

UNIVERSITA' DEGLI STUDI DI PADOVA

FACOLTA' DI SCIENZE STATISTICHE

Corso di laurea triennale in
STATISTICA, POPOLAZIONE E SOCIETA'

TESI DI LAUREA

**Quando le parole erano rosse.
Un'analisi statistica del contenuto attraverso le parole
dei volantini della sinistra padovana degli anni '60.**

Relatore:
Prof. Lorenzo Bernardi

Laureando:
Mattia Peretti
Matricola n. **573157**

ANNO ACCADEMICO 2010/11

*ad Alberto Landi,
che mi ha insegnato che anche con le figurine si impara la matematica.*

INDICE

Capitolo primo

INTRODUZIONE

1. Il contesto pag. 1
2. L'occasione: lo stage pag. 3
3. Obiettivi e risultati attesi pag. 4
4. Lo strumento: TaLTaC² pag. 5
5. Riassunto pag. 6

Capitolo secondo

ANALISI DEL CORPUS PER FORME GRAFICHE SEMPLICI

1. Creazione del vocabolario pag. 9
2. Divisione del *corpus* in *subcorpora* pag. 11
3. Presentazione del vocabolario pag. 16
4. Conteggio e lettura delle forme grafiche ed esclusive pag. 21
 - 4.1 Forme grafiche comuni pag. 23
 - 4.2 Forme grafiche esclusive pag. 28
5. Conteggio e lettura delle specificità pag. 30
6. Estrazione di informazione per parole-chiave pag. 35
 - 6.1 La violenza attraverso le armi pag. 36
 - 6.2 I “verbi violenti” pag. 38

Capitolo terzo

ANALISI DEL CORPUS PER UNITA' TESTUALI COMPLESSE

1. Nuova normalizzazione e novità nel vocabolario pag. 43
2. Estrazione dei segmenti ripetuti e lessicalizzazione pag. 44
3. Sub-occorrenze, segmenti comuni ed esclusivi pag. 50
 - 3.1 La violenza attraverso le armi pag. 51
 - 3.2 I “verbi violenti” pag. 53
4. Segmenti specifici pag. 55

Capitolo quarto

ANALISI DELLE CORRISPONDENZE pag. 61

Capitolo quinto

CONCLUSIONE pag. 65

Bibliografia pag. 69

Siti internet consultati pag. 69

Ringraziamenti pag. 71

CAPITOLO PRIMO:

Introduzione

1. IL CONTESTO

Padova, fine anni '60.

In un contesto politico e sociale caratterizzato da un dualismo opprimente che contrappone lavoratori e padroni, studenti e potenti, comunisti e democristiani, creando i presupposti per proteste, lotte e richieste di innovazione, la comunicazione si presenta come un'arma, uno strumento dalle infinite possibilità sul quale riporre speranze e desideri di cambiamento.

Un cambiamento fortemente sentito da molti come necessario, per modificare una situazione dove niente sembra operare per il bene comune, dove i rappresentanti non rappresentano e la società non si occupa del benessere sociale, dove chi sta ai piani bassi trova in questo dissesto sociale e politico la spinta giusta per arrabbiarsi e salire le scale per andare ad incontrare, o meglio 'scontrare', i piani alti e farli scendere, almeno di qualche gradino.

La parola-chiave diventa allora 'partecipazione': cittadini che stanchi di essere spettatori si trasformano in attori per contribuire al cambiamento e si muovono alla ricerca di un palcoscenico adatto, un teatro con un'acustica perfetta, per farsi sentire davvero da tutti.

La partecipazione però non può agire, non può realizzarsi, senza 'organizzazione' e cresce quindi nel cuore di molti anche la necessità impellente di ritrovarsi, dentro uno spazio e dietro un nome o una sigla, per creare e confrontare idee da comunicare "fuori", ad un esterno a volte disponibile e impaziente di ricevere quelle idee sotto forma di parole, altre volte sordo, per pigrizia o più spesso per scelta, perché sintonizzato su frequenze differenti.

In questo contesto politico/sociale/comunicativo, in città e nella provincia padovana (come in tutta Italia) il Partito Comunista Italiano tenta di diffondere il proprio pensiero e di acquisire consenso, attraverso una voce che parte da numerose fonti per raggiungere un bacino più vasto possibile, grazie ad un fondamentale ed unico “mezzo”: la carta. Ed ecco allora che il volantino d’informazione, o di propaganda, può decollare da diverse mani: da quelle di partito, nazionale, provinciale, o locale, da quelle dei giovani e dei lavoratori, degli studenti e dei cittadini comuni, dalle loro organizzazioni, assemblee e comitati.

Grazie al fondamentale impegno di alcuni, tutto questo resta nella memoria collettiva, a livello di ricordo, ma soprattutto di archivio, offrendo oggi a tutti la possibilità di analizzare e scoprire i contenuti, le forme e il linguaggio della comunicazione politica, sociale e sindacale della sinistra padovana e nazionale e di tutte quelle associazioni e organizzazioni che orbitavano intorno ad essa.

Resta come ricordo di un periodo gravido di pensieri nuovi, di voglia di cambiamento e di una giustizia sociale nuova; resta negli archivi, ma anche nei cuori di tanti e ha segnato la società e la cultura degli anni successivi.

La partecipazione ha dato vita a nuovi soggetti sociali (soprattutto i giovani) che hanno occupato il palcoscenico dando vita a nuovi linguaggi che hanno colorato le speranze di cambiamento e, attraverso le mani sporche di inchiostro da ciclostile, hanno creato le premesse ad un periodo di riforme.

Le speranze di quegli anni hanno dato vita, purtroppo, anche a esperienze di violenza negli anni ‘70 (brigate rosse e nere) che comunque non possono oscurare il cambiamento culturale che ha portato alla ribalta i soggetti politici che hanno fatto la storia dei successivi decenni.

L’analisi della parola quindi non vuole avere un significato agiografico di quegli anni: non è necessario per chi guarda con occhio distaccato partecipare e condividere, serve invece capire perché le parole erano rosse anche se, ascoltando le storie di chi quel periodo lo ha vissuto, la nostalgia e la passione che trasmettono fanno nascere un po’ di invidia guardando un mondo in cui la vita è sempre più declinata sul successo e l’apparenza.

2. L'OCCASIONE: LO STAGE

Stabilito e precisato il contesto che la presente relazione si propone di studiare, è necessario partire da lontano per descrivere il percorso che ha portato, passo dopo passo, alla stesura definitiva, cioè dalla scelta e dalla ricerca dell'esperienza di stage, alla quale la relazione finale è strettamente e curricularmente collegata.

Per effettuare tale scelta, ho cercato un'opportunità che fosse di per sé interessante come esperienza di stage e che mi permettesse inoltre di applicare le conoscenze acquisite in uno degli ambiti di studio che hanno maggiormente destato il mio interesse nei tre precedenti anni di istruzione universitaria, cioè l'analisi statistica del contenuto, conosciuta e studiata durante il corso di "Metodi qualitativi d'indagine".

L'idea di partenza è stata appunto quella di collegare l'esperienza di stage e l'interesse per questa disciplina, ricercando una struttura dove poter concentrare il mio operato sulla creazione di un *corpus* costituito da testi interessanti e adatti per un'analisi statistica, ed essere nel frattempo utile alla struttura stessa che avrebbe dovuto ospitarmi.

Con l'aiuto dell' Ufficio Stage di Facoltà, questa idea si è potuta concretizzare, trovando nel "Centro Studi Ettore Luccini" di Padova, una struttura disponibile ad assumermi per tre mesi (350 ore) in qualità di stagista.

Lo stesso centro studi, attraverso il suo sito, così si presenta ai visitatori:

“Siamo, orgogliosamente, l' Archivio Storico e la Biblioteca del movimento operaio, contadino e popolare veneto.

Il nostro **Centro Studi Ettore Luccini (CSEL)** raccoglie infatti da più di vent'anni ogni documentazione reperibile sulla storia del movimento operaio e popolare veneto.”

<http://www.centrostudiluccini.it/associazione/presentazione.htm>

Dal "Centro Studi Ettore Luccini" mi è stata offerta un'opportunità idonea a ciò che cercavo: procedere all'immissione nel loro archivio telematico, di testi di una serie di volantini presenti nel fondo archivistico relativo alla Federazione Provinciale del Partito Comunista di Padova.

I volantini dei quali mi sono occupato sono una piccola parte di tutti quelli presenti, all'interno dell'archivio, in una serie che raccoglie volantini redatti dal 1962 al 1997 dalla Federazione Provinciale di Padova del P.C.I. o da altri organi (politici, sindacali, del mondo civile, ecc.) riconducibili a movimenti operai, contadini, popolari, della provincia padovana.

La serie è composta da 64 buste (faldoni) e io mi sono occupato di tre di esse, relative agli anni: 1962-1968 la prima, 1969 la seconda e 1970 la terza, quindi le tre buste più "vecchie" della serie, per un totale di più di seicento volantini. Il mio incarico è stato quello di ricopiarli, trascrivendoli manualmente parola per parola al computer, oppure, quando possibile, acquisendo il testo per mezzo del sistema *O.C.R. (optical character recognition)*, a partire dai file *pdf* dei volantini precedentemente scannerizzati. In seguito, dopo averli ordinati per data all'interno delle tre buste, ho inserito i testi dei volantini nel catalogo dell'archivio del centro studi, svolgendo così il mio ruolo di stagista e ottenendo nel frattempo tutto il materiale necessario per l'analisi.

3. OBIETTIVI E RISULTATI ATTESI

Questa relazione si propone di analizzare i volantini della sinistra padovana degli anni '60, con l'intento di ricavare la maggior quantità possibile di informazioni sulle tematiche trattate nella comunicazione politica e sociale dell'epoca e sul linguaggio e le forme utilizzate come veicolo dei pensieri e delle posizioni degli scriventi.

Si tratta di un periodo storico caratterizzato dal cambiamento: in seguito al "miracolo economico" si amplia e guadagna spazio il movimento sindacale e inizia a farsi sentire la voce dei lavoratori, come del resto quella dei giovani che con il '68 cercano nuovi spazi e nuove forme di comunicazione.

È il periodo della partecipazione e dell'impegno, aumenta il numero degli attori sul palco e si alza il volume del confronto, cambiano i toni utilizzati per proporsi alla pubblica opinione e cambia il ruolo della "gente", partecipe e non più passivo spettatore.

Tutti questi cambiamenti, uniti all'aumento della comunicazione che diventa sempre più fondamentale per la trasmissione di idee nuove, rendono estremamente interessante l'analisi dei volantini, strumento principe di questo nuovo modo di comunicare e di proporsi, per riconoscere e rilevare le parole che come “parole d'ordine” hanno colorato di rosso un decennio della storia italiana.

4. LO STRUMENTO: TaLTaC²

A livello introduttivo, prima di proseguire con la descrizione delle operazioni che hanno reso possibile questa analisi, è necessario presentare lo “strumento” attraverso il quale è stato possibile svolgere concretamente il lavoro: si tratta di TaLTaC² (Trattamento Automatico Lessicale e Testuale per l'Analisi del Contenuto), un “software per l'analisi automatica di una collezione di testi (*corpus*), finalizzata a descrivere e interpretare il suo contenuto e/o alcune sue proprietà”, sviluppato da Sergio Bolasco, Francesco Baiocchi e Adolfo Morrone (<http://www.taltac.it/it/index.shtml>).

La presentazione del programma utilizzato durante il lavoro era essenziale per poter poi fare riferimento ad esso nei vari passi dell'analisi, riassunti schematicamente nel prossimo paragrafo.

5. RIASSUNTO

Il vero e proprio punto di partenza dell'analisi è costituito dalla creazione del *corpus*, svolta durante lo stage presso il “Centro Studi Ettore Luccini” con la copiatura - a volte manuale, altre volte assistita da strumenti informatici - dei testi dei volantini presenti nei tre faldoni selezionati nel fondo archivistico relativo alla Federazione Provinciale di Padova del P.C.I..

Alla raccolta del materiale, segue poi la “pulizia” dello stesso, necessaria per le esigenze del software utilizzato e per poter creare un vocabolario il più possibile privo di errori (Capitolo 2, paragrafo 1).

Il lavoro preparatorio procede con la divisione del *corpus* in più *subcorpora*, partizioni dell'intero testo create a partire da una variabile precedentemente scelta e categorizzata in un certo numero di modalità.

Completate queste operazioni preliminari, inizia effettivamente l'analisi: prima con il calcolo delle dimensioni del vocabolario composto da tutte le forme grafiche semplici presenti nel *corpus*, poi con l'osservazione della composizione di quest'ultimo in relazione alle partizioni create, per finire con il confronto dimensionale dei *subcorpora* stessi.

Da quantitativa, l'analisi diventa qualitativa, nel momento in cui si vanno ad osservare le forme grafiche comuni a tutti i *subcorpora* e quelle esclusive di ciascuno, iniziando così il confronto riguardante le tematiche trattate - condivise e non - e il linguaggio utilizzato.

Il confronto viene approfondito, anche con l'aiuto di operazioni statistiche, con il calcolo delle specificità: forme grafiche non esclusive, ma particolarmente rilevanti per un *subcorpus*, in virtù di un (più o meno) evidente sovrautilizzo rispetto alle altre partizioni.

L'analisi prosegue poi con l'estrazione di informazione peculiare attraverso una lista di parole-chiave, operazione svolta con il calcolo dell'indice TF-IDF, che permette appunto di trovare i frammenti (volantini) più rilevanti rispetto ad un tema precedentemente scelto.

A questo punto, la prima parte è conclusa e l'analisi si sposta su di un livello differente e più approfondito con il riconoscimento nel *corpus* delle unità testuali complesse (CTU).

Il primo passo consiste nella normalizzazione dei gruppi nominali e delle polirematiche: forme linguistiche composte da più parole, che hanno significato se tenute insieme, mentre perdono gran parte della loro portata informativa se separate.

La disambiguazione delle forme grafiche semplici prosegue poi con l'estrazione dei segmenti ripetuti, sequenze di parole che si ripetono più volte nel *corpus* e vanno quindi considerate anch'esse come unità testuali complesse, unendo insieme le forme che le compongono.

La nuova normalizzazione e l'estrazione dei segmenti ripetuti modificano le dimensioni del *corpus*: variano di conseguenza le misure lessicometriche che devono quindi essere ricalcolate, sia per il corpus intero, che per i subcorpora, per i quali vengono ricalcolate anche le sub-occorrenze.

Gli ultimi due passi di questo secondo livello di analisi consistono infine nel conteggio e lettura dei segmenti comuni, esclusivi e specifici, con le stesse modalità utilizzate in precedenza per le forme grafiche semplici.

A completare il percorso, un approfondimento rappresentato dall'analisi delle corrispondenze, una tipologia di analisi fattoriale adatta alle variabili qualitative (*subcorpora* e segmenti ripetuti nel nostro caso) che ha come scopo quello di sintetizzare e rappresentare sotto forma geometrica osservazioni complesse contenute in una tabella di contingenza, per permettere una visualizzazione grafica dell'associazione tra le variabili considerate.

CAPITOLO SECONDO:

Analisi del *corpus* per forme grafiche semplici

1. CREAZIONE DEL VOCABOLARIO

La descrizione comincia con le operazioni preliminari, svolte immediatamente dopo la conclusione dello stage, necessarie per poter eseguire l'analisi: "pulire" e sistemare il *corpus*, in questo caso la raccolta di testi ricopiati dai volantini, è il primo passo fondamentale, poiché TaLTaC è piuttosto "sensibile" rispetto al materiale che gli si chiede di analizzare. Si potrà capire il significato in questo contesto del termine "sensibile" con il proseguimento della spiegazione.

Innanzitutto, il software richiede, se il *corpus* è stato scritto manualmente, come nel nostro caso, che quest'ultimo sia in formato *.txt* ed è per questo che in fase di trascrizione dei volantini è stato utilizzato il semplice programma Microsoft "Blocco note", adatto proprio per la sua semplicità: utilizza ad esempio un solo semplice apice per tutti i casi in cui la lingua italiana richiede un apostrofo, una virgoletta unica, o il segno dei minuti primi, mentre in altri programmi, come "Microsoft Word" si trovano più tipi di apostrofi, di trattini, di virgolette, che creano confusione a TaLTaC, costringendo il software ad una lettura più complessa del *corpus*.

L'utilizzo del "Blocco note" ha permesso quindi di iniziare, fin dalla prima trascrizione dei volantini, a rendere leggibile il materiale da immettere in TaLTaC; tuttavia gli errori non erano completamente evitabili, sia per quanto riguarda i volantini trascritti manualmente, sia per quelli acquisiti con il sistema O.C.R.: nel primo caso perché è davvero difficile non commettere errori di battitura scrivendo il più velocemente possibile con la tastiera di un computer, nel secondo caso perché il programma, pur riconoscendole come lettere e simboli, legge in realtà dai file *pdf* delle immagini e spesso non è in grado di distinguere, per fare un esempio, una "L" minuscola (l) da una barra verticale (|), in particolare se la qualità del volantino non è delle migliori. Bisogna infatti tener conto che i volantini analizzati sono pur sempre fogli di carta stampati più di quarant'anni fa, a volte sbiaditi, squalciti, o spiegazzati.

Per proseguire l'opera di "pulizia" dei testi, un altro strumento di fondamentale importanza ed efficacia è stato il sistema di controllo ortografico offerto da "Word", grazie al quale è stato possibile riconoscere e correggere molti errori di battitura (decisamente più numerosi di quanto si potrebbe pensare), l'uso errato degli accenti (in primis, l'accento grave al posto di quello acuto in "perché", "né", ecc.) e gli ancor più numerosi errori dovuti al sistema O.C.R. descritti in precedenza.

A questo punto, il *corpus* è pronto per essere affidato a TaLTaC, ma questo non vuol dire che sia del tutto sistemato e privo di errori e imperfezioni: una delle primissime risposte che il software offre dopo la lettura del *corpus* infatti è un file ("caratteri.txt") che mostra la lista dei caratteri presenti nel testo, riportando per ciascuno la rispettiva frequenza, e permette quindi di notare se sono ancora presenti (ed era proprio così) caratteri sbagliati per poterli correggere o eliminare, tornando ad utilizzare il "Blocco note".

Citare subito il file "caratteri.txt" era necessario al fine di completare la descrizione del percorso svolto per arrivare ad un *corpus* "pulito" e pronto per l'analisi, tuttavia prima di ottenere tale file, ci sono altre due operazioni da svolgere all'inizio di una sessione di lavoro con TaLTaC: *parsing* e normalizzazione.

La procedura del *parsing* permette di identificare i caratteri alfabetici e i separatori (Giuliano e La Rocca, 2008), accettando la suddivisione scelta convenzionalmente da TaLTaC, oppure modificandola a seconda delle esigenze.

Con la normalizzazione invece si va a correggere ulteriormente, semplificare, e in un certo senso chiarificare, il contenuto del testo da analizzare. Le azioni svolte con la normalizzazione sono principalmente di tre tipi (Giuliano e La Rocca, 2008): trasformare gli apostrofi in accenti quando necessario e regolarizzare gli accenti gravi e acuti in base alle regole della lingua italiana (procedimento già svolto grazie al controllo ortografico di "Word"), eliminare le fonti più frequenti di sdoppiamento del dato (presenza di maiuscole superflue, o comunque non significative), uniformare la grafia di nomi propri, sigle, società e altro ancora.

Di queste funzioni, la terza soprattutto è stata di fondamentale importanza nel nostro caso, perché nei volantini sono estremamente frequenti nomi propri di politici e di altre personalità, sigle di partiti (P.C.I. e D.C. in primis), organizzazioni sindacali,

popolari e studentesche, nomi di Comuni della provincia padovana, nomi di giornali (l'Unità), eccetera.

Per le necessità relative al nostro lavoro di analisi del contenuto, ci si è limitati inizialmente ad una normalizzazione “leggera”, evitando cioè di spuntare le prime due voci del menu presentato da TaLTaC, relative alle liste contenenti le locuzioni grammaticali e i gruppi nominali (e Polirematiche). “Inizialmente” perché procedendo nell'analisi si valuterà l'opportunità di una normalizzazione più approfondita comprendente anche queste due voci.

Il vocabolario con la lista di tutte le forme grafiche presenti nel *corpus* a questo punto sarebbe finalmente pronto, tuttavia il *corpus* stesso, anche se finalmente “pulito” e privo di errori ed imperfezioni, non può ancora considerarsi definitivo ed è quindi consigliabile posticipare la presentazione del contenuto e delle dimensioni del vocabolario, per non essere costretti a correggere i dati presentati, dopo poche pagine.

2. DIVISIONE DEL *CORPUS* IN *SUBCORPORA*

Come si è detto poco fa, il *corpus* non può ancora considerarsi definitivo, questo perché per analizzare più in profondità il contenuto dei volantini, è possibile procedere alla suddivisione del *corpus* in più *subcorpora*, creando cioè delle partizioni dell'intero testo di partenza, in base ad una o più variabili d'interesse. Scopriremo presto che durante il percorso di creazione di questi *subcorpora*, il *corpus*, e di conseguenza il vocabolario, subirà alcune ulteriori modifiche, motivo che ha portato a posticipare la presentazione del vocabolario stesso.

Date le caratteristiche del nostro *corpus*, una possibile variabile d'interesse era facilmente rintracciabile: i volantini infatti, come abbiamo detto in fase d'introduzione, provengono da tre diversi faldoni, omogenei al loro interno proprio in relazione alla data di redazione dei volantini stessi.

Inizialmente quindi, la prima variabile d'interesse presa in considerazione è stata proprio l'anno, presto però si è evidenziata l'inadeguatezza di questa variabile per la creazione dei *subcorpora*, in relazione alle necessità di ricerca: le date infatti sono molto

(troppo?) vicine tra loro, dato che i faldoni raccolgono volantini redatti negli anni 1962-1968 il primo, 1969 il secondo, 1970 il terzo; “troppo” vicine perché, dal punto di vista del contenuto, difficilmente si potranno osservare variazioni di rilievo tra una busta e l'altra. Oltretutto, anche la disomogeneità degli archi temporali descritti dai tre faldoni sconsiglia l'utilizzo di questa variabile, perché mentre il secondo e il terzo faldone raccolgono volantini di un unico anno, il primo faldone è composto invece da volantini di ben sette anni diversi. Questa situazione è dovuta al fatto che la conservazione e soprattutto la raccolta di materiale di così vecchia data, è chiaramente più difficile, quanto più indietro si cerca di andare nel tempo: il primo faldone raccoglie volantini di così tanti anni diversi, proprio per il fatto che sarebbe stato uno spreco dividerli anno per anno in diverse buste che sarebbero rimaste parzialmente vuote. Con il '68 e gli anni immediatamente successivi invece, l'aumento della partecipazione alla vita politica, causato dalle tensioni sociali e dalle lotte operaie e studentesche, ha fatto aumentare notevolmente il numero di organizzazioni politiche e sociali partecipi della situazione nazionale e di conseguenza, aumentando il numero delle fonti, è cresciuto anche il numero di volantini prodotti.

L'unione di questi due fattori - la difficoltà nel raccogliere volantini che è direttamente proporzionale alla “vecchiaia” del materiale stesso, e l'aumento di partecipazione politica e sociale alla fine degli anni '60 - crea un trend riconoscibile non solo nei tre faldoni protagonisti di quest'analisi, ma anche in tutto il fondo archivistico del centro studi: presto si vedrà infatti che anche nella terza busta (1970) si verifica un cospicuo aumento del numero di volantini rispetto alla seconda (1969), mentre una visita all'archivio del centro studi permetterebbe di osservare che l'aumento prosegue anche oltre, tanto che per alcuni degli anni successivi, non basta nemmeno un unico faldone per contenere tutti i volantini.

L'inadeguatezza dell'anno di redazione dei volantini per la creazione dei *subcorpora*, ha reso quindi necessaria la ricerca di un'altra caratteristica più adatta: a inizio relazione si è già spiegato che, all'interno dell'archivio del centro studi, i volantini analizzati appartengono al fondo archivistico relativo alla Federazione Provinciale del Partito Comunista di Padova, ma che sono redatti non solo dalla Federazione stessa, ma anche da altri organi (politici, sindacali, del mondo civile, ecc.) riconducibili a movimenti operai, contadini, popolari, della provincia padovana (pag. 4).

Tutto questo ha suggerito di utilizzare la firma, cioè il/gli autore/i dei volantini, come variabile d'interesse: sfogliando i volantini stessi, si è osservato che era anche possibile ricondurre le numerose firme diverse a quattro gruppi, abbastanza omogenei al loro interno, da utilizzare quindi come modalità della variabile d'interesse per la creazione di quattro *subcorpora*:

1. Federazione Provinciale del P.C.I. di Padova e altri organi di partito, locali e nazionali.

Molti volantini portano la firma della Federazione stessa, mentre tra le altre firme “di partito” che compaiono più spesso, troviamo la Federazione Giovanile Comunista Italiana (F.G.C.I.) e circoli e comitati di zona del partito. Un buon numero di volantini inoltre presenta solo la sigla P.C.I. come firma e sono quindi stati inseriti in questo primo gruppo.

2. Sezioni locali del P.C.I.

Il Partito Comunista aveva la capacità di coprire molto bene il territorio, infatti erano presenti e attive moltissime sezioni locali, di piccoli comuni di provincia o di quartieri e zone delle grandi città. Questo si verificava anche a Padova, infatti molti volantini sono stati redatti dalle sezioni di comuni come Cadoneghe e Piove di Sacco, giusto per portare due esempi, altri invece da sezioni di zone del comune di Padova: il quartiere dell'Arcella ad esempio aveva addirittura due sezioni (Arcella 1° e Arcella 2°).

3. Altre firme, appartenenti al mondo del lavoro.

Le lotte operaie, e dei lavoratori in genere, erano in pieno svolgimento alla fine degli anni '60, dunque non stupisce ritrovare volantini firmati da sindacati, commissioni di fabbrica, comitati di base e di lotta, collettivi operai-studenti, associazioni di lavoratori, ecc..

4. Altre firme, appartenenti al mondo civile.

Com'è noto, in quegli anni non erano solo i lavoratori ad essere in fermento, ma anche il mondo studentesco per cambiare la scuola e l'università e i cittadini comuni per le riforme o contro di esse, quindi numerosi volantini sono firmati da comitati cittadini, associazioni e movimenti studenteschi, cellule universitarie e vari comitati in lotta per un particolare obiettivo.

Bisogna sottolineare che nel formare questi quattro gruppi, spesso è stato necessario fare delle scelte, a volte soggettive e del tutto personali, poiché non era sempre semplice ed immediato capire in quale gruppo inserire i volantini, principalmente per tre ordini di motivi: assenza di firma, firma congiunta di due o più organi o associazioni, firma difficilmente riconducibile a uno dei quattro gruppi.

Può essere utile portare alcuni esempi per spiegare e permettere di comprendere meglio la grande varietà di firme ritrovate nei 661 volantini analizzati e le difficoltà che ne sono derivate.

- Per quanto riguarda i volantini non firmati, si è cercato, quando possibile, di estrarre indicazioni utili dal contesto e dai temi trattati, finendo spesso per inserire questi volantini nel primo gruppo perché scritti con ogni probabilità da un organo di Partito.

- Nei volantini che presentavano almeno due firme congiunte, si è scelto di procedere con metodi differenti a seconda delle situazioni: se le due firme erano riconducibili ad uno stesso gruppo (es. F.G.C.I. e P.C.I.) non si poneva nessun problema; quando invece le due firme avrebbero portato ad inserire il volantino in due gruppi differenti, si è valutato quale delle due firme potesse essere considerata “predominante”, per importanza o in base al tema trattato. La F.G.C.I. ad esempio spesso scriveva volantini in collaborazione con associazioni studentesche, soprattutto per dare una base politica e più forte alle rivendicazioni, quindi questi volantini sono stati inseriti nel primo gruppo, mentre volantini firmati unitamente da comitati cittadini e organizzazioni di fabbrica sono stati inseriti nel gruppo del “mondo del lavoro” perché generalmente erano i cittadini ad appoggiare i lavoratori nelle loro richieste e non viceversa.

Un caso particolare è rappresentato dai volantini in cui le firme erano più di due, molte più di due, trattandosi sempre di inviti a manifestazioni pubbliche alle quali aderivano partiti politici, organizzazioni sindacali, molte altre sigle, ma soprattutto cittadini comuni: la scelta migliore è sembrata quindi quella di inserire questi volantini nel quarto gruppo, quello cioè del “mondo civile”.

- Infine, nei casi in cui era difficile scegliere in quale gruppo inserire un volantino, sono state adottate due soluzioni: “forzare” in un certo senso la scelta e inserire il volantino in uno dei quattro gruppi, pur riconoscendone la parziale estraneità all’interno del gruppo scelto; oppure eliminare il volantino dal *corpus* se la scelta risultava davvero troppo complessa e sarebbe stata in ogni caso inadeguata.

Un esempio di scelta un po’ forzata è rappresentato da quei volantini che sono stati inseriti nell’ultimo gruppo, perché costituiti interamente da articoli de “l’Unità”: per quanto si trattasse di una fonte di partito, essendo un giornale il linguaggio e le forme utilizzate sono sicuramente differenti rispetto alle comunicazioni degli altri organi di partito e non è sembrata quindi una scelta corretta quella di inserire questi volantini nel primo *subcorpus*. Sono stati eliminati dal *corpus* invece, alcuni volantini che in realtà erano articoli presi dalla “Squilla di Galzignano”, periodico locale religioso, che poco aveva a che fare con tutti gli altri volantini.

Quest’ultima soluzione rimanda ad una questione rimasta aperta dall’inizio del capitolo, dato che ora si può comprendere perché, prima della divisione in *subcorpora*, il *corpus* non poteva ancora considerarsi definitivo. Alcuni elementi infatti (11 alla fine) sono stati estromessi per la motivazione appena descritta, oppure perché, sfogliando i volantini per osservarne e catalogarne le firme, si è notato che alcuni di essi comparivano due volte, o per errore o perché portavano firme differenti: è stata quindi immediata la scelta di eliminare uno dei due, poiché dal punto di vista del contenuto alcuni dati sarebbero risultati inutilmente raddoppiati. In un caso addirittura, due volantini erano completamente identici, tranne che per la data della manifestazione a cui i cittadini erano invitati.

Prima di reinserire in TaLTaC il *corpus*, finalmente definitivo, ad ogni volantino è stato assegnato insieme al titolo, un codice per consentire al programma di riconoscere il gruppo di appartenenza del volantino in relazione alla firma.

Ogni frammento del *corpus*, cioè ognuno dei 650 volantini rimasti, si presenta quindi con questa sintassi:

****Titolo *Firma=...
Testo

Il titolo, preceduto da quattro asterischi identifica il frammento, il successivo “Firma=...”, preceduto da un asterisco, identifica invece le caratteristiche descrittive del frammento secondo la categoria di analisi individuata (Giuliano e La Rocca, 2008). Nel nostro caso, come titoli si è scelto di inserire i codici che, nella catalogazione del centro studi, permettono di rintracciare il volantino all’interno dell’archivio e si presentano in questa forma: 19.01.002 (esempio), dove 19 è il codice di riferimento del fondo archivistico relativo ai volantini, le due cifre successive indicano la busta (faldone) all’interno della quale si trova il volantino, mentre le ultime tre cifre indicano il codice progressivo del volantino, all’interno della busta.

La firma invece ha, chiaramente, le seguenti modalità di risposta:

1. Federazione Provinciale del P.C.I.;
2. Sezioni locali del P.C.I.;
3. Altro: mondo del lavoro;
4. Altro: mondo civile.

3. PRESENTAZIONE DEL VOCABOLARIO

Completata la creazione dei quattro *subcorpora*, è arrivato il momento di presentare il vocabolario con la lista di tutte le forme grafiche presenti nel *corpus* : è possibile analizzarlo (almeno superficialmente) dal punto di vista quantitativo, chiedendo a TaLTaC di calcolare le “misure lessicometriche” e le “gamme di frequenza” in cui suddividere il vocabolario stesso, e dal punto di vista qualitativo, esplorandone il contenuto per osservare quali sono le parole “non vuote” (saltando quindi parole strumentali come articoli, preposizioni, congiunzioni, ecc.) più presenti.

Figura 2.1 *Misure Lessicometriche del corpus.*

Word Token	N = 259.666
Word Type	V = 17.708
Type/Token Ratio (TTR)	V/N % = 6,82%
% hapax	V1/V % = 45,45%
Freq. media generale	N/V = 14,66

Per quanto riguarda le “misure lessicometriche”, i risultati ottenuti sono riportati nella pagina a fianco (Figura 2.1 a pag. 16): i dati che vi sono contenuti verranno analizzati e spiegati nelle prossime righe.

Con la voce *word token* (N) è indicato il totale delle occorrenze presenti, cioè la dimensione del *corpus* (N=259.666): secondo le definizioni convenzionali, il *corpus* a nostra disposizione è quindi da considerarsi “grande”, perché presenta un numero di occorrenze maggiore di 200.000.

Word type (V=17.708) è invece il numero di forme grafiche distinte e misura l’ampiezza del vocabolario. Il rapporto percentuale tra queste due misure (V/N %), noto come *Type/Token Ratio* (TTR) misura la ricchezza lessicale e, nel nostro caso, il dato è decisamente incoraggiante, poiché le parole distinte rappresentano il 6,82% del *corpus*, e si tende a considerare “trattabile” con TaLTaC un *corpus* con un rapporto tra quantità di parole diverse (*word type*) e di parole totali (*word token*) minore del 20%.

L’altro parametro utile per valutare la trattabilità del *corpus* è la percentuale di *hapax* (V1/V %), cioè di forme grafiche che compaiono un’unica volta nel vocabolario: in questo caso la soglia da non superare è considerata quella del 50% (quindi le f.g. presenti una sola volta non devono essere più della metà del totale V) e anche stavolta il *corpus* rispetta l’indicazione, con una percentuale di *hapax* pari al 45,45%.

Ultimo dato presente in tabella (Figura 2.1), è quello riferito alla frequenza media generale (N/V), che è pari a 14,66: significa che ogni forma grafica compare in media tra le 14 e le 15 volte nel vocabolario. Più che un dato “tecnico”, questa frequenza media rappresenta solo una curiosità, poiché non fornisce indicazioni quantitative rilevanti: il valor medio interessa poco a fronte di frequenze reali che variano da 1 per il 45,45% delle forme grafiche, a migliaia per le forme più presenti.

Insieme alle misure lessicometriche, TaLTaC presenta anche le informazioni relative alle gamme (o fasce) di frequenza: osservando il vocabolario infatti, si nota la presenza di una colonna con scritto “Fasce” come intestazione e “Alta”, “Media” e “Bassa” come modalità e inoltre si può osservare che le forme grafiche riportate in ordine lessicometrico (cioè secondo la frequenza), si presentano anche in un ordinamento per ranghi crescenti, dove per “rango” si intende la posizione occupata da una forma grafica nella graduatoria (Giuliano e La Rocca, 2008). Sono proprio i ranghi a segnare

i limiti tra le diverse gamme di frequenza: il passaggio dalla fascia delle alte frequenze a quelle delle medie è determinato dalla prima coppia di forme grafiche che presentano lo stesso numero di occorrenze, mentre il passaggio dalle medie alle basse frequenze è rappresentato dalla prima lacuna nelle classi di occorrenze, cioè dalla prima cifra intera, partendo dal basso (cioè dagli *hapax*), che non rappresenta il numero di occorrenze di nessuna forma grafica del vocabolario.

Nel nostro caso (Figura 2.2), si nota che la fascia delle alte frequenze è composta da 62 forme grafiche, poiché si incontrano per la prima volta due f.g. (*ed* e *comunisti*) con lo stesso numero di occorrenze (420) al rango 63; la fascia delle basse frequenze raccoglie invece tutte le forme grafiche con frequenza minore o uguale a 101 (cioè a partire da *manifestazione, cose, case, studio*), dato che non è presente nessuna forma grafica con numero di occorrenze uguale a 102.

Figura 2.2 *Gamme di frequenza.*

Limite fra alte e medie frequenze:			
Rango	63	Freq. di soglia	420
		Freq. cumulata %	42,4%
Limite fra medie e basse frequenze:			
Rango	322	Freq. di soglia	101
		Freq. cumulata %	60,9%

Nella stessa tabella si può notare anche che le prime 62 forme grafiche (fascia delle alte frequenze) hanno frequenza cumulata totale pari al 42,4% e questo significa che raccolgono da sole tale percentuale delle occorrenze (n.b.: *word token*) presenti nel *corpus*. Sommando la fascia delle alte e quella delle medie frequenze, si ottiene invece il 60,9% delle occorrenze: quasi il 40% delle occorrenze di conseguenza, sono da imputare alle forme grafiche presenti nella fascia delle basse frequenze.

Dopo la descrizione delle dimensioni e delle caratteristiche quantitative del vocabolario, si può procedere con le prime osservazioni a proposito del contenuto del *corpus* che abbiamo scelto di analizzare: tornando ad osservare il vocabolario creato infatti, è possibile scorgere, tra le forme grafiche appartenenti alla fascia delle alte frequenze, le parole “non vuote” (o tematiche) maggiormente presenti nel *corpus* :

lavoratori con 1802 occorrenze è la prima, poi *lotta* (957 occorrenze), *operai* (744), *politica* (682), *lavoro* (636), *P.C.I.* (551), *governo* (505), *D.C.* (485), *padroni* (482), *scuola* (469), per limitare l'elenco alle prime dieci.

Questa osservazione assume particolare rilievo perché permette a chiunque di comprendere, o quantomeno indovinare, a partire da queste parole-chiave (tematiche appunto), i temi trattati nei volantini protagonisti di questa analisi pur non avendoli mai visti, né tantomeno letti.

Per approfondire l'analisi, sono stati creati in precedenza quattro partizioni del *corpus*, a partire dalle firme dei volantini: calcolando le sub-occorrenze possono quindi essere presentati anche i quattro *subcorpora*, misurando le dimensioni di ciascuno di essi.

Prima di confrontare le misure lessicometriche delle partizioni però, un'opzione interessante è quella di valutare quanti volantini sono presenti in ciascuno dei quattro *subcorpora* e quanti sono in percentuale all'interno del *corpus* (Figura 2.3).

Figura 2.3 *Composizione del corpus per subcorpora.*

Firma	Corpus		1962-68	1969	1970
	Num. Volantini	Percentuale	%	%	%
1. Fed. Prov. P.C.I.	244	37,5 %	42,5	32,6	37,1
2. Sezioni locali	177	27,2 %	27,5	30,0	24,9
3. Mondo del lavoro	123	18,9 %	14,5	25,3	17,8
4. Mondo civile	106	16,3 %	15,5	12,1	20,2
Totale	650		207	190	253

Nella prima parte della tabella, si può osservare che il *subcorpus* decisamente più numeroso è il primo (37,5%), quello composto da volantini firmati dalla Federazione Provinciale o da altri organi del P.C.I.; il secondo (Sezioni locali) rappresenta invece poco più di un quarto del *corpus* (27,2%), mentre meno corposi sono gli ultimi due *subcorpora*, che hanno una percentuale abbastanza simile (18,9% e 16,3%).

Nella parte destra della tabella invece, la stessa composizione percentuale è riferita non all'intero *corpus*, ma ai tre ipotetici *subcorpora* che si sarebbero potuti creare utilizzando l'anno come variabile d'interesse. Si nota che le percentuali hanno una

variabilità piuttosto elevata, infatti ad esempio, la percentuale di volantini del primo *subcorpus* è molto alta (42,5%) tra i volantini più “vecchi” (1962-68), e si abbassa nettamente (32,6%) per quelli del 1969, mentre l’esatto contrario (da 14,5% a 25,3%) accade per i volantini del terzo faldone. Inoltre si osserva che quest’ultimo è decisamente il più corposo (253 volantini), mentre gli altri due sono formati da meno volantini, in particolare il secondo ne contiene solo 190.

Queste ultime osservazioni sottolineano che, anche se l’anno di redazione dei volantini non è una variabile adatta per la creazione di *subcorpora*, potrebbe comunque fornire elementi interessanti per un altro studio che volesse occuparsi di analizzare questo folto gruppo di volantini.

Figura 2.4 *Misure lessicometriche dei subcorpora.*

		Fed. Prov. P.C.I. PD	Sezioni locali	Mondo del lavoro	Mondo civile
Word Token	N =	93.907	55.587	60.606	49.566
Word Type	V =	10.143	7.412	7.486	8.010
Type/Token Ratio (TTR)	V/N % =	10,80%	13,33%	12,35%	16,16%
% hapax	V1/V % =	49,06%	51,88%	48,94%	53,87%
Freq. media generale	N/V =	9,26	7,50	8,10	6,19

Le misure lessicometriche riportate in Figura 2.4 consentono di approfondire il confronto tra le dimensioni dei *subcorpora*, già iniziato in precedenza valutando la distribuzione del numero di volantini, ma rivelano anche qualche dato sorprendente: il primo *subcorpus*, oltre che composto dal maggior numero di volantini, presenta anche il maggior numero di occorrenze (93.907) e di forme grafiche (10.143), stupisce però scoprire che il secondo *subcorpus* rispetto al terzo, anche se raccoglie 54 volantini in più e una percentuale nettamente maggiore sul totale, ha più di cinquemila occorrenze in meno e di conseguenza anche un minor numero di forme grafiche. L’ultimo *subcorpus* inoltre, pur confermandosi il più piccolo, sia dal punto di vista del numero di volantini che del numero di occorrenze, presenta una quantità di forme grafiche (8.010) maggiore degli altri *subcorpora*, primo escluso.

Da queste indicazioni possiamo comprendere che: i testi dei volantini scritti da associazioni del mondo del lavoro (firma=3) sono decisamente lunghi rispetto agli

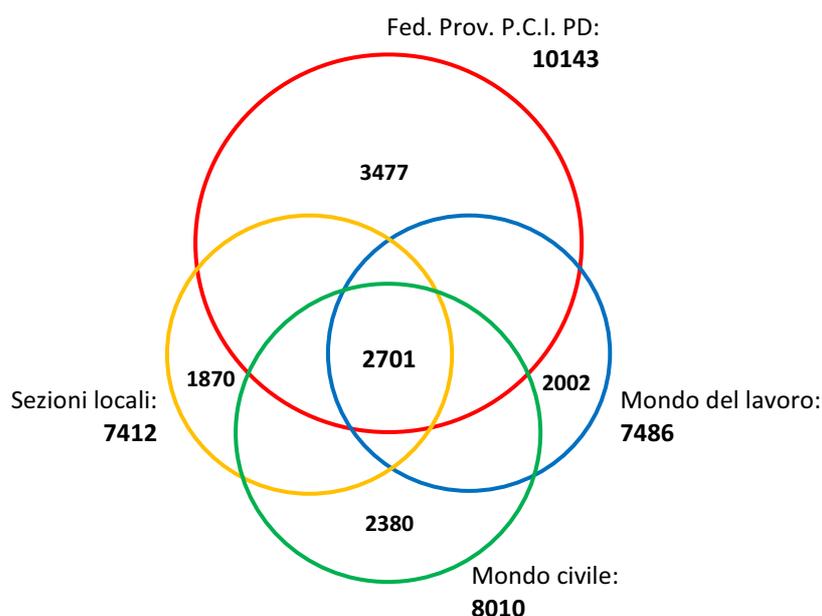
altri e questo spiega il maggior numero di occorrenze rispetto al secondo gruppo di volantini (scritti dalle sezioni locali del P.C.I.); il fatto che i testi scritti da organizzazioni del mondo civile, invece, presentino una quantità di forme grafiche superiore rispetto agli altri è probabilmente specchio di una maggior ricchezza lessicale.

Poco si può dire invece, a proposito dell'estensione lessicale misurata dal T.T.R. ($V/N\%$), poiché si tratta di una misura fortemente influenzata dal numero di occorrenze e infatti è maggiore dove N è minore e viceversa. L'ultima considerazione riguarda la percentuale di *hapax* che è per tutti i *subcorpora* vicina al limite (50%) che ne indica la trattabilità; in due casi è addirittura superiore.

4. CONTEGGIO E LETTURA DELLE FORME GRAFICHE COMUNI ED ESCLUSIVE

L'analisi dei *subcorpora* può ora proseguire con la lettura, a partire dal vocabolario aggiornato dopo il calcolo delle sub-occorrenze, delle forme grafiche comuni a tutti i quattro *subcorpora* e di quelle esclusive per ciascuno.

Figura 2.5 *Rappresentazione grafica dei subcorpora: forme comuni ed esclusive.*



Nella rappresentazione grafica (Figura 2.5 a pag. 21) che si può osservare nella pagina precedente, ad ogni circonferenza corrisponde un *subcorpus*, come si nota anche dalle didascalie che contengono proprio i nomi dei *subcorpora* e il numero di forme grafiche contenute in ciascuno di essi, mentre i diversi colori permettono di distinguere le aree delimitate dalle circonferenze, nonostante le intricate intersezioni. Tali aree rispecchiano, in proporzione, le differenze dei *subcorpora* per quanto riguarda la quantità di forme grafiche che li compongono: si nota infatti che l'area descritta dalla circonferenza rossa (*subcorpus* dei volantini firmati da organi di partito) è decisamente la più grande e infatti deve rappresentare più di 10.000 parole, le altre tre invece sono più piccole e di dimensioni simili, ma, osservando con attenzione, ci si può accorgere che la circonferenza verde (mondo civile) è un po' più grande delle altre due (infatti deve contenere circa 500 forme grafiche in più). Quasi impossibile infine, notare differenze nelle aree descritte dalle altre due circonferenze, gialla e blu (sezioni locali e mondo del lavoro, rispettivamente), perché il *subcorpus* del mondo del lavoro ha solo 74 forme grafiche in più.

Se le aree dei cerchi cercano di rappresentare nel miglior modo possibile le dimensioni dei *subcorpora*, lo stesso non si può dire delle aree delimitate dalle intersezioni, poiché combinare le posizioni delle quattro circonferenze in modo tale da rappresentare a grandezza reale il numero di forme grafiche condivise, di volta in volta, da due o tre *subcorpora*, sarebbe stata un'impresa decisamente ardua.

L'obiettivo quindi è solo quello di chiarire, con l'aiuto di una rappresentazione grafica, il concetto di forme comuni e di forme esclusive: le forme comuni infatti sono rappresentate dall'area posta al centro della figura e delimitata da quattro archi di circonferenze di diversi colori, con la cifra 2701 ad indicare il numero di forme grafiche comuni a tutte le quattro partizioni (circa 15% del vocabolario del *corpus*); le forme esclusive sono invece descritte dalle quattro aree più esterne di ogni circonferenza, non condivise con nessuno degli altri cerchi, ed anche al loro interno, un numero indica la quantità di forme esclusive di ciascun *subcorpus*: dati, quest'ultimi, che verranno analizzati successivamente nel paragrafo dedicato "esclusivamente" alle forme grafiche esclusive.

4.1 Forme grafiche comuni

La lettura delle forme grafiche condivise dai *subcorpora*, cioè presenti almeno una volta in ciascuno di essi, fornisce informazioni interessanti non tanto per quanto riguarda la ricerca di un filo-conduttore che unisca volantini che si presentano con firme diverse, poiché questo si osserva facilmente e banalmente osservando le parole tematiche, cioè quelle maggiormente presenti nel *corpus* (che faranno chiaramente parte del sottoinsieme di forme grafiche definite comuni); piuttosto, osservare le parole presenti in questo sottoinsieme, è molto più interessante al fine di comprendere il linguaggio utilizzato per trasmettere il messaggio veicolato attraverso i volantini, osservando cosa caratterizzi questo linguaggio indipendentemente dalla mano che si è occupata della trascrizione.

Come si è visto, le forme grafiche comuni sono molte, 2701 per la precisione, cioè circa il 15% del vocabolario.

Il problema che si pone è quindi quello di scegliere un metodo adatto ad analizzare le forme comuni, poiché leggerle una per una è impossibile ed oltretutto inutile: impossibile perché sarebbe un'enorme perdita di tempo proporre una lista di 2701 parole, commentandole una per una, inutile perché questa ipotetica lista sarebbe piena di articoli, preposizioni, congiunzioni, ma anche verbi, sostantivi e aggettivi, utilizzati ovviamente in ogni *subcorpus* semplicemente in quanto parte del linguaggio comune o del contesto in cui i volantini si inseriscono.

L'analisi delle forme comuni di conseguenza non può che partire da una scelta soggettiva, cioè dalla ricerca di una parte di esse che possa fornire informazioni rilevanti: si può ad esempio scegliere di osservare quelle forme che compaiono un'unica volta in ciascuno dei *subcorpora* e presentano quindi nel *corpus* un numero di occorrenze pari a quattro. La speranza è che, a partire da questa scelta, sia possibile trovare degli spunti di confronto interessanti riguardo all'utilizzo di una stessa parola, evidentemente "rara", in quattro contesti tematici diversi, dovuti alla differente firma del volantino in cui si è registrata la parola stessa.

Tra le 2701 parole comuni, solo 33 hanno le caratteristiche appena definite e, com'era prevedibile, non sono apparentemente collegate tra loro da alcuna omogeneità di tipo tematico. Si tratta per lo più di verbi (*dormire, testimoniare*, et al.) in varie forme tra cui

molti participi passati (*aggravata, abituati, diffuso*, et al.) e di alcuni interessanti sostantivi (*carriera, contrattacco, motivazione*) e aggettivi (*contenti, drastica, municipalizzate, venete*).

Anche in questo caso TaLTaC offre un comodo strumento per analizzare questi vocaboli, cioè l'analisi delle concordanze, che permette di ritrovare nel *corpus* le parole cercate (in questo caso le 33 forme comuni sopraccitate) e di leggerle all'interno delle frasi in cui sono state inserite.

Si scopre ad esempio che il verbo *testimoniare* ha oggetti molto diversi a seconda del contesto di utilizzo: riguarda una testimonianza ad un processo solo in un caso, cioè in un volantino firmato dal P.C.I. che parla del drammatico processo di Burgos del dicembre 1970, nel corso del quale il governo franchista chiese la condanna a morte per sei partigiani baschi; negli altri casi invece il significato del verbo impiegato è diverso: in un volantino firmato da associazioni sportive (quindi del mondo civile) si chiede a tutti i cittadini di partecipare ad una staffetta ciclistica per *testimoniare* la loro volontà di pace; una sezione locale di Padova invita invece “tutta la popolazione di sentimenti democratici [...] a *testimoniare* la propria fedeltà agli ideali di libertà, di giustizia e di pace della Resistenza italiana”; infine una “Commissione di fabbrica” conclude un lungo volantino sottolineando che “i lavoratori debbono *testimoniare* [...] il rifiuto del ricatto anticomunista”.

Diversa invece la situazione riguardante il *contrattacco*: in questo caso infatti indipendentemente da chi ha redatto il volantino, il termine è utilizzato sempre nella stessa accezione e nello stesso contesto di scontro tra lavoratori in lotta e padroni sempre pronti a passare al *contrattacco* per contrastare il consenso popolare progressivamente guadagnato dai lavoratori con le loro azioni di protesta.

Come ultimo esempio, è curioso il caso dell'aggettivo *contenti*, riferito a quattro soggetti differenti e utilizzato per esprimere concetti molto distanti tra loro.

“Bastonati e *contenti* ”: secondo la Federazione Comunista di Padova, così i padroni vorrebbero che fossero i lavoratori, che nonostante i soprusi subiti dovrebbero portare sempre e comunque pazienza e possibilmente essere anche “inoffensivi sul piano politico”.

Sempre secondo i padroni, ma in un altro volantino firmato stavolta dal “Comitato unitario di base” della Precisa (fabbrica padovana), gli operai dovrebbero essere *contenti* della loro situazione perché vicino a loro c'è chi sta peggio, cioè i giovani

apprendisti costretti a ricevere “salari da fame”, pur producendo come qualunque altro lavoratore. Impossibile non notare la triste somiglianza tra quelli che erano chiamati apprendisti quarant’anni fa e quelli che ora chiamiamo precari: sicuramente anche oggi qualcuno sarebbe pronto a dire loro che farebbero meglio ad accontentarsi perché c’è chi sta peggio, disoccupati e cassaintegrati ad esempio...

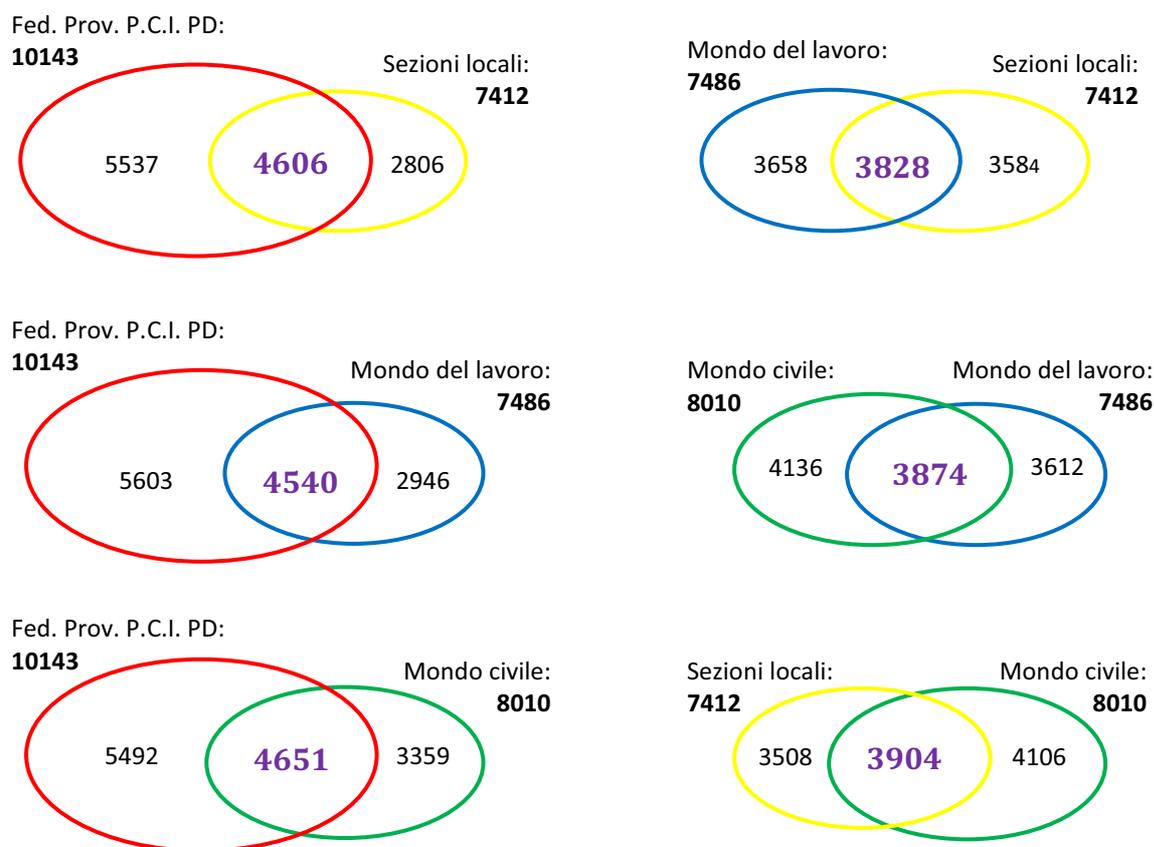
Con l’analisi delle forme grafiche comuni si è potuto osservare come l’utilizzo degli stessi termini nei vari *subcorpora* non significhi sempre omogeneità di linguaggio o di contesto tematico, dato che le stesse parole possono essere usate con modalità molto differenti a seconda dell’intento e del *background* di chi scrive; d’altro canto però, si è capito anche che alcuni temi e alcune forme linguistiche vengono impiegati in maniera trasversale tra gli scriventi, si tratti di un organo di partito, come di un’associazione di cittadini o di lavoratori.

È doveroso comunque ricordare che questa prima parte dell’analisi delle forme comuni, in particolare i risultati ottenuti, dipendono direttamente dalla scelta fatta inizialmente per selezionare una parte delle quasi tremila forme comuni presenti nel *corpus*. Ne consegue che una scelta diversa avrebbe con tutta probabilità portato ad osservazioni differenti ma di pari interesse, e che questa parte dell’analisi potrebbe protrarsi a lungo, fino all’esaurimento di tutte le possibili scelte di partenza, comunque necessarie per ridurre il numero di parole da considerare.

Restando alle forme grafiche comuni, oltre ad osservare la situazione dal punto di vista qualitativo, analizzando le parole utilizzate da tutti i *subcorpora* per trarre qualche conclusione riguardo al linguaggio utilizzato nell’intero *corpus*, può essere interessante anche un confronto quantitativo, per scoprire quali *subcorpora* condividono più forme grafiche e quali invece sono più distanti dal punto di vista del linguaggio utilizzato.

Questo tipo di osservazione può essere svolta e chiarita al meglio confrontando tra loro tutte le possibili coppie di *subcorpora*, per valutare l’entità delle intersezioni tra i loro vocabolari. Nuovamente, una rappresentazione grafica può essere di grande aiuto per la comprensione della situazione (Figura 2.6 a pagina 26).

Figura 2.6 Confronto a coppie tra i vocabolari dei subcorpora, per il conteggio delle forme comuni.



Innanzitutto, è necessario sottolineare che le misure delle aree ellittiche che rappresentano i quattro *subcorpora* (con le circonferenze colorate come in Figura 2.5), e di quelle determinate dalle intersezioni, cercano di rispecchiare, nei limiti del possibile, la reale quantità di forme grafiche che compongono i vocabolari dei quattro gruppi.

Per ciascuna delle sei possibili coppie vengono indicati: esternamente, i numeri di forme grafiche che compongono i vocabolari dei due *subcorpora* confrontati; all'interno invece, ai lati il numero di forme che ogni *subcorpus* non condivide con quello a cui è accoppiato e al centro dell'area determinata dall'intersezione delle due forme ellittiche, il numero di forme comuni ai due *subcorpora* (in viola).

I risultati non evidenziano differenze evidenti nel numero di forme condivise, di volta in volta, da due *subcorpora*: si nota che le tre coppie a sinistra presentano un maggior numero di forme comuni rispetto alle altre tre, ma sono le coppie in cui è coinvolto il gruppo di volantini firmati dalla Federazione Provinciale di Padova del P.C.I. e da altri organi di partito, cioè il *subcorpus* decisamente più corposo, di conseguenza l'alto

numero di forme comuni è - con tutta probabilità - determinato principalmente dal maggior numero di forme “candidate” ad essere condivise.

Per cercare di rendere confrontabile il numero di forme comuni alle coppie di *subcorpora* si è quindi provato a calcolare, per tutti e sei i casi, il rapporto tra le forme comuni (intersezione dei due insiemi) e la somma dei due vocabolari confrontati (unione dei due insiemi): i risultati confermano ancora una volta l'evidente omogeneità dei vocabolari dei quattro *subcorpora*, perché le percentuali ottenute sono estremamente simili e variano davvero di poco, da un minimo pari a 25% nel caso della coppia “Mondo del lavoro” - “Mondo civile” a un massimo di 26,2% per la coppia “Federazione Provinciale del P.C.I. di Padova” - “Sezioni locali del P.C.I.”.

L'unico particolare rilevante per evidenziare differenze di linguaggio tra i *subcorpora*, pur tenendo conto del fatto che le variazioni sono davvero limitate, è rappresentato dal fatto che, ordinando le sei coppie in base al valore (decrescente) del rapporto percentuale calcolato, si trovano agli ultimi tre posti, quindi con meno forme in comune, le coppie in cui sono presenti i volantini scritti da associazioni del mondo civile e studentesco: in un contesto di forte omogeneità di linguaggio tra i *subcorpora* quindi, gli unici volantini che sembrano presentare differenze di lessico, per quanto limitate, sono proprio quelli appartenenti all'ultimo *subcorpus*.

Tutti i volantini presenti nel *corpus* provengono da un unico fondo archivistico del “Centro Studi Ettore Luccini”, infatti in fase di introduzione si è detto che si tratta di una parte (i più vecchi) dei volantini “redatti dal 1962 al 1997 dalla Federazione Provinciale di Padova del P.C.I. o da altri organi (politici, sindacali, del mondo civile, ecc.) riconducibili a movimenti operai, contadini, popolari, della provincia padovana”. La marcata omogeneità di linguaggio che si è osservata anche durante questo ultimo confronto tra *subcorpora* è quindi una diretta conseguenza dell'ambito (unico) da cui sono stati estratti i volantini, che, solo successivamente, sono stati divisi in quattro partizioni per esigenze di analisi.

In conclusione quindi, questa seconda parte dell'osservazione delle forme comuni è servita a sottolineare con forza che i *subcorpora*, considerati due per volta, hanno più o meno la stessa quantità di forme comuni tra di loro, e che solo il *subcorpus* dei volantini scritti da associazioni del mondo civile, sembra aver meno forme in comune con gli altri tre.

4.2 Forme grafiche esclusive

Il passo successivo dell'analisi consiste nel cercare di capire quali sono invece i temi trattati esclusivamente da un *subcorpus* e non dagli altri. A questa domanda è possibile rispondere, almeno parzialmente, attraverso la lettura delle forme grafiche che risultano esclusive per un *subcorpus*, parole che sono quindi utilizzate unicamente in un gruppo di volantini e non negli altri.

Analizzare tutte le forme grafiche esclusive dei quattro *subcorpora* è estremamente difficile, perché si tratta di migliaia di parole (Figura 2.7) ed è quindi necessario occuparsi di un sottoinsieme di forme grafiche scelto *ad hoc*: si può per esempio prendere in considerazione quelle con un numero di occorrenze maggiore di una soglia prefissata, perché tra tutte le forme esclusive saranno quelle più utilizzate all'interno del *subcorpora* preso in esame.

Figura 2.7 *Forme grafiche esclusive per i 4 subcorpora.*

Subcorpus	F.g. Esclusive	% sul totale
1. Fed. Prov. P.C.I. PD	3477	34,3 %
2. Sezioni locali	1870	25,2 %
3. Mondo del lavoro	2002	26,7 %
4. Mondo civile	2380	29,7 %

Scegliendo ad esempio 5 come soglia per selezionare le parole esclusive da considerare, si limita in questo caso l'analisi ad una quindicina (mediamente) di forme grafiche per *subcorpora*:

- nei volantini redatti dalla Federazione Provinciale o da altri organi del P.C.I. compaiono alcune parole esclusive che richiamano tematiche giuridico/legislative come *imputati* (11 occorrenze), *avvocati* e *procuratore* (6 ciascuna), e altre forme che evidenziano invece riferimenti di tipo politico, sia nazionale che internazionale: si possono ritrovare infatti nomi di personalità politiche del tempo come il ministro del lavoro *Donat-Cattin* (7), uno dei pochi politici della D.C. che hanno saputo meritarsi il rispetto anche delle forze di

sinistra e il senatore americano Robert (Bob) *Kennedy* (10) di cui si describe e si denuncia l'assassinio in alcuni volantini, ma anche altre parole riguardanti la politica estera come *Onandia* (7), nome di un partigiano *basco* (8) condannato a morte dal governo franchista nel processo di Burgos (già incontrato), e *Palestina* (6), testimonianza del fatto che il conflitto in medio-oriente tra Israele e Palestina era un tema attuale quarant'anni fa, come lo è oggi;

- le Sezioni locali del partito tendevano invece ad occuparsi, non solo ma principalmente, dei problemi delle loro zone e questo risulta evidente anche dalla lettura delle forme esclusive: si trovano infatti alcuni nomi di quartieri di Padova e di Comuni della provincia, come *Terranegra* (15), *Altichiero* (14), *Chiesanuova* (9) ed altri termini come ad esempio *allacciamento* (7) - dell'acqua o del gas - che indicano la trattazione di problemi concreti e specifici che affliggevano interi quartieri;
- le condizioni di lavoro nelle fabbriche e le lotte tra operai e padroni erano tra le principali protagoniste dei volantini redatti da associazioni del mondo del lavoro e quindi è normale ritrovare tra le forme esclusive da loro utilizzate, i nomi di alcune fabbriche padovane come *Hesco* (22), *Comet* (10) e *Gorena* (8), del sindacato dei metalmeccanici, *F.I.M.* (11), molto attivo in quegli anni per il rinnovo del contratto collettivo di lavoro, di un padrone - *Zamboni* (6) - che si rifiutava di trattare con i sindacati e di un operaio-sindacalista, Franco *Masili* (7), divenuto protagonista, suo malgrado, di volantini dove si denunciava il suo ingiusto licenziamento definito "di rappresaglia". Curioso inoltre, è notare tra queste parole esclusive, la presenza del nome di un prete, don *Borghesi* (6), rappresentante di quella categoria di preti-operai che cercavano di attrarre l'attenzione popolare sulle spesso difficilissime condizioni di lavoro nelle fabbriche, condividendo con gli operai, non solo la messa, ma soprattutto le ore di lavoro alla catena di montaggio e le proteste sindacali per rivendicare condizioni migliori.

- infine, molti dei volantini che fanno parte dell'ultimo gruppo, sono redatti da organizzazioni studentesche e infatti tra le forme grafiche esclusive, sono molte quelle che rimandano al mondo della scuola, riassumendo tutto il percorso formativo: dai *nidi* (7), asili per i più piccoli, al mondo *accademico* (7) che porta al traguardo della *laurea* (11). Una forma esclusiva un po' particolare, perché non si tratta di una parola, è 442 (6), numero di una legge del 12 marzo 1968 che dettava alle università le norme per l'istituzione del quinto anno di corso, necessario per il conseguimento della laurea con valore abilitante.

Leggendo ancora le forme grafiche esclusive di quest'ultimo *subcorpus*, se ne trovano anche alcune che riguardano un tema particolare: *scià* (15), *Persia* (10), *Iran* (7), *Teberan* (7), *iraniani* (7), *Rezai* (6). Il riferimento evidente è alla situazione politica e sociale dell'Iran, chiamato con il nome antico di Persia nella maggior parte dei casi: i volantini che si occupano di questo argomento si riferiscono infatti alle difficili condizioni di vita imposte ai cittadini iraniani dal regime autoritario dello *scià*, ma anche ad episodi particolari, come la fucilazione di 26 prigionieri politici e l'arresto di uno studente, Hussein *Rezai*, rappresentante della "Confederazione degli studenti iraniani".

5. CONTEGGIO E LETTURA DELLE SPECIFICITÀ

Con il riconoscimento e la lettura delle forme grafiche comuni ed esclusive, si è sviluppato un interessante confronto tra *subcorpora* che può essere ulteriormente approfondito con l'analisi delle specificità, processo che permette di individuare con una metodologia di tipo statistico le unità testuali specifiche dei *subcorpora* ricorrendo a modelli probabilistici basati sulla distribuzione ipergeometrica (Lafon, 1980). Affinché un'unità testuale possa essere riconosciuta come specifica di un *subcorpus*, deve essere sovrautilizzata rispetto agli altri *subcorpora* in relazione ad una soglia di probabilità prefissata (generalmente $p > 0,025$) e dev'essere distribuita in modo sufficientemente uniforme all'interno del *corpus* (Tuzzi et al, 2004).

Dopo aver chiesto a TaLTaC di calcolare le specificità, il vocabolario si arricchisce di nuove informazioni: ogni forma grafica viene infatti indicata come specifica o “banale” per ciascun *subcorpora* e per le forme che si dimostrano specifiche viene indicato anche il *p-value* che misura il grado di specificità. Per facilitare la lettura delle sole forme specifiche, è possibile richiedere al programma di creare un file con la lista delle 50 forme che presentano il livello di specificità più alto (quindi con *p-value* minore) per ogni *subcorpora*.

Le informazioni tratte dalla lettura di queste parole confermano ciò che si era già osservato con le forme esclusive, ribadendo quelle che sono le tematiche particolari trattate nei volantini appartenenti ai quattro *subcorpora* creati, ma aggiungono anche alcune indicazioni più approfondite, non solo riguardo ai temi, ma anche a proposito del linguaggio e delle “tecniche comunicative” utilizzate.

- Nel primo gruppo (volantini firmati dalla Federazione Provinciale e da altri organi di partito), tra le forme che - più di tutte - si distinguono per la loro forte specificità, si trovano *federazione, giovanile, italiana*, non a caso tre delle quattro parole che compongono il nome di uno degli organi di partito (F.G.C.I.) che ha prodotto e firmato più volantini: la firma invero è parte integrante del testo del volantino ed è proprio questo il motivo che porta queste tre parole ad essere tra le più specifiche in questo *subcorpus*.

Buona parte delle altre parole specifiche rimanda direttamente alla politica, tra queste *governo, sinistra e destra, centrosinistra, Partito Comunista*, poi *crisi e Rumor*, due parole strettamente legate tra loro nella storia politica italiana poiché in poco più di un anno e mezzo, tra il dicembre 1968 e il luglio 1970, si susseguirono ben tre governi con l'onorevole Mariano Rumor (D.C.) come Primo Ministro in carica, e tutti e tre si chiusero con un periodo di crisi lungo almeno un mese.

Altre parole riguardano il mondo del lavoro, tema caro non solo alle associazioni che firmano i volantini del terzo gruppo, ma anche al P.C.I. stesso, da sempre schierato “dalla parte degli operai e dei lavoratori”: *braccianti, grandi - monopoli, M.E.C.* (Mercato Comune Europeo), ecc..

Abbandonando il discorso delle tematiche trattate, che potrebbe comunque proseguire a lungo vista la presenza di forme specifiche quali *polizia, donne, studenti, Vietnam* e altre ancora, può essere interessante sottolineare altre parole specifiche come *masse, progresso, svolta, sviluppo*, che non sembrerebbero a prima vista direttamente collegate tra loro da un tema, ma che indicano due tendenze particolari del P.C.I. del tempo: da un lato lo sguardo attentamente rivolto al futuro, o meglio alla speranza di un futuro migliore, dove potessero realizzarsi *sviluppo e progresso*, con la consapevolezza però che per poter creare questo futuro era necessaria una *svolta* che non poteva venire che dalle *masse* popolari e lavoratrici; dall'altro, dal punto di vista della comunicazione, la scelta di utilizzare termini forti, come questi, pronti a trasformarsi in slogan capaci di ispirare le masse (appuntamento, ndr) e unirle attorno ad obiettivi comuni.

- Per quanto riguarda i volantini firmati dalle sezioni locali del P.C.I., si possono raccogliere la maggior parte delle forme specifiche in due grandi gruppi che non offrono indicazioni di particolare interesse, per poi osservare più da vicino le forme rimanenti. Il primo gruppo è composto da numerosi nomi di Comuni del padovano, quali *Piazzola* (sul Brenta, ndr), *Cadoneghe, Solesino, Vigodarzere, Camin, Noventa* (et al.), il secondo invece da termini tipici della politica locale: *sezione, giunta, amministrazione, comunale, sindaco, quartiere, consiglio, assessore*, ecc.. Accantonate tutte queste, rimane una dozzina di parole di cui si può dire qualcosa in più: *cittadini* ad esempio segnala l'abitudine delle sezioni di rivolgersi in maniera diretta al lettore del volantino, tecnica utile per aumentare il grado di coinvolgimento; l'assenza o l'inefficacia di un piano *regolatore* e dell'*acquedotto* erano *problemi* tipici di molti piccoli Comuni, mentre diversi, ma comunque numerosi, erano i problemi per gli *alunni* delle scuole; infine *promesse* e solo promesse, era ciò che i comunisti contestavano alle amministrazioni democristiane, incapaci secondo loro di mantenere quelle stesse promesse che venivano continuamente fatte ai cittadini.

Si capisce anche dall'ultimo esempio che tra i due grandi partiti italiani, la contrapposizione era forte e gli scontri molto frequenti, ancor più a livello locale dove interessi e indirizzi politici differenti si affrontavano

quotidianamente nei Consigli comunali: le forme specifiche di questi volantini rispecchiano in modo diretto tale contrapposizione con la presenza di *D.C.* contro *P.C.I.* e quindi *democristiani* contro *comunisti*.

- Un'operazione simile a quella fatta per il *subcorpus* precedente può essere fatta anche per le forme specifiche relative ai volantini scritti da associazioni del mondo del lavoro. In questo caso i due gruppi creati in precedenza possono essere composti dai nomi delle fabbriche padovane da una parte e dai termini riguardanti il mondo del lavoro e le sue componenti dall'altra.

Tra i nomi di fabbrica considerati specifici si trovano *Zedapa*, *Rizzato*, *Hesco* (già incontrata tra le forme esclusive), *Precisa* e *Viscosa*, fabbriche che in molti volantini sono citate tutte insieme, dato che insieme lottavano per ottenere gli stessi diritti e migliori condizioni di lavoro.

Molte altre specificità, come detto, sono invece riferite alle componenti del mondo del lavoro: *lavoratori* e *operai* delle *fabbriche*, perennemente in *lotta* contro i *padroni*, con l'aiuto dei *sindacati* che spingevano in ogni fabbrica per la creazione della *commissione - interna*, fondamentale per organizzare al meglio le lotte. Si nota che tra le forme specifiche molte di queste si presentano in forma sia singolare che plurale (es. fabbrica/-che, padrone/i, sindacato/i), mentre questo non accade per *lavoratori* e *operai* e per la *lotta*: non è un particolare irrilevante, perché per il successo delle lotte operaie era fondamentale mantenere unite le masse lavoratrici, rivolgendosi non ai singoli ma agli *operai* e ai *lavoratori* tutti; allo stesso modo, per ribadire quest'unità la *lotta* doveva essere una sola, al fine di impedire che all'interno della classe operaia ognuno difendesse solo i propri interessi, invece di lottare per obiettivi condivisi come ad esempio lo Statuto dei diritti dei lavoratori, ottenuto nel maggio 1970.

- Per quanto riguarda infine, il *subcorpus* costituito dai volantini scritti da associazioni del mondo civile, si è già potuto notare che molti portano la firma di associazioni del movimento studentesco, particolarmente attivo alla fine degli anni '60: la presenza tra le forme specifiche di molti termini riguardanti il mondo della scuola e dell'università è quindi un'ovvia conseguenza. Ai primi

posti nella graduatoria delle specificità si trovano infatti *università* e *facoltà*, ma sono presenti anche *istituto*, *docenti*, *laurea*, *studenti* e molte altre ancora.

Una particolarità di questo gruppo di forme specifiche è rappresentato dal fatto che, rispetto agli altri tre *subcorpora*, sono molte le forme che non sono solamente specifiche ma anche esclusive, ed è in particolare il caso delle tante parole riguardanti l'Iran, già incontrate in precedenza (*Persia*, *scià*, *Teberan*, ecc.). Per le forme specifiche che non fanno parte né del mondo della scuola, né hanno a che fare con l'Iran, può essere utile riutilizzare l'analisi delle concordanze, com'è stato fatto con le forme comuni, perché non è immediato comprendere il motivo per cui alcune di queste parole risultano specifiche proprio per questo *subcorpus*.

Può essere difficile, ad esempio, ipotizzare una motivazione valida per spiegare il motivo per cui l'aggettivo *militare* appare come specifico per i volantini scritti da associazioni del mondo civile, ma la comprensione diventa più facile leggendo le frasi all'interno delle quali è stato utilizzato: si scopre che in molti casi si parla di “servizio *militare*” e in particolare delle richieste di istituire per legge “la possibilità di prestare un servizio civile alternativo all'obbligo *militare*”. Sempre con l'aiuto dell'analisi delle concordanze si può anche osservare che altre tre forme specifiche, *Marzolla*, *pendolari* e *Siamic*, sono strettamente legate tra loro dato che sono molti i volantini firmati da comitati di lavoratori *pendolari* che lamentano la pessima qualità dei servizi offerti dalle aziende di trasporto pubblico (*Marzolla* e *Siamic* appunto) che devono utilizzare ogni giorno per potersi recare sul luogo di lavoro.

Un ultimo elemento interessante infine è rappresentato dalla presenza del possessivo *vostr*i tra le forme specifiche, che è interessante analizzare per riportare il discorso al linguaggio utilizzato nei volantini per rapportarsi con i “lettori”: l'analisi delle concordanze è fondamentale anche in questo caso perché rivela che *vostr*i è più volte seguito da ‘figli’ o da ‘ragazzi’, nei volantini in cui comitati di studenti si rivolgono ai genitori degli alunni delle scuole per sottolineare le difficili condizioni in cui si trovano a studiare i loro figli, generalmente a causa di strutture assenti o inadeguate.

Come già osservato a proposito dell'utilizzo della forma *cittadini*, specifica per il secondo *subcorpora*, anche in questo caso con *vostr*, si segnala l'abitudine di chi compilava i volantini, di rivolgersi in modo diretto ai riceventi per coinvolgerli il più possibile nelle tematiche trattate.

6. ESTRAZIONE DI INFORMAZIONE PER PAROLE-CHIAVE

L'analisi del *corpus* per forme grafiche semplici si può concludere con la presentazione e il calcolo dell'indice *TFIDF* (*Term Frequency - Inverse Document Frequency*), un'altra risorsa di TaLTaC molto utile nell'ambito dell'estrazione di informazione peculiare dal *corpus* per parole-chiave.

L'indice *TFIDF* si calcola come prodotto tra due fattori, dove il primo (*term frequency*) misura il numero di occorrenze di una forma, mentre il secondo (*inverse document frequency*) è il logaritmo del rapporto tra il numero di documenti costituenti un *corpus* (N) e il numero di documenti che presentano quella forma (n):

$$TFIDF = tf * \log (N/n)$$

Può essere utilizzato in due modi diversi, a seconda del risultato che si cerca di ottenere:

- quando si misura l'indice *TFIDF* di una forma per tutto il vocabolario, l'obiettivo è quello di estrarre il linguaggio rilevante, cioè le forme più significative in quanto concentrate solo in pochi documenti e non disperse in tutto il *corpus*;
- quando, invece, si calcola su un sottoinsieme tematico del vocabolario (cioè su una lista di parole-chiave precedentemente creata) ha l'obiettivo di individuare i soli frammenti significativi rispetto a quelle parole-chiave.

Nel nostro caso lo si vuole utilizzare in questa seconda forma, con l'obiettivo di estrarre i volantini più significativi in relazione ad una lista di parole, scelta per rappresentare una specifica tematica.

6.1 La violenza attraverso le armi

Gli anni in cui sono stati scritti i volantini che compongono il *corpus*, rappresentano un periodo in cui, in Italia, lo scontro politico, sociale e di classe, sfociava spesso in episodi di violenza. Non a caso, in molti vedono negli episodi di quegli anni, come le contestazioni del '68 e la strage di Piazza Fontana (12 dicembre 1969), l'inizio di quel periodo tristemente conosciuto con il nome di "anni di piombo", ed il poliziotto Antonio Annarumma (citato in molti volantini), ucciso il 19 novembre 1969 a Milano durante una manifestazione, viene considerato il primo morto di quel triste periodo.

Questa parentesi storica rappresenta la premessa necessaria per spiegare perché si è scelto di utilizzare l'indice *TFIDF* per estrarre i volantini statisticamente più rilevanti rispetto al tema della violenza. Come si è detto, il processo di estrazione parte dalla creazione di una lista di parole che devono rappresentare nel miglior modo possibile il tema scelto, poiché è proprio dalla selezione adeguata delle parole-chiave che dipende il successo di questo procedimento: la violenza viene perpetrata in genere con l'uso delle armi, ed ecco allora che proprio le armi presenti nel vocabolario, insieme ad altre azioni che indicano violenza, compongono la lista per il calcolo del *TFIDF* (Figura 2.8).

Figura 2.8 *Lista di forme grafiche per il calcolo dell'indice TFIDF e rispettive occorrenze.*

Forma Grafica	Occorrenze
bomb*	40
baston*	11
pistol*	8
pugn*	8
calci	2
tirapugni	5
spranghe	2

Nota: le forme concluse con un asterisco indicano la presenza di più vocaboli con quella radice (es.: baston* sta per 'bastone', 'bastoni', 'bastonato', 'bastonatura', ecc.)

Come si comprende dalla nota alla figura 2.8, le forme appartenenti alla lista non sono solo sette (bensì 19), poiché più forme aventi la stessa radice sono state presentate insieme, mentre le occorrenze totali delle parole appartenenti alla lista sono 76.

Andando a calcolare il valore dell'indice *TFIDF* per i vari frammenti (volantini) che compongono il *corpus*, si potrà ottenere un risultato pari a zero, a uno, o maggiore di uno:

- l'indice sarà pari a zero nel caso in cui nel testo del volantino non sia presente nessuna delle forme che compongono la lista. Il volantino non avrà quindi niente a che fare con il tema della violenza (almeno a partire dalle parole scelte per rappresentarlo);
- sarà pari a uno se nel volantino compare un'unica volta, una sola delle parole in lista, e quindi la rilevanza rispetto alla violenza sarà piuttosto limitata;
- sarà infine maggiore di uno se il testo del volantino presenta più parole appartenenti alla lista creata, o più volte una stessa parola. È questo il caso che dovrebbe permetterci di riconoscere i volantini attinenti al tema della violenza: più alto sarà il valore dell'indice *TFIDF*, maggiore la rilevanza del frammento.

Il numero limitato di occorrenze totali delle parole che compongono la lista, ha fatto sì che i frammenti con indice diverso da zero siano piuttosto pochi: su 650 volantini infatti, solo 32 hanno *TFIDF* uguale a uno, mentre ancor meno - 13 - sono quelli con *TFIDF* maggiore.

Nonostante la quantità di frammenti rilevanti sembri piuttosto limitata, dal punto di vista del contenuto non si può dire che il procedimento non sia riuscito, dato che tutti i volantini estratti (tranne uno) dimostrano di essere realmente pertinenti rispetto al tema della violenza, trattando principalmente tre diverse manifestazioni di essa: agguati e provocazioni di gruppi fascisti, azioni repressive da parte della polizia e i numerosi bombardamenti, opera dell'imperialismo americano nel sud-est asiatico.

Il volantino che presenta il valore maggiore del *TFIDF* (= 2,22), per esempio, è stato scritto dal P.C.I. di Padova per denunciare l'azione di "gruppi di teppisti fascisti", resisi protagonisti di un'aggressione nei confronti di giovani e operai che "vigilavano ai cancelli del Municipio", mentre all'interno il Consiglio comunale discuteva di un attentato incendiario avvenuto poche ore prima contro l'Ateneo padovano.

Il volantino caratterizzato invece dal secondo valore più alto (= 1,95) dell'indice *TFIDF*, scritto dalla Federazione Provinciale del P.C.I., descrive i terribili avvenimenti verificatisi a Roma in occasione della visita del presidente americano Nixon: tra i

manifestanti che protestavano contro la presenza di chi aveva dato il via “all’aggressione imperialista in Vietnam”, si contavano uno studente morto e numerosi feriti in seguito all’azione repressiva di polizia e carabinieri che, “in assetto da guerra, armati di bombe lacrimogene, idranti, bastoni, scudi in plexiglas, si sono scatenati contro i dimostranti, organizzando per le strade della capitale selvagge bastonature, pestaggi sanguinosi, in una caccia all’uomo che è durata alcune ore”.

Proprio il tema dell’imperialismo, come detto, è trattato in altri volantini contraddistinti da un alto valore del *TFIDF*; volantini che descrivono i bombardamenti americani in Vietnam, Laos e Cambogia e criticano questa “aggressione imperialista”, celebrando ed esaltando la strenua resistenza opposta dai partigiani vietnamiti.

Detto dei due volantini più rilevanti rispetto al tema della violenza, è sufficiente sottolineare che gli altri presentano un valore del *TFIDF* compreso tra 1,73 e 1,38 e descrivono fatti simili a quelli già segnalati.

Prima di concludere però, anche per spiegare concretamente come TaLTaC può essere tratto in errore dai termini contenuti nella lista di parole scelte per rappresentare il tema che si vuole analizzare, può essere molto utile ricordare che tutti i volantini, tranne uno, si sono dimostrati realmente inerenti al tema della violenza: un volantino infatti è un estraneo perché non parla di episodi violenti, ma ha ugualmente un *TFIDF* elevato (= 1,41) poiché nel testo si parla di *pugni*. Si tratta però di una metafora perché, citando testualmente, la frase che ha ingannato TaLTaC è la seguente:

“Sarebbe come se il padrone e l’operaio si scontrassero in una gara di boxe sul ring: il padrone usando tutt’e due i *pugni* (quello economico e quello politico), l’operaio usando un *pugno* solo, quello sindacale, e tenendo l’altro dietro la schiena...”

6.2 I “verbi violenti”

Nel paragrafo precedente, l’estrazione di volantini rilevanti rispetto al tema della violenza è stata portata a termine con successo; tuttavia è possibile approfondire ulteriormente l’analisi di questa tematica, che merita davvero grande attenzione per la

centralità che la violenza e la tensione sociale più in generale, hanno avuto nel decennio (ultimi anni in particolare) descritto nei volantini oggetto di questa relazione. L'approfondimento può essere fatto a partire dalla formazione di una nuova lista di parole-chiave, che esprimano nuovamente violenza, ma con un'accezione diversa rispetto alle parole scelte in precedenza che descrivevano armi e azioni violente, nel tentativo stavolta di allargare il campo dell'analisi, non solo alla violenza fisica, ma anche a quelle tensioni violente, palpabili nei numerosi scontri e dualismi del tempo. Per semplificare la scelta delle parole-chiave si è scelto di selezionare, in maniera specifica ed esclusiva, una lista di verbi con i quali viene espressa, nei volantini, la tensione dell'epoca: alcuni sono verbi abbastanza frequenti nel *corpus* come *colpire* e *impedire* (85 occorrenze ciascuno), *lottare* (83) e *conquistare* (66), mentre per quanto riguarda i meno frequenti si è scelto di selezionare solo verbi con almeno 5 occorrenze; alcuni verbi esprimono poi violenza in senso stretto come ad esempio *abbattere* (7), *strappare* (21), *stroncare* (10), mentre altri descrivono una violenza più psicologica come *intimidire* (6), *reprimere* (7), *affrontare* (51). I verbi selezionati sono 44, per un totale di 885 occorrenze. L'obiettivo è quello di estrarre prima di tutto i volantini più rilevanti, come fatto in precedenza, per poi spostare l'analisi a livello di *subcorpora*, nel tentativo di confrontarli, per scoprire anche in questo caso il più rilevante dei quattro, cioè quello in cui sono più presenti i “verbi violenti” selezionati.

Rispetto alla prima analisi del *TFIDF*, questa volta le parole-chiave sono molte di più e soprattutto è decisamente superiore il totale delle loro occorrenze, di conseguenza i volantini con *TFIDF* diverso da zero, cioè rilevanti - almeno un po' - rispetto alla lista creata, sono molti più di prima (per la precisione 376, più della metà dei 650 volantini): 162 presentano infatti un valore del *TFIDF* pari a 1, mentre sono addirittura 214 quelli con un valore maggiore. L'aumento è evidente se si considera che nel caso della lista delle “armi”, solo 45 volantini presentavano un valore diverso da zero: 32 avevano *TFIDF* pari a 1, mentre solo per 13 era maggiore.

Il volantino che presenta il valore più alto del *TFIDF* (= 2,60), scritto dal P.C.I. di Padova, descrive i tragici fatti avvenuti a Pisa nell'ottobre del 1969, quando, nel corso di una manifestazione di protesta dei lavoratori proclamata in seguito ad un'aggressione fascista a due studenti di sinistra avvenuta il giorno precedente, la

polizia “prendendo a pretesto l’azione di un piccolo gruppo di “Potere Operaio” [...], si scatena in una feroce aggressione contro i lavoratori e la popolazione” che sfocia nella morte di un manifestante, lo studente Pardini, “colpito in pieno petto da una bomba lacrimogena”.

Il contenuto di questo volantino ripropone quindi, a proposito di violenza, gli stessi temi che erano emersi nella precedente analisi del *TFIDF*, cioè le provocazioni fasciste e la repressione poliziesca.

Se questo primo volantino non aggiunge quindi novità alle tematiche con cui si sta cercando di spiegare il clima di tensione e violenza degli anni '60, diverso è il discorso riguardante il secondo “in classifica”: questo volantino infatti, redatto dalla sezione del P.C.I. di Battaglia Terme, presenta un valore del *TFIDF* pari a 2,59, vicinissimo quindi a quello del frammento appena descritto, ma non parla dei temi violenti già incontrati, se si esclude un breve accenno all’ “incredibile appoggio di Moro [Aldo, allora Presidente del Consiglio, ndr] all’aggressione americana nel Vietnam”. Il volantino viene infatti redatto periodicamente dalla sezione, insieme ai lavoratori di una fabbrica della zona, la “Galileo”: il titolo è “Unità operaia” e i temi trattati variano dalla politica con il “fallimento del centrosinistra”, alle problematiche interne alla fabbrica stessa. Si tratta quindi di un ottimo esempio di volantino che non tratta di violenza in senso stretto, ma più che altro di tensione, di uno scontro intenso e costante tra le “pretese conservatrici della D.C.” e la “politica popolare” dei comunisti, tra i grandi monopoli, interessati ad “aumentare la produzione, riducendo l’orario di lavoro e pagando meno salari”, e i lavoratori, decisi e instancabili nella difesa del loro posto di lavoro e del loro già basso salario.

La lista degli altri volantini che presentano un alto valore del *TFIDF* ripropone sempre questo doppio binario di interpretazione della violenza: da una parte quella fisica, che si palesa in episodi di repressione, attentati e bombardamenti, dall’altra la violenza delle parole, delle emozioni e soprattutto delle contrapposizioni già descritte molto spesso, che creavano un clima di tensione costante tra le parti in gioco.

Spostando l’analisi a livello di *subcorpora*, cercando cioè di capire in quale delle quattro partizioni compare più di frequente il tema della violenza, che dipende come si è ormai capito, dalla presenza nei volantini delle parole-chiave selezionate, si nota

innanzitutto che tra i dieci frammenti con *TFIDF* più elevato, nessuno proviene dal *subcorpus* del “mondo civile”.

Da questa prima indicazione quindi, sembra che nell’ultimo *subcorpus* si parli poco di tensione e violenza, o quantomeno se ne parli in misura minore rispetto agli altri tre gruppi. Per capire se questa considerazione è giustificata e per ottenere altre informazioni utili per confrontare i *subcorpora* sul tema della violenza, si può calcolare una misura concreta relativa alla presenza di quelli che abbiamo definito “verbi violenti” nei quattro vocabolari con l’aiuto delle sub-occorrenze, rapportando per ogni *subcorpus*, il totale delle frequenze dei “verbi violenti” presenti al suo interno al numero delle sue sub-occorrenze (*word token*): il risultato, moltiplicato per un fattore (mille o cento, ad esempio), indica quante volte uno dei verbi selezionati compare all’interno di ciascun *subcorpus*, ogni mille (o cento) parole.

Si scopre che l’indicazione riferita all’ultimo *subcorpus* era corretta, dato che nei volantini che lo compongono viene utilizzato uno di questi verbi solo 2,62 volte ogni mille parole utilizzate. Di poco più alto invece il risultato per i volantini scritti dalle sezioni locali del P.C.I., nei quali compaiono, sempre ogni mille parole, 2,70 “verbi violenti”.

Questi primi due valori sono simili e piuttosto bassi, per gli altri due *subcorpora* invece i risultati sono praticamente identici, oltre che maggiori ovviamente: 3,91 ogni mille tra i volantini scritti da associazioni del mondo del lavoro, solo un centesimo in più (3,92) invece, per i volantini scritti da organi di partito del P.C.I..

Stando sempre ben attenti a tener presente il fatto che la scelta delle parole-chiave (verbi in questo caso) ha un peso determinante sui risultati ottenuti, sembra che si possa dire che i *subcorpora* che trattano di più il tema della violenza siano il primo e il terzo. Bisogna dire che questo risultato non può considerarsi del tutto inaspettato, poiché si è già detto più volte che le tensioni si sviluppavano principalmente a causa dei due grandi dualismi dell’epoca, tra comunisti e democristiani e tra lavoratori e padroni, e infatti i due *subcorpora* più rilevanti rispetto al tema della violenza/tensione sono composti da volantini firmati proprio dai protagonisti di questi scontri: i comunisti (primo *subcorpus*) e i lavoratori (terzo *subcorpus*).

CAPITOLO TERZO:

Analisi del *corpus* per unità testuali complesse (CTU)

L'analisi per forme grafiche semplici, come si è visto, è molto corposa e stimolante e permette di studiare il *corpus* da diverse angolazioni, tuttavia presenta alcuni svantaggi: considerando singolarmente ogni forma grafica si tendono a perdere i riferimenti dal contesto, perché ogni parola potrebbe avere un senso e un'interpretazione differente se considerata congiuntamente ai termini che la seguono o la precedono (come nel caso della parola *punto* nelle espressioni *mettere_a_punto*, *punto_di_partenza* e *punto_a_capo*); inoltre non si riescono a risolvere alcune ambiguità della lingua italiana, come ad esempio la parola *faccia* che è sia sostantivo che voce del verbo fare (Bernardi e Tuzzi, 2006).

Ricodificare le forme grafiche semplici in unità testuali più complesse è quindi necessario per aumentare la portata informativa delle forme grafiche stesse e approfondire di conseguenza il livello di conoscenza del *corpus*.

1. NUOVA NORMALIZZAZIONE e NOVITA' NEL VOCABOLARIO

Il primo passo per attuare questo secondo livello dell'analisi consiste nel ritornare a TaLTaC per svolgere nuovamente la normalizzazione, spuntando però questa volta, per quanto riguarda la normalizzazione basata su liste, anche la voce "Gruppi nominali e Polirematiche", al fine di riconoscere nel *corpus* quelle sequenze di parole che, nella lingua italiana, acquistano senso o cambiano senso se considerate congiuntamente. Il *software* può così riconoscere queste sequenze di parole come forme grafiche diverse rispetto ai termini che le compongono e calcolarne le occorrenze come per qualunque altra forma grafica, aggiungendole dunque al vocabolario.

In seguito alla normalizzazione aumenta quindi il numero di forme grafiche presenti nel vocabolario e cambiano di conseguenza, in modo più o meno significativo, le misure del *corpus* e i parametri che ne descrivono la ricchezza lessicale. Com'era già

successo nella prima parte dell'analisi, però, è meglio aspettare prima di presentare le nuove misure lessicometriche perché il vocabolario sarà presto soggetto a ulteriori variazioni.

Nel frattempo si può notare che tra le nuove forme grafiche presenti nel vocabolario, la prima che si incontra nell'ordinamento decrescente per numero di occorrenze è *Partito_Comunista* (206 occorrenze), a cui si accompagna come sempre la *Democrazia_Cristiana* (85). Anche tra le altre forme più frequenti se ne incontrano molte decisamente interessanti, che chiariscono meglio di qualunque spiegazione teorica l'utilità di questa operazione di normalizzazione: *Consiglio_comunale* (161), *forze_politiche* (69), *costo_della_vita* (62), *opinione_pubblica* (54), *sciopero_generale* (50), *sviluppo_economico* (50), *condizioni_di_vita* (48), *aumento_dei_prezzi* (45) - e molte altre ancora - unità testuali composte da parole che unite insieme acquistano molto più senso che separate singolarmente, per non parlare dell'evidente aumento, a livello tematico e linguistico, della portata informativa dei dati contenuti nel vocabolario.

2. ESTRAZIONE DEI SEGMENTI RIPETUTI E LESSICALIZZAZIONE

Il processo di disambiguazione delle forme presenti nel vocabolario prosegue con l'estrazione dei segmenti ripetuti, sequenze di forme grafiche composte da tutte le disposizioni a 2, 3, ..., n forme che si ripetono per un certo numero di volte nel *corpus* (Bolasco, 1999).

Per procedere all'estrazione dei segmenti ripetuti, TaLTaC chiede di inserire alcuni valori per comprendere quali e quanti segmenti estrarre: una soglia di frequenza minima (i) delle parole appartenenti al segmento, il numero massimo di parole presenti nei segmenti da estrarre e un'indicazione (facoltativa) riguardo al numero minimo di occorrenze (n) che deve presentare un segmento per essere estratto.

È evidente che non avrebbe senso fissare $i < n$ (un segmento costituito da almeno una parola che ha 3 occorrenze, ad esempio, non può presentarsi con 4 occorrenze) ed è

altrettanto scontato che il numero massimo di parole presenti nei segmenti dev'essere ragionevolmente limitato (generalmente tra 5 e 8).

Nel nostro caso si è scelto di fissare $i = n = 6$, estraendo così solo i segmenti che si ripetono almeno 6 volte nel *corpus*, e il numero massimo di parole presenti nei segmenti pari anch'esso a 6, considerato sufficiente: il risultato è rappresentato da una lista di 3576 segmenti ripetuti.

L'estrazione dei segmenti ha come obiettivo finale la procedura della "lessicalizzazione", attraverso la quale il software viene istruito a riconoscere i segmenti che vogliamo trasformare in unità testuali complesse, fino a modificare il vocabolario (Giuliano e La Rocca, 2008). Questa lessicalizzazione si deve effettuare a partire da una lista di segmenti ripetuti, scelti tra tutti quelli estratti in precedenza: proprio questa scelta e le motivazioni che portano alla selezione sono i successivi problemi da affrontare.

Innanzitutto, già ad una prima lettura della lista di segmenti estratta da TaLTaC, si nota che sono presenti sia segmenti estremamente significativi come ad esempio *scudo_crociato*, simbolo della D.C. e *Porto_Marghera*, grande polo chimico già negli anni '60, ma anche segmenti, come ad esempio *non_superano* e *dei_lavoratori*, che non sembrano rilevanti, dato che si tratta semplicemente di sequenze di parole che si ripetono spesso, ma che non hanno un significato diverso se considerate congiuntamente; segmenti che possono quindi essere ignorati, lasciando separate nel vocabolario le parole che li compongono.

Se una semplice lettura dei segmenti può già permettere di decidere, per alcuni di essi, se esportarli nella lista da lessicalizzare o se ignorarli, tuttavia sarebbe complicato e decisamente dispendioso leggere uno per volta tutte le migliaia di segmenti selezionati. Fortunatamente esiste anche in questo caso la possibilità di calcolare una misura statistica che può servire da guida: si tratta dell'indice *IS*, un indice di significatività dei segmenti per valutare la loro rilevanza all'interno del *corpus*.

L'indice *IS* mostra il grado di assorbimento di ciascun segmento ripetuto rispetto alle parole che lo costituiscono: ad esempio (non riferito al nostro *corpus*), se il segmento *classe_dirigente* ha frequenza 9 e le parole *classe* e *dirigente* hanno rispettivamente 10 e 9 occorrenze, si può dire che il segmento in questione assorbe il 90% delle occorrenze

della parola *classe* e il 100% delle occorrenze della parola *dirigente*, pertanto l'indice vale 0,9 (Guida in linea di TaLTaC²). Questa informazione consente di concludere che è poco informativo trattare le due forme separatamente, perché nel *corpus* si parla quasi esclusivamente di *classe_dirigente* e non ad esempio di *prima_classe* o di *dirigente_scolastico*. Ovviamente, un segmento sarà da considerarsi tanto più rilevante quanto più forte sarà il suo grado di assorbimento rispetto alle parole che lo compongono.

TaLTaC fornisce anche una seconda versione di questo indice, l'indice *IS* relativo, che si ottiene rapportando l'indice assoluto al suo massimo e che presenta l'evidente vantaggio di variare tra 0 e 1.

I due indicatori offrono informazioni diverse. L'indice *IS* assoluto è fortemente condizionato dal numero di parole piene che costituiscono il segmento, pertanto mette in evidenza i segmenti più lunghi e costituiti da un maggior numero di parole, ma anche meno frequenti. L'indice *IS* relativo privilegia invece i segmenti più corti, che spesso rappresentano i termini specialistici del lessico, presentandoli ai primi ranghi, cioè tra quelli particolarmente rilevanti.

A livello di interpretazione dei risultati forniti dai due indici, com'è stato chiarito con l'esempio precedente, i segmenti che presentano un grado di assorbimento più elevato, essendo dei poliformi che conviene trattare come un'unica forma grafica piuttosto che attraverso le forme grafiche che li compongono, sono ottimi candidati per far parte della lista di lessicalizzazione.

Per ridurre la quantità di segmenti da considerare, si possono innanzitutto ordinare gli stessi segmenti in base all'indice *IS* relativo ed eliminare, grazie allo strumento di *Text Data-Mining* offerto da TaLTaC, tutti quelli che presentano un valore minore di una soglia fissata a priori, che può ad esempio essere 0,10 (ricordando che l'indice *IS* relativo varia tra 0 e 1).

Si potrà obiettare che è una misura drastica, tuttavia si dimostra pratica e ragionevole: pratica perché eliminando solo i segmenti con un valore particolarmente basso - minore di 0,10 - dell'indice *IS* relativo, si riesce già a ridurre il numero totale dei segmenti del 75% circa (ne restano solo 879 rispetto ai 3576 iniziali), ragionevole perché scorrendo velocemente i segmenti che si è scelto di eliminare - prima di procedere con l'eliminazione stessa - si osserva che effettivamente non se ne incontrano di particolarmente significativi ai fini della lessicalizzazione.

A questo punto si può ritenere “trattabile” la lista di segmenti rimasti, la cui quantità non è più esagerata, e procedere di conseguenza al controllo manuale della stessa lista, segmento per segmento, spuntando solo quelli ritenuti realmente adatti a diventare parte del vocabolario delle forme grafiche, come entità separate dalle parole che li compongono.

La selezione ha portato alla creazione di una lista contenente 310 segmenti, che diventeranno quindi 310 nuove forme grafiche del vocabolario, con la procedura della lessicalizzazione.

Alcuni tra i segmenti estratti, oltre ai già citati *scudo_crociato* e *Porto_Marghera*, sono davvero molto significativi e aumentano notevolmente la loro portata informativa rispetto alle singole parole da cui sono formati: ad esempio *mercato_comune_europeo*, la cui sigla - M.E.C. - è già stata trovata tra le forme specifiche del primo *subcorpus* ; *classe_operaia* formata da *classe* che poteva essere interpretata in modi diversi se considerata a sé stante (aula, talento, ceto sociale, ecc.) e da *operaia* aggettivo che nel *corpus* è sicuramente riferito ad una moltitudine di sostantivi diversi; segmenti anche molto lunghi come *federazione_giovanile_comunista_italiana*, organo che, come detto, firma una gran quantità di volantini, o segmenti formati da parole vuote, ma che a livello linguistico valgono effettivamente come forme uniche, *per_quanto_riguarda* ad esempio.

Figura 3.1 *Misure Lessicometriche del corpus dopo la lessicalizzazione.*

		Post lessicalizzazione	Post seconda normalizzazione	Forme grafiche semplici
Word Token	N =	250.836	256.587	259.666
Word Type	V =	18.244	17.956	17.708
Type/Token Ratio	V/N % =	7,27%	7,00%	6,82%
% hapax	V1/V % =	44,66%	45,25%	45,45%
Freq. media generale	N/V =	13,75	14,29	14,66

Con la lessicalizzazione si va a modificare per l’ultima volta il vocabolario delle forme grafiche, aggiungendo i segmenti selezionati a tutte le forme già presenti. È ovvio quindi che, per l’ennesima volta - e stavolta davvero ultima - si riscontrino cambiamenti anche nelle misure lessicometriche del *corpus* (Figura 3.1).

Leggendo la tabella (Figura 3.1 a pag. 47) da destra a sinistra si può osservare la successione di questi cambiamenti nel corso dell'analisi (da destra a sinistra) e si nota subito che, rispetto all'inizio, sono presenti più di 500 nuove forme grafiche distinte (*word type*): con la normalizzazione infatti, sono state riconosciute 248 unità testuali complesse formate a partire da forme grafiche semplici già presenti nel *corpus*, mentre con la lessicalizzazione il valore è aumentato di altre 288 unità. Come diretta conseguenza di questo aumento, la quantità di occorrenze totali (*word token*) è scesa quasi fino a 250 mila (da 259.666 a 250.836, cioè quasi novemila in meno), poiché con il riconoscimento delle unità testuali complesse e con l'inserimento di nuovi segmenti, due, tre, o più forme grafiche semplici sono state accorpate più volte in un'unica forma complessa (ritornando all'esempio di prima, *punto_di_partenza* viene conteggiato quindi come una sola occorrenza e non più come tre).

Per gli altri indicatori presenti tra le misure lessicometriche non si notano invece variazioni di particolare rilievo: la ricchezza lessicale misurata dal TTR ad esempio è aumentata di quasi mezzo punto percentuale (da 6,82% a 7,27%), come ovvio risultato dell'aumento delle forme distinte al numeratore (V), unito al decremento delle occorrenze al denominatore (N), mentre la percentuale di *hapax* è leggermente diminuita (da 45,45% a 44,66%).

A un occhio attento non sarà sfuggito che l'aumento delle forme grafiche non è esattamente uguale al numero di segmenti inseriti nel vocabolario con la lessicalizzazione, ma è invece minore (+288 invece che +310); fortunatamente però non si tratta di un errore e c'è una spiegazione anche per questa apparente anomalia: quando una forma grafica semplice viene riconosciuta all'interno di un segmento ripetuto, perde una certa parte delle sue occorrenze, che vengono in un certo senso "affidate" al segmento che viene inserito nel vocabolario. È possibile, ed è evidentemente quello che è accaduto, che per alcune forme semplici, quella "certa parte" appena descritta, non sia solo una parte, bensì la totalità delle occorrenze; significa cioè, che nel *corpus* alcune forme grafiche si presentano sempre come parte di un segmento e di conseguenza dopo la lessicalizzazione, scompaiono dal vocabolario come forme grafiche semplici, per apparire solamente accompagnate ad altre forme.

Lo stesso attento osservatore di poco fa, esaminando la sezione di vocabolario contenente solo i segmenti estratti per la lessicalizzazione, non potrebbe fare a meno di notare altre particolarità e porre di conseguenza alcune domande:

- noterebbe che, anche se avevamo chiesto a TaLTaC di estrarre solamente i segmenti ripetuti con almeno sei occorrenze, sono presenti ben nove segmenti con cinque occorrenze, o addirittura meno;
- osserverebbe poi, che in alcuni casi due o più segmenti sembrano incastrarsi l'uno nell'altro come piccole matriske: si possono leggere ad esempio *collettivo_studenti*, ma anche *collettivo_studenti_medi*, oppure *federazione_giovanile*, *federazione_giovanile_comunista* e *federazione_giovanile_comunista_italiana*; fino a casi ancora più complicati come il terzetto formato da *alleanza_coltivatori_diretti*, *coltivatori_diretti_e_mezzadri* e *alleanza_coltivatori_diretti_e_mezzadri*.

Per fortuna si possono risolvere entrambe le questioni con una sola spiegazione che dovrebbe essere più chiara con l'ausilio di un esempio concreto, scelto tra quelli appena citati: spesso nominando la *federazione_giovanile_comunista* si dava per scontato che si parlasse di quella *italiana* e specificarlo non era quindi necessario; inoltre di *federazione_giovanile* non c'era solo quella comunista, ma anche altre come quella socialista ad esempio e questo spiega il motivo per cui nel *corpus* si ritrova anche il segmento *federazione_giovanile*, senza ulteriori aggettivi.

Nel vocabolario la *federazione_giovanile* risulta avere solo 12 occorrenze, anche se le parole *federazione* e *giovanile* compaiono affiancate anche altre 52 volte, ma con l'aggiunta dell'aggettivo *comunista* e in 23 casi anche di *italiana*: TaLTaC agisce in modo tale da considerare per primo il segmento più lungo, assegnando ad esso le occorrenze che gli competono, per poi passare a quello appena più corto e infine a quello più corto ancora che si dovrà "accontentare" delle occorrenze rimaste.

Questo spiega anche perché tra i segmenti aggiunti al vocabolario, alcuni hanno meno di sei occorrenze, contravvenendo alle nostre richieste: *collettivo_studenti*, per esempio, si presenta addirittura con un'unica occorrenza, nonostante ne avesse inizialmente undici, "per colpa" del segmento *collettivo_studenti_medi* che gliene ha "scippate" ben dieci.

3. SUB-OCCORRENZE, SEGMENTI COMUNI ED ESCLUSIVI

248 unità testuali complesse riconosciute con la normalizzazione dei gruppi nominali e delle polirematiche e 310 segmenti estratti e inseriti nel vocabolario con la lessicalizzazione: queste sono le cifre che rappresentano i cambiamenti nel nostro *corpus* rispetto all'analisi per forme grafiche semplici. Detto degli (scarsi) cambiamenti delle misure lessicometriche, può essere molto interessante anche cercare di scoprire qualcosa di nuovo per quanto riguarda le forme comuni, esclusive e soprattutto specifiche, valutando se tra di esse si trovano ora anche alcune delle forme grafiche complesse che sono state aggiunte al vocabolario. Nei prossimi paragrafi si cercherà di dare una risposta a questi quesiti, non prima però di aver ricalcolato le misure lessicometriche dei *subcorpora* (Figura 3.2) per valutarne le eventuali variazioni.

Figura 3.2 Misure lessicometriche dei subcorpora dopo la lessicalizzazione e variazioni percentuali.

	1	%	2	%	3	%	4	%
N	90.398	-3,74%	53.619	-3,54%	58.590	-3,33%	48.229	-2,70%
V	10.522	+3,74%	7.705	+3,95%	7.740	+3,39%	8.303	+3,66%
V/N%	11,64%	+0,84%	14,37%	+1,04%	13,21%	+0,86%	17,22%	+1,06%
V1/V%	48,40%	-0,66%	51,72%	-0,16%	48,62%	-0,32%	53,98%	+0,11%
N/V	8,59	-0,67	6,96	-0,54	7,57	-0,53	5,81	-0,38

Ripetere per tutti i *subcorpora* gli stessi commenti già fatti due volte per le misure lessicometriche dell'intero *corpus* (aumento delle forme grafiche e diminuzione delle occorrenze) è chiaramente superfluo, può essere utile però valutare se tra i vari *subcorpora* si notano diversità particolari nella natura e nell'intensità delle variazioni: ad esempio si osserva che per il quarto *subcorpus*, la riduzione del numero di occorrenze (N) è minore, in percentuale, rispetto agli altri tre (2,70% in meno, mentre per gli altri è mediamente del 3,5%); la variazione del *Type/Token Ratio* (V/N%) che misura la ricchezza lessicale è invece molto simile tra i *subcorpora*, in particolare tra il primo e il terzo e tra il secondo e il quarto; stupisce un po', infine, notare che la percentuale di *hapax* (V1/V%) è diminuita per tutti i *subcorpora*, tranne che per l'ultimo dove si è verificato un piccolo aumento (+ 0,11%).

3.1 Segmenti comuni

Dei 310 segmenti inseriti nel vocabolario, più di un terzo, 110 per la precisione, sono quelli comuni a tutti i *subcorpora*. Si possono ritrovare tra questi segmenti quelli più frequenti, in particolare i primi due che hanno un numero di occorrenze decisamente maggiore di tutti gli altri e si possono definire tematici per ciò che rappresentano nel *corpus* : *classe_operaia* con 257 occorrenze ben distribuite tra i *subcorpora* e *Partito_Comunista_italiano* (168) rappresentano infatti i primi protagonisti dei volantini e delle tematiche in essi trattate.

I segmenti successivi, anch'essi comuni, presentano molte meno occorrenze rispetto ai primi due e descrivono elementi non altrettanto condivisi da tutti i *subcorpora* : *commissione_interna* ad esempio (89 occorrenze), oltre ad essere tematicamente ascrivibile al terzo *subcorpus* in particolare (mondo del lavoro), non è nemmeno ben distribuito, pur essendo comune a tutti i gruppi; 76 volte su 89 compare infatti proprio tra i volantini del terzo *subcorpus*.

Si può dire lo stesso di *amministrazione_comunale* (84) e di *federazione_di_Padova* (71) che, anche se comuni, compaiono per la maggior parte delle volte nel secondo e nel primo *subcorpus* rispettivamente (sezioni locali e organi di partito).

Tra gli altri segmenti comuni maggiormente presenti nel *corpus* si trovano inoltre alcuni segmenti “lessicali”, cioè vuoti dal punto di vista del contenuto, ma molto utilizzati nella lingua italiana, come ad esempio *ancora_una_volta* (88) e *nello_stesso_tempo* (35).

Anche questa volta i segmenti sono molti e non possono essere elencati uno per uno, quindi dovendo scegliere, dopo aver descritto alcuni dei più presenti, può essere interessante valutare quelli opposti, cioè meno presenti, con il solito aiuto dell'analisi delle concordanze, per scoprirne i contesti di utilizzo: ad esempio gli ultimi due segmenti comuni in classifica, per numero di occorrenze, sono *particolare_attenzione* (6) e *luoghi_di_lavoro* (6) .Viene spontaneo quindi chiedersi a cosa bisognava prestare particolare attenzione e chi doveva farlo, e perché si parla di luoghi di lavoro e quali problemi si verificano al loro interno, se di problemi si tratta.

Partendo da questo secondo caso si scopre che si parla eccome di problemi: dalle “limitazioni di libertà sui *luoghi_di_lavoro*”, alla necessità di difendere il “diritto dei lavoratori di organizzarsi liberamente nei *luoghi_di_lavoro*”, dall’ “accentuato

sfruttamento all'interno dei *luoghi di lavoro*" alle richieste di "ambulatori situati nei *luoghi di lavoro*", che dovevano essere sotto il controllo delle unità sanitarie locali, per monitorare la salute dei lavoratori che operavano spesso in ambienti insalubri e per "rendere a tutti i livelli il servizio di medicina del lavoro autonomo ed indipendente dai padroni".

Come si può notare, le difficoltà sul posto di lavoro erano molteplici e se da un lato è confortante osservare che la consapevolezza di queste difficoltà era ampia e condivisa sia dal mondo politico, nazionale e locale, che dalle associazioni del mondo civile, oltre che ovviamente dalle organizzazioni stesse dei lavoratori, d'altro canto il sentimento è opposto se si considerano i numerosi fatti di cronaca, che anche attualmente riempiono con triste frequenza le pagine dei quotidiani, raccontando incidenti e troppo spesso anche morti sui *luoghi di lavoro*, oltre alle numerose fabbriche (*luoghi di lavoro* appunto) che a causa delle difficoltà economiche sono costrette a chiudere, creando sconforto tra i lavoratori, destinati alla disoccupazione o, nel migliore dei casi, alla cassa-integrazione.

Valutando invece le domande relative a chi doveva prestare *particolare attenzione* e a che cosa, si ottengono risposte decisamente diversificate: il P.C.I. nel tentativo di difendere la salute dei lavoratori, chiedeva al Governo (curiosamente proprio nello stesso volantino dove si promuoveva la presenza di ambulatori situati nei *luoghi di lavoro*) che prestasse *particolare attenzione* "allo studio e classificazione degli agenti chimici nocivi alla salute" tra quelli utilizzati nelle fabbriche, mentre i sindacati nel 1970 reclamavano riforme immediate per risollevare il Paese e a proposito di misure anticongiunturali, invitavano a porre *particolare attenzione* "alle misure tese al consolidamento e alla diffusione più equilibrata nel territorio delle aziende minori". Una forte e improvvisa grandinata scesa nell'agosto del 1970, costringeva invece alcune amministrazioni comunali della zona del Montagnanese a chiedere un intervento pubblico "in favore dei produttori agricoli per la perdita subita dei vari prodotti", con *particolare attenzione* alle piccole e medie aziende; mentre i comunisti di Cadoneghe, ironizzando sull'operato dell'amministrazione comunale democristiana e sulle sue "grandi realizzazioni", invitavano i cittadini a prestare *particolare attenzione* ai "criteri adottati nell'applicazione dell'imposta di famiglia" che "dovrebbe essere applicata democraticamente a partire dai redditi più alti".

Anche in questo caso si nota che in tutti i *subcorpora* si trova un esempio di richiesta di *particolare attenzione*, tuttavia, come sono spesso differenti le tematiche e i problemi trattati, così è molto diverso l'oggetto che, di volta in volta, richiede quella stessa attenzione.

3.2 Segmenti esclusivi

Se i segmenti comuni erano molti - quasi un terzo del totale come si è visto - lo stesso non si può dire dei segmenti esclusivi che sono in tutto 25: nove per il primo *subcorpus*, sei nel secondo e nel terzo e solo quattro nell'ultimo.

- Per descrivere i segmenti esclusivi del primo *subcorpus* (volantini scritti da organi di partito), è fondamentale ricordare la spiegazione riguardante quei gruppi, o coppie di segmenti, che sembrano incastrarsi l'uno nell'altro, infatti tra questi segmenti esclusivi si trovano, ad esempio, *federazione_giovanile_comunista_italiana* (23 occorrenze) e *federazione_giovanile* (12), ma non *federazione_giovanile_comunista*: si è già visto che la F.G.C.I. ha firmato molti dei volantini presenti nel *corpus* ed è quindi comprensibile che scrivendoli abbia citato per intero il proprio nome, inoltre si è già avuto modo di spiegare anche che di federazioni giovanili non c'era solo quella comunista, ma anche altre che in genere firmavano i volantini unitamente alla federazione comunista stessa, quindi non stupisce il fatto che il segmento *federazione_giovanile*, senza ulteriori aggettivi, sia esclusivo per questo *subcorpus*. Il segmento *federazione_giovanile_comunista* è invece presente anche in altri *subcorpora*, dove probabilmente chi firmava il volantino tendeva a dare per scontato l'aggettivo 'italiana'.

Allo stesso modo si trovano tra i segmenti esclusivi anche *collettivo_studenti_medi* (10), ma non *collettivo_studenti*, e *alleanza_coltivatori_diretti_e_mezzagadri* (6), ma non *alleanza_coltivatori_diretti*, né *coltivatori_diretti_e_mezzagadri*.

Gli altri cinque segmenti presenti solo in questo primo *subcorpus* sono: *sede_del_M.S.I.* (9 occorrenze), *governo_quadripartito* (8), *regime_democratico* (7), *1°_maggio* (7), *popolo_basco* (6).

- I sei segmenti esclusivi del secondo *subcorpus* sono decisamente “tematici”. Sono presenti infatti, come per le forme grafiche semplici, nomi di luoghi della provincia padovana, come *colli_Euganei* (8) e *due_Carrare* (7) - i comuni limitrofi di Carrara S. Giorgio e Carrara S. Stefano - e il nome di una sezione del P.C.I. di Padova, *Quattro_Martiri* (6), oltre a *sezioni_comuniste* (6), segmento utilizzato in genere quando più sezioni firmavano insieme un volantino (l'esempio più frequente in tal senso è: “Le sezioni comuniste di Arcella 1° - Arcella 2° - Pontevigodarzere”) e *centri_sociali* (6), altro esempio tipico di associazioni e luoghi di ritrovo e di confronto locali.

In prima battuta può invece stupire l'ultimo segmento esclusivo: infatti forse molti si aspetterebbero di trovare *lavoratori_calzaturieri* (6), tra i segmenti esclusivi del terzo *subcorpus*, cioè tra i volantini scritti da associazioni del mondo del lavoro. In realtà invece, in Veneto come in altre parti d'Italia, accade di frequente che le fabbriche di una zona si specializzino in una stessa produzione, creando un legame forte tra quel tipo di prodotto e il territorio: è questo il caso dei calzaturifici di Noventa Padovana, infatti il segmento *lavoratori_calzaturieri* si trova in volantini firmati dalla sezione del P.C.I. di Noventa, proprio perché, in questo contesto sociale, i problemi di una fabbrica vengono fortemente sentiti da tutta la popolazione della zona.

- Anche per il terzo *subcorpus*, alcuni segmenti esclusivi sono strettamente legati al tema trattato nei volantini, cioè il lavoro: è il caso di *Commissione_fabbriche* (29 occorrenze), *reparto_confezioni* (8) e *coltivatori_diretti_e_mezzadri* (2). Richiedono invece un approfondimento gli altri tre segmenti esclusivi: bisogna chiarire innanzitutto che *10.000_lire* (8) era la richiesta mensile dei lavoratori di alcune fabbriche padovane, relativamente al premio di produzione, mentre *collettivo_studenti* è il segmento, già incontrato in precedenza, rimasto con un'unica occorrenza a causa del segmento *collettivo_studenti_medi*. Per descrivere l'ultimo segmento esclusivo è necessaria, infine, una breve digressione storica: alla fine degli anni '60, per il Partito Comunista era estremamente importante informare i lavoratori sui possibili vantaggi dell'affermazione legislativa dello “Statuto dei diritti dei lavoratori”, in modo tale da ottenere il sostegno

popolare prima di affrontare la questione in Parlamento: per questo furono organizzati numerosi incontri, pubblicizzati in alcuni volantini, tra gli operai delle fabbriche padovane e una *delegazione parlamentare* (6), formata da politici comunisti competenti in materia.

- Concludendo, resta l'ultimo *subcorpus*, formato da volantini redatti da associazioni del mondo civile. C'è però poco da dire in proposito, perché i segmenti esclusivi sono solo quattro e due di essi si presentano una sola volta: *piani_liberi* (8), *numero_442* (6), *criminali_attentati* (1), *giorni_scorsi* (1). Il segmento *piani_liberi* si presenta sette volte su otto all'interno dello stesso volantino, redatto da una cellula universitaria del P.C.I. e intitolato appunto "Piani liberi e svalorizzazione", mentre si è già spiegato il significato di *numero_442* tra le forme grafiche semplici, esclusive per questo *subcorpus*, dove si era incontrato solo il numero - 442 appunto - della legge in questione (legge del 12 marzo 1968 che dettava alle università le norme per l'istituzione del quinto anno di corso, necessario per il conseguimento della laurea con valore abilitante). I due segmenti con occorrenza pari a uno - *criminali_attentati* e *giorni_scorsi* - invece sono entrambi "orfani" delle occorrenze che sono state loro sottratte, per essere conteggiate all'interno dei segmenti: *criminali_attentati_di_Milano* e *nei_giorni_scorsi*.

4. SEGMENTI SPECIFICI

Si è già spiegato, nel corso dell'analisi per forme grafiche semplici, il procedimento che porta a riconoscere le forme specifiche, ora è sufficiente ricordare che affinché un'unità testuale (complessa in questo caso) possa essere riconosciuta come specifica di un *subcorpus*, deve essere sovrautilizzata rispetto agli altri *subcorpora* in relazione ad una soglia di probabilità prefissata e dev'essere distribuita in modo sufficientemente uniforme all'interno del *corpus*.

Rileggere tutte le forme specifiche presenti nel vocabolario non è necessario perché si ritroverebbero le forme già analizzate in precedenza, è sufficiente a questo punto valutare l'eventuale specificità solo delle 310 unità testuali complesse inserite nel vocabolario con la lessicalizzazione.

Bisogna dire innanzitutto che c'è un'evidente disparità tra i *subcorpora* nel numero di segmenti specifici riconosciuti: il *subcorpus* che presenta il maggior numero di specificità positive è il primo, che ne ha 24, per il secondo invece ne vengono riconosciute 18 e per il terzo 22, ma è l'ultimo *subcorpus* a dimostrarsi decisamente differente da questo punto di vista, presentando solo 8 segmenti specifici.

- Nel corso dell'analisi delle forme grafiche semplici specifiche per il primo *subcorpus*, si era già spiegata la presenza tra le forme specifiche delle parole *federazione*, *giovanile* e *italiana*, con il fatto che la firma del volantino è parte integrante del testo. Questa spiegazione ritorna nuovamente utile a questo punto leggendo i segmenti che si presentano con il più alto grado di specificità in questo primo *subcorpus* : *federazione_di_Padova*, il terzetto *federazione_giovanile_comunista_italiana* - *federazione_giovanile_comunista* - *federazione_giovanile*, *Partito_Comunista_italiano* e *collettivo_studenti_medi* sono tutte firme di volantini, precisamente le firme degli organi di partito raccolti in questo primo *subcorpus*. Non proprio tutte a dire il vero, perché il *collettivo_studenti_medi* non viene considerato organo di partito chiaramente, ma la sua presenza qui è giustificata dal fatto che spesso firmava volantini unitamente alla F.G.C.I..

Anche altri segmenti specifici di questo primo gruppo ricordano forme specifiche semplici già incontrate in precedenza, come *studenti_medi* (*studenti* era già una specificità) e *masse_popolari* e *masse_lavoratrici* (in questo caso ad essere specifica era *masse*), ed è chiaro che i segmenti risultano essere decisamente più informativi ed interessanti da analizzare in questo contesto.

Tra gli altri segmenti specifici alcuni rimandano al tema del lavoro, come *movimento_operai*, *lotte_operaie* e *grande_padronato*, ma quando si leggeranno le specificità del terzo *subcorpus* (volantini del mondo del lavoro) si potrà notare l'evidente differenza tra esse e questi tre segmenti.

Le ultime osservazioni sui segmenti specifici di questo primo *subcorpus* rimandano infine alle tematiche riconosciute con il calcolo dell'indice *TFIDF* : si era potuto notare infatti, che i volantini rilevanti rispetto al tema della violenza erano scritti prevalentemente da organi di partito; ritrovare tra questi segmenti specifici l' *imperialismo_americano*, il *popolo_vietnamita*, le *squadre_fasciste* e le *forze_reazionarie* è quindi un risultato del tutto prevedibile.

- Anche per il *subcorpus* formato dai volantini redatti dalle sezioni locali del P.C.I. si riconosce la tendenza a ritrovare, tra le parole che compongono alcuni segmenti specifici, alcune forme grafiche semplici che sono già risultate specifiche nella prima parte dell'analisi: è il caso dell'aggettivo *comunale* in *amministrazione_comunale* e *giunta_comunale* (che sono i due segmenti con il grado di specificità più alto), del termine *comunisti*, che compare stavolta accompagnato dall'ovvia - nel nostro caso - nazionalità (*comunisti_italiani*) e infine, dell'aggettivo *regolatore* riferito al *piano_regolatore*.

Tra le specificità in questo caso si possono anche trovare alcuni temi tipicamente locali, essendo a carico dei Comuni, come le tasse e in particolare l'*imposta_di_famiglia*, il *verde_pubblico*, l'*assistenza_sanitaria* e la *legge_167*, riguardante "l'espropriazione di aree da destinare alla costruzione di abitazioni popolari" e alcune istituzioni e associazioni locali, che a volte si trovano come firma in alcuni volantini, come *comitato_direttivo* (di sezione, ndr), *centro_sociale*, *gruppo_consiliare_comunista*.

Altri due segmenti specifici sono invece interessanti dal punto di vista della comunicazione tra sezioni e cittadinanza: *caro_compagno* ad esempio era la principale forma di saluto con cui la sezione si rivolgeva al lettore del volantino (spesso anche 'cari compagni'), mentre *politica_antipopolare* era uno dei tanti modi per definire e caratterizzare l'operato della Democrazia Cristiana, costantemente accusata di fare solo gli interessi dei ricchi e dei padroni e non certo quelli della popolazione dei Comuni dove governava.

- *Commissione_interna, commissione_fabbriche e lavoratori_italiani*: anche per il terzo *subcorpus*, in tre dei segmenti maggiormente specifici si incontrano forme grafiche semplici, già specifiche anch'esse (*commissione, interna, e lavoratori*). Tra l'altro i primi due rientrano nella solita categoria dei segmenti riconosciuti specifici in quanto firme di volantini e lo stesso dicasi per *comitato_unitario* (curiosamente il plurale - *comitati_unitari* - era specifico per il *subcorpus* precedente).

Analizzando i segmenti specifici del primo gruppo, era stato annunciato un confronto tra alcuni di essi e altri che risultano invece specifici per questo *subcorpus*: *movimento_operai, lotte_operaie* e *grande_padronato* infatti, riguardavano effettivamente il tema del lavoro, ma da un punto di vista più teorico e reclamistico rispetto ai problemi reali e concreti che si incontrano ora tra queste specificità, problemi quali il *premio_di_produzione*, la *medicina_del_lavoro* e la *tutela_della_salute*, gli *assegni_familiari* e le *imposte_indirette*, l'*unità_sindacale*, per concludere poi, con il diritto per tutti ad una *casa_civile* in cui abitare. È evidente da questo confronto la differenza tra politica e mondo del lavoro vero e proprio: la politica, rappresentata dal P.C.I. in questo caso, pur volendo affrontare i problemi dei lavoratori e il tema delle lotte operaie, resta sul vago utilizzando slogan utili e necessari per animare le masse, le associazioni dei lavoratori invece cercano di trovare soluzioni per i problemi reali che vivono tutti i giorni in fabbrica, oppure ogni mese con l'arrivo della - magra - busta-paga.

Concludendo con l'analisi relativa al terzo *subcorpus*, tra i segmenti specifici se ne incontrano anche di "linguistici", come *fare_i_conti* e *andare_avanti*, vuoti dal punto di vista del contenuto, ma non per questo meno interessanti da analizzare per scoprire a cosa si riferiscono. Nelle fabbriche ad esempio, padroni e lavoratori dovevano spesso *fare_i_conti* gli uni con gli altri, in particolare i padroni erano più volte invitati (minacciati in alcuni casi) a *fare_i_conti* con la volontà dei lavoratori e con le loro proposte, viceversa i lavoratori dovevano "continuamente *fare_i_conti* con la prepotenza dei padroni"; a livello politico invece il P.C.I. sosteneva le lotte dei lavoratori,

ribadendo spesso che “con la volontà operaia e popolare dovranno *fare_i_conti* il Governo Rumor e i suoi alleati!”.

La situazione è simile anche per quanto riguarda l’ *andare_avanti*, dato che lo scontro si muove sui due soliti binari: da un lato il Paese che vuole *andare_avanti* anche se i governi D.C. “ [...] tentano di riportarlo indietro”, dall’altra la “lotta unitaria di massa” che deve e vuole *andare_avanti* nonostante la resistenza di padroni e Governo, “per far entrare la volontà dei lavoratori nello Stato” e per proseguire “sulla strada di una sempre più forte avanzata verso nuove conquiste sociali e politiche”.

- Come anticipato a inizio paragrafo, i segmenti specifici dell’ultimo *subcorpus* (volantini scritti da associazioni del mondo civile) sono solamente otto e di conseguenza l’analisi che li riguarda si concluderà in breve tempo.

I segmenti con il grado di specificità più alto sono, anche in questo caso, firme dei volantini - *comitato_cittadino*, *cellula_universitaria* e *movimento_studentesco* - e si nota per l’ennesima volta come i temi trattati in questo *subcorpus* riguardino spesso l’istruzione, la scuola e l’università. Oltre alle due firme infatti, tra i segmenti specifici ci sono anche *scuole_medie* e *asili_nido*, mentre completano il quadro due segmenti “politici”, quali *parlamentari_comunisti* e *consigliere_comunale*, e uno dei principali strumenti di aggregazione per la cittadinanza, nonché fonte di molte idee e proposte, cioè l’ *assemblea_generale* dei cittadini.

CAPITOLO QUARTO

Analisi delle corrispondenze

Volendo approfondire l'esame delle unità testuali complesse, si può cercare di sintetizzare le informazioni sul rapporto tra i *subcorpora* e le *CTU* per mezzo dell'analisi delle corrispondenze (Greenacre, 1984): questa tecnica statistica permette di ricostruire il sistema di relazioni tra parole e autori (nel nostro caso, forme grafiche complesse e firme dei volantini) attraverso la trasformazione di una tabella di contingenza in una rappresentazione grafica, con l'impiego di assi cartesiani utili a collocare univocamente ciascuna parola e ciascun autore mediante coordinate. Sul grafico le posizioni assunte nello spazio da unità testuali e autori sono fondamentali per l'interpretazione, in quanto le parole maggiormente rappresentative di un autore si collocano in posizioni vicine all'autore stesso.

Si è scelto di applicare l'analisi delle corrispondenze solo alle unità testuali complesse e non alle forme grafiche semplici, in quanto le *CTU* hanno dimostrato, nel corso del terzo capitolo, di essere estremamente interessanti e di fornire indicazioni molto più chiare delle semplici parole. Ciò non toglie che un'analisi più corposa e approfondita avrebbe potuto applicare, senza alcun problema, l'analisi delle concordanze anche alle forme semplici più interessanti.

Inizialmente la lista dei 310 segmenti inseriti nel vocabolario con la lessicalizzazione è stata sistemata e sfoltita: era necessario infatti ridurre il numero delle *CTU* da inserire nel grafico per renderlo comprensibile, e si è deciso quindi di non considerare i segmenti vuoti dal punto di vista del contenuto, per limitarsi all'analisi delle tematiche, e di accorpate in uno solo, i segmenti che si presentavano sia nella forma singolare che in quella plurale (ad esempio si potranno osservare nel grafico segmenti come *lotta/e_unitaria/e* e *centro/i_sociale/i*) e anche quelli, già a lungo descritti, che si "incastrano" l'uno nell'altro: il segmento *federazione_giovanile_comunista_(italiana)*, per esempio, rappresenta sia il segmento intero, che quello senza l'aggettivo *italiana*.

L'analisi delle corrispondenze è stata effettuata per mezzo del *software* statistico R che con pochi semplici comandi fornisce un risultato grafico che necessita però di una revisione per poter essere compreso e interpretato in maniera adeguata.

Questo perché con la sistemazione e la riduzione dei segmenti descritte in precedenza, ne sono rimasti 245, ancora troppi per poter essere contenuti tutti nel grafico senza sovrapporsi. Per risolvere il problema è stato necessario agire direttamente sul grafico stesso, valutando ulteriori eliminazioni di segmenti quando due o più si sovrapponevano, o erano troppo vicini per essere leggibili.

Per scegliere quali segmenti eliminare, ci si è basati su alcuni dati presenti nell'*output* prodotto da R : per ogni segmento infatti, viene indicata la “massa”, cioè la frequenza nel vocabolario e i “contributi” che rappresentano quanto una parola è rilevante per ciascuna dimensione: in entrambi i casi, il suggerimento è di eliminare i segmenti che presentano un valore minore, cioè quelli meno frequenti nel *corpus* e quelli che, “contribuendo” in misura minore, hanno minor rilevanza per le dimensioni.

In assenza di indicazioni evidenti di questo tipo, cioè nel caso in cui nel gruppo di segmenti sovrapposti, le differenze in termini di massa e contributi siano minime, si sono scelti discrezionalmente i segmenti più interessanti.

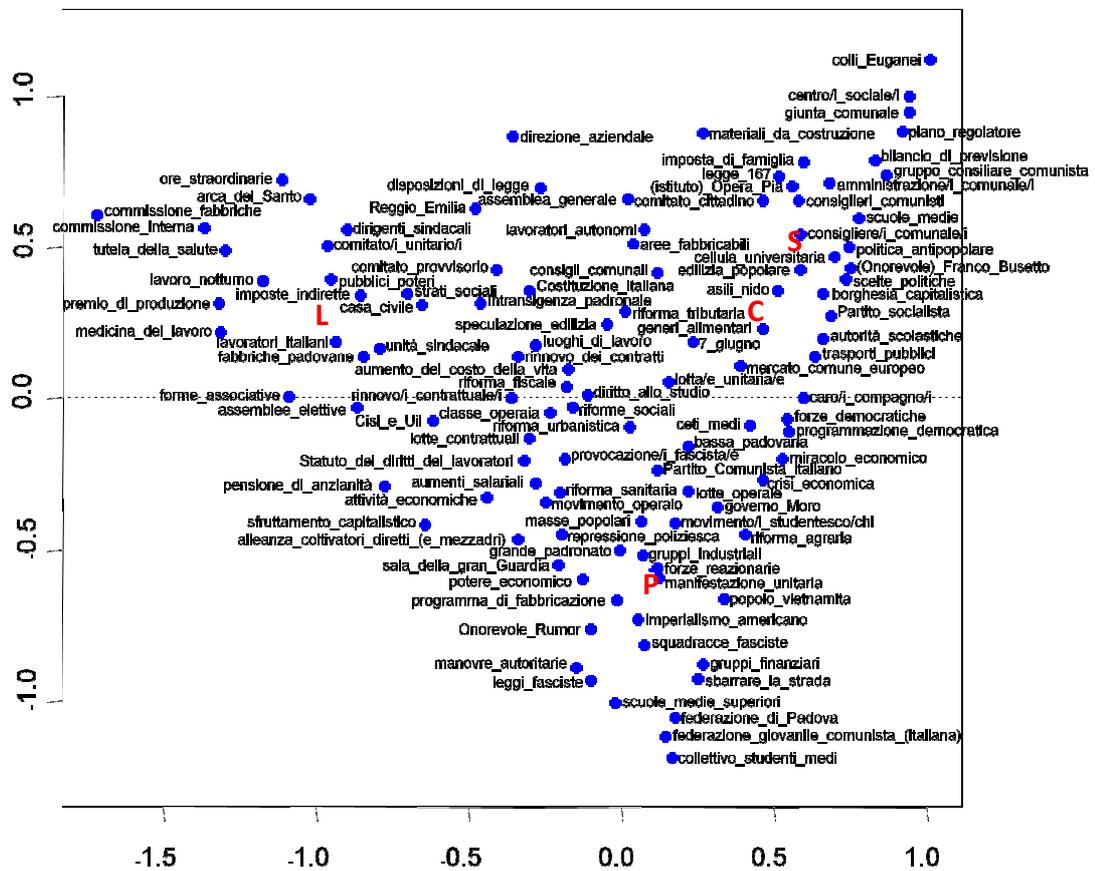
Le sovrapposizioni erano piuttosto frequenti (a volte anche cinque segmenti con identiche coordinate), di conseguenza nel grafico finale (Figura 4.1 presentata qui a fianco, pag. 63) sono presenti “solo” 113 segmenti, cioè meno della metà di quelli selezionati inizialmente.

Per quanto riguarda l'interpretazione grafica, i quattro *subcorpora* sono rappresentati nel grafico dalle lettere rosse: **P**=1: organi di Partito, **S**=2: Sezioni locali, **L**=3: mondo del Lavoro, **C**=4: mondo Civile. Ogni pallino blu indica invece un segmento che è scritto, in linea col pallino stesso, a destra o a sinistra, a seconda di dove c'era lo spazio più adatto per renderlo leggibile.

Decisamente lontano rispetto agli altri tre *subcorpora*, soprattutto in relazione al primo fattore, misurato sull'asse delle ascisse, è il gruppo di volantini firmati dalle associazioni del mondo del lavoro (**L**). Le CTU più vicine ad esso sono *imposte_indirette*, *pubblici_poteri*, *lavoratori_italiani*, *lavoro_notturmo*, *fabbriche_padovane*, *unità_sindacale*, *comitato/i_unitario/i*, *strati_sociali*, *forme_associative*, *dirigenti_sindacali*. Nessuna sorpresa quindi, dato che “vicino” al mondo del lavoro si parla di fabbriche

e lavoratori, mondo sindacale, associazioni di fabbrica, ma anche di *pubblici poteri*, generalmente criticati per l'appoggio ai padroni, a volte palese altre volte solo nella forme del silenzio/consenso; come sempre si trovano anche questioni concrete quali le tasse (*imposte indirette*), che “rubano” una fetta troppo grossa del salario dei lavoratori, e anche esempi della condivisione delle richieste e della lotta, che “deve essere estesa a tutti gli *strati sociali*”.

Figura 4.1 *Analisi delle corrispondenze sulle CTU (113 unità) per subcorpora.*



NOTA: i quattro *subcorpora* sono rappresentati dalle lettere rosse:

P=1: organi di Partito, **S**=2: Sezioni locali, **L**=3: mondo del Lavoro, **C**=4: mondo Civile

N.B.: il commento prosegue nella pagina seguente (pag. 64).

Il secondo fattore (asse verticale) segnala invece la lontananza dagli altri *subcorpora*, dei volantini scritti dalla Federazione Provinciale del P.C.I. di Padova e da altri organi di partito (**P**), posizionati molto più in basso.

Il *subcorpus* è posizionato praticamente in perfetta sovrapposizione con i segmenti *forze_reazionarie* e *manifestazione_unitaria*, mentre, appena più in là, si trovano *gruppi_industriali*, *grande_padronato* e *programma_di_fabbricazione*. Già da questi segmenti si nota l'importanza che il tema del lavoro aveva per il P.C.I., che non a caso, si autodefiniva "partito dei lavoratori", oltre al fondamentale ruolo della lotta, che si esprimeva concretamente nelle manifestazioni, spesso definite 'unitarie' per sottolineare sempre e comunque che solo con l'unità, i lavoratori potevano sperare di ottenere ciò per cui si impegnavano. Altri segmenti molto vicini al Partito sono: *masse_popolari*, *movimento/i_studentesco/chi*, *imperialismo_americano* e *popolo_vietnamita*, *squadre_fasciste* : segnalano altri temi che stavano molto a cuore al P.C.I., come il desiderio di pace con la conseguente denuncia dell'aggressione americana al Vietnam e la denuncia degli atti vandalici e provocatori, messi in atto con triste frequenza da giovani fascisti, in genere durante le numerose manifestazioni di protesta; sono citati inoltre gli studenti, perché anche il tema la scuola e le proteste del '68 e degli anni successivi rivestivano una grande importanza per i comunisti.

Molto vicini, infine, sono i *subcorpora* dei volantini scritti dalle sezioni locali del P.C.I. (**S**) e da associazioni del mondo civile (**C**) che condividevano evidentemente molte tematiche, come ad esempio l'*edilizia_popolare* - posizionata giusto in mezzo - e gli *asili_nido*.

Quasi sulle stesse coordinate del *subcorpus* delle sezioni locali, si posiziona il segmento *consigliere/i_comunale/i*, mentre sono molto vicini anche *cellula_universitaria*, *politica_antipopolare* (quella della D.C., ndr), *consiglieri_comunisti* e *comitato_cittadino*. Quest'ultimo segmento, insieme a *cellula_universitaria*, è un ulteriore esempio della condivisione delle tematiche tra sezioni e associazioni del mondo civile, dato che si tratta proprio di due di queste associazioni, una di cittadini e una di studenti.

Oltre agli *asili_nido* e all'*edilizia popolare* infine, tra i segmenti più vicini al *subcorpus* del mondo civile, si notano: *generi_alimentari*, *7_giugno*, *mercato_comune_europeo*, *borghesia_capitalistica*, *Partito_socialista*, *scelte_politiche*, *autorità_scolastiche*, *trasporti_publici*.

CAPITOLO QUINTO

Conclusioni

Partecipazione e impegno, convinzione e unità, organizzazione e presenza, collettività e condivisione.

Queste le parole che possono riassumere i risultati di questa lunga ed articolata analisi. Dai volantini della sinistra padovana degli anni '60 emerge un quadro di profondo fermento sociale, una situazione nella quale sembra che davvero tutti siano impazienti di lasciare un segno tangibile, di dire la loro per migliorare il Paese e risolvere i numerosi problemi sociali che lo affliggono.

Una partecipazione realmente capillare, dove tutte i personaggi sulla scena si impegnano per difendere e rivendicare i loro interessi, senza dimenticare però gli altri - dimensione diventata apparentemente invisibile nella società odierna - perché è forte la consapevolezza che solo con la condivisione dell'impegno e della lotta, si possono ottenere i miglioramenti richiesti e necessari.

Il Partito Comunista Italiano cerca di imporre i propri ideali di pace, di giustizia civile, di benessere per il popolo tutto e non per pochi; cerca il consenso di tutti, dei lavoratori in particolare, non per fare i propri interessi, ma per difendere i lavoratori stessi, nel tentativo di creare un circolo virtuoso nel quale l'impegno di ognuno può andare a beneficio di tutti.

A testimoniare tutto questo ci sono le parole. Parole che rappresentano sempre qualcosa, parole che non sono flebile voce, ma forte e ad ampio contenuto, parole che creano una catena, un anello che unisce organi politici nazionali a piccoli rappresentanti politici locali, che a loro volta si legano stretti ai cittadini che desiderano rappresentare, cittadini capaci anch'essi di organizzarsi e ritrovarsi, per sostenere sé stessi e tutti i lavoratori, gli stessi lavoratori che nelle fabbriche e sui luoghi di lavoro più in generale, lottano fianco a fianco con i rappresentanti politici, spesso lavoratori anch'essi, e il cerchio, l'anello saldo ed ampio, si chiude.

A creare questa catena sono appunto le parole, lo si è visto nel corso dell'analisi: mai le parole esclusive o specifiche dei *subcorpora* riguardavano solo i problemi tipici degli scriventi; il Partito non parla solo di politica, ma anche di lavoro perché i lavoratori

stessi sono il Partito; parla, anzi denuncia, anche la violenza contro manifestanti che sono prima di tutto cittadini, appartenenti a quel mondo civile che si organizza a livello locale per rivendicare un futuro per i giovani, che si battono per una scuola ed un'università realmente per tutti, e nelle loro lotte affiancano gli operai, tutti insieme in piazza, perché nessuno otterrà i miglioramenti da solo, mentre con l'unità le strade possibili si aprono e il limite si allontana.

A livello di contenuti, il risultato più evidente di questo lavoro è proprio questo: tutti si impegnano per tutti. Emerge una fortissima interdipendenza tra i soggetti protagonisti di quest'analisi, un'organizzazione e un movimento sociale condiviso da politica, mondo del lavoro, cittadini, studenti, associazioni locali, tutti insieme a lottare per realizzare un unico e reale cambiamento.

Ciò che cambia, che presenta delle differenze, è invece il linguaggio utilizzato per comunicare all'esterno la volontà di cambiamento: si nota una divisione, tutt'altro che netta, solo accennata ma in alcuni casi abbastanza evidente, tra gli organi di partito e il mondo civile da una parte e il mondo del lavoro e le sezioni dall'altra.

L'analisi delle parole utilizzate infatti evidenzia spesso una tendenza, soprattutto da parte del P.C.I. e a volte anche dei componenti del mondo civile, a rivolgersi alle masse con lessico e toni tipicamente propagandistici; un modo, forse necessario, per trovare il consenso, per animare "il popolo" e creare quella base sociale che permette poi a lavoratori e politica locale di parlare con più chiarezza, per risolvere i problemi concreti.

Nelle fabbriche ci sono gli inviti alla lotta, le minacce ai padroni e ai grandi monopoli, è vero, ma non solo, perché le difficoltà sono più concrete che mai, palpabili ogni giorno sul posto di lavoro e di conseguenza altrettanto concrete sono le proposte e le richieste per migliorare la situazione e rivendicare dei diritti che dovrebbero essere scontati.

Lo stesso succede a livello di politica locale: lo scontro tra P.C.I. e D.C. in tutte le amministrazioni comunali è ideologico, almeno in partenza, ma si fonda sul modo diverso di gestire il potere, di governare per il bene di molti o per il bene di pochi, il modo diverso di affrontare i problemi, perché diverse sono spesso le priorità.

Alla fine di questo lungo percorso, ciò che rimane, dopo essere cresciuta parola dopo parola, è la nostalgia. Una nostalgia che può provare anche chi non ha vissuto quegli anni, non ha ascoltato e letto quelle parole, né tantomeno ha potuto partecipare alle grandi richieste di cambiamento. Forse, a dire il vero, non si può parlare di nostalgia per qualcosa che non si è vissuto, ma, comunque lo si voglia chiamare, questo sentimento resta, per non aver avuto l'opportunità di conoscere le parole, al tempo in cui erano rosse, potendole collocare opportunamente nel vero contesto di tensioni sociali e di richieste espresse dagli schieramenti contrapposti, dando loro quindi un significato più misurato.

Bibliografia

BERNARDI L. (2008), “Il delta dei significati. Uno studio interdisciplinare sull’espressione autistica”.

BERNARDI L., TUZZI A. (2006), “Lo scontro a parole. Analisi lessico-testuale dei discorsi di quattro candidati all’elezione del Rettore di un prestigioso Ateneo italiano”.

BOLASCO S. (1999), “Analisi multidimensionale dei dati”.

GIULIANO L., LA ROCCA G. (2008), “L’analisi automatica e semi-automatica dei dati testuali”.

GREENACRE M. J. (1984), “*Theory and Application of Correspondence Analysis*”.

LAFON P. (1980), “*Sur la variabilité de la fréquence des formes dans un corpus*” (traduzione italiana).

TUZZI A. et al. (2004), “Molto sono emozionato. Il linguaggio autistico nei testi prodotti con la C.F.”.

TUZZI A. (2003), “Le parole di Antonio (statisticamente parlando)”.

TUZZI A. (2003), “L’analisi del contenuto. Introduzione ai metodi e alle tecniche di ricerca”.

Siti internet consultati

<http://www.centrostudiluccini.it/associazione/presentazione.htm> “Chi siamo”,
consultato il 13.10.2010

<http://www.taltac.it/it/index.shtml> “TaLTaC²”,
consultato il 18.10.2010

<http://www.taltac.it/help/index.html> “Guida in linea di TaLTaC²”,
consultato il 15.11.2010

<http://www.jstatsoft.org/v20/a03/paper> “*Journal of Statistical Software*”,
consultato il 19.11.2010

http://it.wikipedia.org/wiki/Pagina_principale “Wikipedia. L’enciclopedia libera”,
consultato il 21.11.2010

<http://www.econ.upf.edu/~michael/> “*The life and times of Michael Greenacre*”,
consultato il 21.11.2010

<http://cran.r-project.org> “*R-project*”,
consultato il 22.11.2010

RINGRAZIAMENTI

Ringrazio mamma e papà, perché se sono arrivato a scrivere queste righe è soprattutto grazie alla loro fiducia.

Ringrazio Laura - sorella Luna(tica) - per i sorrisi, gli sberleffi, le risate, fondamentali in questo periodo di lunga concentrazione.

Ringrazio (la) Yle, per il continuo ed immutabile sostegno, nei bei momenti di esaltazione come in quelli bui, di delusione e sfiducia.

Ringrazio il Centro Studi Ettore Luccini di Padova, nelle persone di Dario, Giorgio, Mirko e Sandro - in rigoroso ordine alfabetico - perché questa tesi è anche frutto del loro impegno e della loro passione.

Ringrazio il Prof. Bernardi, per la fiducia e la disponibilità, per il tempo concessomi e per la pazienza e la cordialità sempre dimostrate nelle mie numerose visite.

Ringrazio la Prof.ssa Tuzzi, perché può non essere necessario, ma non è di certo di cattivo gusto (... checché ne dica Umberto Eco ...), ringraziare chi dimostra illimitata disponibilità e fornisce aiuti e consigli ben oltre il necessario.

Ringrazio infine le parole, con la loro forza e le loro sfumature, perché senza di loro non solo non ci sarebbe questa tesi, ma saremmo anche privati di buona parte della Bellezza.