'associazione "Vi Comunico Che Penso": l'obiettivo è quello di aggiungere un ulteriore contributo alla discussione sull'autorialità dei testi scritti con la Comunicazione Facilitata attraverso la conferma dei risultati trovati e delle ipotesi prodotte. A una prima parte dedicata all'analisi semi-qualitativa, seguirà una seconda che, sfruttando alcune delle più moderne strategie di analisi quantitativa dei testi, cercherà, a propria volta, di affrontare il tema dell'attribuzione dei testi scritti in CF.

4.1.1 Materiali.

4.1.1.1 Il corpus 2018

Il corpus, che denomineremo *corpus 2018*, utilizzato per l'analisi semi-qualitativa è composto da 28 testi; ciascun testo è dato dal raggruppamento di tutte le conversazioni prodotte da un utilizzatore della CF con un facilitatore, in un periodo che va al massimo dal 2004 al 2018 (sono molto rari i casi in cui un facilitato abbia scritto consecutivamente per 14 anni con lo stesso facilitatore). La scelta di utilizzare questo tipo di testi nasce dalla necessità di reperire in tempi brevi un corpus di ampie dimensioni, in grado di

soddisfare le esigenze di dimensione sia dell'analisi semi-qualitativa sia, e soprattutto, dell'analisi quantitativa. La decisione di non procedere con una raccolta controllata dei testi, a causa della dispendiosità in termini di fatica e tempo per i facilitati e i facilitatori, ha portato con sé dei vantaggi e alcuni svantaggi. Tra gli aspetti positivi vi è la possibilità di utilizzare dei testi genuini e spontanei, frutto del conversare settimanale di ciascun utilizzatore della CF con i propri facilitatori, libero da qualsiasi possibile bias determinato dalla pressione della ricerca o dalla difesa della propria posizione sull'argomento (già di per sé spinoso). I testi che compongono il corpus ritraggono la Comunicazione Facilitata normale e quotidiana del centro; rappresentano un oggetto puro di indagine. Dall'altro lato gli aspetti negativi riguardano l'omogeneità e l'equilibrio del corpus: non vi è infatti nel corpus una omogeneità nella distribuzione dei facilitatori diversi per ciascun facilitato; vi sono facilitati che hanno scritto con buona frequenza con 4 facilitatori diversi e altri che invece hanno scritto la maggior parte dei testi con un unico facilitatore di riferimento. Allo stesso modo il disequilibrio all'interno del corpus si manifesta a livello della dimensione dei testi: vi sono alcune coppie facilitato-facilitatore che contano testi di poco più di 5000 parole mentre altre coppie presentano testi di 50000 parole. Questi disequilibri e disomogeneità, che una raccolta controllata dei dati potrebbe evitare, ostacolano, in parte, il confronto dei testi, complicando soprattutto il lavoro di analisi quantitativa.

La selezione dei testi ha seguito principalmente due criteri: la dimensione dei testi e la distribuzione del numero di facilitatori diversi per ciascun facilitato. Riferendoci al primo criterio sono stati esclusi tutti quei testi che contassero meno di 5000 parole, secondo le istruzioni di Eder (2015). Per quanto riguarda il secondo criterio, sono stati selezionati i testi di ciascun utilizzatore della CF che avesse scritto con almeno 2 facilitatori diversi.

4.1.1.2 Partecipanti

Il risultato della creazione del *corpus 2018* ha consentito di isolare da un gruppo di 112 testi i 28 testi del nostro corpus, in cui sono compresi 10 utilizzatori della tecnica, con diverse patologie e 7 facilitatori diversi, come descritto dalla tabella di seguito (tabella 2). I facilitati, dunque, non sono stati selezionati in base alla diagnosi; tra il gruppo di 10 facilitati, 7 femmine e 3 maschi, troviamo infatti alcuni partecipanti con Sindrome di

Down, altri con paralisi cerebrale infantile e altri con diagnosi di autismo. Di seguito la tabella 2 illustra le caratteristiche principali dei partecipanti (facilitati e facilitatori) e del corpus in generale.

Facilitato	Cod. Facilitato/facilitatore	Types	Tokens
A (F, 45anni sindrome di Down)	A_1	2727	17710
	A_2	1430	5007
	A_3	2834	18091
B (F, 42 anni, rara alterazione cromosomica)	B_1	3494	28862
	B_2	1721	6438
	B_4	1454	6065
C (F, 32 anni, sclerosi tuberosa)	C_1	3035	16229
	C_2	6113	47566
	C_4	1940	6977
	C_8	3913	25745
D (F,48 anni tetraparesi spastica)	D_1	3300	21838
	D_2	2560	11598
E (M,54 anni, insufficienza mentale di grado elevato)	E_1	3892	22934
	E_2	3566	15427
	E_5	1229	5100
	E_7	4962	27651
	E_8	1700	6395
G (M, 34 anni, encefalopatia PV)	G_1	2177	12955
	G_3	1663	7750
I (M,50 anni, ritardo mentale grave)	I_1	5738	64202
	I_5	1321	5464
N (F,21 anni, ritardo mentale grave)	N_1	3531	38549
	N_3	1172	5021
O (F, 26 anni, tetraparesi spastica)	O_1	2061	14578
	O_2	2174	9675
	O_3	3273	24655
R (F, 37 anni, tetraparesi spastica)	R_1	1492	5707
	R_8	1470	5009
TOTALE	28	20021	483198

Tabella 2: descrizione del corpus utilizzato per l'analisi semi-qualitativa.

4.1.2 Il corpus- uno sguardo d'insieme.

Si consideri il *corpus 2018* nella sua totalità, come un unico grande testo, spostando per un istante il focus dall'indagine sull'autorialità e concentrandosi sugli aspetti globali della scrittura degli utilizzatori della tecnica. Il corpus è composto da 483198 occorrenze

(*tokens*) che risultano dall'utilizzo di 20021 parole (*types*). Come infatti già sottolineato dagli studi raccolti in Bernardi (2008), la produzione linguistica, esito di un processo di comunicazione facilitata, presenta delle particolarità che la smarcano dalla comunicazione neurotipica; caratteristiche specifiche che sembrano non avere nulla di idiosincratico o individuale (Di Benedetto, C. 2008) ma sembrano essere un tratto comune alla popolazione con autismo, che scrive con la CF, come confermano gli studi condotti in altre lingue. Le peculiarità del linguaggio della comunicazione facilitata, approfondite in Cortelazzo (2008) e Di Benedetto (2008), semplificando, si possono così riassumere:

- la presenza di un lessico vario e che non si ripete, con la presenza di numerosi Hapax
- la ricorrenza di parole non comuni della lingua italiana
- l'utilizzo di parole comuni isolate dal loro contesto di occorrenza naturale
- la presenza di parole non esistenti nella lingua italiana ma possibili
- l'utilizzo frequente di forme con il prefisso /in-/
- l'uso intensificato di avverbi in /-mente/
- la presenza di strutture sintattiche marcate accanto a strutture sintattiche non marcate
- forte incidenza di dislocazioni a sinistra.

Prima di riferirci alle caratteristiche del corpus oggetto di questa ricerca è opportuno fare alcune precisazioni riguardo alla metodologia utilizzata. Lo studio diretto da Bernardi (2008) fondava le sue affermazioni, prima di tutto, sulla comparazione e sul confronto dei testi scritti dalle persone con autismo in CF con i testi dei rispettivi facilitatori; in questo senso ogni osservazione è basata sulla constatazione sia qualitativa che quantitativa di uno scarto tra lo stile di scrittura degli utilizzatori della tecnica rispetto a quello dei loro assistenti. L'analisi condotta sul corpus preso in esame in questa tesi, invece, è considerarsi in una prospettiva confermativa dei risultati presentati in Bernardi (2008); per tanto il confronto sarà da intendersi non con le produzioni dei facilitatori ma con le produzioni dei facilitati del progetto EASIEST. La seconda precisazione riguarda invece i partecipanti allo studio. Mentre all'interno del

progetto EASIEST i facilitati partecipanti allo studio erano persone con autismo, gli scriventi del *corpus 2018* non sono stati selezionati in base alla diagnosi, ma in funzione della quantità dei loro scritti. Queste due specificazioni permettono di delineare la portata dei risultati dell'analisi semi-qualitativa sul *corpus 2018*. Da un lato, infatti, consentono di confermare le peculiarità della tecnica attraverso l'analisi di testi scritti in un diverso centro consentendo inoltre una osservazione della variazione e della modificazione della tecnica nel corso degli anni. Allo stesso modo, se confermati gli stessi risultati, le successive e ulteriori analisi condotte sul *corpus 2018* potranno verosimilmente riferirsi anche ai risultati delle precedenti ricerche, offrendo un quadro completo sulla CF. Dall'altro lato consentono di valutare se gli aspetti peculiari osservati dagli studi precedenti siano propri della popolazione con autismo o se, viceversa, siano caratteristici della tecnica utilizzata.

I testi del *corpus 2018* presentano molti tratti in comune con i testi del progetto EASIEST, con alcune, sostanziali differenze.

Così come già osservato da Cortelazzo (2008), anche nel caso del *corpus 2018* ci si trova di fronte a un lessico poco ripetitivo e ricco di *Hapax*. Delle 9613 forme diverse che compongono il corpus ben 2981 sono *Hapax*, a conferma di una forte creatività e un'ampia disponibilità lessicale. Anche per quanto riguarda i lemmi di uso non comune il *corpus 2018* si comporta in maniera simile ai testi utilizzati in Bernardi (2008). L'analisi rispetto alla frequenza dei diversi lemmi, condotta con il software Col FIS mostra come siano numerose le forme non comuni all'interno del corpus; si tratta, tra le altre, delle seguenti parole: *abbarbicato, annaspate, aggrappare, appesantito, arreso, astioso, astruso, autistico, bacilli, bagordi, baldoria, balia, bambinesco, benvenuto, biochimica, bisbetico, brumoso, burrascoso, carrozzone, chimera, coetaneo, colloquiare, coltre, condiviso, cronicizzare, cruccio, delittuoso, deriso, delucidare, demordere, disabilità, disagiato, dolente, dotto, elettrizzato esterrefatto, frastuono, ferito, gioiosamente gioioso, greve, guarito, illogico, impertinente, impietosire, ingiusto, invalidabile, nervosamente, nullità, operoso, ottuso, pioniere, rissoso, rallentato, salpare, scettico, turbare, udito, vortice, zoppicare, zoppicante...*

Così tra le parole di uso non comune, come in quelle di uso comune, si conferma la presenza di numerose forme introdotte dal prefisso *in-* negativo tra le quali, *Illibati, illimitata, illimitatamente, illogica, illogicamente, immancabilmente, immobili, immobilizzata, impalpabile, imparagonabile, impaziente, impensabili, impercettibilmente, imperdonabile, impermeabilizzare, implacabile, impossibile, impraticabile, imprevedibile, inaccettabili, inadeguata, inaffidabili, inagibile, inarrestabile, incivile, incolpevole, incomprendibile, incomprendione, inconcepibile, inconfutabile, incontenibile, incontrollabile, incredibile, incredibilmente, incurante, indescrivibile, indimenticabile, indiscutibile, indispensabile, indisponibile, indomabile, ineccepibile, inesauribile, inesorabile, inestimabile, inevitabile, infallibile, informe, ingiuste, ingiustificata, inguaribile, inimmaginabile, inopportuno, insane, insapore, insensato, insensibile, insicura, insignificante, insoddisfatta, insoddisfazione, insolita, insopportabile, insostenibile, insostituibile, inspiegabile, instabilità, instancabile, insufficiente, insuperabile, interminabile, interminate, intoccabile, intollerante, intrattabile, inusuale, inutile, inutilmente, invalicabile, invincibilità, invivibile, involontario, irraggiungibile, irrespirabile, irresponsabile, irrilevante, irrimediabili, irritabile; accanto a queste, una folta occorrenza (286 forme diverse) di avverbi in -mente, elencati di seguito: *veramente, finalmente, sicuramente, seriamente, assolutamente, fortemente, certamente, lentamente, serenamente, facilmente, ultimamente, realmente, decisamente, fisicamente, gioiosamente, tranquillamente, felicemente, grandemente, semplicemente, pienamente, velocemente, esattamente, totalmente, fermamente, diversamente, particolarmente, potentemente, talmente, continuamente, improvvisamente, ottimamente, profondamente, ugualmente, opportunamente, piacevolmente, caramente, estremamente, ovviamente, completamente, definitivamente, difficilmente, effettivamente, nuovamente, positivamente, sinceramente, unicamente, chiaramente, diligentemente, intensamente, opprimente, personalmente, politicamente, praticamente, precisamente, raramente, solamente, umilmente, vivamente, attivamente, delicatamente, fremente, generalmente, liberamente, perfettamente, probabilmente, dolcemente, duramente, erroneamente, infinitamente, moralmente, pazientemente, pesantemente, saggiamente, tremendamente, ufficialmente, attualmente, bruscamente,**

confusamente, correttamente, curiosamente, direttamente, dolorosamente, doppiamente, faticosamente, giustamente, inizialmente, inutilmente, letteralmente, lietamente, minuziosamente, miseramente, normalmente, principalmente, prontamente, rabbiosamente, ragionevolmente, silenziosamente, specialmente, terribilmente, tristemente, violentemente, affettuosamente, apparentemente, astronomicamente, attentamente, automaticamente, brillantemente, calorosamente, civilmente, comunemente, degnamente, discretamente, fantasticamente, frettolosamente, furbescamente, gravemente, illogicamente, immediatamente, interiormente, internamente, leggermente, lungamente, maggiormente, mentalmente, meravigliosamente, nettamente, ottimisticamente, passivamente, perennemente, pigramente, placidamente, popolarmente, possibilmente, pubblicamente, rapidamente, realisticamente, severamente, socialmente, solitamente, splendidamente, stancamente, stupidamente, teneramente, tradizionalmente, umanamente, vertiginosamente, accuratamente, adeguatamente, allegramente, altamente, amaramente, amichevolmente, ampiamente, appositamente, ardentemente, artificialmente, artisticamente, aspramente, assonnatamente, astutamente, autonomamente, avidamente, banalmente, beatamente, bonariamente, brevemente, casualmente, comodamente, concretamente, consapevolmente, contemporaneamente, coraggiosamente, costantemente, culturalmente, delittuosamente, eccessivamente, eccezionalmente, economicamente, egregiamente, elegantemente, energicamente, enormemente, equamente, espressamente, essenzialmente, esteticamente, eternamente, eticamente, fastidiosamente, fedelmente, fiduciosamente, fieramente, fondamentalmente, fortunatamente, forzatamente, freddamente, freneticamente, frequentemente, furbamente, furiosamente, gaiamente, giuridicamente, goffamente, gradatamente, gradualmente, graziosamente, illimitatamente, immancabilmente, immensamente, impercettibilmente, incessantemente, incredibilmente, indirettamente, individualmente, ininterrottamente, intelligentemente, intimamente, ironicamente, largamente, lateralmente, legalmente, lestamente, lievemente, magnificamente, maleducatamente, massimamente, mestamente, minimamente, miracolosamente, misteriosamente, monotonomamente, narcisamente, naturalmente, necessariamente, negativamente, nervosamente, ossequiamente, ossessivamente, pacificamente,

parzialmente, penosamente, permalosamente, polemicamente, poveramente, precocemente, premurosamente, propriamente, prossimamente, psicologicamente, puramente, razionalmente, recentemente, reciprocamente, regolarmente, relativamente, responsabilmente, rigorosamente, rudemente, rumorosamente, saldamente, saltuariamente, saporitamente, scarsamente, scioccamente, sensibilmente, signorilmente, simultaneamente, sommariamente, spiritualmente, spontaneamente, sportivamente, stupendamente, successivamente, superficialmente, tempestivamente, tipicamente, ulteriormente, universalmente, urgentemente, vagamente, validamente, vergognosamente, vivacemente, volutamente, vorticosamente, zittamente. Infine, si riscontra una nutrita rappresentanza di participi presenti utilizzati come aggettivi: *aleggiante, danzante, brontolante, urlante, esultante, brillante, tremante, respirante, rassicurante, disturbante, titubante, rinfrescante, illuminante, spaccante, soffocante, raspante, tremante, traballante (...).*

Così come descritto in Bernardi (2008) è possibile individuare anche all'interno del *corpus 2018* alcune parole che non fanno parte del lessico della lingua italiana ma che sono possibili in quanto costruite secondo le regole di formazione delle parole. Tra queste individuiamo a titolo esemplificativo: *caramente, vacanzare, vulcanoso, teatrare, trafficoso, trionfoso, tristeggiare, solezza, smonatura, saltoso, pollinoso, eccitoso, narcisamente, zittamente.*

Come si può notare è parecchio produttiva l'aggettivazione a partire dai nomi attraverso il suffisso *-oso*, utilizzata, nella maggior parte dei casi, correttamente per indicare la presenza e l'abbondanza della qualità espressa dal nome da cui derivano. Allo stesso modo, si può notare la tendenza a formare verbi denominali (*vacanzare/ teatrare*) o deaggettivali (*tristeggiare*). Infine, la produzione di nuove parole possibili riguarda la creazione di avverbi in *-mente*: nel corpus possiamo infatti osservare forme come *narcisamente*, non presente in alcun vocabolario, ma che conta 6 occorrenze in Internet o come *zittamente* presente in 4 pagine Internet.

Fa invece eccezione la parola *ossequiamente*; secondo le regole di formazione degli avverbi in *-mente*, il suffisso va legato alla forma femminile singolare dell'aggettivo. La forma corretta dovrebbe infatti essere *ossequiosamente*. L'occorrenza di questo singolo

caso di parola non esistente e non possibile non ci consente tuttavia di fare generalizzazioni sull'intero corpus. La forma in questione potrebbe essere infatti il frutto di un semplice refuso o di un lapsus.

Dalla prospettiva delle scelte lessicali e delle costruzioni morfologiche, dunque, il *corpus 2018* riporta le stesse peculiarità individuate nel delta dei significati. Non si può dire altrettanto se si confrontano i due gruppi dalla prospettiva della sintassi. Ciò che colpisce, già a prima vista, è una sostanziale differenza tra i testi analizzati in Bernardi (2008) e quelli del *corpus 2018*. Sebbene in quest'ultimo siano ancora presenti frasi con struttura sintattica marcata, a sorprendere è la netta diminuzione dei casi in cui si osserva una sintassi "a motore semantico" (cfr. Di Benedetto 2008: pag. 87); aspetto che accomuna tutti i partecipanti. Pur essendo presenti costruzioni che si liberano dalla struttura SVO attraverso l'anticipazione del secondo argomento o la posposizione del predicato (esempio 1,2,3,4), che invertono l'ordine nome aggettivo (es. 5, 6) o che fanno a meno degli articoli e di altri elementi grammaticali (es. 7,8,9), l'incidenza di queste deviazioni dalla norma è inferiore a quella descritta all'interno della ricerca EASIEST. Nel *corpus 2018* la maggior parte delle proposizioni infatti dispongono di una struttura non marcata, del tutto simile a quella utilizzata correntemente dalla popolazione cosiddetta "neurotipica" (es. 10, 11, 12, 13).

1. *"suoni buoni faccio"* (N_1)
2. *"forte molto grande mi sembra"* (I_5)
3. *"bravi e testardi operatori ho trovato"* (R_8)
4. *"io donna silenziosa sono"* (B_2)
5. *"sono come vulcanosi monti ormai svuotati che hanno bruciato le loro emozioni"*
(B_4)
6. *"forte senso di piacere nei rigidi muscoli"* (D_2)
7. *"io dico che importante problema si verifica"* (E_5)
8. *"rabbioso momento interno mi ha colpito"* (G_3)
9. *"gambe non aiutano ma mani sì"* (O_2)
10. *"io mi sento tranquilla e sto studiando quello che di nuovo mi circonda comprese le nuove stanze"* (C_1)

11. *“Tutti danno giudizi senza appello, però nessuno vede dietro all'apparenza che ci sono le colpe di tutti”* (E_7)
12. *“Ho sempre fatto paura alla gente per il mio aspetto e per i miei comportamenti”* (B_4)
13. *“La lettera che mi hai scritto è molto bella e mi ha permesso di conoscere di più una persona che ricordo da bambino e basta”* (D_2)

La diminuzione dell'incidenza di queste particolarità sintattiche permette di fare alcune osservazioni. La prima ha che fare con la constatazione che si tratta di una diminuzione delle peculiarità e non di una assenza; in questo senso si possono confermare i dati descritti in Bernardi (2008), confermando la tendenza di chi scrive con la comunicazione facilitata ad ammettere costruzioni sintattiche marcate. Allo stesso tempo, la netta diminuzione di questi tratti all'interno del *corpus 2018* fa pensare a un processo di “normalizzazione” della scrittura avvenuto all'interno del centro dell'associazione “Vi comunico che penso”. Un processo che, se dal punto di vista dell'integrazione e dell'inclusione sociale può essere utile, dal momento che avvicina nello stile e nella forma la produzione linguistica dei facilitati a quella utilizzata comunemente, favorendo in questo modo uno scambio comunicativo più tipico, dal punto di vista dell'analisi linguistica lascia soltanto intravedere quei tratti e quegli aspetti stilistici che viceversa emergerebbero naturalmente. Concludendo questa prima parte di analisi semi-qualitativa, riferendoci agli scopi che si erano proposti, possiamo affermare che il *corpus 2018* presenta numerose somiglianze con il corpus utilizzato dal progetto EASIEST; queste somiglianze sono più forti dal punto di vista lessicale, mentre sul piano sintattico sembrano essere annacquate da un processo di normalizzazione e avvicinamento a forme meno marcate della scrittura. Sulla base di questa dichiarazione possiamo sostenere che la tipicità stilistica descritta in Bernardi (2008) non sia propria solamente dei soggetti con autismo ma sia invece caratteristica di tutti coloro che utilizzano la Comunicazione Facilitata. A livello di approfondimento dei dati sarà dunque più opportuno concentrarsi sulla facilitazione in sé (effetti del tocco, dinamica facilitatore-facilitato) e sul meccanismo che porta alla formulazione del pensiero per iscritto anziché sulla patologia delle persone che utilizzano la CF. Riguardo al secondo obiettivo di questa prima analisi, vale a dire la possibilità di estendere i risultati quantitativi realizzati in

questo progetto agli studi condotti in seno al progetto EASIEST, è opportuno fare delle distinzioni. Se infatti i due corpora presentano numerosi tratti in comune, le differenze riscontrate a livello sintattico, determinate, secondo la nostra ipotesi, da un processo di “normalizzazione” dello stile, potrebbero essere il segnale di una diversa partecipazione e influenza, indiretta, del facilitatore nella scrittura. I risultati che si otterranno con le analisi successive potranno, dunque, essere estesi al corpus del delta dei significati solo al netto delle differenze riscontrate.

4.1.3 Corrispondenze all'interno del corpus.

La seconda parte dell'analisi semi-qualitativa adotta una prospettiva di attribuzione d'autore. Sulla scia degli studi proposti da Emerson (2010) si sono confrontate le scelte lessicali, morfologiche e sintattiche di ciascuna coppia facilitato-facilitatore per vedere se queste scelte fossero dipendenti dal facilitatore o al contrario fossero proprie dello stile del facilitato. L'analisi semi-qualitativa proposta da Emerson (vedi cap. 3) aveva dimostrato come all'interno del suo corpus vi fossero sia elementi riconducibili al facilitatore, sia elementi ricorrenti esclusivamente in testi scritti dallo stesso utilizzatore della CF ma con facilitatori diversi. Così, come i tratti dipendenti dal facilitatore esprimono la natura collaborativa e co-costruttiva della CF, la prova dell'esistenza di termini e sintagmi esclusivi di un facilitato che occorrono con più di un facilitatore sono da considerarsi come indicatori dell'autorialità degli scriventi. Allo stesso modo l'analisi proposta di seguito ha cercato di confrontare i diversi testi sulla base del lessico, della morfologia, dei sintagmi e della lunghezza dell'enunciato per identificare sia elementi caratteristici del facilitatore sia elementi caratteristici del facilitato.

4.1.3.1 Lunghezza media dell'enunciato (LME)

La lunghezza media dell'enunciato per i testi del *corpus 2018* è stata calcolata dividendo il numero di parole di ciascun testo per il numero di frasi. I risultati ottenuti, descritti dalla tabella 3, mostrano come un facilitato tenda ad avere valori di LME più simili ai valori di sé stesso quando scrive con un altro facilitatore rispetto ai valori di LME degli altri facilitati quando scrivono con il suo stesso facilitatore. In particolar modo i casi di I, N e O, dimostrano una forte coerenza interna lungo il dominio della lunghezza della

frase. Per quanto riguarda i facilitatori, soltanto il numero 8 sembra essere correlato a una lunghezza media dell'enunciato più alta, comune a tutti gli utilizzatori della tecnica che scrivono con lui. Se nel caso di R_8 la differenza rispetto a R_1 è trascurabile, in quanto in entrambi i casi la lunghezza dell'enunciato presenta valori alti, nel caso di C_8 ed E_8 la LME si assesta su valori molto più alti rispetto a quelli rispettivamente di C_1, C_2, C_4 e di E1, E_2, E_5, E_7. A parte queste eccezioni, nei restanti casi il valore della lunghezza media dell'enunciato sembra essere costante per ciascun facilitato.

	A	B	C	D	E	G	I	N	O	R
1	7,9	9,7	10,5	9,8	9,3	7,1	8,2	7,2	5,7	11,4
2	6,4	11,7	8,8	8,6	8,8				6,1	
3	7,8					6,8		7,4	6,7	
4		8,9	10,3							
5					8,0		8,6			
7					9,4					
8			12,9		12,3					12,4
media	7,4	10,2	10,6	9,2	8,7	7,0	8,4	7,3	6,2	11,9

Tabella 3: valori della lunghezza media dell'enunciato. Sull'asse x i facilitati, sull'asse y i facilitatori

I dati esposti qui sopra, per quanto non risolvano da soli la questione CF, sono importanti per due diversi motivi; da un punto di vista dell'argomentazione sull'autorialità, infatti, si conferma, almeno nella maggior parte dei casi, una indipendenza dal facilitatore su un aspetto della lingua. La lunghezza media dell'enunciato può essere infatti considerata come una misura indiretta della sintassi utilizzata dallo scrivente o, dal momento che stiamo parlando sostanzialmente di dialoghi, come una misura delle modalità e degli approcci alla conversazione da parte del facilitato. Dal punto di vista della "meccanica" della CF, dimostrare una coerenza interna ai facilitati nel dominio della LME significa anche affermare l'indipendenza del gesto di digitazione; un gesto che non è guidato passivamente ma che sa quando interrompersi e considerare un periodo concluso.

4.1.3.2 Indicatori lessicali ei sintagmatici di influenzamento e indicatori di autorialità.

Il risultato descritto da Emerson (2010) si propone come una mediazione tra le posizioni degli oppositori della tecnica e le posizioni sostenute da chi è favorevole alla CF. Ammettere l'esistenza di tratti facilitatore-specifici significa confermare la presenza di qualche forma di influenzamento nella scrittura. Similmente l'individuazione di tratti facilitato-specifici confermano una forma di autorialità e di indipendenza da parte del facilitato. L'integrazione di questi due aspetti determina una serie di riflessioni sulla tecnica; riflessioni che avranno connotazione opposta a seconda della cornice interpretativa che si utilizza: che si intenda la comunicazione come solo scambio informativo o che la si consideri invece dal punto di vista dell'interazione sociale (Duchan, 1999).

Anche all'interno del *corpus 2018*, l'analisi semi-qualitativa dei testi ha individuato aspetti linguistici che ricorrono esclusivamente nei testi scritti con uno stesso facilitatore in unione con tratti linguistici che al contrario si trovano esclusivamente in più testi di un utilizzatore della CF.

4.1.3.2.1 Metodologie di indagine

I testi del *corpus 2018*, ai fini della analisi delle corrispondenze, sono stati trattati con il *tokenizer* implementato nel pacchetto "Stylo" di R. In questo per ciascun testo è stata creata una lista di tutte le parole e le sequenze di due e tre parole presenti con le rispettive frequenze. A partire da questa lista sono state individuate tutte quelle forme o sequenze di forme che risultavano essere presenti, esclusivamente, in almeno due testi scritti dallo stesso facilitato o in almeno due testi scritti con lo stesso facilitatore. Una volta individuate queste occorrenze si è proceduto con una analisi per lemmi, condotta manualmente. Questo tipo di operazione ha consentito di individuare solamente i lemmi tipici di un facilitato o di un facilitatore escludendo tutte quelle occorrenze inizialmente ritenute esclusive per motivi legati alla concordanza o alla flessione; ad esempio, la parola *silenziosa* compare unicamente nei testi scritti con il facilitatore 1; il lemma *silenzioso* invece, è presente nei testi B1, B2, C1, C2, C4, C8, E1, E2, E5, N1, N3. Non ci sono quindi gli estremi per ritenere il lemma *silenzioso* come forma esclusiva di chi scrive con 1 e pertanto non è stato incluso nella nostra analisi.

4.1.3.2.2 Tipicità legate al facilitatore: indicatori di influenzamento?

I lemmi e le sequenze di parole condivise da coloro che scrivono con uno stesso facilitatore sono descritte nelle tabelle a pag. 78 e seguenti. Partendo da un semplice sguardo d'insieme, ciò che salta all'occhio è l'abbondanza di occorrenze tipiche di chi scrive con il facilitatore 1 e in parte con il facilitatore 2; i lemmi e i sintagmi condivisi esclusivamente da chi scrive con i facilitatori 3, 4, 5 e 8 sono invece in numero minore. Questa differenza potrebbe però dipendere dalla diversa rappresentazione dei diversi facilitatori all'interno del corpus. Non è infatti un caso che i due facilitatori con più parole esclusive siano gli stessi che contano il maggior numero di testi, rispettivamente 10 e 6. È dunque possibile ipotizzare che, progressivamente, con l'aumentare delle dimensioni dei testi aumentino anche le parole che quei testi condividono in maniera esclusiva. Non è tuttavia da scartare l'ipotesi, proposta da Emerson, secondo la quale alcuni facilitatori potrebbero essere più inclini a questo fenomeno rispetto ad altri e che quindi esercitino un'influenza maggiore sulla scrittura, sia essa legata realmente a un effetto ideomotorio sia invece dovuta all'interazione e allo scambio dialogico. Si ritornerà su questi aspetti più avanti, nel capitolo dedicato alla discussione. Per il momento ritorniamo alle liste di parole e di sintagmi per analizzarle nel dettaglio.

Riferendoci alle parole condivise unicamente da chi scrive con il facilitatore 1, accanto a parole di ordine comune come *gridare, esibire, esplodere, imporre, organizzare, materno*, troviamo altri lemmi a bassa frequenza come *indolenzito, infermo, irto, lacrimante, lagna, piagnucolare pensoso, rabbiosamente, risatina, rognoso, spiacente, succulento, viaggiante, tensivo* che mantengono anche all'interno del corpus lo status di parola a bassa frequenza dal momento che nella maggior parte dei casi occorrono solamente due volte, in due testi diversi, eccezion fatta per *spiacente* che compare 39 volte in 9 facilitati diversi. È interessante notare come sia frequente il ricorso al verbo *attivare* e *azionare* per chi scrive con 1, così come al verbo *organizzare*. Osservando i lemmi dal punto di vista della classe grammaticale si possono individuare tre categorie principali:

1. gli aggettivi in *-oso* e avverbi costruiti in *-mente* costruiti su questa base, (*affettuosamente, disastroso, dispettoso, grintoso, maestoso, pensoso, rabbiosamente, scontroso*)

2. i participi presenti utilizzati come aggettivi (*barcollante, ferente, impegnante, lacrimante, luccicante, rifiutante, saltante, tormentante, viaggiante*)
3. i verbi introdotti dal prefisso *ri-* (*ri-accogliere, ri-avere, ri-confermare, ri-controllare, ri-fiorire, ri-attivare, ri-girare, ri-mettere, ri-organizzare*).

Si può far notare come il ricorrere ad aggettivi in *-oso* o a participi presenti con funzione aggettivale sia piuttosto frequente e comune a tutti i facilitati indipendentemente dai loro facilitatori. Per questo motivo, in riferimento alle parole elencate al punto 1 e al punto 2 sarebbe più opportuno considerare la base dell'aggettivo, rispettivamente il nome e il verbo da cui derivano, basi che non risultano esclusive di un facilitatore ma sono condivise da più di una coppia facilitato-facilitatore. Anche i verbi introdotti dal prefisso *ri-* trovano una diffusione piuttosto ampia anche nei testi scritti con facilitatori diversi da 1 anche se non rappresentano una particolarità così eclatante del corpus. Indipendentemente da ciò, non va sottovalutato il fatto che questi lemmi risultino dipendenti da un facilitatore. Dal punto di vista della discussione sulla comunicazione facilitata i risultati proposti rappresentano indubbiamente un argomento a favore di coloro che si oppongono alla tecnica. Se invece si prende in considerazione la prospettiva proposta da Duchan, Emerson e Zanobini, questi dati possono essere interpretati all'interno di un'ottica di collaborazione e di co-costruzione del messaggio, in cui alcuni momenti sono a maggior carico del facilitatore e altri a maggior carico del facilitato.

Lo stesso si può dire per i dati riguardanti i sintagmi (pag. 78 e seguenti). È interessante soffermarci su alcune costruzioni verbali e coppie aggettivo sostantivo. Si può, tra le altre cose, osservare:

l'utilizzo dei verbi *azionare* e *attivare* utilizzati nella medesima accezione:

- io aziono:
 - buon umore (G1)
 - brutti modi (I1)
 - comportamento consono (D1)
 - lamenti (R1)
 - gesti bruschi (A1)

- io attivo
 - capacità (B1)
 - forza (N1)
 - mente e corpo (I1)
 - la coscienza (E1)

il verbo *esibire*, utilizzato con la prima persona ritorna spesso nei testi scritti con 1:

- io esibisco:
 - accordo con questo... (E1)
 - forte stanchezza (D1)
 - momenti di umore differente (B1)
 - una forte interiore gioia (C1)
 - una salute buona (A1)

la costruzione del verbo *salire* seguito da un sostantivo indicante uno stato d'animo:

- sale rabbia (B1, D1, G1, I1)
- sale ansia (A1, E1, I1, R1)
- nervoso sale (N1, B1, I1)

la costruzione *diventare* o *fare il mulo*, nell'accezione di *diventare cocciuto*:

- divento un mulo (G1, I1)
- fare il mulo (E1, I1)

la costruzione *desiderio/desideroso* di

- sapere (D1, I1)
- festa (D1, N1)
- imparare (B1, I1)
- scrivere (N1, G1)
- chiacchiera (E1, O1)

il ricorrere di coppie formate da sostantivo + aggettivo:

- calma interiore (B1, G1, I1, N1)
- gesti nervosi (B1, N1, O1)
- umore variabile (B1, I1, D1)
- umore ballerino (A1, B1, D1, R1) e solo in una occasione da C2
- aria frizzante (E1, I1, R1)

- brezza leggera (A1, I1)
- occhi lacrimanti (A1, N1)
- morte inflitta [come sinonimo di omicidio] (E1, I1).

È interessante notare come una buona parte di queste costruzioni ricorrenti abbiano a che fare con la descrizione del proprio stato interiore, ci si riferisce spesso al nervosismo, alla rabbia al proprio umore: argomenti che sono spesso al centro dello scambio verbale tra facilitatore e facilitato all'interno del contesto quotidiano e dell'interazione dinamica della facilitazione. È possibile ipotizzare un'adesione da parte dei facilitati alle forme comunemente utilizzate dai facilitatori nella loro interazione verbale.

Riferendoci al facilitatore 2 (tabella pag. 78 e seguenti) il numero di lemmi unicamente presenti nei testi scritti con lui è significativamente minore; per questo è complesso identificare delle classi o individuare dei tratti che li accomunano. All'interno della lista è di rilevanza la parola *rammarico* condivisa, almeno una volta, da tutti e sei gli utilizzatori della CF nei testi scritti con 2; tra le parole a bassa frequenza si contano i verbi *demordere* e *garbare* condivisi da E e O, l'aggettivo *giocondo* condiviso da C e O, utilizzati in contesti simili, in posizione preposta al sostantivo e infine i participi presenti *esultante* e *tribolante* utilizzati con funzione aggettivale.

Tra i sintagmi che ricorrono in chi scrive con 2, i più significativi riguardano le coppie aggettivo + sostantivo:

- si rifrange nell'aria e crea *atmosfera frizzante* (B2)
- paesaggio lunare con *atmosfera frizzante* e lucente (C2)
- vedo ragazza seduta e *sguardo assente* (O2)
- ...momenti di interruzione con *sguardo assente* (C2)
- Poter fermare con più facilità i *miei bollenti spiriti* (D2)
- Cronache da Roma gettano *bollenti spiriti* sulle nostre teste (E2)

Interessante anche il ricorrere della costruzione *vedere nuovi soli*, forma figurata per riferirsi a un cambiamento:

- Fiducia che si può migliorare e *vedere nuovi soli* (D2)
- Credo che tutti dobbiamo passare per il male, per poter riaffiorare a riva e vedere *nuovi soli* albeggiare. (C2)

Oppure la costruzione *tornare a sorridere*:

- Perché a tutti è concesso di *tornare a sorridere* (C2)
- No, però so tornare a sorridere (D2)

Per quanto riguarda i facilitatori 3, 4, 5 e 8 il numero di parole e sintagmi unicamente condiviso è fortemente ridotto, probabilmente in virtù di una diversa rappresentazione in termini di lunghezza dei testi all'interno del *corpus 2018* o, se vale l'ipotesi presentata nel caso del facilitatore 1, di una minor esposizione alle abitudini linguistiche del facilitatore.

Tra i lemmi vale la pena sottolineare nel caso di 3 la forma *debolina*, diminutivo di *debole* che compare sia in A che in O e l'aggettivo di uso non comune *estasiante*.

Nel caso di 4 non sono stati individuati lemmi unicamente condivisi, mentre nel caso di 5 l'unico lemma condiviso risulta essere *visione*.

Il facilitatore 8 presenta tra i lemmi condivisi dalle persone che facilita, un paio di avverbi in *-mente* (*maggiormente* e *splendidamente*) insieme a una coppia di parole a bassa frequenza (*apice* e *ristoratore* usato come aggettivo).

Per quanto riguarda i sintagmi in 3 si osserva un frequente ricorso a costruzioni verbali con il verbo *dire* flesso alla prima persona senza espressione del pronome soggetto seguito o preceduto da un avverbio o un aggettivo che si rifanno allo stato emotivo di chi scrive (*dico seria*, *seriamente dico*, *dico felice*). Simile a queste costruzioni il sintagma *sento seria* (O3, A3).

In 4 l'unico sintagma condiviso sia da B che da C, *vedere la luce*, dall'analisi del contesto di occorrenza risulta essere utilizzato con accezioni diverse. In B *vedere la luce* è utilizzato con una intenzione letterale e si riferisce al bisogno di B di stare in una stanza luminosa considerata la paura del buio (es. 1). In C invece lo stesso sintagma è utilizzato in forma figurata, similmente alla forma *nuovi soli* nel testo scritto con il facilitatore 2 (es. 2).

1. Le stanze piccole e senza luce...mi fanno paura, io devo vedere la luce.
2. Io ho pregato tanto per tutti quelli che hanno operato l'accaduto, hanno più bisogno di vedere la luce e trovare il pentimento.

In 5 si possono notare tre sintagmi in cui alla parola *uomo* segue una specificazione sotto forma di aggettivo (*uomo liberato*, *uomo pensante*, *uomo più grande*).

In 8 infine si osservano costruzioni come *non tollero* che compare in R8 e E8, il riferimento all'inizio del caldo con la formula *estate scoppiata all'improvviso* e la coppia sostantivo aggettivo *pensieri affettuosi* che compare sia in C8 che in R8.

4.1.3.2.3 Tipicità legate al facilitato: indicatori di autorialità?

Così come la presenza di parole tipiche di chi scrive con un facilitatore possono rappresentare un indice dell'influenzamento del facilitatore stesso, la dimostrazione dell'esistenza di parole che vengono utilizzati solo da un utilizzatore della tecnica con diversi facilitatori può rappresentare un indizio circa l'autorialità dei facilitati. Le tabelle presentate alla fine del capitolo indicano le parole e le espressioni tipiche di ciascun facilitato.

A una prima vista si può notare una diversa corposità delle liste di parole. In particolar modo C ed E mostrano un numero di parole e sintagmi "propri" di gran lunga più alto rispetto a quello degli altri facilitati. A seguire A, B e O, per concludere con D, G e I. È interessante notare come i facilitati che presentano un maggior numero di parole e sintagmi propri siano gli stessi che hanno scritto in maniera consistente con più di un facilitatore. È possibile leggere questi dati nella prospettiva dell'ipotesi, precedentemente discussa, secondo la quale le abitudini linguistiche di un facilitatore possano fungere da modello di riferimento per il facilitato; una maggior variazione del tipo di facilitatore potrebbe determinare un minor ancoraggio alle scelte linguistiche del facilitatore e di conseguenza una maggior libertà e indipendenza nella scrittura.

Riferendoci ad A possiamo notare come i lemmi che vengono utilizzati con più di un facilitatore appartengono al campo lessicale della cucina *castagnata, cipolla, cottura, erbe, fragola, marmellata, pomodoro* oppure alla sfera della geografia *Grecia, Sicilia, Marocco*. La natura di questi lemmi condivisi è sicuramente legata all'attività svolta da A all'interno del centro. Vista infatti la sua passione per la cucina non è raro che durante le sedute di colloquio individuale si parli di ricette. All'interno della stessa dimensione lessicale il sintagma *prelibati piatti* che si ritrova sia in A1 che in A2. Tra gli altri sintagmi interessanti che ricorrono in A troviamo *pensiero positivo* e *deciso colore*. Infine, ritorna un paio di volte il sintagma *mia impresa* con cui A si riferisce a una sfida che deve essere superata con difficoltà:

- Mondo sembra guardarmi storto e senza speranze, è solo mia impresa dovermi accettare (A2)
- Decisa veste nuova che mi rende soddisfatta della mia impresa (A1, in riferimento agli esami).

Tra i lemmi di B si osserva la parola a bassa frequenza *iroso* che compare 4 volte nel corpus con cui B si riferisce a sé stessa e al suo modo di essere.

- Triste e *iriosa* insomma incazzata (B4)
- Donna scontrosa e *iriosa* sono e avanzo con rabbia (B1)
- Il potente nervoso mi sovrasta rendendomi *iriosa* e senza controllo (B4)
- Il mio vagare lento, gli scatti *iroso* lo fanno arrabbiare (B1).

Allo stesso modo l'aggettivo *lineare* è utilizzato da B in maniera figurata per riferirsi alla propria vita o ai rapporti con le persone:

- Questa vita condotta così. *Così poco lineare* (B4)
- Questo rapporto mai chiarito, che non è *lineare* ma molto tormentato (B4)

Tra i sintagmi di B i più significativi si possono indicare:

- *forte nervosismo* che compare sia in B2 che in B4,
- *porto preoccupazione* (B1, B2) utilizzato nell'accezione di *sono preoccupata*
- *sono paurosa* con l'accezione di "aver paura" (B1, B4).
- Il caso di C, insieme a quello di E, è sicuramente il più interessante. Le parole che sembrano indicare una coerenza interna a C e indipendente dai suoi facilitatori sono infatti molte e significative.

Tra i lemmi di uso non comune si contano *accudiente, baraonda, congratularsi, macello, meravigliosamente, rilento, scosso, scuotimento, venerando*. Il caso di *accudiente* è particolarmente interessante dal momento che la forma del participio presente comunemente considerata corretta in italiano è *accudente* di cui tuttavia non ci sono tracce nel corpus. Altrettanto degni di nota sono i lemmi *concretizzare, frenesia, mistico, prassi, stasi, rockettaro* e i sostantivi alterati *piccolino, fratellone, sorellina* che rappresentano una unicità dei testi di C all'interno del corpus di N.

Concretizzare occorre tre volte con tre facilitatori diversi:

- Lo stato promette ma non *concretizza* (C1)

- Desiderio di generare che si *concretizza* in un figlio (C2)
- *Concretizza* in me riflessioni dopo ogni incontro (C8)

Frenesia occorre invece con tutti i facilitatori di C:

- ...dove la dimensione della *frenesia* aveva lasciato spazio al riposo (C4)
- Uomo inquinatore preso dalla *frenesia* della costruzione (C1)
- Rovine romane immerse nella *frenesia* della vita (C8)
- Il tutto dipende dalla *frenesia* delle cose (C2)

Stasi ricorre in entrambi i casi in riferimento a un proprio stato

- Sono un po' in *stasi* (C2)
- Mi sento risucchiare in un paralitico momento di *stasi* (C4)

È possibile inoltre individuare una classe di parole riferita all'ambito della religione; fanno parte di questa classe lemmi come *catechisti, confessione, cristiana, domenicale, mistico, peccati, pentecoste*.

Un altro tratto distintivo di C sta nel suo utilizzo frequente di forme figurate; *catene* è utilizzato ad esempio con l'accezione di impedimento, così come con *ingabbiata* si riferisce a uno stato di mancanza di libertà:

- Ora sei libera di volare senza *catene* o fardelli (C1)
- Chi soffre è costretto forse a *catene* pesanti (C2)
- Mi sento *ingabbiata* da pesante muro di imbarazzo (C2)
- *Ingabbiata* dentro a regole di conforme società (C1)
- Mi chiamo C e *ingabbiata* in questo corpo scrivo (C2)
- Ovatta fastidiosa. Mi teneva la mente *ingabbiata* (C4)

Allo stesso modo *nutrimento* è usato con l'accezione di essenzialità per lo spirito;

- Natale è *nutrimento* dell'anima (C2)
- L'anima ha trovato *nutrimento* nelle preghiere (C4)

la *sete* di cui parla non risponde a un bisogno del corpo ma della mente:

- Sete di conoscenza (C8)
- Sete di verità (C2)

Le *strade nuove* e il *nuovo volto* sono utilizzate per riferirsi rispettivamente a un cambiamento di prospettive nella vita e a un rinnovamento interiore;

- Si forma davanti a voi una *nuova strada* (C1)

- Felici di iniziare una “*nuova strada insieme*” (C1)
- Ti auguro che *strade nuove* si aprano (C2)
- Peso che... da un *nuovo volto* a quello che ci turba (C1)
- Fare ordine e dare al popolo un *nuovo volto* (C2)

Tra le sequenze di parole esclusive di C si osserva l’utilizzo di espressioni non comuni quali *animare i cuori, fiduciose speranze, molteplici pensieri, amorevole sguardo*.

- Principio che *anima i cuori* sotto la croce (C2)
- Malvagio spirito che *anima i cuori* di chi non ha parola (C4)
- Le vostre proposte mi danno *fiduciosa speranza* (C2)
- Ho imparato ad ascoltare a tramutare in *fiduciosa speranza* (C1)
- *molteplici pensieri* da adulti (C2)
- Io sono sovrastata da *molteplici pensieri* di dolorosa condizione familiare (C4)

Il verbo chiedere è frequentemente utilizzato da C in costruzioni come *chiedere di essere/avere/vivere/imparare*.

La costruzione *essere in fase di* è un altro elemento esclusivo di C:

- Sono *in fase di* risveglio (C1)
- Sono *in fase di* risveglio e aspetto di dare merenda (C2)
- Sono entrata a pieno ritmo e *in fase di* studio (C8).

Il caso di D si colloca invece all’opposto. Tra le parole esclusive a bassa frequenza d’uso troviamo unicamente *mutilato* utilizzato in entrambi i casi in riferimento alla sfera familiare e in particolar modo alla perdita di un caro:

- C’è una figlia *mutilata* che sogna ancora papà (D2)
- C’è il dolore per la mia *famiglia mutilata* (D1).

Non si riscontrano invece sintagmi particolarmente significativi.

Il facilitato E invece presenta un numero significativamente maggiore di lemmi e sintagmi esclusivi. Tra le parole non comuni che occorrono unicamente in E troviamo *assordanti, arso, estremista, sessuoso*.

Quest’ultima parola ricorre frequentemente in E, così come frequenti sono i riferimenti alla sfera della sessualità.

- Ho gustosi pensieri sessuosi che pace non danno (E1)

- Resto sempre con risolini per sessuosi pensieri (E5)
- Io rido per stato d'ansia e sessuosi pensieri (E1)
- Ho dormito poco, io ho pensieri sessuosi (E1)

L'aggettivo *sanguinoso* è utilizzato da E in forma figurata; mentre in B1 è riferito a una ferita procurata di recente in E1 ed E2 è utilizzato per riferirsi a fatti o avvenimenti cruenti:

- Disperati *sanguinosi* fatti (E1)
- I *sanguinosi* fatti sulla striscia di Gaza (E2)
- Serio rancore per la crudele *sanguinosa* mano umana (E1).

E è inoltre l'unico a utilizzare *x* al posto della preposizione per:

- Mi sentivo molto esposto ai numeri *x* ciò... (E2)
- Tutto *x* per darmi regole civili per vivere accettato (E1)
- *X* me fermo aiuto per buoni compiti (E2)

Tra i sintagmi che ricorrono solamente in E citiamo la costruzione *essere fremente di*

- Vera donna che era *fremente* di essere intelligente (E5)
- *Fremente* sono di stare in tua piacevole compagnia (E7)

Il riferirsi a sé stesso come *uomo rigido* e *uomo urlante*

- Le mie idee di *uomo rigido* (E8)
- Giornate difficili per me *uomo rigido* nel pensiero (E8)
- *Uomo rigido* sono (E2)
- Non sarò *uomo urlante* (E1)
- Donne che siete accorse da questo *uomo urlante* (E8)

il menzionare i propri *schemi* mentali; le altre occorrenze di schema sono utilizzate per riferirsi allo studio o al movimento.

- *Schemi* che difficilmente tolgo di dosso (E2)
- Questo non rientra nei miei *schemi* abituarini (E8)
- Le vacanze mi danno ansia e io ho *schemi* precisi (E1)
- Tempo scambussola i miei *schemi* (E5)
- Questo sbalzo disturba i miei *schemi* (E5).
- Mio schema mentale richiede freddo in inverno (E8)

Infine, l'utilizzo del verbo cogliere in riferimento a un messaggio o un significato:

- Mio padre ha colto il valore della mia esperienza scolastica (E2)
- Credo che chi ha ben guardato ha colto il messaggio (E1)
- Le mie orecchie hanno colto (dalla radio) che semestre più piovoso da 200 anni a questa parte (E8)
- Vedo che hai colto il senso di questa radio.

I facilitati G, I N, R e O in maniera simile a D non presentano molti aspetti esclusivi degni di nota né dal punto di vista dei lemmi né da quello dei sintagmi.

Di G citiamo l'uso del sintagma *momenti vuoti* parlando dell'assenza di momenti di noia all'interno del centro:

- Senza troppe pause e senza *moment vuoti* (G3)
- Pochi sono i *momenti vuoti*, io sto bene qua (G1)

Di I il sintagma *mi capita nervoso*:

- Quando non ho pittura mi *capita nervoso* ma poi passa (I5)
- Dico che sono calmo, mi *capita nervoso* solo ogni tanto (I1)

Di N l'uso di gradevole in *gradevole sentire e saluto gradevole* e della coppia sostantivo + aggettivo *giorni vivaci*.

In R si osserva l'occorrenza di una forma dialettale come *ciacolona* utilizzata come sinonimo di *chiacchierina* altra forma unica di R.

La scarsa manifestazione di parole esclusive, per alcuni facilitati, potrebbe essere, come per i facilitatori 3, 4, 5, 8, una diretta conseguenza di una dimensione ridotta dei testi o di un disequilibrio nella dimensione dei testi (come nel caso di I).

Infine, va ricordato come questa analisi si sia occupata dei lemmi e dei sintagmi più particolari, che all'interno del corpus occorrevano in misura minore. L'analisi delle parole e dei sintagmi a maggior frequenza è lasciata ai metodi di analisi quantitativa dei testi che saranno oggetto del prossimo capitolo. Nelle pagine seguenti sono riportati per ciascun facilitatore e per ciascun facilitato le forme lessicali e i sintagmi tipici ed esclusivi.

SINTAGMI DI CHI SCRIVE CON 1

1. In ambiente
2. Luoghi nuovi
3. Solo pioggia
4. Nuove abitudini
5. Piccoli esseri
6. Sale rabbia
7. Pensieri diventano
8. Tesoro prezioso
9. Giorni faticosi
10. Brutti comportamenti
11. Trovare soluzioni
12. Mani controllate
13. Sgridate ricevo
- 14. Calma interiore**
15. Parte nervoso
16. Porta fastidio
17. Giusto modo
18. Buoni complimenti
- 19. Sento nervi**
20. Tranquillo uomo
- 21. Ho sentimenti di**
22. Pensieri importanti
23. Corpo agitato
- 24. Gesti nervosi**
- 25. Umore variabile**
26. Comportamenti sbagliati
27. Giorni nervosi
28. Cuore contento
29. Giornata storta
30. Piacevole momento
- 31. Alterni momenti**
32. Grande cambiamento
- 33. Testa confusa**
- 34. Nervoso sale**
35. Riflessivi pensieri
36. Ritrovato volti
37. Ansiosa agitazione
- 38. Aria frizzante**
39. Brezza leggera
- 40. Dico sincera**
41. Giornata fiacca
42. Graffi pugni
43. Pensiero chiaro
- 44. Umore ballerino**
45. a darmi calma
46. a grandi passi
47. abbastanza bene dico
- 48. abbracci e sorrisi**
49. abitudine a giornate
50. agita rumori voci
- 51. agitato e nervoso**
52. aiuto mi serve
53. ansia e angoscia
54. arriva rimprovero mamma
- 55. aspetto il sole**
- 56. attivo le mani**
- 57. avuto sinistro fitte**
58. bene aria tiepida
59. bisogno di staccare
60. buona salute grazie
61. buono stato fisico
62. calda stagione
63. caldo e aria
- 64. calma e corpo**
65. carico d'amore
66. chiusi nella testa
- 67. cielo grigio porta**
68. come tenerlo buono
69. compagnia di chiacchiere
- 70. con occhi lacrimanti**
71. con occhio attento
72. controllare le mani
73. controllo questo corpo
74. corpo fa fatica
75. critico momento
76. cuore di genitori
77. cuore ricolmo di
- 78. desiderio di chiacchiera**
79. desiderio di festa
80. desiderio di imparare
81. desiderio di sapere
82. desiderio di scrivere
83. buona leggerezza
84. buoni sorrisi
- 85. paesaggi meravigliosi**
86. poco ascolto
87. positivi pensieri
88. riunione familiare

89. serio impegno
90. dico con sincerità
91. dipende sempre dalla
92. dire con certezza
93. disturba gli occhi
94. **divento un mulo**
95. dolce da gustare
96. emozione molto forte
97. energia nel corpo
98. **esplode all' improvviso**
99. **esplodo fastidio sentirti**
100. esprimerlo permette di
101. fa arrabbiare tutto
102. **fantasia di immagini**
103. fare lunghe passeggiate
104. fastidio di pancia
105. fastidio sentirti sulla
106. fatica a entrare
107. fitte lievi
108. forti urli
109. forza per superare
110. fredda e umida
111. fuori controllo dispetti
112. gesti poco femminili
113. giorni di dolore
114. giri al mare
115. grigio di pioggia
116. ha dato gioia
117. ha portato morte
118. ho pensieri particolari
119. ho piacevole gioia
120. ho recuperato energia
121. il caldo sole
122. **il mio grillo**
123. il tempo fugge
124. **in buona sintonia**
125. **in critico momento**
126. **in stato nervoso**
127. incrocio di dita
128. interiore calma
129. io devo calmarmi
130. **io esibisco**
131. **io aziono**
132. io sono attraversata
133. io sono pacifica
134. io sono resistente
135. la mia crescita
136. **la morte inflitta**
137. lamenti da fare
138. le lunghe maniche
139. libera voglio guarire
140. libero i pensieri
141. malessere alle gambe
142. mamma ha paura
143. mani poco controllate
144. maschere e dolci
145. metto mio pensiero
146. mi disturba tutto
147. mi tolgono stabilità
148. mie conosciute stanze
149. **mie giornate scandite**
150. mie quotidiane cose
151. momenti di bagno
152. momenti di unione
153. momento di agitazione
154. momento di confusione
155. morte e distruzione
156. nervoso mi colpisce
157. pensando al mare
158. pensieri e opinioni
159. pensieri fastidiosi
160. pensieri importanti e
161. pensieri sul futuro
162. pensiero e esplodo
163. penseroso sulla vita
164. per devo fare
165. per non sfigurare
166. pesanti pensieri
167. piacevole e divertente
168. questo cielo grigio
169. regole da rispettare
170. ricaricare le batterie
171. ricolmo di affetto
172. ritmo delle giornate
173. ritrovare la strada
174. rotola senza freno
175. **sale ansia**
176. sensazione di tranquillità
177. sensazione piacevole
178. senso di fastidio
179. sento abbastanza bene
180. sento nervi tesi

181. **senza negativi pensieri**
182. sonno negli occhi
183. sonnolenza mi colpisce
184. sonnolenza negli occhi
185. sono in agitazione
186. sono senza nervi
187. **sono spiacente**
188. sono tutta libera
189. speranza nel cuore
190. stato di ansia
191. studiosi del tempo
192. tanta neve arrivata
193. tanto sono nervosa
194. temperatura non troppo
195. tempo di uova
196. testa tutto confuso
197. tiene in allerta
198. tornata in grillo

199. tristezza nel cuore
200. tutto mi agita
201. tutto mi disturba
202. tutto torna calmo
203. umore nero
204. un po' stordita
205. un sorriso sereno
206. un uomo serio
207. un uomo tranquillo
208. un vortice di
209. una gioia forte
210. vedo perduto controllo
211. vero niente altro
212. via della guarigione
213. voce mi distraggono
214. volte passi goffi

SINTAGMI DI CHI SCRIVE CON 2

1. Ed io
2. Mia madre
3. Sono apposto
4. **Accade attorno**
5. **Atmosfera frizzante**
6. **Assente sguardo**
7. **Bollenti spiriti**
8. Coltre di
9. **Prende piede**
10. Respingo facilmente
11. Sano confronto
12. alla mia età
13. bella idea sarebbe
14. bruciore di stomaco
15. che non tradisce
16. condizione di donna
17. cose da risolvere
18. dato tutta me
19. deve essere difficile
20. difficile pensare che
21. essere all'altezza
22. essere una donna
23. faccio tesoro di
24. forte sensazione di
25. giorni si alternano
26. ha dato assicurazione
27. ho grandi aspettative
28. ho molto bisogno
29. imparare nuove cose
30. in mille pezzi
31. io mi aggrego
32. io posso resistere
33. la mia storia
34. la sai apprezzare
35. le vostre proposte
36. lontana da casa
37. mancanza di concentrazione
38. mi piacerebbe moltissimo
39. mi sento partecipe
40. mio corpo riposa
41. muovere le mani
42. occasione per raccontarmi
43. persone semplici che
44. punto di riferimento
45. quando ho male
46. rassicurante per me
47. sarebbe il massimo
48. sempre in tensione
49. sì, molto curiosa
50. sono apposto grazie
51. sono felice oggi
52. sono molto emozionata

53. sono piuttosto nervosa
54. tornare a sorridere
55. tremare le gambe
56. una grande famiglia

- 57. vedere nuovi soli**
58. vera testa dura
59. vorrei essere tranquilla

SINTAGMI DI CHI SCRIVE CON 3

- 1. Seriamente dico**
- 2. Dico seria**
- 3. Sento seria**
4. Energia positiva
5. alla normalità tornato
6. bella idea questa
7. che buon periodo
8. cielo e sole
9. curiosa sarei
- 10. dico felice sono**
11. difficile senza dubbio
12. faccio buoni lavori
13. fanno stare meglio
14. grande bella giornata
15. le voglio bene
16. momenti di tristezza
17. neve ho sentito
18. ricco di emozioni
- 19. rilasso muscoli**
20. sento resto tranquilla
21. serena e rilassata
- 22. sognare e immaginare**
23. sorridente sono anche
24. sta arrivando estate
25. tristi pensieri sono
26. tutto alla normalità
27. voglia di primavera

SINTAGMI DI CHI SCRIVE CON 4

1. Viva io
2. vedere la luce

SINTAGMI DI CHI SCRIVE CON 5

1. Caldo umido
2. Uomo liberato
- 3. Uomo pensante**
4. contento di iniziare
- 5. fatica faccio a fermarmi**
6. porta poca salute
7. uomo più grande

SINTAGMI DI CHI SCRIVE CON 8

1. Non tollero
2. Mi invade
3. Emozione addosso
4. Concordo pienamente
- 5. Estate scoppiata all'improvviso**
- 6. Pensieri affettuosi**
7. Malanni addosso
8. Freddo intenso
9. Credo fermamente
10. brutta influenza
11. che tanto abbiamo
12. emozionanti per me
13. influenza di stagione
14. io mi scarico
15. la propria opinione
16. nostra vita quotidiana
- 17. pensieri dal cuore**
18. sempre grande piacere
19. tra voi donne

SINTAGMI DI A

1. deciso colore
2. mia impresa
3. notte fonda
4. pensiero positivo
5. prelibati piatti

SINTAGMI DI B

1. porto preoccupazione
2. difficile gestire
3. fuori luogo
4. forte nervosismo
5. parole ripetute
6. risposte chiare
7. sono paurosa

SINTAGMI DI C

1. (Dio) ci chiede di essere/avere/esultare/imparare
2. abbraccio caldo
3. essere adorato
4. amore infinito
5. amorevole sguardo
6. anima i cuori
7. aver bisogno di novità
8. avvenimenti orribili
9. brontolona
10. chiedo sostegno
11. chiedono di essere
12. coltivi/sorridi ogni giorno con amore
13. compagnia femminile
14. con animo pacifico/guerriero
15. dare modo
16. dolci sorprese
17. essere in buone mani
18. essere in fase di
19. essere in stasi
20. essere incuriosita
21. essere parte di un gruppo
22. essere trovata
23. famiglia numerosa
24. fiduciose speranze
25. giorni intensi
26. i miei amici
27. il tuo sorriso
28. io non ho voce
29. la tua bellezza
30. meritato riposo
31. mi sono sentita coinvolta
32. molteplici pensieri
33. momento intenso
34. multi + sostantivo
35. non ho voglia di stare seduta
36. nuovo volto/ un nuovo inizio
37. ogni passo (figurato)
38. piena di luminose luci/stelle
39. poter comunicare
40. ritmo scolastico
41. aver sete (figurato)
42. sgranchire le gambe
43. sono fiera
44. sono in pensiero per te
45. spensierata allegria/ingenuità
46. spero di vedervi presto
47. stanchezza addosso
48. strada nuova (figurato)
49. studio che faccio
50. trovare/cercare il mio posto nel mondo

SINTAGMI DI D

1. molto dispiaciuta
2. tanto desidero

SINTAGMI DI E

1. a fatto accaduto/finito
2. Aggettivo + vicende
3. amare la vita
4. ausiliare avere + colto
5. comunque, sempre
6. cuore ferito
7. distribuirsi
8. diventare uomo + aggettivo
9. essere usato da
10. felice uomo
11. forte senso di
12. fremente
13. giovani vite
14. ho sentimento di
15. idea/idee in mente/proposito
16. io dico spero
17. lavaggio del cervello
18. perenne + sostantivo
19. ragazzo pensante
20. riferimenti alla sessualità
21. schemi precisi/personali/miei
22. sembrare stupido
23. trovare/poter essere una soluzione
24. turbare/scuotere le coscienze
25. uomo rigido sono
26. uomo urlante
27. x invece di per

SINTAGMI DI G

1. fratello sono
2. momenti vuoti

SINTAGMI DI I

1. volentieri esercizi
2. caldo soffocante
3. capita nervoso
4. esercizi rilassanti
5. forte chiusura

SINTAGMI DI N

1. giorni vivaci
2. gradevole sentire
3. saluto gradevole
4. tutto esplode

SINTAGMI DI O

1. gioia buona
2. saluto aspetto
3. vera pigrizia

SINTAGMI DI R

1. neve candida
2. sono chiacchierina

Lemmi condivisi da chi scrive con il facilitatore 1											
	Lemma	A	B	C	D	E	G	I	N	O	R
1	affettuosamente			1					1		
2	affollato		1					2			
3	alterno	1	5		2		1	2	6	1	
4	annodato				1			1	1		
5	arzilla			1			1	1	1		
6	attivare		3	1			1	7	1		
7	attorcigliare		3		3			2			
8	avaria				1	1					
9	azionare	1	6					1	1	1	
10	barcollante		1					1			
11	beatitudine		1	1	1						
12	benevoli		1					1	1		
13	buco		1	1		1	1	1	3		
14	caotico	1	2			2		1			
15	capriccio		8		3				6	4	
16	cenone						1	1			
17	chiarire	1						2			
18	concorde		2	2		2		7	1		
19	contessa	1									1
20	coordinato							5	1		
21	crampo				1				1		
22	curvo					1		1	1		
23	declinare		1		2			1			
24	delineare		2	1	1						
25	disastroso			1					1		
26	dispettoso		5	1				5	2		
27	dissenso			1	1						
28	docile				1		1				
29	donnetta		1								1
30	durevole		1	1		2		2			
31	educato		3					12	2		
32	elettrico			1		1			8		
33	energico			1					1		
34	esibire	1		1	4	4			2		
35	esplodere	1	3						1	1	
36	esposto							3	1	1	1
37	esuberante		1		1	1			1		
38	esule				1	2					
39	fagottino				1						1
40	fasullo							1		1	
41	ferente		1				1	2			
42	fitta	1			1		1	2	2		
43	forzuto	1			1			2			
44	frantumato		2			2					
45	frescolino							1	1		
46	generare	1	1	1					1		
47	goffo	2	1					2	2		

Lemmi condivisi da chi scrive con il facilitatore 1											
	Lemma	A	B	C	D	E	G	I	N	O	R
48	grattare		6						4		
49	gridare							3	1		
50	grintoso						1	1	2		
51	groviglio		1	1	1			1	1		
52	impasto		1		1			3			
53	impegnante							2	1		
54	imporre	1	1	2					1		
55	indispensabile				1	1		1			
56	indolenzito				2			2			
57	infermo	1				1					
58	infliggere				2			2		1	1
59	influire					1		1			
60	infuriare						1		1		
61	irto					1		1			
62	lacrimante	1	1	1				1	3		
63	lagna		5	3						2	
64	lamentela	1	1				1	1	1	1	
65	lesto					1			3		
66	letale		1			1					
67	luccicante			1					1		
68	maestoso	1			1			2			
69	malconcio		2	1				1			
70	malmesso	1		1		2		2			
71	manesco						1				
72	mascherato	2	2					2	2		2
73	materno		6					2	3		
74	mesto				1	1		3			
75	miscuglio				1	1					
76	miseramente					1		2			
77	mossa		2					1	2		
78	mostruoso							1	1		
79	mulo					1	1	3			
80	nostalgico	1		2				2			1
81	offendere					1		1			
82	offensivo		2					2			
83	offesa		1			2	1	1			
84	opaco		1		1			1			
85	organizzare	2						3	3	2	
86	ostinato		1	1	1				2		
87	pasticcio								1	1	
88	pelo						1		1		
89	pensoso					1		2	1		
90	pericolante		2					2			
91	pestifero							1	1		
92	piagnucolante		1							1	1
93	possente					1					1
94	rabbiosamente		1	1	1						

Lemmi condivisi da chi scrive con il facilitatore 1											
	Lemma	A	B	C	D	E	G	I	N	O	R
95	rauco			1	1						
96	riaccoglierie		1								1
97	riattivare		1	1					2	1	
98	riavere		2	1		1				1	
99	ricolmo	1		1							
100	riconfermo	1					1				
101	ricontrollo								1		
102	ricoperto	1								1	
103	rifiorire	3			2	1		3			
104	rifiutante					1		1	1		
105	rigirare					1		1	1		
106	rimettere		1	3	2	2		3	4		4
107	riorganizzare		1							1	
108	ripetitive	9							2		2
109	ripetitivo	12					1		1	1	
110	risatina				2	1					
111	rituale		1							1	
112	rognoso		1	1						1	
113	rompi		1				1	1	3	1	
114	saliente		1		1						
115	saltante	1	4			1				2	
116	scatenare		3		2					5	
117	schiacciare								1	1	
118	sciocco				4			15	3		
119	sconnesso		1								1
120	scontroso		1		3			3			
121	sfigurare			1		2		1	1		
122	smaltire		1	1					1		
123	solenne			1	1	2		1			
124	sonnolente				1				1		
125	spiacente		10	1	8	2	4	9	1	1	3
126	stemperato					1		1			
127	stento			1		1	1				
128	stordito		14		2	2	3	7			
129	strampalato	1			1				1		
130	strategia		1	1	2	1		1	3		
131	succulento	2					1	1			
132	tensivo								1	2	
133	tormentante				1					1	
134	tremarella		1	1					1		
135	tristemente					2		1			
136	ugualmente			1	3	1	1	4			
137	variegato		3						1	1	
138	variopinto			1					2	2	
139	viaggiante				1						1

Lemmi condivisi da chi scrive con il facilitatore 2							
	Lemma	A	B	C	D	E	O
1	adagiare			1			2
2	astenersi			1			1
3	badare	1					1
4	colpetto			1	1		1
5	demordere					1	1
6	deridere	1					2
7	egregia			2			1
8	esultante			1		1	1
9	garbare					1	1
10	giocondo			3			1
12	impetuoso			2	1		
13	incatenato			1			1
14	levare			5			1
15	rammarico	1	1	1	2	1	1
16	signorile			1			1
17	sospettoso	1		1			2
18	spazientirsi					1	1
19	sussulto			2			1
20	tribolante		2		1		1
21	viaggetto			3			1

Lemmi condivisi da chi scrive con 3					
	Lemma	A	G	N	O
1	accertarsi		1		1
2	debolino	1			1
3	estasiante	1			1
4	meditare	1		1	
5	pendente			1	1
6	raggiante		1		1

Lemmi condivisi da chi scrive con 5			
	lemma	E	I
1	visione	1	1

Lemmi condivisi da chi scrive con 8				
	lemma	C	E	R
1	apice	1	1	
2	collettivo	1	1	
3	consumismo	1		1
4	domicilio	1	1	
5	maggiormente	1	1	
6	ristoratore	1		1
7	splendidamente	1		1

Facilitato A				
	Lemma	1	2	3
1	abbinamento	1		1
2	castagnata	1		1
3	cipolla	1		1
4	cottura	1		1
5	erbe	1		2
6	fragola	1	1	
7	gesso	2		1
8	gioventù	2		1
9	grezia	1		1
10	marmellata	1		2
11	marocco	1		1
12	pomodoro	3		1
13	presepio	1		1
14	sicilia	1	1	

Facilitato B				
	Lemma	1	2	4
1	bambola	2	1	
2	ferire	2	1	
3	difendersi	1	1	
4	disorientato	1	1	
5	urgente	1	1	
7	riferimento	1	4	2
8	lineare		1	1
9	parolaccia	5		1
10	sberla	5		2
11	confondere	2		2
12	don	3		1
13	iroso	2		2
14	fingere	1		1
15	imprevedibile	1	1	1
16	ostinarsi	1		1
17	tuono	1		1
18	insicurezza	5	2	
19	sottoforma	1	1	
20	scoppio	5		2
21	incerto	2		3

Lemmi esclusivi di C					
	Lemma	1	2	4	8
1	accudiente			1	1
2	allargato		1		1
3	apprezzato	1	2		
4	arido	1	2		
5	aspettativa		1		1
7	aula	1	1		1
8	baraonda				
9	brillantemente		1		1
10	capanna			1	1
11	catechista	1	1		
12	catena		2		1
13	collana	1	1	1	
14	collegare		3	1	1
15	colui		1		1
16	concretizzare	1			1
17	condividere	1	1		1
18	confessione	1	1		
19	congratulare	1			1
20	creativo	1			2
21	cristiano		3		1
22	domenicale	1	1		
23	eccomi	1			1
24	egli		3		1
25	fatina		2		1
26	fondamenta		2	1	
27	formulare		1		1
28	frenesia		1		1
29	grandissimo	1	1	1	1
30	guadagnato		1		1
31	guerriero		1		1
32	importantissim	1	1		
33	incantato		1		1
34	incantevole	1	1		1
35	incredibile	1	2		
36	indisposto	1	2		1
37	informatico		1	1	
38	ingabbiato		1		1
39	ispirare	1	2	1	

Lemmi esclusivi di C					
	Lemma	1	2	4	8
40	laguna		2	1	
41	lanterna		3		5
42	libretto		1	1	
43	macello		2		1
44	memorabile			1	1
45	mentalmente		1		1
46	meravigliosamente		1		1
47	mirato		1		1
48	mistico	1			1
49	multiplo	1		1	
50	nutrimento		1		1
51	offerto		1	2	
52	operato		1		1
53	orribile	1		1	
54	peccato		1	1	1
55	pentecoste	1			1
56	piccolino		3		1
57	prassi		4		1
58	pregiudizio		1	1	
59	preliminare	1		1	1
60	profitto		1		1
61	protettrice		1		1
62	pubblicitario		1	1	
63	rafforzare	1			1
64	rilento	2	5		2
65	rockettaro		1		2
66	scossone			1	1
67	scuotimento			1	2
68	snello		1	1	
69	soave	1		1	
70	stasi		3		1
71	universitario		1	1	
72	venerando	2	3		2
73	volentierissimo		1		1
74	fastidioso		1	1	
75	fratellone	2	1		2
76	sorellina		8	1	3
77	commosso	1		2	3

Facilitato D			
	Lemma	1	2
1	bruscamente	1	1
2	instabilità	1	1
3	livido	1	2
4	malissimo	1	1
5	mutilato	1	1
6	pallina	11	4
7	rabbia	2	1
8	tradito	1	3

Facilitato E						
	Lemma	1	2	5	7	8
1	arso				1	1
2	assordante	2		1	1	
3	astuzia	1			1	
4	avvenire	1			1	
5	bandiera	1	1			
6	bilanci	1	2			
7	calarmi	1	1			
8	captare		1		1	
9	competenza	1	1			
10	comunemente		1		1	
11	concedermi	1	1		1	
12	eccezionale		1		1	
13	epico	5	1			
14	estremista	1	1			
15	frequentatore	1	1		1	
16	gravemente				1	1
17	impressionato	1	1		1	
18	indugio		1		1	
19	indurre	1			1	
20	inevitabile	1	1	1		
21	leader		1		1	
22	leva				1	1
23	possibilmente				1	1
24	regnare	1	1		1	
25	sanguinoso	2	1			
26	sessuoso	3		1		
27	tolleranza	1			1	
28	tragedia	2	3	1	3	
29	tragico	1	2		1	
30	trattativa	1	1			

Facilitato G			
	Lemma	1	3
1	articolo	3	1
2	focus	1	2
3	sfogliare	1	1

Facilitato N			
	Lemma	1	3
1	disordinato	3	2
2	distratto	1	1
3	orologio	1	1
4	severo	1	1

Facilitato O				
	Lemma	1	2	3
1	acquoso	1		1
2	africano	1		2
3	amicizie	1	1	1
4	arto	1		2
5	asso		1	1
6	chiamare	1	2	
7	escluso	1		1
8	gattino		1	1
9	macchia	1		1
10	mezzora	1		1
11	pelo	1		2
12	peloso	1		3
13	risponso	1		1
14	risveglio	1		1
15	smalto	15	1	5
16	sommerso		1	2
17	stendere	1		1
18	versetto	1		1
19	vomitare	1	1	

Facilitato R			
	Lemma	1	8
1	chiacchierina	1	1
2	duchessa	1	1
3	ciacolona	2	2
4	sagra	1	1

Capitolo 5

5.1 Un'analisi quantitativa

L'analisi presentata nel capitolo precedente conferma attraverso strumenti qualitativi la complessa natura dei testi scritti con la Comunicazione Facilitata. La compresenza infatti di tratti stilistici dipendenti dal facilitatore e di altri dipendenti dal facilitato sembra essere a supporto dell'ipotesi, proposta da Emerson, secondo la quale la tecnica consiste in un processo di co-costruzione e di collaborazione tra scrivente e assistente.

Le analisi qualitative, per definizione, si possono occupare di piccoli campioni di testo, degli elementi meno frequenti e più particolari nel tentativo di individuare l'autore attraverso l'identificazione di somiglianze, ricorrenze o differenze.

L'aspetto innovativo di questo progetto consiste nell'applicazione delle più moderne tecniche di analisi quantitativa per l'attribuzione d'autore ai testi scritti in CF per vedere se le stesse considerazioni basate su campioni piccoli e specifici di lingua possono essere estese a campioni più ampi, più difficilmente dominabili dal ricercatore senza il ricorso ad ausili.

5.1.1 Materiali e metodi

5.1.1.1 Hierarchical clustering analysis

L'analisi per gruppi dei diversi corpus è stata eseguita con il software Stylo implementato in R. (Eder et al 2016). La distanza utilizzata è stata la *Cosine Delta*, che, come da analisi della letteratura cfr. Cap 1, risultata la più affidabile. Le analisi sono state condotte sulle 1000 parole più frequenti (1000 MFW) e successivamente sui 1000 trigrammi più frequenti (1000 MFC-3grams).

5.1.1.2 Supported Vector Machine e Random Forest

L'analisi attraverso la Supported Vector Machine e il Random Forest è stata eseguita in R. Sono stati utilizzati per l'analisi le 200 parole più frequenti, le 200 sequenze di 2 parole più frequenti (words bigrams), i 200 bigrammi e trigrammi di caratteri più frequenti. I risultati sono stati rappresentati graficamente all'interno di una confusion matrix. Ciascun testo all'interno dell'analisi con il *machine learning* è stato dapprima suddiviso

in più testi di almeno 500 parole. Ognuno di questi testi è stato utilizzato per costruire il *training set*, secondo un approccio *instance based*. I testi utilizzati per il *test set* sono poi stati scelti in maniera randomizzata tra i testi del *training set*.

5.1.1.3 Il corpus

Il corpus utilizzato in questa analisi è stato raccolto presso uno dei centri dell'associazione "Vi Comunico Che Penso". Sono stati raccolti gli scritti di 19 utilizzatori della CF, senza specificazione della diagnosi, che avessero scritto con costanza con almeno 2 facilitatori diversi, all'interno di un gruppo di 9; i testi corrispondono ai dialoghi e alle conversazioni tra facilitato e facilitatore, raccolti dal centro nel periodo che va dal 2006 al 2018. Per ciascuna conversazione sono stati separati gli interventi dei facilitati (individuati dal carattere minuscolo) dagli interventi dei facilitatori (individuati dal carattere maiuscolo). Infine, si sono uniti in un unico documento testuale tutti gli interventi dell'utilizzatore della CF scritti con un determinato facilitatore; infine ciascun testo è stato codificato sulla base del facilitato e del facilitatore coinvolto nella scrittura. Il codice assegnato ai facilitati corrisponde a una sequenza di 5 lettere identiche, mentre il codice assegnato ai facilitatori è dato da una sequenza di 5 numeri identici (es. AAAAA = facilitato A, 11111 = facilitatore 1 → nome del testo = AAAAA_11111). Il risultato di questa operazione è stato un corpus di 82 testi (gruppo A). In seguito, i testi di ciascun facilitato, indipendentemente dal facilitatore, sono stati uniti in un unico testo, per un totale di 19 testi (gruppo B). A partire dal gruppo B è stata eseguita una suddivisione randomizzata di ciascuna unità testuale in sottotesti di uguali dimensioni, per un totale di 46 testi (gruppo C).

Si è proceduto con lo stesso sistema con gli interventi del facilitatore: va da sé che la codifica del testo ha invertito l'ordine dei codici, antepoendo il codice numerico al codice alfabetico (es. 11111_AAAAA). Si è creato di conseguenza un gruppo 1 di 82 testi e un gruppo 2 di 9 testi.

Per quanto riguarda l'analisi con i sistemi di machine learning, per motivi legati alla dimensione dei testi, è stato creato un ulteriore sottogruppo del gruppo A. In questo caso sono stati selezionati tutti i testi che superassero le 5000 parole; una volta operata questa selezione si sono tenuti i testi di ciascun autore che comparisse almeno due volte

(due facilitatori diversi). Il gruppo D così formato è composto di 28 testi e corrisponde al *corpus 2018* dell'analisi qualitativa. Il gruppo E è composto dagli stessi testi ma codificati in funzione del facilitatore: questo per permettere ai sistemi di machine learning di operare una classificazione in funzione del facilitatore. La tabella 4 qui sotto riassume le operazioni condotte sul corpus.

Nome del gruppo	Numero di testi	Codifica	Descrizione
Gruppo A	82 testi	AAAAA_11111 AAAAA_22222	Testi scritti dai facilitati, per ciascuna coppia facilitato facilitatore
Gruppo B	19 testi	AAAAA_all BBBBB_all	Testi scritti da ciascun facilitato, uniti
Gruppo C	46 testi	AAAAA_all_00001 AAAAA_all_00002	Testi scritti da ciascun facilitato, uniti e divisi in modo casuale
Gruppo D	28 testi (> 5000 parole)	AAAAA_11111 AAAAA_22222	Testi scritti dai facilitati per ciascuna coppia facilitato facilitatore.
Gruppo E	28 testi (>5000 parole)	11111_AAAAA 22222_AAAAA 11111_BBBBB	Testi scritti dai facilitati per ciascuna coppia facilitato facilitatore, codificati in funzione del facilitatore.
Gruppo 1	82 testi	11111_AAAAA 22222_AAAAA	Testi scritti dai facilitatori, per ciascuna coppia facilitatore facilitato
Gruppo 2	9 testi	11111_all 22222_all	Testi scritti da ciascun facilitatore, uniti

Tabella 4: operazioni sul corpus

5.1.2 Risultati

Le indagini condotte all'interno di questa tesi muovono dalle conclusioni e dalle riflessioni prodotte in seno al progetto EASIEST (Bernardi 2008), nel tentativo di validare l'efficacia della comunicazione facilitata, attraverso alcuni nuovi metodi di analisi quantitativa dei testi. Come abbiamo avuto modo di osservare, lo studio coordinato da Lorenzo Bernardi dimostrava come vi fosse uno scarto tra lo stile linguistico dei facilitatori e lo stile linguistico dei facilitati, come anche confermato dalle analisi descritte nel capitolo 4. Questo scarto, ampiamente giustificato dalle riflessioni di natura qualitativa descritte in Bernardi (2008), a livello quantitativo si definiva in una diversa

distribuzione per categorie grammaticali: i testi scritti dai facilitati contavano una maggior occorrenza di aggettivi e avverbi rispetto ai testi scritti dai facilitatori che, a loro volta, mostravano una occorrenza significativamente superiore di parole funzione (parole grammaticali) e verbi. A questa misurazione si aggiungeva l'analisi per gruppi descritta dalla figura 5 (riportata nel capitolo 3).

Il dendrogramma delinea graficamente la distinzione compatta tra il gruppo dei facilitatori e il gruppo dei facilitati; i testi scritti dai facilitatori si raggruppano tra loro, sono simili dal punto di vista stilistico e differiscono dal gruppo dei facilitati. Allo stesso modo i testi scritti dai facilitati presentano delle caratteristiche che li accomunano, sono raggruppati omogeneamente nel dendrogramma e differiscono dai testi dei facilitatori come un unico gruppo.

Il corpus utilizzato e raccolto per questo lavoro presenta delle caratteristiche simili; l'analisi per gruppi dei testi scritti dai facilitati e dei testi scritti dai facilitatori (figura 7) rispecchia l'analisi frutto del progetto EASIEST, polarizzando ancor più la differenza tra i due gruppi. Il dendrogramma, infatti, si dirama esattamente in due direzioni: la prima direzione raggruppa i testi dei facilitati, i testi dei facilitatori la seconda.

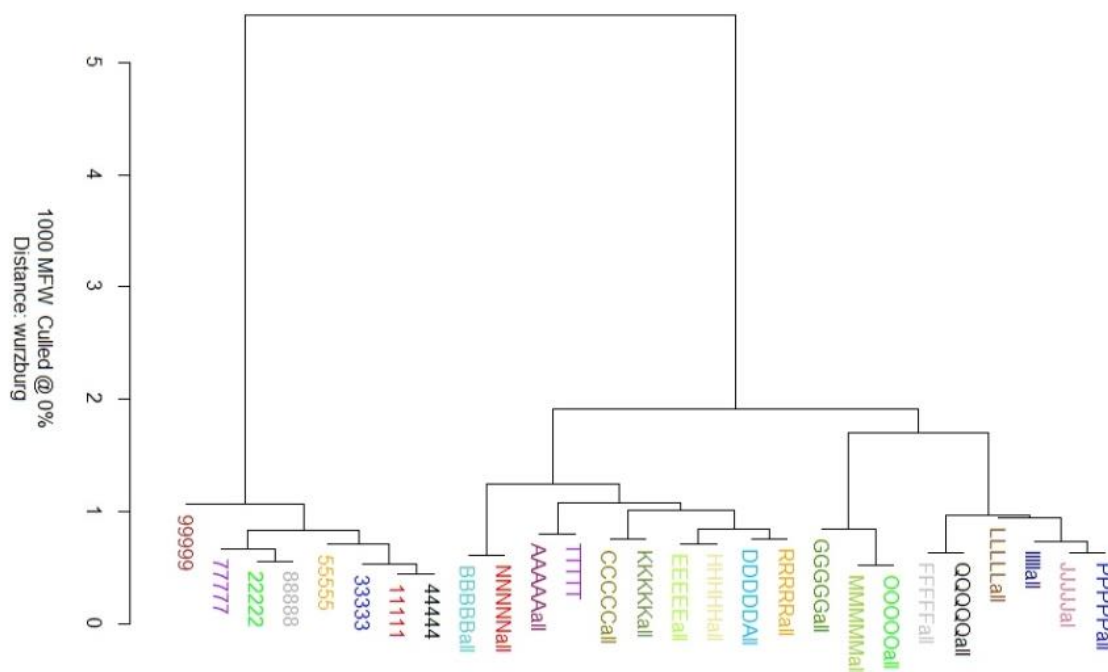


Figura 7. Cluster gerarchico agglomerativo, 1000M FW distanza Cosine Delta. Comparazione tra il gruppo B e il gruppo 2

Le coerenze stilistiche interne ai testi dei facilitatori, confermate dai due studi, permettono, innanzitutto, di affermare come vi sia uno stile proprio di chi scrive con la comunicazione facilitata e, in secondo luogo, come questo stile sia fortemente diverso da quello dei facilitatori ai quali viene invece attribuita l'autorialità. La differenza di stile tra i due gruppi, tuttavia, non può essere considerata, secondo la comunità scientifica, come una prova dell'efficacia della comunicazione facilitata. Il diverso stile dei testi attribuiti ai facilitati potrebbe essere infatti un sottoprodotto del processo stesso di scrittura e dell'effetto ideomotorio (descritto nel capitolo 2). Le analisi condotte finora attestano, secondo i detrattori della CF, uno scarto tra uno stile di scrittura utilizzato consapevolmente dai facilitatori nei loro interventi e uno stile inconsapevole, guidato dall'effetto ideomotorio, che prende il sopravvento negli interventi attribuiti ai facilitati. Sulla base di queste considerazioni, le analisi che vengono presentate qui si preoccupano di approfondire la questione utilizzando metodi di analisi dei testi più recenti e un rinnovato obiettivo di indagine: osservare la persistenza di uno stile linguistico interno al facilitato, indipendente dal facilitatore che lo assiste nel processo di scrittura. Per questo motivo, nella fase di costruzione del corpus, si sono tenuti in considerazione i testi scritti da ciascun facilitato con almeno due facilitatori diversi (vedi materiali e metodi). Individuare una continuità stilistica in testi scritti con facilitatori diversi significherebbe trovare indizi di una identità e di una intenzionalità slegata dal facilitatore e dal suo intervento, ma propria del facilitato; prove che effettivamente i testi scritti in CF sono rappresentativi del pensiero e della volontà dei facilitati. Allo stesso modo, l'eventuale dimostrazione di similarità stilistiche nei testi scritti con lo stesso facilitatore lascerebbe spazio a riflessioni circa il ruolo dell'assistente alla comunicazione e alla sua influenza sul processo di scrittura.

Le nuove metodologie utilizzate in questo lavoro, già presentate nel primo capitolo e riprese nel paragrafo relativo ai metodi, sono l'analisi per gruppi con una distanza delta coseno e una classificazione attraverso SVM (*Supported Vector Machine*) e RF (*Random Forest*).

L'analisi per gruppi condotta con i testi del gruppo B (figura 8) consegna un risultato piuttosto netto. In tutti i casi si osserva una continuità e una coerenza stilistica interna al facilitato.

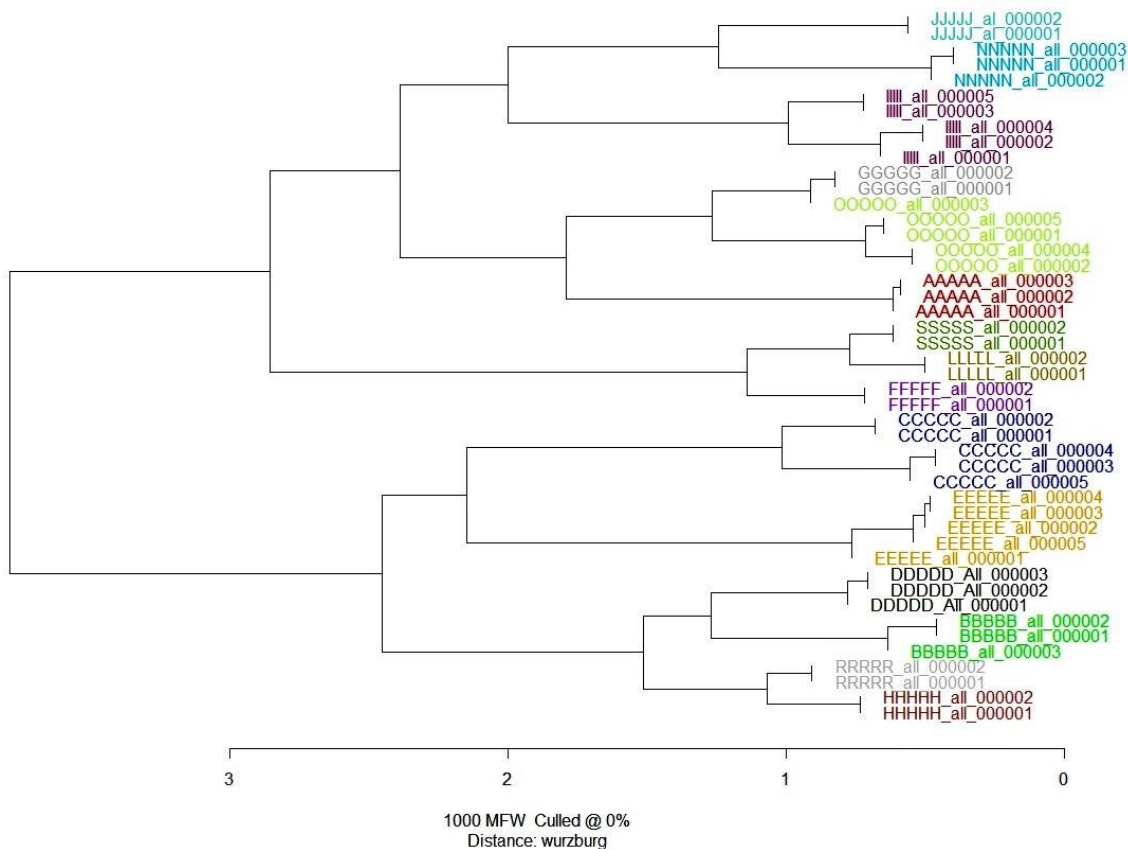


Figura 8: analisi per gruppi del gruppo C, distanza delta coseno, 1000 MFW.

La vicinanza nel dendrogramma dei testi scritti dagli stessi facilitati testimonia la presenza di uno stile, inteso come occorrenza di parole più frequenti, che si mantiene costante e indipendente dal facilitatore presente al momento di scrittura. È un risultato di notevole interesse poiché è in grado di cogliere, riconoscere, discriminare e distinguere con precisione gli elementi di continuità all'interno dell'intera produzione di ciascun utilizzatore della CF e che lo identificano e lo differenziano dagli altri autori.

La significatività di questo risultato è avvalorata dalla perfetta confrontabilità dei diversi elementi che compongono il corpus e dalla dimensione di ciascun testo; ogni elemento testuale del gruppo B supera infatti le 5000 parole, rispettando i parametri di affidabilità proposti da Eder (Eder et al. 2015).

I risultati sono replicati anche con l'analisi per gruppi sulla base dei 1000 trigrammi più frequenti; come si può osservare nella figura 9, si confermano i risultati ottenuti con le parole più frequenti.

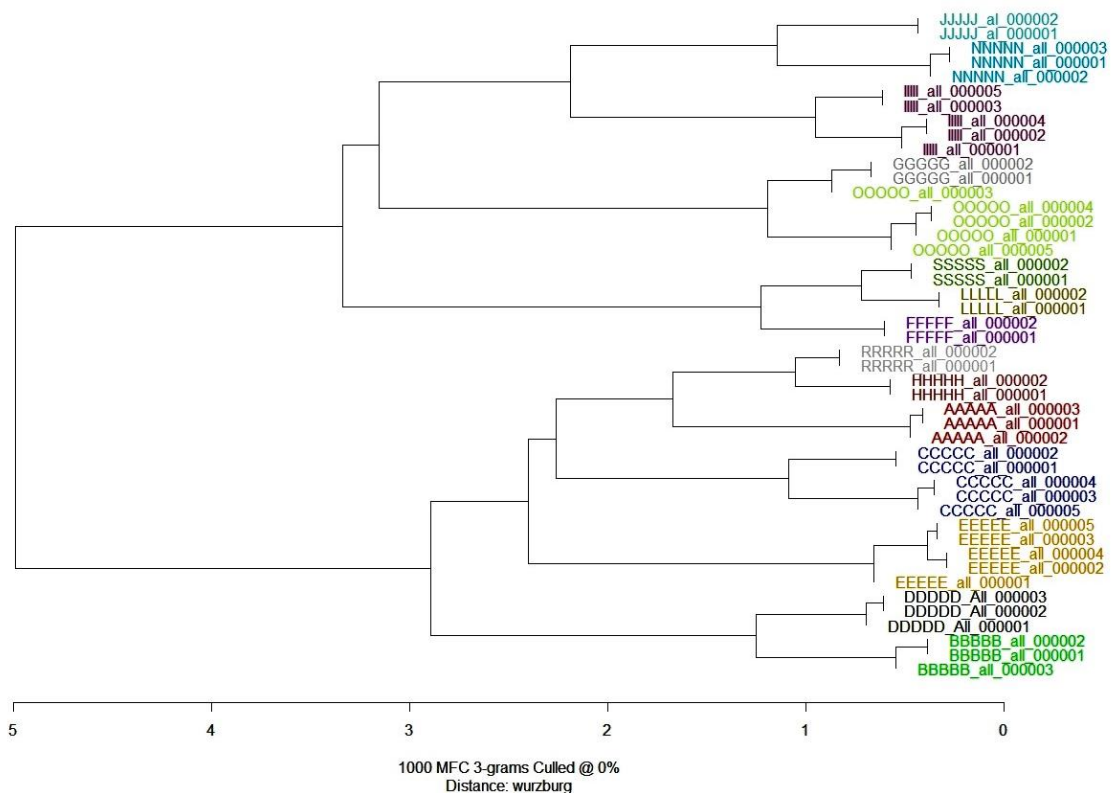


Figura 9: analisi per gruppi del gruppo C, distanza delta coseno, 1000 MFC.

L'interpretazione che si può proporre a partire da questi dati deve saper uscire dalla semplice descrizione del dendrogramma e cercare di leggere il grafico più in profondità, abbracciando anche la prospettiva di coloro che non ritengono la comunicazione facilitata uno strumento efficace per le persone con una disabilità intellettiva e comunicativa. Alcuni detrattori della CF (T. Saloviita 2016) spiegavano la possibile persistenza di uno stesso tratto stilistico in testi scritti con facilitatori diversi come un adeguamento del facilitatore allo stile utilizzato dal facilitato nei testi precedenti. Non è chiaro se l'autore dell'articolo intendesse che questo adeguamento fosse consapevole e intenzionale oppure guidato dall'inconscio. In entrambi i casi sembra tuttavia improbabile che questa tendenza a adeguarsi allo stile delle produzioni precedenti possa esprimersi in maniera così sistematica, soprattutto riuscendo a cogliere le differenti sfumature in termini di occorrenza delle parole grammaticali o delle parole ad alta frequenza. L'ipotesi di un adeguamento all'impronta stilistica del facilitato sembra dunque poco plausibile proprio perché le caratteristiche linguistiche idiosincratice del

facilitato sono difficilmente riconoscibili se non con strumenti di elaborazione quantitativa.

Un'ulteriore osservazione potrebbe essere avanzata a partire da questi risultati; la composizione del corpus è data dalla scomposizione casuale di ciascun testo in sottotesti in cui si intersecano i testi scritti con facilitatori diversi. È possibile, soprattutto nei casi in cui i testi di una coppia facilitato-facilitatore fossero maggiormente rappresentativi per il facilitato stesso, che la scomposizione casuale non abbia fatto altro che suddividere semplicemente in due o più testi diversi gli interventi di una stessa coppia, determinando in questo modo l'appaiamento nel dendrogramma. Secondo questo ragionamento si potrebbe dunque pensare che la similarità dei testi sia il risultato della presenza in testi diversi di uno schema stilistico comune poiché frutto della stessa coppia facilitato-facilitatore e, quindi, che non vi sia una reale continuità e coerenza di stile in testi scritti con facilitatori diversi ma una koinè dipendente dalla coppia facilitato-facilitatore. Questo ragionamento è sicuramente possibile nei casi in cui vi è un effettivo squilibrio nella dimensione dei testi scritti con un facilitatore rispetto ai testi scritti con un altro. In questi casi è tuttavia interessante misurare in che cosa consiste questa koinè, quali sono i tratti che consentono a una coppia facilitato-facilitatore di distinguersi da un'altra coppia in cui il facilitatore non cambia. Lo stesso discorso non può valere in quei casi in cui ogni facilitato è rappresentato da testi di ampie dimensioni per ciascun facilitatore con cui ha scritto (es. CCCCC, EEEEE, NNNNN, OOOOO su tutti).

Le riflessioni a cui ci conducono gli studi con l'analisi per gruppi non possono considerarsi conclusive sul corpus. Gli studi condotti da Emerson e confermati dall'analisi semi-qualitativa presentata nel capitolo 4, mostravano infatti come, accanto a questa coerenza stilistica dipendente dal facilitato, vi fosse una continuità stilistica da riferire al facilitatore. Sulla stessa lunghezza d'onda si sono collocati gli studi condotti attraverso il machine learning, sia con SVM (Supported Vector Machine) sia con RF (Random Forest). Le operazioni sono state condotte sul gruppo D e sul gruppo E per motivi legati alla dimensione dei testi. I risultati del SVM sul gruppo D sono descritti dalla matrice 1. La classificazione di ciascun autore è rappresentata dalle colonne verticali. Nell'intraprendere la descrizione della matrice1 va sottolineata come prima cosa la percentuale di accuratezza. Il 50% di accuratezza media non consente di fare

affermazioni conclusive e definitive su questi risultati. Vanno tuttavia ritenute utili in un’ottica esplorativa, un approfondimento sulla questione CF utile se i dati vengono confrontati con l’indagine qualitativa.

Prediction	Reference									
	AAAAA	BBBBB	CCCCC	DDDDD	EEEEE	GGGGG	IIIII	NNNNN	OOOOO	RRRRR
AAAAA	14.2	0.0	1.4	1.0	2.0	0.4	3.4	0.4	3.6	0.8
BBBBB	0.4	14.0	0.4	1.8	0.8	0.4	2.2	3.0	0.8	0.8
CCCCC	0.8	1.4	35.4	6.0	17.0	0.0	1.0	0.0	4.8	1.2
DDDDD	0.0	0.2	0.4	4.0	1.4	0.0	2.2	0.0	0.0	0.0
EEEEE	0.0	0.2	0.4	0.0	4.2	0.0	1.2	0.0	0.0	0.0
GGGGG	0.0	0.0	0.0	0.0	0.2	4.2	0.8	0.2	3.8	0.0
IIIII	0.0	0.0	0.0	0.2	4.6	1.6	9.0	0.0	0.0	0.4
NNNNN	0.2	0.2	0.0	0.0	0.0	1.2	5.4	4.6	0.6	0.0
OOOOO	0.4	0.0	0.0	0.0	0.0	0.2	2.4	9.0	5.6	0.0
RRRRR	0.0	0.2	0.2	0.2	0.0	0.0	0.0	0.0	0.2	0.8

Accuracy (average) : 0.5053

Matrice 1: analisi con SVM dei testi del gruppo D

Tenuto in considerazione questo, la matrice 1 ci consente di osservare come tre facilitati siano riconosciuti come effettivamente autori dei loro testi con un’ottima frequenza, ci riferiamo ai casi di AAAAA, BBBBB e CCCCC. In altri due casi, GGGGG e IIIII e in parte OOOOO gli utilizzatori sono riconosciuti come autori con una frequenza buona ma con numerose classificazioni erranee. Nei casi di RRRRR, NNNNN e DDDDD si osserva una classificazione erranea della maggior parte dei loro testi. È curioso invece il caso di EEEEE in cui la maggior parte dei suoi testi è riconosciuta come simile a CCCCC. Già nelle analisi per gruppi descritte precedentemente (figure Z e A1) si era constatata una somiglianza nello stile utilizzato da CCCCC e in quello utilizzato da EEEEE. È possibile che il programma, nell’analisi dei dati abbia riconosciuto queste somiglianze e che queste siano state all’origine della classificazione erranea. Se infatti osserviamo i risultati ottenuti con Random Forest sul gruppo D (matrice 2), possiamo osservare una riorganizzazione delle classificazioni dei testi, con un percentuale ampiamente migliore (68%). Il numero dei facilitati correttamente classificati sale a 7 (AAAAA, BBBBB, CCCCC, EEEEE, IIIII, NNNNN, OOOOO). Nel caso della matrice 2 è da sottolineare il forte miglioramento nella classificazione di NNNNN, OOOOO e IIIII oltre al completo ribaltamento nella classificazione di EEEEE che, pur mantenendo alcuni elementi di continuità con CCCCC torna a classificarsi correttamente e con un’ottima frequenza.

	Reference									
Prediction	AAAAA	BBBBB	CCCCC	DDDDD	EEEEE	GGGGG	IIIII	NNNNN	OOOOO	RRRRR
AAAAA	9.4	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.2	0.0
BBBBB	0.0	8.0	0.2	1.0	0.0	0.0	0.0	0.2	0.0	0.0
CCCCC	3.2	3.6	36.2	7.0	5.4	0.0	0.6	0.2	2.8	3.2
DDDDD	0.0	0.6	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0
EEEEE	0.4	0.4	1.0	1.2	22.0	0.0	1.4	0.2	0.6	0.2
GGGGG	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0
IIIII	2.4	1.4	0.8	1.6	2.6	4.8	25.6	2.6	1.0	0.6
NNNNN	0.6	2.2	0.0	0.0	0.0	0.0	0.0	13.6	1.4	0.0
OOOOO	0.0	0.0	0.0	0.4	0.2	3.0	0.0	0.2	13.4	0.0
RRRRR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Accuracy (average) : 0.6863

Matrice 2: analisi con RF dei testi del gruppo D

Le ricerche condotte con il gruppo E sono servite per indagare l'effettiva presenza di uno stile riconducibile al facilitatore. L'obiettivo dell'indagine consisteva nel verificare se i testi scritti con lo stesso facilitatore venissero classificati in funzione del facilitatore, rivelando degli aspetti tipici in tutti i testi scritti con lo stesso assistente, o, al contrario, se non vi fossero dei tratti riconducibili al facilitatore. I risultati ottenuti con SVM (matrice 3) confermati dagli esiti di RF (matrice 4) mostrano come effettivamente vi siano dei tratti riconducibili al facilitatore. Questa continuità stilistica è più forte per alcuni facilitatori (11111, 8888) e più debole con altri (44444, 55555), con notevoli differenze in base al metodo utilizzato (22222, 33333, 77777). È interessante notare come questi risultati coincidano in maniera piuttosto netta con i dati descritti

	Reference						
Prediction	88888	11111	44444	22222	77777	55555	33333
88888	14.4	1.2	0.2	6.8	0.6	0.0	11.6
11111	0.0	94.6	0.4	5.8	0.2	3.8	8.4
44444	0.0	0.6	4.4	20.8	0.0	0.0	0.0
22222	0.0	0.0	0.0	4.2	0.8	0.0	0.0
77777	0.0	0.0	0.0	0.0	5.6	0.0	0.0
55555	0.0	0.0	0.0	0.0	3.8	0.0	0.8
33333	0.0	0.0	0.0	0.0	0.0	0.0	1.0

Accuracy (average) : 0.6537

Matrice 3: analisi con SVM dei testi del gruppo E

	Reference						
Prediction	88888	11111	44444	22222	77777	55555	33333
88888	9.6	0.2	0.0	0.0	0.0	0.0	0.0
11111	3.8	96.0	2.2	0.8	0.0	3.6	1.8
44444	0.0	0.0	0.0	0.0	0.0	0.0	0.0
22222	1.0	0.2	2.8	36.8	0.0	0.2	1.0
77777	0.0	0.0	0.0	0.0	11.0	0.0	0.0
55555	0.0	0.0	0.0	0.0	0.0	0.0	0.0
33333	0.0	0.0	0.0	0.0	0.0	0.0	19.0

Accuracy (average) : 0.9074

Matrice 4: analisi con RF testi del gruppo E

dall'analisi qualitativa in cui il facilitatore 1 e in parte il facilitatore 2 contavano un maggior numero di parole e sintagmi unici.

La presenza di una uniformità stilistica subordinata al facilitatore è, come abbia già detto, una componente che si deve presupporre e ammettere nella relazione dinamica che si instaura tra il facilitato e il facilitatore. Come suggeriscono Duchan (1999), Zanobini e Scopesi (2001) e Emerson (2010) la comunicazione facilitata è da intendersi come un processo di co-costruzione del pensiero scritto, una collaborazione tra utilizzatore della CF e assistente che deve assumere valore in una prospettiva sociale e di interazione e non in un'ottica di semplice scambio informativo. Per questo motivo non deve sorprendere o condurre a rapide conclusioni la presenza di aspetti linguistici comuni ai testi scritti con lo stesso facilitatore. Ciò che emerge con maggior forza da questi risultati è l'ipotesi di una terza via nell'interpretazione della CF, una terza risposta alla domanda "Chi è che scrive quando si usa la CF?", la stessa teorizzata da Duchan e confermata da Zanobini (2001) e Emerson (2010). I testi che risultano da un processo di CF sono frutto in parte dell'intervento del facilitato e in parte del contributo del facilitatore. Un contributo che può essere diverso a seconda di chi partecipa alla scrittura a seconda dell'esperienza del facilitatore o dello stesso facilitato; vi sono facilitati che risentono in misura minore dell'intervento del facilitatore e altri in cui questa "influenza" è più marcata. Il tutto va inquadrato in una prospettiva che vede nella CF non lo strumento di comunicazione definitiva e finale, ma un'occasione per migliorare le

possibilità comunicative di persone con gravi deficit di linguaggio e un graduale processo verso l'autonomia comunicativa.

5.2 Analisi quantitativa di testi tratti da conversazioni WhatsApp

In riferimento all'ipotesi, accennata nel capitolo precedente, secondo cui l'utilizzo da parte dei facilitati di strutture simili a quelle del facilitatore di riferimento, è interessante considerare questa appendice allo studio quantitativo che si propone di misurare se e quanto la natura del dialogo implica un adeguamento da parte degli interlocutori allo stile e alle forme linguistiche delle persone con le quali si sta conversando.

5.2.1 Il corpus WA

Le conversazioni WhatsApp di un gruppo di 14 persone, amiche tra loro, sono state raccolte in un corpus denominato *corpus WA*, per un totale di 30 testi. Le conversazioni fanno riferimento a un lasso temporale piuttosto variabile, parametro che necessariamente influenza la lunghezza dei testi. Per questo esperimento sono stati utilizzati testi che contassero più di 1500 parole. Le operazioni preliminari sul corpus si sono sviluppate in due direzioni per rispondere a due esigenze di ricerca: i) i riferimenti ai nomi propri sono stati eliminati per evitare che, nell'analisi, i testi, soprattutto i più brevi, venissero classificati sulla base della parola che si riferisce al nome, tra le più frequenti all'interno delle conversazioni; ii) dal momento che l'analisi è centrata sulle occorrenze e sulla frequenza delle occorrenze, sono stati cancellati i tratti grafici che davano vita a *emoji*. Ciascun testo, infine, è stato codificato nel modo seguente: nome scrivente_ nome destinatario (i nomi utilizzati sono nomi di fantasia).

5.2.2 Metodi

I testi del corpus WA sono sottoposti a una analisi per gruppi, attraverso il pacchetto *Stylo*, implementato in R. Sono state considerate le mille parole più frequenti (1000 MFW) e, come nei casi precedentemente trattati si è utilizzata la distanza *Cosine Delta*.

5.2.3 Risultati

I risultati dell'analisi per gruppi, riportati dalla figura 10, mostrano una forte tendenza da parte di chi scrive ad assomigliare nella scelta delle parole alla persona con la quale sta intrattenendo una conversazione. Questa somiglianza è espressa graficamente, nel dendrogramma, dall'appaiamento in funzione della chat e quindi del reciproco interlocutore anziché in funzione di una propria identità e coerenza stilistica.

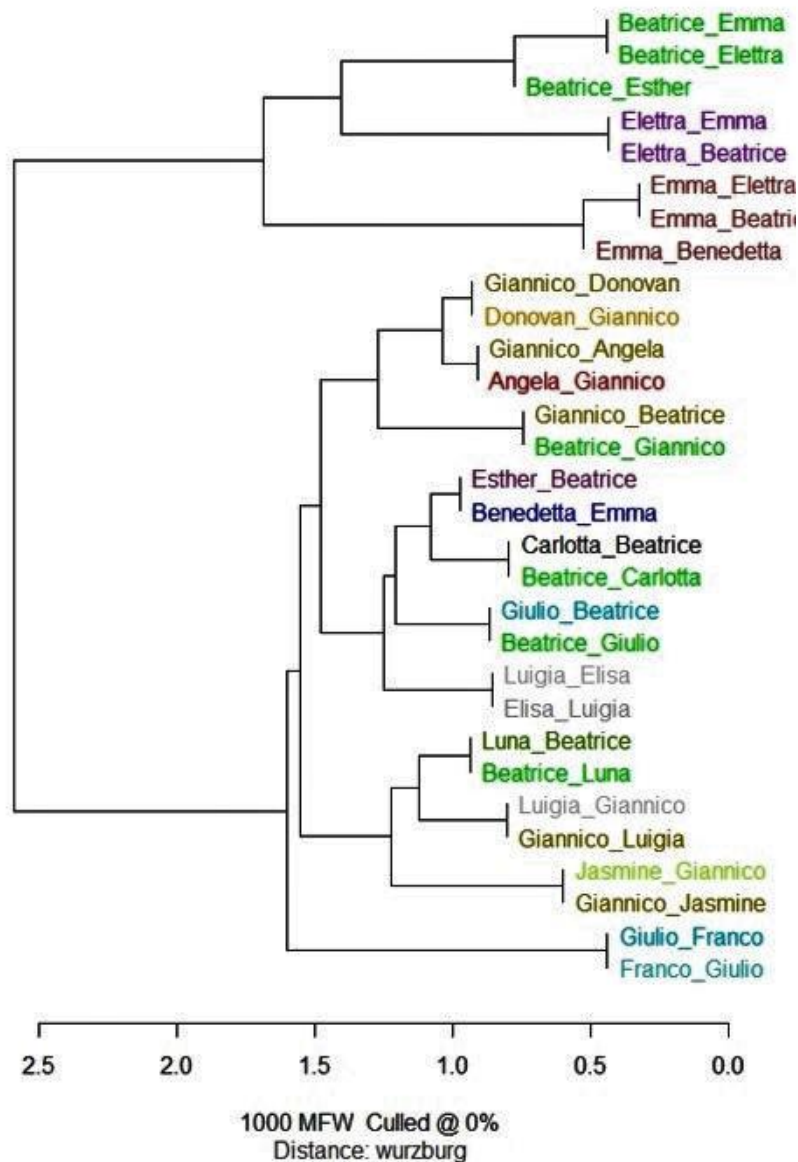


Figura 10: Analisi per gruppi di testi ricavati da conversazioni whatsapp.

Giannico ad esempio dimostra di avere uno stile molto simile a quello di *Donovan* quando scrive a *Donovan*, e uno stile molto simile a *Jasmine* (e molto diverso da sé stesso quando conversa con *Donovan*) quando dialoga con *Jasmine*.

È interessante notare come questo adeguamento allo stile comunicativo del proprio interlocutore si manifesti anche in conversazioni che coprono un arco temporale piuttosto ampio (alcune conversazioni si distendono lungo un periodo di cinque anni) e difficilmente potrebbero essere frutto di un semplice effetto “eco”.

Questi dati sono, *in primis*, importanti in un’ottica di interpretazione dei risultati ottenuti con i testi scritti in CF che, va ricordato, sono composti interamente da dialoghi e conversazioni. Alla luce dei risultati con il *corpus WA* l’ipotesi secondo cui l’uniformità di certe forme lessicali o sintagmatiche in scriventi con lo stesso facilitatore possa essere il segnale di un adeguamento comunicativo assume un forte valore interpretativo.

Allo stesso modo, le operazioni svolte sul *corpus WA* aprono a una nuova possibile fase di indagine sui metodi quantitativi di analisi dei testi. Il dialogo infatti sembra opacizzare i tratti stilistici propri di uno scrivente a favore di uno stile condiviso tra i due interlocutori. Ricerche e studi futuri dovranno approfondire il comportamento e l’efficacia dei sistemi quantitativi di attribuzione d’autore anche in testi costituiti da dialoghi o scambi comunicativi tra più scriventi.

Capitolo 6

6.1 Discussione

6.1.1 Indici di autorialità e indici di influenzamento.

La Comunicazione Facilitata e l'inerente dibattito sulla questione dell'autorialità degli scritti risultano un rebus di difficile soluzione. Numerose sono infatti le variabili che devono essere coinvolte nell'analisi di ciò che sta succedendo durante il processo di facilitazione, molteplici sono le sfaccettature con cui poter caratterizzare questa strategia di comunicazione aumentativa, altrettanto copiose sono le angolazioni da cui poter osservare e descrivere il fenomeno. Nel corso degli anni il giudizio sulla CF si è diviso a seconda della prospettiva di indagine utilizzata, senza riuscire a convergere su un unico quadro interpretativo, esacerbando le difese di ciascuna fazione. L'angolazione adottata da questo lavoro di ricerca si colloca sulla scia degli studi di Ashby (2018), Emerson (2010), Duchan (1999) e Zanobini (2003); per prime infatti queste ricercatrici hanno proposto una modifica del quesito di indagine riguardo alla CF, slegato dalla logica del funziona-non funziona, spesso declinata in termini assoluti, e più orientato a rispondere alla domanda: *“può la CF essere funzionale al miglioramento delle potenzialità comunicative di un individuo?”*. È all'interno di questa cornice che si sono sviluppate le analisi qualitative e quantitative di questo progetto, con l'obiettivo non tanto di proporre una valutazione definitiva e conclusiva sulla questione dell'autorialità, quanto di approfondirla apportando nuovi elementi al dibattito.

I risultati a cui è giunto questo lavoro di tesi, nella parte più conservativa dell'analisi proposta, hanno sostanzialmente confermato le conclusioni a cui erano giunti gli studi precedenti (in particolare progetto EASIEST 2008 e Emerson 2010). I testi scritti con la comunicazione facilitata infatti sembrano aderire a dei caratteri comuni che, dal momento che sono stati osservati in studi internazionali e con lingue diverse dall'italiano, possono essere considerati come degli universali linguistici; l'utilizzo di una scrittura creativa, ricca di forme inusuali, di neologismi, di aggettivi e larga nell'occorrenza degli avverbi (specialmente gli avverbi in *-mente*) si riscontra, infatti, anche nel corpus indagato da questo studio. Va tuttavia sottolineato, dal momento che sarà ripreso nel paragrafo in cui si tratterà dell'influenzamento del facilitatore, come i

testi del *corpus 2018* presentino una diversa tendenza dal punto di vista della sintassi: le strutture sintattiche più marcate, definite in Bernardi (2008) come “a motore semantico”, seppur ancora presenti all’interno del *corpus 2018*, compaiono in misura minore. È stato ipotizzato che questa riduzione della marcatezza della costruzione frasale sia in realtà il frutto di un processo di adeguamento alle forme più convenzionali della lingua, interno al centro in cui sono stati raccolti i dati, con lo scopo di rendere meno drastico lo scarto dalla lingua comunemente utilizzata, in un’ottica di maggior inclusione. Come si è avuto modo di vedere la constatazione di una differenza stilistica dei testi scritti con la CF rispetto a qualsiasi altro tipo di testo non è sufficiente a orientare un’opinione neutra sul tema dell’autorialità (cfr. cap.3). L’analisi semi-qualitativa delle corrispondenze di lemmi e sintagmi ha ribadito le osservazioni proposte da Emerson (2010) dimostrando come nei testi vi siano delle tipicità riconducibili alla persona del facilitato e altri tratti riconducibili al facilitatore. L’importanza di questi dati sta nell’aprire alla possibilità che le due componenti tradizionalmente opposte all’interno del dibattito scientifico, autorialità e influenzamento, siano in realtà compresenti all’interno del processo dinamico che è la facilitazione, come due facce della stessa medaglia. L’individuazione di aspetti linguistici specifici di un facilitato, che occorrono con più di un facilitatore, rappresenta un indizio difficilmente confutabile circa l’effettiva efficacia della tecnica nell’aumentare le possibilità comunicative di chi la utilizza; a maggior ragione se si considera che la maggior parte di questi aspetti occorrono in un numero limitato di volte all’interno di un corpus di grandi dimensioni, riducendo la possibilità che queste occorrenze specifiche siano in realtà frutto di una manipolazione dei facilitatori. La sezione di questa tesi dedicata all’analisi quantitativa porta nuovi elementi a questa argomentazione. Infatti, i testi scritti in CF sono stati sottoposti a un riconoscimento dell’autorialità per mezzo di consolidati metodi quantitativi. L’analisi per gruppi raffigura chiaramente e in misura sorprendentemente netta come dai testi di ciascun utilizzatore della CF, scritti con diversi assistenti, sia possibile identificare una coerenza stilistica e una unicità per ogni facilitato. Similarmente gli studi condotti con SVM e Random Forest, con diversi gradi di accuratezza, dimostrano come per alcuni facilitati sia possibile giungere a un indizio di autorialità.

Il quadro generale, dato dall'integrazione dei diversi esperimenti condotti sul corpus, rende più difficili da argomentare quelle ipotesi secondo cui la CF è unicamente il sottoprodotto dell'influenza del facilitatore. Tutti gli scriventi oggetti di questo studio, infatti, hanno almeno un tipo di analisi, quantitativa o qualitativa che sia, che certifichi la presenza di un tratto proprio, personale e originale all'interno degli scritti. Sulla base di ciò è possibile dunque affermare che per i partecipanti di questa ricerca la Comunicazione Facilitata è stata in grado di aumentare le possibilità e le capacità comunicative.

Giunti a questo primo, fermo caposaldo di questo lavoro è opportuno fare qualche considerazione sull'influenza dei facilitatori negli scritti. L'influenzamento degli assistenti alla comunicazione, alla pari dell'autorialità degli utilizzatori della CF, trova conferme in quasi ogni esperimento condotto (fa eccezione l'analisi per gruppi dove però i facilitatori non rappresentano una variabile di indagine) e per ogni facilitato, anche quelli che mostrano maggiori indizi di autorialità. Così come alcuni facilitati sembrano mostrare un maggior grado di indipendenza nella scrittura, alcuni facilitatori sembrano essere più inclini a influenzare i propri assistiti. Non è chiaro, e dovrebbe essere motivo di ricerche più approfondite, se questa tendenza sia in realtà causata da uno squilibrio nella dimensione del corpus (il facilitatore 1 conta dieci testi mentre il facilitatore 4 solo due) o se invece sia dovuta alle caratteristiche di personalità del facilitatore (come ipotizzato da Emerson, 2010). Non è tuttavia in dubbio che il facilitatore abbia un ruolo all'interno del processo di scrittura. La comunità scientifica è concorde nel classificare l'influenzamento presente nella CF come un esempio dell'effetto ideomotorio. Come già anticipato nel cap. 2 tentare di ridurre l'intero processo di facilitazione a effetto ideomotorio porta con sé dei corollari difficilmente giustificabili. Alla luce dei risultati di questa tesi non è verosimile che tutto il processo di facilitazione sia di natura ideomotoria; come potrebbe rendere conto delle specificità individuali rinvenute nei facilitati che emergono soprattutto dall'analisi della lunghezza media dell'enunciato? Se l'intera produzione scritta fosse governata inconsciamente dal facilitatore come potrebbero essere rispettati per ciascun facilitato i tratti della lunghezza dell'enunciato e conseguentemente di programmazione del pensiero e

costruzione sintattica? Il facilitato deve per forza essere una componente attiva e vigile all'interno del processo di scrittura.

È comunque possibile che in alcuni frangenti il fenomeno si verifichi, soprattutto in quelle situazioni in cui la parola o la lettera sia facilmente intuibile oppure nei contesti linguistici in cui si richiede una formula, come ad esempio nei saluti o nell'introduzione di un concetto (cfr. "io esibisco" che compare solo con chi scrive con 1). I risultati dello studio condotto sulle conversazioni "Whatsapp" introducono un ulteriore piano interpretativo alla questione dell'influenzamento. L'analisi per gruppi condotta con i testi tratti da conversazioni quotidiane dimostra come anche nello scritto si possa verificare un *accomodamento comunicativo* da parte degli interlocutori. Questo fenomeno, naturale e comune a tutti i parlanti, trasportato nel mondo della CF potrebbe verificarsi come una convergenza da parte degli scriventi verso le forme sintattiche e semantiche tipiche del proprio facilitatore. Questa ipotesi potrebbe essere avvalorata dal fatto che i facilitati che risentono in misura minore dell'influenzamento sono gli stessi che non hanno un unico facilitatore di riferimento ma che scrivono o hanno scritto con buona frequenza con più di un assistente (A, C, E). In un'ottica di miglioramento e perfezionamento della tecnica questo potrebbe essere uno spunto prezioso.

6.1.2 Limiti e prospettive future

Il superamento dei limiti che hanno contraddistinto questo studio potrebbe essere un ottimo punto di partenza per ulteriori approfondimenti sul tema della CF. Poter costruire il corpus di indagine fin dall'inizio controllando aspetti quali la dimensione del testo o la distribuzione dei diversi facilitatori permetterebbe di dirimere alcuni dubbi che il *corpus 2018* non può risolvere e dovuti per l'appunto al disequilibrio nella dimensione dei testi del corpus e alla diversa rappresentazione dei facilitatori. In questo progetto non è stato possibile procedere in questo modo dal momento che la raccolta di un nuovo corpus avrebbe richiesto molto più tempo di quello a disposizione.

Sarebbe inoltre di grande interesse il confronto tra i dati qui presentati e i risultati degli esperimenti con il tracciamento oculare che si stanno conducendo presso la Nottingham Trent University per vedere se e come si manifestano a livello dei movimenti oculari l'indipendenza dei facilitati e l'influenzamento dei facilitatori. Infine, la ricerca sulla CF

gioverebbe di un progetto di ricerca basato sull'analisi dei potenziali evocati o più in generale dell'attività cerebrale durante il processo di facilitazione. Uno studio di questo tipo, se integrato con le informazioni raccolte negli ultimi trent'anni, riuscirebbe a fornire un quadro completo e, si auspica, conclusivo sulla questione CF.

Se si considera invece quella branca della linguistica che si occupa dello studio dei sistemi quantitativi di analisi del testo nell'attribuzione d'autore, una delle prospettive che questa tesi sembra aprire riguarda l'applicazione di questi strumenti a testi formati da dialoghi. Come discusso in più di un'occasione, sembra che in un contesto di scambio comunicativo gli stili di scrittura dei due interlocutori si avvicinano, causando una perdita o una attenuazione dell'impronta digitale stilistica di ciascuno scrivente, palesando di conseguenza una scarsa efficacia dell'analisi quantitativa. La ripetizione dell'analisi per gruppi con un corpus di dialoghi più ampio ed equilibrato potrà fornire ulteriori chiarimenti in questa direzione.

6.2 Conclusioni

In conclusione, gli studi presentati in questo progetto sembrano confermare l'esistenza di una terza via all'interno del dibattito sulla CF; una via che ammette sia l'influenzamento dei facilitatori sia l'indipendenza dei facilitati e li considera come aspetti compresenti nella dinamica di interazione scrivente-assistente. Non è possibile dare un giudizio conclusivo sulla tecnica in termini di funziona-non funziona, anche se, contestualizzando G.K. Chesterton, la prima opzione risulta probabilmente la più percorribile dal momento che la prova, anche più piccola, dell'esistenza di indizi di autorialità comporta l'automatica esclusione della seconda. È tuttavia chiara la necessità di una riformulazione e riconsiderazione dei termini con cui riferirsi alla CF. L'effettiva dimostrazione di una possibilità di influenzamento dovrebbe spingere a un atteggiamento di cautela nell'adozione della tecnica, non ad un suo abbandono, considerati gli aspetti positivi possibili, ma ad un utilizzo che tenga presente il rischio e la possibilità di influenzamento e che di conseguenza sia mirato il più possibile all'autonomia e all'indipendenza della persona.

Bibliografia

Abbasi, A. and Chen, H. "Applying authorship analysis to extremist-group web forum messages," *IEEE Intelligent Systems*, vol. 20, no. 6, pp. 67–75, 2005.

American Psychological Association, Nov. 20, 2003, <http://www.apa.org/research/action/facilitated.aspx>

American Speech-Language-Hearing Association. (1995). Facilitated communication [Position Statement]. Available from www.asha.org/policy.

Argamon S, Whitelaw C, Chase P, & Hota S. (2007). Stylistic text classification using functional lexical features. *Journal of the American Society for Information Science and Technology*, 58 (6), 802-822.

Ashby, C. (2018). Il bisogno di riformulare la ricerca scientifica sulla CF. *Comunico ergo sum. La Comunicazione Facilitata: nuove ipotesi sperimentali e prove di validazione attraverso message passing, analisi linguistica, EMG e eye trackers*. Reggio Emilia.

Benedetto, D., Caglioti, E., & Loreto, V. (2002). Language trees and zipping. *Physical Review Letters*, 88(4), 048702

Berger CL (1992): *Facilitated communication guide*. Unknown Binding

Bernardi, L. & Tuzzi, A (2011) Analyzing Written Communication in AAC Contexts: A Statistical Perspective, *Augmentative and Alternative Communication*, 27:3, 183-194.

Bernardi, L. (2008). *Il delta dei significati. Uno studio interdisciplinare dell'espressione autistica*. Roma: Carrocci Faber.

Bernardi, Lorenzo & Tuzzi, Arjuna. (2011). *Statistical Analysis of Textual Data from Corpora of Written Communication. New Results from an Italian Interdisciplinary Research Program (EASIEST)*.

Biklen D., Morton M. W., Saha S. N., Duncan J., Gold D., Hardardottir M., Kama E., O'Connor S., Rao S. (1991). "I AMN NOT A UTISTIVC ON THJE TYP" *Disability, Handicap and Society*, 6, 46–57

Biklen, D (1990). "Communication Unbound: Autism and Praxis". *Harvard Educational Review*. 60 (3): 291–314.

Biklen, D., & Cardinal, D. N. (Eds.). (1997). *Special education series. Contested words, contested science: Unraveling the facilitated communication controversy*. New York, NY, US: Teachers College Press.

Bober, A. (2010). *Zur Wirkungsweise der körperlichen Stütze während der Gestützten Kommunikation (FC). Analyse des Forschungsstands und Ableitung weiterführender Forschungsfragen*. Dissertation.

Bozkurt, D., Baglıoğlu, O., & Uyar, E. (2007), "Authorship Attribution: Performance of Various Features and Classification Methods" *Computer and Information Sciences*

Bruno, M.A., Schakers, C., Vanhauzenhuysse, A., Moonen, G., Laureys, S. (2010). *Facilitated communication in severe traumatic brain injury*. Twentieth Meeting of the European Neurological Society 19-23 June 2010 (p. S84). Berlin: *J Neurol* (2010) 257 (Supp 1): S1:S246.

Burrows, J.F. (1987). *Word patterns and story shapes: The statistical analysis of narrative style*. *Literary and Linguistic Computing*, 2, 61-70.

Burrows, J.F. (1992). *Not unless you ask nicely: The interpretative nexus between analysis and information*. *Literary and Linguistic Computing*, 7(2), 91–109.

Clement, R., & Sharp, D. (2003). *Ngram and Bayesian classification of documents for topic and authorship*. *Literary and Linguistic Computing*, 18(4), 423-447

Cortelazzo, M. (2008). *Ricchezza del lessico nel corpus EASIEST* In "Il delta dei significati" Bernardi, L. 2008, Carrocci Faber. Roma. 66-79

Cortelazzo, M. (2013). *Metodi qualitativi e quantitativi di analisi dei testi*. *Contemporanea: rivista di studi sulla letteratura e sulla comunicazione*, 16(2), 299-310.

De Vel, O., Anderson, A., Corney, M., & Mohay, G. (2001). *Mining e-mail content for author identification forensics*. *SIGMOD Record*, 30(4), 55-64.

- Di Benedetto, C. (2018) Isole semantiche nell'emozionale mare: un'analisi lessicale, retorica e sintattica. In "Il delta dei significati" Bernardi, L. 2008, Carrocci Faber. Roma. 80-96
- Diederich, J., Kindermann, J., Leopold, E., & Paass, G. (2003). Authorship attribution with support vector machines. *Applied Intelligence*, 19(1/2), 109-123.
- Duchan, J. F. (1999). Views of Facilitated Communication: What's the point? *Language, Speech, and Hearing Services*, 30, 401-407
- Eder, M. (2015). Does size matter? Authorship attribution, small samples, big problem. *Digital Scholarship in the Humanities*, 30(2): 167-82.
- Eder, M. and Rybicki, J. (2013). Do birds of a feather really flock together, or how to choose test samples for authorship attribution. *Literary and Linguistic Computing*, 28(2): 229-36.
- Eder, M., Rybicki, J. and Kestemont, M. (2016). Stylometry with R: a package for computational text Analysis. "R Journal", 8(1): 107-121.).
- Emerson, A. (2010). Analyse der bei FC verwendeten Wörter als Indikator für Autorenschaft und Einflussnahme bei der Gestützten Kommunikation. In A. Alfaré, Th. Huber-Kaiser, F. Janz & Th. Klauß (Hrsg.), *Facilitated Communication. Forschung und Praxis im Dialog* (S. 44-50). Karlsruhe: von Loeper
- Emerson, A., Grayson, A., & Griffiths, A. (2001). Can't or won't? Evidence relating to authorship in facilitated communication. *International Journal of Language and Communication Disorders*, 36(1), 98-103.
- F. & Wallace D. L. (1964). *Applied Bayesian and Classical Inference. The Case of Federalist Papers*. Reading (MA): Addison-Wesley.
- G, Stamatatos E, Gritzalis S. & Katsikas, S. (2006). Effective identification of source code authors using byte-level information. *Proceedings of the 28th International Conference on Software Engineering*, (p. 893-896).

Gamon, M. (2004). Linguistic correlates of style: Authorship classification with deep linguistic analysis features. In Proceedings of the 20th International Conference on Computational Linguistics (pp. 611-617).

Grayson, A., Emerson, A., Howard-Jones, P., & O'Neil, L. (2012). Hidden communicative competence: Case study evidence using eye-tracking and video analysis. *Autism*, 16(1), 75-86.

Grieve, J. (2007). Quantitative authorship attribution: An evaluation of techniques. *Literary and Linguistic Computing*, 22(3), 25

Hirst G. & Feiguina O. (2007). Bigrams of syntactic labels for authorship discrimination of short texts. *Literary and linguistics computing*.

Hoover, D. (2004a). Testing Burrows' Delta. *Literary and Linguistic Computing*, 19(4), 453-475.

ISAAC position statement on facilitated communication. International Society for Augmentative and Alternative Communication. (2014). *International Society for Augmentative and Alternative Communication*. (4):357-8

J & Eriksson G. (2007). Authors, genre, and linguistic convention. *Proceeding of the SIGIR Workshop on Plagiarism Analysis, Authorship Attribution, and Near-Duplicate Detection*, (p. 23-28)

Janzen-Wilde, M.L., Duchan, J.F., Higginbotham, D.J. (1995). Successful use of facilitated communication with an oral child. *Journal of Speech & Hearing Research*, 38(3), 658-676.

Joula, P. (2006). Authorship attribution. *Foundation and Trends in Information Retrieval* Vol 1, No.3, 233-334.

Khonji, Mahmoud & Iraqi, Youssef & Jones, Andy. (2015). An Evaluation of Authorship Attribution Using Random Forests.

Kniffka, H. (1996). On forensic 'differential diagnosis'. In H. Kniffka (ed.), *Recent developments in forensic linguistics* (pp. 75-121). Frankfurt/Main: Peter Lang.

Koppel, M., & Schler, J. (2003). Exploiting stylistic idiosyncrasies for authorship attribution. In Proceedings of IJCAI'03 Workshop on Computational Approaches to Style Analysis and Synthesis (pp. 69-72).

Koppel, M., Schler, J., & Bonchek-Dokow, E. (2007). Measuring differentiability: Unmasking pseudonymous authors. *Journal of Machine Learning Research*, 8, 1261-1276.

Megarry J., (1983). *World Yearbook of Education. Computers and Education*

Mendenhall, T. C. (1887). The Characteristic Curves of composition. *Science*, 11, 237-249.

Moretti, F. (2005). *La letteratura vista da lontano*, Torino, Einaudi, 2005, p. 3.

Niemi, J., & Kärnä-Lin, E. (2002). Grammar and lexicon in facilitated communication: A linguistic authorship analysis of a Finnish case. *Mental Retardation*, 40, 347–357.

Peng, F., Shuurmans, D., & Wang, S. (2004). Augmenting naive Bayes classifiers with statistical language models. *Information Retrieval Journal*, 7(1), 317-345.

Rudman, J (1998). "The state of authorship attribution studies: Some problems and solutions," *Computers and the Humanities*, vol. 31, pp. 351–365.

Saloviita T, Leppänen M & Ojalampi U. (2014) Authorship in Facilitated Communication: An Analysis of 11 Cases, *Augmentative and Alternative Communication*, 30 (3), 213-225

Saloviita, T. (2016). Does Linguistic Analysis Confirm the Validity of Facilitated Communication? *Focus on Autism and Other Developmental Disabilities*. 33..

Sanderson C. & Guenter S. (2006). Short text authorship attribution via sequence kernels, Markov chains and author unmasking: An investigation. *Proceedings of the International Conference on Empirical Methods in Natural Language Engineering*, (p. 482 - 491).

Scopesi, A., Zanobini, M., Cresci, L.R. (2003). Aspetti semantici e stilistici della produzione di un bambino autistico in situazione di comunicazione facilitata. *Rivista di psicolinguistica applicata*, 2-3, 83,85.

- Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys*, 34(1).
- Srinivasa, O & Vijaya, V. (2017). Authorship Attribution on Imbalanced English Editorial Corpora. *International Journal of Computer Applications*. 169. 44-47.
- Stamatatos, E. (2006a). Authorship attribution based on feature set subsampling ensembles. *International Journal on Artificial Intelligence Tools*, 15(5), 823-838
- Stamatatos, E. (2008). Author identification: Using text sampling to handle the class imbalance problem. *Information Processing and Management*, 44(2), 790-799
- Stamatatos, E. (2009), A survey of modern authorship attribution methods. *J. Am. Soc. Inf. Sci.*, 60: 538-556.
- Stamatatos, E., Fakotakis, N., & Kokkinakis, G. (2000). Automatic text categorization in terms of genre and author. *Computational Linguistics*, 26(4), 471–495, 2000.
- Stefan Evert, Thomas & Proisl & Jannidis, Fotis & Reger, Isabella & Pielström, Steffen & Schöch, Christof & Vitt, Thorsten. (2017). Understanding and explaining Delta measures for authorship attribution. *Digital Scholarship in the Humanities*. 32. 4-16.
- Teng, G., Lai, M., Ma, J., & Li, Y. (2004). E-mail authorship mining based on SVM for computer forensic. In *Proceedings of the International Conference on Machine Learning and Cybernetics*, 2 (pp. 1204-1207).
- Tuzzi, A. (2003). *L'analisi del contenuto*. Roma: Carrocci.
- Tuzzi, A., Cemin, M., & Castagna, M. (2004). "Moved deeply I am". Autistic language in texts produced with FC. In G. Purnelle, C. Fairon, & A. Dister (Eds.), *JADT 2004. 7es Journées internationales d'Analyse statistique des Données Textuelles* (pp. 1097-1105).
- Tuzzi, A. (2008) "Diversità statistiche. Un'analisi comparata" In "Il delta dei significati" Bernardi, L. 2008, Carrocci Faber. Roma. 37-55.
- Tuzzi, A. (2009). Grammar and lexicon in individuals with autism: A quantitative analysis of a large Italian corpus. *Intellectual and Developmental Disabilities*, 47, 373–385.

Uzuner, O., & Katz, B. (2005). A comparative study of language models for book and author recognition. In *Proceedings of the 2nd International Joint Conference on Natural Language Processing* (pp. 969-980) Springer.

van Halteren, H. (2007). Author verification by linguistic profiling: An exploration of the parameter space. *ACM Transactions on Speech and Language Processing*, 4(1), 1-17.

Wegner, D.M., Fuller V.A., Sparrow B. (2003). Clever hands: Uncontrolled Intelligence in Facilitated Communication. *Journal of Personality and Social Psychology*. 85(1), 5-19.

Wehrenfennig, A., Surian, L. (2008). Autismo e comunicazione facilitata: una rassegna degli studi sperimentali. *Psicologia clinica dello sviluppo*, XII, n.3.

Yule, G. U. (1938). On sentence-length as a Statistical Characteristic of Style in Prose, with Application to Two Cases of Disputed Authorship. *Biometrika*, 30, 363-390.

Yule, G. U. (1944). *The Statistical Study of Literary Vocabulary*. Cambridge: Cambridge University Press.

Zanobini, M., & Scopesi, A. (2001). La comunicazione facilitata in un bambino autistico. *Psicologia Clinica dello Sviluppo*, 3, 395-421.

Zhao Y., & Zobel, J. (2005). Effective and scalable authorship attribution using function words. In *Proceedings of the 2nd Asia Information Retrieval Symposium*.

Zheng R., Qin Y., Huang Z., Chen H. (2003) Authorship Analysis in Cybercrime Investigation, 61. In: Chen H., Miranda R., Zeng D.D., Demchak C., Schroeder J., Madhusudan T. (eds) *Intelligence and Security Informatics*. ISI 2003. *Lecture Notes in Computer Science*, vol 2665. Springer, Berlin, Heidelberg

Zipf, G. (1932). *Selected Studies of the Principle of Relative Frequency in Language*. Cambridge, MA: Harvard University.