

**UNIVERSITÀ DEGLI STUDI DI ROMA  
“La Sapienza”  
FACOLTÀ DI SCIENZE STATISTICHE  
DIPLOMA TRIENNALE**

**Il lessico dei Newsgroups di argomento  
religioso: lo studio di quattro casi esemplari con  
applicazione dello Spad-T®**

Relatore: Prof. Luca C. Giuliano

Diplomando: Alessandro Stabellini  
Matricola N° 04030550

**Anno Accademico 1999/2000  
Seduta di Diploma del 30 maggio 2000**

# SOMMARIO

Introduzione .....	4
I. L'analisi statistica multidimensionale del contenuto.....	6
1.1. La Statistica testuale.....	6
1.2. L'obiettivo dell'analisi dei dati testuali.....	8
1.3. La ricerca qualitativa.....	10
1.4. Definizioni e concetti di linguistica .....	13
1.5. Dalla Linguistica alla Statistica.....	20
1.6. La prima fase dell'indagine.....	23
1.7. Il trattamento del testo.....	33
1.8. Tipi di matrici di dati testuali .....	43
1.9. L'analisi multidimensionale del contenuto .....	46
II. <i>USENET</i> .....	52
2.1. Cos'è <i>USENET</i> .....	52
2.2. Le origini di <i>USENET</i> .....	54
2.3. I messaggi in <i>USENET</i> .....	56
2.4. Il formato dei messaggi <i>USENET</i> .....	62
2.5. I <i>Newsreader</i> .....	73
III. Il caso di studio: il lessico dei Newsgroups di argomento religioso.....	75
3.1. Introduzione all'analisi.....	75
3.2. L'automatizzazione .....	78
3.3. Il software sviluppato per l'automatizzazione .....	84
3.4. I parametri del software applicativo e la normalizzazione .....	87
3.5. Applicazione dello Spad-T® .....	93
Procedura NUMER .....	93
Procedura CORTE .....	94
Procedura SETEX1 .....	94
Procedura MOCAR (parole) .....	95
Output MOCAR (parole).....	96

Procedura SEGME .....	104
Output SEGME .....	105
Procedura MOCAR (segmenti).....	109
Output MOCAR (segmenti).....	110
Procedura APLUM (parole) .....	115
Procedura APLUM (segmenti).....	115
Output APLUM (parole) .....	116
Output APLUM (segmenti).....	126
3.6. Trattamento dell'output dell'APLUM (analisi delle corrispondenze semplici) .....	129
3.7. <b>Commenti e conclusioni</b> .....	153
Appendice A: esempi di messaggi <i>USENET</i> .....	157
Appendice B: tabelle codice .....	163
Appendice C: preparazione del testo e strutturazione del corpus .....	169
Appendice D: alcuni criteri per la corr. degli errori e per la normalizz. del testo .....	173
Appendice E: le espressioni regolari del Perl .....	178
Appendice F: i file di parametri delle procedure dello Spad-T® .....	180
Bibliografia .....	194
Glossario .....	198
Indice Analitico.....	205

## INTRODUZIONE

Il lavoro svolto tratta l'Analisi del Contenuto di alcuni messaggi inviati - in un determinato periodo temporale - in quattro *Newsgroups* di argomento religioso, proponendo nel contempo una metodologia operativa sostanzialmente diversa rispetto a quella adottata da altri studi e ricerche simili a questa. Tale diversità è stata suggerita dalla particolare caratteristica dei dati testuali a disposizione: l'estensione.

L'unione dei 4 files di testo corrispondenti all'insieme dei messaggi inviati nei rispettivi *Newsgroups*, ha originato, infatti, un file di **5,20MB** contenente oltre 750.000 parole. Per un confronto basti pensare che altri lavori - in precedenza - hanno utilizzato dati la cui estensione si aggirava intorno al mezzo Mega byte corrispondente all'incirca a 90.000 parole).

La metodologia adottata in questo studio propone, quindi, un percorso di ricerca che consente di trattare files di testo di notevoli dimensioni in tempi brevi.

L'obiettivo è quello di evidenziare, con un ridotto dispendio di risorse, i contenuti dei discorsi e le tematiche toccate da un gruppo di parlanti senza contravvenire alle caratteristiche dei metodi di ricerca quali la *quantificazione*, la *sistematicità*, *l'esaurività e l'oggettività*.

La perdita d'informazione, pagata a fronte di una maggiore velocità operativa del metodo adottato, è da ritenersi comunque accettabile per lo scopo proposto per i risultati ottenuti.

Nel volume sono riportati due capitoli introduttivi. Uno riguarda la teoria e le tecniche della Statistica testuale in ambito di Analisi del Contenuto, l'altro riporta alcune informazioni sul formato e sullo standard dei messaggi Usenet nonché notizie sulle sue origini e sulla sua struttura funzionale.

Alla fine trovano posto una serie di Appendici contenenti diverse informazioni tecniche sugli argomenti trattati durante il lavoro di ricerca.

## I. L'analisi statistica multidimensionale del contenuto

### 1.1. La Statistica testuale

Vi è un nuovo settore disciplinare della statistica che basa il proprio lavoro di ricerca sull'analisi dei dati testuali: la *Statistica testuale*

Il suo ambito è quello dello studio statistico diretto di fonti di natura linguistica. Il suo impianto metodologico costituisce uno sviluppo naturale degli strumenti quantitativi di analisi qualitativa. I suoi campi d'applicazione sono ampi e fortemente interdisciplinari. Essi vanno dall'analisi dei testi letterari allo studio di raccolte di testi corti (titoli, manifesti), dalle indagini qualitative in ambito psico-sociologico, all'analisi di tipo biografico, dalle ricerche terminologiche alla costruzione di lessici di frequenza [Vedi più avanti in questo capitolo].

Oggi la Statistica testuale è il risultato di una evoluzione che ha visto, durante un *excursus* storico, la logica degli studi lentamente spostarsi da un piano **linguistico quantitativo** (da cui la Statistica testuale eredita alcune regole naturali proprie di qualsiasi idioma quali la legge di Zipf, le fasce di frequenza e i principi di economia della lingua dei quali si parlerà più avanti) ad uno **lessicale**<sup>1</sup>, fino ad arrivare ad un approccio **lessico-testuale** in cui lo studio degli aspetti testuali di un *corpus* è integrato

---

<sup>1</sup> Nell'approccio lessicale lo studio statistico della lingua si concentra sulle caratteristiche di *stock* testuali significativi intesi come campioni particolari della lingua. Gli studi stilometrici sull'intera opera di un autore di E. Brunet ne sono un esempio.

da informazioni linguistiche e da interventi sul testo stesso (*normalizzazione, lemmatizzazione e lessicalizzazione*).

A tutto ciò si deve aggiungere che simili studi oggi sono praticabili grazie alla capacità dei computer, non solo di codificare e riconoscere i caratteri dell'alfabeto di un qualsiasi linguaggio naturale, ma anche di trattare ampie matrici di dati.

## 1.2. L'obiettivo dell'analisi dei dati testuali

La lingua difficilmente potrebbe definirsi un universo statistico: essa sfugge a qualsiasi definizione operativa accettabile.

La parola, infatti, considerata come unità elementare del linguaggio è di per sé non univoca: non è facile cogliere il valore di un insieme di parole, al di là del significato che ne dà il singolo emittitore, come non sempre si riesce ad attribuire un linguaggio ad una situazione/condizione particolare, potendolo poi generalizzare ad altri casi ad esso comparabili, fino ad assumerlo come modello.

<< La lingua è una massa informe di elementi (parole) in continua evoluzione >>  
[Bolasco, 1999: 179].

Tournier<sup>2</sup> con uno schema gerarchico inclusivo d'ordine decrescente cerca di cogliere i rapporti fra lingua e sue concrete possibilità. La lingua, ovvero il “potenziale”, raccoglie l'insieme degli idiomi, ovvero degli usi (l'accettabile). Questi includono tutti i possibili discorsi, ovvero “l'attualizzato”. L'insieme dei discorsi genera l'insieme dei testi (l'osservabile). Da questi ultimi si può trarre un *corpus* (l'osservato) che diviene l'oggetto di studio.

È possibile dunque concepire uno studio dei contenuti di un *corpus* su base statistica basandosi sul confronto di differenti profili lessicali [Vedi pag.13], partendo da una scansione automatica del testo con l'aiuto del computer. Esso – braccio dell'uomo – capace di raccogliere illimitatamente un gran numero di informazioni e soprattutto di

elaborarle velocemente, abbatte i limiti stessi di categorizzazione di un oggetto testuale per far sì che il ricercatore raggiunga lo scopo primario nell'ambito della ricerca: trovare le dimensioni fondamentali di *senso* presenti in un testo.

---

<sup>2</sup> Cfr. Bolasco (1999: 179)

### 1.3. La ricerca qualitativa

La logica che è alla base dell'analisi testuale è appunto quella classica dell'analisi multidimensionale. L'ambito è quello dei metodi fattoriali fondati su di una formalizzazione non probabilistica (utili alla sintesi delle variabili osservate).

Due sono i fattori che fanno preferire l'analisi multidimensionale alle altre tecniche statistiche di analisi:

- il fenomeno oggetto di studio è di tipo *osservazionale*<sup>3</sup>, ovvero il collettivo si compone di singole osservazioni di più variabili su  $n$  unità statistiche;
- l'assenza di riferimenti statistico-probablistici sul tipo di distribuzione delle variabili considerate<sup>4</sup>.

Nell'analisi multidimensionale dei dati il fenomeno viene colto nella sua globalità e complessità: vengono infatti considerate numerose variabili statistiche (leggi dimensioni del fenomeno) eterogenee fra loro, senza vincolarne la scelta alle loro caratteristiche distributive, in modo tale che vi sia un ampio apporto d'informazione. Questa, poi, verrà ridotta in modo da ottenere un guadagno in significato. Il risultato è una lettura più pregnante dei contenuti informativi espressi nei dati.

<< Limitandoci alle tecniche di tipo fattoriale, l'obiettivo principale consiste nel produrre delle dimensioni (fattori) attraverso le quali semplificare, sintetizzare e

---

<sup>3</sup> Esso si contrappone al fenomeno di tipo *sperimentale* dove l'osservazione è di più misure ripetute della stessa grandezza su  $n$  unità statistiche.

<sup>4</sup> La dimensione illimitata e teorica della lingua fa sì che non sia possibile, per definizione, associare alle parole una qualche “frequenza” in senso statistico-probablistico (quest'ultima è invece misurabile su di una raccolta di testi intesi come “campioni” particolari di una lingua [Vedi nota 1]).

rappresentare il fenomeno. Quanto più quest'ultimo deve essere ridefinito o espresso attraverso nuove misure globali (non più elementari) e inosservabili (ossia non direttamente rilevabili), tanto più i risultati saranno soddisfacenti e utili sia come elaborazioni finali che come base per ulteriori trattamenti ... >> [Bolasco, 1999: 22].

La messa in opera di queste tecniche - in ambito statistico-testuale - consente di studiare le relazioni presenti in un *corpus* complesso di dati adottando una vera e propria *strategia di analisi* di tipo *strutturale*. Il risultato non è una misurazione assoluta di tipo quantitativo (stime di singoli indici statistici, tanto per fare un esempio), bensì una misurazione relativa di tipo sistematico-contestuale. In altre parole, tali tecniche non tendono a stabilire quanto valga un insieme di grandezze, bensì quali siano i legami all'interno di un sistema di elementi. E << se gli strumenti intermedi di misura sono strettamente quantitativi, quelli finali di valutazione del "modello" individuato, falsificato o meno che sia, restano pur sempre qualitativi >> [Bolasco, dispensa a.a. 97-98: 6].

Infine un'ultima riflessione. Non vi è alcun dubbio sulla natura "qualitativa" del linguaggio naturale, tuttavia un testo numerizzato (in cui ad ogni parola diversa è attribuito un numero diverso), processato attraverso gli algoritmi di elaborazione informatica con una logica numerica, è un testo che rimane aperto a tutte le sue infinite possibili categorizzazioni. Ciò significa che la natura qualitativa del linguaggio naturale non viene contaminata: la pratica di oggetti quantitativi come la frequenza, il rango, l'ampiezza di un contesto locale, la correlazione e quant'altro possa definirsi sulle parole, grazie all'informatica, si ferma ad un livello di *identificazione* (trattamento

assolutamente qualitativo) che non sovrappone a *priori* alcun livello di misurazione d'ordine superiore all'informazione di base.

Insomma in campo statistico-testuale << L'informatica, quanto la statistica, fonda i suoi calcoli su confronti e conteggi. Gli oggetti testuali che ogni ricercatore vorrà definire per il suo studio saranno inevitabilmente sottoposti a questa logica: l'uso che egli ne farà - pur ricorrendo all'interpretazione di frequenze o all'ausilio di metodiche sofisticate dai processi markoviani alle analisi fattoriali – come genere di misurazioni a *posteriori* resterà peraltro ad un livello squisitamente qualitativo >> [Bolasco, dispensa a.a. 97-98: 6].

## 1.4. Definizioni e concetti di linguistica

Per **corpus** si intende un qualsiasi insieme di testi oggetto di studio. Tali testi possono essere letti secondo diversi punti di interesse come se si sfogliasse un libro in differenti modi (per capitoli, per paragrafi, ecc.) in funzione degli obiettivi prefissati.

Ogni differente lettura secondo diversi criteri genera - dal punto di vista della statistica testuale - un insieme di **profili lessicali** che costituiscono la base per l'analisi.

Questa definizione operativa di *corpus* è applicabile alle più diverse fonti testuali tra cui la trascrizione di discorsi orali e la traduzione di codici specifici.

Per **testo** si intende una fra le tante possibili partizioni del corpus. Il *testo* può essere considerato anche come un insieme di **frammenti** (frasi) di un **discorso** del **parlante** (colui che ha generato il testo) le cui **parole** (componenti elementari) sono denominate **occorrenze** (ovvero ogni sequenza di *parole* che appare in un *testo*). Le *parole* possono anche essere considerate come una sequenza di caratteri di un alfabeto predefinito delimitata da due separatori<sup>5</sup>.

Ogni *frammento*<sup>6</sup> è definito da almeno una *proposizione* o *frase*, delimitata da due separatori "forti"<sup>7</sup>, all'interno dei quali è possibile identificare uno o più **segmenti**

---

<sup>5</sup> Sono considerati, ad esempio, separatori **caratteri non appartenenti all'alfabeto** come lo spazio bianco (*blank*), la punteggiatura (, . : ; ? !), le virgolette, i "trattini" (- / |), le parentesi ( {} () [] ) e ogni carattere speciale (#, @, \$, £, \$ ecc.) presente nel testo. Ma potrebbero considerarsi tali, *ad hoc*, i numeri o quant'altro.

<sup>6</sup> I frammenti possono essere naturalmente già definiti, come nel caso delle risposte libere in un questionario o dei titoli di articoli della stampa, o dei paragrafi e/o commi di un testo giuridico. Ma possono invece non esserlo, come nel caso dei testi letterari e di interviste non direttive. Allora si ricorre all'individuazione dei *separatori* [Vedi nota 5]

<sup>7</sup> In generale, tali separatori sono definiti da segni di punteggiatura, ma, per una buona segmentazione del testo devono essere adottati ulteriori criteri. Nella lettura automatica della seguente sequenza di testo "...

**ripetuti.** Un *segmento ripetuto* è una sequenza di *parole* (*sintagma*), tra tutte le disposizioni di  $2,3,\dots,q$  *parole* che si ripetono più volte nel *corpus*, contenuta interamente in un *frammento*<sup>8</sup>. Tra i *segmenti* si considerano solo i *sintagmi* che costituiscono una *lessia* (*unità di senso* – Vedi pag.21), cioè tutte quelle sequenze di parole che esprimono un contenuto autonomo (**poliformi**). Queste sono di solito il 30% dei *segmenti*. Tra i *poliformi* vi sono le **polirematiche**<sup>9</sup>, ovvero dei *sintagmi* che acquistano un significato diverso rispetto ai significati elementari delle parole semplici che li compongono (es. <disegno\_di\_legge> oppure i costrutti verbali <venir\_meno, andare\_al\_creatore>). È facile che una *lessia* non semplice, in quanto unità semantica, sia spesso una *polirematica*.

Si riporta di seguito un schema di esempio di *segmenti ripetuti*, *poliformi* e *polirematiche* [Bolasco, 1999: 195].

---

in modo, tale che ... " saranno riconosciuti solo <in modo> e <tale che>. Occorre però evitare di scambiare ad esempio dei punti (punti di migliaia [1.15011] punti di abbreviazioni [Sig. Rossi, C.E.E.]) come falsi indicatori di fine frase o fine frammento. In alcuni *software* le frasi vengono identificate mediante algoritmi di riconoscimento di tali situazioni (grammatiche locali) capaci di risolvere tali ambiguità.

<sup>8</sup> Ad esempio una frase del tipo "il governo si propone di" contribuisce alle occorrenze di segmenti quali: <il governo>, <il governo si>, <il governo si propone>, <governo si>, <governo si propone>, <si propone>, <si propone di>, oltre che al segmento <il governo si propone di>. Bisogna notare che i *segmenti ripetuti* sono di per sé assai più numerosi delle stesse *forme grafiche* componenti un *corpus*. Per esempio, ad una soglia di 3 occorrenze, in un *corpus* molto ampio, si ottengono oltre 18.000 segmenti diversi, quando le *forme grafiche* distinte, a quella soglia, sono 9.400: un rapporto di 2 a 1. Al contrario in un *corpus* molto piccolo (minore di 5.000 occorrenze), a soglia di frequenza elevata (ad esempio 10) il numero di segmenti ripetuti può essere inferiore a quello delle parole alla stessa soglia. Nel parlato, o là dove il linguaggio è molto stereotipato, questo rapporto segmenti/parole può essere ancora più elevato: fino ad un valore pari a 4.

<sup>9</sup> La presenza di polirematiche rende molto difficile il giusto riconoscimento di una frase da parte di tutti quegli algoritmi di traduzione automatica dei testi. Un esempio è il passo: <Portare avanti una casa è un'impresa> che equivale ad un modello generale di senso del tipo: <gestire una casa non è facile>. Si ritiene che la presenza di polirematiche in un testo possa arrivare a coprire il 60% dell'intero testo ove si considerano anche tutti i verbi-supporto compresi gli ausiliari essere, avere e fare.

## Schema 1: segmenti ripetuti, poliformi e polirematiche

---

### Segmenti

vuoti: e di, con il, per la, non si, tra le, da tre, ma un, non c'  
incompleti: campo del, è il, quanto si, casa per  
pieni: buona volontà, programma di governo, politica industriale

### Poliformi

locuzioni grammaticali con funzioni di:  
avverbii: di più, non solo, per esempio, di nuovo, in realtà, più o meno, di fatto, del resto  
(luogo) a casa, in chiesa, al di là  
(tempo) di sera, un anno fa, al più presto  
(modo) in particolare, d'accordo, in piedi  
preposizioni: fino a, da parte di, prima di, rispetto a, in modo da, per quanto riguarda  
aggettivi: in punto, di oggi, dei genere, in crisi, di cotone, in fiamme, alla mano  
congiunzioni: il fatto che, dal momento che, prima che, nel senso che, a patto che  
interiezioni: va bene!, grazie a Dio, mamma mia!, hai voglia!, punto e basta  
idiomi e modi di dire: io penso che, è vero che, non è che, per così dire, questo è tutto non c'è niente da fare, è un peccato  
gruppi nominali polirematici: buona fede, lavoro nero, mercato unico, punto di vista, cassa integrazione  
verbi supporto e idiomati: si tratta di, tener conto, portare avanti, far fronte, far parte, prendere atto, dare vita, dare luogo, mettere a punto, venirne fuori, rendersi conto

---

Un *discorso* è caratterizzato da un **linguaggio** le cui componenti sono:

- l'**idioma** della comunità di appartenenza del *parlante* in un dato periodo storico (la lingua del parlante: italiano, inglese ecc.) che è la componente dovuta al *lessico* della lingua;
- il **contesto**, ossia l'ambito o il genere concettuale del discorso, l'aspetto tematico e/o settoriale della terminologia. Il linguaggio è diverso a seconda che tratti di politica, di letteratura, d'informazione di sport;
- la specifica **condizione** di enunciazione del discorso (o di stesura del testo) che riflette la modalità d'interazione fra colui che emette (*E*) e colui che riceve (*R*) il messaggio, per cui si creerà una situazione diversa a seconda che il rapporto si stabilisca fra due soggetti ("uno a uno" o dialogo, lettura) o fra un soggetto e un

gruppo ("uno a molti", manifesto, assemblea), oppure a seconda che il carattere del discorso sia formale o informale e si svolga, ad esempio, in pubblico o in privato oppure vi sia co-presenza o meno fra *E* ed *R* (dialogo a vista o a distanza, via telefono o via mass-media) o che abbia carattere pedagogico/polemico.

Con il termine **contesto locale** si indica un determinato insieme di parole adiacenti ad un termine prefissato (di solito si considerano 5-10 parole prima e 5-10 parole dopo il termine) che funge da polo (*pivot*). Lo studio dei *contesti locali* di una parola viene detto **analisi delle concordanze**.

L'**unità di contesto** è un *frammento* di testo che ha generalmente una lunghezza variabile (da 120 o 200 parole). Spesso corrisponde ad un **enunciato** (proposizione con un senso compiuto) altre volte ad una **frase** (proposizione con una rilevanza sintattica).

La **dimensione** (**N**) o lunghezza di un *corpus* è data dal numero totale delle *occorrenze (parole)*, mentre il **vocabolario** del *corpus* viene definito dall'insieme delle *parole diverse* distinte fra loro<sup>10</sup>. Il numero di *parole diverse* in un testo definisce l'**ampiezza del vocabolario** (**V**).

Vale la seguente relazione:

$$\mathbf{V}_1 + \mathbf{V}_2 + \mathbf{V}_3 + \dots + \mathbf{V}_i + \dots + \mathbf{V}_{f_{\max}} = \mathbf{V}$$

con  $\mathbf{V}_i$  numero di parole diverse che appaiono (o ricorrono) *i* volte ( $\mathbf{V}_1$  rappresenta quindi l'insieme delle parole che appaiono una sola volta  $\mathbf{V}_2$  quelle che ricorrono due volte ecc.. L'insieme può essere costituito anche da una sola *parola*) e  $f_{\max}$  valore delle occorrenze della parola con il maggior numero di occorrenze del vocabolario.

Il **vocabolario** di un *corpus* può essere espresso in **forme grafiche** (ossia parole tali e quali scritte nel testo) o in **lemmi** (ossia riconducendo le parole del testo al

corrispondente vocabolo presente in un *dizionario* della lingua. Ove per *dizionario* si intende l'insieme dei lemmi di un idioma raccolti in un unico "inventario" o *database* lessicale che comprende non solo i *lemmi*, ma anche le *forme flesse* [le voci declinate dei sostantivi o aggettivi, o quelle coniugate dei verbi], le *forme composte*, le *locuzioni* e le forme *idiomatiche*. Per la definizione di *lemma* si veda più avanti).

È logico che il *vocabolario* di un *corpus* espresso in *forme grafiche* avrà un'ampiezza differente del vocabolario espresso in *lemmi*.

Si definisce *lessico* il discorso in *potenza*, ovvero un insieme virtuale di segni linguistici<sup>11</sup> - esistente nella memoria collettiva di una comunità o in quella di un individuo - da cui possono essere tratte tutte le parole di un *potenziale* discorso. È evidente che non tutte le forme possibili - sebbene siano conosciute - saranno di fatto attualizzate (*occorrenze*). Inoltre il *lessico* di un individuo è un riflesso delle sue appartenenze socio-antropologiche, ovvero delle sue origini, dell'esperienza e della cultura acquisita.

Si definisce *rango* il posto occupato da un termine in una graduatoria. Nei *vocabolari* per *occorrenze decrescenti*, ad esempio, un rango "elevato" è indicato da un numero piccolo.

Per *lemma* s'intende la *forma canonica* con cui una parola è presente in un dizionario della lingua (come entrata di una voce)<sup>12</sup>. Ad esempio le occorrenze <scrive> e

---

<sup>10</sup> <casa> diversa da <case> o da <cane>

<sup>11</sup> ovvero uno *stock* mentale di radici lessicali (*lessemi*) da cui ricavare tutte le forme flesse dei corrispondenti *lemmi*. Se un individuo conosce il significato della radice "lavor-" può generare sia il sostantivo <lavoro>, sia il verbo <lavorare>, ma anche forme come <lavorante> o altre flessioni.

<sup>12</sup> L'infinito per i verbi, il singolare per i sostantivi, il singolare-maschile per gli aggettivi.

<scrivevano> individuano due forme grafiche distinte, ovvero due *flessioni*, appartenenti ad uno stesso lemma: il verbo <scrivere>.

Secondo la definizione della grammatica tradizionale, ogni *parola* (in quanto *monema*) può suddividersi in un radicale (*lessema*) e in una desinenza o un affisso (*morfema*): il primo individua l'aspetto lessicale e semantico del termine, il secondo quello grammaticale. La parola "scrivere", ad esempio, si articola in scriv-ere (Queste definizioni variano a seconda dei linguisti che ne discutono).

La *parola* di un *vocabolario* può essere distinta e categorizzata attraverso differenti criteri: rispetto al suo ruolo nella frase, alla sua categoria grammaticale o ad altri criteri.

Molto spesso si fa distinzione tra *parole vuote* e *parole piene*. Le prime sono le *parole* grammaticali o di legame (articoli, preposizioni, congiunzioni ed alcuni aggettivi) che non esprimono in sé un contenuto d'interesse ai fini dell'analisi, ma hanno una funzione *strumentale* in quanto cardini di costrutti lessico-grammaticali<sup>13</sup>. Le seconde sono portatrici di tutti quei *significati* oggetto di studio, delle parti "sostantive" del contenuto di un *discorso* (nomi e aggettivi), delle sue modalità di enunciazione (avverbi) o di azione (verbi) e per questo sono dette anche *parole principali*.

Inoltre due *parole* possono essere *omografe* e per questo avere lo stesso *significante* ma diverso *significato (polisemia)* (ad esempio <stato\_S1> nell'accezione di "istituzione" distinto da stato <stato\_S2> nell'accezione di "situazione/condizione"),

---

<sup>13</sup> Anche le parole grammaticali possono avere importanza nell'interpretare un testo. Ad esempio, il sovrautilizzo di preposizioni come <in> o <di> sottolinea il carattere descrittivo del discorso; una prevalenza di <non>, <per> e <con> sottolinea particolari intenzionalità del parlante, mentre quella dei <ma> e <se> evidenzia elementi legati ad incertezza.

oppure avere differente *significante* ma medesimo *significato (sinonimia)* (ad esempio <stupendo> e <splendido>, <abitanti> e <residenti>, <isolata> e <sperduta>).

## 1.5. Dalla Linguistica alla Statistica

Base di partenza di ogni indagine statistica è la selezione del *collettivo*. Il *collettivo* statistico è in questo caso rappresentato dal *corpus* ovvero da una raccolta di testi (*stock* di materiale testuale), omogenea sotto qualche punto di vista.

L'**unità statistica**, ovvero l'unità di osservazione, può essere di tre tipi a seconda degli obiettivi, del tipo e del livello di analisi del *corpus*.

- L'**unità di testo**, che corrisponde alla *parola* come *forma grafica* (grafia)<sup>14</sup>. È l'unità di analisi elementare per la lettura computerizzata di un testo: La *parola*, vista come <catena di caratteri di un alfabeto delimitata da due separatori che ne definiscono l'inizio e la fine><sup>15</sup>, diviene l'oggetto del *riconoscimento automatico* (*scansione*) di un testo<sup>16</sup>. Il risultato finale è la **numerizzazione** (ad ogni occorrenza diversa si associa un numero diverso) o l'**indicizzazione** del *corpus* (ogni parola è identificata da un codice identificativo e da un indirizzo - pagina, riga - che indica la sua collocazione nel testo). In taluni studi la *forma grafica* può essere “riletta” successivamente come una specifica flessione di un *lemma* (Vedi pag.17) in quel processo che prende appunto il nome di **lemmatizzazione**.
- L'**unità di contesto** che corrisponde ad ogni *frammento* di testo (Vedi pag.13): sia esso una *frase* (proposizione sintatticamente indipendente), un *enunciato*

---

<sup>14</sup> È evidente che una catena di caratteri non è necessariamente una parola di senso. Per cui <carta> e <catra> sono due occorrenze di *parole* diverse ma <catra> non è una parola dell’italiano: lo sarebbe dopo una correzione ortografica.

<sup>15</sup> Di fatto ogni *parola* di un testo è delimitata da spazi bianchi, mentre ogni frase dalla punteggiatura.

(proposizione di senso compiuto), o una *risposta individuale* (al limite, costituita da una sola parola: <sì>) o quant'altro sia da considerarsi unitario sotto qualche punto d'interesse. Ai fini di un'analisi automatica di un *corpus*, può considerarsi come un'unica *unità di contesto* anche un intero testo (un libro, un discorso, un articolo di giornale), oppure una sua parte (un capitolo, un paragrafo, un titolo): ovvero un qualsiasi raggruppamento pertinente di frammenti.

L'*unità di contesto* è l'unità di analisi di studi basati sul confronto di *frammenti* del *discorso* al fine di individuare delle specificità o delle omogeneità rispetto a testi diversi<sup>17</sup>.

- La *forma testuale* che corrisponde alla *lessia*, ovvero alla più piccola unità portatrice di *senso* - non ulteriormente decomponibile - rilevabile in un *corpus*. È l'unità minima significativa del discorso che può essere *semplice* (definita da una sola parola: <cane>, <tavola>), *composta* (costituita da più parole in via d'integrazione: <sangue freddo>) o *complessa* (individuata da una sequenza di parole fra loro connesse: <fare lo gnorri>, <dalla testa ai piedi>). Nell'applicazione dei metodi della *statistica testuale* e *di analisi del contenuto* si adottano unità di tipo misto, ora *semplici*, ora *complesse* che vengono appunto definite *forme testuali*. Così una *forma testuale* potrà essere sia un *lemma* (<scrivere>), sia una riduzione *lessematica* (Vedi pag.18) (ad esempio la radice <attual+> che fonderebbe <attuale>, <attualmente>, distinta da <attu+> frutto delle fusioni <attuare>, <attuazione>, <attuato>), sia un *significante* che rappresenti la fusione di *sinonimie*

---

<sup>16</sup> Il computer infatti legge ogni informazione linguistica come una sequenza di *bytes* ciascuna delimitata dai separatori inizio e fine.

<sup>17</sup> L'obiettivo è quello di individuare differenti “universi lessicali” e lo studio ha interessi terminologici.

accertate nel *corpus* (Vedi pag. 19) (<accordo\$>=accordo+alleanza+patto), ma anche un *poliforme* (locuzione grammaticale o polirematica di contenuto: <in\_cORSO>, <bilancia\_dei\_pagamenti> od una frase fissa *idiomatica*, identificabile come un'entità (<andare\_al\_creatore>).

## 1.6. La prima fase dell'indagine

È il momento della *lettura computerizzata* di un testo<sup>18</sup>. Il dato statistico rilevato è l'*occorrenza* di una *parola* intesa come *forma grafica* (*unità di testo*) e viene considerato lo sviluppo “*orizzontale*” del testo nella sua *dimensione N*. L’interesse è rivolto alla ricostruzione del *discorso* inteso come “*contenuto*” e non come “*modalità in cui si articola tale contenuto*”. Per cui, ad esempio, l'espressione “sono finiti” origina due *forme grafiche* differenti e quindi due *occorrenze* diverse .

In questa fase si parlerà di *frequenza relativa* nel testo considerando ad esempio per la forma “*bella*” il rapporto  $52/7940=0.0065$  (ovvero il rapporto tra il numero delle *occorrenze* della forma sul numero N totale delle *occorrenze* rilevate del testo). In luogo della *frequenza relativa* si preferisce considerare il numero delle *occorrenze normalizzate*, ovvero delle *frequenze relative* espresse non già per 100, ma in unità più ampie (per 10.000 o 100.000) a seconda della dimensione del *corpus*. L’uso delle *occorrenze normalizzate* è dovuta al fatto che, nel caso dei dati linguistici, la *frequenza relativa* è sempre piccolissima. Non solo: la *normalizzazione* consente di effettuare dei confronti anche tra *parole* diverse tra *testi* diversi. Avendo infatti testi di uguale dimensione (stessa base) si può dire se una stessa o diverse *occorrenze* sono più *frequenti* in un testo piuttosto che in un altro.

Bisogna però sottolineare che pur considerando la frequenza come un elemento centrale in quest’ambito, essa non può essere l’unico criterio per selezionare le parole da

analizzare. Basti pensare che in una situazione specifica (in un discorso individuale ad esempio) una stessa parola può avere una frequenza di volta in volta diversa, così da non poter stabilire un valore di probabilità del *lessema* corrispondente. In genere la frequenza di una parola è una sovrastima della sua probabilità. Solo per *corpus* molto grandi (a partire da 500.000 *occorrenze*) il numero di *occorrenze normalizzate* di parole molto frequenti può considerarsi una approssimazione abbastanza buona di una probabilità.

Vi sono però alcune regolarità tipiche di ogni idioma.

I linguisti quantitativi hanno infatti individuato alcune “regole” naturali che sono alla base di ogni distribuzione di frequenza di parole utilizzate in un qualsiasi idioma.

- **Legge di Zipf**

$$f \bullet r = c \quad (\text{dove } c = \text{costante})$$

Essa stabilisce che le parole di un vocabolario si distribuiscono in maniera tale che la frequenza (*f*) e il rango (*r*) di una parola sono inversamente proporzionali.

Un suo classico esempio, fatto dallo stesso Zipf (cfr. Bolasco, 1999: 200) è tratto da uno studio lessicale dell'*Ulysses* di Joyce (dal vocabolario di un corpus di 260.000 *occorrenze*):

al rango 10 la frequenza è 2.653

al rango 100 la frequenza è 265

al rango 1.000 la frequenza è 26

al rango 10.000 la frequenza è 2

---

<sup>18</sup> Nell'Appendice C sono riportate alcune informazioni e istruzioni per la preparazione e per l'acquisizione su supporto elettronico di un testo trattato.

Nota: dato che non tutte le frequenze possibili sono presenti in un vocabolario ed esistono, per il rango, gli *ex aequo*, la regola va intesa in senso statistico, ossia "in media", dal momento che occorre prendere, come valore di frequenza, un valore medio delle occorrenze delle parole appartenenti ad un certo intorno del *rango* considerato.

Ampi dibattiti sulla validità di questa legge hanno fatto sì che essa venisse espressa dalla seguente formula:

$$f \cdot r^a = c \text{ da cui}$$

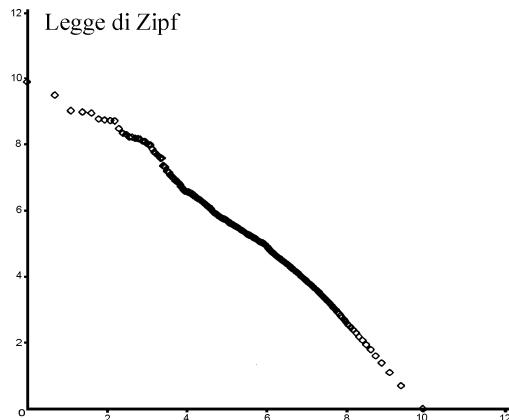
$$\log(f) = -a \cdot \log(r) + c'$$

ovvero l'equazione di una retta (Vedi Figura 1) su di un grafico in coordinate logaritmiche in cui  $y$  descrive il logaritmo della frequenza,  $x$  il logaritmo del rango mentre  $a$  ne indica la pendenza. Si può considerare con buona approssimazione:

$$a = \log N / \log V$$

La pendenza è funzione della cosiddetta *ricchezza del vocabolario*  $V$  (Vedi pag.16) che a sua volta dipende dalla dimensione del *testo*  $N$ . (Vedi più sotto).

**Figura 1**



Fonte: Bolasco, 1999: 202

- **Coefficiente  $G$  di Guiraoud di *ricchezza lessicale***

$$G = \frac{V}{\sqrt{N}}$$

$V$  = numero vocaboli diversi e  $N$  = numero di occorrenze.

In termini empirici è possibile osservare che per testi compresi fra 10.000. e 50.000 occorrenze di parole, il vocabolario  $V$  aumenta come la radice quadrata di  $N$  e  $G$  vale circa 22. Per lo stesso genere di testi il coefficiente angolare della retta di *Zipf* vale  $a = -1.15$ .

Si può assumere anche - come una prima misura di *ricchezza lessicale* - la proporzione di parole diverse sul totale delle occorrenze: ossia  $V/N$  (noto anche come *type/token ratio*) oppure anche  $V_1/V$  ( $V_1$  rappresenta l'insieme delle parole che appaiono una sola volta - chiamate anche *hapax* – mentre  $V$  rappresenta il totale di parole diverse).

- **Fasce di frequenza di un vocabolario**

Si prenda come esempio la seguente tabella:

**Tabella 1**

Esempio di vocabolario per occorrenze decrescenti (in forme grafiche con relativo rango)

Rango	Forma g.	Occ.	Rango	Forma g.	Occ.	Rango	Forma g.	Occ.
1	è	293	29	terra	43	58	dei	22
2	di	217	31	dal	41	58	però	22
3	una	214	31	disoccupazione	41	58	possibilità	22
4	e	197	31	problemi	41	58	potenzialità	22
5	un	194	34	se	38	58	sempre	22
6	la	184	35	ci	35	64	come	21
7	che	149	35	della	35	65	Italia	20
8	regione	147	37	c	33	65	occupazione	20
9	isola	114	37	sviluppo	33	67	bene	19
10	per	113	39	dei	31	...	...	...
11	non	112	39	stato	31	74	potrebbe	18
12	Sardegna	104	41	bellissima	30	76	fare	17
13	più	101	42	deve	28	79	giovani	16
14	i	89	42	isolamento	28	85	perché	15
15	con	87	42	lavoro	28	92	classe	14
16	da	83	45	dell	27	97	anche	13
17	in	82	45	delle	27	100	alla	12
18	ma	78	45	isolata	27	108	abitanti	11
19	le	71	48	gli	26	124	bellezze	10
20	a	68	48	soprattutto	26	139	economica	9
21	molto	67	50	bisogno	25	156	ai	8
22	si	61	50	resto	25	173	aiuti	7
23	l	60	50	sardi	25	199	agricoltura	6
24	il	56	53	ambiente	23	222	abbastanza	5
25	bella	52	53	cui	23	266	acque	4
26	sono	50	53	dove	23	320	oggi	3
27	essere	48	53	poco	23	425	abbiamo	2
28	ha	45	53	troppo	23	657	abbandonate	1
29	ancora	43	58	al	22			

Fonte: Bolasco, 1999: 190

Legenda:

Fascia alta  Fascia media  Fascia bassa 

Le parole di un vocabolario, nella tabella sopra, sono state ordinate per frequenze decrescenti, partendo dalla parola con  $freq_{max}$  fino ad arrivare a quelle con  $freq_{min}$  ossia con una sola occorrenza (*hapax*) [Vedi anche pag.16].

$$\mathbf{V}_{f_{max}}, \mathbf{V}_{f_{max}-1}, \dots, \mathbf{V}_3, \mathbf{V}_2, \mathbf{V}_1$$

Insieme al valore di frequenza è stato indicato anche il *rango* (Vedi pag.17). In questo caso, dato che l'ordinamento è per occorrenze decrescenti, un *rango* elevato è indicato da un numero piccolo.

Si possono individuare tre fasce: le *alte*, le *medie* e le *basse* frequenze.

- Il limite fra le *alte* e le *medie* frequenze si colloca subito sopra la prima parità: ossia la prima coppia di parole che occorrono lo stesso numero di volte (cioè, in Tabella 1 al rango della forma <ha>);
- il confine fra *medie* e *basse* frequenze è individuato dalla prima lacuna nel numero consecutivo di *occorrenze* crescenti (ossia in corrispondenza della parola <sardi> in Tabella 1).

Nelle tre fasce si articolano progressivamente i diversi tipi di parole: parole "vuote" parole strumentali, parole principali (Vedi pag.18).

- La fascia delle *alte* frequenze è quella in cui ogni parola ha un numero diverso di occorrenze da ogni altra. Generalmente è composta all'incirca da 30 o 50 forme (a seconda delle dimensioni del corpus) e, fra queste, al più 4 o 5 sono parole principali, mentre le altre sono parole grammaticali.
- La fascia delle *medie* frequenze si caratterizza per avere al suo interno parole con diverse condizioni sia di parità sia di numero di occorrenze (nella quale cioè non tutte le possibili quantità di occorrenze sono presenti).
- La fascia delle *basse* frequenze possiede tutte le classi di frequenze decrescenti fino a  $V_1$  e contiene la stragrande maggioranza delle parole del vocabolario.

- **Livello di soglia e tasso di copertura del testo**

*Il livello di soglia* è quel valore di frequenza al di sotto del quale si assume che le parole abbandonate non siano rilevanti, mentre il *tasso di copertura del testo* ( $\%cop$ ) è dato dal valore percentuale  $N_{(s)}/N$  (dove  $N_{(s)}$  esprime l'ampiezza del *corpus* sopra il *livello di soglia*, mentre  $N$  il numero delle occorrenze del *corpus* stesso).

Nota: Il rango corrispondente al livello di soglia esprime una stima del numero di dimensioni o variabili attive che definiscono i *profili lessicali*.

Ai fini di *un'analisi del contenuto del corpus*, si riscontra empiricamente che un *livello di soglia di frequenza*, accettabile per selezionare le parole da studiare con tecniche multidimensionali, è posizionato nei primi decili delle *basse frequenze*, ossia intorno ad un 10% o 20% delle parole appartenenti a tale fascia (Vedi pag.27). Infatti, entro questo livello si trovano la maggior parte delle articolazioni del linguaggio riguardanti le parole principali, quanto ad aggettivi e sostantivi (e verbi, ma in misura inferiore, date le loro numerose flessioni).

In diversi studi si è riscontrato, inoltre, che il limite corrispondente al primo decile delle basse frequenze corrisponde spesso al valore  $N/V^{19}$ , mentre il *tasso di copertura del testo*, nell'intorno del primo decile delle basse frequenze, è

---

<sup>19</sup> Valore medio delle occorrenze in un *corpus* che a sua volta rappresenta all'incirca il 12% del vocabolario totale (dato dal valore  $V_{(s)}/V$ ) con un valore di frequenza compreso tra 7 e 12 a seconda che il *corpus* sia piccolo o grande.

compreso fra il 70% (in testi di almeno 8.000 occorrenze) e l'80-85% (in *corpus* medi o medio-grandi)<sup>20</sup>

- **Dimensioni minime del corpus**

Ai fini di un'analisi statistica è importante conoscere quali debbano essere le dimensioni *minime* di un *corpus*. Un criterio empirico che spesso si è rivelato utile è quello di osservare il rapporto *V/N*: se questo supera il 20%, il corpus non può considerarsi sufficientemente esteso<sup>21</sup>. Questa regola vale a meno che il testo non contenga più "idiomi" che si sovrappongono, com'è il caso del linguaggio "globale" delle mailing list in Internet, infarcite di inglese.

<<In generale, un corpus, ai fini di un'analisi su base statistica, è considerato piccolo quando si aggira sulle 50 pagine convenzionali, corrispondenti a circa 15.000 occorrenze (100 Kbytes). È di dimensione media intorno alle 150 pagine (45.000 occorrenze equivalenti a circa 300 Kb), è medio-grande quando raggiunge le 100.000 occorrenze (più di 300 pagine e circa 700 Kb)<sup>22</sup>. Quando il vocabolario di un corpus supera le 500.000 occorrenze (3 Megabytes, equivalenti a oltre 1.500 pagine) la sua estensione raggiunge la dimensione minima di un lessico. Affinché il suo vocabolario possa definirsi un lessico di frequenza occorre che il corpus sia ricavato da fonti eterogenee (tale da approssimare un campione di linguaggio) che

---

<sup>20</sup> Per ottenere quest'ultima percentuale di copertura in testi piccoli, occorre scendere al secondo decile delle basse frequenze.

<sup>21</sup> Il tasso di accrescimento di un vocabolario decresce all'aumentare della dimensione del corpus. In un testo molto corto, ad esempio di sole 100 occorrenze, quasi tutte le parole sono diverse. Via via che il testo cresce in ampiezza, le parole diverse aumentano dapprima velocemente e poi sempre più lentamente. Ciò dipende molto anche dal tipo di testo considerato.

<sup>22</sup> <<Una survey di 800-1.000 risposte non vuote ad una domanda aperta produce generalmente un testo compreso fra le diecimila e le ventimila occorrenze, a seconda dell'ampiezza delle risposte. Occorrono invece almeno 50 interviste libere di oltre mezz'ora di registrazione per riempire 100 pagine di testo

le unità di testo siano i lemmi e che le loro frequenze siano espresse in indice d'uso>>[Bolasco, 1999: 204].

- **Indici d'uso e di dispersione**

In taluni studi risulta utile misurare il grado di diffusione (ripartizione) delle occorrenze nelle varie fonti raccolte. Ciò viene ricavato da un indice **D** di dispersione - valido per ciascuna parola - dato dalla seguente formula:

$$D = 1 - \frac{V}{\sqrt{n-1}}$$

Dove :

$n$  = numero di fonti (testi di ampiezza diversa o subtesti);

$V = \frac{\sigma}{\bar{f}}$  (*coefficiente di variazione*) il cui massimo è pari a  $\sqrt{n-1}$  e lo si ottiene

quando tutte le frequenze parziali (frequenze *normalizzate* della parola nel subtesto *i-esimo*) sono concentrate in un solo testo.  $V$  è nullo, invece, quando tutte le frequenze parziali sono equidistribuite (freq. parziali uguali tra loro).

$\bar{f}$  è la *frequenza media*, pari alla frequenza normalizzata della parola nell'intero *corpus* (uguale alla somma delle frequenze parziali di una parola nei subtesti diviso il numero di tali subtesti).

$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (f_i - \bar{f})^2}$  è lo scarto quadratico medio della frequenza che è pari alla

radice quadrata della somma dei quadrati degli scarti delle frequenze normalizzate

---

trascritto. In sostanza, un testo che non raggiunga le 10.000 occorrenze possiede delle frequenze che nei valori bassi sono inattendibili statisticamente>>[Bolasco, 1999: 204].

della parola nel subtesto *i-esimo* (frequenza parziale) dalla frequenza media. Il tutto diviso  $n$  (numero delle parti, o sub testi, del *corpus*).

Per quanto sopra specificato il rapporto  $\frac{V}{\sqrt{n-1}}$  è sostanzialmente  $\frac{V}{V_{\max}}$

(*coefficiente di variazione* normalizzato) il quale varia nell'intervallo 0-1.

La dispersione  $D$  di una parola, quindi, è il complemento ad 1 del coefficiente normalizzato di variazione<sup>23</sup> della frequenza negli  $n$  subtesti in cui è suddiviso il

*corpus*. Ovvero  $1 - \frac{V}{V_{\max}}$ .

Poiché si può avere una dispersione molto simile con frequenze assai diverse, si preferisce calcolare - per ciascun termine - l'*indice d'uso*  $U$

$$U = D F$$

Ottenuto ponderando il numero assoluto delle occorrenze della parola ( $F$ ), con la dispersione  $D$ . Quando la dispersione è massima l'*uso* è pari alla frequenza.

---

<sup>23</sup> Il complemento ad 1 del *coeff. di variazione* deriva dall'esigenza di ottenere per ciascuna parola una misura di dispersione  $D$  allorquando si abbiano un egual numero di frequenze normalizzate nei vari testi.

## 1.7. Il trattamento del testo

Vi sono diversi principi di **normalizzazione**<sup>24</sup> attraverso i quali un testo viene trattato al fine di scegliere, nel sottoinsieme di parole scelte per l'analisi, quelle sulle quali intervenire per accrescerne il livello informativo.

In poche parole nel trattamento del dato testuale viene ridotta l'ambiguità e migliorata la **monosemia** (univocità dei significati), cercando, nel contempo, di lasciare intatto il contenuto del *testo* con il suo *sistema di variabilità dei significati*.

<<Più in generale, il criterio fondamentale che è alla base di ogni intervento sul testo si può così esprimere: *conservare distinte nel testo le variazioni significative in termini semantici e fondere le forme che costituiscono degli invarianti semantici.*>> [Bolasco, 1999: 213].

Nel paragrafo 1.5. si è già parlato di **lemmatizzazione**<sup>25</sup>. Essa è di fatto un processo di trattamento del testo attraverso cui vi è una trasformazione sistematica delle forme grafiche in *lemmi* (Vedi pag.17). Ciò risulta utile ed opportuno in alcuni casi (ad esempio per i verbi), in altri innocuo (aggettivi), mentre in altri ancora è addirittura dannoso (basti pensare ad alcuni sostantivi per cui il plurale spesso indica dei referenti concettuali diversi dal singolare: comunicazione ≠ comunicazioni; <scienze della comunicazione> e <scienza delle comunicazioni>. Linguaggi e multimedialità nel

---

<sup>24</sup> Nell'appendice D vengono riportati alcuni criteri per la correzione degli errori e per la normalizzazione del testo.

<sup>25</sup> Sono allo studio dei *lemmatizzatori* automatici basati su algoritmi di riconoscimento *markoviani* (Bolasco, dispense a.a. 1997-1998: 42).

primo caso, ingegneria e progettazione dei servizi di trasporto nel secondo. O ancora: paese ≠ paesi: l'uno è “il nostro paese”, l'altro sta per “le altre nazioni”).

È evidente allora che sono necessari altri interventi capaci di massimizzare la ricerca del carattere *monosemico* delle parole di un *corpus* per valorizzarne l'accezione interna pur rimanendo il più possibile ancorati al contenuto del *testo*. La trasformazione delle *forme grafiche* in *forme testuali* (Vedi par.1.5. ) è fra questi.

Nel passaggio dalle *forme grafiche* alle *forme testuali* (Vedi **Tabella 2**) si effettuano sostanzialmente due processi:

**I.** La *disambiguazione*, ovvero la distinzione tra forme *omografe-polisemiche* (Vedi pag.18) che può essere:

- *grammaticale* (lemmi diversi: <posto\_Verbo> e <posto\_Sostantivo>);
- *semantica* (diverse accezioni di uno stesso lemma; <posto-Sostantivo> può significare “luogo”, “impiego”, “spazio”, “sedile”, “centro”);
- *lessico-grammaticale* isolando un *poliforme* (forze\_politiche).

**II.** La *fusione* di forme che costituiscono degli invarianti semanticci (*sinonimia*. Vedi pag.19) che può riguardare:

- equivalenti grammaticali (le diverse flessioni di un aggettivo, o di determinate voci di un verbo);
- equivalenze semantiche (i raggruppamenti di forme diverse, individuanti tratti semanticci o insiemi di sinonimi).

---

**Tabella 2**

Esempio del processo di individuazione delle forme testuali

<i>forme grafiche</i>		<i>analisi lessico-grammaticale</i>			<i>lemmi</i>		<i>forme testuali</i>	
	occ.		occ.	cod.		occ.		occ.
POLITICA	1509	POLITICA_sost	866	1	POLITICA_sost (1+3)	1233	POLITICA_sost	866
		POLITICA_ESTERA_polirem.	114				POLITICA_ESTERA_polirem.	114
		POLITICA_ECONOMICA_polirem.	140				POLITICA_ECONOMICA_polirem.	140
		POLITICA_agg	389	2				
POLITICHE	370	POLITICHE_sost	113	3			POLITICHE_sost	113
		POLITICHE_agg	132	4				
		FORZE-POLITICHE_polrem.	125	5			FORZE-POLITICHE_polrem.	125
POLITICO	241	POLITICO_sost	0	6	POLITICO_agg (2+4+5+6+7+8+9)	1003	POLITICO_agg.	878
		POLITICO_agg	241	7			(2+4+7+9)	
POLITICI	116	POLITICI_sost	0	8				
		POLITICI_agg	116	9				
Total occ.	2236		2236			2236		2236

---

*Fonte:* Bolasco, 1999

Queste *disambiguazioni* o *fusioni* possono essere messe in atto con differenti strumenti.

a) ***L'isofrequenza***

Definiamo *isofrequenza* la condizione di equilibrio o di stabilità - in numero di occorrenze - esistente fra alcune forme flesse di uno stesso *lemma*. Basterà scorrere un vocabolario di un *corpus* secondo l'ordine alfabetico e ci si renderà conto del fenomeno. L'ipotesi che è alla base di questo comportamento è che tanto più un termine è usato con funzioni, significati o forme diverse, tanto più è probabile che esso accumuli un numero maggiore di occorrenze tali da procurare il fenomeno contrapposto alla *isofrequenza*: la *non-isofrequenza*.

Quindi, se è pur vero che l'esistenza dell'*isofrequenza* non può, di per sé, costituire la prova di un'equivalenza di significato nei termini coinvolti, al contrario, il riscontrare una *non-isofrequenza* costituisce spesso l'indizio di un utilizzo plurimo della forma in questione.

Questa circostanza segnala pertanto l'opportunità, vuoi di una *disambiguazione*, ad esempio estraendo una locuzione, vuoi di una *fusione*.

Uno dei casi più evidenti è quello in cui una forma semplice è parte integrante e fondamentale di un *poliforme*. Ad esempio si può osservare in un *corpus* che le due flessioni <locali> e <locale> si presentano con occorrenze assai diverse in virtù della presenza della *polirematica* <enti\_locali>. Una volta decodificate le occorrenze di tale composto, le due flessioni tornano in condizioni di *isofrequenza*.

b) ***La selezione dei poliformi***

Osservando un vocabolario di *forme grafiche* ci si accorge che molte parole comuni sono inspiegabilmente ai primi ranghi del vocabolario. Questa circostanza può essere il riflesso della presenza di *poliformi* (Vedi pag.14), in particolare quelli a contenuto prevalentemente grammaticale, che sono alla base della costruzione stessa del discorso. Si tratta soprattutto di *locuzioni* con funzione *avverbiale* (<in particolare>, <di nuovo>, <a casa>, <una volta>/<tempo fa>, <del tutto>, <alla fine>, <di fatto>), *aggettivale* (<a punto>, <a tempo determinato>, <in mano>, <alla mano>) o *prepositiva* (<fino a>, <in modo da>, <da parte di>, <rispetto a>), o di *congiunzioni composte* (<dal momento che>, <certo che>, <come mai>, <a condizione che>), di *formule idiomatiche* (<tutte queste cose>, <è una cosa che>, <per così dire>, <io credo>) o infine di alcuni *verbi idiomatici* con funzione ausiliare di *verbi supporto* (<rendersi conto>, <andar fatto>, <portare avanti>, <venir meno>, <dare vita>, <va bene/male>, <far parte>, <far fronte>, <fare presto/tardi>, <mettere a punto>, <prendere atto>). Tutti i componenti queste espressioni, viste come frasi fisse, risultano avere nel vocabolario di un *corpus* in *forme grafiche*, un numero di occorrenze alterato rispetto al solo uso ordinario, come parole semplici con il loro significato elementare, diretto od originario. Disambiguare almeno alcune fra queste espressioni diventa necessario e, a volte, essenziale. Si riscontra, infatti, che tali *poliformi* hanno un comportamento (sotto il profilo semantico) assai diverso dalle parole semplici costituenti.

Sarebbe allora interessante poter valutare quanto le occorrenze di un *segmento* (Vedi pag.13) incidono sulle occorrenze delle forme semplici che lo compongono.

Per far questo esiste un indice, denominato **IS**, costruito per selezionare alcune *polirematiche* di contenuto (Vedi pag.14).

$$IS = \left( \sum_{i=1}^L \frac{f_{segm}}{f_{fg_i}} \right) \cdot P$$

ove, date le  $L$  forme grafiche componenti il *segmento*, si pone a rapporto *la*  $f_{segm}$  (n° occorrenze del segmento) a *ciascuna*  $f_{fg_i}$  (occorrenze delle forme grafiche componenti), moltiplicando poi la somma di tutti questi quozienti per  $P$ , quantità che esprime il numero di *parole piene* (Vedi pag.18) presenti nel *segmento*<sup>26</sup>. Tale indice è sempre positivo, si annulla quando il segmento è composto solo da *parole vuote*<sup>27</sup> ed ha il suo massimo pari a  $L^2$ . Condizione quest'ultima in cui tutte le occorrenze della *forma singola* sono date proprio dalla frequenza del *segmento*.

L'indice **IS**<sup>28</sup> appena visto ci dà un aiuto per valutare l'opportunità della *lessicalizzazione* consentendo di valutare l'impatto della frequenza su alcune delle parole chiave coinvolte nel processo di trasformazione delle unità.

La *lessicalizzazione* è quel processo che porta a considerare un *sintagma* (o un qualunque raggruppamento di *parole*) come un solo elemento lessicale. In altri termini la lessicalizzazione consiste nella trasformazione del testo, dovuta al

---

<sup>26</sup> la ricerca automatica di tutti i segmenti ripetuti in un testo è per costruzione ridondante: avviene infatti cercando tutte le sequenze identiche di qualsiasi lunghezza 2, 3, 4 o 5 parole. Per cui ad esempio avremo <punto di> <di vista> <punto di vista> <dal punto di vista> <sotto il punto di vista> (vedi pag.13). Occorre quindi eliminare la ridondanza e selezionare solo quelli “pieni” (*polirematica*) <punto di vista>, prendendo atto che l'ordine di grandezza della frequenza con cui essi occorrono è assai inferiore a quello delle corrispondenti *forme grafiche* elementari. Comunque si apprezza la presenza di *segmenti* quando la loro frequenza è superiore o uguale a 3 o 4 occorrenze.

<sup>27</sup> Presupponendo di avere definito una lista di parole vuote, l'indice consente di scartare i segmenti vuoti o irrilevanti in termini di grado d'assorbimento; questi, generalmente, sono oltre l'80% dell'intero inventario

<sup>28</sup> Esiste anche un altro metodo per isolare sistematicamente i *poliformi* di un testo. Esso si basa sul confronto dell'inventario dei segmenti ripetuti di un *corpus* con una qualche lista significativa di

riconoscimento di una sequenza di forme grafiche, come una sola unità di senso o *lessia*. Ad esempio < capo dello stato> verrà modificata in un'unica unità lessicale <capo\_dello\_stato><sup>29</sup>.

Si prenda come esempio il seguente *segmento*.

“teste rasate” con frequenza  $F = 18$  ed  $IS = 3,636$ , con numero di occorrenze della parola “teste” nel *corpus* pari ad  $f_1 = 22$  e con numero di occorrenze della parola “rasate” pari ad  $f_2 = 18$ . Il valore di  $IS$  vicino al suo massimo rivela in effetti che vi è un buon assorbimento della *forma singola* (“pivot”) da parte del *poliforme* che la contiene. Infatti l’81% della forma <teste> ed il 100% della forma <rasate> viene “catturato” dal segmento.

Se la sequenza avesse assorbito circa il 50% della frequenza della parola, entrambi avrebbero apportato informazioni utili e differenti.

c) ***L'individuazione del linguaggio peculiare***

L’individuazione del linguaggio peculiare può essere vista come la ricerca dell’insieme minimo di parole massimamente rappresentativo del *vocabolario* che consente di ridurre le ambiguità presenti nel *corpus*.

Si parte dal presupposto che le parole più frequenti in un *corpus* (anche dette **parole tema**) non sempre sono **parole chiave** (*peculiar, tipiche*) del *corpus* stesso. Ove per *parola chiave* si intende una parola *sovra/sotto-utilizzata* rispetto alla sua frequenza standard nei normali contesti d’uso. E si sceglie un modello di riferimento

---

*poliformi*, specifica di un settore o di un genere di linguaggio. L’intersezione delle due liste - quella del *corpus* e quella specifica - consente di isolare i segmenti pieni.

<sup>29</sup> Il carattere “\_” *underscore* dovrà essere cancellato dall’elenco dei separatori (Vedi pag. 13).

(rappresentato da un tipo di lessico<sup>30</sup>) rispetto al quale calcolare il sovra/sotto-uso delle *parole chiave*. Così facendo ci si affida ad un criterio che consente di selezionare le parole di un corpus non soltanto sulla base del loro più o meno elevato numero assoluto di occorrenze.

Tale criterio misura la *peculiarità* in termini di *specificità* sia positiva che negativa.

La prima correlata con le parole più frequenti, mentre la seconda con quelle così rare da essere “quasi assenti”, forse perché volutamente evitate dal locutore.

La misura di specificità, per ciascuna parola, è allora data, ad esempio, dal seguente rapporto:

$$z_i = \frac{f_i - f_i^*}{\sqrt{f_i^*}}$$

che costituisce uno scarto standardizzato della frequenza relativa, dove  $f_i$  è il numero di occorrenze normalizzate della *i-esima* parola nel corpus ed  $f_i^*$  il corrispondente valore nel lessico assunto come modello<sup>31</sup>, mentre la quantità al denominatore è lo scarto quadratico medio della frequenza relativa (Vedi pag.46). Come è facile notare, questo rapporto è pari alla radice quadrata dell'i-esimo contributo ad un *chi-quadrato*.

In assenza di un modello di linguaggio di riferimento, si potrebbe lo stesso giungere all’individuazione delle parole chiave del corpus effettuando, in via preliminare,

---

<sup>30</sup> Con la crescita delle potenzialità informatiche di calcolo, attualmente non è difficile effettuare raccolte di testi per la messa a punto di liste di frequenza, assemblando stock sempre più ampi (anche milioni di occorrenze) di materiali riguardanti periodi, generi e situazioni differenti (un tempo solo testi scritti, più recentemente anche testi parlati). Queste liste permettono di costruire i cosiddetti *lessici di frequenza*: in pratica, dei vocabolari ordinati per numero decrescente di occorrenze, o meglio, secondo il loro rango in termini di frequenza d’uso. I lessici di frequenza possono essere utilizzati come modelli di riferimento per la valutazione del sovra/sottouso delle parole nel corpus oggetto di studio.

un'analisi delle corrispondenze sul *corpus* in *forme grafiche*, a soglia di frequenza elevata.

Così facendo si evidenziano sul primo piano fattoriale alcuni “punti cardinali” della struttura del contenuto. Si procede poi ad altre analisi, con soglie di frequenza via via decrescenti, per scoprire quali siano le parole che restano stabili in queste simulazioni e quali siano i contenuti che si definiscono come “sottocampi” o dettagli semantici di tali punti cardinali. Così facendo, si identificano i termini sui quali è opportuno concentrare gli interventi di disambiguazione o di fusione.

Nonostante siano stati appena descritti metodi e criteri per intervenire sul testo con una certa sistematicità, l'atteggiamento che occorrerebbe comunque assumere è quello di procedere con parsimonia: alcuni interventi, infatti, potrebbero procurare più danni che vantaggi (caduta di frequenza, frammentazione delle occorrenze e perdita della *forma* perché al di sotto della *soglia di frequenza* ecc..).

Uno dei criteri fondamentali che guidano gli interventi sul testo consiste da un lato nel tendere a ridurre il numero delle unità lessicali da considerare per l'analisi e dall'altro nel cercare di aumentare il tasso di copertura del testo (Vedi pag.29), a parità di numero di unità considerate<sup>32</sup>.

---

<sup>31</sup> Tali quantità possono essere espresse anche in termini d'indice d'uso (Vedi pag.38). Tale confronto è tanto più valido quanto più il *corpus* è connesso al lessico.

<sup>32</sup> Si ricorda che gli interventi riguardano circa un 10% delle forme del vocabolario da analizzare (che a sua volta potrebbe aggirarsi intorno al 12 % di *V* [Vedi pag.29]).

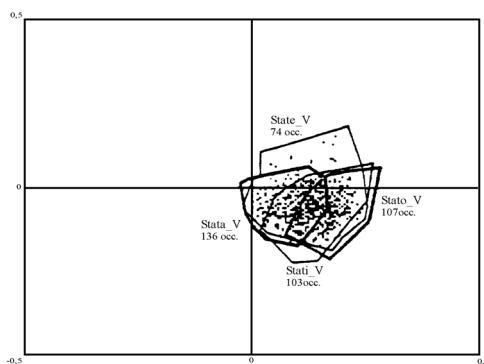
Esiste un test in grado di legittimare la scelta di fondere/non fondere o disambiguare più termini basato sulla ricostruzione - mediante simulazione - delle “regioni di confidenza” sul piano fattoriale<sup>33</sup>.

Si considerano le parole che sarebbero oggetto di pretrattamento.

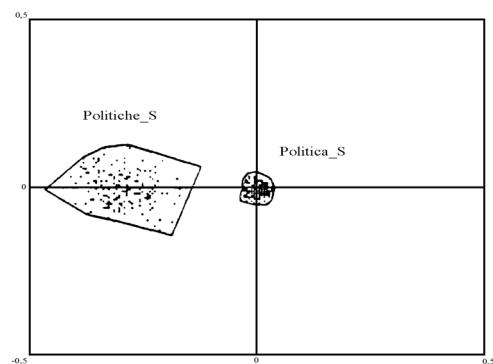
- Si effettuano le disambiguazioni e si analizza come quest’ultime si comportano sul piano fattoriale. Se i loro punti producono regioni disgiunte, la loro disambiguazione è legittima;
- si effettuano le fusioni e, come sopra, si studia il loro comportamento sul piano fattoriale. Se esse hanno regioni di confidenza fortemente incluse, è evidente che una loro fusione sotto un unico lemma non inficerebbe l’analisi.

In Figura 2 ed in Figura 3 si riportano alcuni esempi tratti dalle dispense del Prof. Bolasco da cui si evince come sia legittimo fondere le quattro voci del participio passato del verbo essere, mentre come sia inopportuno mischiare il singolare ed il plurale del nome politica.

**Figura 2:** Particípio passato del verbo essere



**Figura 3:** Singolare/plurale del nome politica



<sup>33</sup> Si sfrutta in pratica la proprietà vicinanza=somiglianza dei punti sui piani. Le matrici sono matrici di frequenza <parole x subtesti>.

## 1.8. Tipi di matrici di dati testuali

Al fine di un'analisi statistica del contenuto, ogni *corpus* deve tradursi in un'opportuna matrice di dati.

Fondamentalmente, si possono avere tre diversi tipi di matrici di dati testuali<sup>34</sup>:

- {frammenti x forme}, in cui in riga si hanno i *frammenti* di testo, da considerarsi come unità statistiche d'analisi (od "osservazioni") e in colonna si hanno le *forme* selezionate per lo studio, da considerarsi come variabili relative a ciascuna unità.

		$F_{\max}$	1	2	3	...	$j$	...	$F_{(s)}$
	<i>Forma</i>	<i>di</i>	<i>è</i>	<i>il</i>	<i>...</i>	<i>...</i>		<i>era</i>	$V_{(s)}$
	<i>Framm.</i>								
1		1	1	0	...	0	...	0	
2		1	0	1	...	1	...	0	
...		...	...	...	...	...	...	...	
<i>i</i>		0	1	1	...	1	...	1	
...		...	...	...	...	...	...	...	
<i>n</i>		1	1	0	...	0	...	0	

Dove le righe possono essere: singole risposte degli intervistati, le singole proposizioni, i singoli versi, i commi. Mentre le colonne sono costituite dalle *unità lessicali* (Vedi pag.38) selezionate dal vocabolario del *corpus*, espresso ora in *forme grafiche*, *lemmi* o *segmenti*, ora in forme miste (*lessie*, *forme testuali*). Le colonne della matrice costituiscono le informazioni elementari di ogni unità e sono quindi le componenti di un *profilo lessicale*, che sarà oggetto dello studio. In ogni casella  $(i,j)$

---

<sup>34</sup> Le matrici conterranno solo le unità lessicali,  $V_{f\max}, \dots, V_s$ , che superano una prescelta soglia di frequenza  $s$ , dal momento che si analizza non l'intero vocabolario di un corpo, ma un sua parte (Vedi pag. 29).

della matrice è riportato il numero di occorrenze dell'unità lessicale *j-esima* presente nell'osservazione *i-esima*. Di fatto questa matrice è in prevalenza una tabella booleana (1 = presenza, 0 = assenza) poiché solo raramente vi sono più occorrenze di una stessa parola in un frammento. Sarà comunque una matrice di tipo sparso, cioè con oltre il 95% delle caselle nulle, poiché ciascun frammento è composto di qualche decina di parole, mentre l'ampiezza del vocabolario è solitamente di diverse centinaia - se non migliaia - di unità lessicali. In questa matrice, si perde l'informazione relativa alla disposizione dei termini all'interno di ciascun frammento, mentre se ne conosce la loro combinazione. Ad esempio, il frammento <una casa bella veramente> coinciderebbe con <veramente una bella casa>. Tali "perdite d'informazione" sono tuttavia trascurabili rispetto al contenuto informativo dovuto alla combinazione dei termini stessi; se, al contrario, dovessero ritenersi fondamentali, non si dovrà fare ricorso a questo genere di matrici.

A questa matrice può venire associata, per ciascun frammento, una serie di categorizzazioni (A, B, C, ... ) che registrano le modalità di altrettante variabili qualitative numeriche. Ad esempio, nel caso di un corpus di una raccolta di articoli di giornale (ogni articolo costituisce un frammento), è possibile associare all'articolo categorizzazioni sull'autore (A), sulla posizione nella pagina (B), sull'argomento trattato (C) ecc.

Si avrà così la seguente matrice delle variabili categoriali:

A	B	C	D
1	3	5	2
2	1	3	1
...			
2	3	2	2
...			
1	3	1	2

- {forme x testi}, in cui si ha in riga il vocabolario selezionato allo scopo e in colonna i testi (o parti) secondo cui si considera suddiviso il *corpus*. *L'unità lessicale* è quindi l'unità statistica d'analisi e il *testo* costituisce la variabile di studio. L'informazione statistica interna alla matrice è la frequenza (numero di occorrenze assolute) con cui una “parola” (*forma, segmento o lessia*) ricorre in ciascun testo. Il profilo lessicale d'interesse è spesso il profilo colonna, dal momento che si confronteranno i diversi testi sulla base della differente presenza (frequenza) delle parole. In ogni caso, la lettura diretta di tali profili richiederebbe di trasformare le occorrenze assolute in occorrenze normalizzate.

	<i>Testo</i>	1	2	...	<i>j</i>	...	<i>T</i>
<i>Forma</i>							
1	<i>di</i>	82	35	...	40	...	145
2	<i>è</i>	56	77	...	19	...	70
3	<i>il</i>	49	62		33	...	12
...		...	...	...	...		...
<i>i</i>	...	29	10		25	...	56
...		...	...	...	...		...
<i>V<sub>(s)</sub></i>	<i>era</i>	3	1	...	5	...	7

- {forme x forme}, in cui sia le righe che le colonne descrivono il "vocabolario" prescelto. L'informazione statistica interna alla matrice è una misura di similarità (espressa come sola presenza/assenza o come grado di correlazione) che registra il tipo o livello di co-occorrenza fra le forme, all'interno dei testi. Questa matrice può venire utilizzata ad esempio per ponderare i profili colonna nelle matrici del tipo {frammenti x forme}.

	cosa	madre	casa	vita	dare	...
cosa	1	0	1	0	0	...
madre	0	1	1	1	1	...
casa	1	1	1	1	0	...
vita	0	1	1	1	1	...
dare	0	1	0	1	1	...
...	...	...	...	...	...	...

## 1.9. L'analisi multidimensionale del contenuto

Una volta raggiunto un buon livello del dato linguistico come dato statistico, ossia una volta ricercata una qualche approssimazione della *monosemia* per le parole significative ai fini dell'analisi, si procede all'applicazione delle tecniche statistiche per differenti livelli di analisi.

Si possono individuare due livelli di studio. In un primo livello si affronta per così dire uno studio esterno o “verticale” del *testo*: l'analisi è di tipo *lessicale* perché l'interesse si basa sulla terminologia utilizzata (*vocabolario*). Il risultato finale è l'*analisi delle specificità*.

In un secondo livello si fa ricorso a tecniche statistiche multidimensionali (ad esempio di tipo fattoriale) capaci di studiare il contesto generale delle varie co-occorrenze delle parole attraverso lo studio dei profili lessicali descritti dalle matrici dei dati, fino alla ricostruzione dei sintagmi fondamentali presenti nel *corpus*. Il risultato finale è la ricostruzione dei principali modelli di comportamento del *senso*.

- **Specificità di forme e frasi in un testo**

Con il termine *specificità* si intende indicare se e quanto una parola sia tipica o specifica di un sub-testo, nell'ambito di uno stesso *corpus*, o - più in generale - quanto una forma sia specifica rispetto ad un qualche modello di linguaggio di riferimento.

Una misura di *specificità* di una parola in un testo, di solito ottenuta a livello di forme testuali, viene calcolata a partire dalla tabella di frequenza che ripartisce le occorrenze totali di una forma del corpus nei vari sub-testi in cui essa occorre.

Essa può esser data semplicemente da:

$$z = \frac{(x - x_{teor})}{\sigma_x}$$

Che è del tutto simile alla formula già vista a pag.40.

Si arriva a tale formula per il calcolo della *specificità* partendo da alcune semplici considerazioni.

<<Come è noto si possono indicare con:

$$E(x) = n \cdot p \quad \text{e} \quad \sigma_x = \sqrt{n \cdot p \cdot q}$$

rispettivamente la media del numero assoluto di occorrenze di una parola e il suo scarto quadratico medio, ove  $p$  (e  $q$ ) è la probabilità, come frequenza relativa, dell'apparire della parola (e rispettivamente del suo non apparire) in un testo, ed  $n$  è il numero di “prove” che si immagina di effettuare per ottenere la parola in oggetto.

Nel nostro caso  $n$  è pari al numero totale di occorrenze nel sub-testo: ipotizzando ogni *tranche* di corpus della stessa dimensione,  $n$  è costante in tutto il corpus.

Questo schema teorico sottintende, nell'ipotesi di indipendenza fra eventi, che l'apparire delle occorrenze di una parola in ciascun sub-testo possa essere immaginato come un evento aleatorio, ove  $p$  appunto è la probabilità di ottenere quella parola ogni  $n$  “prove”.

Ogni qualvolta si ottiene una proporzione di occorrenze di molto superiore (o inferiore) a questa quantità  $n \cdot p$  si può supporre che ciò non sia dovuto al caso ma

piuttosto sia l'espressione di una qualche "causa" specifica. Ha senso allora voler misurare in termini di uno scarto relativo questa differenza. Tale scarto prenderà la forma seguente:

$$z = \frac{(x - x_{teor})}{\sigma_x}$$

Ora in ambito linguistico, la frequenza relativa  $p$  di una parola in un testo è di fatto sempre bassissima, per cui, volendo semplificare il calcolo, possiamo esprimere  $\sigma_x$  come  $\sigma_x = \sqrt{n \cdot p}$ , in quanto il prodotto di  $p \cdot q$  è praticamente sempre uguale a  $p$ .

Ma il tal modo lo s.q.m. della frequenza assoluta di una parola è pari alla radice quadrata della frequenza assoluta teorica. In questo senso lo scarto standardizzato [...] assume la forma

$$z = \frac{(x - x_{teor})}{\sqrt{x_{teor}}} \gg [\text{Bolasco, 1999: 227}]$$

Questo rapporto può essere valutato utilizzando i criteri classici della significatività statistica con alcune considerazioni.

<< [...] Assumendo il *corpus* come una popolazione e ogni sua parte (sub-testo) come un campione, il modello distributivo di riferimento - per valutare in termini probabilistici il numero di occorrenze di una parola presenti in questo campione - è quello di una legge *ipergeometrica*, legge vicina alla distribuzione *multinomiale* quando le frequenze relative sono molto piccole rispetto alla popolazione. Sotto particolari condizioni (frequenze assolute osservate non inferiori ad una certa frequenza) a sua volta quest'ultima è ben approssimata da una variabile casuale normale. In pratica quando si stabilisce un livello di soglia minimo sul numero di

occorrenze di una parola, per considerarla "in analisi", il calcolo delle specificità avviene attraverso un *valore-test* che confronta la frequenza relativa di una parola nella parte, con la corrispondente frequenza relativa nel corpus totale. Questo test è effettuato sotto l'ipotesi di un'approssimazione normale, per cui è possibile assumere i classici limiti degli intervalli di confidenza di una variabile standardizzata  $z$  e assumere le regole ben conosciute della distribuzione di Gauss.

Quando  $z$  è all'incirca intorno allo zero ciò significa che la parola è presente nel sub-testo in proporzioni puramente aleatorie, ossia tanto quanto in media ci si può aspettare. In tal caso la parola non è significativa, quindi in qualche modo è "banale", come dire che fa parte del vocabolario di base (necessario alla costruzione) del testo.

Se  $z$  è superiore, in valore assoluto, a 2 sappiamo che la sua presenza è significativamente diversa da quella attesa (sotto una ben determinata ipotesi teorica, che è quella dell'equidistribuzione e quindi di indipendenza, all'interno di un certo schema di estrazione e di un modello probabilistico di riferimento). Quindi il numero delle sue occorrenze è significativo, sia in termini positivi che negativi.

Nel primo di questi due ultimi casi si dirà che il numero di occorrenze della parola in esame nel sub-testo supera largamente il valore atteso per puro effetto del caso e che la parola è *caratteristica* del testo (specificità positiva). Nel secondo caso si dirà che la sua così bassa frequenza è anch'essa significativa, per cui vi sarà una qualche causa per la quale la parola non è presente nel testo quanto ci si potesse aspettare. La parola si dice allora anti-caratteristica o "rara" o anche mal rappresentata.

Una selezione di parole con specificità positive  $S^+$  o negative  $S^-$  consente di individuare alcuni tratti salienti del sub-testo, in modo da identificarne i principali contenuti>> [Bolasco, dispense anno accademico 1997-1998: 52].

Una estensione del criterio di selezione delle *forme caratteristiche* è quello dell'**estrazione di frasi significative** che consiste nell'identificare alcuni contesti locali che appunto contengono tali *forme*. Ciò si basa sul principio che una frase è tanto più caratteristica quante più parole ad alta specificità essa contiene. Pertanto se si considera di calcolare il valor medio dei *valori-test* delle parole che formano la frase, più è elevata questa quantità, più significativa è la frase<sup>35</sup>.

Con una semplificazione si può immaginare di adottare come informazione il rango associato alle forme caratteristiche per ciascun sub-testo. Ovvero data la lista delle forme di un sub-testo, secondo la loro specificità positiva decrescente, e consideratone il rango (ranghi bassi = alta specificità) si calcola il rango medio delle parole della frase. Se il rango medio è piccolo vuol dire che essa contiene solo parole caratteristiche.

- **Gli assi fattoriali: una interpretazione in chiave linguistica**

Un insieme di unità lessicali - ordinate lungo un asse fattoriale - può essere concepito e interpretato alla stregua di un *sintagma* (Vedi pag.14). Il concetto che c'è alla base è che l'ordinamento degli elementi su di un asse crea significato. La lettura di quest'ordine genera il senso (d'un enunciato) da attribuire all'asse. Sarà

---

<sup>35</sup> Naturalmente questa misura è influenzata dal numero di parole, in quanto tende a privilegiare le frasi corte. Infatti, a parità di forme caratteristiche, ogni parola banale che si aggiunge nel calcolo tende ad abbassare la media

così possibile trarre il senso latente e globale del sistema di significati elementari messi in gioco dalle forme oggetto di analisi su un fattore, limitandosi alla interpretazione delle sole "parole" significative per l'asse considerato.

Il differente disporsi degli elementi su di un nuovo asse genera un diverso *sintagma* e conseguentemente un differente punto di vista dal quale leggere i contenuti del *corpus*.

## II. USENET

### 2.1. Cos'è USENET

<< *Usenet* riguarda le persone [...] *Usenet* è l'insieme delle persone che sanno cos'è *Usenet* >>. [ Dern, 1995: 197]

*NetNews*, più comunemente chiamata *Usenet*, è un sistema condiviso (*sharing system*) di messaggi che provengono da tutto il mondo in un formato standard. In poche parole *Usenet* è una comunità mondiale di bacheche elettroniche (*BBS*)<sup>36</sup>, strettamente associate ad Internet, le cui informazioni sono costituite da singoli messaggi, ciascuno dei quali può essere letto e condiviso da molti utenti. I messaggi sono organizzati in gruppi di argomenti o “*Newsgroup*”.

*Usenet* comprende molti computer in molte organizzazioni e sedi diverse.

Gli utenti di *Usenet* leggono e forniscono contributi (affiggono messaggi) nella sede *Usenet* locale. Ogni sede *Usenet* distribuisce le affissioni dei suoi clienti alle altre sedi *Usenet* in base ai vari parametri di configurazione impliciti ed esplicativi e a sua volta riceve affissioni dalle altre sedi.

Tramite *Usenet* milioni di utenti di computer in tutto il mondo condividono informazioni, presentano domande e risposte, conducono discussioni tra più utenti.

---

<sup>36</sup> Le *BBS*, a differenza della posta elettronica, consentono di organizzare le informazioni come *risorse comuni condivise*, in directory che non appartengono al singolo utente. I messaggi arrivano in aree di proprietà del software della bachecca elettronica: molti utenti possono leggere contemporaneamente la stessa copia, come se si trattasse di un giornale comune o di un avviso affisso ad un muro, mediante programmi software progettati per organizzare e leggere i numerosi messaggi. Invece di inviare un messaggio ad una persona, lo si invia, o “affigge”, in un gruppo della *BBS*.

Esistono pacchetti software speciali per gli utenti di *Usenet* (*Newsreader* per la lettura e l'affissione di messaggi).

<< *Usenet* non è un'organizzazione. Nessuna persona o gruppo ha autorità su *Usenet* nel suo insieme. Nessuno controlla chi riceve un contributo, quali articoli sono propagati nelle varie sedi, chi può affiggere gli articoli o qualsiasi altra cosa. Non esiste un'azienda *Usenet*, né un gruppo di utenti *Usenet*: ogni utente è solo >>. [ Dern, 1995: 198].

I messaggi *Usenet* viaggiano con qualsiasi mezzo, che può essere *UUCP* (*Unix to Unix copy*), in cui i computer effettuano chiamate da modem a modem e inoltrano le copie di tutti i messaggi appropriati, oppure attraverso *Internet*<sup>37</sup> e anche via radio, dischetto, nastro di *backup* e *CD-Rom*.

---

<sup>37</sup> Nel settore delle reti, il termine tecnico per designare il collegamento di reti è *internetworking*, mentre per indicare una rete di reti viene impiegato il termine *internetwork o internet*. La nuova internet work, incentrata su ARPAnet fu soprannominata Internet, con la "I" maiuscola.

## 2.2. Le origini di *USENET*

*Usenet* nacque nel 1979 quando due studenti della *Duke University* nella Carolina del Nord – *Tom Truscott* e *Jim Ellis* – pensarono di collegare i rispettivi computer per scambiare informazioni avvalendosi di servizi UUCP (*Unix to Unix copy*) che erano appena stati aggiunti in una nuova versione del sistema operativo *Unix*. *Steve Bellovin*, a quei tempi studente presso la *University of North Carolina*, scrisse un software che gestiva lo scambio dei messaggi e che fu installato nei computer delle due università. Il servizio prese il nome “*News*” e le sedi che lo fornivano divennero note come *Usenet* (reti di utenti *Unix*).

Negli anni successivi questi ed altri membri della comunità *Unix*, tra cui *Steve Daniel*, *Mark Harton*, *Matt Glickman* e *Rick Adams*, scrissero nuove versioni del software ‘*News*’ per la lettura, l’affissione e la distribuzione di messaggi *News* attraverso la rete *Usenet* in continua espansione.

Nella metà degli anni Ottanta, il servizio *Usenet* si era diffuso nei sistemi *Unix* di migliaia di università e organizzazioni operanti nel settore della fabbricazione e dell’uso dei computer. Vennero adottate convenzioni per coordinare il collegamento delle sedi *Usenet*, l’avviamento di nuovi gruppi *News* e le regole sociali.

Alla fine degli anni Ottanta, la rete *Usenet* si era già affermata come risorsa principale per discussioni e sviluppi tecnici all'interno di *Unix*, di *TCP/IP*<sup>38</sup> e delle comunità *Internet*, oltre che come comunità pienamente autonoma.

Nei primi anni Novanta *Usenet* è disponibile in tutti i continenti, compresa l'Antartide.

---

<sup>38</sup> *TCP/IP* è l'acronimo di (*Transmission Control Protocol/Internetworking Protocol*).

*ARPA* (*Advanced Research Projects Agency*), negli anni sessanta, attuò una serie di progetti basati sulla creazione di hardware e software per realizzare una piccola rete sperimentale basata sulla commutazione di pacchetto.

Dal progetto *ARPAnet* e dal concetto di *internetworking* (ovvero del collegamento di singole reti in un'entità più grande) si sviluppò un nuovo *protocollo* di comunicazione che permise l'interconnessione di reti a commutazione di pacchetto eterogenee: il protocollo *TCP/IP* (*Transmission Control Protocol/Internetworking Protocol*).

*La commutazione di pacchetto* è un metodo di condivisione di una rete di computer, per cui ogni flusso di dati viene suddiviso in pacchetti. Ogni pacchetto è formato da una porzione di dati e da una sezione denominata *intestazione* contenente informazioni quali l'indirizzo del mittente e del destinatario, il numero di ID del pacchetto per consentire il riassemblaggio delle porzioni di dati nell'ordine esatto, l'ora della creazione del pacchetto e le informazioni per verificare che il contenuto del pacchetto sia stato trasmesso correttamente.

Il processo di trasmissione di un pacchetto prende il nome di *commutazione*. I pacchetti vengono commutati attraverso la rete (definita come un insieme di *linee* e *nodi*, dove le linee sono i percorsi nei quali scorrono i dati, mentre i nodi sono i punti in cui le linee si intersecano e le risorse, ovvero i dati, vengono trasferiti ad altre linee). Nella rete i pacchetti sono trasferiti da un commutatore di pacchetto (nodo) di una linea al successivo che esamina l'intestazione per decidere a quale linea commutare (o inoltrare) il pacchetto. Con la commutazione di pacchetto molti flussi di dati possono essere inseriti e mescolati su di una linea comune per essere in seguito riordinati nella loro forma originaria all'altra estremità. Grazie alle informazioni contenute in ciascun pacchetto, i pacchetti provenienti da una sorgente possono anche essere trasmessi attraverso percorsi (route) diversi e arrivare tutti a destinazione.

## 2.3. I messaggi in *Usenet*

La suddivisione nei vari *Newsgroup*<sup>39</sup> delle migliaia di articoli affissi ogni giorno in *Usenet* contribuisce ad identificare quali messaggi vale la pena di leggere. Tuttavia anche questo può non essere adeguato: un gruppo di notizie può arrivare facilmente a decine e anche centinaia di affissioni quotidiane.

Un metodo per ovviare a questo inconveniente è il “concatenamento” dei messaggi, ovvero interconnettendo gli articoli secondo l’ordine delle risposte. Ogni catena di discussioni è un albero di articoli, in cui tutti gli articoli di risposta (figli) si diramano dai rispettivi articoli di origine (genitori).

L’argomento di un articolo è definito come il contenuto del relativo campo “Subject”.

L’*RFC (Request for Comments)*<sup>40</sup> 1036 (*Standard for Interchange of USENET Messages*) reperibile all’indirizzo Internet <http://www.ietf.org/> è un documento che contiene tutti i riferimenti per il formato standard dei messaggi *Usenet* e per alcuni standard di trasmissione delle “*News*”.

---

<sup>39</sup> Un gruppo di notizie o *Newsgroup* può essere diviso in sottogruppi in una sorta di gerarchia. L’assegnazione dei nomi nei *Newsgroup* di *Usenet* segue determinate convenzioni: i nomi dei gruppi e dei sottogruppi sono separati da un punto (.). Il nome della parte sinistra indica la gerarchia di massimo livello. Per esempio *it.discussioni.auto* indica il sottogruppo auto nel gruppo discussioni della gerarchia *it* (italiano).

<sup>40</sup> I documenti *RFC (Request for Comments)* sono le serie ufficiali di documenti che definiscono gli standard e gli altri aspetti di Internet, identificati sovente con un numero di *RFC* (per esempio: *RFC 1822*) oltre che con un titolo. Scritte da sviluppatori di spicco di Internet, le *Request for Comments* vanno dalle specifiche dei protocolli alle proposte, alle analisi, alle specifiche, ai resoconti generali, fino ad argomenti più leggeri come saggi umoristici occasionali o argomenti culturali quali poesie.

La regola è che tutti i messaggi *Usenet* devono essere formattati come messaggi di posta Internet in accordo con l'*RFC 822*<sup>41</sup>.

Lo standard *Usenet* è però più restrittivo rispetto lo standard Internet. Si hanno infatti alcune necessità in più e l'impossibilità di usare alcune caratteristiche di Internet.

In generale, comunque, è possibile scrivere un messaggio *Usenet* con un *tool* per comporre messaggi di posta Internet.

I messaggi *Usenet* sono composti da una intestazione, costituita da una serie di righe di testo (*header lines*) seguita dal “corpo” del messaggio (*body*).

L'intestazione è separata dal corpo da una riga vuota (*null line*).

Ciascuna riga dell'intestazione è formata da:

- una parola chiave o nome del campo (*field-name*) scritta in caratteri stampabili *ASCII*<sup>42</sup> che hanno valore tra il 33 ed il 126, eccetto il carattere “:”
- il carattere “.”
- uno spazio
- altre informazioni addizionali.

Ciascuna riga termina con il carattere di controllo di ritorno carrello (*carriage-return*) o di avanzamento riga (*line-feed*).

---

<sup>41</sup> L'*RFC 822* è un documento che contiene i riferimenti per il formato standard dei messaggi di testo “*ARPA Internet*”.

<sup>42</sup> *ASCII* è l'acronimo di *American Standard Code for Information Interchange*, ovvero un sistema di codifica che permette una corrispondenza biunivoca tra i numeri binari ed un insieme di caratteri (alfanumerici e di servizio). Ciò che ne deriva è una tabella su cui viene riportato il numero binario con il corrispondente carattere di riferimento. Nelle tabelle *ASCII* si usa indicare il valore numerico in formato decimale o esadecimale. La codifica *ASCII* può essere a 7 bit oppure a 8 bit. Nel primo caso comprende  $2^7=128$  caratteri dei quali, quelli fino al valore decimale 31, sono caratteri di controllo non stampabili. Nel secondo caso comprende  $2^8=256$  caratteri ovvero i caratteri di controllo, quelli stampabili più gli estesi. Oltre a quella *ASCII* esistono altre tabelle codici nate per soddisfare esigenze diverse come quella di gestire un sistema in una lingua diversa dall'inglese.

Alcune righe di intestazione sono obbligatorie, altre opzionali.

Le righe obbligatorie sono:

- **From:** contenente l'indirizzo di posta Internet della persona che ha spedito il messaggio. Esso può contenere il nome completo della persona fra parentesi **dopo** l'indirizzo di posta. L'*RFC 822* infatti specifica che tutto il testo fra parentesi è interpretato come un commento. In questo caso, il nome completo non è considerato come un commento ma come una parte opzionale della riga d'intestazione.

Il campo From può anche contenere il nome tra parentesi angolari, inoltre il nome dell'utente dovrebbe essere riportato tutto in minuscolo in quanto “case sensitive” (sensibile alle lettere minuscole/maiuscole).

Tutti validi sono i seguenti esempi:

- From: alessandro@mclink.it
  - From: alessandro@mclink.it (Alessandro Stabellini)
  - From: Alessandro Stabellini <alessandro@mclink.it>
- mentre l'indirizzo Alessandro@mclink.it è differente, ad esempio da alessandro@mclink.it.

Il nome completo può contenere tutti i caratteri stampabili *ASCII* eccetto i seguenti:

“)”, “(“, “<”, “>”, “;”, “.:”, “@”, “!”, “/”, “=”, “,.”.

- **Date:** contenente la data in cui il messaggio è stato originariamente spedito, l'ora, i minuti i secondi della sua spedizione ed un *offset* (compensazione) di +/- ore e minuti.
- **Newsgroup:** nel quale è riportato il nome del *Newsgroup* o dei *Newsgroups* di appartenenza del messaggio. Diversi nomi sono separati da una virgola.
- **Subject:** ovvero l'oggetto del messaggio. Se il messaggio è stato spedito in risposta ad un altro messaggio il *Subject* inizierà con quattro caratteri: "Re: ". In questo caso la riga d'intestazione "**References:** " è richiesta.
- **Message-ID:** contenente un identificatore del messaggio. Posto tra parentesi angolari deve avere questo formato: <numero identificativo@numero completo del dominio o dell '*host* attraverso cui il messaggio è entrato nel *Network*>
- **Path:** indica il percorso preso dal messaggio per arrivare nel sistema. I segni di punteggiatura (tranne il punto) separano i vari percorsi oppure i commenti.

Le righe opzionali sono:

- **Organization:** contenente l'organizzazione alla quale il sottoscrittore (o la sua macchina) del messaggio appartiene;

- **References:** contenente un numero di riferimento, ovvero il “**Message-ID**” del messaggio del quale il presente ne è una risposta.
- **Reply-to:** contenente l’indirizzo presso il quale devono essere indirizzati i messaggi di risposta. Per questo campo ritroviamo la stessa sintassi vista per il campo **From:**
- **Sender:** contenente il riferimento della persona o ente responsabile dell’accesso nella rete.
- **Follow-up:** contenente i *Newsgroups* o il *Newsgroup* presso cui una copia del messaggio è stata spedita.
- **Expires:** contenente la data in cui il messaggio scade e può essere cancellato. Se non presente si assume come data di scadenza quella impostata di *default*.
- **Content:** se presente indica che il messaggio è un messaggio di controllo, non rivolto all’utente, derivante dal meccanismo del *Newsgroup*. Se non presente, il campo **Subject:** può essere usato per gli stessi scopi: se i primi quattro caratteri del *Subject* sono “*cmsg*” significa che il messaggio è di controllo.

- **Distribution:** in aggiunta al campo **Newsgroup:** altera la distribuzione dei messaggi. Potrebbe essere ad esempio usato per limitare la distribuzione dei messaggi ad utenti di alcune zone. Diversi nomi sono separati da una virgola.
- **Summary:** contenente un breve sommario del messaggio.
- **Approved:** è presente per i messaggi indirizzati ai *Newsgroup* moderati<sup>43</sup>. Potrebbe essere aggiunto dal moderatore e contiene il suo indirizzo di *E-mail*.
- **Lines:** contiene il numero di righe del corpo del messaggio.
- **Xref:** contiene il nome dell'*host* (senza dominio) ed il nome del *Newsgroup* (separati da uno spazio vuoto) insieme ad altre informazioni circa il numero progressivo del presente messaggio ed il numero totale dei messaggi presenti nel *Newsgroup* stesso (questi ultimi separati dal carattere “;”).

---

<sup>43</sup> I *Newsgroup* moderati sono *Newsgroup* in cui tutti i messaggi sono sottoposti al vaglio di una persona definita il moderatore del gruppo. Il moderatore dovrà decidere della pertinenza di ciascun messaggio all'argomento del *Newsgroup*.

## 2.4. Il formato dei messaggi *USENET*

Mentre l'*RFC* 822 specifica i dettagli circa l'intestazione dei messaggi, l'*RFC* 1521<sup>44</sup> definisce il formato del “corpo” del messaggio.

Lo standard della posta elettronica così come descritto nell'*RFC* 821 e 822, limita il contenuto dei messaggi a linee di testo *ASCII* a 7 bit non più lunghe di 1000 caratteri, consentendo l'uso dei caratteri stampabili fino al valore *ASCII* 128.

Tale limitazione non permetterebbe l'adozione di un particolare insieme (*set*) di caratteri diverso da quello *ASCII* a 7 bit, né l'invio di dati in formato binario (non testuale), come immagini e audio.

Per ovviare a questo inconveniente i programmi di invio e ricezione delle “*News*” o delle “*Mail*” (“*Local mail user Agent*”, ovvero programmi, con i quali gli utenti spediscono e ricevono messaggi) implementano un sistema di conversione (codifica) standard di tutti i dati non testuali in caratteri stampabili *ASCII* a 7 bit.

A ciò si unisce l'estensione del servizio *SMTP* “*Simple Mail Transfer Protocol*” (protocollo di trasmissione delle *mail*) - così come riportato nell'*RFC* 1426<sup>45</sup> - che consente il trasporto di messaggi *MIME* a 8 bit.

L'estensione *SMTP* lascia, però, la possibilità di limitare la lunghezza delle linee di testo le quali possono continuare ad essere non più lunghe di 1000 caratteri.

---

<sup>44</sup> L'*RFC* 1521 decreta lo standard *MIME* (*Multipurpose Internet Mail Extensions*).

<sup>45</sup> L'*RFC* 1426 decreta l'estensione ad 8 bit del servizio *MIME*.

L'*RFC* 1521 introduce nuovi campi d'intestazione che definiscono il formato del corpo del messaggio e consentono una compatibilità tra tutti i programmi di posta elettronica.

Eventuali campi o valori dei campi stessi, al di fuori dello standard decretato dall'*RFC* 1521, devono essere indicati facendoli precedere dai caratteri “X-”.

I campi standard sono i seguenti:

- **Mime-Version:** che riporta la versione dello standard *MIME* adottato per il corpo del messaggio. Il campo ha la seguente struttura:

*Version*: = “**Mime-Version**”: “1*digit* (intero).”1*digit* (intero)” .

I commenti possono essere inseriti tra parentesi dopo il numero di versione *MIME*, in accordo con l'*RFC* 822.

In assenza del campo *Mime-Version* il corpo del messaggio viene interpretato come testo in caratteri *US-ASCII*.

- **Content-Type:** che descrive il tipo di dati contenuti nel corpo del messaggio di modo che il programma di posta (*Local mail user Agent*) possa rappresentare in modo appropriato il messaggio stesso.

Il *Content-Type* può riportare 7 valori, ciascuno con alcune varianti. Tali valori possono essere scritti sia in caratteri maiuscoli che minuscoli.

Se il *Content-Type* non è specificato si assume che il corpo del messaggio è rappresentato da informazioni testuali in caratteri *US-ASCII*.

I valori sono i seguenti:

<b>text</b>	Informazioni testuali	/plain <sup>46</sup>	Testo non formattato per cui non
-------------	-----------------------	----------------------	----------------------------------

---

<sup>46</sup> Un parametro aggiuntivo, indicato insieme al valore */plain* e separato da questo dal carattere “;” è *charset*, ovvero il set di caratteri adottato per comporre il messaggio. Un set di caratteri è sostanzialmente

			è richiesto <i>software</i> speciale per tradurre il corpo del messaggio
		<b>/richtext</b>	Testo formattato secondo le specifiche riportate nell' <i>RFC 1341</i> <sup>47</sup> e compatibile con il linguaggio <i>SGML</i> <sup>48</sup> . È richiesto <i>software</i> speciale per tradurre il corpo del messaggio. Tale <i>software</i> può essere un <i>SGML editor/viewer</i> <sup>49</sup> .
		<b>/enriched</b>	Testo formattato compatibile con le specifiche riportate nell' <i>RFC 1563</i> <sup>50</sup> . La formattazione <i>enriched</i> è un raffinamento di quella <i>richtext</i> specificata sopra.
		<b>/html</b>	Testo formattato compatibile con le specifiche riportate nell' <i>RFC 1866</i> <sup>51</sup> . È richiesto software speciale per tradurre il corpo del messaggio.
<b>multipart</b> <sup>52</sup>	Nel corpo del messaggio si	<b>/mixed</b>	Ciascuna parte (ordinata) del corpo del messaggio è

una tabella in cui, per ogni carattere è riportato l'equivalente codice binario o esadecimale. La composizione e interpretazione del testo del messaggio avviene ad opera del *Local mail user Agent* che deve avere a disposizione, nel sistema, la tabella codice sono le seguenti:

US-ASCII (ASCII); ISO-88591 (Latina: Europa occidentale); ISO-88592 (Latina: Europa centrale/est). Per una lista esaustiva riferirsi alla pag. 28 dell'*RFC 1521*. Per la tabella ISO-88591 vedere Appendice B.

<sup>47</sup> L'*RFC 1341* decreta lo standard *MIME*. È stato sostituito dall'*RFC 1521*. Ma la parte che riguarda il testo *richtext* è ancora valida.

<sup>48</sup> *SGML* è l'acronimo di *Standard Generalized Mark-up Language*. Un documento in linguaggio *SGML* contiene due tipi d'informazione: la struttura ed il contenuto. La struttura è determinata da particolari istruzioni sempre in formato testo (*Murk-up tags*) racchiuse tra parentesi angolari . Il contenuto è rappresentato dal testo racchiuso tra i *Murk-up tags*. Per struttura si intendono le parti logiche di un documento (Capitoli, sezioni, paragrafi, ecc.) [[http://www.techapps.co.uk/iihb\\_sgml.html](http://www.techapps.co.uk/iihb_sgml.html)]

<sup>49</sup> Sono programmi per comporre o semplicemente visualizzare documenti scritti in *SGML*. Nell'*RFC 1341* (sostituito dall'*RFC 1521*) (Appendice D) viene proposto un semplice programma in 44 linee per tradurre un testo in "richtext" in "plain text".

<sup>50</sup> L'*RFC 1563* (*The text/enriched MIME Content-type*) specifica la sintassi del testo formattato "enriched".

<sup>51</sup> L'*RFC 1866* definisce la sintassi dell'*HTML* (*Hypertext Markup Language*, simile all'*SGML* di cui sopra) nonché il valore "/html" del campo "Content-Type".

<sup>52</sup> Il corpo del messaggio in questo caso è costituito da diverse parti ciascuna delimitata da precisi confini (*boundary*) espressi sotto forma di caratteri alfanumerici identificativi indicati anche nell'intestazione del messaggio nel campo Content-Type. La voce *boundary*= "caratteri alfanumerici", infatti, segue con il carattere ";" la dicitura *multipart* . Ciascuna parte del corpo del messaggio - incapsulata tra il valore *boundary* specificato nell'intestazione e preceduto dai caratteri "--" (codice *ASCII* decimale 45) - contiene a sua volta una intestazione ed un corpo separati da una riga vuota, in maniera simile a quella specificata nell'*RFC 822* e nell'*RFC 1521*. Il valore del *boundary* preceduto e terminato dai caratteri "--" indica la fine dell'ultima parte in cui è diviso il corpo del messaggio. Il testo contenuto tra i *boundary* può terminare con una linea vuota (in tal caso trattasi di testo *esplicito* in formato *ASCII*. Ciò viene indicato anche nell'intestazione del *boundary* stesso) oppure no (in questo caso trattasi di testo *implicito* in formato *ASCII* e non è necessario indicarlo nell'intestazione del *boundary*). [Vedere Appendice A per un esempio]

	possono trovare diversi tipi di dati - testuali e non - combinati assieme.		indipendente e deve essere interpretata secondo un preciso ordine. Ordine dato dai <i>boundary</i> (vedere nota 52).
		<b>/alternative</b>	Sintatticamente identica al valore <i>/mixed</i> , indica che le parti (ordinate) in cui è diviso il corpo del messaggio sono “alternative” perché trattasi di uno stesso tipo d’informazione in formati intercambiabili. In tal caso si può scegliere, tra i diversi tipi della stessa informazione, quella migliore per le caratteristiche del sistema in uso. L’ultima parte è quella più fedele rispetto all’originale. [Vedere Appendice A per un esempio].
		<b>/digest</b>	Ciascuna parte (ordinata) in cui è diviso il corpo del messaggio è a sua volta un messaggio formattato secondo lo standard decretato dall’ <i>RFC 822</i> . Il valore del <i>Content-Type</i> è automaticamente portato al valore <i>message/rfc822</i> .[Vedere Appendice A per un esempio].
		<b>/parallel</b>	Non ha importanza l’ordine delle parti in cui è diviso il corpo del messaggio, così come avveniva per i valori precedenti: esse vengono rappresentate simultaneamente nel sistema.
<b>message</b>	Indica la incapsulazione di un messaggio: il corpo di un messaggio è esso stesso un messaggio che segue lo standard dell’ <i>RFC 822</i> o parte di esso.	<b>/rfc822</b>	Il messaggio incapsulato segue lo standard decretato dall’ <i>RFC 822</i> . Non sono richiesti i campi “ <i>From</i> ”, “ <i>Subject</i> ” e “ <i>To</i> ” per il messaggio incapsulato.
		<b>/partial</b> <sup>53</sup>	È usato quando il messaggio incapsulato è molto grande. In questo caso il programma di posta incapsula il messaggio spezzandolo in diverse mail che verranno riassemblate dal

<sup>53</sup> Tre parametri sono richiesti nel campo *Content-Type* (i parametri sono indicati successivamente al valore *message/partial* e separati da questo e tra loro dal carattere “;”). Il parametro “*id*” è un identificatore unico usato per riassemblare le parti del messaggio. Il parametro “*number*” è un numero intero che indica la sequenza delle parti del messaggio da assemblare (la numerazione inizia da 1 e non da 0). Il parametro “*total*” che indica il numero totale delle parti in cui è stato scomposto il messaggio. [Vedere Appendice A per un esempio]

			programma che le riceverà.
		<b>/external-body</b> <sup>54</sup>	Il corpo del messaggio non è incluso, bensì esterno, ovvero disponibile altrove sotto forma di file.
<b>application</b>	Le informazioni sono di tipo binario: sono necessari dei programmi esterni per la interpretazione.	<b>/octet-stream</b> <sup>55</sup>	Dati binari non interpretabili: è consigliabile scrivere le informazioni su di un file.
		<b>/postscript</b>	Indica la presenza di un programma <i>postscript</i> .
<b>image</b>	Il corpo del messaggio contiene una immagine	<b>/gif</b> <b>/jpeg</b> <b>/...</b>	Sono i vari formati che si possono trovare. La lista non è esaustiva.
<b>audio</b>	Il corpo del messaggio contiene	<b>/basic</b>	Ad indicare un audio codificato ad 8 bit mono <i>PCM</i> .

<sup>54</sup> Quando il valore del campo “Content-Type: ” è “message/external-body” troviamo dei valori aggiuntivi per alcuni campi atti a descrivere determinate caratteristiche del file esterno che rappresenta il corpo del messaggio. Tali valori, separati tra loro dal carattere “;”, sono i seguenti:

<b>Content-Type:</b> <b>message/external-body;</b>	<b>access-type=</b>	<b>ftp</b>	Il corpo del messaggio è accessibile come file usando il protocollo <i>ftp</i> [RFC 959].
<b>name=</b>	Il nome del file che contiene il corpo del messaggio.	<b>tftp</b>	Il corpo del messaggio è accessibile tramite file usando il protocollo <i>tftp</i> [RFC 783].
<b>site=</b>	Nome del dominio della macchina o delle macchine che contengono il file	<b>anon-ftp</b>	accesso ftp anonimo: non sono richiesti nome e <i>password</i> .
<b>directory=</b>	La directory dalla quale il file è disponibile (campo disponibile solo per l' <i>ftp</i> e l' <i>tftp access-type</i> )	<b>local-file</b>	Il file è accessibile nella macchina locale come file.
<b>mode=</b>	La modalità di ricezione del file (campo disponibile solo per l' <i>ftp</i> e l' <i>tftp access-type</i> ). Riferirsi all’RFC 959 e RFC 783 per le specifiche.	<b>afs</b>	Il file è accessibile tramite l’AFS <i>file system</i>
<b>server=</b>	L’indirizzo <i>email</i> del <i>server</i> che contiene il file (campo disponibile solo per il <i>mail server access-type</i> )	<b>mail-server</b>	Il file è accessibile tramite un <i>mail-server</i> .
<b>subject=</b>	L’oggetto della mail spedita per ottenere i dati (campo disponibile solo per il <i>mail server access-type</i> )	<b>expiration</b>	Data oltre la quale il file potrebbe non essere più disponibile. L’RFC 1123 estende le possibilità dell’RFC 822 in termini di numeri per la data: permette 4 <i>digit</i> per l’anno.
<b>size=</b>	La lunghezza del file		
<b>permission=</b>	Indica se l’utente può accedere al file o solo in lettura ( <i>read</i> ), oppure in lettura/scrittura ( <i>read-write</i> ).		

Nota: alla fine dei valori spiegati appena sopra, trova posto, separata da due linee vuote, l’intestazione del corpo incapsulato nel file. Nel senso che il corpo disponibile nel file ha un’intestazione che viene riportata nel messaggio d’origine. [Vedere Appendice A per un esempio]

<sup>55</sup> Il valore *octet-stream* può avere i seguenti parametri indicati in successione e separati dal carattere “;”:

**type:** tipo o categoria del dato.

**padding:** numero di bit di “riempimento” usati laddove i bit totali dai dati inclusi nel corpo non sono multipli di un byte.

**name:** parametro introdotto già dall’RFC 1341 suggerisce il nome del file quando appunto i dati vengono salvati in un file.

	dell'audio		
<b>video</b>	Il corpo contiene delle immagini in sequenza con audio	<b>/mpeg</b>	I dati sono codificati secondo lo standard <i>mpeg</i>

- **Content-Transfer-Encoding:** indica il tipo di trasformazione che i dati hanno subito per il loro trasporto. All'inizio del paragrafo si è specificato che i programmi, con i quali gli utenti spediscono e ricevono messaggi implementano un sistema di conversione (codifica) standard di tutti i dati non testuali in caratteri stampabili *ASCII* a 7 bit. E a ciò, si è visto, si unisce l'estensione del servizio *SMTP* “*Simple Mail Transfer Protocol*” (protocollo di trasmissione delle *mail*) che consente il trasporto di messaggi *MIME* a 8 bit.

Il campo *Content-Transfer-Encoding* indica quale “mappa” di conversione è stata adottata per convertire i dati originari per renderli “adatti” al loro trasporto.

I valori per il campo *Content-Transfer-Encoding* sono i seguenti:

<b>7bit:</b>	Significa che i dati sono tutti rappresentati in linee lunghe non più di mille caratteri <i>US-ASCII</i> .
<b>8bit:</b>	Significa che i dati sono rappresentati in linee lunghe mille caratteri i quali possono essere anche <i>non-ASCII</i> ;
<b>binary:</b>	Significa che i dati sono rappresentati in linee più lunghe di mille caratteri i quali possono essere anche <i>non-ASCII</i>
<b>base64:</b>	Significa che è stato adottato un algoritmo di codifica per i dati. L'algoritmo è spiegato nell' <i>RFC 1521</i> a pag.21. I dati codificati sono il 33% più lunghi dei dati non codificati.
<b>quoted-printable</b>	Significa che è stata apportata una codifica per cui si sono adottati i soli caratteri stampabili <i>ASCII</i> . La codifica investe gli spazi vuoti e le interruzioni di linea. Ciò assicura che i dati stessi non vengano cambiati durante il loro trasporto. Le regole di questa codifica sono indicate a pag.18 dell' <i>RFC 1521</i> . Si riporta un piccolo esempio esplicativo. Il carattere “=” ( <i>ASCII</i> 61) è rappresentato come “=3D” (corrispondente al suo valore esadecimale).

Nota: uno dei metodi di codifica dei dati (ora poco usato, perché soppiantato dallo standard *MIME*) è lo *UUENCODE*. Lo *UUENCODE* è un metodo di

compressione *UNIX* che converte un file binario in caratteri *ASCII*. Tale conversione è effettuata dal *Local mail user Agent* che provvederà ad inserire il file codificato sotto forma di caratteri *ASCII* nel corpo del messaggio. Il file codificato inizierà con la scritta “*begin 644*” più il nome del file e terminerà con la scritta “*end*”.

- **Content-Description:** (opzionale) In cui viene riportata una breve descrizione testuale del corpo del messaggio sottostante.
- **Content-ID:** (opzionale) Sintatticamente identico al “*Message-ID*” specificato nell’RFC 822, serve ad identificare il corpo del messaggio. Tale identificatore è utile laddove si vuole creare un riferimento ad un altro corpo.

L’RFC 2183 specifica un nuovo campo (opzionale) d’intestazione per i messaggi MIME: il **Content-disposition**. Di seguito si riportano i valori per questo campo ed i loro parametri. I parametri sono separati tra loro dal carattere “;”.

<b>Attachment</b>	La parte del corpo sottostante è posta in allegato, ovvero è separata dal resto del corpo e la sua visione non è immediata, ma soggetta ad alcune azioni proposte dal <i>Local mail user Agent</i> .	<b>filename</b> = <b>creation-date</b> = <b>modification-date</b> = <b>read date</b> = <b>size parameter</b> =	Nome del file suggerito.  Data di creazione del file.  Data in cui il file è stato modificato  Data in cui il file è stato letto l’ultima volta.  Lunghezza approssimativa del file.
<b>Inline</b>	La parte del corpo sottostante può essere visionata insieme al resto del corpo del messaggio. Tale opzione risulta utile quando bisogna includere in un messaggio del testo. Informazioni non testuali verranno interpretate separatamente dal <i>Local mail user Agent</i> .		

## Note

- Alcuni *Local mail user Agent* hanno come opzione quella di inserire un file direttamente nel corpo del messaggio come se fosse del testo. In tal caso viene effettuato un algoritmo di codifica a livello locale che traduce le informazioni binarie in testo. Ciò risulta utile, appunto, per l'accorpamento di *file* di testo.

Il processo è irreversibile.

- L'*RFC 1740*<sup>56</sup> decreta che in genere i documenti di testo *Apple Macintosh* possono essere spediti con tutti i criteri specificati dallo standard *MIME*. I *File* binari *Apple Macintosh* vengono codificati secondo i formati *AppleSingle* o *AppleDouble*. Ciò viene indicato nell'intestazione del messaggio *MacMIME* secondo le specifiche indicate nell'*RFC 1740*.
- L'*RFC 2387*<sup>57</sup> introduce un nuovo valore per il campo “**Content-type:**”. Il **multipart/related**. Esso indica un meccanismo per aggregare le parti in cui il corpo del messaggio è diviso. È usato nelle applicazioni *MIME-PEM* (vedere più sotto) e *MIME-Macintosh*.

I suoi parametri separati dal carattere “;” sono:

**type="dichiarazione"** (obbligatorio) Permette al *Local mail user Agent* di determinare il valore del *Content-type* della parte del corpo sottostante senza riferirsi alla sua intestazione.

**start="informazioni"** (opzionale) Contiene il *Content-ID* della parte del corpo che deve essere processata per prima.

---

<sup>56</sup> L'*RFC 1740 (MIME Encapsulation of Macintosh files – MacMIME)* decreta lo standard per l'incapsulazione dei *file* *Apple Macintosh*

**start-info="informazioni"** (opzionale) Contiene delle stringhe che costituiscono dei parametri (*command line parameters*) necessari per processare le parti del corpo aggregate. Un esempio è un file eseguibile che necessita di alcuni parametri prima della sua esecuzione.

- Esistono alcune applicazioni che servono per assicurare la privacy della posta inviata. Le *MIME-PEM*<sup>58</sup> ed il formato *PGP*<sup>59</sup> dei messaggi, nascono con questo intento: sono infatti dei sistemi di criptazione ed autenticazione delle *mail*.

Le *MIME-PEM* seguono un approccio di criptazione sia simmetrico<sup>60</sup> che asimmetrico, mentre il formato PGP si basa su di un sistema di criptazione asimmetrico<sup>61</sup>.

---

<sup>57</sup> L'*RFC 2387 (The MIME Multipart/Related Content-type)* decreta le specifiche per un nuovo valore del campo “Content-Type:”.

<sup>58</sup> *PEM* sta per *privacy-enhanced mail*. I dettagli del servizio PEM sono contenuti nell'*RFC 1421, 1422, 1423, 1424*.

<sup>59</sup> *PGP* sta per *Pretty Good Privacy*. I dettagli del formato PGP sono contenuti nell'*RFC 1991, 2015 e 1847*.

<sup>60</sup> Con il sistema a chiave simmetrica, un messaggio è crittato e decrittato usando la stessa chiave ; questa deve essere conosciuta sia dal mittente che dal ricevente. La chiave passa dall'uno all'altro, attraverso una transazione separata, ma il sistema è vulnerabile perché essa può venir rubata nel momento in cui attraversa la rete.

<sup>61</sup> PGP utilizza una tecnica di criptazione basata su chiavi pubbliche asimmetriche, che prevede la presenza di due chiavi: una pubblica, che l'autore può mettere a disposizione di chiunque inviandola a un apposito server Web e una privata, conservata solo dall'utente e protetta da un codice di accesso personale. Per trasmettere un messaggio in forma protetta è anzitutto necessario che il mittente scarichi dal server Web la chiave pubblica del destinatario e cripti con questa il documento; il destinatario potrà a questo punto decriptare la comunicazione ricevuta tramite la propria chiave privata, utilizzando il codice di accesso personale. Chiaramente è virtualmente impossibile risalire dalla chiave pubblica a quella privata, poiché gli algoritmi di criptazione utilizzati hanno raggiunto livelli di sicurezza paragonabili a quelli impiegati in campo militare, impedendo qualsiasi tentativo di decifrazione operato tramite sistemi di analisi matematica.

Per firmare elettronicamente un messaggio PGP utilizza invece un metodo inverso a quello impiegato per la criptazione e cioè adopera la chiave privata (ed il relativo codice di accesso) per apporre la firma e quella pubblica per verificare la conformità del messaggio con l'impronta generata durante la sigla.

Un messaggio criptato MIME-PEM lo si riconosce perché il corpo è encapsulato tra le due stringhe

"-----BEGIN PRIVACY-ENHANCED MESSAGE-----" e  
"-----END PRIVACY-ENHANCED MESSAGE-----". Esso, perciò avrà la seguente struttura:

*Pre-Encapsulation Boundary (Pre-EB)* Ovvero il *boundary* (numero di confine) indicati anche nell'intestazione del messaggio

-----BEGIN PRIVACY-ENHANCED MESSAGE-----

Righe d'intestazione in formato testo contenenti informazioni circa la criptazione e la chiave di criptazione

Una riga vuota di separazione  
Testo del messaggio criptato

*Post-Encapsulation Boundary (Post-EB)*

-----END PRIVACY-ENHANCED MESSAGE-----

Un messaggio criptato PGP è più complesso. L'*RFC 2015* specifica la struttura di un messaggio criptato PGP. In particolare il campo d'intestazione **Content-Type**: si arricchisce di due nuovi valori con relativi parametri. I parametri sono separati tra loro dal carattere “;”.

Essi sono:

<b>multipart /encrypted</b> <sup>62</sup>	Indica che il messaggio è stato criptato	<b>protocol=</b> <b>"application/pgp-encrypted"</b>	Protocollo usato per la criptazione
<b>multipart /signed</b> <sup>63</sup>	Indica un messaggio firmato	<b>protocol=</b> <b>"application/pgp-signature"</b>	Protocollo usato per la firma
<b>application</b>	Usato solo per la distribuzione		

<sup>62</sup> Un messaggio criptato è formato da due parti: la prima costituita da una informazione di controllo indicante la versione dell' *application/pgp-encrypted*, la seconda contenente i dati criptati avente “*application/octet-stream*” come valore del *Content-Type* nell'intestazione. Ogni parte è ovviamente racchiusa tra i *boundary*. [Vedere l'appendice A per un esempio]

<sup>63</sup> Un messaggio firmato è costituito da due parti: la prima contenente i dati in formato MIME, la seconda contenente la firma con “ *application/pgp-signature*” come valore del campo *Content-Type*. [Vedere l'appendice A per un esempio]

/pgp-keys	della chiave pubblica.		
-----------	------------------------	--	--

Bisogna notare che esiste anche un metodo combinato in cui in una unica operazione il messaggio viene firmato e criptato. In questo caso si incapsula un messaggio nell'altro [si rimanda all'appendice A per un esempio].

Comunque tutte queste operazioni sono trasparenti all'utente perché affidate interamente al *Local mail user Agent*.

Quello che l'utente può incontrare è un testo che inizia con la scritta “-----*BEGIN PGP MESSAGE*-----” oppure “-----*BEGIN PGP SIGNED MESSAGE*-----” in cui può incontrare un insieme di caratteri incomprensibili racchiusi tra le scritte “-----*BEGIN PGP SIGNATURE*-----” e “-----*END PGP SIGNATURE*-----” oppure ancora una volta “-----*BEGIN PGP MESSAGE*-----” e “-----*END PGP MESSAGE*-----” che rappresentano la firma PGP del messaggio.

## 2.5. I *Newsreader*

Nel paragrafo 1.3 è stato specificato che i messaggi *Usenet* devono essere formattati come messaggi di posta elettronica conformemente con quanto specificato con l'*RFC 822*.

In generale, quindi, è possibile scrivere un messaggio *Usenet* con un *tool* per comporre messaggi di posta Internet, detto anche *Local mail user Agent* (programma demandato alla gestione della posta che lavora nella macchina locale, ovvero nel Pc dell'utente).

Esistono però dei programmi sviluppati appositamente per gestire i messaggi *Usenet*. Essi vengono denominati *Newsreader* (programmi di lettura di notizie).

Tali programmi nascondono all'utente il modo in cui i gruppi di notizie ed i messaggi *Usenet* sono effettivamente memorizzati, facilitando la selezione, la lettura e la risposta ai messaggi. Essi sono in grado infatti di leggere e interpretare l'intestazione dei messaggi così da poter rappresentare le informazioni in maniera corretta.

Sono disponibili molti programmi di lettura di notizie per vari tipi di computer per sfruttare le varie interfacce utente di tipo grafico e le altre caratteristiche specifiche del computer. Molti di questi programmi sono disponibili gratuitamente, altri vengono commercializzati. Il canale di diffusione principale è Internet.

I *Newsreader* hanno le seguenti caratteristiche comuni:

- **Registrazioni delle catene dei messaggi.** I programmi di lettura delle notizie di *Usenet* prevedono il concatenamento (vedere pag. 56), ovvero possono esaminare

ordinare, selezionare e presentare all'interno di un gruppo articoli suddivisi in catene di argomenti. Questo facilita il reperimento degli articoli desiderati, consentendo di saltare quelli che non interessano.

- **File di esclusione.** Si tratta di file (anche detti filtri) creati e gestiti dal programma di lettura delle notizie, per indicare un elenco di persone e di argomenti a proposito dei quali l'utente non vuole leggere affissioni. Alcuni programmi di lettura di notizie consentono anche di specificare il periodo di tempo, durante il quale un comando di esclusione deve rimanere attivo.
- **Formattazione dei messaggi di risposta.** Nel creare commenti ad un messaggio precedente, sovente può essere utile inserire parte di tale messaggio (*quoting*<sup>64</sup>), contrassegnandolo in qualche modo per indicare che sono parole altrui e specificando la fonte. Questa funzione consente di inserire il messaggio desiderato, solitamente rientrato e con un carattere quale “>” oppure “|” sul margine sinistro di ciascuna riga, anteponendo una riga al messaggio stesso per indicare l'autore e la data d'affissione.

---

<sup>64</sup> Dall'inglese *to quote*: citare

### III. Il caso di studio: il lessico dei Newsgroups di argomento religioso

#### 3.1. Introduzione all'analisi

Il materiale raccolto comprende i messaggi inviati, in un determinato periodo temporale, sui *Newsgroups* (Vedi il capitolo su *Usenet*) di argomento religioso.

Nella Tabella 3 si riportano alcuni elementi utili per identificare i files dai quali ha tratto origine l'analisi.

**Tabella 3:** Quadro degli elementi identificativi dei files oggetto di studio

Newsgroup	Riferim. Temporale	Lunghezza File di testo originale
<i>it.cultura.cattolica</i> Server: <a href="http://www.caspur.it/servizi/usenet/news/it.cultura.html">http://www.caspur.it/servizi/usenet/news/it.cultura.html</a> news://news.caspur.it:it.cultura.cattolica/ Newsreader: <b>Netscape Communicator per Windows®</b> Parametri di salvataggio: <b>Seleziona tutto/salva</b>	17/09/99 al 05/10/99	2.356kB (2,29MB)
<i>it.cultura.ebraica</i> Server: <a href="http://www.caspur.it/servizi/usenet/news/it.cultura.html">http://www.caspur.it/servizi/usenet/news/it.cultura.html</a> news://news.caspur.it:it.cultura.ebraica/ Newsreader: <b>Netscape Communicator per Windows®</b> Parametri di salvataggio: <b>Seleziona tutto/salva</b>	18/09/99 al 06/10/99	152kB
<i>it.cultura.religioni</i> Server: <a href="http://www.caspur.it/servizi/usenet/news/it.cultura.html">http://www.caspur.it/servizi/usenet/news/it.cultura.html</a> news://news.caspur.it:it.cultura.religioni/ Newsreader: <b>Netscape Communicator per Windows®</b> Parametri di salvataggio: <b>Seleziona tutto/salva</b>	18/09/99 al 06/10/99	555kB
<i>it.cultura.ateismo</i> Server: <a href="http://www.caspur.it/servizi/usenet/news/it.cultura.html">http://www.caspur.it/servizi/usenet/news/it.cultura.html</a> news://news.caspur.it:it.cultura.ateismo/ Newsreader: <b>Netscape Communicator per Windows®</b> Parametri di salvataggio: <b>Seleziona tutto/salva</b>	18/09/99 al 06/10/99	2.269kB (2,21MB)
	<b>TOTALE</b>	<b>5,20MB</b>

Il primo elemento che salta subito all'occhio è la quantità di dati a disposizione: altri studi e ricerche - simili a questa - effettuate nello stesso campo, con gli stessi obiettivi, hanno trattato, infatti, dati la cui estensione, per il file di testo, era al di sotto del Mega byte. Per dare un'idea della corrispondenza tra testo e byte, basti pensare che il file in oggetto, di 5,20MB, contiene un numero di parole che si aggira intorno alle 750.000 occorrenze, mentre un antecedente studio su di un file di circa 600KB conteneva al suo interno poco più di 90.000 parole.

Oltre a ciò è importante considerare anche un ulteriore dato che dà una misura circa i tempi di elaborazione per l'ottenimento degli elementi necessari alla valutazione dei risultati: su di una macchina di potenza medio alta (Computer Pentium® 400 MHz, 128MB di Ram, dischi SCSI da 7200 Rpm) l'applicativo Spad-T® per Dos, su di un file di testo di poco meno di 5MB, ha impiegato complessivamente all'incirca 2 ore cumulative per terminare correttamente tutte le elaborazioni<sup>65</sup>.

L'estensione del file di dati, il periodo temporale in cui tali dati sono stati raccolti (di soli 19 giorni), i tempi e le risorse stimate per l'elaborazione hanno suggerito quindi l'uso, per tale ricerca, di una metodologia sostanzialmente diversa da quella adottata fino ad ora dagli altri studi con i quali, però, essa condivide lo stesso scopo.

Se è pur vero infatti che - in questa sede - l'obiettivo finale è quello di individuare, attraverso l'uso di tecniche statistiche, il *significato* di un insieme ordinato di parole (*sintagma*. Vedi pag.14) per evidenziare il contenuto di un *corpus*, i mezzi adottati per raggiungerlo sono però orientati all'ottenimento di una maggiore "velocità operativa", cercando, nel contempo, di non contravvenire ad alcune caratteristiche fondamentali dei

---

<sup>65</sup> Lo Spad-T® fa molto uso della memoria di massa.

metodi di ricerca stessi quali la *quantificazione*, la *sistematicità*, l'*esaustività* e l'*oggettività*.

Questo lavoro propone quindi una metodologia che, a fronte di una perdita d'informazione, si propone di trattare dati molto ampi in tempi ridotti con un risultato senza dubbio più “sporco”, ma ritenuto comunque accettabile per determinati fini.

Si potrebbe ad esempio pensare di estendere, per i *Newsgroups* o per altri *forum* di discussione, il periodo di rilevazione ad un mese, a tre mesi e, perché no, ad un anno e trattare files di decine di Mega byte per stabilire molto più velocemente, con un ridotto dispendio di risorse, i contenuti dei discorsi o le tematiche toccate dai parlanti durante tale periodo.

### 3.2. L'automatizzazione

Automatizzare determinate operazioni attraverso l'uso del computer rappresenta, senza dubbio, un mezzo efficace per snellire e velocizzare alcune fasi dell'analisi.

In un processo di ricerca, però, l'automatizzazione non è sempre sinonimo di *precisione*. Questo perché uno strumento automatizzato, tarato in base ad uno *standard*, può non fornire dei risultati accurati ed attendibili se applicato ad un *set* di dati che non rispetta tale *standard*.

Un mezzo per ovviare a questo inconveniente potrebbe essere quello di individuare delle regolarità che sono alla base del fenomeno oggetto di studio.

Il fenomeno *Newsgroups*, per come è stato definito nei primi paragrafi del capitolo *USENET* (Vedi pag.52) individua un gruppo di persone che, con regole stabilite, colloquiano tra di loro, su determinati temi, scambiandosi messaggi in forma elettronica attraverso una rete di computer (*Internet*). Tali regole riguardano non solo la *netiquettes* (insieme di norme di buona educazione e di comportamento in rete), ma anche il formato che i messaggi devono avere per conformarsi ai *protocolli* di comunicazione di *Internet* e poter essere così instradati.

Ebbene, l'individuazione delle regole che stabiliscono il formato elettronico dei messaggi, costituiscono la base di partenza per una **prima fase** del processo di automatizzazione in cui vengono selezionati gli elementi pertinenti all'analisi.

Vi sono numerosissimi documenti ufficiali disponibili in rete (gli *RFC* Vedi pag.56 e *Bibliografia*) che definiscono gli *standard* e gli altri aspetti di *Internet*.

La lettura e la traduzione di alcuni di questi documenti (tutto il materiale accuratamente tradotto è riportato nel capitolo dedicato a *USENET*) ha permesso di estrapolare gli elementi necessari per la messa a punto degli strumenti di indagine.

Ogni messaggio, infatti, scambiato in Internet, è costituito da una *intestazione* seguita da un *corpo* (Vedi pag.56 per una accurata descrizione). L'*intestazione* contiene le istruzioni in formato testo che servono più che altro al *server* e al *newsreader-client* (Vedi pag.73) per poter individuare instradare e formattare il messaggio, mentre il *corpo* è rappresentato dal contenuto vero e proprio del messaggio stesso.

Qui di seguito si riporta un esempio di un messaggio con allegato.

```
Received: from net133-132.mclink.it (net133-132.mclink.it
[195.110.133.132])
  by mail1.mclink.it (8.9.1/8.9.0) with SMTP id QAA07050
  for <nome@mclink.it>; Thu, 13 Apr 2000 16:53:52 +0200 (CEST)
From: Alessandro Stabellini <nome@mclink.it>
To: Alessandro Stabellini<nome@mclink.it>
Subject: Messaggio di esempio
Date: Thu, 13 Apr 2000 16:51:21 +0200
Reply-To: alfader@mclink.it
Message-ID: <thnbfsg7ftra03mbcgqoio3mct5a77719@4ax.com>
X-Mailer: Forte Agent 1.7/32.534
MIME-Version: 1.0
Content-Type: multipart/mixed; boundary="-
=_9nnbfsc9pj6ifjnnuitlqih5tcmnivs18a.MFSBCHJLHS"
-----_9nnbfsc9pj6ifjnnuitlqih5tcmnivs18a.MFSBCHJLHS
Content-Type: text/plain; charset=us-ascii
Content-Transfer-Encoding: quoted-printable

Questo e' il corpo del messaggio
in cui viene riportato il testo dello scrivente
-----_9nnbfsc9pj6ifjnnuitlqih5tcmnivs18a.MFSBCHJLHS
Content-Type: application/octet-stream; name=daily.zip
Content-Transfer-Encoding: base64
Content-Disposition: attachment; filename=daily.zip

UEsDBBQAAAIAAAAYiyj96t3ftwAAAEgBAAAHAAAAYXZwLnNldFXP3QrcIBTA8fvB3qH7I
HSuLeiq
[...]
```

**Intestazione**

**Intestazione**

**Intestazione con allegato**

Scartare a priori le *intestazioni* e considerare il solo *corpo* come oggetto di studio potrebbe essere un criterio di selezione automatizzato, rapido, ma "pericoloso" per diversi aspetti.

Un primo aspetto riguarda la perdita d'informazione: l'intestazione contiene infatti, oltre allo "oggetto", dati circa la provenienza del messaggio e l'identità dello scrivente. Tali dati potrebbero essere importanti laddove si volessero - ad esempio - incrociare i risultati dell'elaborazione con gli autori dei contenuti in analisi.

Un secondo aspetto riguarda l'operazione di individuazione e discriminazione del *corpo* del messaggio e della sua *intestazione*. Come è già stato premesso, tra i compiti del *newsreader-client* c'è quello di interpretare i contenuti dell'*intestazione* sotto forma di istruzioni al fine di restituire all'utente un messaggio più fruibile in termini di lettura. Bene, il *newsreader* potrebbe non interpretare correttamente l'intestazione riportando alcuni errori (è il caso ad esempio della erronea codifica del set di caratteri ISO 88591 [Vedi Nota 46 e Appendice B] per cui le accentate vengono visualizzate sotto forma dei corrispondenti valori esadecimali), o addirittura, l'utente stesso, potrebbe essere non capace di utilizzare correttamente gli strumenti messi a disposizione dal *newsreader* (è il caso ad esempio degli allegati codificati, inseriti direttamente nel corpo del messaggio sotto forma di caratteri incomprensibili [Vedi pag.68]).

In questi ultimi casi eliminare l'*intestazione* significherebbe, quindi, privarsi della possibilità di individuare, di correggere e di controllare alcune variabili ed errori.

Le considerazioni appena fatte hanno suggerito determinate scelte di tipo operativo: si è infatti deciso - in un primo momento - di scartare gli allegati binari e mantenere, unite

al *corpo* del messaggio, alcune righe d'*intestazione* che potessero apportare informazioni aggiuntive in sede di analisi. Tali righe sono :

- **Subject:**
- **Date:**
- **From:**
- **Organization:**
- **Newsgroups:**

Per una descrizione dei campi si rimanda a pag.56.

E ciò è avvenuto in automatico sfruttando la possibilità che la maggior parte dei *newsreader* offrono: il salvataggio su supporto elettronico, di determinate parti ben precise e selezionabili di un insieme di messaggi.

Ecco come compare uno dei messaggi, tra quelli considerati per l'analisi, dopo il salvataggio automatizzato:

Subject: R: "Qui il Tibet soffre ancora"  
Date: Mon, 27 Sep 1999 18:02:01 GMT  
From: "Lev" <Lev@streghe.com>  
Organization: [Infostrada]  
Newsgroups: it.cultura.newage,it.cultura.religioni

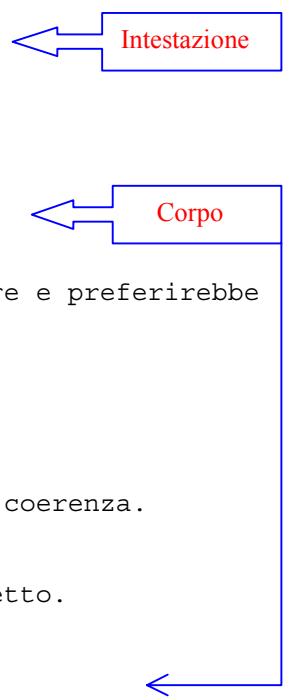
Mandala <mandhala@freemail.it> wrote in message  
F5TE3.15942\$E3.107581@typhoon.libero.it...  
>  
>Qualunque tibetano inorridirebbe a sentirti parlare e preferirebbe  
morire  
al grido di  
> W il Dalai Lama.

Non si preoccupino, moriranno tutti, in silenzio.

> Mi dispiace che tu non comprenda il valore della coerenza.

La coerenza e' per i VIVI.  
Loro sono gia' morti, solo che nessuno glielo ha detto.  
[...]

Lev.



A questo punto si è proceduto con la **seconda fase** del processo di automatizzazione. Questa è stata la fase più complicata da operativizzare. Lo scopo, infatti, è sempre quello di individuare gli elementi pertinenti all'analisi attraverso strumenti di selezione il più possibile precisi ed oggettivi, con la difficoltà maggiore qui, però, dell'individuazione di regole ben precise alle quali riferirsi, così come era avvenuto nella prima fase dell'automatizzazione.

Abbiamo visto che ciascun messaggio contenuto nei 4 files di testo oggetto di studio (Vedi Tabella 3), dopo la prima fase di selezione automatizzata, risulta costituito da alcune righe d'intestazione che ne stabiliscono provenienza ed "oggetto" e da un *corpo* che esprime il *discorso* dell'autore. Ed è su quest'ultimo che, nella seconda fase del processo di automatizzazione, si è soffermata l'attenzione.

Il *corpo* di un messaggio è rappresentato, di fatto, da uno scritto contenente un linguaggio molto vicino al parlato sia per la struttura del mezzo di comunicazione (*Internet*), che per la modalità intrinseca del tipo di colloquio tenuto con gli altri "avventori" (botta e risposta [Vedi pag. 74]). Ciò fa sì che molte regole sintattico-grammaticali del linguaggio scritto vengano, nei messaggi *USENET*, stravolte. Non solo. Nel paragrafo 1.2. si è messa in evidenza proprio la difficoltà di cogliere il significato che un singolo emittitore dà ad un insieme di parole e l'impossibilità di attribuire un linguaggio ad una situazione/condizione particolare a tal punto da non poter essere poi generalizzato ad altri casi ad esso comparabili.

Tutto questo ha suggerito una linea operativa basata sostanzialmente sulla ricerca di regole che stabiliscono la composizione di tutti quegli elementi, contenuti nel *corpo* del messaggio, che possano essere rintracciati in tutti i messaggi inviati nei *Newsgroups* e,

allo stesso tempo, siano superflui ai fini dell'analisi. Tali elementi sono rappresentati dagli indirizzi di posta e di Internet, dalle firme digitali (Vedi pag.70) e da quant'altro (per una descrizione completa vedere Tabella 4) identificabile da una ben precisa struttura di forma all'interno del file di testo (Vedi pag. 90).

Ovviamente la procedura di individuazione di eliminazione e/o sostituzione di questi elementi nei files non poteva avvenire manualmente: ciò avrebbe vanificato il tentativo di ridurre i tempi di elaborazione.

Per l'occasione è stato allora scritto, compilato e sviluppato un software che sotto forma di tool applicativo permettesse in automatico quanto appena descritto.

Si rimanda al capitolo successivo per una sua esauriente trattazione.

### 3.3. Il software sviluppato per l'automatizzazione

Il programma, scritto in Delphi<sup>66</sup>, implementa una libreria - sviluppata e distribuita in licenza *Freeware*<sup>67</sup> dal russo Andrey Sorokin - che consente di sfruttare, per la ricerca del testo in un file, le potenzialità delle espressioni regolari del Perl<sup>68</sup>. Windows<sup>®</sup> è la piattaforma su cui gira il programma.

Le principali sue caratteristiche sono:

- possibilità di usare le espressioni regolari del Perl per la ricerca/sostituzione del testo [Vedi nota 68];
- possibilità di operare su più files di testo contemporaneamente;
- possibilità di salvare i parametri di lavoro in un database editabile e richiamabile alla bisogna;
- possibilità di operare passo passo su uno o più files di testo;
- possibilità di eliminare le righe che contengono o iniziano con una determinata stringa;
- possibilità, su uno o più files contemporaneamente, di rimuovere o sostituire il testo che soddisfa determinate condizioni.

---

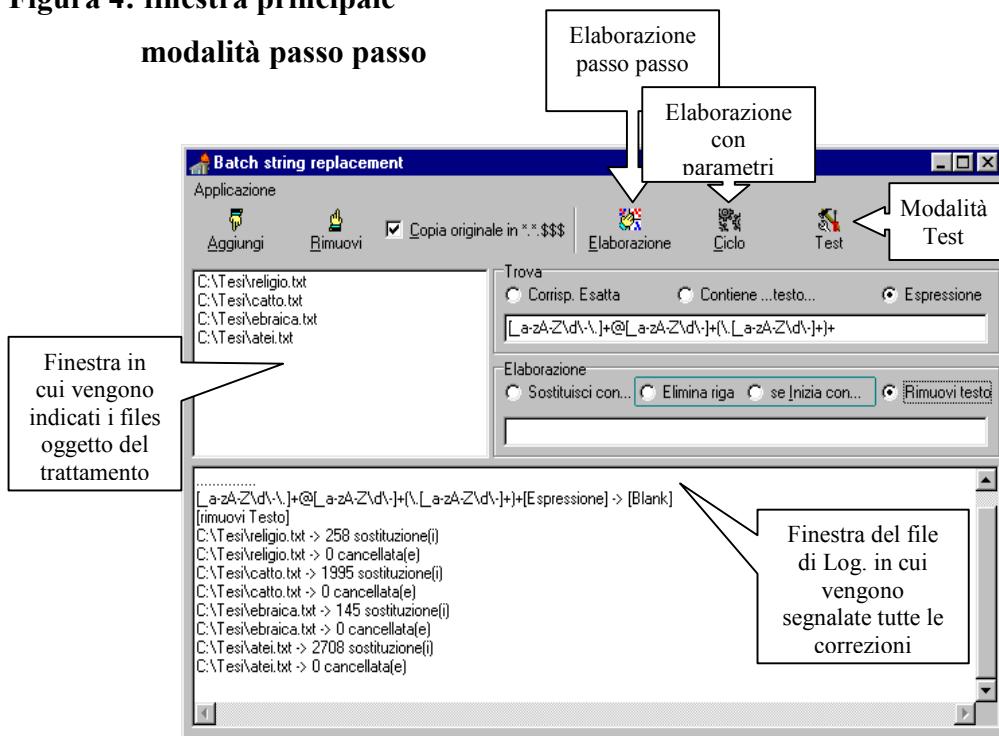
<sup>66</sup> Della Borland/Inprise<sup>®</sup>, è un applicativo per piattaforma Windows<sup>®</sup> che offre un ambiente di sviluppo software basato su di un linguaggio ad “oggetti”. Il codice generato rispecchia i costrutti del Pascal.

<sup>67</sup> La licenza *freeware* consente di usare liberamente un prodotto da essa protetto.

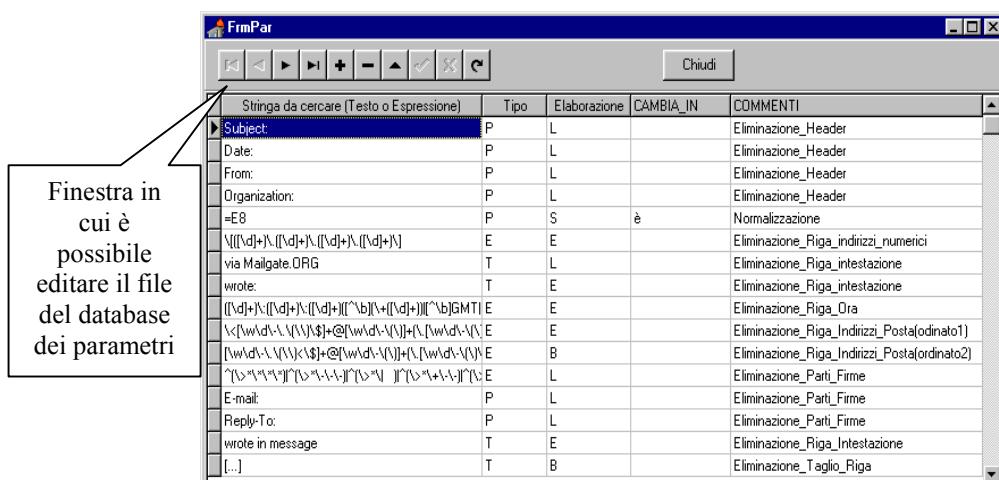
<sup>68</sup> Il Perl (*Practical Extraction Report Language*) è un linguaggio di programmazione di uso generale che al momento della sua nascita venne orientato all’elaborazione dei testi. Un punto di forza del Perl è costituito dalle espressioni regolari: estremamente potenti ed articolate. Si tratta di particolari costrutti che servono per analizzare le stringhe. In sostanza una espressione regolare è una sequenza di caratteri - che segue determinate regole - di solito usata per cercare le occorrenze di una o più stringhe in un file [Vedi anche appendice E].

In Figura 4 e Figura 5 si riporta lo *screenshot* del programma da cui, con un colpo d'occhio, si possono cogliere le sue caratteristiche e funzionalità.

**Figura 4: finestra principale  
modalità passo passo**



**Figura 5 : finestra dei parametri  
modalità ciclo**



Come si nota dalle sue caratteristiche, il programma è stato pensato e sviluppato all'insegna della massima versatilità e potenza. Ciò si è tradotto nella possibilità di disporre di uno strumento che potesse essere applicato, con la minima difficoltà, ai dati testuali oggetto di studio. Non solo: il file di parametri usato per l'elaborazione, una volta creato, può essere applicato successivamente ad un altro *set* di dati simili o addirittura modificato e adattato a differenti situazioni o reso più efficace (*preciso*) attraverso un processo di "raffinamento" dei parametri stessi.

Con ciò si ritiene di aver colmato, almeno in parte, una lacuna: l'eccessiva dipendenza dell'automatizzazione dagli elementi utilizzati per la messa a punto delle sue fasi.

### **3.4. Il parametri del software applicativo e la normalizzazione**

Si riportano in Tabella 4 tutte le sostituzioni/eliminazioni effettuate contemporaneamente su tutti e 4 i files di testo oggetto di studio.

**Nota.** Dal momento che il software a disposizione, descritto nel paragrafo precedente, consente operazioni di sostituzione/eliminazione di stringhe di testo di qualsiasi tipo si è ritenuto opportuno inserire, nel file di parametri, anche una serie d'istruzioni volte alla *normalizzazione* del testo (Vedi paragrafo 1.7. e Appendice D).

**Tabella 4:** Elenco dei parametri di sostituzione/eliminazione adottati nel software di automatizzazione

N.	TROVA	COMMENTO	AZIONE
1	Spazio inizio riga	Rimuovi lo spazio ad inizio riga con l'Espressione: ^\x20 (Normalizzazione)	Rimuovi testo
2	"	Carattere ANSI Windows 147 (Vedi Appendice B) (Normalizzazione)	Sost. con → <<
3	"	Carattere ANSI Windows 148 (Vedi Appendice B) (Normalizzazione)	Sost. con → >>
4	<<	Carattere ANSI Windows 171 (Vedi Appendice B) (Normalizzazione)	Sost. con → <<
5	>>	Carattere ANSI Windows 187 (Vedi Appendice B) (Normalizzazione)	Sost. con → >>
6	'	Carattere ANSI Windows 146 (Vedi Appendice B) (Normalizzazione)	Sost. con → '
7	~	Carattere ANSI Windows 96 (Vedi Appendice B) (Normalizzazione)	Sost. con → `
8	'	Carattere ANSI Windows 145 (Vedi Appendice B) (Normalizzazione)	Sost. con → '
9	...	Carattere ANSI Windows 133 (Vedi Appendice B) (Normalizzazione)	Sost. con → ...
10	•	Carattere ANSI Windows 149 (Vedi Appendice B) (Normalizzazione)	Sost. con → -
11	-	Carattere ANSI Windows 150 (Vedi Appendice B) (Normalizzazione)	Sost. con → -
12	§	Carattere ANSI Windows 167 (Vedi Appendice B) (Normalizzazione)	Rimuovi testo
13	.	Carattere ANSI Windows 183 (Vedi Appendice B) (Normalizzazione)	Rimuovi testo
14	É	Carattere ANSI Windows 201 (Vedi Appendice B) (Normalizzazione)	Sost. con → È
14	é		Sost. con → è
16	þ	Carattere ANSI Windows 254 (Vedi Appendice B) (Normalizzazione)	Sost. con → )
17	o'	Tale sostituzione potrebbe essere non corretta (parole tra virgolette trasformate in parole con lettere accentate) (Normalizzazione)	Sost. con → ò
18	a'	" " "	Sost. con → à
19	e'	" " "	Sost. con → è
20	u'	" " "	Sost. con → ù
21	i'	" " "	Sost. con → ì
22	o'	" " "	Sost. con → Ò
23	A'	" " "	Sost. con → À
24	E'	" " "	Sost. con → È
25	U'	" " "	Sost. con → Ù
26	I'	" " "	Sost. con → Ì
27	Subject:	Eliminazione Header	Elimina riga se all'inizio
28	Date:	Eliminazione Header	Elimina riga se all'inizio
29	From:	Eliminazione Header	Elimina riga se all'inizio
30	Organization:	Eliminazione Header	Elimina riga se all'inizio
31	=E8	Tabella codici Iso-88591 (Vedi Appendice B)	Sost. con → è

32	[nnn.nnn.nnn.nnn]	Trova gli indirizzi DNS del tipo [192.111.11.1]. Usa l'espressione \([(\d{1,3})\.](\d{1,3})\.](\d{1,3})\.](\d{1,3})\]	Elimina riga
33	via Mailgate.ORG	Eliminazione riga d'intestazione	Elimina riga se all'inizio
34	wrote:	Eliminazione riga d'intestazione	Elimina riga
35	ore:minuti:secondi + offset GMT/CET	Eliminazione dell'ora d'invio del messaggio. Usa l'espressione (\d{1,2}:\d{1,2}:\d{1,2})\:(\d{1,2}):\d{1,2}(\d{1,2}:\d{1,2}:\d{1,2})[^\s]GMT [\^\s]CET [\^\s](-(\d{1,2}:\d{1,2}:\d{1,2}))	Elimina riga contenente testo
36	<rossi@provider.it>	Eliminazione riga indirizzi di posta Internet del tipo <aldo@iol.it>. Usa l'espressione \<[\w\d\-\.\(\)]\\$+\@[\w\d\-\.\(\)]+\([\.\w\d\-\.\(\)]+\)>	Elimina riga contenente testo (ordinato1)
37	rossi@provider.it	Eliminazione indirizzi di posta Internet del tipo <aldo@iol.it>. Usa l'espressione [\w\d\-\.\(\)]<\\$]+\@[\w\d\-\.\(\)]+\([\.\w\d\-\.\(\)]+\)>+	Rimuovi testo (ordinato2)
38	*** Firma ... et similia	Eliminazione di diverse tipologie di firme. Usa l'espressione ^(>*/*/* ^\(>*-\ - -\) ^\(>*\ \ ) ^\(>*+\ - -\) ^\(>*\#\#\#) ^\(>*\ \~\ \~\) ^\(>*= = = )=	Elimina riga se all'inizio
39	E-mail:	Elimina riga di indirizzi di posta	Elimina riga se all'inizio
40	Replay-To:	Eliminazione Header	Elimina riga se all'inizio
41	wrote in message	Eliminazione riga con determinati caratteri	Elimina riga contenente testo
42	[...]		Rimuovi testo
43	lsci		Rimuovi testo
44	jno		Rimuovi testo
45	[snip]		Rimuovi testo
46	☺ ☺...	Eliminazione Emoticons. Usa l'espressione (:\-[\\\\ \])(DP9\+)(:\-[\\\\ \])(DP9\+)(8\-[\\\ \])(DP9\+)(^\[\-\_]+\^)	Rimuovi testo
47	it.comp.hardware	Eliminazione riferimento ad altri Newsgroups. Usa l'espressione \<it\.\w\+\.\.\?\w\+\>	Rimuovi testo
48	--	Da effettuare dopo le espressioni	Elimina riga se all'inizio
49	**	Da effettuare dopo le espressioni	Rimuovi testo
50	>__ Firma ...	Eliminazione firma. Usa l'espressione ^(\*\_\_\_)	Elimina riga se all'inizio
51	****	Ordinato dopo le espressioni	Elimina riga se all'inizio
52	http:\URL et similia	Eliminazione indirizzi Internet. Usa l'espressione (( \<( ?([Ff][Tt][Pp][Hh][Tt][Pp])) ://([\w\d\-\-]+([\w\d\-\>])+)+)(([\w\d\-\ \.\~\>])+)+)*?/	Rimuovi testo (ordinato 1)
53	www.indirizzoInternet	Eliminazione indirizzi internet. Usa l'espressione (( \<( ?([Ww][Ww][Ww])) .\([\w\d\-\-]+([\w\d\-\>])+)+)(([\w\d\-\ \.\~\>])+)+)*?/	Rimuovi tetto (ordinato 1)
54	doppio spazio		Sost. con → uno spazio
55	Newsgroups:	Sostituisci il testo Newsgroups: ad inizio riga. Usa l'espressione ^Newsgroups:	Sost. con → \$\$\$
56	****		Elimina riga se all'inizio
57	=====		Elimina riga se all'inizio

Si notino le eliminazioni con il numero progressivo dal 27 al 30. Con esse, "scartando" i campi d'*intestazione*, selezionati nella prima fase dell'automatizzazione, si è deciso di perdere le informazioni circa l'identità dell'autore, l'oggetto e la provenienza del messaggio, rinunciando, di fatto, alla possibilità di eventuali incroci con questi dati. La sostituzione N°55 è stata, invece, operata allo scopo di stabilire un marcatore di inizio/fine messaggio utile anche per il riconoscimento di eventuali firme all'interno del testo (vedi più avanti).

Le righe con il numero 32,35,36,37,38,46,47,50,52,53,55 contraddistinguono delle operazioni effettuate usando le espressioni regolari del Perl. Per costruire tali espressioni è stato necessario uno studio preliminare che ha consentito di identificare la struttura (forma) degli elementi da ricercare (indirizzi Internet, indirizzi di posta, firme ecc...) e di massimizzare il *pattern matching* effettuato dalla libreria del programma che utilizza i costrutti delle espressioni regolari del Perl. Con il termine *pattern matching* si intende l'operazione di verifica se una certa stringa o una sua sottostringa corrisponde ad un certo *pattern*, ovvero se essa è costruita secondo un determinato schema. Nel caso degli indirizzi Internet ad esempio nella forma *http://www.mmlink.it/* si cercano stringhe che iniziano con uno schema del tipo *http://* seguito da qualsiasi carattere alfanumerico (tranne alcuni) ripetuto più volte, seguito da un punto, seguito da altri caratteri alfanumerici (tranne alcuni) ripetuti più volte, seguito ancora da un punto e da altri caratteri alfanumerici. Naturalmente per abbracciare tutti i casi possibili di indirizzi individuabili nei messaggi Internet si è fatto riferimento alle regole che sono alla base della loro costruzione formale e sintattica. Ciò ha trovato realizzazione nella seguente espressione:

(([\<\()?( [Ff][Tt][Pp][Hh][Tt][Pp]))://([\w\d\-\-]+(\.\[\w\d\-\>\])\+)(([\w\d\-\-\.\~\>])\+)+)\*/?.

Stesso ragionamento è stato effettuato per gli indirizzi di posta, per gli *Emoticons* (segni che esprimono lo stato d'animo dello scrivente, tipo faccetta sorridente: utente felice) e per gli indirizzi numerici.

Per le firme alla fine dei messaggi si è percorsa una strada diversa. Data l'impossibilità di identificare uno schema di composizione univoco (una firma è rappresentata da una porzione di testo la cui struttura [forma] è fortemente dipendente dallo scrivente), si è ritenuto opportuno eliminare solamente quelle righe di testo che iniziavano con caratteri (asterischi, tilde ecc...) ripetuti per più di 3-4 volte (cfr. riga 38, 56 e 57 di Tabella 4).

Un'ultima nota. Differentemente da quanto effettuato su di una ricerca simile (Etudes des échanges électronique sur Internet et Intranet: forums et courriers électroniques [JADT 2000]), si è deciso di non eliminare il *quoting* (Vedi pag.74), ovvero quella parte di messaggio che l'autore decide di riportare da un altro messaggio per riferire il proprio commento. I motivi sono gli stessi di quelli visti in occasione delle firme dei messaggi: non poter riscontrare delle regolarità che avrebbero permesso un trattamento di eliminazione automatico.

Fatto girare il programma con i parametri appena spiegati, il file è stato infine trasformato in "tutto maiuscolo". Operazione quest'ultima che ha provocato una perdita d'informazione (impossibilità di identificazione dei nomi, delle sigle, ecc...) ma si è dimostrata necessaria per ridurre il numero di parole distinte in analisi in assenza di quei criteri, quali la *lemmatizzazione* la *disambiguazione* (Vedi il paragrafo successivo), utili proprio allo scopo.

In Tabella 5 sono riportati, in termini di estensione in byte, gli effetti dell'automatizzazione sul file di dati originario.

**Tabella 5:** Effetti dell'automatizzazione  
sui files originali

<b>Prima</b>	<b>File</b>	<b>Dopo</b>
2,21MB	atei.txt	1,87MB
2,29MB	catto.txt	2,01MB
152KB	ebraica.txt	125KB
555KB	religio.txt	506KB
<b>5,20MB</b>	<b>Totale</b>	<b>4,50MB</b>

### 3.5. Applicazione dello Spad-T®

I dati, passati attraverso le due fasi di trattamento automatizzato del testo, sono stati accorpati in un unico file<sup>69</sup> secondo lo schema *A*, riportato nell'APPENDICE C<sup>70</sup> e subito dopo sottoposti alle procedure dello Spad-T® per la loro analisi, saltando due passaggi fondamentali per l'Analisi del Contenuto: la *lemmatizzazione* (Vedi pag.20) e la *disambiguazione* (Vedi pag.34).

La loro applicazione avrebbe significato, infatti, aumentare in maniera considerevole i tempi di elaborazione e provocare uno sfruttamento oltre misura delle risorse a disposizione.

- **Procedura NUMER**

La procedura NUMER ha consentito di numerizzare il testo e di contare il numero di occorrenze (Vedi Appendice F per il file di parametri adottato). In Tabella 6 sono riportati, in forma succinta, i risultati di questa operazione per varie soglie di frequenza.

**Tabella 6:** Quadro riassuntivo dell'output della procedura NUMER dello Spad-T®

Soglia	Risposte	Occorrenze	N.ro par. dist.	Par. distinte	%COP
0	117.529	747.859	40.852	5,5%	100%
2	"	716.975	17.644	2,5%	96%
5	"	687.370	9.683	1,4%	92%
10	"	658.034	5.805	0,9%	88%
20	"	622.453	3.350	0,54%	83,2%

<sup>69</sup> Il file è stato salvato limitando il margine destro ("a capo") ad 80 colonne, in quanto lo Spad-T® non considera il testo oltre questo valore.

<sup>70</sup> La divisione dei messaggi attraverso i caratteri "----" è stata omessa. Il testo risulta così diviso in 4 subtesti rappresentati dai rispettivi *Newsgroups* in cui ciascuna linea (riga) è un *individuo*.

30	"	597.520	2.353	0,4%	80%
50	"	565.393	1.533	0,3%	75,6%

Si noti la bassa percentuale di copertura del testo già a soglia zero. Ciò a denotare la connotazione del linguaggio adottato nei *Newsgroups*: molto vicino al parlato.

- **Procedura CORTE**

La procedura CORTE è stata utilizzata per correggere il vocabolario, ovvero per eliminare\accorpare assieme determinate forme. Operazione questa che non può considerarsi una vera e propria *lemmatizzazione*. Infatti qui si è adottato come criterio quello di eliminare senza discriminazione<sup>71</sup> i pronomi, gli articoli, i verbi ausiliari, alcuni aggettivi ed alcune parole<sup>72</sup> di uso comune in Internet (come Email, wrote ecc...) nonché di riportare a forma canonica solo quelle forme che non davano adito a significati ambigui<sup>73</sup> (facciamo = fare; chiese = chiesa).

Si rimanda all'Appendice F (parametri della procedura CORTE) per l'elenco completo delle parole trattate.

- **Procedura SETEX1**

Essa ha consentito di aumentare a 50 la soglia di frequenza delle parole considerate e di impostare a 3 il valore di lunghezza delle parole in analisi (Vedi Appendice F

---

<sup>71</sup> L'attenzione si è soffermata solo sulle parole di lunghezza 3. Vedi procedura SETEX1.

<sup>72</sup> Tali parole si sarebbe potuto eliminarle attraverso l'automatizzazione vista nel paragrafo 3.2. . Ciò non è stato fatto semplicemente per non ridurre già dall'inizio il contenuto informativo dei dati e per poter controllare l'operazione eventualmente in funzione di alcuni *contesti locali* (Vedi pag.16).

<sup>73</sup> La disambiguazione avrebbe comportato la lettura di tutto il testo aumentando considerevolmente le risorse a disposizione.

per il file di parametri adottato per questa procedura). Ciò ha consentito di ottenere un numero totale di parole pari a 161.261 ed un numero di parole distinte pari a 764 (Valore extrapolato dopo un *porting* dei dati dall'*output* dello Spad-T® all'applicativo Excel® per Windows®).

- **Procedura MOCAR (Parole)**

Grazie alla MOCAR (Vedi Appendice F per il file di parametri adottato) è stata generata una lista delle prime 50 parole caratteristiche attraverso le quali si è potuto analizzare le specificità delle forme rispetto ai 4 testi (Vedi par.1.9. ). Si riporta qui di seguito l'*output* della procedura.

# Output procedura MOCARM

SELECTION DES FORMES LEXICALES CARACTERISTIQUES

LIBELLE DE LA FORME GRAPHIQUE	TEXTE NUMERO	1	ATEI		FREQUENCE INTERNE	V. TEST	PROBA
			POURCENTAGE INTERNE	POURCENTAGE GLOBAL			
1 ATEO			.73	.33	488.	528.	24.738 .000
2 CREDERE			.92	.53	621.	853.	18.394 .000
3 CREDENTE			.55	.29	373.	464.	17.032 .000
4 RELIGIONE			.75	.46	502.	737.	14.442 .000
5 ATEISMO			.19	.08	130.	135.	13.653 .000
6 ESISTERE			.66	.40	443.	649.	13.620 .000
7 SCIENTIFICO			.38	.21	259.	338.	13.015 .000
8 SCIENZA			.35	.18	234.	298.	12.917 .000
9 ESPERIENZA			.50	.30	339.	480.	12.766 .000
10 LOGICA			.22	.11	150.	175.	11.988 .000
11 DIOS			.11	.05	76.	76.	11.243 .000
12 UNIVERSO			.24	.13	159.	202.	10.665 .000
13 WILDE			.10	.04	65.	65.	10.359 .000
14 SCHIANTO			.09	.04	63.	63.	10.191 .000
15 CERVELLO			.13	.06	87.	95.	10.156 .000
16 LAMENTO			.10	.04	64.	66.	9.611 .000
17 IPOTESI			.28	.16	187.	263.	9.575 .000
18 MATEMATICA			.16	.08	108.	131.	9.492 .000
19 TEORIA			.39	.25	263.	408.	9.203 .000
20 IMPROBABILE			.11	.05	73.	82.	8.843 .000
21 PRE			.14	.07	96.	118.	8.736 .000
22 VOLARE			.08	.03	54.	56.	8.702 .000
23 REGIME			.12	.06	79.	92.	8.666 .000
24 CERTO			.59	.42	396.	680.	8.633 .000
25 DOGMA			.26	.16	174.	257.	8.339 .000
26 NERO			.10	.05	66.	75.	8.232 .000
27 ARBITRIO			.09	.04	59.	66.	7.976 .000
28 DEFINIZIONE			.14	.08	97.	126.	7.967 .000
29 SPAZIO			.16	.09	107.	143.	7.951 .000
30 IMPOSSIBILE			.15	.09	104.	141.	7.627 .000
31 GIORNALI			.11	.05	72.	88.	7.609 .000
32 FOTO			.09	.04	61.	71.	7.589 .000
33 ESISTENZA			.25	.16	168.	257.	7.574 .000
34 FARE		3.14	2.79	2114.	4494.	7.284	.000
35 SATANA			.22	.14	149.	227.	7.193 .000
36 PERSONA			.73	.57	493.	922.	7.163 .000
37 TELEVISIONE			.08	.04	54.	63.	7.095 .000
38 SISTEMA			.15	.09	100.	140.	7.023 .000
39 MIRACOLO			.16	.10	108.	156.	6.855 .000
40 DEMOCRAZIA			.07	.04	49.	57.	6.778 .000
41 VEDERE			.87	.71	586.	1142.	6.518 .000
42 BAMBINO			.23	.15	154.	247.	6.459 .000
43 TOGLIERE			.09	.05	61.	78.	6.452 .000
44 SCIENZIATI			.07	.04	49.	60.	6.211 .000
45 INFINITO			.10	.06	69.	95.	5.994 .000
46 PROVA			.33	.24	222.	390.	5.990 .000
47 MATERIA			.12	.07	83.	120.	5.976 .000
48 COSCIENZA			.16	.10	107.	165.	5.903 .000
49 LIBERO			.20	.14	136.	222.	5.789 .000
50 SOCIALE			.15	.10	104.	163.	5.596 .000
50 POPOLO			.03	.09	23.	144.	-6.557 .000
49 SALVEZZA			.03	.09	22.	141.	-6.590 .000
48 NOME			.13	.22	90.	361.	-6.619 .000
47 SECOLO			.07	.14	47.	227.	-6.631 .000
46 PRATICA			.03	.09	21.	139.	-6.681 .000
45 UOMO			.71	.90	477.	1449.	-6.885 .000
44 DOLORE			.02	.07	15.	120.	-6.891 .000
43 DEVOCIONE			.00	.04	1.	57.	-6.958 .000
42 SUKKOT			.00	.03	0.	52.	-7.107 .000
41 LUCA			.01	.06	9.	100.	-7.169 .000

40	SPECIE	.10	.19	68.	310.	-7.289	.000
39	COPPIA	.01	.05	5.	88.	-7.574	.000
38	RAPPORTO	.02	.08	15.	134.	-7.678	.000
37	LITURGIA	.00	.04	0.	60.	-7.682	.000
36	CADUTA	.00	.05	3.	80.	-7.730	.000
35	MEDITAZIONE	.00	.05	2.	75.	-7.778	.000
34	TRADIZIONE	.05	.13	32.	207.	-8.078	.000
33	AMORE	.22	.36	147.	581.	-8.237	.000
32	APOSTOLO	.01	.07	9.	120.	-8.329	.000
31	VERITÀ	.17	.30	113.	482.	-8.387	.000
30	GRECO	.02	.09	13.	139.	-8.391	.000
29	SABATO	.01	.06	4.	97.	-8.434	.000
28	TESTI	.15	.27	99.	442.	-8.514	.000
27	SPIRITO	.09	.20	63.	329.	-8.659	.000
26	SELEZIONE	.01	.07	7.	119.	-8.796	.000
25	SAN	.02	.10	16.	162.	-8.901	.000
24	TORAH	.01	.07	6.	117.	-8.961	.000
23	GIOVANNI	.04	.13	26.	204.	-8.966	.000
22	CUORE	.04	.14	29.	227.	-9.457	.000
21	ISRAELE	.00	.07	3.	111.	-9.534	.000
20	SERMONTI	.00	.06	0.	95.	-9.816	.000
19	EFISIO	.00	.06	0.	95.	-9.816	.000
18	SANTO	.12	.26	79.	418.	-9.897	.000
17	PACE	.06	.18	41.	288.	-10.063	.000
16	PAROLA	.27	.47	181.	757.	-10.260	.000
15	ANGELO	.03	.14	23.	223.	-10.310	.000
14	FRATELLO	.03	.14	22.	220.	-10.361	.000
13	TNM	.00	.07	0.	112.	-10.703	.000
12	SPIRITUALE	.02	.13	15.	206.	-11.079	.000
11	SCRITTURA	.03	.13	17.	216.	-11.108	.000
10	PREGHIERA	.01	.10	4.	154.	-11.335	.000
9	MUTAZIONE	.01	.11	7.	182.	-11.786	.000
8	GEOVA	.00	.10	3.	161.	-11.929	.000
7	PAOLO	.04	.17	24.	280.	-12.354	.000
6	TDG	.01	.11	4.	184.	-12.621	.000
5	CRISTO	.22	.47	151.	765.	-12.906	.000
4	SIGNORE	.04	.20	30.	324.	-12.969	.000
3	MARIA	.08	.26	51.	415.	-13.098	.000
2	TRADUZIONE	.00	.15	2.	245.	-15.437	.000
1	GESÙ	.21	.63	139.	1021.	-19.581	.000

TEXTE NUMERO 2 CATTOLCI

LIBELLE DE LA FORME GRAPHIQUE	--- POURCENTAGE ---		FREQUENCE		V.TEST	PROBA
	INTERNE	GLOBAL	INTERNE	GLOBALE		
1 TRADUZIONE	.32	.15	240.	245.	18.048	.000
2 GESÙ	1.01	.63	752.	1021.	17.925	.000
3 TDG	.24	.11	179.	184.	15.302	.000
4 MUTAZIONE	.24	.11	175.	182.	14.743	.000
5 GEOVA	.21	.10	158.	161.	14.656	.000
6 MARIA	.45	.26	335.	415.	14.497	.000
7 TNM	.15	.07	112.	112.	12.921	.000
8 SIGNORE	.35	.20	262.	324.	12.849	.000
9 SCRITTURA	.25	.13	185.	216.	12.067	.000
10 EFISIO	.13	.06	95.	95.	11.863	.000
11 SELEZIONE	.15	.07	112.	119.	11.230	.000
12 SERMONTI	.13	.06	93.	95.	11.103	.000
13 PAOLO	.29	.17	219.	280.	10.957	.000
14 SANTO	.41	.26	303.	418.	10.914	.000
15 CRISTO	.67	.47	501.	765.	10.798	.000
16 PREGHIERA	.18	.10	135.	154.	10.753	.000
17 SPECIE	.32	.19	234.	310.	10.499	.000
18 GRECO	.16	.09	121.	139.	10.032	.000
19 COPPIA	.11	.05	83.	88.	9.668	.000
20 PACE	.29	.18	213.	288.	9.563	.000
21 FRATELLO	.23	.14	169.	220.	9.245	.000
22 PAROLA	.64	.47	473.	757.	9.054	.000
23 LITURGIA	.08	.04	59.	60.	8.853	.000
24 NOME	.33	.22	246.	361.	8.415	.000
25 APOSTOLO	.13	.07	100.	120.	8.357	.000
26 TESTIMONI	.12	.07	92.	109.	8.210	.000
27 TESTI	.39	.27	288.	442.	8.036	.000
28 SCRITTO	.62	.47	457.	760.	7.759	.000
29 LUCA	.11	.06	83.	100.	7.533	.000
30 VERITÀ	.41	.30	303.	482.	7.371	.000

31	SACERDOTE	.14	.09	107.	140.	7.240	.000
32	CUORE	.21	.14	159.	227.	7.237	.000
33	DEVOZIONE	.07	.04	52.	57.	7.098	.000
34	DNA	.10	.05	73.	88.	7.042	.000
35	EVOLUZIONE	.17	.11	126.	178.	6.596	.000
36	CONTESTO	.11	.06	78.	100.	6.426	.000
37	DOLORE	.12	.07	90.	120.	6.354	.000
38	PARE	.29	.21	214.	338.	6.321	.000
39	DIRE	2.14	1.90	1587.	3069.	6.311	.000
40	OGGI	.39	.30	292.	485.	6.209	.000
41	CAMMINO	.11	.07	80.	106.	6.063	.000
42	CARNE	.12	.08	92.	126.	6.045	.000
43	PASTORI	.06	.03	48.	56.	6.033	.000
44	CEI	.06	.03	45.	52.	5.940	.000
45	FRASE	.15	.10	111.	160.	5.874	.000
46	BIBBIA	.28	.21	208.	335.	5.838	.000
47	INGLESE	.08	.04	57.	71.	5.796	.000
48	RAPPORTO	.13	.08	94.	134.	5.546	.000
49	SABATO	.10	.06	72.	97.	5.530	.000
50	TRADIZIONE	.18	.13	135.	207.	5.475	.000

50	IPOTESI	.10	.16	74.	263.	-5.894	.000
49	MAESTRO	.02	.06	14.	90.	-6.011	.000
48	MANIFESTO	.01	.05	8.	73.	-6.337	.000
47	SOGGETTO	.01	.05	8.	73.	-6.337	.000
46	INDIRIZZO	.02	.07	18.	108.	-6.340	.000
45	GIORNALI	.02	.05	12.	88.	-6.370	.000
44	SOLDI	.02	.05	12.	88.	-6.370	.000
43	CINA	.01	.04	6.	66.	-6.397	.000
42	FOTO	.01	.04	7.	71.	-6.481	.000
41	SPAZIO	.04	.09	28.	143.	-6.532	.000
40	CADUTA	.01	.05	9.	80.	-6.584	.000
39	ARBITRIO	.01	.04	5.	66.	-6.724	.000
38	IMPROBABILE	.01	.05	9.	82.	-6.731	.000
37	MATEMATICA	.03	.08	23.	131.	-6.770	.000
36	IMPOSSIBILE	.04	.09	26.	141.	-6.792	.000
35	VITA	.59	.75	437.	1203.	-6.826	.000
34	POTERE	.06	.12	42.	191.	-6.836	.000
33	LIBERO	.07	.14	52.	222.	-6.914	.000
32	NERO	.01	.05	6.	75.	-7.092	.000
31	MATERIA	.02	.07	18.	120.	-7.124	.000
30	EBREI	.06	.13	48.	217.	-7.255	.000
29	ESISTENZA	.08	.16	61.	257.	-7.342	.000
28	PRE	.02	.07	16.	118.	-7.437	.000
27	PERSONA	.42	.57	313.	922.	-7.451	.000
26	VOLARE	.00	.03	1.	56.	-7.454	.000
25	SUKKOT	.00	.03	0.	52.	-7.644	.000
24	COSCENZA	.04	.10	28.	165.	-7.798	.000
23	CERVELLO	.01	.06	8.	95.	-7.912	.000
22	LAMENTO	.00	.04	1.	66.	-8.210	.000
21	REGIME	.01	.06	6.	92.	-8.277	.000
20	SCHIANTO	.00	.04	0.	63.	-8.474	.000
19	WILDE	.00	.04	0.	65.	-8.617	.000
18	FILM	.01	.06	7.	104.	-8.760	.000
17	ADAMO	.02	.08	13.	132.	-8.945	.000
16	EVA	.01	.08	10.	125.	-9.243	.000
15	DIOS	.00	.05	0.	76.	-9.364	.000
14	LOGICA	.03	.11	22.	175.	-9.425	.000
13	SCIENZA	.08	.18	57.	298.	-9.686	.000
12	UNIVERSO	.04	.13	28.	202.	-9.694	.000
11	SCIENTIFICO	.09	.21	69.	338.	-9.792	.000
10	ESISTERE	.23	.40	171.	649.	-10.307	.000
9	TORAH	.00	.07	3.	117.	-10.674	.000
8	ATEISMO	.01	.08	5.	135.	-11.059	.000
7	ESPERIENZA	.14	.30	101.	480.	-11.389	.000
6	SATANA	.02	.14	17.	227.	-12.747	.000
5	CREDENTE	.11	.29	81.	464.	-12.949	.000
4	CREDERE	.27	.53	198.	853.	-13.825	.000
3	MORTE	.20	.45	147.	721.	-14.383	.000
2	RELIGIONE	.18	.46	135.	737.	-15.820	.000
1	ATEO	.05	.33	34.	528.	-20.217	.000

TEXTE NUMERO 3 EBRAICA

----- LIBELLE DE LA ----- POURCENTAGE ----- FREQUENCE ----- V.TEST ----- PROBA -----

## FORME GRAPHIQUE

## INTERNE GLOBAL INTERNE GLOBALE

1	TORAH	4.29	.07	108.	117.	28.886	.000
2	SUKKOT	2.06	.03	52.	52.	20.618	.000
3	EBRAICO	2.02	.09	51.	148.	15.343	.000
4	ISRAELE	1.51	.07	38.	111.	13.194	.000
5	EBREI	1.07	.13	27.	217.	8.152	.000
6	LIBRO	1.35	.31	34.	501.	6.949	.000
7	DIVINITÀ	.56	.05	14.	84.	6.444	.000
8	HOME	.48	.04	12.	61.	6.268	.000
9	ITALIA	.79	.16	20.	254.	5.724	.000
10	MEMORIA	.40	.04	10.	61.	5.376	.000
11	CORSO	.44	.06	11.	89.	5.111	.000
12	VECCHIO	.40	.05	10.	76.	4.977	.000
13	GRAZIE	.75	.19	19.	303.	4.917	.000
14	STORIA	.79	.21	20.	341.	4.844	.000
15	UMANITÀ	.44	.07	11.	109.	4.706	.000
16	BIBLICO	.32	.03	8.	53.	4.646	.000
17	GOVERNO	.40	.06	10.	91.	4.639	.000
18	RISPOSTA	.83	.25	21.	406.	4.565	.000
19	PRESENZA	.40	.06	10.	99.	4.477	.000
20	CONCETTO	.64	.16	16.	258.	4.465	.000
21	TRADIZIONE	.56	.13	14.	207.	4.382	.000
22	MESSAGGIO	.91	.31	23.	505.	4.344	.000
23	GRUPPO	.64	.18	16.	292.	4.123	.000
24	ASSOCIAZIONE	.32	.05	8.	73.	4.115	.000
25	INDIRIZZO	.36	.07	9.	108.	3.873	.000
26	AFFERMARE	.44	.10	11.	165.	3.826	.000
27	TESTI	.75	.27	19.	442.	3.744	.000
28	SERVE	.32	.06	8.	102.	3.526	.000
29	SCRITTO	1.03	.47	26.	760.	3.513	.000
30	GUERRA	.32	.06	8.	103.	3.508	.000
31	LINGUA	.28	.05	7.	79.	3.485	.000
32	ATTUALE	.28	.05	7.	82.	3.423	.000
33	PRODOTTO	.24	.04	6.	60.	3.386	.000
34	RISPONDERE	.52	.17	13.	268.	3.382	.000
35	DIALOGO	.32	.07	8.	111.	3.371	.000
36	INFORMAZIONE	.32	.07	8.	111.	3.371	.000
37	SCRITTI	.28	.05	7.	86.	3.343	.000
38	UNICO	.56	.21	14.	332.	3.132	.001
39	STATO	1.15	.62	29.	997.	3.009	.001
40	MODERNO	.28	.07	7.	110.	2.915	.002
41	POPOLO	.32	.09	8.	144.	2.873	.002
42	MILANO	.20	.04	5.	58.	2.856	.002
43	SCUSATE	.20	.04	5.	58.	2.856	.002
44	1999	.24	.05	6.	86.	2.827	.002
45	DOMANDA	.52	.21	13.	331.	2.817	.002
46	SCRITTA	.20	.04	5.	62.	2.761	.003
47	FINE	.48	.19	12.	299.	2.755	.003
48	NESSUNO	.95	.51	24.	828.	2.708	.003
49	CINA	.20	.04	5.	66.	2.671	.004
50	DUEMILA	.20	.04	5.	66.	2.671	.004
50	PARADISO	.00	.13	0.	211.	-1.810	.035
49	PICCOLO	.00	.13	0.	215.	-1.838	.033
48	DIMOSTRARE	.00	.14	0.	218.	-1.860	.031
47	DETTO	.16	.39	4.	631.	-1.868	.031
46	VIVERE	.08	.28	2.	444.	-1.881	.030
45	GUARDARE	.00	.14	0.	223.	-1.894	.029
44	DOVERE	.60	.97	15.	1562.	-1.919	.028
43	CUORE	.00	.14	0.	227.	-1.922	.027
42	SATANA	.00	.14	0.	227.	-1.922	.027
41	VANGELO	.00	.15	0.	241.	-2.016	.022
40	ORIGINALE	.00	.15	0.	241.	-2.016	.022
39	SENSO	.12	.36	3.	574.	-2.037	.021
38	SEMBRARE	.12	.36	3.	581.	-2.069	.019
37	VERO	.12	.37	3.	589.	-2.106	.018
36	DOGMA	.00	.16	0.	257.	-2.120	.017
35	STATI	.00	.16	0.	262.	-2.152	.016
34	CONTO	.00	.17	0.	271.	-2.208	.014
33	UNO	.32	.67	8.	1083.	-2.243	.012
32	NATURA	.00	.17	0.	280.	-2.263	.012
31	QUESTIONE	.00	.17	0.	281.	-2.269	.012
30	EGLI	.00	.17	0.	282.	-2.275	.011
29	SANTO	.04	.26	1.	418.	-2.310	.010
28	DISCORSO	.00	.18	0.	292.	-2.335	.010
27	OPERA	.00	.18	0.	295.	-2.353	.009

26	SCIENZA	.00	.18	0.	298.	-2.370	.009
25	FIGLIO	.08	.34	2.	553.	-2.414	.008
24	CREDENTE	.04	.29	1.	464.	-2.540	.006
23	SPIRITO	.00	.20	0.	329.	-2.546	.005
22	CORPO	.00	.20	0.	330.	-2.552	.005
21	SCIENTIFICO	.00	.21	0.	338.	-2.596	.005
20	POTER	.83	1.43	21.	2300.	-2.608	.005
19	DONNA	.00	.23	0.	365.	-2.739	.003
18	TEORIA	.00	.25	0.	408.	-2.955	.002
17	MARIA	.00	.26	0.	415.	-2.989	.001
16	RELIGIOSO	.00	.26	0.	417.	-2.999	.001
15	AMORE	.04	.36	1.	581.	-3.071	.001
14	ESPERIENZA	.00	.30	0.	480.	-3.290	.001
13	PADRE	.00	.30	0.	484.	-3.308	.000
12	FEDE	.12	.56	3.	908.	-3.374	.000
11	PERSONA	.12	.57	3.	922.	-3.423	.000
10	ATEO	.00	.33	0.	528.	-3.498	.000
9	CREDERE	.08	.53	2.	853.	-3.613	.000
8	MORTE	.04	.45	1.	721.	-3.630	.000
7	CERTO	.00	.42	0.	680.	-4.095	.000
6	CRISTO	.00	.47	0.	765.	-4.396	.000
5	FARE	1.31	2.79	33.	4494.	-4.948	.000
4	VITA	.04	.75	1.	1203.	-5.180	.000
3	CHIESA	.04	.86	1.	1389.	-5.683	.000
2	DIRE	.48	1.90	12.	3069.	-6.153	.000
1	DIO	.00	1.78	0.	2863.	-9.206	.000

TEXTE NUMERO 4 RELIGIONI

LIBELLE DE LA FORME GRAPHIQUE	---POURCENTAGE---		FREQUENCE		V.TEST	PROBA
	INTERNE	GLOBAL	INTERNE	GLOBALE		
1 MORTE	1.36	.45	234.	721.	15.792	.000
2 CADUTA	.40	.05	68.	80.	15.379	.000
3 EVA	.49	.08	85.	125.	15.196	.000
4 MEDITAZIONE	.35	.05	61.	75.	14.187	.000
5 MAESTRO	.38	.06	65.	90.	13.713	.000
6 THOMAS	.26	.03	44.	52.	12.290	.000
7 ADAMO	.41	.08	70.	132.	11.974	.000
8 PRINCIPIO	.41	.09	70.	149.	11.132	.000
9 ALBERO	.31	.06	54.	97.	10.797	.000
10 ATTENZIONE	.33	.08	56.	126.	9.597	.000
11 GIORNO	.75	.33	129.	528.	8.947	.000
12 VITA	1.36	.75	233.	1203.	8.922	.000
13 NASCITA	.27	.07	47.	108.	8.659	.000
14 SCOPO	.30	.08	51.	127.	8.555	.000
15 CREAZIONE	.31	.09	54.	145.	8.352	.000
16 UOMO	1.51	.90	259.	1449.	8.228	.000
17 FRUTTO	.24	.06	42.	97.	8.149	.000
18 SPIRITUALE	.38	.13	65.	206.	8.069	.000
19 MORIRE	.21	.05	36.	76.	7.974	.000
20 ANGELO	.37	.14	64.	223.	7.368	.000
21 SPIRITO	.46	.20	79.	329.	6.842	.000
22 AMORE	.69	.36	118.	581.	6.781	.000
23 LIBERTÀ	.44	.19	75.	308.	6.769	.000
24 SATANA	.35	.14	61.	227.	6.754	.000
25 NUOVO	.58	.32	100.	517.	5.792	.000
26 SALVEZZA	.23	.09	40.	141.	5.732	.000
27 RELAZIONE	.13	.04	23.	57.	5.701	.000
28 POPOLO	.23	.09	40.	144.	5.619	.000
29 OPERA	.38	.18	65.	295.	5.588	.000
30 FORZA	.33	.15	56.	240.	5.554	.000
31 PERIODO	.21	.08	36.	127.	5.426	.000
32 ANIMA	.26	.11	44.	174.	5.369	.000
33 GIOVANNI	.29	.13	49.	204.	5.363	.000
34 VOLONTÀ	.24	.10	41.	160.	5.251	.000
35 SAN	.24	.10	41.	162.	5.182	.000
36 CONOSCENZA	.28	.13	48.	204.	5.182	.000
37 ANNO	.75	.47	129.	764.	5.170	.000
38 EBREI	.29	.13	50.	217.	5.162	.000
39 SETTE	.13	.04	22.	61.	5.139	.000
40 PRIMO	.42	.23	73.	368.	5.128	.000
41 CAUSA	.24	.10	41.	164.	5.115	.000
42 RELIGIOSO	.47	.26	80.	417.	5.103	.000
43 CADERE	.12	.03	20.	53.	5.055	.000
44 MILANO	.12	.04	21.	58.	5.030	.000

45	GRANDE	.62	.38	107.	614.	5.005	.000
46	INCONTRO	.16	.06	27.	89.	4.971	.000
47	PRATICA	.21	.09	36.	139.	4.965	.000
48	OTTOBRE	.14	.05	24.	74.	4.948	.000
49	ROMA	.22	.09	38.	152.	4.918	.000
50	VIA	.42	.23	72.	372.	4.903	.000
50	CERVELLO	.00	.06	0.	95.	-4.091	.000
49	EFISIO	.00	.06	0.	95.	-4.091	.000
48	VOLERE	.67	.95	115.	1527.	-4.114	.000
47	SCUSA	.03	.12	5.	199.	-4.197	.000
46	SENTIRE	.11	.26	19.	414.	-4.339	.000
45	PUNTI	.22	.40	37.	648.	-4.348	.000
44	MENO	.17	.35	30.	561.	-4.367	.000
43	PROVA	.10	.24	17.	390.	-4.392	.000
42	LOGICA	.02	.11	3.	175.	-4.433	.000
41	PARADISO	.03	.13	5.	211.	-4.433	.000
40	EVOLUZIONE	.02	.11	3.	178.	-4.495	.000
39	TNM	.00	.07	0.	112.	-4.516	.000
38	TORAH	.00	.07	0.	117.	-4.634	.000
37	SCELTA	.03	.15	6.	239.	-4.639	.000
36	SELEZIONE	.00	.07	0.	119.	-4.680	.000
35	RIPETO	.00	.08	0.	122.	-4.749	.000
34	FEDE	.32	.56	55.	908.	-4.780	.000
33	DEFINIZIONE	.00	.08	0.	126.	-4.840	.000
32	SAPERE	.21	.42	36.	681.	-4.896	.000
31	MATEMATICA	.00	.08	0.	131.	-4.951	.000
30	SEMBRARE	.16	.36	28.	581.	-4.958	.000
29	NESSUNO	.27	.51	47.	828.	-4.978	.000
28	ATEISMO	.00	.08	0.	135.	-5.038	.000
27	SCIENTIFICO	.06	.21	10.	338.	-5.196	.000
26	SCIENZA	.04	.18	7.	298.	-5.358	.000
25	TDG	.01	.11	1.	184.	-5.473	.000
24	SPECIE	.04	.19	7.	310.	-5.551	.000
23	GEOVA	.00	.10	0.	161.	-5.573	.000
22	MAGARI	.02	.15	3.	237.	-5.608	.000
21	CERTO	.18	.42	31.	680.	-5.660	.000
20	PARE	.05	.21	8.	338.	-5.713	.000
19	SOLO	1.06	1.55	182.	2495.	-5.731	.000
18	DOGMA	.02	.16	3.	257.	-5.948	.000
17	ESISTERE	.16	.40	27.	649.	-5.950	.000
16	MUTAZIONE	.00	.11	0.	182.	-5.972	.000
15	SCRITTO	.20	.47	34.	760.	-6.093	.000
14	DETTO	.14	.39	24.	631.	-6.248	.000
13	POTER	.91	1.43	156.	2300.	-6.398	.000
12	CAPIRE	.16	.43	27.	689.	-6.408	.000
11	TRADUZIONE	.01	.15	1.	245.	-6.545	.000
10	TANTO	.20	.50	34.	804.	-6.564	.000
9	IPOTESI	.01	.16	1.	263.	-6.832	.000
8	VEDERE	.33	.71	56.	1142.	-6.917	.000
7	CREDENTE	.05	.29	9.	464.	-7.212	.000
6	TEORIA	.03	.25	6.	408.	-7.250	.000
5	CREDERE	.19	.53	32.	853.	-7.367	.000
4	CREDO	.08	.38	13.	614.	-8.114	.000
3	ATEO	.03	.33	6.	528.	-8.733	.000
2	DIRE	.91	1.90	157.	3069.	-10.996	.000
1	FARE	1.27	2.79	219.	4494.	-14.036	.000

SELECTION DES INDIVIDUS OU REPONSES CARACTERISTIQUES (CRITERE DE FREQUENCE DES MOTS)

-----  
TEXTE NUMERO 1 ATEI  
-----

-----  
CRITERE DE REPONSE OU INDIVIDU CARACTERISTIQUE  
CLASSIFICATION

-----  
24.738 -- 1 > ATEI,  
24.738 -- 2 ATEI".  
24.738 -- 3 ATEI.  
24.738 -- 4 > ATEI.  
24.738 -- 5 ATEI,  
24.738 -- 6 ATEO  
18.394 -- 7 CREDE.  
18.394 -- 8 CREDI.  
18.394 -- 9 CREDE.  
18.394 -- 10 CREDONO

17.032 -- 11 CREDENTI.  
 17.032 -- 12 CREDENTI.  
 17.032 -- 13 >CREDENTE ...  
 17.032 -- 14 > >CREDENTE ...  
 17.032 -- 15 > CREDENTI.  
 14.442 -- 16 >RELIGIONE.  
 14.442 -- 17 > RELIGIONE... ?  
 14.442 -- 18 RELIGIONI.  
 14.442 -- 19 >RELIGIONE.  
 14.442 -- 20 RELIGIONE.

TEXTE NUMERO 2 CATTOLICI

CRITERE DE REPONSE OU INDIVIDU CARACTERISTIQUE  
CLASSIFICATION

18.048 -- 1 TRADUZIONE  
 17.925 -- 2 GESÙ...  
 17.925 -- 3 GESÙ  
 17.925 -- 4 >, GESÙ  
 17.925 -- 5 GESÙ.  
 17.925 -- 6 GESÙ,  
 17.925 -- 7 GESÙ?")  
 17.925 -- 8 GESÙ  
 17.925 -- 9 GESÙ?  
 17.925 -- 10 GESÙ.  
 17.925 -- 11 GESÙ,  
 14.743 -- 12 MUTAZIONE.  
 14.743 -- 13 MUTAZIONI...  
 14.656 -- 14 GEOVA  
 14.656 -- 15 GEOVA  
 14.656 -- 16 GEOVA.  
 14.656 -- 17 GEOVA  
 14.656 -- 18 GEOVA  
 14.497 -- 19 MARIA:  
 14.497 -- 20 MARIA.

TEXTE NUMERO 3 EBRAICA

CRITERE DE REPONSE OU INDIVIDU CARACTERISTIQUE  
CLASSIFICATION

28.886 -- 1 TORAH?  
 20.618 -- 2 SUKKOT  
 20.618 -- 3 SUKKOT  
 15.343 -- 4 > EBRAICO"  
 15.343 -- 5 EBRAICO"  
 14.443 -- 6 SIMCHAT TORAH  
 14.443 -- 7 SIMCHAT TORAH  
 10.309 -- 8 YOUR SUKKOT -  
 10.309 -- 9 EREV SUKKOT  
 9.901 -- 10 "YOUR SUKKOT - SIMCHAT TORAH GUIDE,"  
 9.901 -- 11 THE SUKKOT - SIMCHAT TORAH CALENDAR  
 9.901 -- 12 YOUR SUKKOT - SIMCHAT TORAH GUIDE  
 9.901 -- 13 YOUR SUKKOT - SIMCHAT TORAH GUIDE  
 9.629 -- 14 SIMCHAT TORAH GUIDE  
 7.476 -- 15 THE SUKKOT - SIMCHAT TORAH CALENDAR 5760/1999  
 7.222 -- 16 WHAT IS SIMCHAT TORAH?  
 7.222 -- 17 WHAT IS SIMCHAT TORAH?  
 7.222 -- 18 FEET OF THE TORAH  
 7.222 -- 19 FEET OF THE TORAH  
 6.873 -- 20 OBSERVANCES OF SUKKOT

TEXTE NUMERO 4 RELIGIONI

CRITERE DE REPONSE OU INDIVIDU CARACTERISTIQUE  
CLASSIFICATION

15.792 -- 1 MORTE.  
 15.792 -- 2 MORTE?  
 15.792 -- 3 MORTE.  
 15.792 -- 4 MORTE.  
 15.792 -- 5 MORTE.  
 11.132 -- 6 PRINCIPIO.

9.057 -- 7 ADAMO ED EVA.  
8.947 -- 8 GIORNO."  
8.947 -- 9 GIORNI  
8.947 -- 10 GIORNI.  
8.922 -- 11 VITA.  
8.922 -- 12 VITA  
8.922 -- 13 VITA  
8.922 -- 14 VITA  
8.922 -- 15 VITA".  
8.922 -- 16 VITA.  
8.659 -- 17 NASCITA.  
8.352 -- 18 CREAZIONE.  
8.352 -- 19 CREAZIONE.  
8.228 -- 20 UOMINI.

- **Procedura SEGME**

Con la SEGME è avvenuto il calcolo della frequenza dei differenti *segmenti ripetuti* (Vedi pag.13) e la costruzione<sup>74</sup> della tabella di contingenza che incrocia le risposte (linee di testo) in riga ed i *segmenti* in colonna. Si rimanda all'Appendice F per il file di parametri adottato.

Nella pagina successiva si riportano i passi salienti dell'Output della SEGME.

---

<sup>74</sup> La costruzione avviene in un file interno allo Spad-T®

## Output procedura SEGME

---

### TABLEAU DES SEGMENTS REPETES

---

SEUILS MINIMUM DE FREQUENCE DE REPETITION:

SEUIL GENERAL 10  
SEGMENTS DE LONGUEUR 2 10  
SEGMENTS DE LONGUEUR 3 10

---

### SEG FREQ LONG TEXTE DU SEGMENT

---

-----AFFERMAZIONE		
1	11	2 AFFERMAZIONE DIO
2	18	2 AMARE DARE
3	13	2 ANNO PRIMA
4	12	2 ANNO SANTO
5	16	2 BAMBINO MORTI
6	15	4 BAMBINO MORTI SENZA BATTESIMO
7	12	2 BASE COMUNE
8	12	2 BELLO DIFFERENZA
9	11	2 BENE DIVERSO
10	11	2 BENE FARE
11	41	2 BENE POCO
12	14	2 BENE PRESTO
13	11	2 BIBBIA TDG
14	41	2 BUONO FEDE
15	11	2 BUONO LETTURA
16	15	2 BUONO NOTIZIA
17	20	2 BUONO SENSO
18	19	2 BUONO VOLONTÀ
19	27	2 CAPIRE BENE
20	14	2 CAPIRE MEGLIO
21	12	2 CARO PIETRO
22	12	2 CARO SALUTO
23	12	2 CASO PARTICOLARE
24	12	2 CERTI ARGOMENTO
25	20	2 CERTO PUNTI
26	28	2 CERTO SENSO
27	39	2 CHIEDERE SCUSA
28	142	2 CHIESA CATTOLICO
29	11	2 CHIESA CRISTIANO
30	13	2 CONOSCERE BENE
31	11	2 CONOSCENZA SCIENTIFICO
32	15	2 CORPO FISICO
33	13	2 CRISTO GESÙ
34	13	2 CRISTO STORICO
35	33	2 CULTURA GENERALE
36	18	2 DOVERE ANDARE
37	14	2 DOVERE CREDERE
38	12	2 DOVERE DARE
39	37	2 DOVERE DIRE
40	82	2 DOVERE FARE
41	11	2 DOVERE RISPONDERE

42 11 2 DOVERE SENTIRE  
 43 15 2 DOVERE SOLO  
 -----DIRE  
 44 18 2 DIRE NULLA  
 45 25 2 DIRE OGGI  
 46 11 2 DIRE SOLO  
 -----DIECI  
 47 14 2 DIECI ANNO  
 -----DIO  
 48 57 2 DIO ESISTERE  
 49 18 2 DIO PADRE  
 50 13 2 DIO POTER  
 51 11 2 DIO UOMO  
 -----DIRITTO  
 52 21 2 DIRITTO UMANO  
 -----DUE  
 53 14 2 DUE ANNO  
 54 16 2 DUE PARTI  
 55 11 2 DUE PERSONA  
 56 14 2 DUE VOLTE  
 -----DUEMILA  
 57 47 2 DUEMILA ANNO  
 -----ESPERIENZA  
 58 15 2 ESPERIENZA PERSONALE  
 -----FARE  
 59 31 2 FARE BENE  
 60 40 2 FARE CAPIRE  
 61 17 2 FARE CREDERE  
 62 20 2 FARE MALE  
 63 13 2 FARE NASCERE  
 64 37 2 FARE NOTARE  
 65 17 2 FARE NULLA  
 66 42 2 FARE PARTE  
 67 13 2 FARE PASSARE  
 68 13 2 FARE RIFERIMENTO  
 69 22 2 FARE SOLO  
 70 11 2 FARE TANTO  
 71 24 2 FARE VEDERE  
 -----FEDE  
 72 16 2 FEDE CATTOLICO  
 73 15 2 FEDE CRISTIANO  
 -----FORSE  
 74 11 2 FORSE DOVERE  
 -----GENERE  
 75 14 2 GENERE UMANO  
 -----GESÙ  
 76 89 2 GESÙ CRISTO  
 77 18 2 GESÙ DIRE  
 -----GIOVANNI  
 78 36 2 GIOVANNI PAOLO  
 -----GRANDE  
 79 17 2 GRANDE PARTE  
 -----GRAZIE  
 80 11 2 GRAZIE GRAZIE  
 -----GUARDARE  
 81 12 2 GUARDARE CASO  
 -----HOME  
 82 43 2 HOME PAGE  
 -----IPOTESI  
 83 17 2 IPOTESI DIVINA  
 -----LASCIARE  
 84 13 2 LASCIARE PERDERE  
 -----LIBERO  
 85 56 2 LIBERO ARBITRIO  
 -----LIBERTÀ  
 86 29 2 LIBERTÀ RELIGIOSO  
 -----LIBRO  
 87 13 2 LIBRO SACRO  
 -----LINGUA  
 88 12 2 LINGUA ORIGINALE  
 -----MAGGIORE  
 89 42 2 MAGGIORE PARTE  
 -----MENO  
 90 16 2 MENO MALE  
 -----METODO  
 91 13 2 METODO SCIENTIFICO  
 -----MODO  
 92 17 2 MODO DIVERSO  
 93 11 2 MODO MENO  
 -----MONDO

94	12	2 MONDO CATTOLICO	-----MORTI
95	16	3 MORTI SENZA BATTESSIMO	-----NATURA
96	17	2 NATURA DIVINA	
97	24	2 NATURA UMANO	-----NESSUNO
98	11	2 NESSUNO DIO	
99	14	2 NESSUNO MODO	
100	24	2 NESSUNO PARTE	-----NORMALE
101	12	2 NESSUNO TIPO	-----NUOVO
102	12	2 NORMALE PADRE	-----NUOVO
103	21	2 NUOVO MONDO	-----OTTOBRE
104	17	2 NUOVO TESTAMENTO	
105	15	2 OTTOBRE 1999	-----PADRE
106	13	2 PADRE FIGLIO	-----PARTIRE
107	11	2 PADRE NORMALE	-----PECCATO
108	14	2 PADRE PIO	-----POCO
110	122	2 PECCATO ORIGINALE	-----POTER
111	13	2 POCO PAROLA	-----PRE
112	32	2 POTER CAPIRE	-----PRIMA
113	12	2 POTER CONTINUARE	-----PRIMA
114	13	2 POTER CREDERE	-----PRIMO
115	27	2 POTER DARE	-----RELIGIONE
116	125	2 POTER DIRE	-----RENDERE
117	12	2 POTER DIMOSTRARE	-----RENDERE
118	18	2 POTER DIVENTARE	-----SAPERE
119	132	2 POTER FARE	-----SAN
120	14	2 POTER LEGGERE	-----SCUOLE
121	16	2 POTER PARLARE	-----SELEZIONE
122	17	2 POTER PENSARE	-----SEMPICE
123	13	2 POTER RISONDERE	-----SEMPRE
124	31	2 POTER TROVARE	-----SENTIRE
125	30	2 POTER VEDERE	-----SENZA
126	33	2 POTER VOLARE	-----SETTEMBRE
127	93	2 PRE MORTE	-----SIGNORE
128	17	2 PRIMA PERSONA	
129	21	2 PRIMA VOLTA	
130	11	2 PRIMO CASO	
131	14	2 RELIGIONE CATTOLICO	
132	15	2 RELIGIONE CRISTIANO	
133	26	2 RENDERE CONTO	
134	11	2 SAPERE BENE	
135	17	2 SAN GIOVANNI	
136	34	2 SAN PAOLO	
137	21	2 SAN PIETRO	
138	12	2 SCUOLE PUBBLICA	
139	15	2 SELEZIONE NATURALE	
140	11	2 SEMPLICE CURIOSITÀ	
141	14	2 SEMPLICE FARE	
142	11	2 SEMPRE DETTO	
143	31	2 SENTIRE PARLARE	
144	19	2 SENZA BATTESSIMO	
145	13	2 SENZA DUBBIO	
146	19	2 SETTEMBRE 1999	
147	12	2 SIGNORE GESÙ	

148	26	2 SITO INTERNET	SITO
149	15	2 SOCIETÀ CIVILE	SOCIETÀ
150	13	2 SOLO DETTO	SOLO
151	12	2 SOLO DIO	
152	14	2 SOLO DUE	
153	25	2 SOLO FARE	
154	11	2 SOLO PAROLA	
155	22	2 SPAZIO TEMPO	SPAZIO
156	12	2 SPECIE DIVERSA	SPECIE
157	66	2 SPIRITO SANTO	SPIRITO
158	14	2 STATO FARE	STATO
159	12	2 STATO USARE	
160	11	2 TANTO TEMPO	TANTO
161	42	2 TEORIA SCIENTIFICO	TEORIA
162	24	2 TERZO MILLENNIO	TERZO
163	11	2 TESTI BIBLICO	TESTI
164	17	2 TESTI GRECO	
165	14	2 TRADUZIONE LETTERALE	TRADUZIONE
166	17	2 TRE GIORNO	TRE
167	63	2 UNO SCHIANTO	UNO
168	26	2 UNO SOLO	
169	55	2 UNO STATO	
170	16	2 UNO STUDIO	
171	11	2 UOMO DIO	UOMO
172	17	2 UOMO NUOVO	
173	15	2 VENIRE DARE	VENIRE
174	21	2 VENIRE FARE	
175	21	2 VEDERE SOLO	VEDERE
176	15	2 VEDERE SOPRA	
177	14	2 VERGINE MARIA	VERGINE
178	31	2 VERO DIO	VERO
179	20	2 VIA DIRE	VIA
180	11	2 VIA VIA	
181	24	2 VITA ETERNA	VITA
182	13	2 VITA SPIRITUALE	
183	14	2 VOLERE ANDARE	VOLERE
184	12	2 VOLERE ARRIVARE	
185	21	2 VOLERE CAPIRE	
186	15	2 VOLERE CREDERE	
187	12	2 VOLERE DARE	
188	169	2 VOLERE DIRE	
189	19	2 VOLERE DISCUTERE	
190	77	2 VOLERE FARE	
191	15	2 VOLERE METTERE	
192	11	2 VOLERE PARLARE	
193	25	2 VOLERE SAPERE	
194	26	2 VOLERE SOLO	
195	19	2 VOLERE VEDERE	

- **Procedura MOCAR (Segmenti)**

Grazie alla MOCAR (Vedi Appendice F per il file di parametri adottato) è stata generata una lista dei primi 50 segmenti caratteristici attraverso i quali si è potuto analizzare ulteriormente, in riferimento ai 4 testi, alcune specificità (Vedi par.1.9. ).

Si riporta qui di seguito l'output della procedura.

# Output della procedura MOCARS

## SELECTION DES FORMES LEXICALES CARACTERISTIQUES

	TEXTE NUMERO	1	ATEI	POURCENTAGE		FREQUENCE		V.TEST	PROBA
	LIBELLE DE LA FORME GRAPHIQUE			INTERNE	GLOBAL	INTERNE	GLOBALE		
1	127-PRE MORTE	4.10	2.03	91.	93.	10.634	.000		
2	167-UNO SCHIANTO	2.84	1.37	63.	63.	9.276	.000		
3	85-LIBERO ARBITRIO	2.34	1.22	52.	56.	7.046	.000		
4	110-PECCATO ORIGINALE	4.32	2.66	96.	122.	6.851	.000		
5	35-CULTURA GENERALE	1.49	.72	33.	33.	6.514	.000		
6	48-DIO ESISTERE	2.30	1.24	51.	57.	6.456	.000		
7	126-POTER VOLARE	1.44	.72	32.	33.	5.950	.000		
8	155-SPAZIO TEMPO	.95	.48	21.	22.	4.541	.000		
9	83-IPOTESI DIVINA	.77	.37	17.	17.	4.447	.000		
10	91-METODO SCIENTIFICO	.59	.28	13.	13.	3.775	.000		
11	106-PADRE FIGLIO	.59	.28	13.	13.	3.775	.000		
12	87-LIBRO SACRO	.59	.28	13.	13.	3.775	.000		
13	82-HOME PAGE	1.49	.94	33.	43.	3.644	.000		
14	102-NORMALE PADRE	.54	.26	12.	12.	3.590	.000		
15	1-AFFERMAZIONE DIO	.50	.24	11.	11.	3.396	.000		
16	107-PADRE NORMALE	.50	.24	11.	11.	3.396	.000		
17	31-CONOSCENZA SCIENTIFICO	.50	.24	11.	11.	3.396	.000		
18	98-NESSUNO DIO	.50	.24	11.	11.	3.396	.000		
19	119-POTER FARE	3.69	2.88	82.	132.	3.107	.001		
20	189-VOLERE DISCUTERE	.72	.41	16.	19.	2.983	.001		
21	138-SCUOLE PUBBLICA	.50	.26	11.	12.	2.836	.002		
22	32-CORPO FISICO	.59	.33	13.	15.	2.800	.003		
23	191-VOLERE METTERE	.59	.33	13.	15.	2.800	.003		
24	25-CERTO PUNTI	.72	.44	16.	20.	2.659	.004		
25	50-DIO POTER	.50	.28	11.	13.	2.395	.008		
26	159-STATO USARE	.45	.26	10.	12.	2.177	.015		
27	152-SOLO DUE	.50	.31	11.	14.	2.018	.022		
28	71-FARE VEDERE	.77	.52	17.	24.	2.009	.022		
29	143-SENTIRE PARLARE	.95	.68	21.	31.	1.985	.024		
30	192-VOLERE PARLARE	.41	.24	9.	11.	1.946	.026		
31	9-BENE DIVERSO	.41	.24	9.	11.	1.946	.026		
32	51-DIO UOMO	.41	.24	9.	11.	1.946	.026		
33	161-TEORIA SCIENTIFICO	1.22	.92	27.	42.	1.912	.028		
34	185-VOLERE CAIRE	.68	.46	15.	21.	1.906	.028		
35	44-DIRE NULLA	.59	.39	13.	18.	1.798	.036		
36	150-SOLO DETTO	.45	.28	10.	13.	1.797	.036		
37	149-SOCIETÀ CIVILE	.50	.33	11.	15.	1.684	.046		
38	186-VOLERE CREDERE	.50	.33	11.	15.	1.684	.046		
39	6-BAMBINO MORTI SENZA BATTESSIMO	.50	.33	11.	15.	1.684	.046		
40	66-FARE PARTE	1.17	.92	26.	42.	1.599	.055		
41	65-FARE NULLA	.54	.37	12.	17.	1.594	.056		
42	122-POTER PENSARE	.54	.37	12.	17.	1.594	.056		
43	8-BELLO DIFFERENZA	.41	.26	9.	12.	1.563	.059		
44	81-GUARDARE CASO	.41	.26	9.	12.	1.563	.059		
45	133-RENDERE CONTO	.77	.57	17.	26.	1.539	.062		
46	174-VENIRE FARE	.63	.46	14.	21.	1.458	.072		
47	175-VEDERE SOLO	.63	.46	14.	21.	1.458	.072		
48	95-MORTI SENZA BATTESSIMO	.50	.35	11.	16.	1.380	.084		
49	5-BAMBINO MORTI	.50	.35	11.	16.	1.380	.084		
50	169-UNO STATO	1.44	1.20	32.	55.	1.316	.094		
50	187-VOLERE DARE	.09	.26	2.	12.	-1.966	.025		
49	38-DOVERE DARE	.09	.26	2.	12.	-1.966	.025		
48	24-CERTI ARGOMENTO	.09	.26	2.	12.	-1.966	.025		
47	105-OTTOBRE 1999	.14	.33	3.	15.	-1.992	.023		
46	49-DIO PADRE	.18	.39	4.	18.	-2.033	.021		
45	100-NESSUNO PARTE	.27	.52	6.	24.	-2.135	.016		
44	111-POCO PAROLA	.09	.28	2.	13.	-2.176	.015		
43	63-FARE NASCERE	.09	.28	2.	13.	-2.176	.015		
42	67-FARE PASSARE	.09	.28	2.	13.	-2.176	.015		
41	18-BUONO VOLONTÀ	.18	.41	4.	19.	-2.213	.013		
40	135-SAN GIOVANNI	.14	.37	3.	17.	-2.371	.009		

39	37-DOVERE CREDERE	.09	.31	2.	14.	-2.376	.009
38	59-FARE BENE	.36	.68	8.	31.	-2.392	.008
37	15-BUONO LETTURA	.05	.24	1.	11.	-2.424	.008
36	188-VOLERE DIRE	2.97	3.69	66.	169.	-2.424	.008
35	139-SELEZIONE NATURALE	.09	.33	2.	15.	-2.567	.005
34	184-VOLERE ARRIVARE	.05	.26	1.	12.	-2.630	.004
33	78-GIOVANNI PAOLO	.41	.79	9.	36.	-2.712	.003
32	27-CHIEDERE SCUSA	.45	.85	10.	39.	-2.753	.003
31	68-FARE RIFERIMENTO	.05	.28	1.	13.	-2.825	.002
30	168-UNO SOLO	.23	.57	5.	26.	-2.878	.002
29	177-VERGINE MARIA	.05	.31	1.	14.	-3.011	.001
28	12-BENE PRESTO	.05	.31	1.	14.	-3.011	.001
27	136-SAN PAOLO	.32	.74	7.	34.	-3.181	.001
26	163-TESTI BIBLICO	.00	.24	0.	11.	-3.206	.001
25	154-SOLO PAROLA	.00	.24	0.	11.	-3.206	.001
24	80-GRAZIE GRAZIE	.00	.24	0.	11.	-3.206	.001
23	13-BIBBIA TDG	.00	.24	0.	11.	-3.206	.001
22	4-ANNO SANTO	.00	.26	0.	12.	-3.392	.000
21	88-LINGUA ORIGINALE	.00	.26	0.	12.	-3.392	.000
20	7-BASE COMUNE	.00	.26	0.	12.	-3.392	.000
19	96-NATURA DIVINA	.05	.37	1.	17.	-3.524	.000
18	137-SAN PIETRO	.09	.46	2.	21.	-3.567	.000
17	182-VITA SPIRITUALE	.00	.28	0.	13.	-3.571	.000
16	33-CHRISTO GESÙ	.00	.28	0.	13.	-3.571	.000
15	77-GESÙ DIRE	.05	.39	1.	18.	-3.683	.000
14	165-TRADUZIONE LETTERALE	.00	.31	0.	14.	-3.741	.000
13	16-BUONO NOTIZIA	.00	.33	0.	15.	-3.905	.000
12	173-VENIRE DARE	.00	.33	0.	15.	-3.905	.000
11	172-UOMO NUOVO	.00	.37	0.	17.	-4.216	.000
10	166-TRE GIORNO	.00	.37	0.	17.	-4.216	.000
9	164-TESTI GRECO	.00	.37	0.	17.	-4.216	.000
8	2-AMARE DARE	.00	.39	0.	18.	-4.364	.000
7	162-TERZO MILLENNIO	.05	.52	1.	24.	-4.543	.000
6	157-SPIRITO SANTO	.59	1.44	13.	66.	-4.745	.000
5	14-BUONO FEDE	.23	.89	5.	41.	-4.773	.000
4	103-NUOVO MONDO	.00	.46	0.	21.	-4.783	.000
3	45-DIRE OGGI	.00	.55	0.	25.	-5.294	.000
2	148-SITO INTERNET	.00	.57	0.	26.	-5.415	.000
1	76-GESÙ CRISTO	.54	1.94	12.	89.	-6.930	.000

TEXTE NUMERO 2 CATTOLICI

LIBELLE DE LA FORME GRAPHIQUE	--- POURCENTAGE ---		FREQUENCE		V.TEST	PROBA
	INTERNE	GLOBAL	INTERNE	GLOBALE		
1 45-DIRE OGGI	1.26	.55	25.	25.	6.060	.000
2 103-NUOVO MONDO	1.06	.46	21.	21.	5.489	.000
3 14-BUONO FEDE	1.77	.89	35.	41.	5.451	.000
4 76-GESÙ CRISTO	3.19	1.94	63.	89.	5.211	.000
5 2-AMARE DARE	.91	.39	18.	18.	5.022	.000
6 164-TESTI GRECO	.86	.37	17.	17.	4.857	.000
7 16-BUONO NOTIZIA	.76	.33	15.	15.	4.511	.000
8 165-TRADUZIONE LETTERALE	.71	.31	14.	14.	4.328	.000
9 162-TERZO MILLENNIO	1.06	.52	21.	24.	4.302	.000
10 172-UOMO NUOVO	.81	.37	16.	17.	4.188	.000
11 33-CHRISTO GESÙ	.66	.28	13.	13.	4.138	.000
12 157-SPIRITO SANTO	2.28	1.44	45.	66.	4.004	.000
13 88-LINGUA ORIGINALE	.61	.26	12.	12.	3.940	.000
14 80-GRAZIE GRAZIE	.56	.24	11.	11.	3.733	.000
15 163-TESTI BIBLICO	.56	.24	11.	11.	3.733	.000
16 13-BIBBIA TDG	.56	.24	11.	11.	3.733	.000
17 59-FARE BENE	1.16	.68	23.	31.	3.329	.000
18 168-UNO SOLO	1.01	.57	20.	26.	3.307	.000
19 77-GESÙ DIRE	.76	.39	15.	18.	3.256	.001
20 139-SELEZIONE NATURALE	.66	.33	13.	15.	3.209	.001
21 184-VOLERE ARRIVARE	.56	.26	11.	12.	3.199	.001
22 4-ANNO SANTO	.56	.26	11.	12.	3.199	.001
23 188-VOLERE DIRE	4.70	3.69	93.	169.	3.085	.001
24 37-DOVERE CREDERE	.61	.31	12.	14.	2.999	.001
25 177-VERGINE MARIA	.61	.31	12.	14.	2.999	.001
26 12-BENE PRESTO	.61	.31	12.	14.	2.999	.001
27 27-CHIEDERE SCUSA	1.32	.85	26.	39.	2.808	.002
28 67-FARE PASSARE	.56	.28	11.	13.	2.780	.003
29 111-POCO PAROLA	.56	.28	11.	13.	2.780	.003
30 24-CERTI ARGOMENTO	.51	.26	10.	12.	2.549	.005
31 187-VOLERE DARE	.51	.26	10.	12.	2.549	.005

32	38-DOVERE DARE	.51	.26	10.	12.	2.549	.005
33	96-NATURA DIVINA	.66	.37	13.	17.	2.542	.006
34	158-STATO FARE	.56	.31	11.	14.	2.421	.008
35	182-VITA SPIRITUALE	.51	.28	10.	13.	2.187	.014
36	173-VENIRE DARE	.56	.33	11.	15.	2.103	.018
37	26-CERTO SENSO	.91	.61	18.	28.	2.067	.019
38	21-CARO PIETRO	.46	.26	9.	12.	1.939	.026
39	22-CARO SALUTO	.46	.26	9.	12.	1.939	.026
40	156-SPECIE DIVERSA	.46	.26	9.	12.	1.939	.026
41	120-POTER LEGGERE	.51	.31	10.	14.	1.867	.031
42	11-BENE POCO	1.21	.89	24.	41.	1.833	.033
43	49-DIO PADRE	.61	.39	12.	18.	1.775	.038
44	148-SITO INTERNET	.81	.57	16.	26.	1.694	.045
45	10-BENE FARE	.40	.24	8.	11.	1.676	.047
46	60-FARE CAPIRE	1.16	.87	23.	40.	1.675	.047
47	68-FARE RIFERIMENTO	.46	.28	9.	13.	1.617	.053
48	63-FARE NASCERE	.46	.28	9.	13.	1.617	.053
49	18-BUONO VOLONTÀ	.61	.41	12.	19.	1.528	.063
50	137-SAN PIETRO	.66	.46	13.	21.	1.514	.065

50	186-VOLERE CREDERE	.20	.33	4.	15.	-1.030	.152
49	6-BAMBINO MORTI SENZA BATTESSIMO	.20	.33	4.	15.	-1.030	.152
48	44-DIRE NULLA	.25	.39	5.	18.	-1.082	.140
47	185-VOLERE CAPIRE	.30	.46	6.	21.	-1.134	.128
46	66-FARE PARTE	.71	.92	14.	42.	-1.135	.128
45	71-FARE VEDERE	.35	.52	7.	24.	-1.184	.118
44	150-SOLO DETTO	.15	.28	3.	13.	-1.190	.117
43	192-VOLERE PARLARE	.10	.24	2.	11.	-1.391	.082
42	9-BENE DIVERSO	.10	.24	2.	11.	-1.391	.082
41	29-CHIESA CRISTIANO	.10	.24	2.	11.	-1.391	.082
40	51-DIO UOMO	.10	.24	2.	11.	-1.391	.082
39	152-SOLO DUE	.15	.31	3.	14.	-1.392	.082
38	65-FARE NULLA	.20	.37	4.	17.	-1.408	.080
37	122-POTER PENSARE	.20	.37	4.	17.	-1.408	.080
36	143-SENTIRE PARLARE	.46	.68	9.	31.	-1.422	.078
35	149-SOCIETÀ CIVILE	.15	.33	3.	15.	-1.583	.057
34	105-OTTOBRE 1999	.15	.33	3.	15.	-1.583	.057
33	113-POTER CONTINUARE	.10	.26	2.	12.	-1.601	.055
32	81-GUARDARE CASO	.10	.26	2.	12.	-1.601	.055
31	181-VITA ETERNA	.30	.52	6.	24.	-1.617	.053
30	50-DIO POTER	.10	.28	2.	13.	-1.799	.036
29	25-CERTO PUNTI	.20	.44	4.	20.	-1.918	.028
28	86-LIBERTÀ RELIGIOSO	.35	.63	7.	29.	-1.923	.027
27	135-SAN GIOVANNI	.15	.37	3.	17.	-1.940	.026
26	191-VOLERE METTERE	.10	.33	2.	15.	-2.166	.015
25	119-POTER FARE	2.23	2.88	44.	132.	-2.240	.013
24	138-SCUOLE PUBBLICA	.05	.26	1.	12.	-2.275	.011
23	159-STATO USARE	.05	.26	1.	12.	-2.275	.011
22	189-VOLERE DISCUTERE	.10	.41	2.	19.	-2.812	.002
21	107-PADRE NORMALE	.00	.24	0.	11.	-2.880	.002
20	1-AFFERMAZIONE DIO	.00	.24	0.	11.	-2.880	.002
19	98-NESSUNO DIO	.00	.24	0.	11.	-2.880	.002
18	31-CONOSCENZA SCIENTIFICO	.00	.24	0.	11.	-2.880	.002
17	7-BASE COMUNE	.00	.26	0.	12.	-3.054	.001
16	102-NORMALE PADRE	.00	.26	0.	12.	-3.054	.001
15	87-LIBRO SACRO	.00	.28	0.	13.	-3.220	.001
14	91-METODO SCIENTIFICO	.00	.28	0.	13.	-3.220	.001
13	106-PADRE FIGLIO	.00	.28	0.	13.	-3.220	.001
12	82-HOME PAGE	.40	.94	8.	43.	-3.239	.001
11	32-CORPO FISICO	.00	.33	0.	15.	-3.532	.000
10	169-UNO STATO	.51	1.20	10.	55.	-3.786	.000
9	155-SPAZIO TEMPO	.05	.48	1.	22.	-3.811	.000
8	83-IPOTESI DIVINA	.00	.37	0.	17.	-3.821	.000
7	126-POTER VOLARE	.00	.72	0.	33.	-5.663	.000
6	35-CULTURA GENERALE	.00	.72	0.	33.	-5.663	.000
5	85-LIBERO ARBITRIO	.10	1.22	2.	56.	-6.661	.000
4	110-PECCATO ORIGINALE	.91	2.66	18.	122.	-6.714	.000
3	48-DIO ESISTERE	.10	1.24	2.	57.	-6.739	.000
2	167-UNO SCHIANTO	.00	1.37	0.	63.	-8.117	.000
1	127-PRE MORTE	.05	2.03	1.	93.	-9.568	.000

TEXTE NUMERO 3 EBRAICA

LIBELLE DE LA FORME GRAPHIQUE --- POURCENTAGE --- FREQUENCE V. TEST PROBA

INTERNE GLOBAL INTERNE GLOBALE

1	154-SOLO PAROLA	14.00	.24	7.	11.	6.846	.000
2	123-POTER RISONDERE	6.00	.28	3.	13.	3.411	.000
3	173-VENIRE DARE	6.00	.33	3.	15.	3.287	.001
4	89-MAGGIORE PARTE	8.00	.92	4.	42.	3.080	.001
5	57-DUEMILA ANNO	8.00	1.03	4.	47.	2.952	.002
6	100-NESSUNO PARTE	6.00	.52	3.	24.	2.862	.002
7	15-BUONO LETTURA	4.00	.24	2.	11.	2.511	.006
8	41-DOVERE RISONDERE	4.00	.24	2.	11.	2.511	.006
9	105-OTTOBRE 1999	4.00	.33	2.	15.	2.284	.011
10	52-DIRITTO UMANO	4.00	.46	2.	21.	2.024	.021
11	137-SAN PIETRO	4.00	.46	2.	21.	2.024	.021
12	74-FORSE DOVERE	2.00	.24	1.	11.	1.207	.114
13	94-MONDO CATTOLICO	2.00	.26	1.	12.	1.158	.123
14	30-CONOSCERE BENE	2.00	.28	1.	13.	1.112	.133
15	20-CAPIRE MEGLIO	2.00	.31	1.	14.	1.069	.143
16	53-DUE ANNO	2.00	.31	1.	14.	1.069	.143
17	56-DUE VOLTE	2.00	.31	1.	14.	1.069	.143
18	170-UNO STUDIO	2.00	.35	1.	16.	.989	.161
19	36-DOVERE ANDARE	2.00	.39	1.	18.	.917	.179
20	193-VOLERE SAPERE	2.00	.55	1.	25.	.705	.240
21	124-POTER TROVARE	2.00	.68	1.	31.	.556	.289
22	112-POTER CAPIRE	2.00	.70	1.	32.	.533	.297
23	78-GIOVANNI PAOLO	2.00	.79	1.	36.	.447	.327
24	39-DOVERE DIRE	2.00	.81	1.	37.	.427	.335
25	27-CHIEDERE SCUSA	2.00	.85	1.	39.	.387	.349
26	66-FARE PARTE	2.00	.92	1.	42.	.331	.370

10	167-UNO SCHIANTO	.00	1.37	0.	63.	-.003	.499
9	157-SPIRITO SANTO	.00	1.44	0.	66.	-.044	.482
8	190-VOLERE FARE	.00	1.68	0.	77.	-.185	.427
7	40-DOVERE FARE	.00	1.79	0.	82.	-.244	.403
6	76-GESÙ CRISTO	.00	1.94	0.	89.	-.324	.373
5	127-PRE MORTE	.00	2.03	0.	93.	-.367	.357
4	110-PECCATO ORIGINALE	.00	2.66	0.	122.	-.651	.258
3	116-POTER DIRE	.00	2.73	0.	125.	-.678	.249
2	28-CHIESA CATTOLICO	.00	3.10	0.	142.	-.822	.205
1	188-VOLERE DIRE	.00	3.69	0.	169.	-.1031	.151

TEXTE NUMERO 4 RELIGIONI

LIBELLE DE LA FORME GRAPHIQUE	---POURCENTAGE---		FREQUENCE		V.TEST	PROBA	
	INTERNE	GLOBAL	INTERNE	GLOBALE			
1	7-BASE COMUNE	3.58	.26	12.	12.	7.566	.000
2	135-SAN GIOVANNI	3.28	.37	11.	17.	5.864	.000
3	166-TRE GIORNO	2.99	.37	10.	17.	5.339	.000
4	148-SITO INTERNET	2.99	.57	10.	26.	4.347	.000
5	105-OTTOBRE 1999	2.09	.33	7.	15.	3.942	.000
6	29-CHIESA CRISTIANO	1.79	.24	6.	11.	3.894	.000
7	169-UNO STATO	3.88	1.20	13.	55.	3.679	.000
8	86-LIBERTÀ RELIGIOSO	2.69	.63	9.	29.	3.625	.000
9	136-SAN PAOLO	2.69	.74	9.	34.	3.265	.001
10	28-CHIESA CATTOLICO	5.97	3.10	20.	142.	2.736	.003
11	76-GESÙ CRISTO	4.18	1.94	14.	89.	2.596	.005
12	34-CRISTO STORICO	1.19	.28	4.	13.	2.263	.012
13	125-POTER VEDERE	1.79	.65	6.	30.	2.068	.019
14	15-BUONO LETTURA	.90	.24	3.	11.	1.738	.041
15	171-UOMO DIO	.90	.24	3.	11.	1.738	.041
16	78-GIOVANNI PAOLO	1.79	.79	6.	36.	1.706	.044
17	137-SAN PIETRO	1.19	.46	4.	21.	1.535	.062
18	182-VITA SPIRITUALE	.90	.28	3.	13.	1.522	.064
19	68-FARE RIFERIMENTO	.90	.28	3.	13.	1.522	.064
20	73-FEDE CRISTIANO	.90	.33	3.	15.	1.332	.091
21	181-VITA ETERNA	1.19	.52	4.	24.	1.320	.093
22	157-SPIRITO SANTO	2.39	1.44	8.	66.	1.248	.106
23	79-GRANDE PARTE	.90	.37	3.	17.	1.161	.123
24	96-NATURA DIVINA	.90	.37	3.	17.	1.161	.123
25	104-NUOVO TESTAMENTO	.90	.37	3.	17.	1.161	.123
26	36-DOVERE ANDARE	.90	.39	3.	18.	1.081	.140
27	18-BUONO VOLONTÀ	.90	.41	3.	19.	1.004	.158
28	195-VOLERE VEDERE	.90	.41	3.	19.	1.004	.158
29	62-FARE MALE	.90	.44	3.	20.	.931	.176

30	160-TANTO TEMPO	.60	.24	2.	11.	.879	.190
31	55-DUE PERSONA	.60	.24	2.	11.	.879	.190
32	129-PRIMA VOLTA	.90	.46	3.	21.	.860	.195
33	113-POTER CONTINUARE	.60	.26	2.	12.	.782	.217
34	151-SOLO DIO	.60	.26	2.	12.	.782	.217
35	3-ANNO PRIMA	.60	.28	2.	13.	.690	.245
36	63-FARE NASCERE	.60	.28	2.	13.	.690	.245
37	131-RELIGIONE CATTOLICO	.60	.31	2.	14.	.604	.273
38	20-CAPIRE MEGLIO	.60	.31	2.	14.	.604	.273
39	53-DUE ANNO	.60	.31	2.	14.	.604	.273
40	108-PADRE PIO	.60	.31	2.	14.	.604	.273
41	56-DUE VOLTE	.60	.31	2.	14.	.604	.273
42	32-CORPO FISICO	.60	.33	2.	15.	.522	.301
43	72-FEDE CATTOLICO	.60	.35	2.	16.	.444	.329
44	49-DIO PADRE	.60	.39	2.	18.	.297	.383
45	77-GESÙ DIRE	.60	.39	2.	18.	.297	.383
46	118-POTER DIVENTARE	.60	.39	2.	18.	.297	.383
47	178-VERO DIO	.90	.68	3.	31.	.259	.398
48	146-SETTEMBRE 1999	.60	.41	2.	19.	.229	.410
49	52-DIRITTO UMANO	.60	.46	2.	21.	.098	.461
50	69-FARE SOLO	.60	.48	2.	22.	.036	.486
50	141-SEMPLICE FARE	.00	.31	0.	14.	-.399	.345
49	75-GENERE UMANO	.00	.31	0.	14.	-.399	.345
48	183-VOLERE ANDARE	.00	.31	0.	14.	-.399	.345
47	165-TRADUZIONE LETTERALE	.00	.31	0.	14.	-.399	.345
46	152-SOLO DUE	.00	.31	0.	14.	-.399	.345
45	143-SENTIRE PARLARE	.30	.68	1.	31.	-.450	.327
44	16-BUONO NOTIZIA	.00	.33	0.	15.	-.469	.320
43	139-SELEZIONE NATURALE	.00	.33	0.	15.	-.469	.320
42	191-VOLERE METTERE	.00	.33	0.	15.	-.469	.320
41	6-BAMBINO MORTI SENZA BATTESIMO	.00	.33	0.	15.	-.469	.320
40	176-VEDERE SOPRA	.00	.33	0.	15.	-.469	.320
39	186-VOLERE CREDERE	.00	.33	0.	15.	-.469	.320
38	112-POTER CAPIRE	.30	.70	1.	32.	-.497	.309
37	188-VOLERE DIRE	2.99	3.69	10.	169.	-.530	.298
36	95-MORTI SENZA BATTESIMO	.00	.35	0.	16.	-.535	.296
35	90-MENO MALE	.00	.35	0.	16.	-.535	.296
34	5-BAMBINO MORTI	.00	.35	0.	16.	-.535	.296
33	126-POTER VOLARE	.30	.72	1.	33.	-.544	.293
32	83-IPOTESI DIVINA	.00	.37	0.	17.	-.599	.275
31	164-TESTI GRECO	.00	.37	0.	17.	-.599	.275
30	2-AMARE DARE	.00	.39	0.	18.	-.661	.254
29	44-DIRE NULLA	.00	.39	0.	18.	-.661	.254
28	144-SENZA BATTESIMO	.00	.41	0.	19.	-.720	.236
27	39-DOVERE DIRE	.30	.81	1.	37.	-.722	.235
26	179-VIA DIRE	.00	.44	0.	20.	-.778	.218
25	25-CERTO PUNTI	.00	.44	0.	20.	-.778	.218
24	85-LIBERO ARBITRIO	.60	1.22	2.	56.	-.801	.212
23	103-NUOVO MONDO	.00	.46	0.	21.	-.833	.202
22	174-VENIRE FARE	.00	.46	0.	21.	-.833	.202
21	175-VEDERE SOLO	.00	.46	0.	21.	-.833	.202
20	185-VOLERE CAPIRE	.00	.46	0.	21.	-.833	.202
19	60-FARE CAPIRE	.30	.87	1.	40.	-.848	.198
18	155-SPAZIO TEMPO	.00	.48	0.	22.	-.887	.188
17	14-BUONO FEDE	.30	.89	1.	41.	-.889	.187
16	66-FARE PARTE	.30	.92	1.	42.	-.929	.177
15	71-FARE VEDERE	.00	.52	0.	24.	-.990	.161
14	45-DIRE OGGI	.00	.55	0.	25.	-1.040	.149
13	40-DOVERE FARE	.90	1.79	3.	82.	-1.084	.139
12	133-RENDERE CONTO	.00	.57	0.	26.	-1.089	.138
11	19-CAPIRE BENE	.00	.59	0.	27.	-1.136	.128
10	116-POTER DIRE	1.49	2.73	5.	125.	-1.307	.096
9	59-FARE BENE	.00	.68	0.	31.	-1.315	.094
8	124-POTER TROVARE	.00	.68	0.	31.	-1.315	.094
7	35-CULTURA GENERALE	.00	.72	0.	33.	-1.399	.081
6	190-VOLERE FARE	.60	1.68	2.	77.	-1.466	.071
5	119-POTER FARE	1.49	2.88	5.	132.	-1.469	.071
4	89-MAGGIORE PARTE	.00	.92	0.	42.	-1.743	.041
3	161-TEORIA SCIENTIFICO	.00	.92	0.	42.	-1.743	.041
2	167-UNO SCHIANTO	.00	1.37	0.	63.	-2.416	.008
1	127-PRE MORTE	.30	2.03	1.	93.	-2.472	.007

- **Procedura APLUM (parole)**

L'APLUM ha permesso l'analisi delle corrispondenze semplici "parole x testi" nonché il calcolo di alcuni valori utili alla sua interpretazione quali la massa, i contributi assoluti e relativi. Nell'Appendice F sono riportati i parametri adottati per l'elaborazione.

Si riporta a pag.116 un passo dell'output dello Spad-T®

- **Procedura APLUM (segmenti)**

L'APLUM applicata ai *segmenti* ha consentito l'analisi delle corrispondenze semplici "*segmenti* x testi".

Il discorso è il medesimo di quello già visto per l'APLUM applicata alle parole.

A pag.126 si riporta una selezione dell'output della procedura.

# Output procedura APLUMM

EDITION DES COORDONNEES ET CONTRIBUTIONS DES LIGNES

	MOTS OU SEGMENTS DE MOTS	MASSES DISTO2 *	COORDONNEES	* CONTR.ABSOLUES *	COSINUS CARRES *
1999		.001 .29 *	.37 -.28 .28 *	.08 .05 .07 *	.47 .27 .26 *
2000		.000 .09 *	.02 .28 .07 *	.00 .04 .00 *	.00 .94 .05 *
ACCETTARE		.001 .03 *	-.16 -.03 -.01 *	.02 .00 .00 *	.96 .03 .00 *
ACCORDO		.001 .07 *	-.16 -.07 -.20 *	.04 .01 .11 *	.36 .06 .58 *
ACQUA		.001 .05 *	-.13 -.18 .07 *	.01 .02 .01 *	.33 .57 .10 *
ADAMO		.001 1.99 *	.04 .10 1.41 *	.00 .01 2.87 *	.00 .00 .99 *
AFFERMARE		.001 .19 *	.29 -.31 -.09 *	.09 .12 .01 *	.45 .50 .04 *
AFFERMAZIONE		.002 .14 *	-.20 -.29 -.12 *	.07 .16 .04 *	.30 .61 .10 *
AIUTARE		.000 .08 *	.15 .01 .24 *	.01 .00 .04 *	.28 .00 .72 *
AIUTO		.000 .05 *	.07 .15 .16 *	.00 .01 .02 *	.09 .42 .49 *
ALBERO		.001 2.13 *	.21 .41 1.38 *	.03 .12 2.04 *	.02 .08 .90 *
ALTO		.000 .10 *	.05 .08 .31 *	.00 .00 .08 *	.02 .07 .91 *
AMARE		.001 .11 *	-.01 .28 -.19 *	.00 .10 .07 *	.00 .69 .31 *
AMBITO		.000 .00 *	-.03 -.04 -.04 *	.00 .00 .00 *	.21 .36 .42 *
AMICO		.001 .01 *	.04 .10 -.01 *	.00 .02 .00 *	.10 .88 .01 *
AMORE		.004 .18 *	.08 .35 .22 *	.03 .52 .32 *	.04 .68 .28 *
ANALISI		.001 .32 *	-.01 -.52 .21 *	.00 .19 .05 *	.00 .86 .14 *
ANDARE		.001 .04 *	.09 .17 -.04 *	.01 .04 .00 *	.19 .76 .04 *
ANGELO		.001 .59 *	.29 .58 .41 *	.12 .54 .42 *	.14 .57 .29 *
ANIMA		.001 .25 *	.04 .28 .42 *	.00 .10 .33 *	.01 .31 .69 *
ANNO		.005 .06 *	.16 .04 .17 *	.13 .01 .25 *	.45 .03 .52 *
APERTO		.001 .06 *	.02 -.24 -.08 *	.00 .04 .01 *	.01 .88 .11 *
APOCALISSE		.000 .57 *	.36 .64 -.20 *	.04 .15 .02 *	.22 .71 .07 *
APOSTOLO		.001 .60 *	.21 .70 -.27 *	.03 .42 .09 *	.07 .81 .12 *
ARBITRIO		.000 .94 *	-.56 -.79 .04 *	.14 .30 .00 *	.34 .66 .00 *
ARGOMENTO		.002 .02 *	-.02 -.12 -.09 *	.00 .04 .03 *	.01 .62 .37 *
ARRIVARE		.000 .20 *	.02 .24 .38 *	.00 .03 .10 *	.00 .28 .72 *
ARTE		.000 .32 *	-.37 -.42 -.03 *	.06 .08 .00 *	.44 .56 .00 *
ARTICOLO		.000 .28 *	.06 -.53 .06 *	.01 .11 .00 *	.01 .97 .01 *
ASPETTO		.002 .10 *	.07 .29 -.09 *	.01 .15 .02 *	.06 .85 .09 *
ASSOCIAZIONE		.000 .61 *	.61 -.47 .07 *	.18 .12 .00 *	.62 .37 .01 *
ASSOLUTA		.000 .23 *	.12 -.41 .23 *	.01 .07 .04 *	.06 .71 .23 *
ASSURDO		.000 .40 *	-.41 -.48 -.06 *	.07 .10 .00 *	.42 .58 .01 *
ATEO		.003 1.06 *	-.60 -.84 .00 *	1.25 2.70 .00 *	.34 .66 .00 *
ATEISMO		.001 1.23 *	-.64 -.91 -.01 *	.36 .81 .00 *	.33 .67 .00 *
ATTEGGIAMENTO		.001 .02 *	-.10 .08 -.01 *	.01 .00 .00 *	.58 .42 .00 *
ATTENZIONE		.001 1.23 *	.31 .40 .99 *	.08 1.14 1.36 *	.08 .13 .80 *
ATTESA		.000 .06 *	-.03 .23 -.08 *	.00 .02 .00 *	.01 .87 .11 *
ATTI		.000 .05 *	-.09 .08 .18 *	.00 .00 .02 *	.16 .13 .71 *
ATTIVITÀ		.000 .47 *	.05 .25 .64 *	.00 .04 .34 *	.01 .14 .86 *
ATTO		.001 .05 *	-.08 -.14 .17 *	.01 .02 .05 *	.13 .34 .53 *
ATTUALE		.001 .33 *	.51 -.22 -.15 *	.14 .03 .02 *	.78 .15 .07 *
AUTORE		.001 .06 *	-.01 .24 .00 *	.00 .07 .00 *	.00 1.00 .00 *
AUTORITÀ		.000 .20 *	.26 .04 .35 *	.04 .00 .11 *	.35 .01 .64 *
AVVENIRE		.000 .04 *	.10 .17 .02 *	.00 .01 .00 *	.25 .74 .01 *
AVVISO		.000 .13 *	.09 .24 -.25 *	.00 .03 .05 *	.07 .45 .49 *
AZIONE		.001 .11 *	-.12 -.01 .30 *	.02 .00 .18 *	.14 .00 .85 *
BAMBINO		.002 .18 *	-.28 -.32 -.01 *	.13 .18 .00 *	.44 .56 .00 *
BASA		.000 .09 *	.22 .09 -.18 *	.02 .00 .02 *	.54 .09 .38 *
BASE		.001 .10 *	-.03 .17 .26 *	.00 .04 .14 *	.01 .30 .69 *
BATTESIMO		.001 .29 *	-.36 -.40 .02 *	.10 .13 .00 *	.44 .56 .00 *
BELLO		.002 .07 *	-.15 .19 -.10 *	.05 .09 .04 *	.31 .53 .16 *
BENE		.008 .01 *	-.04 .03 -.09 *	.01 .01 .11 *	.15 .09 .75 *
BIBBIA		.002 .11 *	.04 .27 -.17 *	.00 .18 .10 *	.02 .72 .26 *
BIBLICO		.000 1.69 *	1.25 .07 -.35 *	.55 .00 .07 *	.93 .00 .07 *
BISOGNO		.001 .03 *	-.09 -.13 .04 *	.01 .03 .01 *	.32 .60 .08 *
BOCCA		.000 .09 *	-.01 .27 -.15 *	.00 .03 .02 *	.00 .76 .24 *
BRAVO		.000 .16 *	-.28 -.22 -.19 *	.03 .02 .02 *	.48 .30 .22 *
BREVE		.000 .10 *	.06 .08 .31 *	.00 .00 .07 *	.04 .06 .90 *
BUONO		.003 .00 *	.02 .04 -.02 *	.00 .01 .00 *	.15 .68 .16 *
CADERE		.000 .81 *	.15 .42 .78 *	.01 .07 .36 *	.03 .21 .76 *
CADUTA		.000 5.81 *	.42 .66 2.28 *	.10 .26 4.55 *	.03 .08 .89 *
CAMBIARE		.001 .04 *	.04 .20 .00 *	.00 .03 .00 *	.04 .96 .00 *
CAMMINO		.001 .36 *	.12 .53 -.26 *	.01 .21 .08 *	.04 .77 .19 *
CAMPO		.001 .09 *	-.09 -.22 .17 *	.01 .04 .04 *	.10 .56 .34 *
CAPACITÀ		.001 .05 *	-.19 -.10 .09 *	.02 .01 .01 *	.66 .19 .16 *
CAPIRE		.004 .07 *	-.14 -.14 -.17 *	.09 .10 .21 *	.30 .30 .40 *
CAPITOLO		.000 .18 *	.06 -.29 .31 *	.00 .03 .06 *	.02 .45 .53 *
CAPO		.000 .02 *	-.11 .07 -.04 *	.01 .00 .00 *	.68 .23 .09 *

CARATTERE	.000	.23 *	.20	.06	.44	*	.02	.00	.14 *	.17	.02	.81 *
CARATTERISTICHE	.000	.11 *	-.17	-.10	.27	*	.02	.01	.06 *	.27	.09	.64 *
CARITÀ	.001	.07 *	-.19	-.04	-.19	*	.02	.00	.04 *	.47	.03	.51 *
CARNE	.001	.41 *	.45	.34	-.29	*	.17	.11	.11 *	.51	.29	.21 *
CARO	.001	.14 *	.09	.23	-.28	*	.01	.08	.17 *	.06	.39	.55 *
CASA	.001	.07 *	.21	-.12	.12	*	.05	.02	.03 *	.62	.18	.20 *
CASO	.004	.01 *	-.05	-.03	-.10	*	.01	.01	.08 *	.20	.08	.72 *
CASINI	.000	.28 *	.13	.51	-.05	*	.01	.11	.00 *	.06	.93	.01 *
CATTOLICO	.007	.01 *	-.08	-.09	-.04	*	.05	.06	.02 *	.40	.51	.09 *
CATTOLICESIMO	.001	.01 *	.04	-.07	.05	*	.00	.00	.00 *	.17	.56	.27 *
CAUSA	.001	.29 *	.12	-.04	.52	*	.02	.00	.49 *	.05	.01	.94 *
CEI	.000	.66 *	.15	.63	-.49	*	.01	.15	.14 *	.04	.61	.36 *
CENTRO	.000	.40 *	.18	.01	.61	*	.01	.00	.29 *	.08	.00	.92 *
CERCARE	.002	.03 *	.12	.00	-.13	*	.04	.00	.07 *	.48	.00	.51 *
CERTA	.001	.02 *	-.11	.07	-.04	*	.01	.00	.00 *	.63	.27	.11 *
CERTE	.001	.11 *	-.07	.18	-.27	*	.01	.04	.13 *	.05	.29	.66 *
CERTEZZA	.000	.11 *	-.25	-.21	.04	*	.02	.02	.00 *	.59	.40	.01 *
CERTI	.001	.07 *	-.01	.16	-.20	*	.00	.03	.07 *	.00	.39	.61 *
CERTO	.004	.13 *	-.27	-.22	-.10	*	.33	.24	.08 *	.56	.37	.08 *
CERVELLO	.001	1.03 *	-.59	-.82	-.04	*	.22	.47	.00 *	.34	.66	.00 *
CHIAMA	.001	.10 *	.02	.31	-.04	*	.00	.13	.00 *	.00	.98	.02 *
CHIARO	.001	.13 *	-.05	.23	-.28	*	.00	.06	.13 *	.02	.39	.59 *
CHIEDERE	.002	.04 *	.00	.11	-.16	*	.00	.04	.11 *	.00	.33	.67 *
CHIESA	.009	.03 *	-.07	.13	-.08	*	.05	.17	.09 *	.20	.60	.21 *
CIELO	.001	.25 *	.21	.43	.13	*	.05	.22	.03 *	.18	.75	.07 *
CINA	.000	.72 *	.17	-.76	.32	*	.01	.28	.07 *	.04	.82	.14 *
CITATO	.000	.28 *	.49	-.01	-.20	*	.09	.00	.03 *	.85	.00	.15 *
CITTÀ	.001	.02 *	-.07	-.05	.10	*	.00	.00	.01 *	.28	.13	.59 *
CIVILE	.000	.03 *	-.15	-.03	.08	*	.01	.00	.00 *	.78	.03	.19 *
CLERO	.000	.20 *	-.32	-.31	-.05	*	.04	.04	.00 *	.50	.49	.01 *
COERENZA	.000	.06 *	-.18	-.09	.13	*	.02	.00	.01 *	.58	.15	.27 *
COLPA	.001	.10 *	-.20	-.07	-.24	*	.04	.00	.10 *	.39	.04	.56 *
COMMESSO	.000	.53 *	-.42	-.55	.23	*	.06	.12	.03 *	.33	.57	.10 *
COMPITO	.000	.16 *	.00	.31	-.26	*	.00	.04	.04 *	.00	.59	.41 *
COMPORTEMENTO	.001	.12 *	-.26	-.20	-.09	*	.04	.03	.01 *	.58	.35	.07 *
COMPRENDERE	.001	.24 *	.06	-.11	.47	*	.00	.01	.20 *	.02	.05	.93 *
COMUNE	.001	.12 *	.02	.13	.32	*	.00	.02	.15 *	.00	.14	.86 *
COMUNIONE	.000	.55 *	.18	.66	-.30	*	.01	.16	.05 *	.06	.78	.16 *
COMUNITÀ	.001	.09 *	.27	.14	.05	*	.05	.01	.00 *	.78	.20	.02 *
CONCETTO	.002	.25 *	.17	-.47	-.05	*	.05	.42	.01 *	.11	.88	.01 *
CONCEZIONE	.000	.14 *	-.02	.27	-.26	*	.00	.03	.05 *	.00	.52	.48 *
CONCLUSIONE	.001	.05 *	-.13	.04	-.16	*	.01	.00	.03 *	.39	.04	.57 *
CONDIZIONE	.001	.03 *	-.09	.08	.14	*	.01	.01	.03 *	.23	.19	.58 *
CONFRONTO	.001	.05 *	.10	-.16	.10	*	.01	.04	.02 *	.22	.56	.23 *
CONOSCERE	.003	.01 *	-.05	-.05	.08	*	.01	.01	.03 *	.22	.19	.59 *
CONOSCENZA	.001	.19 *	-.01	.00	.44	*	.00	.00	.43 *	.00	.00	1.00 *
CONSAPEVOLEZZA	.000	.33 *	-.25	-.27	.43	*	.03	.03	.12 *	.19	.23	.58 *
CONSEGUENZA	.001	.16 *	-.26	-.25	.18	*	.07	.07	.05 *	.43	.38	.19 *
CONSIDERARE	.001	.08 *	-.11	-.25	.01	*	.01	.08	.00 *	.16	.83	.00 *
CONSIDERAZIONE	.000	.14 *	-.19	-.29	-.12	*	.02	.05	.01 *	.27	.62	.11 *
CONSIGLIO	.000	.30 *	-.27	-.23	-.41	*	.03	.02	.11 *	.24	.19	.57 *
CONTA	.000	.25 *	-.34	-.35	-.09	*	.05	.05	.01 *	.47	.50	.03 *
CONTENTO	.000	.21 *	-.32	-.31	-.09	*	.04	.04	.00 *	.49	.47	.04 *
CONTESTO	.001	.45 *	.34	.45	-.37	*	.08	.14	.15 *	.26	.44	.30 *
CONTINUARE	.001	.15 *	.15	-.34	.14	*	.02	.13	.04 *	.14	.72	.13 *
CONTO	.002	.04 *	-.17	-.03	-.09	*	.05	.00	.02 *	.77	.02	.20 *
CONTRARIO	.001	.03 *	-.15	-.02	.03	*	.02	.00	.00 *	.94	.02	.04 *
CONTRO	.002	.08 *	.04	.04	-.27	*	.00	.00	.22 *	.02	.02	.96 *
CONVERSIONE	.000	.26 *	.23	.41	-.19	*	.02	.07	.02 *	.21	.65	.15 *
CONVINTO	.001	.17 *	-.30	-.28	-.05	*	.09	.09	.00 *	.52	.46	.01 *
CONVINZIONE	.001	.05 *	-.19	-.09	.00	*	.03	.01	.00 *	.82	.18	.00 *
COPPIA	.001	.94 *	.20	.74	-.59	*	.02	.35	.34 *	.04	.59	.37 *
CORPO	.002	.10 *	-.15	-.05	.27	*	.05	.01	.27 *	.21	.02	.76 *
CORRETTO	.000	.08 *	-.11	.11	-.24	*	.00	.00	.03 *	.15	.14	.71 *
CORSO	.001	.89 *	.88	-.24	.23	*	.46	.04	.05 *	.88	.07	.06 *
COSCIENZA	.001	.36 *	-.26	-.39	.38	*	.08	.18	.26 *	.19	.41	.40 *
CREARE	.000	.02 *	-.09	.10	-.01	*	.00	.01	.00 *	.42	.57	.00 *
CREATO	.001	.09 *	.23	.16	.08	*	.04	.02	.01 *	.64	.29	.08 *
CREATORE	.000	.19 *	.08	.38	.18	*	.00	.07	.02 *	.04	.79	.18 *
CREAZIONE	.001	.76 *	.17	.09	.85	*	.03	.01	1.15 *	.04	.01	.95 *
CREDERE	.005	.40 *	-.39	-.50	-.04	*	.88	1.52	.01 *	.39	.61	.00 *
CREDENTE	.003	.62 *	-.47	-.63	-.05	*	.68	1.34	.01 *	.36	.64	.00 *
CREDO	.004	.10 *	-.17	-.13	-.23	*	.12	.07	.35 *	.31	.17	.53 *
CRISTIANO	.005	.02 *	.10	.03	.10	*	.05	.00	.09 *	.45	.03	.51 *
CRISTIANESIMO	.001	.10 *	.17	.09	.26	*	.04	.01	.14 *	.28	.08	.64 *
CRISTO	.005	.23 *	.11	.47	.00	*	.06	1.21	.00 *	.05	.95	.00 *
CRITICA	.000	.09 *	.12	-.03	.28	*	.01	.00	.06 *	.15	.01	.84 *
CROCE	.001	.38 *	.18	.57	.14	*	.02	.26	.02 *	.08	.87	.05 *
CULTO	.000	.58 *	.25	.70	.19	*	.03	.21	.03 *	.11	.83	.06 *
CULTURA	.002	.07 *	-.06	-.25	.05	*	.01	.12	.01 *	.05	.92	.04 *

CULTURALE	.001	.00	*	-.01	.06	.02	*	.00	.00	.00	*	.02	.89	.09	*
CUORE	.001	.38	*	.18	.59	.04	*	.05	.58	.00	*	.08	.91	.00	*
CURIOSITÀ	.000	.42	*	-.30	-.56	-.11	*	.04	.14	.01	*	.22	.76	.03	*
DARE	.006	.00	*	.04	-.01	-.01	*	.01	.00	.00	*	.90	.03	.06	*
DARWIN	.000	.12	*	-.12	.10	-.30	*	.00	.00	.05	*	.12	.09	.79	*
DECISO	.000	.18	*	-.21	-.18	.32	*	.01	.01	.06	*	.25	.18	.58	*
DEFINITIVA	.000	.10	*	-.05	.21	-.23	*	.00	.02	.03	*	.03	.44	.53	*
DEFINIZIONE	.001	.54	*	-.46	-.56	-.14	*	.17	.28	.03	*	.39	.58	.04	*
DEMOCRAZIA	.000	.81	*	-.54	-.72	-.02	*	.11	.22	.00	*	.36	.64	.00	*
DERIVA	.000	.06	*	-.06	.19	-.16	*	.00	.01	.02	*	.05	.55	.39	*
DESIDERIO	.000	.41	*	.00	.16	.62	*	.00	.01	.29	*	.00	.06	.94	*
DETTO	.004	.06	*	-.13	-.04	-.20	*	.07	.01	.27	*	.28	.03	.69	*
DOVERE	.010	.01	*	-.07	-.01	-.05	*	.05	.00	.04	*	.66	.03	.32	*
DEVOZIONE	.000	.85	*	.26	.80	-.38	*	.02	.27	.09	*	.08	.76	.17	*
DIALOGO	.001	.22	*	.37	-.29	.04	*	.10	.07	.00	*	.61	.38	.01	*
DIRE	.019	.04	*	-.10	.04	-.18	*	.21	.04	1.10	*	.23	.04	.73	*
DIECI	.000	.16	*	-.12	-.02	.38	*	.01	.00	.08	*	.09	.00	.91	*
DIFFERENZA	.001	.07	*	-.08	.07	-.24	*	.01	.01	.13	*	.10	.07	.82	*
DIFFICILE	.001	.04	*	.04	-.19	-.03	*	.00	.04	.00	*	.05	.92	.03	*
DIFFICOLTÀ	.000	.40	*	.14	.45	.43	*	.01	.08	.11	*	.05	.50	.46	*
DIMENSIONE	.000	.08	*	-.16	-.07	.22	*	.01	.00	.03	*	.33	.06	.61	*
DIMOSTRARE	.001	.08	*	-.23	-.14	-.10	*	.08	.03	.02	*	.64	.24	.12	*
DIO	.018	.02	*	-.11	.05	.09	*	.22	.06	.25	*	.52	.12	.36	*
DIOS	.000	1.40	*	-.67	-.97	.01	*	.23	.52	.00	*	.32	.68	.00	*
DIPENDE	.000	.20	*	-.13	-.43	.00	*	.01	.10	.00	*	.08	.92	.00	*
DIRETTA	.000	.06	*	-.02	.22	.09	*	.00	.02	.01	*	.00	.84	.16	*
DIREZIONE	.000	.16	*	.04	-.04	.40	*	.00	.00	.09	*	.01	.01	.98	*
DIRITTO	.002	.04	*	.19	-.07	-.03	*	.06	.01	.00	*	.86	.12	.03	*
DISCORSO	.002	.05	*	-.13	.06	-.17	*	.03	.01	.09	*	.34	.07	.59	*
DISCUSSIONE	.001	.07	*	.26	.01	.02	*	.08	.00	.00	*	.99	.00	.01	*
DISCUTERE	.001	.13	*	-.13	-.32	-.11	*	.01	.07	.01	*	.14	.76	.10	*
DIVENTA	.001	.03	*	-.06	.14	.08	*	.00	.01	.01	*	.13	.65	.21	*
DIVENTARE	.001	.02	*	-.11	.05	.08	*	.01	.00	.01	*	.58	.13	.29	*
DIVERSA	.002	.03	*	-.06	-.15	.02	*	.01	.05	.00	*	.15	.84	.01	*
DIVERSO	.002	.02	*	.04	-.13	.07	*	.00	.03	.02	*	.06	.71	.23	*
DIVINA	.001	.05	*	-.09	.07	.18	*	.01	.01	.07	*	.18	.11	.72	*
DIVINITÀ	.001	1.53	*	1.04	-.66	.11	*	.61	.26	.01	*	.71	.28	.01	*
DNA	.001	.55	*	.12	.57	-.45	*	.01	.21	.20	*	.03	.60	.38	*
DOGMA	.002	.31	*	-.37	-.39	-.16	*	.23	.28	.07	*	.44	.49	.08	*
DOLORE	.001	.41	*	.17	.60	-.12	*	.02	.32	.02	*	.07	.89	.04	*
DOMANDA	.002	.10	*	.05	-.30	-.08	*	.01	.21	.02	*	.03	.91	.07	*
DOMANI	.000	.16	*	-.01	.29	-.27	*	.00	.04	.05	*	.00	.55	.45	*
DONNA	.002	.04	*	-.17	-.03	-.08	*	.07	.00	.03	*	.79	.03	.19	*
DONO	.000	.05	*	-.09	.13	-.15	*	.00	.01	.02	*	.15	.36	.49	*
DOTTRINA	.001	.04	*	-.06	.12	.16	*	.00	.02	.05	*	.09	.33	.57	*
DUBBIO	.002	.04	*	.08	.09	-.17	*	.01	.01	.08	*	.13	.17	.70	*
DUE	.004	.00	*	-.03	-.02	.02	*	.00	.00	.00	*	.65	.17	.19	*
DEUAMILA	.000	.32	*	.29	-.48	-.08	*	.04	.11	.01	*	.26	.72	.02	*
EBRAICO	.001	7.10	*	2.44	-1.07	-.09	*	5.86	1.23	.01	*	.84	.16	.00	*
EBREI	.001	1.03	*	.75	-.53	.42	*	.82	.44	.43	*	.55	.27	.17	*
EDUCAZIONE	.000	.21	*	-.30	-.32	.13	*	.03	.04	.01	*	.44	.48	.08	*
EFFETTO	.001	.08	*	-.11	-.25	.04	*	.02	.10	.00	*	.16	.82	.02	*
EFISIO	.001	1.17	*	.26	.85	-.63	*	.04	.49	.41	*	.06	.61	.33	*
EGLI	.002	.07	*	-.03	.18	.19	*	.00	.06	.11	*	.01	.46	.52	*
ELEMENTI	.000	.11	*	-.18	-.12	.25	*	.01	.01	.04	*	.31	.12	.57	*
ENORME	.000	.05	*	.16	-.06	.15	*	.01	.00	.01	*	.50	.08	.42	*
ENTRA	.001	.08	*	-.23	-.15	-.03	*	.03	.02	.00	*	.69	.30	.01	*
ENTRAMBI	.000	.08	*	-.01	.22	.18	*	.00	.02	.02	*	.00	.59	.41	*
ENTRARE	.001	.03	*	-.06	.16	-.02	*	.00	.02	.00	*	.10	.88	.02	*
EPOCA	.000	.17	*	-.25	-.23	.23	*	.03	.03	.04	*	.36	.31	.33	*
ERRORI	.001	.08	*	.26	-.07	-.02	*	.10	.01	.00	*	.93	.07	.00	*
ESEMPIO	.004	.00	*	-.03	.02	-.02	*	.00	.00	.00	*	.63	.13	.24	*
ESISTERE	.004	.29	*	-.28	-.46	-.05	*	.35	1.00	.02	*	.27	.72	.01	*
ESISTENZA	.002	.25	*	-.20	-.45	.09	*	.07	.37	.02	*	.16	.81	.03	*
ESPERIENZA	.003	.36	*	-.38	-.45	.10	*	.46	.71	.06	*	.40	.57	.03	*
ESPRESSIONE	.000	.29	*	.08	.46	-.26	*	.00	.11	.06	*	.02	.74	.24	*
ETERNA	.000	.04	*	-.15	-.04	.11	*	.01	.00	.01	*	.65	.04	.31	*
ETICA	.000	.11	*	-.20	-.07	-.25	*	.01	.00	.04	*	.39	.04	.57	*
EVA	.001	3.49	*	.20	.32	1.83	*	.03	.09	4.60	*	.01	.03	.96	*
EVENTO	.001	.09	*	-.24	-.17	.01	*	.05	.03	.00	*	.65	.35	.00	*
EVIDENTE	.000	.05	*	-.12	-.15	-.13	*	.01	.01	.01	*	.27	.44	.30	*
EVIDENZA	.000	.12	*	-.13	.09	-.32	*	.01	.00	.09	*	.13	.06	.81	*
EVITARE	.000	.02	*	.04	.14	-.06	*	.00	.01	.00	*	.06	.77	.17	*
EVOLUZIONE	.001	.27	*	.01	.34	-.39	*	.00	.15	.30	*	.00	.43	.57	*
EVOLUZIONISMO	.000	.14	*	.03	.34	-.17	*	.00	.05	.02	*	.00	.80	.19	*
FARE	.028	.04	*	-.12	-.05	-.16	*	.42	.08	1.30	*	.33	.06	.61	*
FACILE	.001	.14	*	.29	-.20	-.12	*	.05	.03	.02	*	.59	.30	.11	*
FAMIGLIA	.001	.03	*	.00	.16	.03	*	.00	.03	.00	*	.00	.98	.02	*
FAVORE	.001	.16	*	-.25	-.30	-.11	*	.05	.08	.02	*	.37	.55	.08	*
FEDE	.006	.03	*	-.13	.00	-.13	*	.10	.00	.18	*	.47	.00	.53	*

FEDELI	.000	.28	*	.09	.47	-.22	*	.00	.10	.03	*	.03	.80	.17	*
FELICE	.001	.35	*	-.38	-.45	.07	*	.08	.12	.00	*	.41	.58	.01	*
FELICITÀ	.001	.02	*	-.10	.07	.03	*	.01	.00	.00	*	.60	.33	.07	*
FENOMENO	.001	.14	*	-.28	-.24	-.09	*	.09	.07	.01	*	.55	.40	.05	*
FESTA	.000	.51	*	.31	.51	.39	*	.04	.12	.11	*	.19	.51	.30	*
FIGLIO	.003	.01	*	-.09	.04	.01	*	.03	.01	.00	*	.78	.21	.01	*
FIGURA	.000	.13	*	-.05	.11	.34	*	.00	.01	.10	*	.02	.10	.88	*
FILM	.001	.65	*	-.02	-.66	.46	*	.00	.33	.25	*	.00	.67	.33	*
FILOSOFIA	.001	.27	*	-.35	-.38	.04	*	.07	.09	.00	*	.45	.55	.01	*
FINALMENTE	.000	.05	*	.17	-.12	.05	*	.01	.01	.00	*	.63	.31	.06	*
FINE	.002	.11	*	.28	.07	.17	*	.16	.01	.10	*	.69	.04	.26	*
FINISCE	.000	.03	*	-.09	.12	-.11	*	.00	.01	.01	*	.22	.42	.35	*
FISICA	.001	.21	*	-.13	-.22	.38	*	.01	.03	.15	*	.08	.23	.69	*
FISICO	.001	.37	*	-.38	-.46	.10	*	.09	.14	.01	*	.40	.58	.03	*
FONDAMENTALE	.000	.05	*	-.11	.09	-.18	*	.01	.00	.03	*	.24	.15	.61	*
FONDO	.001	.13	*	.03	.24	-.26	*	.00	.06	.11	*	.01	.46	.54	*
FONTE	.000	.10	*	.28	.15	-.06	*	.04	.01	.00	*	.75	.21	.04	*
FONTI	.000	.15	*	.05	.37	-.09	*	.00	.05	.01	*	.01	.92	.06	*
FORMA	.003	.01	*	-.02	.07	.02	*	.00	.01	.00	*	.04	.89	.06	*
FORSE	.004	.05	*	.12	.02	-.18	*	.06	.00	.22	*	.29	.01	.70	*
FORTE	.001	.07	*	-.06	.12	.24	*	.00	.01	.05	*	.05	.19	.76	*
FORZA	.001	.18	*	.03	.19	.38	*	.00	.06	.37	*	.01	.20	.80	*
FOTO	.000	.81	*	-.53	-.72	.06	*	.13	.27	.00	*	.35	.65	.00	*
FRASE	.001	.24	*	.21	.28	-.34	*	.05	.09	.21	*	.18	.33	.49	*
FRATELLO	.001	.45	*	.25	.60	-.15	*	.09	.59	.05	*	.14	.81	.05	*
FRONTE	.001	.00	*	.00	.05	.04	*	.00	.00	.00	*	.00	.59	.41	*
FRUTTO	.001	1.13	*	.11	.30	1.01	*	.01	.06	1.09	*	.01	.08	.91	*
FUNZIONE	.000	.39	*	.12	.54	-.29	*	.01	.16	.07	*	.04	.74	.22	*
FUTURO	.001	.15	*	-.13	-.04	.37	*	.01	.00	.14	*	.12	.01	.87	*
GENERALE	.001	.16	*	-.29	-.28	.01	*	.09	.09	.00	*	.53	.47	.00	*
GENERE	.001	.07	*	-.16	-.21	-.06	*	.03	.06	.01	*	.33	.63	.04	*
GENESI	.001	.37	*	-.25	-.44	.34	*	.03	.12	.11	*	.16	.52	.31	*
GENITORI	.000	.05	*	-.20	-.09	-.07	*	.02	.00	.00	*	.77	.15	.08	*
GENTE	.001	.04	*	-.16	-.11	-.02	*	.04	.02	.00	*	.68	.32	.01	*
GEOVA	.001	1.09	*	.24	.81	-.61	*	.06	.77	.67	*	.05	.60	.35	*
GERUSALEMME	.000	.57	*	.56	.27	.43	*	.11	.03	.11	*	.55	.13	.32	*
GESÙ	.006	.36	*	.22	.54	-.14	*	.32	2.18	.21	*	.13	.82	.05	*
GIOCO	.000	.04	*	.06	.11	-.17	*	.00	.01	.02	*	.07	.28	.65	*
GIOIA	.000	.40	*	.13	.42	.46	*	.01	.10	.17	*	.04	.44	.52	*
GIORNALI	.001	.66	*	-.42	-.70	.01	*	.10	.31	.00	*	.26	.74	.00	*
GIORNO	.003	.23	*	.11	.27	.38	*	.04	.28	.84	*	.05	.32	.64	*
GIOVANI	.000	.08	*	.01	.26	.09	*	.00	.03	.00	*	.00	.89	.11	*
GIOVANNI	.001	.44	*	.23	.56	.27	*	.07	.46	.16	*	.12	.72	.17	*
GIRO	.001	.02	*	-.09	.08	.08	*	.00	.00	.01	*	.38	.31	.31	*
GIUBILEO	.001	.22	*	-.02	.30	-.37	*	.00	.06	.14	*	.00	.40	.60	*
GIUDICARE	.000	.12	*	-.08	.16	-.29	*	.00	.01	.06	*	.06	.23	.71	*
GIUDIZIO	.001	.04	*	-.14	.03	-.14	*	.02	.00	.03	*	.49	.03	.48	*
GIUSTA	.000	.02	*	-.11	.07	.01	*	.01	.00	.00	*	.70	.30	.00	*
GIUSTIZIA	.001	.09	*	-.24	-.18	-.01	*	.03	.02	.00	*	.65	.35	.00	*
GIUSTO	.001	.03	*	-.09	.04	-.14	*	.01	.00	.04	*	.25	.06	.69	*
GOVERNO	.001	.85	*	.72	-.35	.46	*	.31	.08	.21	*	.60	.15	.25	*
GRADO	.001	.03	*	-.07	-.16	-.01	*	.01	.04	.00	*	.16	.84	.00	*
GRANDE	.004	.08	*	.11	.21	.16	*	.05	.20	.16	*	.14	.56	.30	*
GRAVE	.000	.10	*	.19	-.16	-.19	*	.02	.01	.03	*	.36	.26	.38	*
GRAZIA	.000	.11	*	.03	.29	.14	*	.00	.04	.01	*	.01	.79	.20	*
GRAZIE	.002	.18	*	.39	-.06	-.17	*	.30	.01	.09	*	.82	.02	.16	*
GRECO	.001	.68	*	.18	.67	-.44	*	.03	.45	.30	*	.05	.66	.29	*
GRUPPO	.002	.22	*	.28	-.20	.31	*	.15	.09	.32	*	.36	.18	.45	*
GUARDARE	.001	.14	*	-.28	-.25	-.05	*	.12	.10	.01	*	.56	.43	.02	*
GUERRA	.001	.26	*	.40	-.30	-.03	*	.11	.07	.00	*	.64	.36	.00	*
HOME	.000	2.34	*	1.08	-1.09	.00	*	.47	.52	.00	*	.50	.50	.00	*
IDEA	.003	.07	*	-.08	-.25	-.02	*	.02	.21	.00	*	.10	.89	.01	*
IERI	.000	.14	*	.20	.31	-.08	*	.01	.04	.00	*	.28	.68	.04	*
IGNORANZA	.000	.25	*	.46	-.10	-.17	*	.08	.00	.02	*	.84	.04	.12	*
IMHO	.001	.19	*	.26	.02	-.35	*	.05	.00	.14	*	.36	.00	.64	*
IMMAGINE	.001	.09	*	-.22	-.16	.11	*	.05	.03	.02	*	.56	.30	.14	*
IMPEGNO	.000	.33	*	.15	.51	.21	*	.01	.12	.03	*	.07	.79	.14	*
IMPLICA	.000	.22	*	-.06	-.43	-.18	*	.00	.07	.02	*	.02	.83	.15	*
IMPORTA	.000	.13	*	.17	-.30	.09	*	.01	.04	.00	*	.23	.71	.06	*
IMPORTANTE	.002	.02	*	-.05	-.11	.08	*	.00	.02	.02	*	.10	.60	.30	*
IMPORTANZA	.001	.19	*	.17	.40	.04	*	.02	.12	.00	*	.15	.84	.01	*
IMPOSSIBILE	.001	.43	*	-.41	-.51	.11	*	.16	.26	.02	*	.38	.59	.03	*
IMPROBABILE	.001	.93	*	-.57	-.77	-.06	*	.18	.36	.00	*	.35	.65	.00	*
INCONTRO	.001	.54	*	.40	.34	.51	*	.09	.08	.25	*	.30	.22	.48	*
INDIRIZZO	.001	.58	*	.28	-.67	.23	*	.06	.35	.06	*	.14	.77	.09	*
INDIVIDUO	.000	.19	*	-.19	-.39	-.04	*	.01	.06	.00	*	.19	.81	.01	*
INFERNO	.001	.05	*	-.20	-.11	.02	*	.03	.01	.00	*	.76	.23	.01	*
INFINITO	.001	.40	*	-.40	-.48	.01	*	.10	.16	.00	*	.41	.59	.00	*
INFORMAZIONE	.001	.24	*	.32	-.37	.04	*	.08	.11	.00	*	.43	.56	.01	*
INGLESE	.000	.53	*	.39	.55	-.29	*	.07	.16	.06	*	.28	.57	.15	*

INIZIO	.001	.02 *	.08	.12	-.04	*	.01	.01	.00 *	.32	.62	.06 *
INSEGNA	.000	.14 *	.26	.22	-.14	*	.03	.02	.01 *	.48	.37	.15 *
INSEGNAMENTI	.000	.81 *	.79	.34	.27	*	.21	.04	.04 *	.76	.14	.09 *
INSIEME	.001	.11 *	.32	.06	.05	*	.13	.01	.01 *	.94	.04	.03 *
INTENDERE	.001	.06 *	-.12	-.02	-.21	*	.01	.00	.08 *	.23	.01	.76 *
INTERA	.000	.16 *	-.02	.16	.37	*	.00	.01	.08 *	.00	.16	.83 *
INTERESSA	.001	.07 *	.04	-.24	-.07	*	.00	.04	.00 *	.02	.90	.08 *
INTERESSANTE	.001	.18 *	-.30	-.30	.01	*	.05	.06	.00 *	.51	.49	.00 *
INTERNET	.001	.15 *	.34	.08	.16	*	.08	.00	.03 *	.79	.04	.17 *
INTERNO	.001	.13 *	.13	.33	-.05	*	.01	.08	.00 *	.12	.86	.02 *
INTERPRETAZIONE	.001	.14 *	.34	-.01	-.16	*	.10	.00	.03 *	.82	.00	.18 *
INTERVENTO	.000	.28 *	.13	.52	-.01	*	.01	.10	.00 *	.06	.94	.00 *
INUTILE	.001	.08 *	-.23	-.17	.00	*	.04	.02	.00 *	.66	.34	.00 *
INVITO	.000	.14 *	-.04	.25	-.29	*	.00	.02	.05 *	.01	.42	.57 *
IPOTESI	.002	.38 *	-.38	-.47	-.16	*	.25	.42	.08 *	.37	.57	.07 *
ISRAELE	.001	7.32 *	2.58	-.81	.12	*	4.90	.53	.02 *	.91	.09	.00 *
ITALIA	.002	.37 *	.29	-.52	-.11	*	.14	.50	.03 *	.23	.74	.03 *
ITALIANA	.001	.05 *	.14	.09	-.15	*	.03	.01	.05 *	.40	.16	.44 *
LAMENTO	.000	1.26 *	-.64	-.92	.04	*	.18	.41	.00 *	.32	.67	.00 *
LASCIARE	.001	.01 *	-.09	-.05	.00	*	.01	.00	.00 *	.76	.23	.00 *
LATINO	.000	.44 *	.25	.52	-.33	*	.03	.13	.08 *	.14	.61	.25 *
LAVORO	.001	.20 *	.07	.11	.42	*	.01	.01	.30 *	.03	.06	.92 *
LEGGE	.002	.03 *	-.02	.17	-.02	*	.00	.06	.00 *	.01	.97	.02 *
LEGGERE	.001	.03 *	.04	-.13	-.12	*	.00	.02	.03 *	.05	.53	.42 *
LEGGI	.001	.16 *	-.15	-.37	-.03	*	.02	.14	.00 *	.14	.85	.00 *
LEI	.001	.09 *	.20	.20	.08	*	.04	.05	.01 *	.47	.45	.08 *
LETTERA	.001	.23 *	.22	.42	-.06	*	.04	.16	.01 *	.21	.77	.02 *
LETTERALE	.000	.48 *	.17	.63	-.23	*	.01	.15	.03 *	.06	.82	.11 *
LETTO	.001	.04 *	.02	.00	-.19	*	.00	.00	.08 *	.01	.00	.98 *
LETTURA	.001	.25 *	.31	.38	-.11	*	.07	.11	.01 *	.38	.58	.05 *
LIBERO	.001	.22 *	-.22	-.33	.25	*	.07	.18	.15 *	.21	.50	.29 *
LIBERTÀ	.002	.24 *	.01	-.07	.48	*	.00	.01	.78 *	.00	.02	.98 *
LIBRO	.003	.18 *	.34	-.27	-.01	*	.38	.26	.00 *	.62	.38	.00 *
LIMITI	.000	.02 *	-.15	-.01	-.03	*	.01	.00	.00 *	.96	.00	.03 *
LINGUA	.000	.68 *	.77	.22	-.21	*	.31	.03	.04 *	.87	.07	.07 *
LINGUAGGIO	.000	.02 *	.11	.06	-.06	*	.01	.00	.00 *	.66	.17	.17 *
LISTA	.000	.08 *	.24	.11	.08	*	.02	.01	.00 *	.74	.17	.09 *
LITURGIA	.000	1.10 *	.26	.84	-.57	*	.03	.31	.21 *	.06	.64	.29 *
LIVELLO	.001	.01 *	-.07	.01	-.02	*	.00	.00	.00 *	.91	.03	.06 *
LOGICA	.001	.80 *	-.53	-.72	-.02	*	.33	.65	.00 *	.36	.64	.00 *
LUCA	.001	.57 *	.26	.63	-.33	*	.04	.29	.12 *	.12	.69	.19 *
LUCE	.001	.14 *	-.14	-.06	.34	*	.03	.01	.25 *	.15	.02	.82 *
LUNGO	.001	.05 *	-.03	-.03	.23	*	.00	.00	.09 *	.01	.01	.97 *
LUOGO	.001	.01 *	.01	.08	-.06	*	.00	.01	.01 *	.02	.57	.41 *
MADONNA	.001	.25 *	.04	.40	-.29	*	.00	.11	.09 *	.01	.64	.35 *
MADRE	.001	.13 *	.05	.36	.01	*	.00	.16	.00 *	.02	.98	.00 *
MAESTRO	.001	4.00 *	.48	.46	1.88	*	.14	.14	3.51 *	.06	.05	.89 *
MAGARI	.001	.10 *	.03	-.13	-.29	*	.00	.03	.22 *	.01	.16	.83 *
MAGGIORI	.001	.09 *	.29	-.01	-.06	*	.09	.00	.01 *	.95	.00	.05 *
MAGGIORANZA	.001	.24 *	-.16	-.45	.07	*	.01	.12	.00 *	.11	.87	.02 *
MALE	.003	.02 *	-.07	-.04	.10	*	.01	.01	.04 *	.33	.11	.57 *
MANCA	.000	.05 *	-.12	.08	-.18	*	.01	.00	.02 *	.27	.12	.61 *
MANCANZA	.001	.07 *	-.22	-.13	-.04	*	.03	.01	.00 *	.72	.26	.02 *
MANO	.001	.03 *	.07	.14	.03	*	.01	.03	.00 *	.19	.77	.04 *
MANIERA	.001	.14 *	.05	.37	-.08	*	.00	.10	.01 *	.02	.94	.05 *
MANIFESTO	.000	.52 *	-.03	-.57	.44	*	.00	.17	.16 *	.00	.62	.38 *
MARIA	.003	.50 *	.16	.61	-.31	*	.07	1.13	.44 *	.05	.75	.19 *
MATEMATICA	.001	.70 *	-.51	-.65	-.10	*	.22	.41	.02 *	.37	.62	.02 *
MATERIA	.001	.41 *	-.29	-.47	.32	*	.07	.19	.14 *	.21	.54	.25 *
MATRIMONIO	.000	.27 *	.10	.47	-.19	*	.00	.11	.03 *	.03	.83	.14 *
MATTEO	.000	.23 *	.09	.45	-.14	*	.00	.08	.01 *	.03	.89	.08 *
MECCANISMO	.000	.14 *	-.01	.29	-.24	*	.00	.03	.03 *	.00	.59	.41 *
MEDITAZIONE	.000	5.27 *	.42	.69	2.15	*	.09	.26	3.79 *	.03	.09	.88 *
MEGLIO	.003	.02 *	-.08	.02	-.10	*	.02	.00	.05 *	.36	.02	.62 *
MEMORIA	.000	1.50 *	1.17	-.35	-.04	*	.56	.05	.00 *	.92	.08	.00 *
MENO	.003	.04 *	-.12	-.04	-.15	*	.05	.01	.14 *	.36	.05	.59 *
MENTE	.001	.18 *	-.15	-.08	.39	*	.03	.01	.32 *	.13	.03	.84 *
MERITO	.001	.14 *	.14	.27	-.23	*	.01	.06	.07 *	.13	.51	.36 *
MESE	.000	.04 *	-.05	.15	.14	*	.00	.01	.01 *	.06	.48	.46 *
MESSA	.000	.13 *	.04	.35	-.06	*	.00	.07	.00 *	.01	.96	.03 *
MESSAGGIO	.003	.07 *	.15	-.22	-.06	*	.08	.18	.02 *	.30	.66	.04 *
METTERE	.002	.04 *	.01	.18	-.11	*	.00	.07	.04 *	.00	.72	.28 *
METODO	.001	.03 *	-.17	-.04	-.06	*	.03	.00	.01 *	.86	.04	.10 *
MEZZO	.001	.18 *	.07	.41	-.11	*	.01	.22	.02 *	.03	.91	.06 *
MIGLIAIA	.000	.19 *	-.05	-.43	.11	*	.00	.08	.01 *	.01	.93	.06 *
MIGLIORE	.001	.02 *	-.09	.10	-.07	*	.01	.01	.01 *	.37	.42	.21 *
MILANO	.000	1.08 *	.74	.05	.72	*	.21	.00	.33 *	.51	.00	.48 *
MILIARDI	.001	.24 *	-.32	-.35	.11	*	.06	.07	.01 *	.43	.51	.05 *
MILIONI	.001	.13 *	.01	-.15	.33	*	.00	.02	.11 *	.00	.18	.82 *
MILLE	.000	.09 *	.06	.02	.29	*	.00	.00	.05 *	.04	.01	.95 *

MILLENNIO	.000	.51 *	.16	.62	-.32	*	.01	.16	.06 *	.05	.75	.20 *
MINIMO	.000	.07 *	-.02	-.26	.04	*	.00	.03	.00 *	.01	.97	.02 *
MIRACOLO	.001	.32 *	-.37	-.42	.07	*	.14	.20	.01 *	.42	.56	.02 *
MISSIONE	.001	.08 *	-.07	.09	.25	*	.00	.00	.06 *	.07	.10	.83 *
MISURA	.000	.12 *	-.02	.27	-.21	*	.00	.03	.03 *	.00	.63	.37 *
MODELLO	.001	.15 *	-.28	-.24	-.10	*	.05	.04	.01 *	.54	.39	.07 *
MODERATORE	.000	.07 *	-.15	.03	-.22	*	.01	.00	.03 *	.31	.01	.68 *
MODERAZIONE	.000	.43 *	-.32	-.57	.03	*	.05	.17	.00 *	.24	.76	.00 *
MODERNO	.001	.36 *	.54	.21	-.17	*	.21	.03	.04 *	.80	.12	.08 *
MODO	.005	.00 *	.03	.04	.01	*	.01	.01	.00 *	.37	.59	.03 *
MOGLIE	.000	.29 *	.27	.44	.14	*	.03	.08	.01 *	.26	.68	.07 *
MOMENTO	.002	.07 *	-.16	-.19	.11	*	.04	.06	.03 *	.36	.48	.16 *
MONDO	.005	.02 *	.06	.03	.14	*	.02	.01	.18 *	.14	.04	.82 *
MORALE	.001	.09 *	-.24	-.16	-.09	*	.06	.03	.01 *	.63	.29	.09 *
MORIRE	.000	1.43 *	.18	.17	1.17	*	.02	.02	1.14 *	.02	.02	.96 *
MORTE	.004	.61 *	-.09	-.06	.77	*	.04	.02	4.72 *	.01	.01	.98 *
MORTI	.001	.02 *	-.11	.04	.08	*	.01	.00	.01 *	.62	.08	.30 *
MORTO	.001	.16 *	.12	-.03	.39	*	.01	.00	.19 *	.08	.00	.91 *
MOTIVO	.002	.02 *	.10	-.02	-.12	*	.02	.00	.04 *	.38	.02	.60 *
MOVIMENTO	.000	.52 *	.53	-.12	.47	*	.10	.01	.13 *	.54	.03	.43 *
MUORE	.000	.05 *	-.05	.14	.16	*	.00	.01	.02 *	.06	.42	.52 *
MUSULMANI	.000	.50 *	.27	.41	.51	*	.03	.08	.18 *	.15	.34	.51 *
MUTAZIONE	.001	1.01 *	.22	.78	-.60	*	.06	.80	.72 *	.05	.59	.36 *
NASCERE	.001	.09 *	.16	.17	.19	*	.04	.05	.09 *	.30	.31	.39 *
NASCITA	.001	1.15 *	.12	.16	1.05	*	.01	.02	1.32 *	.01	.02	.97 *
NATURA	.002	.10 *	.00	.23	.22	*	.00	.11	.15 *	.00	.52	.48 *
NATURALE	.001	.02 *	-.12	.06	-.07	*	.01	.00	.00 *	.63	.16	.21 *
NECESSARIO	.001	.01 *	.10	-.04	-.06	*	.01	.00	.01 *	.68	.10	.22 *
NECESSITÀ	.001	.04 *	-.05	.16	.10	*	.00	.02	.01 *	.06	.68	.25 *
NEGARE	.000	.02 *	.05	.10	-.07	*	.00	.00	.00 *	.14	.55	.31 *
NERO	.000	.89 *	-.55	-.76	.07	*	.15	.32	.00 *	.34	.65	.01 *
NESSUNO	.005	.05 *	.01	-.19	-.12	*	.00	.22	.13 *	.00	.72	.28 *
NEW	.000	.22 *	.20	.41	-.13	*	.02	.09	.01 *	.17	.76	.07 *
NOME	.002	.20 *	.10	.35	-.26	*	.02	.31	.27 *	.05	.61	.35 *
NOMI	.000	.12 *	-.03	.26	-.24	*	.00	.04	.05 *	.01	.52	.47 *
NORMALE	.000	.20 *	-.31	-.32	.03	*	.04	.05	.00 *	.49	.51	.01 *
NOTA	.000	.01 *	.00	.04	-.08	*	.00	.00	.01 *	.00	.20	.80 *
NOTARE	.000	.11 *	-.24	-.14	-.19	*	.03	.01	.03 *	.50	.18	.32 *
NOTIZIA	.000	.03 *	.13	.08	-.10	*	.01	.00	.01 *	.52	.18	.29 *
NOTO	.000	.32 *	-.02	-.41	.38	*	.00	.07	.09 *	.00	.54	.46 *
NULLA	.003	.05 *	-.12	-.18	.03	*	.06	.14	.01 *	.30	.67	.02 *
NUMERO	.001	.00 *	.03	.03	.00	*	.00	.00	.00 *	.49	.51	.00 *
NUOVO	.003	.11 *	.22	.10	.23	*	.16	.04	.30 *	.42	.09	.48 *
OCCHI	.001	.02 *	-.09	-.02	.09	*	.01	.00	.01 *	.49	.03	.49 *
ODIO	.000	.11 *	.24	.03	.23	*	.02	.00	.03 *	.52	.01	.47 *
OGGETTO	.000	.45 *	.16	.25	.60	*	.01	.03	.31 *	.06	.14	.80 *
OGGI	.003	.10 *	.19	.16	-.20	*	.11	.09	.22 *	.35	.24	.41 *
OLTRE	.002	.01 *	.01	-.10	-.01	*	.00	.02	.00 *	.01	.97	.02 *
OMOSESSUALI	.000	.14 *	-.18	-.02	-.33	*	.02	.00	.09 *	.24	.00	.76 *
OPERA	.002	.22 *	.08	.37	.28	*	.01	.29	.25 *	.03	.62	.36 *
OPINIONE	.001	.03 *	.00	-.16	-.05	*	.00	.03	.00 *	.00	.92	.08 *
ORA	.003	.04 *	.13	.11	-.09	*	.06	.04	.04 *	.48	.32	.20 *
ORDINE	.001	.24 *	.09	.39	.27	*	.00	.10	.07 *	.04	.65	.31 *
ORE	.000	.10 *	.13	.24	-.15	*	.01	.02	.01 *	.18	.59	.23 *
ORGANIZZAZIONE	.000	.02 *	-.13	.02	.02	*	.01	.00	.00 *	.95	.02	.03 *
ORIGINALE	.001	.05 *	-.12	.02	.19	*	.02	.00	.10 *	.27	.01	.72 *
ORIGINE	.001	.06 *	-.03	.19	.15	*	.00	.03	.03 *	.01	.60	.39 *
OSSERVAZIONE	.000	.23 *	-.33	-.31	-.16	*	.04	.04	.02 *	.46	.43	.11 *
OTTENERE	.000	.36 *	.02	-.10	.59	*	.00	.00	.21 *	.00	.03	.97 *
OTTOBRE	.000	.68 *	.48	.37	.56	*	.11	.07	.25 *	.34	.20	.46 *
PACE	.002	.36 *	.17	.55	-.15	*	.06	.65	.07 *	.08	.86	.06 *
PADRE	.003	.05 *	-.02	.23	.05	*	.00	.18	.02 *	.01	.94	.05 *
PAESE	.001	.08 *	.06	.00	.28	*	.01	.00	.18 *	.04	.00	.96 *
PAGE	.000	.16 *	-.29	-.25	-.12	*	.03	.03	.01 *	.52	.39	.09 *
PAGINA	.000	.02 *	.08	-.08	-.10	*	.00	.00	.01 *	.31	.25	.44 *
PANE	.000	.33 *	.51	.09	-.25	*	.10	.00	.04 *	.79	.03	.19 *
PAOLO	.002	.50 *	.39	.54	-.22	*	.29	.60	.15 *	.31	.59	.10 *
PAPA	.002	.01 *	-.04	.05	-.06	*	.00	.01	.01 *	.22	.30	.47 *
PARADISO	.001	.10 *	-.12	.08	-.28	*	.02	.01	.18 *	.15	.07	.78 *
PARE	.002	.15 *	.08	.17	-.34	*	.01	.07	.43 *	.04	.19	.77 *
PARERE	.000	.10 *	-.01	-.15	-.28	*	.00	.01	.06 *	.00	.23	.77 *
ARI	.000	.14 *	.21	.18	-.24	*	.02	.02	.04 *	.32	.25	.44 *
PARLARE	.006	.01 *	-.05	-.04	-.10	*	.02	.01	.11 *	.17	.11	.72 *
PAROLA	.005	.14 *	.20	.30	-.08	*	.21	.49	.06 *	.30	.64	.05 *
PARROCCHIA	.000	.04 *	-.13	.05	-.15	*	.01	.00	.01 *	.39	.06	.54 *
PARTE	.005	.00 *	.02	-.01	-.03	*	.00	.00	.01 *	.23	.12	.65 *
PARTI	.001	.02 *	-.09	-.08	.09	*	.00	.00	.01 *	.35	.31	.34 *
PARTICOLARE	.001	.05 *	-.08	.08	-.19	*	.01	.01	.09 *	.12	.13	.74 *
PARTIRE	.001	.13 *	.28	-.17	.14	*	.07	.03	.03 *	.61	.23	.15 *
PASSAGGIO	.000	.26 *	.23	.41	-.19	*	.02	.07	.02 *	.21	.65	.15 *

PASSARE	.001	.02 *	.07	-.01	-.10	*	.01	.00	.02 *	.29	.01	.70 *
PASSI	.001	.09 *	.02	.29	.01	*	.00	.05	.00 *	.00	.99	.00 *
PASSO	.001	.11 *	.11	.32	-.04	*	.01	.08	.00 *	.11	.88	.02 *
PASTORI	.000	.64 *	.19	.68	-.39	*	.01	.19	.09 *	.05	.71	.23 *
PAURA	.001	.13 *	-.20	-.15	.25	*	.04	.02	.09 *	.32	.18	.50 *
PECCATO	.003	.06 *	-.13	-.11	.19	*	.05	.04	.18 *	.25	.19	.56 *
PEGGIO	.000	.05 *	-.08	-.17	-.12	*	.00	.01	.01 *	.13	.59	.28 *
PENA	.001	.06 *	.24	-.07	.00	*	.03	.00	.00 *	.92	.08	.00 *
PENSARE	.005	.02 *	-.09	-.01	-.11	*	.04	.00	.09 *	.39	.01	.60 *
PENSIERO	.002	.06 *	.00	.01	.24	*	.00	.00	.19 *	.00	.00	1.00 *
PERDERE	.001	.13 *	.22	-.01	-.29	*	.03	.00	.08 *	.38	.00	.62 *
PERDONO	.000	.23 *	.06	.42	-.23	*	.00	.07	.03 *	.02	.76	.22 *
PERFETTO	.000	.35 *	.17	.55	.13	*	.01	.13	.01 *	.08	.87	.05 *
PERIODO	.001	.36 *	.19	-.01	.57	*	.03	.00	.45 *	.10	.00	.90 *
PERMESSO	.000	.28 *	.52	-.09	.02	*	.09	.00	.00 *	.97	.03	.00 *
PERMETTE	.000	.02 *	-.14	.01	-.02	*	.01	.00	.00 *	.98	.00	.02 *
PERSONA	.006	.08 *	-.19	-.16	.13	*	.22	.17	.17 *	.46	.33	.22 *
PERSONALE	.001	.01 *	-.10	.01	-.03	*	.02	.00	.00 *	.91	.00	.09 *
PERSONALMENTE	.001	.04 *	-.03	-.08	-.17	*	.00	.00	.04 *	.02	.17	.81 *
PESO	.000	.16 *	-.28	-.27	.06	*	.03	.03	.00 *	.51	.46	.02 *
PIACERE	.001	.04 *	-.10	-.05	-.16	*	.02	.00	.06 *	.27	.07	.66 *
PIANO	.001	.12 *	.04	.34	.05	*	.00	.09	.00 *	.02	.97	.02 *
PICCOLO	.001	.05 *	-.03	.21	.00	*	.00	.07	.00 *	.02	.98	.00 *
PIEDI	.000	.16 *	.06	.36	.15	*	.00	.05	.01 *	.03	.83	.15 *
PIENO	.001	.03 *	-.06	-.15	-.08	*	.00	.02	.01 *	.09	.71	.20 *
PIETRO	.001	.11 *	.10	.31	.03	*	.02	.17	.00 *	.10	.89	.01 *
PIO	.000	.21 *	.10	.45	.04	*	.00	.09	.00 *	.05	.94	.01 *
POCO	.003	.05 *	.18	.06	-.11	*	.11	.01	.07 *	.67	.07	.25 *
POLITICA	.001	.18 *	-.13	-.21	.34	*	.01	.03	.12 *	.10	.25	.65 *
POLITICO	.000	.05 *	-.16	-.06	.15	*	.01	.00	.02 *	.49	.06	.45 *
POPOLO	.001	.54 *	.55	.26	.42	*	.29	.07	.27 *	.55	.13	.32 *
PORTA	.001	.02 *	-.10	-.10	.01	*	.01	.01	.00 *	.54	.45	.01 *
PORTARE	.001	.03 *	-.06	.14	.08	*	.00	.01	.01 *	.13	.67	.20 *
POSIZIONE	.001	.15 *	.08	.29	.24	*	.01	.12	.12 *	.05	.57	.38 *
POTER	.014	.02 *	-.10	-.05	-.10	*	.14	.04	.26 *	.42	.11	.48 *
POSSIBILE	.002	.06 *	.04	-.20	-.15	*	.00	.10	.09 *	.03	.61	.36 *
POSSIBILITÀ	.001	.03 *	-.14	-.08	.06	*	.02	.01	.01 *	.63	.24	.13 *
POSTA	.000	.07 *	-.16	.00	-.20	*	.01	.00	.02 *	.39	.00	.61 *
POSTO	.001	.01 *	-.02	-.04	.08	*	.00	.00	.01 *	.03	.16	.81 *
POTENZA	.000	.02 *	-.10	.08	.00	*	.00	.00	.00 *	.58	.42	.00 *
POTERE	.001	.24 *	-.21	-.35	.28	*	.06	.17	.17 *	.18	.50	.32 *
PRATICA	.001	.44 *	.18	.54	.35	*	.03	.29	.18 *	.07	.65	.28 *
PRATICHE	.000	.26 *	-.29	-.07	.42	*	.03	.00	.10 *	.32	.02	.66 *
PRE	.001	.65 *	-.43	-.68	.03	*	.15	.40	.00 *	.29	.71	.00 *
PREFERISCO	.000	.33 *	.41	-.27	-.30	*	.06	.03	.05 *	.51	.22	.27 *
PREGHIERA	.001	.76 *	.26	.78	-.28	*	.07	.69	.13 *	.09	.81	.10 *
PREGO	.000	.04 *	-.09	.11	-.12	*	.00	.01	.01 *	.24	.34	.42 *
PRENDERE	.001	.01 *	-.06	.05	-.02	*	.00	.00	.00 *	.54	.43	.03 *
PRESENTE	.001	.03 *	.17	.07	.02	*	.04	.01	.00 *	.86	.13	.01 *
PRESENZA	.001	.53 *	.72	-.12	-.02	*	.34	.01	.00 *	.97	.03	.00 *
PRESO	.001	.04 *	.04	-.19	-.03	*	.00	.03	.00 *	.04	.93	.03 *
PRESTO	.000	.33 *	.54	.17	-.11	*	.12	.01	.01 *	.88	.09	.04 *
PRETE	.001	.04 *	-.09	-.06	-.18	*	.01	.00	.07 *	.18	.07	.75 *
PRIMA	.005	.01 *	.00	.08	.01	*	.00	.04	.00 *	.00	.98	.02 *
PRIMO	.002	.15 *	.28	.14	.23	*	.19	.05	.21 *	.52	.13	.35 *
PRINCIPALE	.000	.14 *	.06	.37	.02	*	.00	.06	.00 *	.02	.97	.00 *
PRINCIPI	.000	.23 *	-.20	-.17	.41	*	.02	.02	.14 *	.17	.13	.70 *
PRINCIPIO	.001	1.41 *	.20	.45	1.08	*	.04	.22	1.91 *	.03	.15	.83 *
PROBLEMA	.004	.05 *	-.10	-.18	-.08	*	.04	.15	.05 *	.21	.66	.13 *
PROCESSI	.001	.10 *	.03	.29	.13	*	.00	.10	.03 *	.01	.83	.16 *
PRODOTTO	.000	.47 *	.61	-.29	-.11	*	.15	.04	.01 *	.79	.18	.03 *
PROFONDO	.000	.08 *	.20	.04	.20	*	.02	.00	.03 *	.48	.02	.50 *
PROPOSITO	.001	.08 *	-.10	.12	-.23	*	.01	.01	.07 *	.13	.18	.68 *
PROSSIMO	.001	.05 *	.12	-.11	.14	*	.01	.01	.02 *	.32	.25	.43 *
PROVA	.002	.11 *	-.21	-.23	-.12	*	.11	.15	.06 *	.39	.48	.12 *
PUBBLICA	.000	.11 *	-.22	-.17	.20	*	.02	.01	.03 *	.41	.25	.34 *
PUBBLICO	.001	.11 *	-.22	-.11	-.23	*	.04	.01	.06 *	.44	.11	.46 *
PUNTI	.004	.06 *	-.14	-.16	-.10	*	.09	.13	.07 *	.35	.47	.18 *
QUALITÀ	.000	.07 *	-.04	-.01	-.26	*	.00	.00	.05 *	.03	.00	.97 *
QUATTRO	.000	.31 *	.27	.14	.47	*	.03	.01	.15 *	.23	.07	.70 *
QUESTIONE	.002	.05 *	-.20	-.11	.00	*	.07	.02	.00 *	.78	.22	.00 *
RAGIONAMENTO	.000	.18 *	-.30	-.28	-.10	*	.03	.03	.01 *	.51	.43	.06 *
RAGIONE	.002	.12 *	-.22	-.24	-.12	*	.09	.11	.04 *	.40	.48	.12 *
RAGIONI	.000	.09 *	.18	.19	-.14	*	.02	.02	.02 *	.36	.41	.23 *
RAPPORTO	.001	.39 *	.29	.55	.03	*	.08	.29	.00 *	.22	.78	.00 *
REALE	.000	.28 *	-.29	-.32	.31	*	.04	.05	.08 *	.29	.36	.35 *
REALTÀ	.003	.04 *	-.13	-.12	.06	*	.05	.04	.02 *	.51	.38	.11 *
REGIME	.001	.83 *	-.38	-.82	.10	*	.09	.46	.01 *	.17	.81	.01 *
REGNO	.001	.19 *	.03	.25	.35	*	.00	.07	.21 *	.00	.34	.65 *
REGOLE	.000	.30 *	.03	-.54	-.10	*	.00	.17	.01 *	.00	.97	.03 *

RELAZIONE	.000	1.03 *	.22	.54	.83	*	.02	.12	.43 *	.05	.28	.67 *
RELIGIONE	.005	.34 *	-.24	-.48	.21	*	.27	1.26	.36 *	.17	.70	.13 *
RELIGIOSO	.003	.13 *	-.14	-.05	.33	*	.06	.01	.49 *	.15	.02	.82 *
RENDERE	.003	.02 *	-.08	-.01	-.13	*	.02	.00	.10 *	.26	.01	.74 *
RESPONSABILE	.000	.37 *	-.33	-.40	.33	*	.04	.07	.07 *	.29	.42	.29 *
RESPONSABILITÀ	.001	.07 *	-.12	-.10	.21	*	.02	.01	.08 *	.21	.14	.65 *
RESURREZIONE	.000	.27 *	.13	.50	.02	*	.01	.09	.00 *	.06	.94	.00 *
RETE	.000	.05 *	.08	-.06	.21	*	.00	.00	.04 *	.11	.08	.81 *
RICERCA	.001	.00 *	.04	.02	.01	*	.00	.00	.00 *	.68	.26	.05 *
RICEVUTO	.000	.21 *	.09	.44	-.04	*	.00	.07	.00 *	.04	.95	.01 *
RICHIESTA	.000	.07 *	.11	.23	-.09	*	.01	.03	.01 *	.16	.73	.11 *
RICORDA	.000	.24 *	.34	.35	-.07	*	.05	.05	.00 *	.48	.50	.02 *
RICORDO	.001	.11 *	.16	.15	-.25	*	.02	.02	.07 *	.24	.20	.56 *
RIUSCIRE	.001	.09 *	-.16	-.20	-.17	*	.03	.05	.06 *	.28	.41	.30 *
RIFERIMENTO	.001	.03 *	-.11	.09	-.10	*	.01	.01	.01 *	.40	.26	.34 *
RIFLESSIONE	.000	.09 *	-.01	.21	.21	*	.00	.02	.03 *	.00	.50	.50 *
RIGUARDA	.001	.01 *	-.05	-.03	-.10	*	.00	.00	.02 *	.19	.06	.75 *
RIMANE	.001	.04 *	-.08	.14	-.10	*	.00	.01	.01 *	.16	.55	.29 *
RINGRAZIO	.000	.35 *	.09	.49	-.32	*	.00	.10	.07 *	.02	.68	.30 *
RIPETO	.001	.14 *	-.01	.04	-.37	*	.00	.00	.18 *	.00	.01	.99 *
RISCHIO	.000	.12 *	.02	.32	-.13	*	.00	.04	.01 *	.00	.86	.14 *
RISPETTO	.002	.06 *	-.11	-.19	-.11	*	.02	.08	.04 *	.20	.60	.21 *
RISPONDERE	.002	.10 *	.15	-.27	-.09	*	.04	.14	.02 *	.21	.71	.08 *
RISPOSTA	.003	.09 *	.22	-.19	-.09	*	.13	.10	.03 *	.54	.38	.08 *
RISPOSTO	.001	.26 *	.05	.41	-.29	*	.00	.11	.08 *	.01	.66	.33 *
RISULTA	.000	.10 *	-.23	-.13	-.17	*	.02	.01	.02 *	.53	.17	.30 *
RISULTATI	.000	.17 *	-.28	-.28	.11	*	.03	.03	.01 *	.48	.46	.07 *
RITENERE	.001	.07 *	.09	.05	-.25	*	.01	.00	.14 *	.11	.03	.86 *
RIVELAZIONE	.000	.25 *	.03	.38	-.33	*	.00	.08	.09 *	.00	.57	.43 *
ROBERTO	.000	.11 *	-.03	.24	-.23	*	.00	.02	.03 *	.01	.51	.48 *
ROMA	.001	.29 *	.33	.03	.42	*	.11	.00	.30 *	.38	.00	.62 *
ROMANO	.000	.04 *	-.05	.18	-.04	*	.00	.01	.00 *	.07	.90	.04 *
RUOLO	.001	.20 *	.05	.39	-.21	*	.00	.13	.05 *	.01	.77	.21 *
SABATO	.001	.64 *	.27	.74	.13	*	.05	.39	.02 *	.11	.86	.03 *
SACERDOTE	.001	.39 *	.13	.55	-.26	*	.02	.31	.10 *	.04	.78	.18 *
SACRO	.000	.04 *	-.18	-.06	.02	*	.01	.00	.00 *	.88	.11	.01 *
SAPERE	.004	.06 *	-.07	-.20	-.12	*	.02	.19	.11 *	.09	.67	.24 *
SALUTO	.002	.09 *	.02	.14	-.26	*	.00	.04	.18 *	.00	.24	.76 *
SALVEZZA	.001	.50 *	.18	.52	.43	*	.03	.28	.29 *	.06	.55	.38 *
SAN	.001	.51 *	.35	.55	.29	*	.13	.35	.15 *	.24	.59	.17 *
SANGUE	.000	.04 *	.10	.12	.13	*	.00	.01	.01 *	.23	.36	.41 *
SANTO	.003	.29 *	.12	.48	-.22	*	.04	.70	.23 *	.05	.78	.17 *
SANTITÀ	.000	.57 *	.37	.65	.07	*	.05	.19	.00 *	.24	.75	.01 *
SATANA	.001	.72 *	-.29	-.38	.70	*	.12	.24	1.22 *	.11	.20	.68 *
SBAGLIATO	.000	.31 *	.28	-.46	-.14	*	.04	.10	.02 *	.26	.67	.07 *
SCELTA	.001	.08 *	.03	-.13	-.25	*	.00	.03	.16 *	.01	.21	.77 *
SCIANTO	.000	1.40 *	-.67	-.97	.01	*	.19	.44	.00 *	.32	.68	.00 *
SCIENTIFICO	.002	.51 *	-.45	-.55	-.04	*	.45	.75	.01 *	.39	.60	.00 *
SCIENZA	.002	.56 *	-.47	-.59	-.05	*	.43	.75	.01 *	.38	.61	.00 *
SCIENZIATI	.000	.67 *	-.50	-.64	-.11	*	.10	.18	.01 *	.37	.61	.02 *
SCOPO	.001	.97 *	.37	.12	.90	*	.11	.01	1.14 *	.14	.02	.84 *
SCRITTA	.000	.35 *	.56	-.04	-.18	*	.13	.00	.02 *	.90	.00	.09 *
SCRITTI	.001	.28 *	.47	-.24	.01	*	.13	.04	.00 *	.79	.21	.00 *
SCRITTO	.005	.12 *	.20	.10	-.27	*	.20	.06	.61 *	.32	.08	.60 *
SCRITTURA	.001	.65 *	.32	.61	-.42	*	.15	.59	.42 *	.16	.57	.27 *
SCRIVERE	.002	.05 *	-.07	-.14	.16	*	.01	.04	.08 *	.10	.40	.51 *
SCUOLE	.001	.08 *	.00	.28	-.04	*	.00	.11	.00 *	.00	.98	.02 *
SCUSA	.001	.09 *	-.03	.08	-.29	*	.00	.01	.19 *	.01	.07	.92 *
SCUSATE	.000	.35 *	.50	-.25	-.21	*	.09	.03	.03 *	.70	.18	.13 *
SECOLO	.001	.19 *	.19	.39	.06	*	.06	.25	.01 *	.19	.79	.02 *
SECONDO	.004	.01 *	-.02	.10	.04	*	.00	.05	.01 *	.03	.82	.15 *
SEDE	.000	.10 *	.22	-.21	-.05	*	.02	.02	.00 *	.52	.45	.03 *
SEGNO	.000	.08 *	.01	.28	-.05	*	.00	.04	.00 *	.00	.97	.03 *
SEGUIRE	.001	.08 *	.23	.11	-.11	*	.03	.01	.01 *	.68	.16	.16 *
SEGUITO	.001	.18 *	.04	.29	.32	*	.00	.06	.11 *	.01	.45	.55 *
SELEZIONE	.001	.93 *	.20	.74	-.59	*	.03	.47	.45 *	.04	.58	.37 *
SEMBRARE	.004	.05 *	-.15	-.06	-.16	*	.08	.02	.15 *	.43	.08	.49 *
SEMPLICE	.002	.05 *	-.18	-.12	-.01	*	.07	.04	.00 *	.68	.32	.00 *
SEMPRE	.007	.00 *	-.04	.00	-.05	*	.01	.00	.03 *	.30	.00	.70 *
SENSI	.000	.06 *	-.09	.12	-.18	*	.00	.01	.02 *	.16	.28	.57 *
SENSO	.004	.04 *	-.06	.09	-.18	*	.01	.04	.20 *	.08	.20	.72 *
SENTIRE	.003	.04 *	-.05	-.05	-.18	*	.01	.01	.15 *	.07	.06	.86 *
SENTIMENTI	.000	.08 *	-.23	-.15	-.05	*	.03	.01	.00 *	.68	.28	.04 *
SENZA	.006	.00 *	-.04	.01	-.06	*	.01	.00	.03 *	.30	.01	.70 *
SERA	.000	.20 *	.33	.23	.19	*	.04	.02	.02 *	.55	.27	.18 *
SERIE	.001	.12 *	-.26	-.21	-.08	*	.04	.03	.01 *	.58	.37	.05 *
SERMONTI	.001	1.08 *	.26	.84	-.55	*	.04	.49	.32 *	.06	.65	.28 *
SERVE	.001	.32 *	.54	-.04	-.18	*	.20	.00	.04 *	.89	.01	.10 *
SERVIZIO	.001	.06 *	-.07	.16	-.18	*	.00	.02	.03 *	.08	.41	.51 *
SESSUALE	.000	.26 *	-.06	.07	.51	*	.00	.00	.19 *	.01	.02	.97 *

SETTE	.000	.71 *	-.01	.11	.84	*	.00	.01	.47 *	.00	.02	.98 *
SETTEMBRE	.000	.16 *	.17	.36	.00	*	.02	.07	.00 *	.18	.82	.00 *
SICURO	.001	.07 *	-.18	-.03	-.18	*	.02	.00	.03 *	.48	.02	.50 *
SIGNIFICA	.001	.04 *	.09	.10	-.16	*	.01	.02	.05 *	.19	.25	.57 *
SIGNIFICATO	.001	.09 *	.01	.29	-.06	*	.00	.12	.01 *	.00	.95	.05 *
SIGNORE	.002	.52 *	.23	.64	-.25	*	.12	.95	.23 *	.10	.77	.12 *
SILENZIO	.000	.67 *	.09	.31	.75	*	.00	.04	.35 *	.01	.14	.84 *
SIMILE	.002	.02 *	-.05	-.11	.04	*	.00	.02	.00 *	.17	.73	.10 *
SISTEMA	.001	.37 *	-.39	-.46	.00	*	.14	.22	.00 *	.42	.58	.00 *
SITO	.001	.24 *	.28	.35	.21	*	.09	.15	.09 *	.32	.49	.19 *
SITUAZIONE	.001	.09 *	.09	.20	.20	*	.00	.02	.04 *	.09	.44	.47 *
SOCIALE	.001	.23 *	-.31	-.33	.15	*	.10	.13	.04 *	.42	.48	.10 *
SOCIETÀ	.001	.10 *	-.19	-.21	.12	*	.04	.05	.03 *	.38	.47	.15 *
SOFFERENZA	.000	.41 *	.10	.36	.52	*	.00	.06	.19 *	.02	.31	.66 *
SOGGETTO	.000	.53 *	-.30	-.58	.31	*	.04	.18	.08 *	.18	.64	.18 *
SOGNO	.000	.24 *	-.21	-.40	.20	*	.02	.08	.03 *	.18	.65	.17 *
SOLO	.015	.01 *	-.01	-.07	-.09	*	.00	.09	.24 *	.00	.36	.64 *
SOLDI	.001	.45 *	-.14	-.61	.24	*	.01	.24	.06 *	.04	.83	.13 *
SOLITO	.001	.17 *	-.30	-.27	-.08	*	.06	.06	.01 *	.52	.44	.04 *
SOLUZIONE	.000	.23 *	.01	-.47	-.09	*	.00	.11	.01 *	.00	.97	.03 *
SOPRA	.001	.04 *	-.10	.02	-.17	*	.01	.00	.06 *	.24	.01	.75 *
SORTA	.000	.14 *	.05	.37	-.07	*	.00	.05	.00 *	.02	.95	.03 *
SOSTANZA	.001	.06 *	-.12	.08	-.20	*	.01	.00	.04 *	.24	.10	.67 *
SOSTENERE	.000	.18 *	.22	.15	-.32	*	.02	.01	.07 *	.28	.12	.59 *
SOTTO	.001	.03 *	.10	.12	.03	*	.01	.02	.00 *	.39	.58	.03 *
SPAZIO	.001	.45 *	-.42	-.52	.04	*	.17	.29	.00 *	.39	.60	.00 *
SPECIE	.002	.36 *	.08	.43	-.41	*	.01	.41	.57 *	.02	.51	.47 *
SPERANZA	.000	.47 *	.17	.50	.44	*	.01	.13	.15 *	.06	.53	.41 *
SPERO	.001	.10 *	.10	.08	-.29	*	.01	.01	.15 *	.10	.06	.84 *
SPIEGARE	.002	.03 *	.02	-.13	-.11	*	.00	.03	.04 *	.02	.57	.41 *
SPIEGAZIONE	.001	.05 *	-.18	-.05	-.12	*	.04	.00	.03 *	.65	.04	.31 *
SPIRITO	.002	.33 *	.14	.46	.31	*	.04	.52	.35 *	.06	.65	.29 *
SPIRITUALE	.001	.76 *	.26	.67	.49	*	.10	.67	.54 *	.09	.59	.32 *
STAMPA	.000	.31 *	-.32	-.38	.25	*	.04	.06	.04 *	.34	.46	.21 *
STATI	.002	.03 *	-.16	-.04	.04	*	.05	.00	.00 *	.88	.06	.06 *
STATO	.006	.03 *	.07	-.12	.11	*	.03	.10	.12 *	.17	.46	.37 *
STORIA	.002	.12 *	.30	-.18	.00	*	.20	.08	.00 *	.75	.25	.00 *
STORICO	.001	.19 *	.24	.37	-.01	*	.08	.19	.00 *	.31	.69	.00 *
STRADA	.001	.04 *	-.04	.20	-.03	*	.00	.03	.00 *	.04	.94	.02 *
STRANO	.000	.16 *	-.09	.17	-.35	*	.00	.02	.10 *	.05	.18	.77 *
STUDI	.001	.08 *	.04	-.28	.07	*	.00	.05	.00 *	.02	.92	.06 *
STUDIARE	.000	.12 *	.00	-.34	.03	*	.00	.05	.00 *	.00	.99	.01 *
STUDIO	.001	.17 *	.32	-.17	-.20	*	.07	.02	.04 *	.60	.17	.23 *
STUDIOSI	.000	.02 *	-.15	.00	-.05	*	.01	.00	.00 *	.91	.00	.09 *
SUCCESSO	.000	.16 *	.05	.32	.24	*	.00	.05	.04 *	.02	.64	.34 *
SUFFICIENTE	.001	.05 *	-.20	-.09	-.06	*	.02	.01	.00 *	.75	.17	.08 *
SUKKOT	.000	63.02 *	6.96	-3.76	-.64	*	16.75	5.36	.23 *	.77	.22	.01 *
SUPERIORE	.000	.04 *	-.15	-.03	.14	*	.01	.00	.01 *	.52	.02	.46 *
TANTO	.005	.05 *	-.07	-.02	-.20	*	.03	.00	.36 *	.11	.01	.88 *
TDG	.001	1.06 *	.24	.80	-.59	*	.07	.87	.71 *	.05	.61	.33 *
TELEVISIONE	.000	.80 *	-.53	-.72	-.03	*	.12	.24	.00 *	.36	.64	.00 *
TEMA	.001	.34 *	.45	.17	.33	*	.16	.02	.14 *	.60	.08	.32 *
TEMPI	.001	.03 *	.01	.18	-.05	*	.00	.04	.00 *	.00	.93	.07 *
TEMPIO	.000	.56 *	.31	.53	.43	*	.05	.14	.14 *	.18	.50	.33 *
TEMPO	.004	.02 *	.05	-.11	.00	*	.01	.06	.00 *	.18	.82	.00 *
TENDENZA	.000	.09 *	-.22	-.10	-.17	*	.02	.00	.02 *	.54	.12	.34 *
TEOLOGIA	.000	.08 *	-.01	.26	-.09	*	.00	.04	.01 *	.00	.90	.10 *
TEORIA	.003	.25 *	-.34	-.33	-.17	*	.31	.32	.12 *	.45	.43	.11 *
TERMINI	.001	.07 *	.12	.21	-.09	*	.02	.05	.01 *	.22	.67	.11 *
TERMINI	.001	.01 *	-.08	-.05	.03	*	.00	.00	.00 *	.62	.28	.10 *
TERRA	.002	.09 *	.15	.16	.21	*	.04	.05	.14 *	.25	.27	.48 *
TERZO	.000	.16 *	.29	.26	-.04	*	.04	.03	.00 *	.56	.43	.01 *
TESI	.001	.02 *	.03	.07	.10	*	.00	.00	.02 *	.07	.27	.66 *
TESTA	.001	.07 *	-.09	-.24	-.07	*	.01	.04	.01 *	.11	.82	.07 *
TESTAMENTO	.000	.27 *	.14	-.46	.18	*	.01	.08	.02 *	.08	.80	.12 *
TESTI	.003	.22 *	.36	.23	-.21	*	.37	.17	.21 *	.57	.24	.19 *
TESTIMONI	.001	.59 *	.14	.61	-.45	*	.01	.29	.24 *	.03	.62	.34 *
TESTIMONIANZA	.000	.08 *	.00	.28	-.07	*	.00	.03	.00 *	.00	.94	.06 *
THOMAS	.000	5.75 *	.40	.63	2.28	*	.06	.15	2.96 *	.03	.07	.90 *
THREAD	.001	.20 *	-.02	.29	-.34	*	.00	.07	.14 *	.00	.42	.58 *
TIMOR	.000	.33 *	-.06	.06	.57	*	.00	.00	.18 *	.01	.01	.98 *
TIPO	.002	.09 *	-.18	-.23	.01	*	.07	.12	.00 *	.37	.63	.00 *
TITOLO	.001	.01 *	.10	.04	.02	*	.01	.00	.00 *	.84	.12	.03 *
TNM	.001	1.17 *	.26	.85	-.63	*	.05	.58	.48 *	.06	.61	.33 *
TOGLIERE	.000	.56 *	-.45	-.59	.09	*	.11	.20	.01 *	.37	.62	.01 *
TORAH	.001	53.56 *	6.40	-3.50	-.60	*	31.8310.45	47 *	.76	.23	.01	*
TOTALE	.000	.09 *	-.16	-.25	.03	*	.01	.04	.00 *	.29	.70	.01 *
TRADIZIONE	.001	.42 *	.60	.24	-.10	*	.49	.08	.02 *	.84	.13	.03 *
TRADUZIONE	.002	1.09 *	.30	.79	-.61	*	.15	1.12	.99 *	.09	.58	.34 *
TRAMITE	.000	.22 *	.03	-.10	.46	*	.00	.00	.12 *	.00	.04	.95 *

TRATTA	.001	.01 *	.00	.08	-.06	*	.00	.01	.01 *	.00	.62	.38 *
TRE	.001	.12 *	.28	.08	.19	*	.11	.01	.08 *	.66	.06	.29 *
TROPPO	.001	.06 *	.11	-.07	-.21	*	.02	.01	.11 *	.19	.08	.73 *
TROVARE	.003	.03 *	.09	.01	-.13	*	.03	.00	.10 *	.33	.00	.67 *
UFFICIALE	.000	.03 *	-.07	.14	.01	*	.00	.01	.00 *	.17	.83	.00 *
ULTIMO	.002	.08 *	.24	.09	.12	*	.11	.02	.04 *	.73	.10	.17 *
UMANO	.003	.02 *	-.09	-.03	.11	*	.02	.00	.06 *	.37	.04	.59 *
UMANITÀ	.001	.66 *	.72	-.18	.34	*	.37	.03	.14 *	.78	.05	.17 *
UNICO	.002	.05 *	.18	-.11	-.06	*	.07	.03	.01 *	.67	.26	.07 *
UNIONE	.000	.48 *	.20	.57	.34	*	.01	.13	.07 *	.08	.69	.23 *
UNITÀ	.000	.42 *	-.07	.03	.64	*	.00	.00	.28 *	.01	.00	.99 *
UNIVERSALE	.000	.16 *	.07	.36	.17	*	.00	.06	.02 *	.03	.80	.17 *
UNIVERSO	.001	.58 *	-.45	-.60	.13	*	.28	.52	.03 *	.36	.62	.03 *
UNO	.007	.01 *	-.09	-.03	.04	*	.05	.01	.02 *	.76	.07	.16 *
UOMO	.009	.07 *	.08	.16	.19	*	.06	.28	.57 *	.10	.38	.52 *
USA	.001	.02 *	-.07	-.05	-.10	*	.00	.00	.01 *	.31	.16	.53 *
USARE	.001	.08 *	-.15	-.20	-.14	*	.02	.05	.03 *	.28	.50	.23 *
USO	.001	.07 *	.13	.22	-.01	*	.01	.04	.00 *	.26	.74	.00 *
UTILE	.000	.06 *	-.06	-.07	-.23	*	.00	.00	.04 *	.06	.08	.86 *
VENIRE	.007	.00 *	.01	.03	-.05	*	.00	.01	.03 *	.06	.22	.73 *
VALE	.001	.06 *	.18	-.16	-.01	*	.02	.02	.00 *	.54	.45	.00 *
VALORE	.001	.17 *	.40	-.05	.06	*	.10	.00	.00 *	.96	.02	.03 *
VALORI	.000	.02 *	.01	.03	.14	*	.00	.00	.01 *	.00	.03	.96 *
VANGELO	.001	.20 *	.09	.40	.18	*	.01	.29	.09 *	.04	.80	.16 *
VARI	.001	.04 *	.09	-.17	.06	*	.01	.04	.01 *	.20	.72	.08 *
VATICANO	.001	.03 *	-.11	-.08	.10	*	.01	.01	.01 *	.41	.24	.35 *
VECCHIO	.000	.90 *	.78	-.54	.04	*	.30	.16	.00 *	.67	.33	.00 *
VEDERE	.007	.06 *	-.10	-.17	-.13	*	.07	.24	.23 *	.16	.52	.32 *
VERA	.001	.08 *	.17	.17	.14	*	.04	.05	.05 *	.38	.37	.25 *
VERBO	.000	.20 *	.10	.42	.12	*	.00	.10	.01 *	.05	.88	.07 *
VERGINE	.000	.10 *	.15	.27	.10	*	.01	.03	.01 *	.21	.69	.10 *
VERI	.000	.12 *	.00	.21	.28	*	.00	.02	.05 *	.00	.37	.63 *
VERIFICA	.000	.21 *	-.30	-.25	-.24	*	.03	.02	.04 *	.43	.29	.29 *
VERIFICARE	.000	.08 *	.19	.03	-.21	*	.01	.00	.03 *	.44	.01	.55 *
VERITÀ	.003	.15 *	.11	.37	-.03	*	.04	.49	.01 *	.08	.91	.01 *
VERO	.004	.02 *	-.12	-.03	-.06	*	.05	.00	.02 *	.77	.04	.19 *
VERSIONE	.001	.36 *	.45	.35	-.18	*	.12	.08	.03 *	.57	.34	.09 *
VERSO	.002	.11 *	.03	.13	.29	*	.00	.03	.23 *	.01	.16	.83 *
VESCOVO	.000	.47 *	.21	.65	.07	*	.02	.16	.00 *	.10	.89	.01 *
VIA	.002	.13 *	.22	.18	.21	*	.12	.09	.18 *	.39	.26	.35 *
VICINO	.000	.24 *	.37	.16	.28	*	.06	.01	.06 *	.57	.10	.32 *
VISIONE	.001	.08 *	-.13	-.25	-.04	*	.01	.05	.00 *	.20	.78	.02 *
VISSUTO	.000	.12 *	-.16	-.09	.29	*	.01	.00	.06 *	.23	.07	.71 *
VISTA	.002	.09 *	-.22	-.21	.00	*	.08	.08	.00 *	.52	.48	.00 *
VISTO	.003	.04 *	-.06	-.13	-.13	*	.01	.05	.08 *	.11	.44	.44 *
VITA	.007	.11 *	-.10	.01	.31	*	.08	.00	1.28 *	.10	.00	.90 *
VIVERE	.003	.02 *	-.13	-.06	.01	*	.05	.01	.00 *	.83	.16	.01 *
VOCE	.001	.25 *	.10	.47	-.14	*	.01	.22	.03 *	.04	.89	.07 *
VOGLIA	.001	.05 *	-.14	.04	-.18	*	.02	.00	.05 *	.37	.03	.60 *
VOLERE	.009	.02 *	-.09	-.07	-.07	*	.08	.05	.09 *	.45	.25	.30 *
VOLARE	.000	1.23 *	-.63	-.91	.05	*	.15	.34	.00 *	.32	.67	.00 *
VOLONTÀ	.001	.26 *	.03	.25	.44	*	.00	.08	.34 *	.00	.25	.74 *
VOLTA	.003	.07 *	.16	.12	.18	*	.07	.05	.15 *	.35	.21	.44 *
VOLTE	.002	.03 *	.00	.04	-.18	*	.00	.00	.12 *	.00	.04	.96 *
WILDE	.000	1.40 *	-.67	-.97	.01	*	.19	.45	.00 *	.32	.68	.00 *

# Output della procedura APLUMS

EDITION DES COORDONNEES ET CONTRIBUTIONS DES LIGNES

MOTS OU SEGMENTS DE MOTS	MASSES DISTO2 *	COORDONNEES			*CONTR. ABSOLUES *			COSINUS CARRES *				
*****												
1-AFFERMAZIONE DIO	.002	1.06 *	1.03	.07	-.03	*	.86	.01	.00 *	.99	.01	.00 *
2-AMARE DARE	.004	1.32 *	-.97	-.32	-.53	*	1.24	.24	.75 *	.71	.08	.22 *
3-ANNO PRIMA	.003	.12 *	-.18	-.12	.27	*	.03	.02	.14 *	.28	.11	.61 *
4-ANNO SANTO	.003	1.04 *	-.96	-.29	-.20	*	.81	.13	.07 *	.88	.08	.04 *
5-BAMBINO MORTI	.003	.20 *	.41	-.05	-.19	*	.19	.01	.08 *	.81	.01	.17 *
6-BAMBINO MORTI SENZA	.003	.27 *	.50	-.03	-.16	*	.27	.00	.06 *	.90	.00	.10 *
7-BASE COMUNE	.003	12.68 *	-.87	.01	3.45	*	.67	.0021	1.00 *	.06	.00	.94 *
8-BELLO DIFFERENZA	.003	.31 *	.53	-.03	-.16	*	.25	.00	.04 *	.92	.00	.08 *
9-BENE DIVERSO	.002	.46 *	.67	.00	-.12	*	.36	.00	.02 *	.97	.00	.03 *
10-BENE FARE	.002	.38 *	-.42	-.21	-.40	*	.14	.07	.25 *	.47	.12	.41 *
11-BENE POCO	.009	.11 *	-.28	-.16	-.07	*	.23	.14	.03 *	.71	.24	.04 *
12-BENE PRESTO	.003	.78 *	-.82	-.27	-.21	*	.69	.13	.09 *	.85	.09	.06 *
13-BIBBIA TDG	.002	1.32 *	-.97	-.32	-.53	*	.76	.15	.46 *	.71	.08	.22 *
14-BUONO FEDE	.009	.73 *	-.72	-.26	-.37	*	1.57	.37	.84 *	.71	.10	.19 *
15-BUONO LETTURA	.002	3.54 *	-.89	1.56	.56	*	.64	3.50	.51 *	.22	.69	.09 *
16-BUONO NOTIZIA	.003	1.32 *	-.97	-.32	-.53	*	1.03	.20	.62 *	.71	.08	.22 *
17-BUONO SENSO	.004	.03 *	.14	-.09	-.06	*	.03	.02	.01 *	.63	.26	.11 *
18-BUONO VOLONTÀ	.004	.36 *	-.53	-.19	-.20	*	.40	.09	.11 *	.79	.10	.11 *
19-CAPIRE BENE	.006	.11 *	.22	-.09	-.23	*	.09	.03	.22 *	.43	.07	.50 *
20-CAPIRE MEGLIO	.003	.60 *	-.57	.50	.13	*	.34	.46	.03 *	.55	.42	.03 *
21-CARO PIETRO	.003	.46 *	-.63	-.23	-.12	*	.35	.08	.02 *	.86	.11	.03 *
22-CARO SALUTO	.003	.43 *	-.47	-.22	-.41	*	.19	.08	.29 *	.50	.11	.38 *
23-CASO PARTICOLARE	.003	.17 *	.36	-.06	-.20	*	.12	.01	.07 *	.76	.02	.22 *
24-CERTI ARGOMENTO	.003	.67 *	-.63	-.25	-.45	*	.35	.10	.35 *	.60	.10	.30 *
25-CERTO PUNTI	.004	.41 *	.63	-.01	-.13	*	.59	.00	.05 *	.96	.00	.04 *
26-CERTO SENSO	.006	.20 *	-.39	-.18	-.10	*	.31	.12	.04 *	.77	.17	.06 *
27-CHIEDERE SCUSA	.009	.26 *	-.47	.05	-.20	*	.63	.01	.24 *	.83	.01	.16 *
28-CHIESA CATTOLICO	.031	.09 *	-.14	-.11	.23	*	.20	.24	1.14 *	.22	.15	.63 *
29-CHIESA CRISTIANO	.002	3.30 *	-.37	-.03	1.78	*	.11	.00	5.11 *	.04	.00	.96 *
30-CONOSCERE BENE	.003	.54 *	.21	.67	-.24	*	.04	.75	.11 *	.08	.82	.11 *
31-CONOSCENZA SCIENTIF	.002	1.06 *	1.03	.07	-.03	*	.86	.01	.00 *	.99	.01	.00 *
32-CORPO FISICO	.003	.79 *	.77	.06	.43	*	.66	.01	.42 *	.76	.01	.24 *
33-CRISTO GESÙ	.003	1.32 *	-.97	-.32	-.53	*	.89	.17	.54 *	.71	.08	.22 *
34-CRISTO STORICO	.003	.83 *	-.32	-.10	.85	*	.10	.02	1.37 *	.13	.01	.86 *
35-CULTURA GENERALE	.007	1.06 *	1.03	.07	-.03	*	2.58	.02	.00 *	.99	.01	.00 *
36-DOVERE ANDARE	.004	.35 *	-.32	.40	.29	*	.14	.38	.22 *	.30	.46	.24 *
37-DOVERE CREDERE	.003	.75 *	-.68	-.26	-.46	*	.48	.13	.44 *	.62	.09	.28 *
38-DOVERE DARE	.003	.67 *	-.63	-.25	-.45	*	.35	.10	.35 *	.60	.10	.30 *
39-DOVERE DIRE	.008	.09 *	-.17	.11	-.23	*	.08	.06	.28 *	.32	.13	.55 *
40-DOVERE FARE	.018	.04 *	.16	-.09	-.10	*	.15	.08	.13 *	.57	.18	.26 *
41-DOVERE RISPONDERE	.002	2.87 *	-.55	1.54	-.43	*	.24	3.42	.30 *	.10	.83	.07 *
42-DOVERE SENTIRE	.002	.12 *	-.06	-.14	-.30	*	.00	.03	.15 *	.03	.17	.80 *
43-DOVERE SOLO	.003	.06 *	.24	-.06	.03	*	.06	.01	.00 *	.92	.06	.02 *
44-DIRE NULLA	.004	.26 *	.47	-.04	-.17	*	.30	.00	.08 *	.88	.01	.11 *
45-DIRE OGGI	.005	1.32 *	-.97	-.32	-.53	*	1.72	.33	1.04 *	.71	.08	.22 *
46-DIRE SOLO	.002	.09 *	.12	-.11	-.26	*	.01	.02	.11 *	.16	.12	.72 *
47-DIECI ANNO	.003	.10 *	.03	-.12	-.28	*	.00	.03	.16 *	.01	.16	.83 *
48-DIO ESISTERE	.012	.72 *	.82	.06	.20	*	2.86	.02	.32 *	.94	.00	.05 *
49-DIO PADRE	.004	.30 *	-.51	-.20	.02	*	.35	.09	.00 *	.87	.13	.00 *
50-DIO POTER	.003	.53 *	.72	.01	-.11	*	.50	.00	.02 *	.98	.00	.02 *
51-DIO UOMO	.002	.46 *	.67	.00	-.12	*	.36	.00	.02 *	.97	.00	.03 *
52-DIRITTO UMANO	.005	.68 *	-.26	.78	.02	*	.11	1.67	.00 *	.10	.90	.00 *
53-DUE ANNO	.003	.42 *	-.15	.59	.24	*	.02	.63	.12 *	.05	.81	.13 *
54-DUE PARTI	.003	.03 *	.16	-.08	.00	*	.03	.01	.00 *	.81	.19	.00 *
55-DUE PERSONA	.002	.24 *	.14	-.05	.47	*	.02	.00	.35 *	.08	.01	.91 *
56-DUE VOLTE	.003	.60 *	-.57	.50	.13	*	.34	.46	.03 *	.55	.42	.03 *
57-DUEMILA ANNO	.010	.52 *	-.04	.72	-.05	*	.01	3.14	.02 *	.00	.99	.00 *
58-ESPERIENZA PERSONAL	.003	.02 *	.10	-.09	.00	*	.01	.02	.00 *	.58	.42	.00 *
59-FARE BENE	.007	.41 *	-.45	-.22	-.40	*	.46	.19	.74 *	.49	.12	.39 *
60-FARE CAPIRE	.009	.11 *	-.17	-.15	-.23	*	.08	.12	.32 *	.26	.23	.51 *
61-FARE CREDERE	.004	.02 *	-.02	-.12	-.06	*	.00	.03	.01 *	.03	.76	.22 *
62-FARE MALE	.004	.27 *	-.45	-.17	.19	*	.30	.08	.11 *	.76	.11	.13 *
63-FARE NASCERE	.003	.48 *	-.64	-.21	.16	*	.40	.07	.05 *	.86	.09	.05 *
64-FARE NOTARE	.008	.06 *	.23	-.07	-.02	*	.14	.02	.00 *	.91	.08	.01 *
65-FARE NULLA	.004	.20 *	.45	-.02	.06	*	.25	.00	.01 *	.98	.00	.02 *
66-FARE PARTE	.009	.11 *	.25	.16	-.13	*	.20	.14	.11 *	.60	.24	.16 *
67-FARE PASSARE	.003	.71 *	-.66	-.26	-.46	*	.42	.11	.40 *	.61	.10	.29 *

68-FARE RIFERIMENTO	.003	.85 *	-.79	-.21	.43	*	.60	.08	.35 *	.73	.05	.21 *
69-FARE SOLO	.005	.08 *	-.23	-.15	.01	*	.09	.06	.00 *	.71	.29	.00 *
70-FARE TANTO	.002	.09 *	.12	-.11	-.26	*	.01	.02	.11 *	.16	.12	.72 *
71-FARE VEDERE	.005	.23 *	.45	-.04	-.18	*	.35	.01	.11 *	.86	.01	.13 *
72-FEDE CATTOLICO	.003	.06 *	.04	-.08	.22	*	.00	.01	.11 *	.03	.12	.85 *
73-FEDE CRISTIANO	.003	.25 *	-.15	-.10	.47	*	.02	.02	.48 *	.09	.04	.87 *
74-FORSE DOVERE	.002	.77 *	.24	.81	-.23	*	.05	.94	.09 *	.08	.85	.07 *
75-GENERE UMANO	.003	.14 *	-.11	-.15	-.32	*	.01	.04	.21 *	.09	.17	.74 *
76-GESÙ CRISTO	.019	.54 *	-.68	-.22	.16	*	3.05	.54	.34 *	.86	.09	.05 *
77-GESÙ DIRE	.004	.79 *	-.84	-.26	-.06	*	.95	.16	.01 *	.91	.09	.00 *
78-GIOVANNI PAOLO	.008	.30 *	-.47	.10	.25	*	.59	.05	.33 *	.75	.03	.21 *
79-GRANDE PARTE	.004	.20 *	-.25	-.12	.35	*	.08	.03	.30 *	.31	.08	.62 *
80-GRAZIE GRAZIE	.002	1.32 *	-.97	-.32	-.53	*	.76	.15	.46 *	.71	.08	.22 *
81-GUARDARE CASO	.003	.32 *	.54	.00	.18	*	.26	.00	.05 *	.90	.00	.10 *
82-HOME PAGE	.009	.33 *	.57	.00	.04	*	1.03	.00	.01 *	1.00	.00	.00 *
83-IPOTESI DIVINA	.004	1.06 *	1.03	.07	-.03	*	1.33	.01	.00 *	.99	.01	.00 *
84-LASCIARE PERDERE	.003	.12 *	.26	-.08	-.22	*	.07	.01	.10 *	.55	.05	.40 *
85-LIBERO ARBITRIO	.012	.80 *	.89	.06	.08	*	3.27	.02	.05 *	.99	.00	.01 *
86-LIBERTÀ RELIGIOSO	.006	.87 *	-.04	-.04	.93	*	.00	.01	3.68 *	.00	.00	1.00 *
87-LIBERO SACRO	.003	1.06 *	1.03	.07	-.03	*	1.01	.01	.00 *	.99	.01	.00 *
88-LINGUA ORIGINALE	.003	1.32 *	-.97	-.32	-.53	*	.83	.16	.50 *	.71	.08	.22 *
89-MAGGIORE PARTE	.009	.73 *	-.08	.79	-.31	*	.02	3.39	.60 *	.01	.86	.13 *
90-MENO MALE	.003	.20 *	-.22	-.17	-.34	*	.06	.06	.28 *	.24	.15	.61 *
91-METODO SCIENTIFICO	.003	1.06 *	1.03	.07	-.03	*	1.01	.01	.00 *	.99	.01	.00 *
92-MODO DIVERSO	.004	.02 *	.10	-.09	-.03	*	.01	.02	.00 *	.49	.46	.06 *
93-MODO MENO	.002	.21 *	-.41	-.18	-.03	*	.14	.05	.00 *	.83	.16	.01 *
94-MONDO CATTOLICO	.003	.65 *	-.36	.62	-.38	*	.11	.59	.26 *	.20	.58	.22 *
95-MORTI SENZA BATTESI	.003	.20 *	.41	-.05	-.19	*	.19	.01	.08 *	.81	.01	.17 *
96-NATURA DIVINA	.004	.79 *	-.83	-.24	.20	*	.87	.13	.10 *	.88	.07	.05 *
97-NATURA UMANO	.005	.03 *	.12	-.08	.07	*	.03	.02	.02 *	.57	.24	.19 *
98-NESSUNO DIO	.002	1.06 *	1.03	.07	-.03	*	.86	.01	.00 *	.99	.01	.00 *
99-NESSUNO MODO	.003	.03 *	-.10	-.13	-.03	*	.01	.03	.00 *	.38	.58	.04 *
100-NESSUNO PARTE	.005	1.37 *	-.55	1.00	-.27	*	.54	3.13	.25 *	.22	.73	.05 *
101-NESSUNO TIPO	.003	.17 *	.36	-.06	-.20	*	.12	.01	.07 *	.76	.02	.22 *
102-NORMALE PADRE	.003	1.06 *	1.03	.07	-.03	*	.94	.01	.00 *	.99	.01	.00 *
103-NUOVO MONDO	.005	1.32 *	-.97	-.32	-.53	*	1.44	.28	.87 *	.71	.08	.22 *
104-NUOVO TESTAMENTO	.004	.20 *	.11	-.05	.44	*	.01	.01	.48 *	.06	.01	.93 *
105-OCTOBRE 1999	.003	3.78 *	-.62	1.20	1.40	*	.42	2.82	4.32 *	.10	.38	.52 *
106-PADRE FIGLIO	.003	1.06 *	1.03	.07	-.03	*	1.01	.01	.00 *	.99	.01	.00 *
107-PADRE NORMALE	.002	1.06 *	1.03	.07	-.03	*	.86	.01	.00 *	.99	.01	.00 *
108-PADRE PIO	.003	.08 *	-.10	-.10	.25	*	.01	.02	.13 *	.11	.13	.76 *
109-PARTIRE POLITICO	.002	.12 *	.31	-.04	.15	*	.08	.00	.04 *	.80	.01	.18 *
110-PECCATO ORIGINALE	.027	.39 *	.61	.01	.12	*	3.34	.00	.28 *	.96	.00	.04 *
111-POCO PAROLA	.003	.71 *	-.66	-.26	-.46	*	.42	.11	.40 *	.61	.10	.29 *
112-POTER CAPIRE	.007	.07 *	-.05	.18	-.18	*	.01	.13	.15 *	.04	.47	.49 *
113-POTER CONTINUARE	.003	.36 *	.38	.00	.47	*	.13	.00	.38 *	.40	.00	.60 *
114-POTER CREDERE	.003	.21 *	.41	-.05	-.18	*	.17	.00	.07 *	.83	.01	.16 *
115-POTER DARE	.006	.04 *	.15	-.09	-.11	*	.04	.03	.04 *	.53	.20	.28 *
116-POTER DIRE	.027	.03 *	.08	-.10	-.11	*	.05	.17	.23 *	.20	.37	.44 *
117-POTER DEMOSTRARE	.003	.10 *	.03	-.12	-.28	*	.00	.02	.14 *	.01	.16	.83 *
118-POTER DIVENTARE	.004	.04 *	.04	-.09	.16	*	.00	.02	.07 *	.05	.21	.74 *
119-POTER FARE	.029	.08 *	.27	.01	-.07	*	.72	.00	.10 *	.93	.00	.06 *
120-POTER LEGGERE	.003	.35 *	-.40	-.21	-.39	*	.16	.08	.31 *	.45	.12	.43 *
121-POTER PARLARE	.003	.03 *	-.09	-.13	-.06	*	.01	.03	.01 *	.27	.58	.15 *
122-POTER PENSARE	.004	.20 *	.45	-.02	.06	*	.25	.00	.01 *	.98	.00	.02 *
123-POTER RISPONDERE	.003	4.49 *	-.35	2.09	-.08	*	.12	7.38	.01 *	.03	.97	.00 *
124-POTER TROVARE	.007	.14 *	.17	.22	-.25	*	.07	.20	.28 *	.21	.35	.44 *
125-POTER VEDERE	.007	.25 *	-.15	-.10	.47	*	.05	.04	.95 *	.09	.04	.87 *
126-POTER VOLARE	.007	.95 *	.97	.07	.08	*	2.29	.02	.03 *	.99	.01	.01 *
127-PRE MORTE	.020	.98 *	.99	.07	.00	*	6.68	.06	.00 *	1.00	.00	.00 *
128-PRIMA PERSONA	.004	.05 *	-.14	-.14	-.09	*	.02	.04	.02 *	.41	.41	.18 *
129-PRIMA VOLTA	.005	.08 *	-.10	-.10	.25	*	.01	.03	.20 *	.11	.13	.76 *
130-PRIMO CASO	.002	.08 *	-.23	-.15	.01	*	.04	.03	.00 *	.71	.29	.00 *
131-RELIGIONE CATTOLICO	.003	.09 *	.04	-.08	.29	*	.00	.01	.17 *	.02	.06	.91 *
132-RELIGIONE CRISTIANO	.003	.02 *	.10	-.09	.00	*	.01	.02	.00 *	.58	.42	.00 *
133-RENDERE CONTO	.006	.16 *	.34	-.06	-.20	*	.22	.01	.16 *	.72	.02	.26 *
134-SAPERE BENE	.002	.21 *	-.41	-.18	-.03	*	.14	.05	.00 *	.83	.16	.01 *
135-SAN GIOVANNI	.004	4.86 *	-.55	-.04	2.13	*	.38	0.011	.37 *	.06	.00	.94 *
136-SAN PAOLO	.007	.70 *	-.53	-.15	.63	*	.71	.10	1.95 *	.40	.03	.56 *
137-SAN PIETRO	.005	1.23 *	-.82	.70	.25	*	1.05	1.34	.20 *	.55	.40	.05 *
138-SCUOLE PUBBLICA	.003	.75 *	.86	.04	-.07	*	.66	.00	.01 *	.99	.00	.01 *
139-SELEZIONE NATURALE	.003	.78 *	-.70	-.27	-.47	*	.54	.14	.48 *	.63	.09	.28 *
140-SEMPLICE CURIOSITÀ	.002	.26 *	.48	-.03	-.17	*	.19	.00	.05 *	.89	.00	.11 *
141-SEMPLICE FARE	.003	.10 *	.17	-.10	-.25	*	.03	.02	.12 *	.30	.09	.60 *
142-SEMPRE DETTO	.002	.26 *	.48	-.03	-.17	*	.19	.00	.05 *	.89	.00	.11 *
143-SENTIRE PARLARE	.007	.16 *	.39	-.04	-.06	*	.34	.01	.02 *	.96	.01	.03 *
144-SENZA BATTESIMO	.004	.14 *	.29	-.07	-.22	*	.12	.01	.13 *	.63	.04	.34 *
145-SENZA DUBBIO	.003	.08 *	.27	-.05	.08	*	.07	.00	.01 *	.88	.03	.08 *
146-SETTEMBRE 1999	.004	.13 *	-.33	-.16	.05	*	.15	.06	.01 *	.79	.19	.02 *
147-SIGNORE GESÙ	.003	.11 *	-.29	-.16	-.03	*	.08	.04	.00 *	.76	.23	.01 *

148-SITO INTERNET	.006	1.90 *	-.93	-.19	1.00	*	1.66	.13	3.82 *	.45	.02	.53 *
149-SOCIETÀ CIVILE	.003	.26 *	.50	-.01	.10	*	.28	.00	.02 *	.96	.00	.04 *
150-SOLO DETTO	.003	.34 *	.57	-.02	-.15	*	.31	.00	.04 *	.94	.00	.06 *
151-SOLO DIO	.003	.19 *	-.29	-.13	.30	*	.07	.03	.16 *	.43	.10	.47 *
152-SOLO DUE	.003	.38 *	.60	-.01	-.14	*	.37	.00	.04 *	.95	.00	.05 *
153-SOLO FARE	.005	.12 *	-.24	-.17	-.19	*	.11	.09	.14 *	.48	.22	.30 *
154-SOLO PAROLA	.002	36.43 *	-1.41	5.83	-.66	*	1.6148	.72	.71 *	.05	.93	.01 *
155-SPAZIO TEMPO	.005	.88 *	.94	.06	-.05	*	1.43	.01	.01 *	.99	.00	.00 *
156-SPECIE DIVERSA	.003	.43 *	-.47	-.22	-.41	*	.19	.08	.29 *	.50	.11	.38 *
157-SPIRITO SANTO	.014	.36 *	-.56	-.20	.05	*	1.54	.35	.02 *	.88	.11	.01 *
158-STATO FARE	.003	.53 *	-.54	-.24	-.42	*	.30	.10	.37 *	.55	.11	.34 *
159-STATO USARE	.003	.54 *	.70	.04	.22	*	.44	.00	.08 *	.91	.00	.09 *
160-TANTO TEMPO	.002	.24 *	.14	-.05	.47	*	.02	.00	.35 *	.08	.01	.91 *
161-TEORIA SCIENTIFICO	.009	.15 *	.32	-.07	-.21	*	.31	.02	.27 *	.67	.03	.30 *
162-TERZO MILLENNIO	.005	.87 *	-.87	-.28	-.18	*	1.35	.24	.11 *	.88	.09	.04 *
163-TESTI BIBLICO	.002	1.32 *	-.97	-.32	-.53	*	.76	.15	.46 *	.71	.08	.22 *
164-TESTI GRECO	.004	1.32 *	-.97	-.32	-.53	*	1.17	.23	.71 *	.71	.08	.22 *
165-TRADUZIONE LETTERAL	.003	1.32 *	-.97	-.32	-.53	*	.96	.19	.58 *	.71	.08	.22 *
166-TRE GIORNO	.004	4.13 *	-.91	-.13	1.81	*	1.04	.04	8.19 *	.20	.00	.80 *
167-UNO SCHIANTO	.014	1.06 *	1.03	.07	-.03	*	4.92	.04	.01 *	.99	.01	.00 *
168-UNO SOLO	.006	.47 *	-.58	-.23	-.28	*	.64	.18	.31 *	.71	.12	.17 *
169-UNO STATO	.012	.54 *	.22	-.01	.70	*	.19	.00	3.98 *	.09	.00	.91 *
170-UNO STUDIO	.003	.25 *	-.13	.48	-.08	*	.02	.47	.01 *	.07	.91	.02 *
171-UOMO DIO	.002	.60 *	-.22	-.09	.74	*	.04	.01	.88 *	.08	.01	.91 *
172-UOMO NUOVO	.004	1.10 *	-.96	-.30	-.30	*	1.16	.20	.22 *	.84	.08	.08 *
173-VENIRE DARE	.003	3.97 *	-1.10	1.63	-.31	*	1.34	5.22	.21 *	.30	.67	.02 *
174-VENIRE FARE	.005	.17 *	.36	-.06	-.20	*	.21	.01	.12 *	.76	.02	.22 *
175-VEDERE SOLO	.005	.17 *	.36	-.06	-.20	*	.21	.01	.12 *	.76	.02	.22 *
176-VEDERE SOPRA	.003	.16 *	-.17	-.16	-.33	*	.03	.05	.24 *	.17	.16	.67 *
177-VERGINE MARIA	.003	.78 *	-.82	-.27	-.21	*	.69	.13	.09 *	.85	.09	.06 *
178-VERO DIO	.007	.09 *	-.25	-.15	.03	*	.14	.09	.00 *	.73	.26	.01 *
179-VIA DIRE	.004	.16 *	-.17	-.16	-.33	*	.04	.07	.32 *	.17	.16	.67 *
180-VIA VIA	.002	.08 *	-.23	-.15	.01	*	.04	.03	.00 *	.71	.29	.00 *
181-VITA ETERNA	.005	.23 *	.21	-.04	.42	*	.08	.00	.64 *	.20	.01	.79 *
182-VITA SPIRITUALE	.003	1.10 *	-.94	-.24	.39	*	.85	.10	.29 *	.81	.05	.14 *
183-VOLERE ANDARE	.003	.15 *	.32	-.07	-.21	*	.10	.01	.09 *	.67	.03	.30 *
184-VOLERE ARRIVARE	.003	.96 *	-.80	-.29	-.49	*	.57	.13	.42 *	.66	.09	.25 *
185-VOLERE CAPIRE	.005	.24 *	.46	-.04	-.17	*	.33	.00	.09 *	.87	.01	.12 *
186-VOLERE CREDERE	.003	.27 *	.50	-.03	-.16	*	.27	.00	.06 *	.90	.00	.10 *
187-VOLERE DARE	.003	.67 *	-.63	-.25	-.45	*	.35	.10	.35 *	.60	.10	.30 *
188-VOLERE DIRE	.037	.06 *	-.18	-.15	-.10	*	.41	.48	.25 *	.51	.34	.16 *
189-VOLERE DISCUTERE	.004	.53 *	.72	.03	.10	*	.72	.00	.03 *	.98	.00	.02 *
190-VOLERE FARE	.017	.05 *	.07	-.11	-.17	*	.03	.12	.32 *	.12	.26	.63 *
191-VOLERE METTERE	.003	.59 *	.76	.02	-.10	*	.64	.00	.02 *	.98	.00	.02 *
192-VOLERE PARLARE	.002	.46 *	.67	.00	-.12	*	.36	.00	.02 *	.97	.00	.03 *
193-VOLERE SAPERE	.005	.15 *	.21	.32	-.08	*	.08	.32	.02 *	.29	.67	.04 *
194-VOLERE SOLO	.006	.04 *	-.04	-.13	-.15	*	.00	.05	.08 *	.04	.40	.55 *
195-VOLERE VEDERE	.004	.14 *	-.22	-.12	.28	*	.07	.04	.22 *	.33	.11	.56 *

### **3.6. Trattamento dell'output dell'APLUM (analisi delle corrispondenze semplici)**

Come si può vedere nelle pagine dedicate all'output della procedura APLUM per le parole e per i *segmenti*, gli elementi in analisi sono talmente numerosi (764 parole e 195 *segmenti*) da rendere difficoltosa l'interpretazione del grafico fattoriale.

Si è ritenuto opportuno, quindi, adottare un criterio che consentisse di selezionare le parole ed i *segmenti* più significativi allo scopo di aumentare la fruibilità della rappresentazione grafica.

Si riportano, qui di seguito, le strategie adottate per le parole e per i *segmenti* che risultano essere fra loro differenti per alcuni punti.

Una nota: per entrambi i processi è stato necessario l'uso di un *Text-editor* (Textpad® reperibile alla URL [www.textpad.com](http://www.textpad.com)) per il trattamento dei file di testo dell'output dello Spad-T® al fine di rendere questi ultimi idonei ad un *porting* negli applicativi Excel® e SPSS® attraverso i quali si è potuto concretizzare il criterio di selezione.

- **Output APLUM (parole)**

1. Eliminazione delle parole con contributi assoluti  $\leq 0,05$ ;
2. Ordinamento delle parole secondo il contributo assoluto dell'asse 1, dell'asse 2 e dell'asse 3;

3. Eliminazione delle parole dal basso contributo assoluto e dal significato ritenuto troppo generico per la valutazione del contenuto del *discorso* (Vedi Tabella 7 a pag. 132);
4. Selezione delle parole con una condizione filtro basata sugli operatori logici che desse come effetto l'unione di tre insiemi<sup>75</sup>: quello delle parole con contributo assoluto sull'asse 1  $\geq 0,4$ , quello delle parole con contributo assoluto sull'asse 2  $\geq 0,4$  e quello delle parole con contributo assoluto sull'asse 3  $\geq 0,4$ ;
5. Rappresentazione grafica su ( $F_1$ ,  $F_2$ ) delle parole di cui al punto 4. (Cfr. il Grafico 1 più avanti);
6. Selezione delle parole con un'altra condizione filtro<sup>76</sup> che desse come risultato l'insieme unione delle parole con contributi assoluti sull'asse 1, 2 e 3  $\geq 0,1$  intersecato con il negato dell'insieme di cui al punto 4.;
7. Rappresentazione grafica su ( $F_1$ ,  $F_2$ ) delle parole di cui al punto 6 (Cfr. il Grafico 3 più avanti).

- **Output APLUM (*segmenti*)**

1. Eliminazione dei *segmenti* con contributi assoluti  $\leq 0,03$ ;
2. Ordinamento dei segmenti secondo il contributo assoluto dell'asse 1, dell'asse 2 e dell'asse 3;

---

<sup>75</sup> La condizione nell'SPSS® è la seguente:

(contr.ass.asse\_1 $\geq 0,4$ |contr.ass.asse\_2 $\geq 0,4$ |contr.ass.asse\_3 $\geq 0,4$ )

<sup>76</sup> La condizione nell'SPSS® è la seguente:

[(contr.ass.asse\_1 $\geq 0,1$ |contr.ass.asse\_2 $\geq 0,1$ |contr.ass.asse\_3 $\geq 0,1$ ) AND NOT (contr.ass.asse\_1 $\geq 0,4$ |contr.ass.asse\_2 $\geq 0,4$ |contr.ass.asse\_3 $\geq 0,4$ )]

3. Eliminazione dei *segmenti* dal basso contributo assoluto e dal significato ritenuto troppo generico per la valutazione del contenuto del *discorso* (Vedi Tabella 8 a pag.138);
4. Rappresentazione grafica su ( $F_1$ ,  $F_2$ ) dei *segmenti* di cui al punto 3. (Cfr. il Grafico 5 più avanti).

Nel paragrafo 3.7 a pag. 153 sono contenuti i commenti ai grafici e alle tabelle di seguito riportate.

**Tabella 7:** Elenco delle parole eliminate contraddistinte da un basso contributo assoluto e dal significato ritenuto troppo generico per la valutazione del contenuto del *discorso*.

Parola	Contr. assol. F1	Contr. assol. F2	Contr. assol. F3
MISSIONE	0,00	0,00	0,06
BREVE	0,00	0,00	0,07
ALTO	0,00	0,00	0,08
LETTO	0,00	0,00	0,08
LUNGO	0,00	0,00	0,09
DIREZIONE	0,00	0,00	0,09
TRAMITE	0,00	0,00	0,12
VOLTE	0,00	0,00	0,12
RIPETO	0,00	0,00	0,18
TIMOR	0,00	0,00	0,18
OTTENERE	0,00	0,00	0,21
CONTRO	0,00	0,00	0,22
PARERE	0,00	0,01	0,06
GIUDICARE	0,00	0,01	0,06
INTERA	0,00	0,01	0,08
FIGURA	0,00	0,01	0,10
SCUSA	0,00	0,01	0,19
SETTE	0,00	0,01	0,47
STRANO	0,00	0,02	0,10
MILIONI	0,00	0,02	0,11
COMUNE	0,00	0,02	0,15
CAPITOLO	0,00	0,03	0,06
CERTI	0,00	0,03	0,07
ARRIVARE	0,00	0,03	0,10
SCELTA	0,00	0,03	0,16
MAGARI	0,00	0,03	0,22
VERSO	0,00	0,03	0,23
CHIEDERE	0,00	0,04	0,11
BASE	0,00	0,04	0,14
SALUTO	0,00	0,04	0,18
ATTIVITÀ	0,00	0,04	0,34
LEGGE	0,00	0,06	0,00
PRINCIPALE	0,00	0,06	0,00
UNIVERSALE	0,00	0,06	0,02
FONDO	0,00	0,06	0,11
SEGUITO	0,00	0,06	0,11
EGLI	0,00	0,06	0,11
CHIARO	0,00	0,06	0,13
MESSA	0,00	0,07	0,00
RICEVUTO	0,00	0,07	0,00
PICCOLO	0,00	0,07	0,00
AUTORE	0,00	0,07	0,00
IMPLICA	0,00	0,07	0,02
METTERE	0,00	0,07	0,04
NOTO	0,00	0,07	0,09
THREAD	0,00	0,07	0,14
MIGLIAIA	0,00	0,08	0,01

PIANO	0,00	0,09	0,00
PIO	0,00	0,09	0,00
SOLO	0,00	0,09	0,24
MANIERA	0,00	0,10	0,01
VERBO	0,00	0,10	0,01
FEDELI	0,00	0,10	0,03
PROCESSI	0,00	0,10	0,03
RINGRAZIO	0,00	0,10	0,07
POSSIBILE	0,00	0,10	0,09
SCUOLE	0,00	0,11	0,00
ARTICOLO	0,00	0,11	0,00
SOLUZIONE	0,00	0,11	0,01
MATRIMONIO	0,00	0,11	0,03
ESPRESSIONE	0,00	0,11	0,06
RISPOSTO	0,00	0,11	0,08
CHIAMA	0,00	0,13	0,00
RUOLO	0,00	0,13	0,05
MANIFESTO	0,00	0,17	0,16
ANALISI	0,00	0,19	0,05
NESSUNO	0,00	0,22	0,13
FILM	0,00	0,33	0,25
CRITICA	0,01	0,00	0,06
VISSUTO	0,01	0,00	0,06
SOPRA	0,01	0,00	0,06
DIECI	0,01	0,00	0,08
INTENDERE	0,01	0,00	0,08
EVIDENZA	0,01	0,00	0,09
FUTURO	0,01	0,00	0,14
RITENERE	0,01	0,00	0,14
PAESE	0,01	0,00	0,18
MORTO	0,01	0,00	0,19
CENTRO	0,01	0,00	0,29
CONDIZIONE	0,01	0,01	0,03
CONOSCERE	0,01	0,01	0,03
ELEMENTI	0,01	0,01	0,04
MALE	0,01	0,01	0,04
DECISO	0,01	0,01	0,06
DIVINA	0,01	0,01	0,07
PROPOSITO	0,01	0,01	0,07
CASO	0,01	0,01	0,08
DUBBIO	0,01	0,01	0,08
PARTICOLARE	0,01	0,01	0,09
BENE	0,01	0,01	0,11
DIFFERENZA	0,01	0,01	0,13
SENTIRE	0,01	0,01	0,15
SPERO	0,01	0,01	0,15
LAVORO	0,01	0,01	0,30
SIGNIFICA	0,01	0,02	0,05
ATTO	0,01	0,02	0,05
MANO	0,01	0,03	0,00
RICHIESTA	0,01	0,03	0,01
BISOGNO	0,01	0,03	0,01

POLITICA	0,01	0,03	0,12
FISICA	0,01	0,03	0,15
OGGETTO	0,01	0,03	0,31
GRADO	0,01	0,04	0,00
USO	0,01	0,04	0,00
TOTALE	0,01	0,04	0,00
IERI	0,01	0,04	0,00
IMPORTA	0,01	0,04	0,00
ANDARE	0,01	0,04	0,00
VARI	0,01	0,04	0,01
TESTA	0,01	0,04	0,01
CONFRONTO	0,01	0,04	0,02
CAMPO	0,01	0,04	0,04
SCRIVERE	0,01	0,04	0,08
CERTE	0,01	0,04	0,13
SENSO	0,01	0,04	0,20
DIVERSA	0,01	0,05	0,00
VISTO	0,01	0,05	0,08
INDIVIDUO	0,01	0,06	0,00
TEMPO	0,01	0,06	0,00
MERITO	0,01	0,06	0,07
DISCUTERE	0,01	0,07	0,01
ASSOLUTA	0,01	0,07	0,04
CADERE	0,01	0,07	0,36
PARE	0,01	0,07	0,43
PASSO	0,01	0,08	0,00
INTERNO	0,01	0,08	0,00
CONSIDERARE	0,01	0,08	0,00
DIFFICOLTÀ	0,01	0,08	0,11
CARO	0,01	0,08	0,17
DIPENDE	0,01	0,10	0,00
INTERVENTO	0,01	0,10	0,00
GIOIA	0,01	0,10	0,17
CASINI	0,01	0,11	0,00
MAGGIORANZA	0,01	0,12	0,00
CULTURA	0,01	0,12	0,01
IMPEGNO	0,01	0,12	0,03
POSIZIONE	0,01	0,12	0,12
PERFETTO	0,01	0,13	0,01
UNIONE	0,01	0,13	0,07
ASPETTO	0,01	0,15	0,02
LETTERALE	0,01	0,15	0,03
FUNZIONE	0,01	0,16	0,07
DOMANDA	0,01	0,21	0,02
MEZZO	0,01	0,22	0,02
VOCE	0,01	0,22	0,03
SOLDI	0,01	0,24	0,06
UMANO	0,02	0,00	0,06
RENDERE	0,02	0,00	0,10
ORIGINALE	0,02	0,00	0,10
CARATTERE	0,02	0,00	0,14
AZIONE	0,02	0,00	0,18

CAUSA	0,02	0,00	0,49
LISTA	0,02	0,01	0,00
SUFFICIENTE	0,02	0,01	0,00
CAPACITÀ	0,02	0,01	0,01
POSSIBILITÀ	0,02	0,01	0,01
RISULTA	0,02	0,01	0,02
PUBBLICA	0,02	0,01	0,03
GRAVE	0,02	0,01	0,03
CARATTERISTICHE	0,02	0,01	0,06
SOSTENERE	0,02	0,01	0,07
PARLARE	0,02	0,01	0,11
SEDE	0,02	0,02	0,00
VALE	0,02	0,02	0,00
RAGIONI	0,02	0,02	0,02
PARI	0,02	0,02	0,04
RICORDO	0,02	0,02	0,07
PRINCIPI	0,02	0,02	0,14
CONSIDERAZIONE	0,02	0,05	0,01
TERMINE	0,02	0,05	0,01
USARE	0,02	0,05	0,03
SETTEMBRE	0,02	0,07	0,00
PASSAGGIO	0,02	0,07	0,02
SOGNO	0,02	0,08	0,03
NEW	0,02	0,09	0,01
EFFETTO	0,02	0,10	0,00
IMPORTANZA	0,02	0,12	0,00
CONTINUARE	0,02	0,13	0,04
LEGGI	0,02	0,14	0,00
IDEA	0,02	0,21	0,00
PERDERE	0,03	0,00	0,08
TROVARE	0,03	0,00	0,10
PRATICHE	0,03	0,00	0,10
TANTO	0,03	0,00	0,36
PERIODO	0,03	0,00	0,45
DISCORSO	0,03	0,01	0,09
QUATTRO	0,03	0,01	0,15
LATINO	0,03	0,13	0,08
CERCARE	0,04	0,00	0,07
PENSARE	0,04	0,00	0,09
COLPA	0,04	0,00	0,10
AUTORITÀ	0,04	0,00	0,11
PUBBLICO	0,04	0,01	0,06
ACCORDO	0,04	0,01	0,11
REALE	0,04	0,05	0,08
TERRA	0,04	0,05	0,14
MOMENTO	0,04	0,06	0,03
STAMPA	0,04	0,06	0,04
RESPONSABILE	0,04	0,07	0,07
SBAGLIATO	0,04	0,10	0,02
DUEMILA	0,04	0,11	0,01
CURIOSITÀ	0,04	0,14	0,01
RISPONDERE	0,04	0,14	0,02

PROBLEMA	0,04	0,15	0,05
LETTERA	0,04	0,16	0,01
SOGGETTO	0,04	0,18	0,08
GIORNO	0,04	0,28	0,84
IMHO	0,05	0,00	0,14
COMUNITÀ	0,05	0,01	0,00
MENO	0,05	0,01	0,14
INTERESSANTE	0,05	0,06	0,00
FAVORE	0,05	0,08	0,02
BELLO	0,05	0,09	0,04
FRASE	0,05	0,09	0,21
GRANDE	0,05	0,20	0,16
FORSE	0,06	0,00	0,22
DIRITTO	0,06	0,01	0,00
VICINO	0,06	0,01	0,06
PREFERISCO	0,06	0,03	0,05
ORA	0,06	0,04	0,04
SOLITO	0,06	0,06	0,01
MILIARDI	0,06	0,07	0,01
COMMESSO	0,06	0,12	0,03
NULLA	0,06	0,14	0,01
INDIRIZZO	0,06	0,35	0,06
DETTO	0,07	0,01	0,27
QUESTIONE	0,07	0,02	0,00
STUDIO	0,07	0,02	0,04
UNICO	0,07	0,03	0,01
PARTIRE	0,07	0,03	0,03
SEMPLICE	0,07	0,04	0,00
VOLTA	0,07	0,05	0,15
CONSEGUENZA	0,07	0,07	0,05
ASSURDO	0,07	0,10	0,00
LETTURA	0,07	0,11	0,01
TIPO	0,07	0,12	0,00
AFFERMAZIONE	0,07	0,16	0,04
INGLESE	0,07	0,16	0,06
MATERIA	0,07	0,19	0,14
DISCUSSIONE	0,08	0,00	0,00
INTERNET	0,08	0,00	0,03
SEMBRARE	0,08	0,02	0,15
1999	0,08	0,05	0,07
VOLERE	0,08	0,05	0,09
VISTA	0,08	0,08	0,00
INFORMAZIONE	0,08	0,11	0,00
CONTESTO	0,08	0,14	0,15
PERMESSO	0,09	0,00	0,00
MAGGIORE	0,09	0,00	0,01
CITATO	0,09	0,00	0,03
SCUSATE	0,09	0,03	0,03
FENOMENO	0,09	0,07	0,01
GENERALE	0,09	0,09	0,00
CONVINTO	0,09	0,09	0,00
AFFERMARE	0,09	0,12	0,01

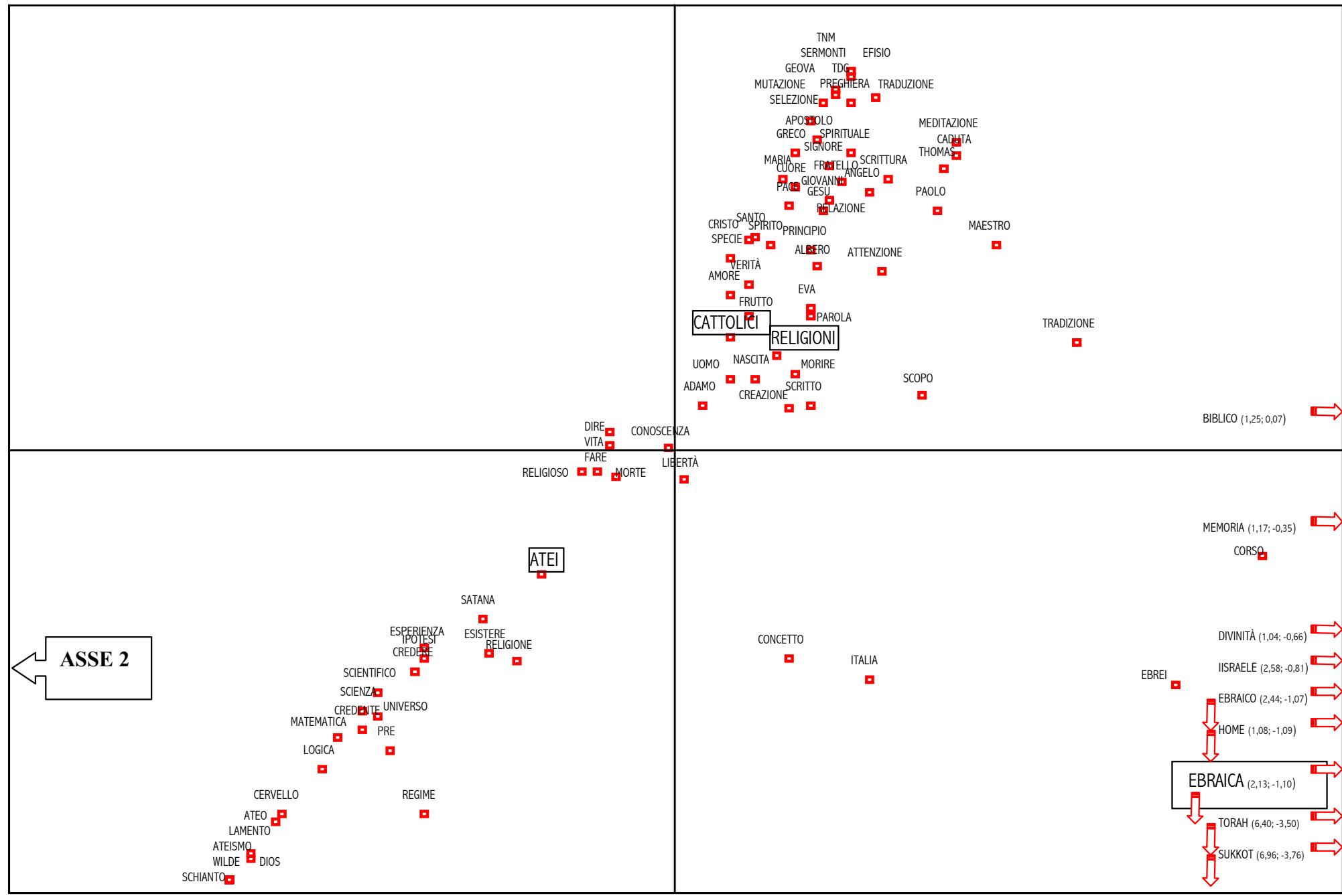
PUNTI	0,09	0,13	0,07
FISICO	0,09	0,14	0,01
SITO	0,09	0,15	0,09
VALORE	0,10	0,00	0,00
INTERPRETAZIONE	0,10	0,00	0,03
MOVIMENTO	0,10	0,01	0,13
POCO	0,11	0,01	0,07
TRE	0,11	0,01	0,08
ULTIMO	0,11	0,02	0,04
OTTOBRE	0,11	0,07	0,25
OGGI	0,11	0,09	0,22
PROVA	0,11	0,15	0,06

**Tabella 8:** Elenco dei *segmenti* eliminati contraddistinti da un basso contributo assoluto e dal significato ritenuto troppo generico per la valutazione del contenuto del *discorso*.

Segmenti	Contr. assol. F1	Contr. assol. F2	Contr. assol. F1
POTER DIVENTARE	0	0,02	0,07
POTER DEMONSTRARE	0	0,02	0,14
DOVERE SENTIRE	0	0,03	0,15
DIECI ANNO	0	0,03	0,16
VOLERE SOLO	0	0,05	0,08
DIRE SOLO	0,01	0,02	0,11
FARE TANTO	0,01	0,02	0,11
TANTO TEMPO	0,02	0	0,35
PRIMA PERSONA	0,02	0,04	0,02
UNO STUDIO	0,02	0,47	0,01
DUE ANNO	0,02	0,63	0,12
MAGGIORE PARTE	0,02	3,39	0,6
SEMPLICE FARE	0,03	0,02	0,12
ANNO PRIMA	0,03	0,02	0,14
VEDERE SOPRA	0,03	0,05	0,24
VOLERE FARE	0,03	0,12	0,32
VIA VIA	0,04	0,03	0
PRIMO CASO	0,04	0,03	0
POTER DARE	0,04	0,03	0,04
VIA DIRE	0,04	0,07	0,32
FORSE DOVERE	0,05	0,94	0,09
DOVERE SOLO	0,06	0,01	0
MENO MALE	0,06	0,06	0,28
SENZA DUBBIO	0,07	0	0,01
LASCIARE PERDERE	0,07	0,01	0,1
VOLERE VEDERE	0,07	0,04	0,22
GRANDE PARTE	0,08	0,03	0,3
DOVERE DIRE	0,08	0,06	0,28
VOLERE SAPERE	0,08	0,32	0,02
FARE SOLO	0,09	0,06	0
VOLERE ANDARE	0,1	0,01	0,09
SOLO FARE	0,11	0,09	0,14
CASO PARTICOLARE	0,12	0,01	0,07
NESSUNO TIPO	0,12	0,01	0,07
MODO MENO	0,14	0,05	0
BENE FARE	0,14	0,07	0,25
DOVERE ANDARE	0,14	0,38	0,22
SETTEMBRE 1999	0,15	0,06	0,01
DOVERE FARE	0,15	0,08	0,13
POTER LEGGERE	0,16	0,08	0,31
POTER CREDERE	0,17	0	0,07
SEMPRE DETTO	0,19	0	0,05
CARO SALUTO	0,19	0,08	0,29
VENIRE FARE	0,21	0,01	0,12
VEDERE SOLO	0,21	0,01	0,12
FARE NULLA	0,25	0	0,01
POTER PENSARE	0,25	0	0,01

BELLO DIFFERENZA	0,25	0	0,04
GUARDARE CASO	0,26	0	0,05
DIRE NULLA	0,3	0	0,08
FARE MALE	0,3	0,08	0,11
STATO FARE	0,3	0,1	0,37
SOLO DETTO	0,31	0	0,04
CERTO SENSO	0,31	0,12	0,04
VOLERE CAPIRE	0,33	0	0,09
SENTIRE PARLARE	0,34	0,01	0,02
POTER VEDERE	0,05	0,04	0,95
POTER DIRE	0,05	0,17	0,23
POTER TROVARE	0,07	0,2	0,28
PARTIRE POLITICO	0,08	0	0,04
CAPIRE BENE	0,09	0,03	0,22
FARE NOTARE	0,14	0,02	0
SAPERE BENE	0,14	0,05	0
SEMPLICE CURIOSITÀ	0,19	0	0,05
BENE POCO	0,23	0,14	0,03
VOLERE CREDERE	0,27	0	0,06
DUE VOLTE	0,34	0,46	0,03
FARE VEDERE	0,35	0,01	0,11
DOVERE DARE	0,35	0,1	0,35
VOLERE DARE	0,35	0,1	0,35
SOLO DUE	0,37	0	0,04
VOLERE DIRE	0,41	0,48	0,25
FARE PASSARE	0,42	0,11	0,4
OTTOBRE 1999	0,42	2,82	4,32
FARE BENE	0,46	0,19	0,74
BENE PRESTO	0,69	0,13	0,09
POTER FARE	0,72	0	0,1
GRAZIE GRAZIE	0,76	0,15	0,46
HOME PAGE	1,03	0	0,01
TRE GIORNO	1,04	0,04	8,19
SITO INTERNET	1,66	0,13	3,82
DIRE OGGI	1,72	0,33	1,04
POTER VOLARE	2,29	0,02	0,03

# Grafico 1 Analisi delle corrispondenze lessicali: parole ad alto contributo assoluto



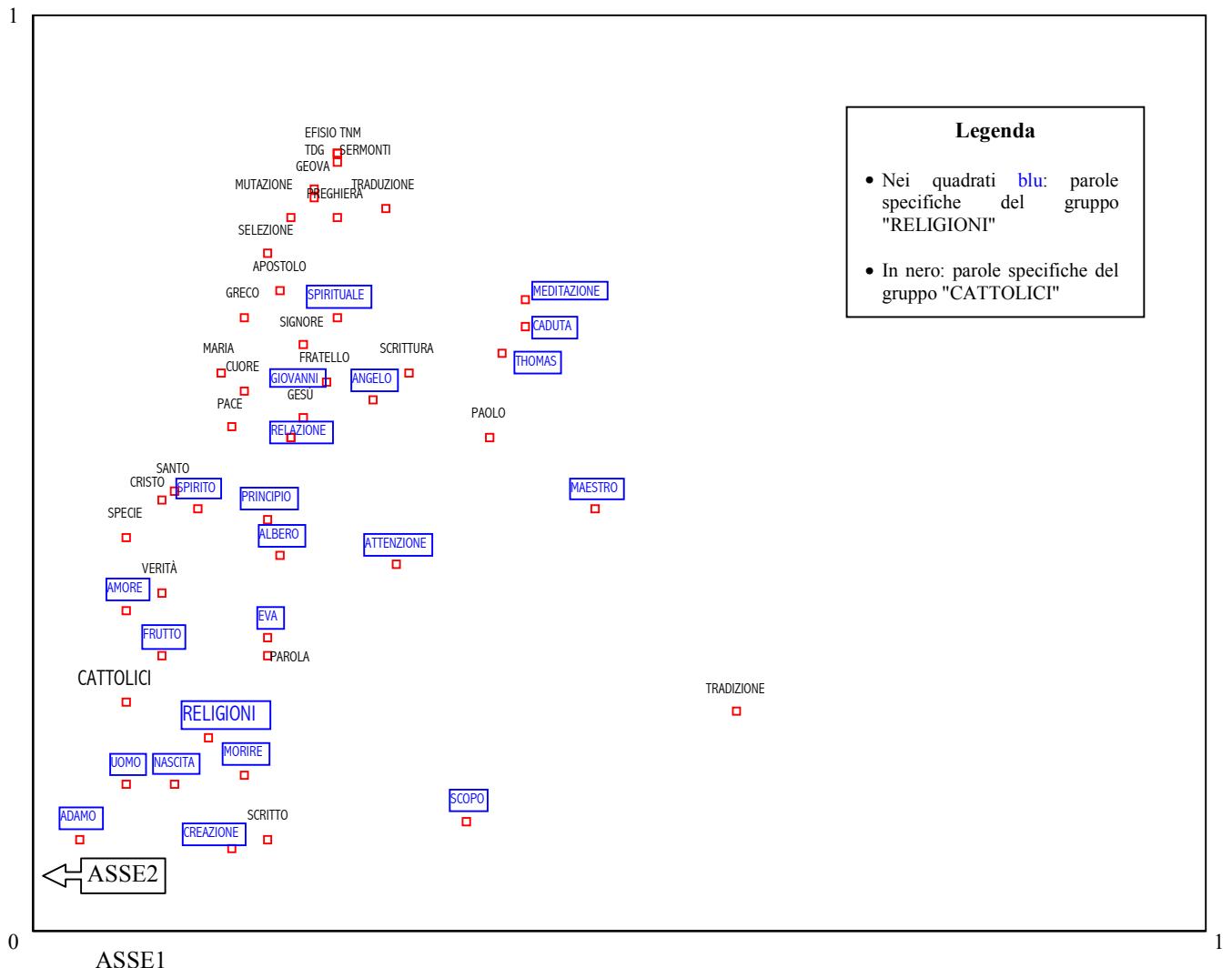
**Tabella 9:** Descrizione dell'ASSE 1 Grafico 1. Parole ordinate rispetto alla distanza dall'origine

Parola	Massa	Coord.	Parola	Massa	Coord.
DIOS	0	-0,67	SCRITTO	0,01	0,2
WILDE	0	-0,67	APOSTOLO	0	0,21
SCHIANTO	0	-0,67	ALBERO	0	0,21
ATEISMO	0	-0,64	GESÙ	0,01	0,22
LAMENTO	0	-0,64	MUTAZIONE	0	0,22
ATEO	0	-0,6	RELAZIONE	0	0,22
CERVELLO	0	-0,59	SIGNORE	0	0,23
LOGICA	0	-0,53	GIOVANNI	0	0,23
MATEMATICA	0	-0,51	TDG	0	0,24
CREDENTE	0	-0,47	GEOVA	0	0,24
SCIENZA	0	-0,47	FRATELLO	0	0,25
SCIENTIFICO	0	-0,45	PREGHIERA	0	0,26
UNIVERSO	0	-0,45	SPIRITUALE	0	0,26
PRE	0	-0,43	TNM	0	0,26
CREDERE	0,01	-0,39	EFISIO	0	0,26
ESPERIENZA	0	-0,38	SERMONTI	0	0,26
REGIME	0	-0,38	ANGELO	0	0,29
IPOTESI	0	-0,38	ITALIA	0	0,29
SATANA	0	-0,29	TRADUZIONE	0	0,3
ESISTERE	0	-0,28	ATTENZIONE	0	0,31
RELIGIONE	0,01	-0,24	SCRITTURA	0	0,32
ATEI	0,42	-0,2	SCOPO	0	0,37
RELIGIOSO	0	-0,14	PAOLO	0	0,39
FARE	0,03	-0,12	THOMAS	0	0,4
DIRE	0,02	-0,1	CADUTA	0	0,42
VITA	0,01	-0,1	MEDITAZIONE	0	0,42
MORTE	0	-0,09	MAESTRO	0	0,48
CONOSCENZA	0	-0,01	TRADIZIONE	0	0,6
<b>ZONA CENTRALE</b>			EBREI	0	0,75
LIBERTÀ	0	0,01	CORSO	0	0,88
ADAMO	0	0,04	DIVINITÀ	0	1,04
CATTOLICI	0,46	0,08	HOME	0	1,08
AMORE	0	0,08	MEMORIA	0	1,17
SPECIE	0	0,08	BIBLICO	0	1,25
UOMO	0,01	0,08	<u>EBRAICA</u>	0,02	2,13
CRISTO	0,01	0,11	EBRAICO	0	2,44
VERITÀ	0	0,11	ISRAELE	0	2,58
FRUTTO	0	0,11	TORAH	0	6,4
SANTO	0	0,12	SUKKOT	0	6,96
NASCITA	0	0,12			
SPIRITO	0	0,14			
RELIGIONI	0,11	0,15			
MARIA	0	0,16			
PACE	0	0,17			
CONCETTO	0	0,17			
CREAZIONE	0	0,17			
CUORE	0	0,18			
GRECO	0	0,18			
MORIRE	0	0,18			
PAROLA	0,01	0,2			
SELEZIONE	0	0,2			
PRINCIPIO	0	0,2			
EVA	0	0,2			

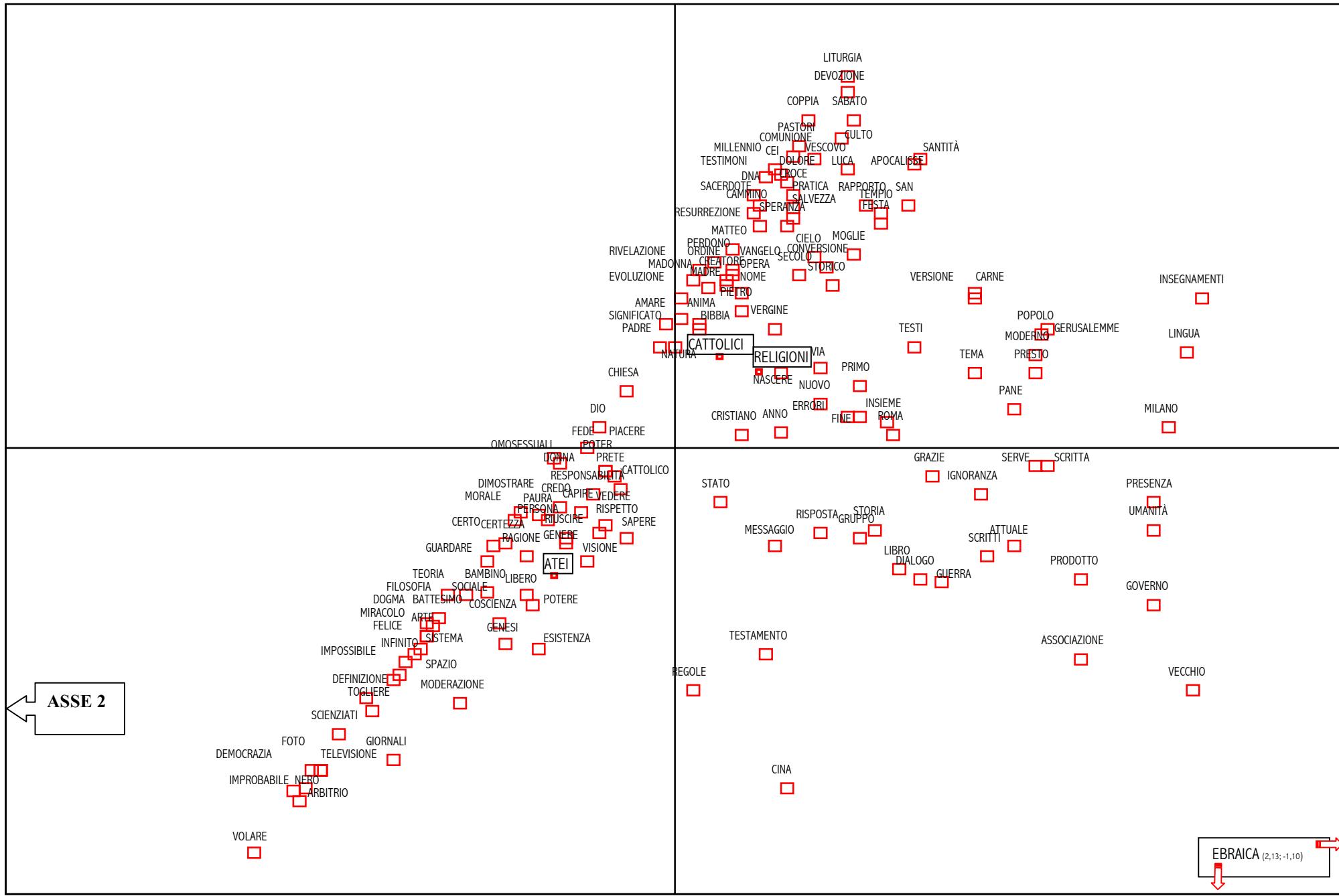
**Tabella 10:** Descrizione dell'ASSE 2 Grafico 1. Parole ordinate rispetto alla distanza dall'origine

Parola	Massa	Coord.	Parola	Massa	Coord.
SUKKOT	0	-3,76	PAROLA	0,01	0,3
TORAH	0	-3,5	FRUTTO	0	0,3
EBRAICA	0,02	-1,1	EVA	0	0,32
HOME	0	-1,09	AMORE	0	0,35
EBRAICO	0	-1,07	VERITÀ	0	0,37
DIOS	0	-0,97	ATTENZIONE	0	0,4
WILDE	0	-0,97	ALBERO	0	0,41
SCHIANTO	0	-0,97	SPECIE	0	0,43
LAMENTO	0	-0,92	PRINCIPIO	0	0,45
ATEISMO	0	-0,91	SPIRITO	0	0,46
ATEO	0	-0,84	MAESTRO	0	0,46
CERVELLO	0	-0,82	CRISTO	0,01	0,47
REGIME	0	-0,82	SANTO	0	0,48
ISRAELE	0	-0,81	GESÙ	0,01	0,54
LOGICA	0	-0,72	PAOLO	0	0,54
PRE	0	-0,68	RELAZIONE	0	0,54
DIVINITÀ	0	-0,66	PACE	0	0,55
MATEMATICA	0	-0,65	GIOVANNI	0	0,56
CREDENTE	0	-0,63	ANGELO	0	0,58
UNIVERSO	0	-0,6	CUORE	0	0,59
SCIENZA	0	-0,59	FRATELLO	0	0,6
SCIENTIFICO	0	-0,55	MARIA	0	0,61
EBREI	0	-0,53	SCRITTURA	0	0,61
ITALIA	0	-0,52	THOMAS	0	0,63
CREDERE	0,01	-0,5	SIGNORE	0	0,64
RELIGIONE	0,01	-0,48	CADUTA	0	0,66
IPOTESI	0	-0,47	SPIRITUALE	0	0,67
CONCETTO	0	-0,47	GRECO	0	0,67
ESISTERE	0	-0,46	MEDITAZIONE	0	0,69
ESPERIENZA	0	-0,45	APOSTOLO	0	0,7
SATANA	0	-0,38	SELEZIONE	0	0,74
MEMORIA	0	-0,35	MUTAZIONE	0	0,78
ATEI	0,42	-0,28	PREGHIERA	0	0,78
CORSO	0	-0,24	TRADUZIONE	0	0,79
LIBERTÀ	0	-0,07	TDG	0	0,8
MORTE	0	-0,06	GEOVA	0	0,81
FARE	0,03	-0,05	SERMONTI	0	0,84
RELIGIOSO	0	-0,05	TNM	0	0,85
<b>ZONA CENTRALE</b>			EFISIO	0	0,85
CONOSCENZA	0	0			
VITA	0,01	0,01			
DIRE	0,02	0,04			
BIBLICO	0	0,07			
CREAZIONE	0	0,09			
SCRITTO	0,01	0,1			
ADAMO	0	0,1			
SCOPO	0	0,12			
UOMO	0,01	0,16			
NASCITA	0	0,16			
MORIRE	0	0,17			
RELIGIONI	0,11	0,21			
TRADIZIONE	0	0,24			
CATTOLICI	0,46	0,25			

**Grafico 2** Analisi delle corrispondenze lessicali: parole ad alto contributo assoluto, quadrante positivo e specificità



### Grafico 3 Analisi delle corrispondenze lessicali: parole a basso contributo assoluto



**Tabella 11:** Descrizione dell'ASSE 1 Grafico 3. Parole ordinate rispetto alla distanza dall'origine

Parola	Massa	Coord.	Parola	Massa	Coord.
VOLARE	0	-0,63	PRETE	0	-0,09
IMPROBABILE	0	-0,57	CATTOLICO	0,01	-0,08
ARBITRIO	0	-0,56	SAPERE	0	-0,07
NERO	0	-0,55	CHIESA	0,01	-0,07
DEMOCRAZIA	0	-0,54	PADRE	0	-0,02
FOTO	0	-0,53	AMARE	0	-0,01
TELEVISIONE	0	-0,53	<b>ZONA CENTRALE</b>		
SCIENZIATI	0	-0,5	NATURA	0	0
DEFINIZIONE	0	-0,46	EVOLUZIONE	0	0,01
TOGLIERE	0	-0,45	SIGNIFICATO	0	0,01
GIORNALI	0	-0,42	REGOLE	0	0,03
SPAZIO	0	-0,42	RIVELAZIONE	0	0,03
IMPOSSIBILE	0	-0,41	BIBBIA	0	0,04
INFINITO	0	-0,4	MADONNA	0	0,04
SISTEMA	0	-0,39	ANIMA	0	0,04
FELICE	0	-0,38	MADRE	0	0,05
DOGMA	0	-0,37	PERDONO	0	0,06
MIRACOLO	0	-0,37	STATO	0,01	0,07
ARTE	0	-0,37	CATTOLICI	0,46	0,08
BATTESIMO	0	-0,36	OPERA	0	0,08
FILOSOFIA	0	-0,35	CREATORE	0	0,08
TEORIA	0	-0,34	VANGELO	0	0,09
MODERAZIONE	0	-0,32	ORDINE	0	0,09
SOCIALE	0	-0,31	MATTEO	0	0,09
BAMBINO	0	-0,28	NOME	0	0,1
GUARDARE	0	-0,28	PIETRO	0	0,1
CERTO	0	-0,27	CRISTIANO	0,01	0,1
COSCIENZA	0	-0,26	DNA	0	0,12
GENESI	0	-0,25	CAMMINO	0	0,12
CERTEZZA	0	-0,25	SACERDOTE	0	0,13
MORALE	0	-0,24	RESURREZIONE	0	0,13
DIMOSTRARE	0	-0,23	TESTIMONI	0	0,14
LIBERO	0	-0,22	TESTAMENTO	0	0,14
RAGIONE	0	-0,22	RELIGIONI	0,11	0,15
POTERE	0	-0,21	MESSAGGIO	0	0,15
ATEI	0,42	-0,2	CEI	0	0,15
ESISTENZA	0	-0,2	VERGINE	0	0,15
PAURA	0	-0,2	MILLENNIO	0	0,16
PERSONA	0,01	-0,19	NASCERE	0	0,16
OMOSESSUALI	0	-0,18	ANNO	0,01	0,16
CREDO	0	-0,17	DOLORE	0	0,17
DONNA	0	-0,17	CINA	0	0,17
GENERE	0	-0,16	SPERANZA	0	0,17
RIUSCIRE	0	-0,16	PRATICA	0	0,18
CAPIRE	0	-0,14	SALVEZZA	0	0,18
VISIONE	0	-0,13	CROCE	0	0,18
FEDE	0,01	-0,13	COMUNIONE	0	0,18
RESPONSABILITÀ	0	-0,12	SECOLO	0	0,19
RISPETTO	0	-0,11	PASTORI	0	0,19
DIO	0,02	-0,11	COPPIA	0	0,2
VEDERE	0,01	-0,1	CIELO	0	0,21
POTER	0,01	-0,1	VESCOVO	0	0,21
PIACERE	0	-0,1	RISPOSTA	0	0,22
			VIA	0	0,22

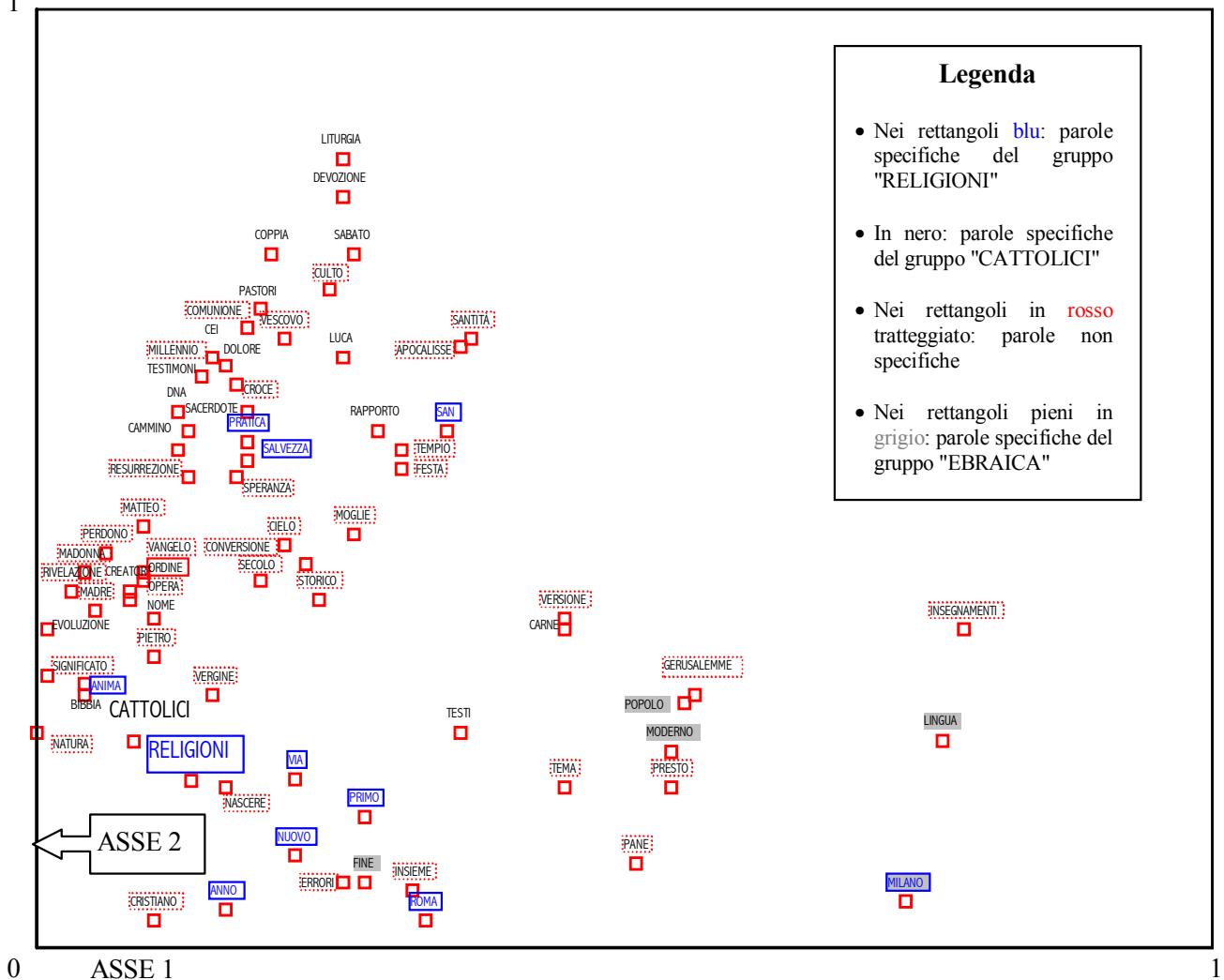
			NUOVO	0	0,22
			CONVERSIONE	0	0,23
			STORICO	0	0,24
			CULTO	0	0,25
			LITURGIA	0	0,26
			LUCA	0	0,26
			DEVOZIONE	0	0,26
			ERRORI	0	0,26
			SABATO	0	0,27
			MOGLIE	0	0,27
			GRUPPO	0	0,28
			PRIMO	0	0,28
			FINE	0	0,28
			RAPPORTO	0	0,29
			STORIA	0	0,3
			TEMPIO	0	0,31
			FESTA	0	0,31
			INSIEME	0	0,32
			ROMA	0	0,33
			LIBRO	0	0,34
			SAN	0	0,35
			TESTI	0	0,36
			APOCALISSE	0	0,36
			SANTITÀ	0	0,37
			DIALOGO	0	0,37
			GRAZIE	0	0,39
			GUERRA	0	0,4
			CARNE	0	0,45
			VERSIONE	0	0,45
			TEMA	0	0,45
			IGNORANZA	0	0,46
			SCRITTI	0	0,47
			ATTUALE	0	0,51
			PANE	0	0,51
			MODERNO	0	0,54
			PRESTO	0	0,54
			SERVE	0	0,54
			POPOLO	0	0,55
			GERUSALEMME	0	0,56
			SCRITTA	0	0,56
			ASSOCIAZIONE	0	0,61
			PRODOTTO	0	0,61
			GOVERNO	0	0,72
			UMANITÀ	0	0,72
			PRESENZA	0	0,72
			MILANO	0	0,74
			LINGUA	0	0,77
			VECCHIO	0	0,78
			INSEGNAMENTI	0	0,79
			<u>EBRAICA</u>	0	2,13

**Tabella 12:** Descrizione dell'ASSE 2 Grafico 3. Parole ordinate rispetto alla distanza dall'origine

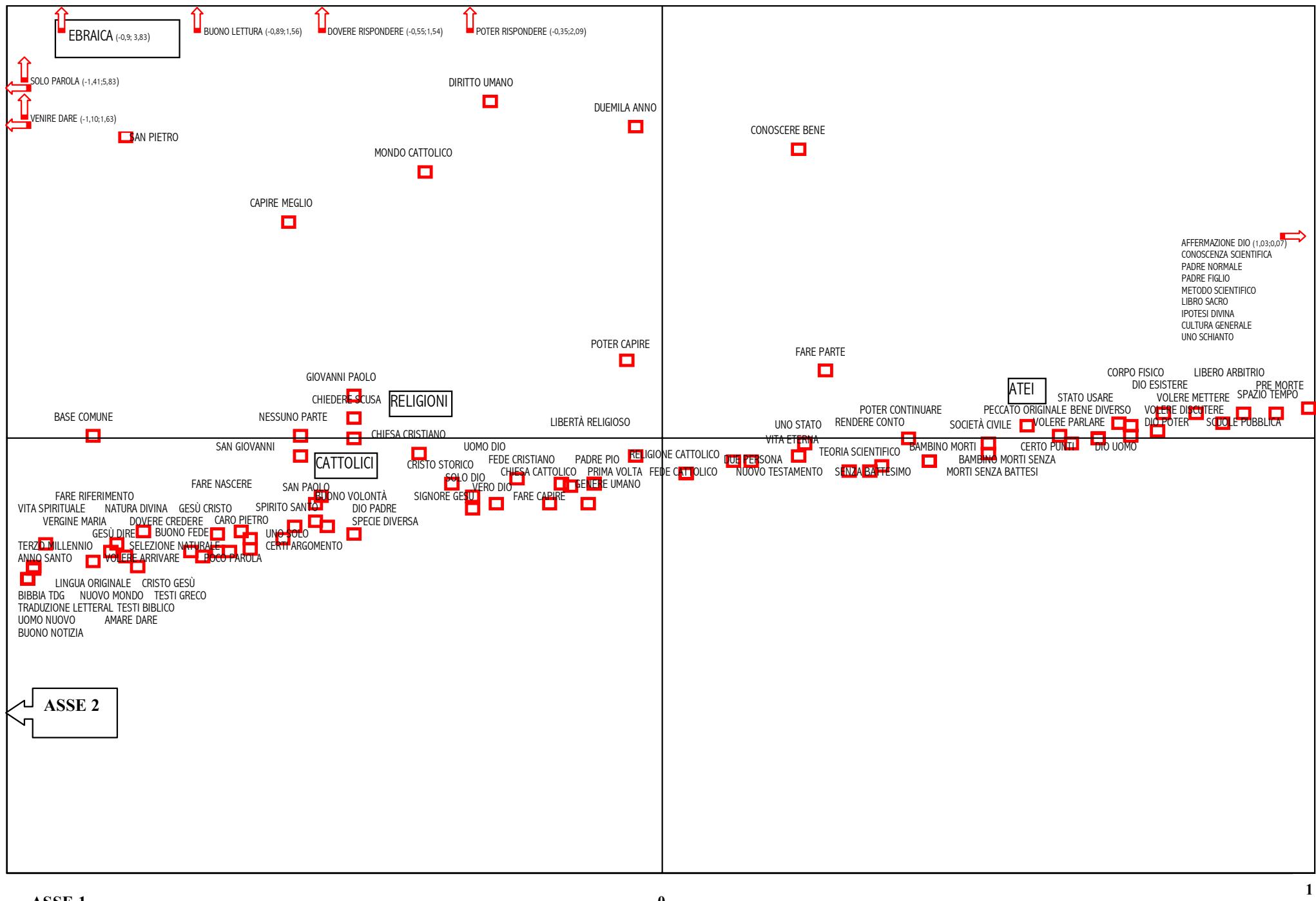
Parola	Massa	Coord.	Parola	Massa	Coord.
EBRAICA	0,02	-1,1	RIUSCIRE	0	-0,2
VOLARE	0	-0,91	RISPOSTA	0	-0,19
ARBITRIO	0	-0,79	RISPETTO	0	-0,19
IMPROBABILE	0	-0,77	STORIA	0	-0,18
NERO	0	-0,76	UMANITÀ	0	-0,18
CINA	0	-0,76	VEDERE	0,01	-0,17
FOTO	0	-0,72	PERSONA	0,01	-0,16
TELEVISIONE	0	-0,72	MORALE	0	-0,16
DEMOCRAZIA	0	-0,72	PAURA	0	-0,15
GIORNALI	0	-0,7	CAPIRE	0	-0,14
SCIENZIATI	0	-0,64	DIMOSTRARE	0	-0,14
TOGLIERE	0	-0,59	CREDO	0	-0,13
MODERAZIONE	0	-0,57	STATO	0,01	-0,12
DEFINIZIONE	0	-0,56	PRESENZA	0	-0,12
REGOLE	0	-0,54	RESPONSABILITÀ	0	-0,1
VECCHIO	0	-0,54	IGNORANZA	0	-0,1
SPAZIO	0	-0,52	CATTOLICO	0,01	-0,09
IMPOSSIBILE	0	-0,51	GRAZIE	0	-0,06
INFINITO	0	-0,48	PRETE	0	-0,06
ASSOCIAZIONE	0	-0,47	POTER	0,01	-0,05
SISTEMA	0	-0,46	PIACERE	0	-0,05
TESTAMENTO	0	-0,46	SERVE	0	-0,04
ESISTENZA	0	-0,45	SCRITTA	0	-0,04
FELICE	0	-0,45	DONNA	0	-0,03
GENESI	0	-0,44	OMOSESSUALI	0	-0,02
MIRACOLO	0	-0,42	<b>ZONA CENTRALE</b>		
ARTE	0	-0,42	FEDE	0,01	0
BATTESIMO	0	-0,4	ROMA	0	0,03
DOGMA	0	-0,39	CRISTIANO	0,01	0,03
COSCIENZA	0	-0,39	ANNO	0,01	0,04
FILOSOFIA	0	-0,38	DIO	0,02	0,05
POTERE	0	-0,35	MILANO	0	0,05
GOVERNO	0	-0,35	INSIEME	0	0,06
TEORIA	0	-0,33	FINE	0	0,07
LIBERO	0	-0,33	ERRORI	0	0,07
SOCIALE	0	-0,33	PANE	0	0,09
BAMBINO	0	-0,32	NUOVO	0	0,1
GUERRA	0	-0,3	CHIESA	0,01	0,13
ATEI	0,42	-0,28	PRIMO	0	0,14
DIALOGO	0	-0,29	NASCERE	0	0,17
PRODOTTO	0	-0,29	TEMA	0	0,17
LIBRO	0	-0,27	PRESTO	0	0,17
GUARDARE	0	-0,25	VIA	0	0,18
VISIONE	0	-0,25	RELIGIONI	0,11	0,21
RAGIONE	0	-0,24	MODERNO	0	0,21
SCRITTI	0	-0,24	LINGUA	0	0,22
CERTO	0	-0,22	PADRE	0	0,23
MESSAGGIO	0	-0,22	TESTI	0	0,23
ATTUALE	0	-0,22	NATURA	0	0,23
GENERE	0	-0,21	CATTOLICI	0,46	0,25
CERTEZZA	0	-0,21	POPOLO	0	0,26
SAPERE	0	-0,2	BIBBIA	0	0,27
GRUPPO	0	-0,2	GERUSALEMME	0	0,27
			VERGINE	0	0,27

			ANIMA	0	0,28
			AMARE	0	0,28
			SIGNIFICATO	0	0,29
			PIETRO	0	0,31
			EVOLUZIONE	0	0,34
			CARNE	0	0,34
			INSEGNAMENTI	0	0,34
			NOME	0	0,35
			VERSIONE	0	0,35
			MADRE	0	0,36
			OPERA	0	0,37
			STORICO	0	0,37
			RIVELAZIONE	0	0,38
			CREATORE	0	0,38
			SECOLO	0	0,39
			ORDINE	0	0,39
			VANGELO	0	0,4
			MADONNA	0	0,4
			CONVERSIONE	0	0,41
			PERDONO	0	0,42
			CIELO	0	0,43
			MOGLIE	0	0,44
			MATTEO	0	0,45
			SPERANZA	0	0,5
			RESURREZIONE	0	0,5
			FESTA	0	0,51
			SALVEZZA	0	0,52
			CAMMINO	0	0,53
			TEMPIO	0	0,53
			PRATICA	0	0,54
			SAN	0	0,55
			SACERDOTE	0	0,55
			RAPPORTO	0	0,55
			CROCE	0	0,57
			DNA	0	0,57
			DOLORE	0	0,6
			TESTIMONI	0	0,61
			MILLENNIO	0	0,62
			LUCA	0	0,63
			CEI	0	0,63
			APOCALISSE	0	0,64
			SANTITÀ	0	0,65
			VESCOVO	0	0,65
			COMUNIONE	0	0,66
			PASTORI	0	0,68
			CULTO	0	0,7
			SABATO	0	0,74
			COPPIA	0	0,74
			DEVOZIONE	0	0,8
			LITURGIA	0	0,84

1 **Grafico 4** Analisi delle corr. lessicali: parole dal basso contributo assoluto, quadrante positivo e specificità



**Grafico 5** Analisi delle corrispondenze lessicali: segmenti caratteristici



**Tabella 13:** Descrizione dell'ASSE 1 Grafico 5. Segmenti ordinati rispetto alla distanza dall'origine

Segmenti	Massa	Coord.	Segmenti	Massa	Coord.
SOLO PAROLA	0,002	-1,41	VERO DIO	0,007	-0,25
VENIRE DARE	0,003	-1,1	UOMO DIO	0,002	-0,22
TESTI BIBLICO	0,002	-0,97	FARE CAPIRE	0,009	-0,17
BIBBIA TDG	0,002	-0,97	FEDE CRISTIANO	0,003	-0,15
LINGUA ORIGINALE	0,003	-0,97	CHIESA CATTOLICO	0,031	-0,14
CRISTO GESÙ	0,003	-0,97	GENERE UMANO	0,003	-0,11
TRADUZIONE LETTERAL	0,003	-0,97	PADRE PIO	0,003	-0,1
BUONO NOTIZIA	0,003	-0,97	PRIMA VOLTA	0,005	-0,1
TESTI GRECO	0,004	-0,97	POTER CAPIRE	0,007	-0,05
AMARE DARE	0,004	-0,97	LIBERTÀ RELIGIOSO	0,006	-0,04
NUOVO MONDO	0,005	-0,97	DUEMILA ANNO	0,01	-0,04
ANNO SANTO	0,003	-0,96	<b>ZONA CENTRALE</b>		
UOMO NUOVO	0,004	-0,96	FEDE CATTOLICO	0,003	0,04
VITA SPIRITUALE	0,003	-0,94	RELIGIONE CATTOLICO	0,003	0,04
EBRAICA	0,011	-0,9	NUOVO TESTAMENTO	0,004	0,11
BUONO LETTURA	0,002	-0,89	DUE PERSONA	0,002	0,14
BASE COMUNE	0,003	-0,87	CONOSCERE BENE	0,003	0,21
TERZO MILLENNIO	0,005	-0,87	VITA ETERNA	0,005	0,21
GESÙ DIRE	0,004	-0,84	UNO STATO	0,012	0,22
NATURA DIVINA	0,004	-0,83	FARE PARTE	0,009	0,25
VERGINE MARIA	0,003	-0,82	SENZA BATTESIMO	0,004	0,29
SAN PIETRO	0,005	-0,82	TEORIA SCIENTIFICO	0,009	0,32
VOLERE ARRIVARE	0,003	-0,8	RENDERE CONTO	0,006	0,34
FARE RIFERIMENTO	0,003	-0,79	POTER CONTINUARE	0,003	0,38
BUONO FEDE	0,009	-0,72	BAMBINO MORTI	0,003	0,41
SELEZIONE NATURALE	0,003	-0,7	MORTI SENZA BATTESI	0,003	0,41
DOVERE CREDERE	0,003	-0,68	BAMBINO MORTI SENZA	0,003	0,5
GESÙ CRISTO	0,019	-0,68	SOCIETÀ CIVILE	0,003	0,5
POCO PAROLA	0,003	-0,66	ATEI	0,485	0,56
FARE NASCERE	0,003	-0,64	PECCATO ORIGINALE	0,027	0,61
CARO PIETRO	0,003	-0,63	CERTO PUNTI	0,004	0,63
CERTI ARGOMENTO	0,003	-0,63	DIO UOMO	0,002	0,67
UNO SOLO	0,006	-0,58	VOLERE PARLARE	0,002	0,67
CAPIRE MEGLIO	0,003	-0,57	BENE DIVERSO	0,002	0,67
SPIRITO SANTO	0,014	-0,56	STATO USARE	0,003	0,7
DOVERE RISONDERE	0,002	-0,55	DIO POTER	0,003	0,72
SAN GIOVANNI	0,004	-0,55	VOLERE DISCUTERE	0,004	0,72
NESSUNO PARTE	0,005	-0,55	VOLERE METTERE	0,003	0,76
BUONO VOLONTÀ	0,004	-0,53	CORPO FISICO	0,003	0,77
SAN PAOLO	0,007	-0,53	DIO ESISTERE	0,012	0,82
CATTOLICI	0,431	-0,52	SCUOLE PUBBLICA	0,003	0,86
DIO PADRE	0,004	-0,51	LIBERO ARBITRIO	0,012	0,89
SPECIE DIVERSA	0,003	-0,47	SPAZIO TEMPO	0,005	0,94
GIOVANNI PAOLO	0,008	-0,47	PRE MORTE	0,02	0,99
CHIEDERE SCUSA	0,009	-0,47	AFFERMAZIONE DIO	0,002	1,03
RELIGIONI	0,073	-0,47	CONOSCENZA SCIENTIF	0,002	1,03
CHIESA CRISTIANO	0,002	-0,37	PADRE NORMALE	0,002	1,03
MONDO CATTOLICO	0,003	-0,36	NESSUNO DIO	0,002	1,03
POTER RISONDERE	0,003	-0,35	NORMALE PADRE	0,003	1,03
CRISTO STORICO	0,003	-0,32	PADRE FIGLIO	0,003	1,03
SOLO DIO	0,003	-0,29	METODO SCIENTIFICO	0,003	1,03
SIGNORE GESÙ	0,003	-0,29	LIBRO SACRO	0,003	1,03
DIRITTO UMANO	0,005	-0,26	IPOTESI DIVINA	0,004	1,03
			CULTURA GENERALE	0,007	1,03
			UNO SCHIANTO	0,014	1,03

**Tabella 14:** Descrizione dell'ASSE 2 Grafico 5. Segmenti ordinati rispetto alla distanza dall'origine

Segmenti	Massa	Coord.	Segmenti	Massa	Coord.
TESTI BIBLICO	0,002	-0,32	LIBERTÀ RELIGIOSO	0,006	-0,04
BIBBIA TDG	0,002	-0,32	VITA ETERNA	0,005	-0,04
LINGUA ORIGINALE	0,003	-0,32	SAN GIOVANNI	0,004	-0,04
CRISTO GESÙ	0,003	-0,32	CHIESA CRISTIANO	0,002	-0,03
TRADUZIONE LETTERAL	0,003	-0,32	BAMBINO MORTI SENZA	0,003	-0,03
BUONO NOTIZIA	0,003	-0,32	UNO STATO	0,012	-0,01
TESTI GRECO	0,004	-0,32	SOCIETÀ CIVILE	0,003	-0,01
AMARE DARE	0,004	-0,32	CERTO PUNTI	0,004	-0,01
NUOVO MONDO	0,005	-0,32	<b>ZONA CENTRALE</b>		
UOMO NUOVO	0,004	-0,3	POTER CONTINUARE	0,003	0
VOLERE ARRIVARE	0,003	-0,29	DIO UOMO	0,002	0
ANNO SANTO	0,003	-0,29	VOLERE PARLARE	0,002	0
TERZO MILLENNIO	0,005	-0,28	BENE DIVERSO	0,002	0
SELEZIONE NATURALE	0,003	-0,27	RELIGIONI	0,073	0
VERGINE MARIA	0,003	-0,27	DIO POTER	0,003	0,01
POCO PAROLA	0,003	-0,26	BASE COMUNE	0,003	0,01
DOVERE CREDERE	0,003	-0,26	PECCATO ORIGINALE	0,027	0,01
GESÙ DIRE	0,004	-0,26	VOLERE METTERE	0,003	0,02
BUONO FEDE	0,009	-0,26	VOLERE DISCUTERE	0,004	0,03
CERTI ARGOMENTO	0,003	-0,25	ATEI	0,485	0,03
VITA SPIRITUALE	0,003	-0,24	STATO USARE	0,003	0,04
NATURA DIVINA	0,004	-0,24	SCUOLE PUBBLICA	0,003	0,04
CARO PIETRO	0,003	-0,23	CHIEDERE SCUSA	0,009	0,05
UNO SOLO	0,006	-0,23	CORPO FISICO	0,003	0,06
SPECIE DIVERSA	0,003	-0,22	SPAZIO TEMPO	0,005	0,06
GESÙ CRISTO	0,019	-0,22	DIO ESISTERE	0,012	0,06
FARE NASCERE	0,003	-0,21	LIBERO ARBITRIO	0,012	0,06
FARE RIFERIMENTO	0,003	-0,21	AFFERMAZIONE DIO	0,002	0,07
DIO PADRE	0,004	-0,2	CONOSCENZA SCIENTIF	0,002	0,07
SPIRITO SANTO	0,014	-0,2	PADRE NORMALE	0,002	0,07
BUONO VOLONTÀ	0,004	-0,19	NESSUNO DIO	0,002	0,07
SIGNORE GESÙ	0,003	-0,16	NORMALE PADRE	0,003	0,07
GENERE UMANO	0,003	-0,15	PADRE FIGLIO	0,003	0,07
FARE CAPIRE	0,009	-0,15	METODO SCIENTIFICO	0,003	0,07
VERO DIO	0,007	-0,15	LIBRO SACRO	0,003	0,07
SAN PAOLO	0,007	-0,15	IPOTESI DIVINA	0,004	0,07
SOLO DIO	0,003	-0,13	CULTURA GENERALE	0,007	0,07
CATTOLICI	0,431	-0,13	UNO SCHIANTO	0,014	0,07
CHIESA CATTOLICO	0,031	-0,11	PRE MORTE	0,02	0,07
PADRE PIO	0,003	-0,1	GIOVANNI PAOLO	0,008	0,1
PRIMA VOLTA	0,005	-0,1	FARE PARTE	0,009	0,16
FEDE CRISTIANO	0,003	-0,1	POTER CAPIRE	0,007	0,18
CRISTO STORICO	0,003	-0,1	CAPIRE MEGLIO	0,003	0,5
UOMO DIO	0,002	-0,09	MONDO CATTOLICO	0,003	0,62
FEDE CATTOLICO	0,003	-0,08	CONOSCERE BENE	0,003	0,67
RELIGIONE CATTOLICO	0,003	-0,08	SAN PIETRO	0,005	0,7
SENZA BATTESIMO	0,004	-0,07	DUEMILA ANNO	0,01	0,72
TEORIA SCIENTIFICO	0,009	-0,07	DIRITTO UMANO	0,005	0,78
RENDERE CONTO	0,006	-0,06	NESSUNO PARTE	0,005	1
NUOVO TESTAMENTO	0,004	-0,05	DOVERE RISONDERE	0,002	1,54
DUE PERSONA	0,002	-0,05	BUONO LETTURA	0,002	1,56
BAMBINO MORTI	0,003	-0,05	VENIRE DARE	0,003	1,63
MORTI SENZA BATTESI	0,003	-0,05	POTER RISONDERE	0,003	2,09
			<b>EBRAICA</b>	0,011	3,83
			SOLO PAROLA	0,002	5,83

### 3.7. Commenti e conclusioni

L'analisi delle corrispondenze semplici permette senza dubbio di cogliere il senso latente dei significati di un insieme ordinato di parole (*sintagma*)<sup>77</sup>.

La disposizione degli elementi lungo gli assi, infatti, consente una chiave di lettura dei contenuti del *corpus* oggetto di studio. <<Sarà possibile trarre il senso latente e globale del sistema di significati elementari messi in gioco dalle forme oggetto di analisi su un fattore, limitandosi all'interpretazione delle sole "parole" significative per l'asse considerato>> [Bolasco, 1999: 231].

Nel caso specifico, l'interpretazione dei grafici permette di cogliere le tematiche toccate da un gruppo di parlanti avventori di un *forum* di discussione.

Se si prende in considerazione il **Grafico 1** a pag.140 si vede che parole come "TNM" (Traduzione Nuovo Mondo Sacre scritture), "TDG" (Testimoni Di Geova), "GEOVA", addensate nel quadrante positivo, testimoniano una concentrazione di argomenti riguardanti le varie traduzioni della Bibbia. Sullo stesso quadrante, lo scienziato "SERMONTI" (teorico dell'evoluzionismo)<sup>78</sup> insieme alle parole "MUTAZIONE" e

---

<sup>77</sup> Si privilegia il termine "sintagma" per sottolineare il concetto che l'ordinamento degli elementi su di un asse, come disposizione di forme testuali, crea significato.

<sup>78</sup> Sermonti Giuseppe, da non confondere con Vittorio (scrittore e traduttore), docente universitario e biologo di fama internazionale, è stato autore di ricerche all'avanguardia nel campo della genetica dei microrganismi, scoprendo la ricombinazione genetica parasessuale in *Penicillium* e in *Streptomyces*. Ha scritto importanti testi scientifici tra cui ricordiamo *Genetics of antibiotics producing microorganisms* (Wiley & Sons) e *Genetica generale* (Boringhieri). È autore anche di numerosi libri e saggi di riflessione critica sulla scienza moderna in generale o su alcuni aspetti particolari. Ricordiamo: *La mela di Adamo e la mela di Newton* (Rusconi, 1974), *Dopo Darwin* (Rusconi, 1980), *Le forme della vita* (Armando, 1981) e il recentissimo *Dimenticare Darwin* (Rusconi, 1999).

"SELEZIONE"<sup>79</sup> polarizzano, invece, l'attenzione sulle teorie Darwiniane dello sviluppo della vita sulla terra.

Volendo analizzare le *specificità*, ovvero le parole caratteristiche dei quattro *Newsgroups*, generate dalla procedura MOCAR (Cfr. output a pag. 96) ed esplicitate nel **Grafico 2** a pag.143 per i *Newsgroup* "CATTOLICI" e "RELIGIONI", si vede come le parole appena descritte siano tipiche del gruppo dei "CATTOLICI". D'altronde la vicinanza tra il gruppo "RELIGIONI" e sue *specifiche* parole come "UOMO", "NASCITA", "MORIRE", "FRUTTO", "AMORE", "ALBERO", "PRINCIPIO" lascerebbero presupporre la trattazione di argomenti legati alla figura biblica dell'albero della vita il cui frutto comunica l'immortalità .

Sul quadrante negativo del **Grafico 1**, invece, si nota come nel *Newsgroup* dedicato agli "ATEI", le tematiche riguardano la "MATEMATICA" la "LOGICA" la "ESPERIENZA" e la "SCIENZA". Parole queste, che si trovano molto vicine al termine "RELIGIONE" che risulta essere particolarmente caratteristico del gruppo degli "ATEI" (Cfr. output della procedura MOCARM a pag.96). Nel gruppo "EBRAICA" (quarto quadrante del **Grafico 1**) si parla, invece, di "TORAH" della festa delle capanne (SUKKOT)<sup>80</sup> nonché di "ISRAELE".

---

<sup>79</sup> Le parole TNM, TDG, GEOVA, MUTAZIONE, SERMONTI sono quelle che rispetto all'ASSE 2 sono più distanti dall'origine (Cfr. Tabella 10 a pag. 142 ).

<sup>80</sup> Da "Sukkah" o "Succot" località ad est del Giordano dove Giacobbe costruì una casa circondata da capanne per il bestiame dalle quali questo luogo prese il nome (Cfr. Bibbia Gen 33, 17; Gdc 8, 4-5. È un termine che indica anche una località dove si accamparono gli Ebrei dopo aver lasciato Ramesse (Cfr. Bibbia Es 12, 37; Nm 33, 5-6).

Durante la festa dei Tabernacoli o delle Capanne (che nell'anno 1999 si festeggia il 25 settembre) gli Ebrei dovevano risiedere in una capanna per una settimana sia per festeggiare la vendemmia, sia per ricordare a se stessi di essere stati un popolo nomade sotto la guida di Dio (Cfr. Bibbia Lv 23, 39-43; Ne 8,14).

Se si osserva il **Grafico 3** a pag. 144 e la relativa discriminazione tra gruppi e specificità del quadrante positivo (**Grafico 4** a pag. 149), si noterà che anche le parole "CEI" (una edizione della Bibbia), "DNA" e "TESTIMONI" oltre ad essere specifiche del gruppo dei "CATTOLICI", confermano le teorie su di una trattazione in esso di argomenti inerenti alla Bibbia dei Testimoni Di Geova e allo sviluppo della vita. Per quanto , invece, riguarda il gruppo degli "ATEI" è interessante notare la vicinanza di termini "FILOSOFIA", "DOGMA", "BATTESIMO", "BAMBINO", "LIBERO" che farebbero pensare a discorsi sul valore dei sacramenti o addirittura sul senso di talune scelte in campo religioso. Sullo stesso quadrante le parole "GIORNALI", "TELEVISIONE", "DEMOCRAZIA" vicine l'una all'altra testimonierebbero, sempre nel gruppo degli "ATEI", alcuni argomenti di attualità come base di alcuni discorsi.

Se si passa al **Grafico 5** a pag.150, dal quale si legge l'ordinamento lungo gli assi fattoriali dei *segmenti* caratteristici, non solo si trovano alcune conferme delle ipotesi appena formulate, ma si riescono ad aggiungere ulteriori elementi di valutazione del fenomeno. Nel quadrante negativo, infatti si trovano *segmenti* quali "BIBBIA TDG", "NUOVO MONDO", "TRADUZIONE LETTERALE", "TESTI GRECO", "SELEZIONE NATURALE" ma anche "CHIEDERE SCUSA" e "GIOVANNI PAOLO" la cui specificità , appartenendo al gruppo "EBRAICA" (Cfr. l'output della procedura MOCAR per i *segmenti* a pag.110), lascerebbe intendere il viaggio del Papa in Israele<sup>81</sup> e le sue intenzioni diplomatiche come temi di alcuni discorsi dei parlanti<sup>82</sup>.

---

<sup>81</sup> Viaggio che si è tenuto nei primi mesi del 2000. Periodo, questo, successivo alla data di rilevazione dei messaggi sui *Newsgroups*.

<sup>82</sup> Una conferma potrebbe anche venire dalla parola "MEMORIA", che oltre ad essere specifica del gruppo "EBRAICA" è posizionata vicino ad esso, nel **Grafico 1**, ai limiti esterni del quarto quadrante.

Nel secondo/quarto quadrante, sempre del **Grafico 5**, i *segmenti* "PECCATO ORIGINALE" e "BAMBINI MORTI SENZA BATTESIMO", addensati attorno al gruppo "ATEI", confermano anch'essi quanto detto in relazione al **Grafico 3**, oltre alla teoria, derivante dalla lettura del **Grafico 1**, che riguarda le tematiche scientifiche di alcuni discorsi.

In conclusione, si può affermare che la metodologia adottata si è rivelata efficace. Essa ha consentito, infatti, di trattare dati testuali molto ampi e articolati in tempi ridotti fornendo - nel contempo - al ricercatore gli elementi necessari al raggiungimento dello scopo proposto: evidenziare gli argomenti e le tematiche trattate da un gruppo di parlanti attraverso gli strumenti messi a disposizione dalla Statistica.

## APPENDICE A

### Esempi di messaggi *USENET*

- **Esempio di messaggio in cui compaiono i *boundary* (MIME multipart):**

```
From: Nathaniel Borenstein <nsb@bellcore.com>
To: Ned Freed <ned@innosoft.com>
Subject: Sample message
MIME-Version: 1.0
Content-type: multipart/mixed; boundary="simple
boundary"
```

This is the preamble. It is to be ignored, though it  
is a handy place for mail composers to include an  
explanatory note to non-MIME conformant readers.  
--simple boundary

This is implicitly typed plain ASCII text.  
It does NOT end with a linebreak.  
--simple boundary  
Content-type: text/plain; charset=us-ascii

This is explicitly typed plain ASCII text.  
It DOES end with a linebreak.

--simple boundary--  
This is the epilogue. It is also to be ignored.

- **Esempio di messaggio MIME multipart/alternative:**

```
From: Nathaniel Borenstein <nsb@bellcore.com>
To: Ned Freed <ned@innosoft.com>
Subject: Formatted text mail
MIME-Version: 1.0
Content-Type: multipart/alternative; boundary=boundary42
```

--boundary42

Content-Type: text/plain; charset=us-ascii

...plain text version of message goes here....  
--boundary42  
Content-Type: text/richtext

.... RFC 1341 richtext version of same message goes here ...  
--boundary42

```
Content-Type: text/x-whatever  
.... fanciest formatted version of same message goes here  
...  
--boundary42--
```

- **Esempio di messaggio MIME multipart/digest:**

```
From: Moderator-Address  
To: Recipient-List  
MIME-Version: 1.0  
Subject: Internet Digest, volume 42  
Content-Type: multipart/digest;  
boundary="---- next message ----"  
----- next message ----  
  
From: someone-else  
Subject: my opinion  
  
...body goes here ...  
----- next message ----  
  
From: someone-else-again  
Subject: my different opinion  
  
... another body goes here...  
----- next message -----
```

- **Esempio di messaggio MIME message/partial:**

For example, if an audio message is broken into two parts, the first part might look something like this:

```
X-Weird-Header-1: Foo  
From: Bill@host.com  
To: joe@otherhost.com  
Subject: Audio mail  
Message-ID: <id1@host.com>  
MIME-Version: 1.0  
Content-type: message/partial;  
    id="ABC@host.com";  
    number=1; total=2  
  
X-Weird-Header-1: Bar  
X-Weird-Header-2: Hello
```

```
Message-ID: <anotherid@foo.com>  
MIME-Version: 1.0  
Content-type: audio/basic  
Content-transfer-encoding: base64
```

```
... first half of encoded audio data goes here...
```

and the second half might look something like this:

```
From: Bill@host.com
To: joe@otherhost.com
Subject: Audio mail
MIME-Version: 1.0
Message-ID: <id2@host.com>
Content-type: message/partial;
    id="ABC@host.com"; number=2; total=2
```

```
... second half of encoded audio data goes here...
```

Then, when the fragmented message is reassembled, the resulting message to be displayed to the user should look something like this:

```
X-Weird-Header-1: Foo
From: Bill@host.com
To: joe@otherhost.com
Subject: Audio mail
Message-ID: <anotherid@foo.com>
MIME-Version: 1.0
Content-type: audio/basic
Content-transfer-encoding: base64
```

```
... first half of encoded audio data goes here...
```

```
... second half of encoded audio data goes here...
```

- **Esempio di messaggio MIME message/external-body:**

```
From: Whomever
To: Someone
Subject: whatever
MIME-Version: 1.0
Message-ID: <id1@host.com>
Content-Type: multipart/alternative; boundary=42
Content-ID: <id001@guppylake.bellcore.com>

--42
Content-Type: message/external-body;
    name="BodyFormats.ps";
    site="thumper.bellcore.com";
    access-type=ANON-FTP;
    directory="pub";
    mode="image";
    expiration="Fri, 14 Jun 1991 19:13:14 -0400 (EDT)"

Content-type: application/postscript
Content-ID: <id42@guppylake.bellcore.com>

--42
Content-Type: message/external-body;
```

```

name="/u/nsb/writing/rfcs/RFC-MIME.ps";
site="thumper.bellcore.com";
access-type=AFS
expiration="Fri, 14 Jun 1991 19:13:14 -0400 (EDT)"

Content-type: application/postscript
Content-ID: <id42@guppylake.bellcore.com>

--42
Content-Type: message/external-body;
access-type=mail-server
server="listserv@bogus.bitnet";
expiration="Fri, 14 Jun 1991 19:13:14 -0400 (EDT)"

Content-type: application/postscript
Content-ID: <id42@guppylake.bellcore.com>

get RFC-MIME.DOC

--42--

```

- **Esempio di messaggio MIME-PEM:**

```

-----BEGIN PRIVACY-ENHANCED MESSAGE-----
Proc-Type: 4,ENCRYPTED
Content-Domain: RFC822
DEK-Info: DES-CBC,F8143EDE5960C597
Originator-ID-Symmetric: linn@zendia.enet.dec.com, ,
Recipient-ID-Symmetric: linn@zendia.enet.dec.com,ptf-kmc,3
Key-Info: DES-ECB,RSA-MD2,9FD3AAD2F2691B9A,
           B70665BB9BF7CBCDA60195DB94F727D3
Recipient-ID-Symmetric: pem-dev@tis.com,ptf-kmc,4
Key-Info: DES-ECB,RSA-MD2,161A3F75DC82EF26,
           E2EF532C65CBCFF79F83A2658132DB47

LlrHB0eJzyhP+/fSStdW8okeEnv47jxe7SJ/iN72ohNcUk2jHEUSoH1nvNSIWL9M8tE
jmF/zxB+bATMtPjCUWbz8Lr9wloXIkjHUlBLpvXR0UrUzYbkNpk0agV2IzUpkJ6UiRR
GcDSvzrsoK+oNvqu6z7Xs5Xfz5rDqUcMlK1Z6720dcBWGGsDLpTpSCnpotdXd/H5LMD
WnonNvPCwQUHt==
-----END PRIVACY-ENHANCED MESSAGE-----

```

- **Esempio di messaggio criptato PGP:**

```

From: Michael Elkins <elkins@aero.org>
To: Michael Elkins <elkins@aero.org>
Mime-Version: 1.0
Content-Type: multipart/encrypted; boundary=foo;
protocol="application/pgp-encrypted"

--foo
Content-Type: application/pgp-encrypted

```

```

Version: 1

--foo
Content-Type: application/octet-stream

-----BEGIN PGP MESSAGE-----
Version: 2.6.2

hIwDY32hYGCE8MkBA/wOu7d45aUxF4Q0RKJprD3v5Z9K1YcRJ2fve87lM1Dlx4Oj
eW4GDdBfLbJE7VUpp13N19GL8e/AqbyyjHH4aS0YoTk10QQ9nnRvjY8nZL3MPXSZ
g9VGQxFeGqzykzmykU6A26MSMexR4ApeeON6xzZWfo+0yOqAq6lb46wsvldZ96YA
AABH78hyX7YX4uT1tNCWEIIBoqqvCeIMpp7UQ2IzBrXg6GtukS8NxbukLeamqVW3
1yt21DYOjuLzcMNe/JNsD9vDVCvOOG3OCi8=
=zzaA
-----END PGP MESSAGE-----

--foo--

```

- **Esempio di messaggio firmato PGP**

```

From: Michael Elkins <elkins@aero.org>
To: Michael Elkins <elkins@aero.org>
Mime-Version: 1.0
Content-Type: multipart/signed; boundary=bar; micalg=pgp-md5;
protocol="application/pgp-signature"

--bar
& Content-Type: text/plain; charset=iso-8859-1
& Content-Transfer-Encoding: quoted-printable
&
& =A1Hola!
&
& Did you know that talking to yourself is a sign of senility?
&
& It's generally a good idea to encode lines that begin with
& From=20because some mail transport agents will insert a greater-
& than (>) sign, thus invalidating the signature.
&
& Also, in some cases it might be desirable to encode any =20
& trailing whitespace that occurs on lines in order to ensure =20
& that the message signature is not invalidated when passing =20
& a gateway that modifies such whitespace (like BITNET). =20
&
& me

--bar
Content-Type: application/pgp-signature

-----BEGIN PGP MESSAGE-----
Version: 2.6.2

```

```
iQCVAwUBMJrRF2N9oWBghPDJAQE9UQQAtl7LuRVndBjrk4EqYB1b3h5QXIX/LC//jJV5bNv
kZIGPIcEmI5iFd9boEgvpirHtIREEqLQRkYNoBActFBZmh9GC3C041WGquMbrbxc+nIs1TIK1A
08rVi9ig/2Yh7LFrK5Ein57U/W72vgSxLhe/zhdfolT9BrnHOxEa44b+EI=
=ndaj
-----END PGP MESSAGE-----
```

--bar--

- **Esempio di messaggio firmato e poi criptato PGP**

```
Content-Type: multipart/encrypted;
  protocol="application/pgp-encrypted"; boundary=foo

--foo
Content-Type: application/pgp-encrypted

Version: 1

--foo
Content-Type: application/octet-stream

-----BEGIN PGP MESSAGE-----
& Content-Type: multipart/signed; micalg=pgp-md5
&   protocol="application/pgp-signature"; boundary=bar
&
& --bar
& Content-Type: text/plain; charset=us-ascii
&
& This message was first signed, and then encrypted.
&
& --bar
& Content-Type: application/pgp-signature
&
& -----BEGIN PGP MESSAGE-----
& Version: 2.6.2
&
& iQCVAwUBMJrRF2N9oWBghPDJAQE9UQQAtl7LuRVndBjrk4EqYB1b3h5QXIX/LC//jJV5bNv
kZIGPIcEmI5iFd9boEgvpirHtIREEqLQRkYNoBActFBZmh9GC3C041WGq
& uMbrbxc+nIs1TIK1A08rVi9ig/2Yh7LFrK5Ein57U/W72vgSxLhe/zhdfolT9Brn
& HOxEa44b+EI=
& =ndaj
& -----END PGP MESSAGE-----
&
& --bar--
-----END PGP MESSAGE-----
```

## APPENDICE B

### Tabella codice

- **Tabella caratteri ISO-8859-1**

Name: iso-8859-1  
Description: Latin1: Western Europe  
Version: 1  
Charset: ISO-8859-1, ISO\_8859-1, iso-ir-100, latin1, 11, IBM819, CP819, csISOLatin1

```
0x80 0x120AC #EURO SIGN
#0x81      #UNDEFINED
0x82 0x1201A #SINGLE LOW-9 QUOTATION MARK
0x83 0x10192 #LATIN SMALL LETTER F WITH HOOK
0x84 0x1201E #DOUBLE LOW-9 QUOTATION MARK
0x85 0x12026 #HORIZONTAL ELLIPSIS
0x86 0x12020 #DAGGER
0x87 0x12021 #DOUBLE DAGGER
0x88 0x102C6 #MODIFIER LETTER CIRCUMFLEX ACCENT
0x89 0x12030 #PER MILLE SIGN
0x8A 0x10160 #LATIN CAPITAL LETTER S WITH CARON
0x8B 0x12039 #SINGLE LEFT-POINTING ANGLE QUOTATION MARK
0x8C 0x10152 #LATIN CAPITAL LIGATURE OE
#0x8D      #UNDEFINED
0x8E 0x1017D #LATIN CAPITAL LETTER Z WITH CARON
#0x8F      #UNDEFINED
#0x90      #UNDEFINED
0x91 0x12018 #LEFT SINGLE QUOTATION MARK
0x92 0x12019 #RIGHT SINGLE QUOTATION MARK
0x93 0x1201C #LEFT DOUBLE QUOTATION MARK
0x94 0x1201D #RIGHT DOUBLE QUOTATION MARK
0x95 0x12022 #BULLET
0x96 0x12013 #EN DASH
0x97 0x12014 #EM DASH
0x98 0x102DC #SMALL TILDE
0x99 0x12122 #TRADE MARK SIGN
0x9A 0x10161 #LATIN SMALL LETTER S WITH CARON
0x9B 0x1203A #SINGLE RIGHT-POINTING ANGLE QUOTATION MARK
0x9C 0x10153 #LATIN SMALL LIGATURE OE
#0x9D      #UNDEFINED
0x9E 0x1017E #LATIN SMALL LETTER Z WITH CARON
0x9F 0x10178 #LATIN CAPITAL LETTER Y WITH DIAERESIS

0xA0 0x00A0 # NO-BREAK SPACE
0xA1 0x00A1 # INVERTED EXCLAMATION MARK
0xA2 0x00A2 # CENT SIGN
```

0xA3 0x00A3 # POUND SIGN  
0xA4 0x00A4 # CURRENCY SIGN  
0xA5 0x00A5 # YEN SIGN  
0xA6 0x00A6 # BROKEN BAR  
0xA7 0x00A7 # SECTION SIGN  
0xA8 0x00A8 # DIAERESIS  
0xA9 0x00A9 # COPYRIGHT SIGN  
0xAA 0x00AA # FEMININE ORDINAL INDICATOR  
0xAB 0x00AB # LEFT-POINTING DOUBLE ANGLE QUOTATION MARK  
0xAC 0x00AC # NOT SIGN  
0xAD 0x00AD # SOFT HYPHEN  
0xAE 0x00AE # REGISTERED SIGN  
0xAF 0x00AF # MACRON  
0xB0 0x00B0 # DEGREE SIGN  
0xB1 0x00B1 # PLUS-MINUS SIGN  
0xB2 0x00B2 # SUPERSCRIPT TWO  
0xB3 0x00B3 # SUPERSCRIPT THREE  
0xB4 0x00B4 # ACUTE ACCENT  
0xB5 0x00B5 # MICRO SIGN  
0xB6 0x00B6 # PILCROW SIGN  
0xB7 0x00B7 # MIDDLE DOT  
0xB8 0x00B8 # CEDILLA  
0xB9 0x00B9 # SUPERSCRIPT ONE  
0xBA 0x00BA # MASCULINE ORDINAL INDICATOR  
0xBB 0x00BB # RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK  
0xBC 0x00BC # VULGAR FRACTION ONE QUARTER  
0xBD 0x00BD # VULGAR FRACTION ONE HALF  
0xBE 0x00BE # VULGAR FRACTION THREE QUARTERS  
0xBF 0x00BF # INVERTED QUESTION MARK  
0xC0 0x00C0 # LATIN CAPITAL LETTER A WITH GRAVE  
0xC1 0x00C1 # LATIN CAPITAL LETTER A WITH ACUTE  
0xC2 0x00C2 # LATIN CAPITAL LETTER A WITH CIRCUMFLEX  
0xC3 0x00C3 # LATIN CAPITAL LETTER A WITH TILDE  
0xC4 0x00C4 # LATIN CAPITAL LETTER A WITH DIAERESIS  
0xC5 0x00C5 # LATIN CAPITAL LETTER A WITH RING ABOVE  
0xC6 0x00C6 # LATIN CAPITAL LETTER AE  
0xC7 0x00C7 # LATIN CAPITAL LETTER C WITH CEDILLA  
0xC8 0x00C8 # LATIN CAPITAL LETTER E WITH GRAVE  
0xC9 0x00C9 # LATIN CAPITAL LETTER E WITH ACUTE  
0xCA 0x00CA # LATIN CAPITAL LETTER E WITH CIRCUMFLEX  
0xCB 0x00CB # LATIN CAPITAL LETTER E WITH DIAERESIS  
0xCC 0x00CC # LATIN CAPITAL LETTER I WITH GRAVE  
0xCD 0x00CD # LATIN CAPITAL LETTER I WITH ACUTE  
0xCE 0x00CE # LATIN CAPITAL LETTER I WITH CIRCUMFLEX  
0xCF 0x00CF # LATIN CAPITAL LETTER I WITH DIAERESIS  
0xD0 0x00D0 # LATIN CAPITAL LETTER ETH (Icelandic)  
0xD1 0x00D1 # LATIN CAPITAL LETTER N WITH TILDE  
0xD2 0x00D2 # LATIN CAPITAL LETTER O WITH GRAVE  
0xD3 0x00D3 # LATIN CAPITAL LETTER O WITH ACUTE  
0xD4 0x00D4 # LATIN CAPITAL LETTER O WITH CIRCUMFLEX  
0xD5 0x00D5 # LATIN CAPITAL LETTER O WITH TILDE  
0xD6 0x00D6 # LATIN CAPITAL LETTER O WITH DIAERESIS  
0xD7 0x00D7 # MULTIPLICATION SIGN  
0xD8 0x00D8 # LATIN CAPITAL LETTER O WITH STROKE  
0xD9 0x00D9 # LATIN CAPITAL LETTER U WITH GRAVE  
0xDA 0x00DA # LATIN CAPITAL LETTER U WITH ACUTE

0xDB 0x00DB # LATIN CAPITAL LETTER U WITH CIRCUMFLEX  
0xDC 0x00DC # LATIN CAPITAL LETTER U WITH DIAERESIS  
0xDD 0x00DD # LATIN CAPITAL LETTER Y WITH ACUTE  
0xDE 0x00DE # LATIN CAPITAL LETTER THORN (Icelandic)  
0xDF 0x00DF # LATIN SMALL LETTER SHARP S (German)  
0xE0 0x00E0 # LATIN SMALL LETTER A WITH GRAVE  
0xE1 0x00E1 # LATIN SMALL LETTER A WITH ACUTE  
0xE2 0x00E2 # LATIN SMALL LETTER A WITH CIRCUMFLEX  
0xE3 0x00E3 # LATIN SMALL LETTER A WITH TILDE  
0xE4 0x00E4 # LATIN SMALL LETTER A WITH DIAERESIS  
0xE5 0x00E5 # LATIN SMALL LETTER A WITH RING ABOVE  
0xE6 0x00E6 # LATIN SMALL LETTER AE  
0xE7 0x00E7 # LATIN SMALL LETTER C WITH CEDILLA  
0xE8 0x00E8 # LATIN SMALL LETTER E WITH GRAVE  
0xE9 0x00E9 # LATIN SMALL LETTER E WITH ACUTE  
0xEA 0x00EA # LATIN SMALL LETTER E WITH CIRCUMFLEX  
0xEB 0x00EB # LATIN SMALL LETTER E WITH DIAERESIS  
0xEC 0x00EC # LATIN SMALL LETTER I WITH GRAVE  
0xED 0x00ED # LATIN SMALL LETTER I WITH ACUTE  
0xEE 0x00EE # LATIN SMALL LETTER I WITH CIRCUMFLEX  
0xEF 0x00EF # LATIN SMALL LETTER I WITH DIAERESIS  
0xF0 0x00F0 # LATIN SMALL LETTER ETH (Icelandic)  
0xF1 0x00F1 # LATIN SMALL LETTER N WITH TILDE  
0xF2 0x00F2 # LATIN SMALL LETTER O WITH GRAVE  
0xF3 0x00F3 # LATIN SMALL LETTER O WITH ACUTE  
0xF4 0x00F4 # LATIN SMALL LETTER O WITH CIRCUMFLEX  
0xF5 0x00F5 # LATIN SMALL LETTER O WITH TILDE  
0xF6 0x00F6 # LATIN SMALL LETTER O WITH DIAERESIS  
0xF7 0x00F7 # DIVISION SIGN  
0xF8 0x00F8 # LATIN SMALL LETTER O WITH STROKE  
0xF9 0x00F9 # LATIN SMALL LETTER U WITH GRAVE  
0xFA 0x00FA # LATIN SMALL LETTER U WITH ACUTE  
0xFB 0x00FB # LATIN SMALL LETTER U WITH CIRCUMFLEX  
0xFC 0x00FC # LATIN SMALL LETTER U WITH DIAERESIS  
0xFD 0x00FD # LATIN SMALL LETTER Y WITH ACUTE  
0xFE 0x00FE # LATIN SMALL LETTER THORN (Icelandic)  
0xFF 0x00FF # LATIN SMALL LETTER Y WITH DIAERESIS

- **Tabella codice ANSI Windows® Codice pagina 850 Courier font**

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
00:	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
10:	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
20:	!	"	#	\$	%	&	'	(	)	*	+	,	-	.	/	
30:	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
40:	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	0
50:	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	-
60:	~	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
70:	p	q	r	s	t	u	v	w	x	y	z	{	}	~	DEL	
80:	€	□	„	ƒ	„	…	†	‡	^	„	„	<	€	□	□	□
90:	□	„	„	„	„	„	„	„	„	„	„	„	„	„	„	„
A0:	¡	¢	£	¤	¥	¦	§	¨	®	™	¤	«	»	¬	¬	¬
B0:	¤	±	²	³	‘	µ	¶	·	¸	¹	º	»	¤	¤	¤	¤
C0:	À	Á	Â	Ã	Ä	Ë	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï	Ï
D0:	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	Þ
E0:	à	á	â	ã	ä	ë	ç	è	é	ê	ë	ì	í	î	ï	ï
F0:	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

**Nota:** i caratteri Ansi Windows® sono richiamabili direttamente da tastiera digitando Alt + il numero decimale corrispondente al carattere, secondo la tabella di cui sopra, anteponendo al numero lo zero. I numeri devono essere digitati dal tastierino numerico. Vedere più avanti per una tabella di conversione esadecimale-decimale.

- **Tabella codice Microsoft Dos® Codice pagina 850 Courier font**

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
00:	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
10:	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
20:	!	"	#	\$	%	&	'	(	)	*	+	,	-	.	/	
30:	0	1	2	3	4	5	6	7	8	9	:	:	<	=	>	?
40:	Ø	À	Á	Â	Ã	Ð	È	É	Í	Ó	Ó	Ó	Ñ	À	Á	Ó
50:	Þ	Ù	É	Ã	Ã	Ã	Ã	Ã	Ã	Ã	Ã	Ã	Ã	Ã	Ã	Ã
60:	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘
70:	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘
80:	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘
90:	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘
A0:	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘
B0:	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘
C0:	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘
D0:	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘
E0:	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘
F0:	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘

**Nota:** i caratteri Microsoft Dos® sono richiamabili direttamente da tastiera digitando Alt + il numero decimale corrispondente al carattere, secondo la tabella di cui sopra. I numeri devono essere digitati dal tastierino numerico. Vedere più avanti per una tabella di conversione esadecimale-decimale.

- **Tabella codice IBM ASCII estesa**

0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
Ø	Ø	Ø	Ø	Ø	Ø	Ø	Ø	Ø	Ø	Ø	Ø	Ø	Ø	Ø	Ø
1	►	◀	↕	↔	↕	↔	↑	↓	→	←	↶	↷	▲	▼	?
2	!	"	#	\$	%	&	,	<	>	*	+	,	-	.	/
3	Ø	À	Á	Â	Ã	Ð	È	É	Í	Ó	Ó	Ó	Ñ	À	Ó
4	Þ	Ù	É	Ã	Ã	Ã	Ã	Ã	Ã	Ã	Ã	Ã	Ã	Ã	Ã
5	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘
6	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘
7	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘
8	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘
9	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘
A	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘
B	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘
C	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘
D	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘
E	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘
F	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘	‘

- **Tabella di conversione esadecimale decimale**

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F.
0	000	001	002	003	004	005	006	007	008	009	010	011	012	013	014	015
1	016	017	018	019	020	021	022	023	024	025	026	027	028	029	030	031
2	032	033	034	035	036	037	038	039	040	041	042	043	044	045	046	047
3	048	049	050	051	052	053	054	055	056	057	058	059	060	061	062	063
4	064	065	066	067	068	069	070	071	072	073	074	075	076	077	078	079
5	080	081	082	083	084	085	086	087	088	089	090	091	092	093	094	095
6	096	097	098	099	100	101	102	103	104	105	106	107	108	109	110	111
7	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127
8	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
9	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
A	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
B	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
C	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207
D	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
E	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
F	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255

## APPENDICE C

### Preparazione del testo e strutturazione del corpus

<< Seppure va riconosciuto che è sempre più facile trascrivere o trasportare su computer un testo rispetto ad alcuni anni fa, tuttavia è opportuno prendere qualche prudente attenzione prima di concludere che un testo così "acquisito" sia già "pronto per l'uso". Ciò soprattutto quando il testo non deve essere utilizzato solo per essere letto (cioè come fosse una fotocopia) o riascoltato (come fosse un messaggio), bensì per essere trattato come un'informazione per un'analisi automatica del suo contenuto (ovvero come un dato statistico).

1 - Il testo può essere acquisito su un supporto elettronico:

- su dischetti da 3,5 pollici (doppia densità 720K, alta densità 1,44M, o superdischi ultima generazione da 120MB), su hard *disk*, su *CD* o su cartucce rimovibili (Zip, Jaz, Syquest, Bernoulli, Ditto, Dat ecc.)
- in righe (records) preferibilmente di lunghezza inferiore agli 80 caratteri spazi inclusi (in qualche caso anche fino a 250); è il caso di osservare che la visualizzazione di un testo sul monitor non corrisponde alla lunghezza di memorizzazione dei records fisici su disco<sup>83</sup>.
- in modalità Alto Basso (ossia in minuscolo con eventuali maiuscole), per riconoscere più facilmente i nomi propri, le sigle le accentazioni
- senza sillabazione (ossia senza spezzare alcuna parola con trattini) -senza tabulatori o altri comandi speciali per carattere e/o impaginazione
- senza bloccare ogni riga, quindi con il blocco del record alla fine del paragrafo. In questa prima fase, il testo deve essere registrato in formato testo-ANSI per Windows (Winword) o in formato testo\_ASCII per **MS\_DOS**, o in modalità "solo testo" (nel caso del Macintosh). *Soltanto al termine delle eventuali correzioni e lessicalizzazioni [...], il testo dovrà essere invece salvato con riga bloccata (ovvero con un fine paragrafo alla fine di ciascuna riga, nella situazione in cui questa è visualizzata sullo schermo al momento del salvataggio)*<sup>84</sup>.

Qui di seguito seguono alcuni facsimile di testo, nel caso dell'uso del pacchetto SpadT:

#### ***A. Modello di testo di interviste non direttive o di testi letterari, giornalistici, ecc.***

---

<sup>83</sup> Ad esempio la lunghezza di un paragrafo spesso supera la riga visualizzata a schermo e comprende anche alcune decine di righe, in tal caso un paragrafo equivale ad un solo records. Per questo, a seconda dei programmi utilizzati (ad es. SPAD), occorre salvare il testo in maniera <solo testo con interruzione di linea> in modo da far coincidere la riga visualizzata alla lunghezza del record fisico memorizzato.

<sup>84</sup> Un controllo del buon esito del salvataggio può essere fatto attraverso il comando "TYPE namefile" in DOS: solo in tal caso la visualizzazione su schermo coincide con il reale bloccaggio delle righe (ossia con la lunghezza fisica dei records).

[configurato per il formato di SpadT: TYPE= 1]

\*\*\*\* TESTO n. 1

**Qui comincia il testo della prima intervista che sarà trascritto senza interventi dei traduttori, ma con la sola accortezza di non trascrivere rumori, commenti o altra informazione che non siano le parole pronunciate dall'intervistato.**

**Nel caso di parole (connettivi e simili) ripetute ma senza senso queste si possono omettere; al contrario nei casi di parole di senso compiuto - ma non capite dal traduttore per cattiva registrazione - si può trascrivere convenzionalmente l'acronimo "PMPI" (Parola Mancante Per Incomprensione), una volta per ogni parola che manca.**

**Ad ogni pausa o stacco logico di argomentazione nel discorso si può andare a capo battendo un <INVIO>.**

---

**I trattini che precedono segnano la fine di una fase dell'intervista. I trattini vanno messi solo nel caso in cui l'intervistatore ponga una domanda specifica, che sarà eventualmente anch'essa trascritta nel modo seguente:**

**PPPP**

**Quale aspetto ritiene.....fine domanda**

---

**Inizia qui la risposta dell'intervistato. bla bla bla ... bla bla ...alla fine della intervista n. 1 far seguire di seguito la successiva nel modo seguente**

\*\*\*\*TESTO n.2

**segue come sopra con le stesse regole sullo stesso file.**

Trattandosi di un colloquio, le varie domande dell'intervistatore non sono necessariamente nello stesso ordine in tutti i testi: quindi il loro riconoscimento serve solo a distinguere le frasi tra loro. Un modo più generale di trascrivere le interviste non direttive, nelle quali le domande dell'intervistatore possono essere con cadenza "casuale", è quello di considerare ogni frase (sia dell'intervistato sia dell'intervistatore) come un frammento (enunciato) sul quale applicare marcature o categorizzazioni: in tal caso si seguiranno le istruzioni date nel punto B che segue. Nel caso di trascrizione di testi letterari o di articoli di giornale, ogni testo viene considerato per intero come un solo enunciato.

**B. Modello di testo delle risposte a domande aperte di un'intervista strutturata con questionario, o di qualunque altra situazione in cui interessa una segmentazione in 'frasi' (frammenti).**

[configurato secondo l'istruzione SpadT: TYPE=2]

Trascrivere tutte le risposte in un solo file, così strutturato, usando le stesse convenzioni di cui sopra, con l'accortezza di dare ad ogni intervistato (unità testuale) la stessa rigorosa numerazione del suo questionario (scheda) d'indagine.

----0001

**Testo della risposta alla prima domanda bla...**

**...bla bla.**

**++++**

**Testo della risposta alla seconda domanda bla... Ala bla bla**

**++++**

**Testo della risposta alla ultima domanda bla..**

**bla bla bla bla bla.**

**----0002**

**segue come sopra con le stesse regole sullo stesso file. al termine dell'ultima risposta dell'ultimo individuo il file deve chiudersi con quattro segni di uguale come segue**

**=====**

In questo modello di testo di tipo B, esisterà un file separato di dati numerici (categoriali e/o quantitativi) relativo alle informazioni raccolte con il questionario, che è possibile associare al file di testo costituente il corpus di studio per l'analisi testuale. Tale associazione avviene mediante l'abbinamento dei codici di individuazione dei frammenti. è indispensabile che i due files abbiano i codici nello stesso ordine.

**Esempio di file numerico associato:**

0001	2	25	1	4
0002	1	19	2	1
0003	3	23	2	

**...ecc.**

Ogni colonna di dati corrisponderà ad una caratteristica utile da incrociare con l'analisi testuale (categorie di analisi di contenuto) da considerarsi come un attributo dell'unità di test.

Una diversa maniera di marcare un testo, nelle sue differenti categorizzazioni possibili, è ad esempio quello richiesto dal programma Lexico, che inserisce i valori categoriali, direttamente nel file testuale. In tal caso, all'inizio di ogni parte del corpus (testo o frammento) si utilizza una chiave di lettura (ad es. ch1) diversa per ogni variabile di categorizzazione. Tale chiave viene scritta fra due segni rispettivamente minore-maggiore <ch1=x>, ponendo dopo la parola chiave il valore assunto dalla variabile per quel frammento. Ad esempio

**<ch1 =2> <s1 =4>**

**Testo del primo frammento bla...**

**...bla bla bla bla bla.**

**<ch1 =1 <s1=2>**

**Testo del secondo frammento segue come sopra**

**... bla bla bla bla bla**

**<ch 1 =2> <s 1 =3> Testo del terzo frammento bla...**

**... bla bla bla bla bla.**

**<ch1 =1 > <s1 =1 >**

**Testo del quarto frammento ... bla bla bla bla bla.**

sta a significare che il primo subtesto appartiene alla categoria 2 per la variabile ch1, e alla categoria 4 per la variabile s1; il secondo frammento è associato invece rispettivamente alla categoria 1 e alla 2; il terzo frammento è classificato analogamente alla 2 e alla 3; il quarto alla

1 in entrambe le variabili. Si noti che in Lexico il testo deve essere espresso in records di lunghezza inferiore a 80 caratteri, mentre non richiede alcun segno di fine file.

Altri software come *Alceste*, *Hyperbase*, *DBT*, *Intex* presentano altre caratteristiche nel formato del testo, per cui è opportuno leggere le relative istruzioni prima dell'uso.>> [Bolasco, dispensa a.a. 1997-1998]

## APPENDICE D

## Alcuni criteri per la correzione degli errori e per la normalizzazione del testo

<< Al fine di una buona preparazione dei testo si deve procedere, *nell'ordine qui descritto, alle* seguenti operazioni. Non rispettando quest'ordine, si possono determinare nuovi "errori" nel testo.

2- Correggere gli errori formali: eventualmente con un correttore ortografico, ma solo se si tratta di linguaggio comune: poiché poiché, può puó puo' ed altri con i ï û ü ecc.

3- Ricercare eventuali date abbreviate (es. '90) per riportarle alla forma intera (1990); attenzione ai numeri separati da slash come 3/4 che possono essere ricodificati automaticamente in date (3 aprile o 4 marzo) da alcuni software, contro la nostra volontà.

4- Gli accenti possono essere trasformati in apostrofo (però diventa *pero'*) solo quando si volesse analizzare il testo tutto in Maiuscolo<sup>85</sup>. L'uso del testo in tutto maiuscolo deve essere ben valutato in quanto in genere fa perdere alcune informazioni utili, come la distinzione fra nomi e parole comuni: ad esempio fra cognomi e colori (il signor Rossi e l'aggettivo comune *rossi*). Se non vi sono ragioni specifiche, è preferibile lasciare il testo in modalità <Alto Basso>, con tutti gli accenti e fare attenzione alle operazioni intermedie di salvataggio del file.

5- Correggere l'apostrofo facendolo seguire da un <blank> (spazio) in maniera da evitare le confusioni fra apostrofi veri e accenti (esempio "dall'altra" diventa "dall' altra", "pero'sono" diventa "pero' sono").

6- Correggere e verificare ulteriori eventuali casi di accenti che, se trascurati, renderebbero omografe due parole non omofone [è il caso di *pèsca* (Frutto) per distinguerlo da *pésca* (atto del pescare), un altro caso è quello di "princípi" da distinguere da *principi*], o sdoppierebbero una stessa parola [affinché e affinchè, sé e sè, ecc.]. Più in particolare ciò può accadere anche per parole straniere o per parole dialettali, in cui gli accenti a volte sono situati all'interno della parola. La sostituzione dell'accento con l'apostrofo in questi casi è impossibile, poiché ad esempio *principi* diventando "princi'pi" sarebbe successivamente spezzato in "princi' pi" a causa

<sup>85</sup> Il passaggio al maiuscolo, infatti, può creare caratteri anomali sulle lettere accentate, per quanto detto al primo punto dei paragrafo.

della scelta di un blank dopo l'apostrofo e quindi trasformato erroneamente in due parole. Un'altra possibilità consiste nel riportare a lettera semplice la vocale eventualmente accentata e nell'aggiungere alla forma grafica un <\_1>, ovvero "pèsc"= "pesca\_1", per distinguerla dalla seconda accezione. Questo punto non è sempre necessario, grazie al forte effetto di contesto che rende altamente improbabili certe coesistenze nello stesso corpus.

7- Eliminare ogni <doppio spazio> bianco (due blanks consecutivi), sostituendolo con un solo spazio, al fine di non inficiare le successive ricerche di parole o di frasi, in modalità di testo integrale.

La maggior parte di queste operazioni possono essere svolte in maniera automatica, attraverso un programma, in via di perfezionamento, denominato TALTAC (Bolasco, Morrone, Baiocchi 1997).

#### *[...] La formazione di acronimi composti, poliformi e altri step di normalizzazione del testo*

8- Riconoscere gli acronimi (PIL = prodotto interno lordo; CE = Comunità Europea) e i nomi (di luoghi, persone, o istituzioni) nonché alcune parole/espressioni fortemente ricorrenti (con alta pertinenza tematica: TLC = telecomunicazioni, PMI = piccole e medie imprese). Se si analizzerà il testo in modalità Alto Basso è preferibile scrivere gli acronimi in tutto maiuscolo. Altrimenti è possibile aggiungere un carattere (ad esempio <\_!>) agli acronimi e ai nomi propri il cui significante diventa omografo di parole comuni (verdi\_!, bianco\_!, nato\_!, ce\_!, ecc.): solo in tal modo è possibile distinguerli dai corrispondenti verdi, bianco, nato, ce.

9- Da uno screening preliminare sui "segmenti ripetuti" (da effettuarsi senza alcun tipo di separatore) si selezionano le frasi fisse più frequenti da identificare e da rimodificare con slash "/" (esempio: in/modo) o underscore "\_" (in\_particolare); nel fare queste lessicalizzazioni occorre ricostruire le sequenze composte dalla più lunga alla più corta [ad esempio: 1) non solo, 2) non so, 3) non lo so, 4) lo so]. Se queste correzioni fossero fatte direttamente in Word, fare attenzione a selezionare l'opzione "parola intera", per evitare sgradevoli fusioni improprie (es. "alla/finestra" nel corso della correzione di "alla/fine").

In particolare, è importante lessicalizzare le espressioni "polirematiche", solo se esse esistono in forma di frase fissa e non possono trovarsi cioè nel testo in forma troncata o implicita: come è il caso di bilancio/dello/stato che può trovarsi citato anche solo come il Bilancio.

10- Riportare tutti i prefissi (extra, socio ecc.) adiacenti alla parola cui sono associati, creando sempre quindi una sola parola senza trattino (extraparlamentare, socioeconomico); segue una lista di prefissi e *prefissoidi* più frequenti (non tutti in realtà in italiano lasciano dubbi di scrittura, costituendo in molti casi già una sola parola): A, AGRI, AGRO, ANGLO, ANTI, ARABO, AUTO, BI, BIO, CO, COE, COM, CON, CONTRA/O, DE, DECA, DI, ECONOMICO, ELETTRO, EURO, EX, EXTRA, FRANCO, GEO, GIURIDICO, IN, ITALO, LATINO, MACRO, MEDIO, MEGA, META, METAL, MICRO, MINI, NEO, NORD, OLTRE, ORTO, PAN, PARA, POLI, POLITICO, POST, PRE, PSEUDO, PUBBLICO, QUASI, RADIO, RE, REGIO, RI, SAN, SANT, SEMI, SOCIAL, SOCIO, SOPRA, SOTTO, SOVIETICO, SOVRA, STORICO, SU, SUB, SUD, SUPER, SUR, TECNICO, TELE, TERMO, TERZO, TRI, ULTRA, VICE, VIDEO.

Fra questi, esistono prefissi che scontano la presenza di intere parole nella formazione di composti: per la sola iniziale C, da un buon dizionario, si possono ricavare ad esempio: CACCIA, CACO, CALCI/IO/O, CALLI/O, CALZA, CAMBIA, CANTA, CAPO/I, CARBO, CARDIO, CARICA, CARO, CARPO, CARTAIO, CASEI, CASSA, CATA, CATTO, CADA, CAVA, CECO, CEFALO, CENTI/O, CENTRO/I, CEREAL, CEREBRO, CHEMIO, CHERO, CHILO, CHIRO, CIANO, CICLO, CINE, CINO, CIRCUM, CIS, CITO, CLAU, CLAV, CLIMA, CLINO, CLORO, COLEI/O, COPRI/O, CORTO, COSMO, CRANIO, CRE, CRINO, CRIOS, CRIPTO, CRISO, CRISTALLO, CRITTO, CROMO, CRONO, CRUCI..

11- Eliminare tutti i trattini "-" presenti nel testo, dal momento che la lineetta viene dal computer interpretata molto spesso come separatore per spezzare una riga in due o per sillabare una parola.

12- Nella trascrizione dei nomi, seguire questi esempi: DeMartino, DeMartino/G, DeMartino/on, DeMartino/on/G e non adottare soluzioni del tipo: G. De Martino, G/DeMartino, on/DeMartino/G, on/G/DeMartino, ecc. che inficiano il riconoscimento dei nome o l'ordine alfabetico.

13- Correggere, per le disambiguazioni di tipo grammaticale o di tipo semantico, in automatico prima tutte le concordanze più numerose, facendo attenzione ad usare l'opzione "parola intera", *senza distinzione di M/minuscole* per catturare ad esempio politiche e Politiche, o sia capo dello Stato che capo dello stato, o forze sociali e Forze Armate ecc.

14- Eliminare in modo automatico la maiuscola dopo il punto, per evitare lo sdoppiamento improprio di una stessa parola in due forme grafiche diverse: Quando e quando, La e la, Allora e allora<sup>86</sup> ecc. In tale ambito, esiste il problema più generale di riconoscere in quali situazioni il punto non identifica una fine frase: ad es. <il sig. Rossi> o <1.256 lire>, o <art. 3 della Costituzione>. Per fare questo, esistono diversi approcci: vedi il grafo che individua una grammatica locale per la costruzione della frase in INTEX (Silberstein 1998) o alcune procedure di normalizzazione proposte da Labb   (1992).

15- L'uso sistematico ed esclusivo, nei processi di lessicalizzazione, di alcuni caratteri come lo <\_> o lo </> consente di tornare al testo originario senza difficoltà. E' sufficiente infatti considerare questi due caratteri come dei separatori e non come "caratteri dell'alfabeto del corpus", per spezzare i segmenti ripetuti o le polirematiche o per riunificare le varie accezioni di un termine (stato\_S1 come "istituzione", stato\_S2 come "status, condizione" stato\_V1 come participio passato del verbo "essere" o stato\_V2 del verbo "stare").

#### *[...] Scelta dei separatori per le operazioni di categorizzazione delle parole*

Al fine di poter selezionare il tipo di lettura del corpus che di volta in volta si desidera (forme grafiche, lemmi, poliformi, forme testuali),    utile immaginare una gerarchia fra i separatori da utilizzare, ci   consente, ad esempio, di ritornare alle forme grafiche dopo una correzione fisica del testo.

Una proposta, dal pi   debole al pi   forte dei separatori, pu   essere la seguente:  
usare

---

<sup>86</sup> La sola situazione in cui questa correzione non    opportuna    quella nella quale si voglia studiare le differenti forme di inizio frase.

- la sottolineatura <\_> per esprimere sia le disambiguazioni grammaticali \_S (sost) \_V (verbi) \_G (aggettivi) \_B (avverbi) \_P (preposizioni) \_C (congiunzioni) \_N (pronomi); che le disambiguazioni semantiche apponendo un numero per ciascuna accezione \_1 \_2 o abbinandolo alla categoria grammaticale \_S1 \_S2
- lo slash </> per l'identificazione di poliformi o altri costrutti lessico-grammaticali (in/modo, disegno/di/legge, andare/al/creatore)
- il <+> per comporre dei sotto-articoli dei dizionario (art+14 della Costituzione) l'asterisco <\*> per separare numeri da loro attributi o misure (2\*mila, 2\*anni)
- altri separatori specifici di riserva da definire di volta in volta (ad esempio <\$>, o <&>, o <^>)

Secondo quanto già detto al punto 2.1, è bene evitare di scegliere come separatori altri simboli (~ £) che non sono invarianti al variare di tastiere, di sistemi operativi o di applicazioni (ad esempio i segni minore (<) o maggiore (>) sono interdetti nel *software Lexico*, perché utilizzati come delimitatori delle variabili di categorizzazione dei subtesti, come illustrato precedentemente al punto 1).

Ai fini delle analisi statistiche, i caratteri aggiuntivi qui proposti possono essere virtualmente eliminati, considerandoli come un carattere separatore (blank): in tal caso si torna al testo iniziale senza dover effettuare ulteriori correzioni "a ritroso".

#### *[...] Registrazione del testo per l'analisi statistica e ulteriori scelte in corso d'elaborazione*

Solo a questo punto, è opportuno registrare il file del corpus in modo <testo con righe bloccate> ("solo testo con layout", o "testo con LF", o "testo con interruzione di linea" o "MS-DOS con layout": quest'ultimo è raccomandato per l'uso di Spad\_T). Per dettagli si rimanda anche il punto 1.1. E' appena il caso di sottolineare che salvataggi incoerenti nel passaggio fra diversi sistemi operativi rischiano di "sporcare" irrimediabilmente il testo: occorre prestare molta attenzione. *Per un approfondimento vedi il file -Icww.doc*

#### *[...] Scelta di simboli per fusioni o equivalenze*

Al termine delle reali correzioni "fisiche" del testo, possono essere predisposte delle correzioni "virtuali", che agiscono cioè sul testo solo attraverso un programma specifico (ad esempio con Spad\_T o Sphinx), in funzione di scelte dell'utilizzatore nel corso di un'elaborazione.

A tal fine, ad esempio, si possono predisporre via via delle equivalenze fra parole (cfr. la nozione *di forma testuale* nei paragrafi precedenti) per ricodificare gruppi di significati d'interesse (famiglie di parole, campi semanticici ecc.). Sarà opportuno scegliere dei simboli per distinguere tali termini dalle altre forme grafiche non raggruppate. E' preferibile apporre tali simboli<sup>87</sup> dietro il significante della forma, per salvaguardare sempre l'ordine alfabetico nelle liste di forme, ad es.:

---

<sup>87</sup> Si noti che tali simboli sono inseriti solo come "Istruzioni di programma" e non sul testo.

- <#> per le flessioni di uno stesso lemma aggettivo o sostantivo
- <&> per le flessioni dei verbi (al fine di poterli riselectonare tutti direttamente, ad esempio attraverso un'istruzione generica di 'find')
- <@> per le forme omologhe (sinonimi)
- <\$> per le riduzioni lessematiche (raggruppamenti di una stessa radice): coer-enza, coer-ente; consapevol-e, consapevol-mente; razionalizz-azione, razionalizz-are, razionalizz-ato (in *Alceste* per le riduzioni viene utilizzato il <+>).

### *[...] Sulle parole vuote*

Spesso si lavora con una lista di parole che s'intendono considerare come vuote cioè prive d'interesse e quindi da abbandonare (ossia porre in delete ai fini dell'analisi statistica). Tale lista è totalmente soggettiva, e funzionale alle necessità dell'analisi; non esistono criteri generali per definirle: ad esempio è appena il caso di sottolineare che le parole grammaticali (preposizioni, congiunzioni, articoli) sono molto importanti per interpretare un testo (si pensi per alcune di loro alla funzione di "determinanti"). Pertanto non può essere questo un criterio valido comunque per definirle come vuote alcune forme del corpus.

Parimenti potrebbero essere considerate vuote anche parole piene (come sostantivi, aggettivi o verbi) posizionate ai primissimi posti nella graduatoria per numero decrescente di occorrenze del vocabolario del corpus (fascia delle alte frequenze). Al contrario, un'opportuna disambiguazione di tali forme (a partire dall'analisi dei segmenti e da varie analisi di concordanze o di contesti locali) mette spesso in evidenza le diverse accezioni in cui queste vengono utilizzate. In particolare, la disambiguazione dei verbi comuni può far emergere gli usi idiomati (phrasal verbs) di più alta pertinenza per il corpus considerato.>> [Bolasco, dispensa a.a. 1997-1998].

## APPENDICE E

### Le espressioni regolari del Perl

<< In questo contesto parlando di *espressioni regolari* intenderemo una espressione costruita secondo una sintassi ben precisa che ci permette di descrivere uno schema di stringa.

Senza entrare in ulteriori dettagli cerchiamo di capire questo concetto mediante un esempio elementare. Supponiamo di voler "descrivere" tutte le stringhe composte secondo il seguente schema: un numero con un qualsiasi numero di cifre, un carattere di spaziatura o tabulazione ed una parola composta da caratteri qualsiasi. Con una espressione regolare diventa banale descrivere questa "ricetta" di composizione di una stringa:

`\d+\s.+`

I primi due caratteri ``\d'' indicano la presenza di un carattere numerico (0, 1, 2, ..., 9); il carattere ``+'' che segue una certa sequenza indica che il carattere rappresentato può essere ripetuto una o più volte. La sequenza ``\s'' indica un qualsiasi carattere di spaziatura o di tabulazione. Infine il punto ``.'' indica un carattere qualsiasi e il simbolo ``+'' finale sta ad indicare che questo carattere può essere ripetuto una o più volte.

Questa espressione regolare descriverà quindi stringhe del tipo ``1234 pippo'', ``1 ab\\$%\&xy'', ``1 2''. Le seguenti stringhe invece non risultano "descritte" dalla precedente espressione regolare: ``a b'', ``pippo'', ``albero casa'', ``1+3=4''.

La seguente tabella descrive sinteticamente i termini che possono comporre una espressione regolare:

Tab. 11: Termini per la composizione di una espressione regolare

.	qualsiasi carattere escluso il new line (``\n'')
[a-z0-9]	qualsiasi carattere di questo insieme
[^a-z0-9]	qualsiasi carattere esclusi quelli di questo insieme
\d	una cifra qualsiasi; equivalente a ``[0-9]''
\D	un carattere che non sia una cifra; equivalente a ``[^0-9]''
\w	un carattere alfanumerico; equivalente a ``[a-zA-Z0-9]''
\W	un carattere non alfanumerico; equivalente a ``[^a-zA-Z0-9]''
\s	un carattere di spaziatura (spazio, tabulazione, new line, ecc.)
\S	un carattere non di spaziatura
\n	il carattere <i>new line</i>

\r	il carattere return (ritorno carrello)
\t	il carattere di tabulazione
\f	form feed, carattere di avanzamento di pagina
\b	backspace, cancellazione di un carattere a sinistra
\0	null, il carattere nullo
\	il carattere `` ''
\\"	il carattere ``\''
\*	il carattere ``*''
<i>x</i> ?	il carattere <i>x</i> ripetuto 0 o 1 volta
<i>x</i> *	il carattere <i>x</i> ripetuto 0 o più volte
<i>x</i> +	il carattere <i>x</i> ripetuto una o più volte
pippo	la stringa ``pippo''
aa bb cc	la stringa ``aa'' oppure la stringa ``bb'' oppure la stringa ``cc''
^	la stringa inizia con l'espressione regolare seguente
\$	la stringa termina con l'espressione regolare precedente

>> [Liverani, 1996].

## APPENDICE F

### I file di parametri delle procedure dello Spad-T®

- **Procedura NUMER: numerizzazione del testo, conteggio parole**

- **Soglia di freq.= 0**
- **Lunghezza parole=20**
- **Memoria=1200000**
- **Parametri/Output: Religio.pat, Religio.lst**

```
NXLPA=60 LISTP=1 LISTF=1 LRESM=OUI NAPEL=0 MEM=1200000
:=====Lecture des textes
NTEXZ='RELIGI~1.TXT' NTEXM='RELIGIO.TXM'

PROC ARTEX
=====
ITYP=1 LIREP=1

PROC SELOX
=====
NUMQ=1 LDONA=NON

NSPA='RELIGIO.SPA'

PROC NUMER
=====
:Num,risation des textes
NSEU=0 NMOMI=0 LEDIT=TOT NXLET =20 NXSIG =150 NXMAX=47000 COEF =1
HISTO 90 680 1530 2620 4300 5120 6600 6600 5900 5000 3500 2300 >
1300 690 410 180 130 60 27 80
FAIBLE '"$ | ><_ - [] *~{ }#+%/
FORT .;:(!)!?,,
FIN

STOP
```

- **Procedura CORTE: correzione del vocabolario**

- **Eliminazione di 347 parole vuote**
- **Equivalenze**
- **Memoria=3000000**

- **Parametri/Output: Corte.pat, Corte.lst**

```
LISTP=OUI,LISTF=OUI,NXLPA=59,LRESM=NON,NAPEL=0,MEM=3000000
NSPA = 'RELIGIO.SPA'
NSPC = 'RELIGIO.SPC'

PROC CORTE
:=====
LELET=1
DELET 000
DELET ^^SPAWN^^
DELET ABBASTANZA
DELET ABBIA
DELET ABBIAMO
DELET ABBIANO
DELET ADDIRITTURA
DELET ADESSO
DELET AFFATTO
DELET AFFINCHÈ
DELET AGLI
DELET ALCUN
DELET ALCUNA
DELET ALCUNE
DELET ALCUNI
DELET ALESSANDRO
DELET ALEX
DELET ALL
DELET ALLA
DELET ALLE
DELET ALLO
DELET ALLORA
DELET ALMENO
DELET ALTRA
DELET ALTRE
DELET ALTRETTANTO
DELET ALTRI
DELET ALTRIMENTI
DELET ALTRO
DELET ALTROVE
DELET ALTRUI
DELET ANCH
DELET ANCHE
DELET ANCORA
DELET AND
DELET ANDREA
DELET ANTONIO
DELET ANZI
DELET APPENA
DELET APPUNTO
DELET ARE
DELET ASSOLUTAMENTE
DELET ATTRAVERSO
DELET AVANTI
DELET AVENDO
DELET AVER
DELET AVERE
```

DELET AVESSE  
DELET AVETE  
DELET AVEVA  
DELET AVEVANO  
DELET AVEVO  
DELET AVRAI  
DELET AVREBBE  
DELET AVREI  
DELET AVRÀ  
DELET AVUTO  
DELET BASTA  
DELET BEH  
DELET BENSÌ  
DELET BISOGNA  
DELET BLA  
DELET BODINI  
DELET CAN  
DELET CERTAMENTE  
DELET CHE  
DELET CHI  
DELET CHIARAMENTE  
DELET CHISSÀ  
DELET CHIUNQUE  
DELET CHRISTOS  
DELET CIAO  
DELET CIASCUNO  
DELET CIOÈ  
DELET CIRCA  
DELET CIÒ  
DELET CLAUDIO  
DELET CMQ  
DELET COL  
DELET COLORO  
DELET COLUI  
DELET COM  
DELET COME  
DELET COMPLETAMENTE  
DELET COMUNQUE  
DELET CON  
DELET COS  
DELET COSA  
DELET COSE  
DELET COSÌ  
DELET CUI  
DELET CUT  
DELET DAGLI  
DELET DAI  
DELET DAL  
DELET DALL  
DELET DALLA  
DELET DALLE  
DELET DALLO  
DELET DANIELE  
DELET DAVANTI  
DELET DAVVERO  
DELET DEGLI

DELET DEI  
DELET DEL  
DELET DELL  
DELET DELLA  
DELET DELLE  
DELET DELLO  
DELET DENTRO  
DELET DIETRO  
DELET DIRETTAMENTE  
DELET DON  
DELET DOPO  
DELET DOVE  
DELET DUNQUE  
DELET DURANTE  
DELET ECC  
DELET ECCO  
DELET EMAIL  
DELET EPPURE  
DELET ERA  
DELET ERANO  
DELET ERO  
DELET ESATTAMENTE  
DELET ESSA  
DELET ESSE  
DELET ESSENDOD  
DELET ESSERE  
DELET ESSERI  
DELET ESSI  
DELET ESSO  
DELET ETC  
DELET FIN  
DELET FINCHÈ  
DELET FINO  
DELET FINORA  
DELET FOR  
DELET FOSSE  
DELET FOSZERO  
DELET FRA  
DELET FRAKZ  
DELET FROM  
DELET FUORI  
DELET FURONO  
DELET GERACI  
DELET GIANNI  
DELET GIÀ  
DELET GLI  
DELET HAI  
DELET HANNO  
DELET ICC  
DELET INFATTI  
DELET INFINE  
DELET INOLTRE  
DELET INSOMMA  
DELET INTORNO  
DELET INVECE  
DELET LAS

DELET LEO  
DELET LEV  
DELET LIRIOFLOR  
DELET LIST  
DELET LORO  
DELET LOS  
DELET LUI  
DELET MAH  
DELET MAI  
DELET MAIL  
DELET MASSIMO  
DELET MAURO  
DELET MENTRE  
DELET MIA  
DELET MICA  
DELET MIE  
DELET MIEI  
DELET MIO  
DELET MOLTE  
DELET MOLTI  
DELET MOLTO  
DELET MSG  
DELET NATURALMENTE  
DELET NEANCHE  
DELET NECESSARIAMENTE  
DELET NEGLI  
DELET NEI  
DELET NEL  
DELET NELL  
DELET NELLA  
DELET NELLE  
DELET NELLO  
DELET NEMMENO  
DELET NEPPURE  
DELET NEWS  
DELET NEWSGROUP  
DELET NIENTE  
DELET NOI  
DELET NON  
DELET NONOSTANTE  
DELET NOSTRA  
DELET NOSTRE  
DELET NOSTRI  
DELET NOSTRO  
DELET NOT  
DELET OBJECTZONE  
DELET OGNI  
DELET OGNUNO  
DELET ONE  
DELET OPPURE  
DELET ORMAI  
DELET OUR  
DELET OVVERO  
DELET OVVIAMENTE  
DELET OVVIO  
DELET PARTICOLARMENTE

DELET PER  
DELET PERCHÈ  
DELET PERCIÒ  
DELET PERFETTAMENTE  
DELET PERFINO  
DELET PERÒ  
DELET PIUTTOSTO  
DELET PIÙ  
DELET POI  
DELET POICHIÈ  
DELET POR  
DELET POST  
DELET PRECEDENTE  
DELET PRESSO  
DELET PROBABILMENTE  
DELET PROPRI  
DELET PROPRIA  
DELET PROPRIE  
DELET PROPRIO  
DELET PUR  
DELET PURE  
DELET PURTROPPO  
DELET PUÒ  
DELET QUA  
DELET QUAL  
DELET QUALCHE  
DELET QUALCOSA  
DELET QUALCUNO  
DELET QUALE  
DELET QUALI  
DELET QUALSIASI  
DELET QUALUNQUE  
DELET QUANDO  
DELET QUANTE  
DELET QUANTI  
DELET QUANTO  
DELET QUASI  
DELET QUE  
DELET QUEGLI  
DELET QUEI  
DELET QUEL  
DELET QUELL  
DELET QUELLA  
DELET QUELLE  
DELET QUELLI  
DELET QUELLO  
DELET QUEST  
DELET QUESTA  
DELET QUESTE  
DELET QUESTI  
DELET QUESTO  
DELET QUI  
DELET QUINDI  
DELET RAFFAELE  
DELET SARAI  
DELET SARANNO

DELET SAREBBE  
DELET SAREBBERO  
DELET SAREI  
DELET SARÀ  
DELET SEBBENE  
DELET SEI  
DELET SEMPLICEMENTE  
DELET SERGIO  
DELET SIA  
DELET SIAMO  
DELET SIANO  
DELET SICCOME  
DELET SICURAMENTE  
DELET SIETE  
DELET SOLAMENTE  
DELET SOLTANTO  
DELET SON  
DELET SONO  
DELET SOPRATTUTTO  
DELET SPECIALMENTE  
DELET SPESO  
DELET STA  
DELET STAI  
DELET STANNO  
DELET STARE  
DELET STESSO  
DELET STESSA  
DELET STESSE  
DELET STESSI  
DELET STIAMO  
DELET STO  
DELET STATA  
DELET STATE  
DELET STEFANO  
DELET STEVE  
DELET SUA  
DELET SUBITO  
DELET SUE  
DELET SUGLI  
DELET SUI  
DELET SUL  
DELET SULL  
DELET SULLA  
DELET SULLE  
DELET SUO  
DELET SUOI  
DELET TAL  
DELET TALE  
DELET TALI  
DELET THAT  
DELET THE  
DELET THIS  
DELET TOTALMENTE  
DELET TRA  
DELET TUA  
DELET TUE

DELET TUO  
DELET TUOI  
DELET TUTT  
DELET TUTTA  
DELET TUTTAVIA  
DELET TUTTE  
DELET TUTTI  
DELET TUTTO  
DELET UNA  
DELET VERAMENTE  
DELET VEROLINI  
DELET VITTORIO  
DELET VOI  
DELET VOSTRA  
DELET VOSTRE  
DELET VOSTRI  
DELET VOSTRO  
DELET WAS  
DELET WITH  
DELET YOU  
EQUIV AFFERMARE AFFERMA  
EQUIV AFFERMAZIONE AFFERMAZIONI  
EQUIV AMARE AMA  
EQUIV AMICO AMICI  
EQUIV ANGELO ANGELI  
EQUIV ANNO ANNI  
EQUIV APOSTOLO APOSTOLI  
EQUIV ARGOMENTO ARGOMENTI ARGOMENTAZIONI  
EQUIV ASPETTO ASPETTI  
EQUIV ATEO ATEI  
EQUIV AUTORE AUTORI  
EQUIV AZIONE AZIONI  
EQUIV BAMBINO BAMBINI  
EQUIV BELLO BELLA BEL  
EQUIV BENE BEN BENISSIMO  
EQUIV BUONO BUONI BUONA BUON  
EQUIV CAMBIARE CAMBIA  
EQUIV CAPIRE CAPISCI CAPISCO CAPITO  
EQUIV CASO CASI  
EQUIV CATTOLICO CATTOLICI CATTOLICA  
EQUIV CERCARE CERCA CERCO CERCATO  
EQUIV CHIAMA CHIAMATI CHIAMATO  
EQUIV CHIEDERE CHIEDE CHIEDO CHIESTO  
EQUIV CHIESA CHIESE  
EQUIV CIELO CIELI  
EQUIV CONCETTO CONCETTI  
EQUIV CONCLUSIONE CONCLUSIONI  
EQUIV CONDIZIONE CONDIZIONI  
EQUIV CONFRONTO CONFRONTI  
EQUIV CONOSCERE CONOSCI CONOSCE CONOSCIUTO CONOSCO  
EQUIV CONSEGUENZA CONSEGUENZE  
EQUIV CONSIDERARE CONSIDERATO  
EQUIV CONTINUARE CONTINUA  
EQUIV CONVINZIONE CONVINZIONI  
EQUIV CREDENTE CREDENTI  
EQUIV CREDERE CREDE CREDERÀ CREDONO CREDI

EQUIV CRISTIANO CRISTIANI CRISTIANA CRISTIANE  
EQUIV DARE DANNO DATA DATE DATI DATO  
EQUIV DIMOSTRARE DIMOSTRA  
EQUIV DIRE DICE DICENDO DIMMI DICEVA DICCI DICIAMO DICO DICONO >  
DIREI DISSE  
EQUIV DIFFERENZA DIFFERENZE  
EQUIV DIRITTO DIRITTI  
EQUIV DISCORSO DISCORSI  
EQUIV DISCUSSIONE DISCUSSIONI  
EQUIV DIVERSA DIVERSE  
EQUIV DIVERSO DIVERSI  
EQUIV DIVINA DIVINO  
EQUIV DOGMA DOGMI  
EQUIV DOMANDA DOMANDE  
EQUIV DONNA DONNE  
EQUIV DOVERE DEVE DEVI DEVO DOVEVA DOVREBBE DOVREBBERO DOVREMBO >  
DOVRESTI DOVUTO DEVONO DOBBIAMO DOVREI  
EQUIV DUBBIO DUBBI  
EQUIV EBRAICO EBRAICA  
EQUIV EFFETTO EFFETTI  
EQUIV ERRORI ERRORE  
EQUIV ESEMPIO ESEMPI  
EQUIV ESISTERE ESISTA ESISTE ESISTONO  
EQUIV ESPERIENZA ESPERIENZE  
EQUIV EVENTO EVENTI  
EQUIV FARE FACCIA FACCIAMO FACCIO FACENDO FAI FAR FARLO FARSI >  
FARTI FATE FATTA FATTE FATTI FATTO FECE FANNO  
EQUIV FENOMENO FENOMENI  
EQUIV FIGLIO FIGLI  
EQUIV FORMA FORME  
EQUIV FORZA FORZE  
EQUIV FRATELLO FRATELLI  
EQUIV GIORNO GIORNI  
EQUIV GRANDE GRAN GRANDI  
EQUIV GRUPPO GRUPPI  
EQUIV GUARDARE GUARDA  
EQUIV IDEA IDEE  
EQUIV IMMAGINE IMMAGINI  
EQUIV IMPORTANTE IMPORTANTI  
EQUIV INFORMAZIONE INFORMAZIONI  
EQUIV INTENDERE INTENDE INTENDO  
EQUIV ITALIANA ITALIANO  
EQUIV LASCIARE LASCIA LASCIATO LASCIO  
EQUIV LIBERO LIBERA  
EQUIV LIBRO LIBRI  
EQUIV LUNGO LUNGA  
EQUIV LUOGO LUOGHI  
EQUIV MAGGIORE MAGGIOR  
EQUIV MANO MANI  
EQUIV MESSAGGIO MESSAGGI  
EQUIV METODO METODI  
EQUIV METTERE METTE MESSO  
EQUIV MIRACOLO MIRACOLI  
EQUIV MODERNO MODERNA  
EQUIV MODO MODI  
EQUIV MOTIVO MOTIVI

EQUIV MUTAZIONE MUTAZIONI  
EQUIV NASCERE NASCE NATO  
EQUIV NECESSARIO NECESSARIA  
EQUIV NESSUNO NESSUNA NESSUN  
EQUIV NUOVO NUOVE NUOVI NUOVA  
EQUIV OPERA OPERE  
EQUIV OPINIONE OPINIONI  
EQUIV PAESE PAESI  
EQUIV PARLARE PARLA PARLANDO PARLANO PARLATO PARLI PARLO  
EQUIV PAROLA PAROLE  
EQUIV PARTICOLARE PARTICOLARI  
EQUIV PARTIRE PARTITO  
EQUIV PASSARE PASSATO  
EQUIV PECCATO PECCATI  
EQUIV PENSARE PENSA PENSI PENSO  
EQUIV PERSONA PERSONE  
EQUIV PERSONALE PERSONALI  
EQUIV PIACERE PIACE  
EQUIV PICCOLO PICCOLA PICCOLI  
EQUIV POCO POCHE POCHI  
EQUIV POSIZIONE POSIZIONI  
EQUIV POTER POSSA POSSANO POSSIAMO POSSO POSSONO POTEVA POTREBBE >  
POTREBBERO POTREI POTRÀ POTUTO PUOI  
EQUIV PRENDERE PRENDE  
EQUIV PRESENTE PRESENTI  
EQUIV PRETE PRETI  
EQUIV PRIMO PRIMI  
EQUIV PROBLEMA PROBLEMI  
EQUIV PROCESSI PROCESSO  
EQUIV PROVA PROVE  
EQUIV PUNTI PUNTO  
EQUIV QUESTIONE QUESTIONI  
EQUIV RELIGIONE RELIGIONI  
EQUIV RELIGIOSO RELIGIOSA RELIGIOSE RELIGIOSI  
EQUIV RENDERE RENDE RESO RESTA RESTO  
EQUIV RIUSCIRE RIESCE RIESCO  
EQUIV RIGUARDA RIGUARDO  
EQUIV RISONDERE RISPONDE RISPONDO  
EQUIV RISPOSTA RISPOSTE  
EQUIV RITENERE RITENGO RITIENE  
EQUIV SACERDOTE SACERDOTI  
EQUIV SALUTO SALUTI  
EQUIV SANTO SANTI SANTA  
EQUIV SAPERE SAI SAPPIAMO SANNO  
EQUIV SCELTA SCELTE  
EQUIV SCIENTIFICO SCIENTIFICHE SCIENTIFICA  
EQUIV SCRITTURA SCRITTURE  
EQUIV SCRIVERE SCRIVE  
EQUIV SCUOLE SCUOLA  
EQUIV SECOLO SECOLI  
EQUIV SECONDO SECONDA  
EQUIV SEMBRARE SEMBRA SEMBRANO  
EQUIV SEMPLICE SEMPLICI  
EQUIV SENTIRE SENTI SENTO SENTE SENTITO  
EQUIV SIMILE SIMILI  
EQUIV SOCIALE SOCIALI

```

EQUIV SOLO SOLA SOLE
EQUIV SPIEGARE SPIEGA SPIEGATO
EQUIV SPIEGAZIONE SPIEGAZIONI
EQUIV SPIRITO SPIRITI
EQUIV STORICO STORICA
EQUIV TANTO TANTI TANTE
EQUIV TEORIA TEORIE
EQUIV TESTI TESTO
EQUIV TROVARE TROVA TROVATO TROVI TROVO
EQUIV ULTIMO ULTIMA ULTIMI
EQUIV UMANO UMANI UMANA
EQUIV UNICO UNICA
EQUIV UOMO UOMINI
EQUIV USARE USATO
EQUIV VANGELO VANGELI
EQUIV VARI VARIE
EQUIV VEDERE VEDE VEDI VEDIAMO VEDO VEDRAI
EQUIV VENIRE VAI VADO VANNO VENUTO VERRÀ VIENE VENGA VENGONO
EQUIV VIVERE VIVE VIVO VIVONO
EQUIV VOLERE VUOLE VOLEVA VOLEVO VOLUTO VORREI VUOI VUOL >
    VOGLIAMO VOGLIO VOGLIONO
FIN
STOP

```

- **Procedura SETEX1: riduzione di soglia e riduzione lunghezza parole in analisi**
  - **Soglia di frequenza=50**
  - **Lunghezza parole > 2**
  - **Memoria=3000000**
  - **Elaborazione effettuata sul file corretto (Vedi procedura CORTE)**
  - **Output: Setexm.pat, Setexm.lst**

```

LISTP=OUI, LISTF=OUI, NXLPA=59, LRESM=NON, NAPEL=0, MEM=3000000
:Réductions des seuils de fréquence des mots
NSPA ='RELIGIO.SPC'
NSPB ='RELIGIO.SPB'

PROC SETEX
=====
NSEU =50 NMOMI=2 NREMI=0 LEDIT =NEW
STOP

```

- **Procedura MOCAR, RECAR: parole e risposte (righe) caratteristiche**
  - **Prime 50 parole caratteristiche**
  - **Prime 20 risposte (righe) caratteristiche**
  - **Memoria=3000000**

- **Elaborazione effettuata dopo la riduzione di soglia (Vedi procedura SETEX1)**
- **Parametri/Output: Mocarm.pat, Mocarm.lst**

```

LISTP=OUI ,LISTF=OUI ,NXLPA=59 ,LRESM=NON ,NAPEL=0 ,MEM=3000000
NSPA = 'RELIGIO.SPB'
NTEXM= 'RELIGIO.TXM'

PROC MOTEX
:=====T d c m * t=====
LEDIT=OUI NVSEL=0

PROC MOCAR
:=====M C=====
NOMOT=50 NOREP=20

PROC RECAR
:=====R l c=====
NOREP=20
STOP
:=====

```

- **Procedura SEGME: calcolo delle frequenze dei differenti segmenti ripetuti e costruzione della tabella di contingenza che incrocia le risposte (in riga) ed i segmenti (in colonna)**
  - **Lunghezza massima di un segmento=10**
  - **Numero massimo di segmenti=50000**
  - **Soglia di frequenza generale=10; segmenti lunghezza 2=10; segmenti lunghezza 3=10**
  - **Numero massimo di segmenti di lunghezza 2=2200**
  - **Elaborazione effettuata dopo la riduzione di soglia (Vedi procedura SETEX1)**
  - **Parametri/Output: Segme.pat, Segme.lst**

```

LISTP=OUI ,LISTF=OUI ,NXLPA=59 ,LRESM=NON ,NAPEL=0 ,MEM=3000000
NSPA = 'RELIGIO.SPB'
NSME = 'RELIGIO.SME'
PROC SEGME
:=====
NXLON =10 NSEUG =10 NSEU2 =10 NSEU3 =10 NXSEG =50000 LEDIT=100 >
NXLE2 =2200
STOP

```

- **Procedura MOCAR: selezione dei segmenti caratteristici per classe di risposta.**

**Tavola di contingenza parole x testi**

- **50 segmenti caratteristici**
- **Elaborazione effettuata sul file di base (file di base rispetto alla procedura SEGME)**
- **Memoria=3000000**
- **Parametri/Output: Mocars.pat, Mocars.lst**

```
LISTP=OUI , LISTF=OUI , NXLPA=59 , LRESM=NON , NAPEL=0 , MEM=3000000
NTEXM= 'RELIGIO.TXM'
NSPA = 'RELIGIO.SME'
```

```
PROC MOTEX
:=====Table de contingence mots*textes=====
NVSEL=0 LEDIT=OUI
```

```
PROC MOCAR
:=====Mots caractéristiques=====
NOMOT=50 NOREP=0
```

```
STOP
```

- **Procedura APLUM per le parole: analisi delle corr. semplici parole x testi**

- **Numero assi calcolati=3**
- **Numero piani fattoriali=1**
- **Numero pagine piani fattoriali=1**
- **Numero linee per pagina=60**
- **Numero assi illustrati=2**
- **Classificazione per parole**
- **Numero assi utilizzati=4**
- **Lista numeri per classi=4**
- **Memoria=3000000**
- **Parametri/Output: Aplumm.pat, Aplumm.lst**

```
LISTP=OUI , LISTF=OUI , NXLPA=59 , LRESM=NON , NAPEL=0 , MEM=3000000
NSPA = 'RELIGIO.SPB'
NTEXM= 'RELIGIO.TXM'
```

```
PROC MOTEX
:=====
NVSEL=0 LEDIT=1
```

```

PROC APLUM
:=====
NAXE=3 LEDIT=OUI NGRAF=1 NPAGE=1 NLIGN=60

PROC CLAIR
:=====
NAXE=2 LIGN=OUI NMAX=10

PROC RECIP
:=====
NAXU=4 LDEND=DENSE NTERM=50 LDESC=NON

PROC PARTI
:=====
NITER=10 LEDIN=COMPL LEDIT=ACT
4

STOP

```

- **Procedura APLUM per i segmenti: analisi delle corr. semplici segmenti x testi**
  - **Vedi APLUM per le parole**
  - **Memoria=3000000**
  - **Parametri/Output: Aplums.pat, Aplums.lst**

```

LISTP=OUI , LISTF=OUI , NXLPA=59 , LRESM=NON , NAPEL=0 , MEM=3000000
NSPA = 'RELIGIO.SME'
NTEXM= 'RELIGIO.TXM'

```

```

PROC MOTEX
:=====
NVSEL=0 LEDIT=1

PROC APLUM
:=====
NAXE=3 LEDIT=OUI NGRAF=1 NPAGE=1 NLIGN=60

PROC CLAIR
:=====
NAXE=2 LIGN=OUI NMAX=10

STOP

```

## BIBLIOGRAFIA

- Alvisi F. Baldacci A. *Perl*, Università degli Studi di Bologna Facoltà di Scienze  
Cornacchini A. Matematiche, Fisiche e Naturali,  
<http://antares.csr.unibo.it/~alvisi/>
- Amaturo E. (1993) *Messaggio, Simbolo, Comunicazione. Introduzione all'Analisi del Contenuto*, La Nuova Italia Scientifica, Roma
- Giuliano L.C. *Appunti del corso di Metodi Quantitativi per le Scienze Sociali*, a.a. 96-97
- Benedetti C. (1989) *Istituzioni di Statistica*, Veschi, Milano
- Bolasco S. (1999) *Analisi Multidimensionale dei dati*, Carocci, Roma
- Bolasco S. (1997- 1998) *Metodi per l'analisi statistica dei dati testuali*, dispense a.a. 1997-1998 Corso di Statistica III – Facoltà di Economia, Scuola di Specializzazione di Metodi e Tecniche della Ricerca Sociale.
- Cipriani R. Bolasco S. *Ricerca qualitativa e computer*, Franco Angeli, Milano (1995)
- Crescimanni A. *Appunti di analisi dei dati*, dispense del corso di Analisi dei Dati del D.U. in Statistica a.a. 1997/98
- Cutillo Enrica A. *Lezioni di Statistica Sociale I,II*, CISU, Roma (1996)
- Dern D.D. (1995) *Alla scoperta di Internet*, McGraw-Hill, Milano
- Fraire M. (1994) *Metodi di Analisi Multidimensionale dei Dati*, CISU, Roma
- Frosini, Montinaro, Nicolini (1994) *Il Campionamento da popolazioni finite*, UTET, Torino
- JADT (2000) *Actes des 5<sup>es</sup> Journées internationales d'Analyses statisque des Données Textuelles 9-11 Mars2000*, M. Rajman & J.C.

- Chappelier, École Polytechnique fédérale de Lausanne  
 Leti G. (1983) *Statistica descrittiva*, Il Mulino, Bologna
- Liverani M. (1996) *Introduzione al Perl*, Dispense del corso di introduzione alla programmazione in Perl,  
<http://www.faqit.to/docs/perl/perl.html>
- Memoli R. – Saporiti A. (1995) *Disegno della Ricerca e Analisi dei Dati*, Euroma, Roma
- Orsi R. (1985) *Probabilità ed inferenza statistica*, Il Mulino, Bologna
- RFC 783* (1981) *The TFTP protocol (revision 2)*,  
<http://www.pasteur.fr/infosci/RFC/>
- RFC 821* (1982) *Simple mail transfer protocol*,  
<http://www.pasteur.fr/infosci/RFC/>
- RFC 822* (1982) *Standard for the format of Arpa Internet text messages*,  
<http://www.pasteur.fr/infosci/RFC>
- RFC 959* (1985) *File Transfer Protocol (FTP)*,  
<http://www.pasteur.fr/infosci/RFC/>
- RFC 1036* (1987) *Standard for Interchange of USENET Messages*,  
<http://www.pasteur.fr/infosci/RFC/>
- RFC 1123* (1989) *Requirements for Internet Hosts -- Application and Support*,  
<http://www.pasteur.fr/infosci/RFC/>
- RFC 1341* (1989) *MIME (Multipurpose Internet Mail Extensions): Mechanisms for Specifying and Describing the Format of Internet Message Bodies*, <http://www.pasteur.fr/infosci/RFC/>
- RFC 1421* (1993) *Privacy Enhancement for Internet Electronic Mail: Part I: Message Encryption and Authentication Procedures*,  
<http://www.pasteur.fr/infosci/RFC/>
- RFC 1422* (1993) *Privacy Enhancement for Internet Electronic Mail: Part II: Certificate-Based Key Management*,  
<http://www.pasteur.fr/infosci/RFC/>
- RFC 1423* (1993) *Privacy Enhancement for Internet Electronic Mail: Part III*:

- Algorithms, Modes, and Identifiers,*  
<http://www.pasteur.fr/infosci/RFC/>
- RFC 1424 (1993)* *Privacy Enhancement for Internet Electronic Mail: Part IV: Key Certification and Related Services,*  
<http://www.pasteur.fr/infosci/RFC/>
- RFC 1426 (1993)* *SMTP Service Extension for 8bit-MIMEtransport,*  
<http://www.pasteur.fr/infosci/RFC/>
- RFC 1521 (1993)* *MIME (Multipurpose Internet Mail Extensions) Part One: Mechanisms for Specifying and Describing the Format of Internet Message Bodies,* <http://www.pasteur.fr/infosci/RFC/>
- RFC 1563 (1994)* *The text/enriched MIME Content-type,*  
<http://www.pasteur.fr/infosci/RFC/>
- RFC 1740 (1994)* *MIME Encapsulation of Macintosh files – MacMIME,*  
<http://www.pasteur.fr/infosci/RFC/>
- RFC 1847 (1995)* *Security Multiparts for MIME: Multipart/Signed and Multipart/Encrypted,* <http://www.pasteur.fr/infosci/RFC/>
- RFC 1866 (1995)* *Hypertext Markup Language - 2.0,*  
<http://www.pasteur.fr/infosci/RFC/>
- RFC 1991 (1996)* *PGP Message Exchange Formats,*  
<http://www.pasteur.fr/infosci/RFC/>
- RFC 2015 (1996)* *MIME Security with Pretty Good Privacy (PGP),*  
<http://www.pasteur.fr/infosci/RFC/>
- RFC 2183 (1997)* *Communicating Presentation Information in Internet Messages: The Content-Disposition Header Field,*  
<http://www.pasteur.fr/infosci/RFC/>
- RFC 2387 (1998)* *The MIME Multipart/Related Content-type,*  
<http://www.pasteur.fr/infosci/RFC/>
- Ricolfi L. (1997) *La ricerca qualitativa*, NIS, Roma
- Smelser Neil J. (1995) *Manuale di Sociologia*, il Mulino, Bologna

Spad-T®

*Introduction à SPAD-T intégré Version 1.5 P.C, CISIA,  
Saint –Mandé (France)*

*What is SGML?*

*Extract from the OII Technology Handbook,*

[http://www.techapps.co.uk/iihb\\_sgml.html](http://www.techapps.co.uk/iihb_sgml.html)

Crittografia

<http://www.netzapping.com/banks/tesi/pesavento/crittog.htm>

<http://www.newtech.it/utenti/netzapping/banks/tesi/pesavento/crittog.htm>

*What is the UUencode Compression Program?*

<http://www.ualberta.ca/CNS/HELP/filetran/uuencode.html>

## **GLOSSARIO**

### **A**

ampiezza di banda (bandwidth): parte di banda occupata per la trasmissione di un segnale. Maggiore è l'ampiezza di banda, più veloce è la trasmissione. Una buona larghezza di banda permette l'invio di segnali complessi come quelli audio e video.

analogico (segnale): segnale continuo rappresentato con onde sinusoidali.

applet: applicazione Java (v.) di norma richiamata da una pagina HTML

ARPAnet : rete realizzata nel 1969 dal Ministero della difesa statunitense tramite l' Advanced Research Project Agency (ARPA). Ha preceduto Internet. Oggi non è più esistente.

asincrona (comunicazione): si realizza quando i soggetti comunicano ed accedono ad informazioni in tempi diversi ed a loro discrezione. Un esempio di tale modalità è la posta elettronica.

ATM: (Asynchronous Transfer Mode) modalità di trasferimento dati particolarmente veloce.

### **B**

backbone: "dorsale", rete ad alta velocità che connette gli host (v.) più potenti

baud: è l'unità di misura standard per la velocità di trasferimento dati di un modem. Un baud equivale quasi ad un bit

BBS (Bullettin Board System): bacheche elettroniche dove possono essere inviati e letti messaggi (solo testuali) pubblici. La modalità di comunicazione è asincrona.

bit: informazione elementare binaria.

bitnet: (Because It's Time Network: è il momento di mettersi in rete). E' una rete internazionale di computer orientata alle università e alla ricerca.

bounce: "rimbalzo", dicesi di un messaggio di posta elettronica che non arriva a destinazione ma si perde nella rete

BPS ( bit per secondo): unità di misura per la trasmissione dei dati su un canale.

browser: programma che permette di visualizzare pagine web. I più utilizzati sono Netscape Navigator e Internet Explorer.

byte: aggregazione di otto bit.

### **C**

chat: conversazione "via tastiera" tra gli utenti di computer collegati attraverso una rete telematica. La modalità di comunicazione è sincrona.

client: programma posto su un computer in rete che chiede servizi ad un altro computer della rete detto server della rete.

CMC (Computer Mediated Communication): comunicazione mediata dai computer e dalle reti che li collegano.

commutazione di circuito: processo mediante il quale i dati vengono trasferiti creando un appropriato collegamento tra chiamante e chiamato.

commutazione di pacchetto: processo mediante il quale i packet (v) vengono collocati sul canale e viaggiano attraverso la rete fino a destinazione.

computer conferencing system (CCS) : pacchetto software integrato che contiene le funzioni fondamentali per svolgere l'educazione a distanza on line. Queste sono essenzialmente due: la trasmissione dei contenuti da parte dell'insegnante e loro fruizione da parte dei discenti e la disponibilità di un'area di dibattito. Tale area viene realizzata grazie ad un sistema interno di posta elettronica che svolge anche un servizio di e-mail tra individui.

cyberspace: lo spazio virtuale dove si interagisce attraverso le reti telematiche. Il termine è stato coniato da William Gibson.

## D

dial-up: vedi Terminale (emulazione)

digitale (segnaletica): segnale discreto che consiste in una scelta tra due stati possibili. E' il segnale utilizzato per la trasmissione dei dati nelle reti telematiche.

distance education (educazione a distanza): esperienza di insegnamento/apprendimento in cui il docente e i discenti si trovano in una condizione di separazione fisica per supplire alla quale viene fatto uso di un qualsiasi mezzo di comunicazione: dal sistema postale alla televisione, alla telefonia fino alla telematica.

distance education on-line: forma di didattica a distanza basata sul principio della comunicazione mediata dal computer (CMC)

dominio: parte destra di un indirizzo Internet, per esempio "forcom.unito.it" di comunic(forcom.unit.it

down: dicesi down un sito che ha problemi tecnici tali da impedire l'accesso.

download: prelievo di un file da un computer remoto

## E

e-mail (electronic mail): servizio messo a disposizione da Internet che consente lo scambio di messaggi tra utenti che posseggono un indirizzo di posta elettronica. La comunicazione è asincrona e può essere uno a uno o uno a molti.

emoticone (v. smiley)

Eudora: programma ad interfaccia grafica che gestisce la posta elettronica.

## F

F2F (FaceTo Face): forma di didattica che prevede la compresenza fisica degli studenti e dell'insegnante.

FAQ (Frequently Asked Questions): un insieme di domande piu' frequenti relative ad un determinato argomento con relative risposte.

flame: messaggio scambiato in rete dal tono polemico-acceso.

follow-up: è la risposta ad una domanda inviata in un USENET o ad un listserv (v.).

freeware: software gratuito in rete.

FTP (File Transfer Protocol): protocollo per il trasferimento di file attraverso la rete.

## G

gateway: elaboratori specializzati che interconnettono reti di tipologia diversa.

GIF (Graphic Interchange Format): formato sviluppato per l'uso di qualità fotografica in immagini grafiche.

GNU: (Gnu is Not Unix): progetto della Free Software Foundation finalizzato alla realizzazione di una versione gratuita del sistema operativo UNIX.

gopher: protocollo Internet che permette di recuperare in rete dati e informazioni presso le università e i centri di ricerca di tutto il mondo.

hacker: persona che si diverte a portare l'hardware e il software ai loro limiti. Quelli comunemente detti "pirati informatici" sono invece definiti sulla rete come "crackers".

handshake: procedura che seguono due modem che provano a connettersi per stabilire come dovranno trasferire i dati.

hang: ciò che accade quando un modem fallisce l'aggancio della linea.

host: computer connesso alla rete in grado di ospitare utenti o informazioni.

HTML (Hyper Text Markup Language): è il linguaggio con il quale si scrivono le pagine da inserire su Internet

http (Hyper Text Transfer Protocol): protocollo di comunicazione attraverso il quale i client web si scambiano i documenti http con i server web.

## I

Internet: la più grande rete internazionale di computer fatta di piccole reti collegate grazie a protocolli internazionali.

Intranet: rete locale che pur non essendo necessariamente accessibile dall'esterno fa uso di tecnologie Internet.

IP (Internet Protocol): protocollo che in rete gestisce la trasmissione dei dati e la regolazione del traffico fra i vari computer. In particolare ha il compito di impacchettare i dati in uscita e di inviarli trovando la strada migliore per arrivare ad un particolare computer fra tutti quelli connessi nella rete.

IRC (Internet Relay Chat): software accessibile attraverso Internet che permette la comunicazione sincrona multiutente.

ISDN (Integrated Services Digital Network): standard di telecomunicazione che trasmette segnali digitali su linee telefoniche. E' capace di fornire agli utenti servizi diversi (per esempio voce, dati, immagini) utilizzando le stesse risorse hardware e software.

ISO (International Standardization Organization): gruppo di ricercatori che ha un nuovo protocollo Internet chiamato ISO/OSI.

## **J**

Java: linguaggio di programmazione sviluppato dalla Sun per realizzare applicazioni distribuite in rete.

## **L**

linea commutata: linea utilizzata per mettere in comunicazione due calcolatori o un terminale e un calcolatore venendo attivata solo durante lo scambio di dati.

linea dedicata: linea che mantiene in costante comunicazione due computer. Vi si fa ricorso per le comunicazioni di lunga durata o che hanno bisogno di elevata velocità ed affidabilità.

link: collegamento ipermediale fra unità informative.

lista: elenco di indirizzi di posta elettronica.

listserver: programma che risiede sullo stesso computer che ospita una lista e che possiede un proprio indirizzo di posta elettronica. Gestisce i messaggi inviati alla lista dagli iscritti alla stessa.

login: procedura di ingresso in un host computer.

logoff: procedura di uscita in un host computer.

LYNX: www browser (v.) solo testuale.

## **M**

mailing list: conferenza telematica i cui interventi vengono inviati direttamente nelle caselle di posta elettronica degli iscritti alla lista da parte di un listserver.

MIME (Multipurpose Internet Mail Extensions): formato per la codifica, la trasmissione e la gestione di informazioni binarie attraverso la rete.

modem: hardware usato dalla maggior parte dei computer per trasmettere dati su linee telefoniche.

MOO (Multi-user Dimension Object Oriented): sofisticato programma informatico che permette a più utenti di collegarsi via Internet ad un ambiente condiviso (che contiene stanze ed altri oggetti) e di interagire fra loro e con l'ambiente in sincrono.

MOSAIC: uno dei più famosi www browser.

motore di ricerca: programma raggiungibile attraverso la rete e in grado di indicizzare e fornire su richiesta dell'utente informazioni presenti su Internet.

MUD (Multiple User Dungeon): realtà virtuale basata su testo che permette agli utenti di interagire sia con l'ambiente che con gli altri utenti. E' nato alla fine degli anni '70 ed è ispirato ai giochi di ruolo.

multimedia: qualsiasi documento che utilizza molteplici forme di comunicazione quali testo, audio e video.

## N

netiquettes: insieme di norme di buona educazione per regolare il comportamento in rete.

Netscape: uno dei più recenti www browser grafici.

network: sistema di comunicazione che unisce due computer permettendo a questi di scambiare dati.

newsgroup: conferenza telematica la cui partecipazione non avviene tramite posta elettronica come nelle mailing-list (v.) ma attraverso specifici programmi newsreader.

## O

off-line: ciò che è sconnesso da un sistema telematico.

on-line: ciò che è connesso ad un sistema telematico.

OSI (Open System Interconnect): nuovo protocollo Internet creato dalla ISO (v.).

## P

packet: un insieme di dati senza un dimensione definita. E' il risultato della suddivisione da parte del protocollo TCP/IP di messaggi completi in blocchi.

password: parola d'ordine personalizzata di accesso ad un qualsiasi programma o servizio.

Pine: programma con interfaccia a caratteri che gestisce la posta elettronica.

Ping: programma in grado di tracciare il percorso di un messaggio da un sito ad un altro.

POP (Post Office Protocol): protocollo per la gestione della posta elettronica.

Post: comporre un messaggio per un newsgroup (v.) e spedirlo per condividerlo con gli altri.

Postmaster: persona da contattare ad un particolare sito per chiedere informazioni sul sito o su uno degli utenti dello stesso.

PPP (Point to Point Protocol): protocollo avanzato per il collegamento ad Internet.

protocollo: un insieme di regole o uno standard definito per permettere a diversi computer e reti di comunicare a livello internazionale.

provider: società o istituzione che fornisce l'accesso ad Internet:

## R

router: v. gateway

## S

server: programma su una macchina della rete che fornisce servizi.

shareware: software gratuitamente disponibile sulla rete per un periodo limitato dopo il quale è previsto un pagamento.

sincrona (comunicazione): comunicazione che obbliga i soggetti interessati ad interagire nello stesso momento. Un esempio è il chatting.

sito: termine generale con cui si indica un computer collegato ad Internet. Tipicamente è un computer che ospita informazioni, cioè un server.

SLIP (Serial Line Internet Protocol): protocollo per il collegamento alla rete attraverso una porta seriale. Tende ormai ad essere sostituito dal più avanzato PPP (v.).

smiley: modalità di descrivere le emozioni on-line utilizzando i caratteri della tastiera. Si leggono guardando da sinistra a destra. Esempio: :-)

surfing: muoversi fra le informazioni in rete. Si realizza "saltellando" da un link all'altro.

sysadmin: amministratore di sistema; persona che fa funzionare un host system o un sito ad accesso pubblico.

sysop: operatore di sistema.

## T

TCP/IP (Transmission Control Protocol /Internet Protocol): insieme di convenzioni tecniche che stanno alla base di Internet e che consente lo scambio di dati fra i computer in rete.

telnet: protocollo di collegamento che trasforma un computer in un terminale a caratteri di un computer remoto.

terminale (emulazione): fa in modo che un computer sembri correttamente collegato al computer dell'host system.

## **U**

UNIX: sistema operativo che, prima dell'avvento dei browser e dei protocolli quale www, era necessario usare per ricevere dati attraverso Internet.

upload: copiare/spedire un file da un sistema remoto.

URL (Uniform Resource Locator): indirizzo Internet per un gopher, un FTP o un sito web.

Usenet: un newsgroup Internet.

user name: qualsiasi combinazione di lettere e numeri che molti host system chiedono di inserire come nome utente la prima volta che ci si collega.

## **V**

Veronica: strumento per la ricerca di informazioni all'interno dello spazio dei gopher.

virtual community: v. corso!

VRML (Virtual Reality Modelling Language): linguaggio utilizzato per la creazione di oggetti e ambienti tridimensionali su Internet.

VT100: emulatore di terminale. E' probabilmente il più diffuso nei siti raggiungibili viai Telnet.

## **W**

WWW (World Wide Web): un sofisticato protocollo per navigare in Internet attraverso pagine ipertestuali e multimediali.

### Bibliografia

Corrado Petrucco, INTERNET. GUIDA PER I COMUNI MORTALI, Ed. Il cardo  
AAVV, INTERNET '96- MANUALE PER L'USO DELLA RETE, Laterza  
Tosoratti, Collinassi, INTRODUZIONE ALL'INFORMATICA, Ambrosiana

# INDICE ANALITICO

## /

/alternative .....	65
/basic .....	66
/digest.....	65
/enriched .....	64
/gif.....	66
/html.....	64
/jpeg .....	66
/mixed .....	64
/mpeg .....	67
/octet-stream.....	66
/parallel .....	65
/partial.....	65
/plain .....	63
/postscript.....	66
/rfc822.....	65
/richtext.....	64

## 7

7bit:.....	67
------------	----

## 8

8bit:.....	67
------------	----

## A

access-type=.....	66
afs .....	66
<i>ampiezza del vocabolario</i> .....	16
<i>analisi delle concordanze</i> .....	16
<i>analisi delle specificità</i> .....	46
<i>analisi lessicale</i> .....	46
anon-ftp.....	66
<i>Apple Macintosh</i> .....	69
<i>AppleDouble</i> .....	69
<i>AppleSingle</i> .....	69
application.....	66
application/pgp-encrypted.....	71
application/pgp-keys.....	71
application/pgp-signature.....	71
Approved: .....	61
ARPA .....	55
ARPAnet.....	<i>Vedi ARPA</i>
ASCII .....	57
assi fattoriali.....	50
attachment.....	68
audio .....	66

## B

base64:.....	67
<i>BBS</i> .....	52
<i>begin 644</i> .....	68
<i>BEGIN PGP MESSAGE</i> .....	72
<i>BEGIN PGP SIGNATURE</i> .....	72
<i>BEGIN PGP SIGNED MESSAGE</i> .....	72
<i>BEGIN PRIVACY-ENHANCED MESSAGE</i> ..	71
binary: .....	67
<i>body</i> .....	<i>Vedi corpo</i>
<i>boundary</i> .....	64; 157

## C

<i>carriage-return</i> .....	57
<i>charset</i> .....	63
<i>cmsg</i> .....	60
Coefficiente G di Guiraoud .....	26
<i>collettivo</i> .....	20
<i>commutazione di pacchetto</i> .....	55
concatenamento .....	56
condizione .....	15
Content: .....	60
Content-Description: .....	68
Content-disposition .....	68
Content-ID:.....	68
Content-Transfer-Encoding:.....	67
Content-Type:.....	63
contesto .....	15
<i>contesto locale</i> .....	16
corpo.....	57
<i>corpus</i> .....	13
creation-date=.....	68

## D

Date: .....	59
<i>dimensione (N)</i> .....	16
Dimensioni minime del corpus.....	30
directory= .....	66
<i>disambiguazione</i> .....	34
<i>discorso</i> .....	13
Distribution: .....	61
<i>dizionario</i> .....	17

## E

<i>end</i> .....	68
<i>enunciato</i> .....	16
estrazione di frasi significative .....	50

expiration .....	66
Expires: .....	60
external-body .....	66
<b>F</b>	
Fasce di frequenza .....	27
field-name .....	57
filename= .....	68
Follow-up: .....	60
forma testuale .....	21
forme composte .....	17
forme flesse .....	17
forme grafiche .....	16
frammenti .....	13
frase .....	16
frasi significative .....	50
frequenza relativa .....	23
From: .....	58
ftp .....	66
fusione .....	34
<b>G</b>	
Guiraoud .....	Vedi coefficiente G di Guiraoud
<b>H</b>	
hapax .....	26
header lines .....	Vedi intestazione
<b>I</b>	
id .....	65
idioma .....	15
idiomatiche .....	17
image .....	66
Indice D .....	31
Indice d'uso U .....	32
indice IS .....	38
Indici d'uso e di dispersione .....	31
indicizzazione .....	20
inline .....	68
Internet .....	53
internetworking .....	55
intestazione .....	57
IS Vedi Indice IS	
ISO .....	64
ISO-88591 .....	64
ISO-88592 .....	64
Isofrequenza .....	36
<b>L</b>	
Legge di Zipf .....	24
lemma .....	17
lemmatizzazione .....	20
lemmi .....	16
lessema .....	18

lessemi .....	17
lessia .....	21
lessicalizzazione .....	38
lessico .....	17
linee .....	55
line-feed .....	57
Lines: .....	61
linguaggio .....	15
linguaggio peculiare .....	39
Livello di soglia .....	29
Local mail user Agent .....	62; 73
local-file .....	66
locuzioni .....	17
<b>M</b>	
MacMIME .....	69
mail-server .....	66
message .....	65
message/external-body .....	66
message/external-body: .....	159
message/partial: .....	158
Message-ID: .....	59; 60
MIME .....	62
MIME a 8 bi .....	67
MIME a 8 bit .....	62
MIME-PEM .....	70
Mime-Version: .....	63
mode= .....	66
modification-date= .....	68
monema .....	18
monosemia .....	33
morfema .....	18
multipart .....	64
multipart/alternative: .....	157
multipart/digest: .....	158
multipart/related .....	69
multipart: .....	157
multipart\encrypted .....	71
multipart\signed .....	71
Murk-up tags .....	64
<b>N</b>	
name: .....	66
name= .....	66
News .....	54
Newsgroup .....	56
Newsgroup moderati .....	61
Newsreader .....	59
nodi .....	73
normalizzazione .....	55
null line .....	33
number .....	57
number .....	65
numerizzazione .....	20

**O**

<i>occorrenze</i> .....	13
<i>occorrenze normalizzate</i> .....	23
<i>omografie</i> .....	18
<i>Organization</i> : .....	59

**P**

<i>padding</i> : .....	66
<i>parlante</i> .....	13
<i>parole</i> .....	13
<i>parole chiave</i> .....	39
<i>parole piene</i> .....	18
<i>parole tema</i> .....	39
<i>parole vuote</i> .....	18
<i>Path</i> : .....	59
<i>permission</i> =.....	66
<i>PGP</i> .....	70; 71
<i>poliformi</i> .....	14
<i>polirematiche</i> .....	14
<i>polisemia</i> .....	18
<i>profili lessicali</i> .....	13
<i>protocol</i> =.....	71

**Q**

<i>gouting</i> .....	74
<i>quoted-printable</i> .....	67

**R**

<i>rango</i> .....	17
<i>Re</i> : .....	59
<i>read date</i> = .....	68
<i>References</i> : .....	59; 60
<i>Reply-to</i> : .....	60
<i>RFC</i> .....	56
<i>riconoscimento automatico</i> .....	20
<i>ricostruzione del senso</i> .....	46
<i>route</i> .....	55

**S**

<i>segmenti ripetuti</i> .....	14
<i>Selezione dei poliformi</i> .....	36
<i>Sender</i> : .....	60
<i>senso</i> .....	46
<i>separatori</i> .....	13
<i>server</i> = .....	66
<i>SGML</i> .....	64
<i>significante</i> .....	18; 19
<i>significato</i> .....	18; 19

<i>Simple Mail Transfer Protocol</i> .....	62
<i>sinonimia</i> .....	19
<i>sintagma</i> .....	14
<i>site</i> .....	66
<i>size</i> = .....	66
<i>SMTP</i> .....	62; 67
Specificità di forme e frasi in un testo .....	46
<i>start</i> = .....	69
<i>start-info</i> .....	70
Statistica testuale .....	6
studio verticale .....	46
<i>Subject</i> : .....	59
<i>subject</i> = .....	66
<i>Summary</i> : .....	61
sviluppo .....	23

**T**

<i>tabelle codici</i> .....	57
<i>tags</i> .....	64
tasso di copertura del testo .....	29
<i>TCP/IP</i> .....	55
<i>testo</i> .....	13
<i>text</i> .....	63
<i>tftp</i> .....	66
<i>total</i> .....	65
<i>Tournier</i> .....	8
<i>type/token ratio</i> .....	26
<i>type</i> : .....	66
<i>type</i> = .....	69

**U**

<i>unità di contesto</i> .....	16; 20
<i>unità di testo</i> .....	20
<i>unità statistica</i> .....	20
<i>US-ASCII</i> .....	64
<i>USENET</i> .....	52
<i>UUCP</i> .....	53
<i>UUENCODE</i> .....	67

**V**

<i>video</i> .....	67
<i>vocabolario</i> .....	16

**X**

<i>X-</i> .....	63
<i>Xref</i> : .....	61

**Z**

<i>Zipf</i> .....	Vedi Legge di Zipf
-------------------	--------------------