

UNIVERSITÀ DEGLI STUDI DI ROMA

“LA SAPIENZA”

FACOLTÀ DI SCIENZE STATISTICHE

Diploma di Statistica

PROGETTO CAMPUS

Alessandro Stabellini

SPAD•T®

Software per l'analisi dei dati testuali

Introduzione alle principali funzioni

Applicazione “Texte”

Procedura CORTE

Procedura SETEX

Procedura SEGME

Procedura MOCAR RECAR

Novembre 2000, n.1

**Dispense didattiche per il corso di
Metodi Quantitativi per le Scienze Sociali
a cura di Luca Giuliano
a.a. 2000-2001**

INDICE

Premessa	4
Introduzione.....	5
Configurazione del sistema operativo	6
DOS®	6
Windows® 9X	7
Windows® NT 4.0 – Windows® 2000.....	9
Windows® ME	10
Il file di dati (<i>corpus</i>).....	11
Le Procedure.....	14
Creazione di un'applicazione Spad-T®	14
Procedura CORTE	20
Procedura SETEX.....	22
Procedura SEGME.....	24
Procedura MOCAR RECAR	26
Appendice A	31
Bibliografia	33

PREMESSA

Questa dispensa si propone come obiettivo quello di fornire, a tutti coloro che vogliono effettuare un'applicazione di statistica testuale, uno strumento agile e sintetico di guida per alcune delle principali funzioni del software Spad-T®.

In particolare verranno esaminate le seguenti applicazioni/procedure:

- Applicazione “Texte”
- Procedura CORTE
- Procedura SETEX
- Procedura SEGME
- Procedura MOCAR RECAR

Contestualmente vengono forniti alcuni utili consigli circa l'installazione del software sui principali sistemi operativi e sull'impostazione dei parametri e del file di dati.

Per eventuali approfondimenti si rimanda il lettore a volumi specifici quali il manuale d'uso: Lebart L. Morineau A. Becue M., Haeusler L., *Introduction à SPAD-T intégré Version 1.5 P.C.* CISIA, Saint – Mandé (France), 1993.

INTRODUZIONE

Lo Spad-T® è un software di origine francese orientato all'analisi statistica di un testo secondo un approccio *lessicometrico*.

Lo Spad-T® riconosce ogni parola di un testo come forma grafica, ovvero come un elemento il cui tratto caratteristico è l'apparenza grafica (parole tali e quali scritte nel testo: poiché diverso da poichè) e basa l'elaborazione:

- sulla tabella lessicale intera riferita al *corpus* oggetto di studio;
- sulla frequenza d'occorrenza delle parole del *corpus*;
- sul confronto fra "profili" lessicali delle differenti parti del *corpus*.

Lo Spad-T® fornisce degli strumenti computazionali che trovano applicazione nella Statistica Linguistica:

- procedura NUMER: numerizzazione delle parole, calcolo della loro lunghezza e frequenza;
- procedura SEGME: generazione dei segmenti ripetuti;
- procedura MOTEX: generazione della tabella di contingenza.

E nella Statistica Testuale:

- procedura CORTE: correzione del vocabolario;
- procedura SETEX: riduzione della soglia di frequenza;
- procedura CORDA: concordanza delle parole;
- procedura MOCAR: individuazione delle parole o dei segmenti caratteristici;
- procedura RECAR: individuazione delle risposte caratteristiche;
- procedura APLUM: analisi delle corrispondenze semplici parole*testi o segmenti*testi;
- procedura ASPAR: analisi delle corrispondenze multiple risposte*parole o risposte*segmenti;
- procedura PARTI: creazione di cluster;
- procedura CLAIR: interpretazione dei fattori.

CONFIGURAZIONE DEL SISTEMA OPERATIVO

Lo Spad-T® è un applicativo per DOS® (3.3 o superiore) che richiede al sistema delle risorse di memoria fisica di cui i computer di oggi dispongono, sicuramente, in misura maggiore: 5.5 MB liberi su disco rigido e 2 MB di memoria RAM minimi.

Bisogna in ogni modo tener conto che – per come lavora lo Spad-T® – la velocità operativa di un processo, all'interno del programma, dipende, non solo dalla potenza di calcolo della CPU, ma anche dalle *performance* del disco rigido. Questo perché lo Spad-T®, durante l'elaborazione, crea alcuni file temporanei necessari allo svolgimento dei calcoli, sfruttando, così, pesantemente l'Hard disk.

Lo Spad-T® può essere usato tranquillamente all'interno di Windows®. Sono richiesti, però, alcuni accorgimenti che saranno proposti al lettore più avanti in questo paragrafo.

Nelle configurazioni che verranno descritte si presuppone che lo Spad-T® sia installato sul disco C nella *directory* C:\SPADT

DOS®

È necessario modificare alcuni parametri¹ in due file che il sistema operativo carica al momento dell'avvio.

- Nel file CONFIG.SYS

Aggiungere le righe: FILES=40 e BUFFERS=30 oppure, qualora si voglia sfruttare l'area di memoria superiore per caricare le due istruzioni, FILESHIGH=40 e BUFFERSHIGH=30. In quest'ultimo caso occorre però assicurarsi che siano presenti, in **ordine**, oltre all'istruzione di caricamento del gestore di memoria estesa HIMEM.SYS, riga DEVICE=C:\WINDOWS\HIMEM.SYS², anche le seguenti istruzioni:

DOS=UMB

DOS=HIGH

¹ Per modificare il CONFIG.SYS e l'AUTOEXEC.BAT sarà sufficiente usare un editor di testi ASCII.

² La riga del DEVICE riporta il percorso completo del file. Il file può avere anche un'altra collocazione. Su sistemi senza Windows®, ad esempio, il file si trova nella *directory* DOS. Il file può trovarsi addirittura nella *directory* WINDOWS/COMMAND

DEVICE=C:\WINDOWS\EMM386.EXE³. Quest'ultimo per la gestione della memoria espansa.

- Nel file AUTOEXEC.BAT aggiungere l'istruzione SET SPADT=C:\SPADT e l'istruzione SET PATH=%PATH%;C:\SPADT.

Qualora si dovessero incontrare dei problemi di scarsa memoria allocata, si provi a cancellare dall'HIMEM.SYS l'eventuale parametro /NOEMS ed aggiungere il parametro /NUMHANDLES=127.

La riga del CONFIG.SYS, cambierà e diventerà:

DEVICE=C:\WINDOWS\HIMEM.SYS /NUMHANDLES=127

Inoltre, sempre in presenza di problemi di gestione di memoria, si provi a risolverli editando la riga DEVICE=C:\WINDOWS\EMM386.EXE del CONFIG.SYS facendola diventare DEVICE=C:\WINDOWS\EMM386.EXE xxx RAM (con xxx = 512 come valore minimo)

WINDOWS® 9X

Si può far "girare" tranquillamente lo Spad-T® in una finestra DOS® di Win9x sfruttando il *multitasking* che questo sistema operativo offre senza, ogni volta, "riavviare in modalità MS-DOS" per eseguire il programma.

Sono necessari, però, anche in questo caso, alcuni accorgimenti.

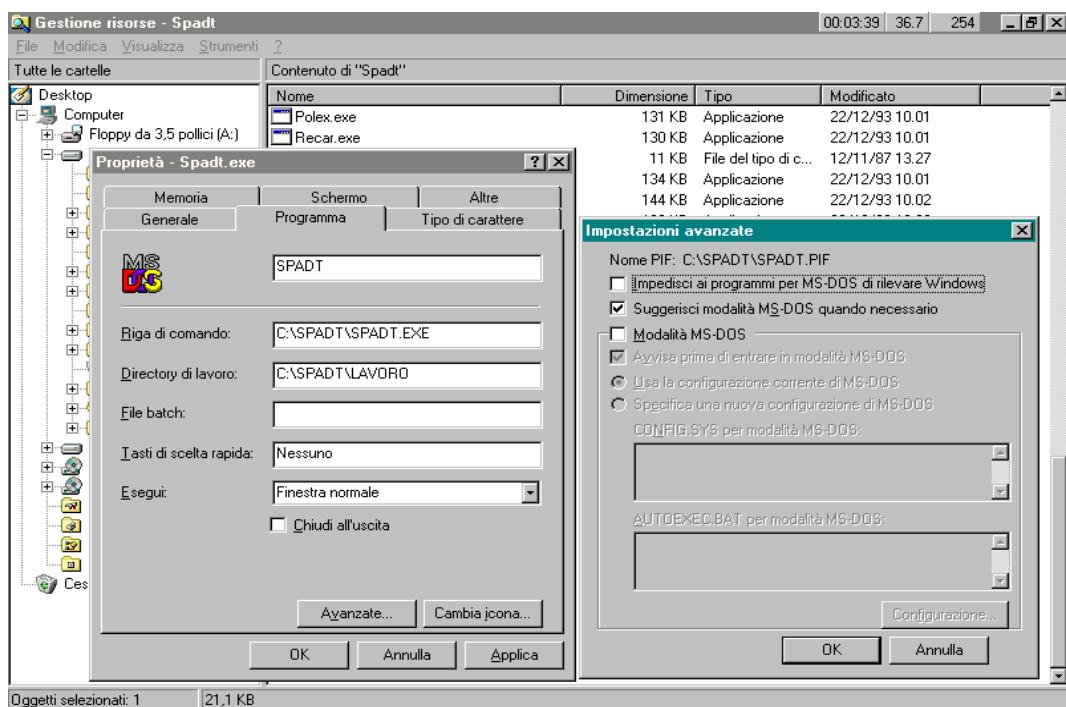
Tramite "Gestione risorse" di Win9x occorrerà aprire la directory C:\SPADT ed evidenziare le proprietà del file SPADT.EXE (tasto destro del mouse sul file e scelta dell'opzione "proprietà" dal menù contestuale che appare). Scegliendo la lingetta "programma" sarà proposta una schermata in cui si potrà scegliere la *directory* (o cartella) di lavoro. Indicarne una a piacimento⁴. Questa sarà la *directory* in cui il programma effettuerà tutti i calcoli, creando e cancellando i file funzionali alle sue procedure interne.

Si consiglia di indicare una *sottodirectory* dello Spad-T® come *directory* di lavoro e non già la sua stessa *directory*. Soprattutto per un fattore di comodità: nella *sottodirectory* indicata, infatti, si potranno ritrovare tutti i file del lavoro che si sta svolgendo.

³ Vedi nota 2

⁴ Nella scelta dei nomi si eviti di scegliere nomi più lunghi di 8 caratteri per problemi di incompatibilità dello Spad-T® stesso con i nomi che non seguono la regola MS-DOS® dell'8+3 (nome file+estensione).

Figura 1: Creazione del file SPADT.PIF



Dal menù "avanzate" scegliere l'opzione: "Suggerisci modalità MS-DOS quando necessario". Dare una serie di OK fino a ritornare alla finestra di partenza. A questo punto verrà automaticamente creato un file SPADT.PIF che è un file di collegamento all'applicazione SPADT.EXE nel quale sono presenti le configurazioni appena effettuate. Basterà "tagliare" e "incollare" il file SPADT.PIF sul *desktop* (o dove si vuole) per avere un collegamento allo Spad-T® che lavora in una particolare *directory*.

L'operazione appena descritta può essere ripetuta quante volte si vuole ed è utilissima in presenza di più *set* di dati per i quali - ad esempio - si vogliono eseguire studi diversi. Per ciascuno studio può esser creata una differente cartella.

In caso di file di dati molto grandi può insorgere - durante l'elaborazione - un problema di incompatibilità che non consente allo Spad-T® di proseguire nei calcoli.

Per risolvere il problema occorre aprire con un editor di testi il file SYSTEM.INI nella *directory* di Windows® ed aggiungere la riga DEVICE=PHARLAP.386 nella sezione [386Enh]. In aggiunta, occorre copiare il file PHARLAP.386⁵ nella *directory* di Windows® e nella sua *sottodirectory* SYSTEM.

⁵ Il file PHARLAP.386 si trova, già presente, nella *directory* dello Spad-T®.

WINDOWS® NT 4.0 – WINDOWS® 2000

Lo Spad-T® può “girare” tranquillamente sotto Windows®NT o Windows®2000 eseguendo degli aggiustamenti nel file SPADT.PIF del collegamento all’eseguibile (Vedi pag.8).

NT - o Win2000 - non hanno una modalità DOS® in cui far girare i programmi a 16 bit, bensì una *macchina virtuale* in cui viene emulato l’ambiente DOS®. Anche se il concetto, rispetto a Win9x è diverso, il risultato non cambia: Lo Spad-T® può girare in NT (o in Win2000) sotto finestra.

Il file SPADT.PIF in NT (o Win2000) è leggermente diverso.

Al posto del menù “avanzate” c’è il menù “Windows NT” (in Win2000 rimane il menù “avanzate”).

Se lo aprite vi accorgerete che di *default* Windows® NT (Win2000) riporta, su di una finestra, due file:

il file AUTOEXEC.NT ed il file CONFIG.NT. Basterà editare questi due file⁶ con un editor di testi in formato ASCII ed **aggiungere**⁷:

- al file AUTOEXEC.NT le righe:

```
SET SPADT=C:\SPADT  
SET PATH=%PATH%;C:\SPADT
```

- al file CONFIG.NT le righe:

```
FILES=408  
BUFFERS=30
```

Salvare i due file con un differente nome, avendo cura di collocarli sempre nella *directory* WINNT/SYSTEM32 e ricordarsi di cambiare i loro nomi nel menù “Windows NT” (o “avanzate” in Win2000) dalle “proprietà” del file SPADT.PIF.

Per quanto riguarda la gestione della memoria, menù “memoria” delle “proprietà” del file “SPADT.PIF”, le impostazioni devono essere le seguenti:

- Memoria convenzionale: Totale “Automatica”; Iniziale “Automatica”;
- Memoria espansa (EMS): “Nessuna”;
- Memoria estesa (XMS): “Nessuna”;

⁶ I file si trovano nella *directory* C:\WINNT\SYSTEM32

⁷ Si dà particolare importanza al termine “aggiungere” perché le istruzioni contenute già di *default* nei due file sono vitali per il corretto funzionamento dello Spad-T®

⁸ Se il valore di FILES è già a 20, portarlo a 40.

- Memoria modalità protetta (DPMI): “Automatica”.

Questo è tutto.

Potrà essere necessario editare, come in Win9x, il file SYSTEM.INI nella directory di WinNt per aggiungere i riferimenti al file PHARLAP.386, ma rimandiamo il lettore alla pag.8 dato che l’operazione è la medesima di quella effettuabile in Win9x.

WINDOWS® ME

Windows® ME – sostituto di Windows® 98 – non dà la possibilità, come il suo predecessore, di far girare i programmi DOS® nell’apposita modalità. Pur tuttavia esiste, nel nuovo sistema operativo di Microsoft®, il “prompt MS-DOS” e quindi la facoltà di far girare lo Spad-T® in “finestra”.

Eseguendo gli stessi passaggi descritti a pag. 7 per Win9x relativamente alla creazione del file SPADT.PIF, si dovrà, in WinME, avere accortezza di settare in maniera corretta i parametri della memoria (tasto destro del mouse sul file SPADT.PIF, cliccare su “proprietà” e poi su “memoria”).

Questi sono i valori che bisogna impostare:

- Memoria convenzionale: “Automatica”;
- Ambiente iniziale: “Automatico”;
- Memoria espansa (EMS): “Nessuna”;
- Memoria Ms-Dos (DPMI): “Automatica”.

In aggiunta è necessario editare con un editor di testi ASCII il file AUTOEXEC.BAT per aggiungervi le due righe:

SET SPADT=C:\SPADT

SET PATH=%PATH%;C:\SPADT

Ed il file CONFIG.SYS per inserire le righe:

FILES=40

BUFFERS=30

IL FILE DI DATI (*CORPUS*)

Per impostare il file di dati in maniera tale che lo Spad-T® possa interpretarlo, sarà necessario usare un editor di testi in grado di salvare il file in formato testo (estensione txt) con righe non più lunghe di 80 colonne⁹ (salvataggio con interruzione di linea).

Bisogna però fare una precisazione. La tabella codice¹⁰ riferita ad un determinato *set* di caratteri è diversa in DOS® ed in Windows®. In poche parole un determinato carattere, sotto DOS®, può avere una forma grafica totalmente diversa in Windows®. È il caso, ad esempio, delle lettere accentate (Vedere Appendice A).

Questo non rappresenta un problema dal momento che lo Spad-T®, nella creazione della tabella lessicale, si riferisce ai codici esadecimali dei caratteri contenuti nel file di testo in formato binario. Ciò significa che se si decide di editare il file di dati con un editor per Windows®, la lettura dei risultati deve avvenire con lo stesso editor o comunque con un editor capace di leggere i codici ANSI Windows®¹¹ (Vedere Appendice A) a meno che non si decida già di salvare il file con un set di caratteri DOS® (salva in modalità MS-DOS)¹² e di usare l'editor interno dello Spad-T® o l'edit del DOS® per ogni operazione successiva.

Avendo ben chiare le poche regole appena menzionate circa i mezzi da utilizzare per editare il file di dati (*corpus*), vediamo come questo deve essere strutturato.

In questa dispensa noi ci riferiamo all'applicazione "Texte" (o testo unico) dello Spad-T® utilizzata nel caso in cui non si abbia esigenza di associare al file di dati un file numerico per la codifica delle variabili.

La "Texte" richiede che ciascun testo, in cui è ripartito il *corpus*, sia inserito tra due righe di testo che inizino con 4 asterischi più l'eventuale etichetta identificativa del testo.

Il *corpus* deve terminare con una riga composta da una sequenza di 4 segni di ugualie (====).

⁹ Tale limitazione è data dal modo in cui lo Spad-T® legge il testo: tutto ciò che si trova dalla 81-esima colonna in poi non viene considerato in alcun modo.

¹⁰ Nella tabella codice, per ogni carattere, è riportato l'equivalente codice decimale o esadecimale.

¹¹ Bisogna fare attenzione a tutte le istruzioni funzionali allo Spad-T® per la formattazione e l'elaborazione - quest'ultime contenute nei file di parametri con estensione "pat" nella directory di lavoro scelta per l'elaborazione (Vedi pag. 31) - che debbono essere riportate in caratteri DOS®.

¹² L'operazione di salvataggio in un *set* di caratteri non è reversibile: se si salva in modalità MS-DOS non si può poi ritornare al *set* ANSI Windows® aprendo e salvando di nuovo il file in formato testo (estensione txt).

Il file di dati avrà quindi la seguente struttura:

Figura 2: Il file di dati nell'applicazione “Texte”. Una riga¹³ un individuo. (Per individuo si intende l'entità nella quale viene ripartito, a sua volta, il testo)

File di dati	
****Testo 1 Questa è la prima riga del testo uno Questa è la seconda riga del testo uno ... Questa è la i -esima riga del testo uno ... Questa è la n -esima riga del testo uno ****Testo 2 Questa è la prima riga del testo due Questa è la seconda riga del testo due ... Questa è la i -esima riga del testo due ... Questa è la n -esima riga del testo due ... ****Testo n Questa è la prima riga del testo n Questa è la seconda riga del testo n ... Questa è la i -esima riga del testo n ... Questa è la n -esima riga del testo n =====	-> Individuo 1_Testo 1 -> Individuo 2_Testo 1 -> Individuo i _Testo 1 -> Individuo n _Testo 1 -> Individuo 1_Testo 2 -> Individuo 2_Testo 2 -> Individuo i _Testo 2 -> Individuo n _Testo 2 -> Individuo 1_Testo n -> Individuo 2_Testo n -> Individuo i _Testo n -> Individuo n _Testo n

Qualora, invece, si volessero ripartire i testi del *corpus* in entità diverse rispetto a ciascuna riga componente, basterà inserire una sequenza di 4 segni meno (----): il testo racchiuso tra questi segni rappresenta un individuo nel senso visto in Figura 2.

Un esempio è riportato in Figura 3.

¹³ Una riga può essere al massimo lunga 80 colonne (Cfr. Nota 9)

Figura 3: Il file di dati nell'applicazione “Texte”. Un individuo su più righe. (Per individuo si intende l'entità nella quale viene ripartito, a sua volta, il testo)

File di dati	
****Testo 1 Questa è la prima riga del testo uno Questa è la seconda riga del testo uno ---- Questa è la terza riga del testo uno ---- ... Questa è la <i>i</i> -esima riga del testo uno ... Questa è la <i>n</i> -esima riga del testo uno	} Individuo 1_Testo 1 } Individuo 2_Testo 1 } Individuo 3_Testo 1
****Testo 2 Questa è la prima riga del testo due Questa è la seconda riga del testo due ... Questa è la <i>i</i> -esima riga del testo due ---- ... Questa è la <i>n</i> -esima riga del testo due	} Individuo 1_Testo 2 } Individuo <i>n</i> _Testo 2
****Testo n Questa è la prima riga del testo n ---- Questa è la seconda riga del testo n ... Questa è la <i>i</i> -esima riga del testo n ... Questa è la <i>n</i> -esima riga del testo n ====	} Individuo 1_Testo n } Individuo 2_Testo n

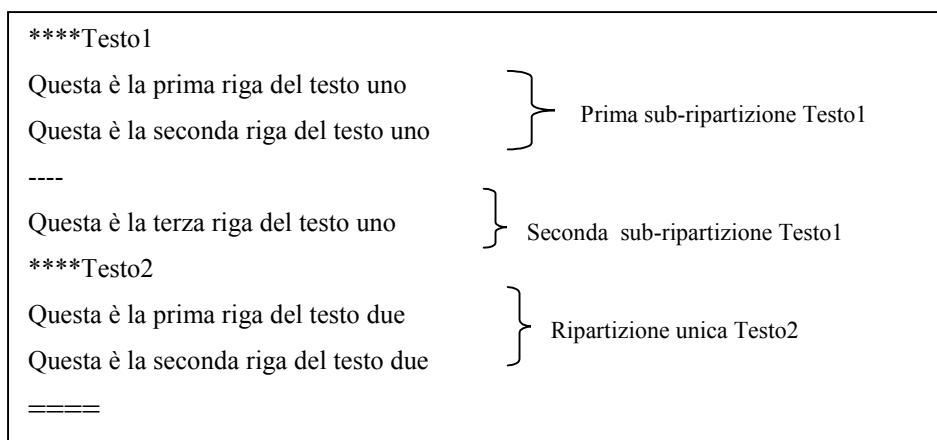
LE PROCEDURE

Creazione di un'applicazione Spad-T®

Figura 4: Schermata principale dello Spad-T®



Immaginiamo di aver registrato il file di dati (*corpus*) sotto il nome corpus.txt¹⁴ e di aver dato ad esso la seguente struttura:



¹⁴ In relazione a quanto detto per la *directory* di lavoro a pag. 7, si consiglia di creare una cartella "nome dell'applicazione" (corpus) nella quale copiare il file "corpus.txt" insieme al collegamento al file "spadt.exe". Avviato il programma tramite il collegamento, si ritroveranno nella *directory* scelta i file generati dallo Spad-T® nonché i listati dei risultati.

Nel *corpus* sono contenuti 2 testi di cui il primo è ripartito in 2 subtesti o paragrafi o entità (lo Spad-T® chiama queste entità “individui”. Cfr. Figura 2 e Figura 3).

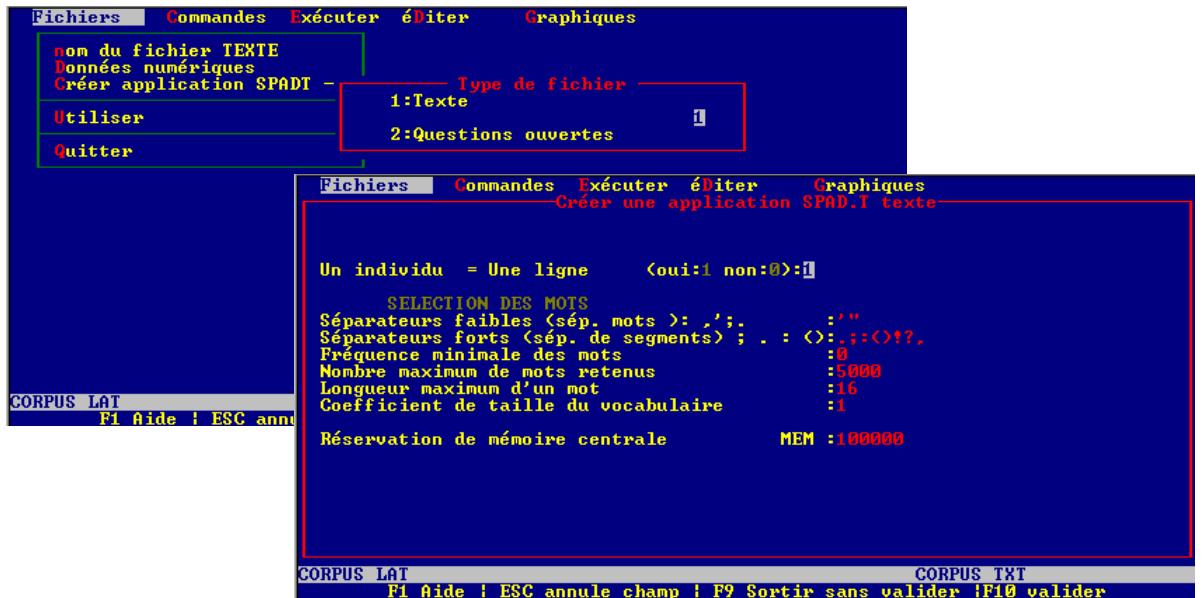
Per dare inizio all’analisi testuale occorrerà aprire il menù “Fichiers” (la combinazione di tasti per aprire i menù e per qualsiasi altra operazione effettuabile sull’interfaccia grafica viene riportata sulla barra delle istruzioni. Cfr. Figura 4) e scegliere “nom du fichier TEXTE”. Indicare il nome del file da analizzare (corpus.txt) e premere il tasto ESC per continuare.

Il passo successivo sarà quello di selezionare, sempre dal menù “Fichiers”, l’opzione “Créer application SPADT”).

Il programma richiederà a questo punto di indicare il tipo di dato testuale che si vuole analizzare. Occorrerà scegliere l’opzione “Texte” e validare con il tasto “F10”.

Successivamente apparirà una nuova schermata in cui si possono indicare i parametri per la lettura del file di dati.

Figura 5: Applicazione Texte



I parametri dell’applicazione “Texte” sono:

- un individuo = una linea (sì/no) [selezionare sì/no a seconda del tipo di dati. Cfr. Figura 2, Figura 3];
- il tipo di separatori deboli (i separatori delle parole o caratteri che non vengono considerati nel momento della generazione della tabella lessicale);

- il tipo di separatori forti (i separatori che non vengono considerati nella costruzione della tabella lessicale ma che costituiscono gli elementi su cui costruire i segmenti ripetuti);
- la soglia minima di frequenza delle parole da includere nello studio;
- il numero massimo di parole da considerare;
- il numero massimo di lettere di ciascuna forma;
- la memoria da allocare per l'elaborazione;

Per file di piccole dimensioni, i parametri indicati di *default*, riguardanti il numero massimo delle parole da considerare e la memoria allocata, sono sufficienti per l'elaborazione. Nel caso di file di dati di grosse dimensioni occorrerà invece aumentare i valori alla bisogna.

Premendo il tasto F10 verrà richiesto di salvare l'applicazione con un nome a scelta. Nell'esempio è stato scelto il nome "corpus".

I parametri potranno essere editati con comodo, in qualsiasi momento, aprendo il file "nome_applicazione.pat" (corpus.pat).

Aprendo il file corpus.pat esso apparirà nella seguente "veste":

Figura 6: Il file di parametri dell'applicazione "Texte"

```

NXLPA=60 LISTP=1 LISTF=1 LRESM=OUI NAPEL=0 MEM=100000
: ===== Lecture des textes
NTEXZ='CORPUS.TXT' NTEXM='CORPUS.TXM'

PROC ARTEX
=====
ITYP=1 LIREP=0

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

NSPA='CORPUS.SPA'

PROC NUMER
=====
:Numérisation des textes
NSEU=0 NMOMI=0 LEDIT=TOT NXLET =16 NXMAX=5000 COEF =1
FAIBLE ''
FORT .;:( )!?,
FIN
STOP

```

Tenendo conto che:

- l'ordine dei parametri è indifferente;
- la forma in cui i parametri vengono indicati è: [parametro]=[valore];
- il formato è libero;
- i separatori ammessi sono gli spazi bianchi, le virgole e le barre oblique;
- il simbolo “>” su di una riga permette di continuare sulla riga successiva;
- il simbolo “:” all'inizio della riga annuncia un commento contenuto sulla riga;
- il parametro NOPAR (no parameter) può essere usato per non indicare quei parametri a cui – comunque – viene dato un valore di *default* dallo Spad-T® nel momento dell'elaborazione;
- l'istruzione di fine “STOP” deve essere in maiuscolo.

Descriviamo i parametri:

NXLPA: numero di linee per pagina dell'elenco dei risultati. [0 o superiore a 20. Valore di *default* 60]. Il valore 0 fa sì che lo Spad-T® non inserisca nel file dei risultati nessuna interruzione di pagine.

LISTP: lista, nell'elenco dei risultati, dei parametri impostati [1=Sì, 0=No. Valore di *default* 0].

LISTF: lista, nell'elenco dei risultati, degli elementi caratteristici del file di dati (numero di risposte, numero di individui ecc.) [1=Sì, 0=No. Valore di *default* 1].

LRESM: Dà indicazioni, nell'elenco dei risultati, circa la memoria allocata, tramite il parametro MEM (vedere più avanti) e richiesta dal programma [1 o OUI = Sì, 0 o NON = No. Valore di *default* 0 o NON].

NAPEL: numero di chiamate preliminari al generatore di numeri aleatori [0 o più. Il valore 0 fa sì che il generatore venga reinizializzato ad ogni chiamata. Valore di *default* 0].

MEM: memoria allocata. [tutti i valori positivi. Valore di *default* 50 000].

NTEXZ: nome del file di dati sorgente (*corpus*).

NTEXM: nome del file di dati utile.

PROC ARTEX: procedura di archiviazione del testo.

ITYP: natura dei dati testuali [1 o TEX = testi, 2 o ENQ = domande aperte].

LIREP: indicatore di definizione di un individuo (Cfr.Figura 2, Figura 3) [0 o NON = se esiste un marcitore di fine “individuo” – linea che inizia con “----” – 1 o OUI = se ogni linea è un “individuo”. Valore di *default* 0]

PROC SELOX: procedura per selezionare le domande aperte da trattare. Valido nel caso di **ITYP=2**. Nel caso di **ITYP=1** (testi unici), **NUMQ** deve essere uguale ad uno (Vedi sotto)

NUMQ: numero della domanda selezionata (Cfr. la **PROC SELOX**)

LDONA: presenza di un file numerico da associare al file di dati [0 o NON = assenza; 1 o OUI = presenza. Valore di *default* 0]

NSPA: nome del file dati testuali numerizzato (individui x forme)

PROC NUMER: procedura di numerizzazione del testo

NSEU: soglia di frequenza [Tutti i valori positivi o 0. Valore di *default* 0]

NMOMI: lunghezza minima di una forma [Valori inferiori a **NXLET**. Valore di *default* 0]

LEDIT: parametri per la tabella di frequenza delle forme [0 o NON = nessuna, 1 o ALPHA = tabella per ordine alfabetico, 2 o TRI = tabella per ordine di frequenza, 3 o TOT = tabella per ordine alfabetico e per ordine di frequenza. Valore di *default* 1]

NXLET: numero massimo di lettere di un forma grafica [Valori da 10 a 20. Valore di *default* 16]

NXMAX: numero massimo di forme distinte considerate [Valori > 50. Valore di *default* 60]

COEF: coefficiente moltiplicatore dell’istogramma di frequenza delle forme per lunghezza. [valori minori di 11. Valore di *default* 1].

L’istogramma di frequenza restituisce, per lunghezza di ciascuna forma grafica, il valore di frequenza massima prevista.

Di seguito viene riportato l’istogramma di *default*:

Lunghezza:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Frequenza:	40	110	220	430	550	550	600	530	450	300	210	110	75	60	50	25

Il parametro **COEF** sostanzialmente moltiplica il valore di frequenza di *default*.

Qualora si avesse bisogno di variare singolarmente i valori, occorrerà editare una propria lista da inserire nel file “*.pat”.

L’istruzione sarà del tipo:

HISTO $n_1 \ n_2 \ n_3 \ \dots \ n_k$

Dove con n_k si indica il valore di frequenza massima per la forma di lunghezza K.

Nota: se il numero dei valori da inserire dopo l'istruzione **HISTO** fosse tale da superare le 80 colonne di testo, inserire nella colonna 79 il simbolo “>” e continuare l'elenco alla riga successiva.

FAIBLE: lista dei separatori deboli

FORT: lista dei separatori forti

Per mandare in esecuzione i parametri appena visti occorrerà aprire il menù “Exécuter” e scegliere l'opzione “Exécuter”. Per la combinazione di tasti da utilizzare, riferirsi alle istruzioni contenute nella barra delle istruzioni (Cfr. Figura 4).

I risultati verranno registrati in un file con lo stesso nome dato all'applicazione, ma avente estensione “.lst”. Nel nostro caso corpus.lst

In questo file sono contenuti, oltre ai risultati, anche i parametri impostati e gli eventuali errori che il programma ha incontrato durante l'elaborazione.

Importantissimi i passi del testo del file in cui vengono riportati i dati circa la memoria allocata e la memoria utilizzata per ciascuna procedura, l'istogramma di frequenza e la costituzione del testo.

Di seguito vengono riportati alcuni di questi passi:

RESERVATION DE MEMOIRE CENTRALE	VOUS AVEZ DEMANDE	100000 → Memoria allocata
-----	VOUS AVEZ BESOIN DE ...	4923 → Memoria necessaria
EFFECTIFS SELON LE NOMBRE DE LETTRES → istogramma di frequenza : controllare che i valori osservati siano minori di quelli previsti		
NOMBRES DE LETTRES	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16	
EFFECTIFS OBSERVES	1 1 3 1 3 1 1 0 0 0 0 0 0 0 0 0	
EFF. MAXIMAUX PREVUS	40 110 220 430 550 550 600 530 450 300 210 110 75 60 50 25	
BILAN DU TRAITEMENT		

NOMBRE TOTAL DE REPONSES =	3	→ Numero di risposte (summa delle ripartizioni)
NOMBRE TOTAL DE MOTS =	40	→ Numero totale delle parole
NOMBRE DE MOTS DISTINCTS =	11	→ Numero parole distinte
POURCENT. MOTS DISTINCTS =	27.5	→ Perc. Parole distinte
SELECTION DES MOTS		

SEUIL DE FREQUENCE =	0	→ Soglia di freq. (0, ovvero tutte le parole)
TOTAL DES MOTS RETENUS =	40	→ Totale delle parole considerate
MOTS DISTINCTS RETENUS =	11	→ Parole distinte considerate

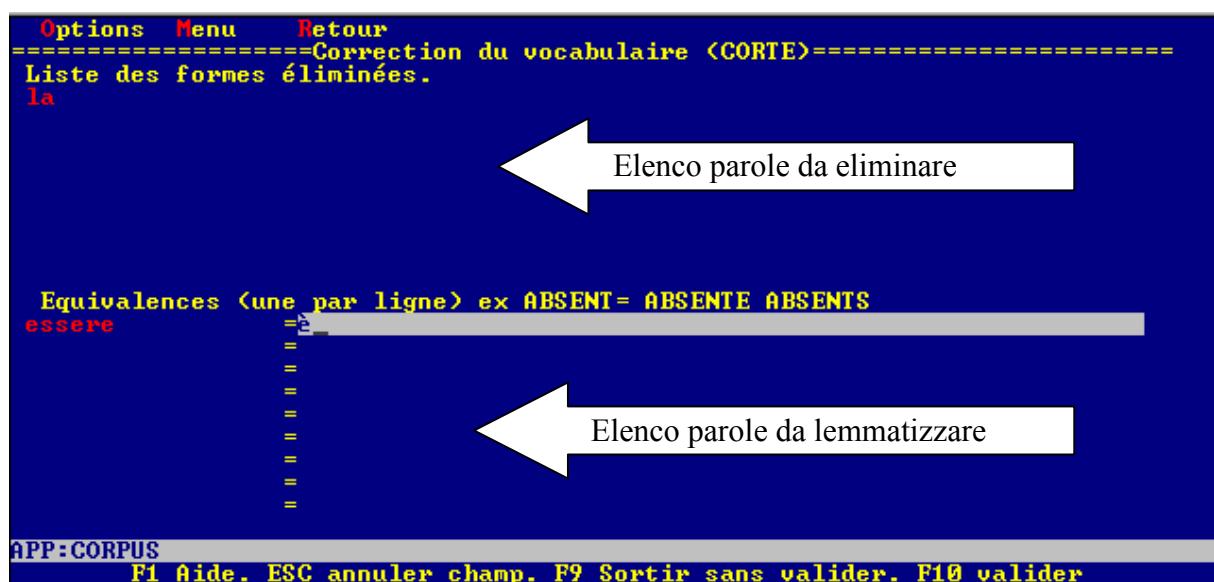
Procedura CORTE

A questo punto, qualora fosse necessario correggere il vocabolario (elenco delle parole distinte) occorrerà avviare la procedura **CORTE**. Attraverso la **CORTE** si possono così effettuare sul testo delle operazioni di lemmatizzazione¹⁵ semplicemente facendo equivalere una forma grafica al proprio *lemma*.

Al menù della **CORTE** si accede scegliendo “Segment-corretion-seuils” dal menù “Commandes” ed aprodo “Menu” e di nuovo “Correction Corte” dalla schermata che compare (fare sempre riferimento alla barra delle istruzioni in Figura 4 per utilizzare la giusta combinazione di tasti che permette l’apertura delle finestre).

In Figura 7 è riportata la schermata delle procedure in cui si può osservare che lo Spad-T® mette a disposizione dell’utente un editor DOS® in cui si possono inserire le forme grafiche da eliminare e quelle da lemmatizzare.

Figura 7: Schermata della procedura **CORTE**



L’elenco delle forme eliminate/lemmatizzate verrà salvato in un file dopo la pressione del tasto F10 e la scelta , per esso, di un nome. In questo esempio si è lasciato il nome che lo

¹⁵ La lemmatizzazione è un procedimento in base al quale una forma grafica viene “riletta”, successivamente, come una specifica flessione di un *lemma*. Ove per *lemma* s’intende la *forma canonica* con cui una parola è presente in un dizionario della lingua (come entrata di una voce: l’infinito per i verbi, il singolare per i sostantivi, il singolare-maschile per gli aggettivi.) Ad esempio le occorrenze <scrive> e <scrivevano> individuano due forme grafiche distinte, ovvero due *flessioni*, appartenenti ad uno stesso lemma: il verbo <scrivere>.

Spad-T® mette di *default* (corte). Il file avrà estensione “.pat” è potrà essere editato in un secondo momento usando, magari, un editor più funzionale¹⁶ rispetto a quello dello Spad-T®.

Ecco come compare il file corpus.pat dell'esempio:

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

PROC CORTE
=====
LEDIT=1
DELET la
EQUIV è essere
FIN
STOP
```

Nell'editare il file bisogna ricordarsi che il primo valore o forma del parametro **EQUIV** è quello in cui verranno cambiati tutti i valori o parole successive contenute nella riga. Ricordarsi sempre di non inserire l'elenco in modo tale da andare oltre la colonna 80. In caso di elenchi lunghi porre in colonna 79 il simbolo “>” e continuare sulla riga successiva.

Gran parte dei parametri che compaiono nel file corte.pat già sono stato spiegati per il file corpus.pat (Cfr. pag. 16).

I nuovi parametri sono:

NSPC: il nuovo file di testo numerizzato (individui x forme).

LEDIT: edizione di un vocabolario [1 o NON = no, 1 o NSPC = edizione di un nuovo vocabolario, 3 o TOT = edizione di un nuovo e del vecchio vocabolario. Valore di *default* 0].

DELET: comando di eliminazione [Lista delle forme].

EQUIV: comando di equivalenza [lista delle forme].

¹⁶ Usando editor per Windows® bisogna fare attenzione alla differente tabella codice che questo utilizza rispetto allo Spad-T® [Cfr. il capitolo “File di dati (*corpus*)” a pag. 11].

Per mandare in esecuzione la procedura occorre ritornare alla schermata iniziale (“Retour”) e mandare in esecuzione “Exécuter” dal menù “Exécuter”.

I risultati saranno contenuti nel file corte.lst.

Riportiamo un passo del file di testo dei risultati relativo alla tabella di frequenza delle parole del vocabolario corretto attraverso la **CORTE**:

	EFFECTIFS	(CODES)	FORMES LEXICALES
1	5	(1)	Questa
2	5	(2)	del
3	2	(3)	due
4	2	(4)	prima
5	5	(5)	riga
6	2	(6)	seconda
7	1	(7)	terza
8	5	(8)	testo
9	3	(9)	uno
10	5	(10)	essere

Procedura SETEX

La procedura **SETEX** consente di ridurre la soglia di frequenza delle parole. Le forme grafiche che ricorrono più frequentemente del valore di soglia impostato saranno conservate. Non solo: la **SETEX** consente anche di impostare la lunghezza minima delle parole e delle risposte al di sopra della quale le parole o le risposte verranno considerate ai fini dell’analisi.

Per avviare la procedura occorre scegliere il menù “Segment-correction-seuils” dal menù “Commandes” della finestra principale e successivamente selezionare “SETEX1” da “Menu” (fare sempre riferimento alla barra delle istruzioni in Figura 4 per utilizzare la giusta combinazione di tasti che permette l’apertura delle finestre).

Di seguito viene riportata la schermata della **SETEX** in cui si possono impostare i valori appena visti per la frequenza e per la lunghezza minima di parole (mots) e risposte (reponses).

Viene altresì proposta la scelta di poter effettuare l’operazione sul file di Base (Il file di dati di partenza) oppure su quello corretto (Corrigès) dalla procedura **CORTE**.

Figura 8: Procedura SETEX1

```

Options   Menu   Retour
=====Réductions des seuils de fréquence des mots (SETEX)=====
Mots utilisés(1=Base, 2=Corrigés) :2
Nouveau seuil de fréquence des mots :3
Nouvelle longueur minimale des mots retenus :0
Nouvelle longueur minimale des réponses retenues :0

APP:CORPUS
F1 Aide. ESC annuler champ. F9 Sortir sans valider. F10 valider

```

Il file setexm.pat¹⁷, creato con la validazione dei valori impostati (tasto F10) genera il seguente listato:

```

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

PROC SETEX
=====
NSEU =3 NMOMI=0 NREMI=0 LEDIT =NEW

```

I nuovi parametri sono:

NSPB: il nuovo file di testo numerizzato (individui x forme).

NSEU: nuova soglia di frequenza [Valori >0. Valore di *default* soglia precedente].

NREMI: valore di soglia della lunghezza delle risposte [Valori ≥ 0 . Valore di *default* soglia precedente].

NMOMI: valore di soglia della lunghezza delle risposte [Valori da 0 a 20. Valore di *default* soglia precedente]¹⁸.

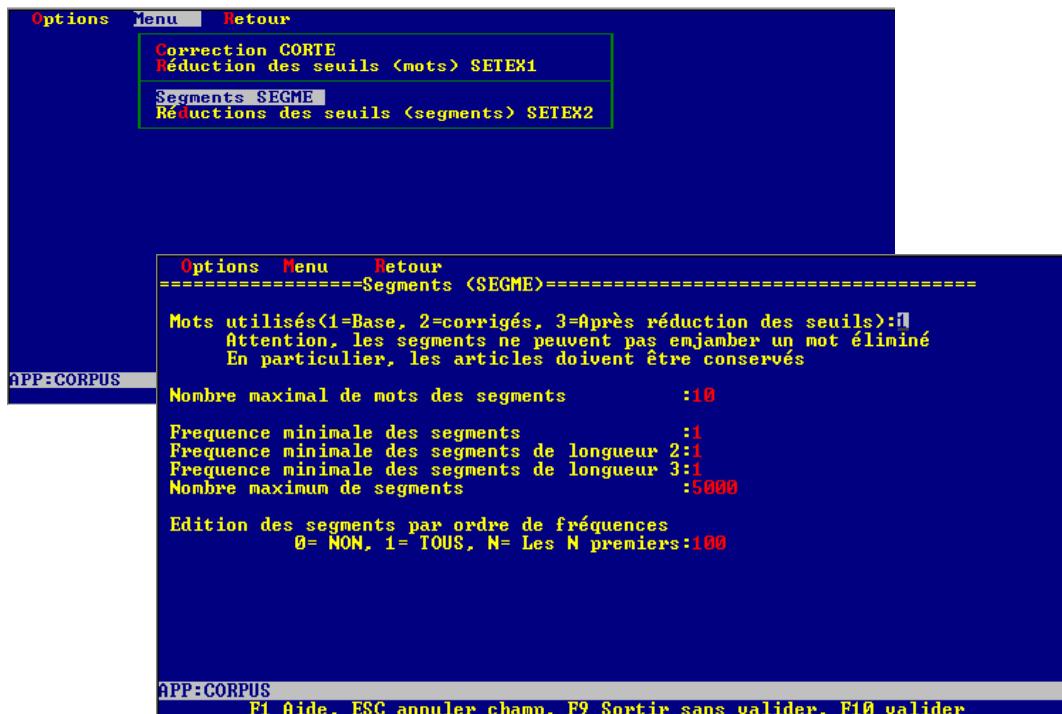
¹⁷ Lo Spad-T dà la possibilità di poter scegliere un nome da dare al file. Di *default* viene indicato il nome setexm.

Procedura SEGME

La procedura **SEGME** parte dal file di dati testuali per generare i *segmenti 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 suo *frammento*¹⁹.

Per accedere alla procedura occorre scegliere il menù “Segment-correction-seuils” dal menù “Commandes” della finestra principale e successivamente selezionare “SEGME” da “Menu” (fare sempre riferimento alla barra delle istruzioni in Figura 4 per utilizzare la giusta combinazione di tasti che permette l’apertura delle finestre).

Figura 9: Selezione della proc. SEGME ed opzioni



La schermata che appare dà la possibilità di selezionare diversi parametri:

- il file sul quale si vuole applicare la SEGME (su quello di base, su quello corretto dalla CORTE, oppure su quello trattato dalla SETEX1);
- il numero massimo delle parole contenute nel *segmento*:

¹⁸ Bisogna fare attenzione perché il parametro **NREMI** ha effetto sul testo corretto e dopo l’applicazione di **NSEU** e **NMOMI**. Quindi il valore di **NREMI** è in funzione del numero di parole del nuovo testo numerizzato e corretto.

- la frequenza minima di un *segmento*;
- la frequenza minima di un *segmento* di lunghezza due;
- la frequenza minima di un *segmento* di lunghezza tre;
- il numero massimo dei *segmenti*;
- la scelta del numero di *segmenti* da visualizzare nella tabella di frequenza.

Il file segme.pat²⁰, creato con la validazione dei valori impostati (tasto F10) genera il seguente listato:

```

LISTP=OUI , LISTF=OUI , NXLPA=59 , LRESM=NON , NAPEL=0 , MEM=3000000
NSPA = 'CORPUS . SPA'
NSME = 'CORPUS . SME'
PROC SEGME
=====
NXLON =10 NSEUG =1 NSEU2 =1 NSEU3 =1 NXSEG =5000 LEDIT=100 NXLE2 =200
STOP

```

I parametri nuovi sono:

NXLON: Lunghezza massima di un segmento in numero di forme [Valori da 2 ad N. Valore di *default* 10].

NSEUG: Soglia di frequenza (generale) dei segmenti ripetuti [Valori > 0. Valore di *default* 1].

NSEU2 : Soglia di frequenza dei segmenti ripetuti di lunghezza 2 [Valori > **NSEUG**. Valore di *default* **NSEUG**].

NSEU3 : Soglia di frequenza dei segmenti ripetuti di lunghezza 3 [Valori > **NSEUG**. Valore di *default* **NSEUG**].

NXSEG : Numero massimo di segmenti ripetuti considerati [Valori >0. Valore di *default* 10]

LEDIT : Lista dei segmenti ripetuti per ordine di frequenza [Valori > 0; 0 o NON = nessuna; 1 o TOT = Lista completa per ordine di frequenza. Valore di *default* 100]

NXLE2 : Numero massimo di segmenti di lunghezza 2 [Valori >0. Valore di *default* a calcolo].

NSME: il nuovo file di testo numerizzato (individui x forme).

¹⁹ 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 di>, oltre che al segmento <il governo si propone di>.

²⁰ Lo Spad-T dà la possibilità di poter scegliere un nome da dare al file. Di *default* viene indicato il nome segme.

Di seguito un estratto dal file dei risultati con estensione “.lst” al quale è stato dato il nome “segme”.

```
[...]  
-----  
SEG FREQ LONG TEXTE DU SEGMENT  
-----  
----- Questa  
1 5 3 Questa è la  
2 2 7 Questa è la prima riga del testo  
3 2 7 Questa è la seconda riga del testo  
-----riga  
4 5 3 riga del testo  
5 2 4 riga del testo due  
6 3 4 riga del testo uno  
[...]  
+-----+  
! SEGMENTS REPETES PAR ORDRE DE FREQUENCE !  
+-----+  
+-----+  
! FREQUENCE TEXTE DU SEGMENT !  
+-----+  
5 4-rida del testo  
5 1-Questa è la  
3 6-rida del testo uno  
2 5-rida del testo due  
2 3-Questa è la seconda riga del testo  
2 2-Questa è la prima riga del testo  
[...]
```

Si può decidere di applicare la **SETEX** anche ai segmenti ripetuti. Per far questo occorrerà entrare di nuovo nel menù “Segment-correction-seuils” per poi scegliere da “Menu” l’opzione “SETEX2”.

La SETEX2 consente di selezionare la nuova soglia di frequenza dei segmenti e la lunghezza minima delle forme contenute in una risposta.

Procedura MOCAR RECAR

Le procedure **MOCAR RECAR** forniscono delle misure di *specificità*, ove per *specificità* si intende indicare se e quanto una parola, od un frammento, sia tipico o specifico di un sub-testo, nell’ambito di uno stesso *corpus*.

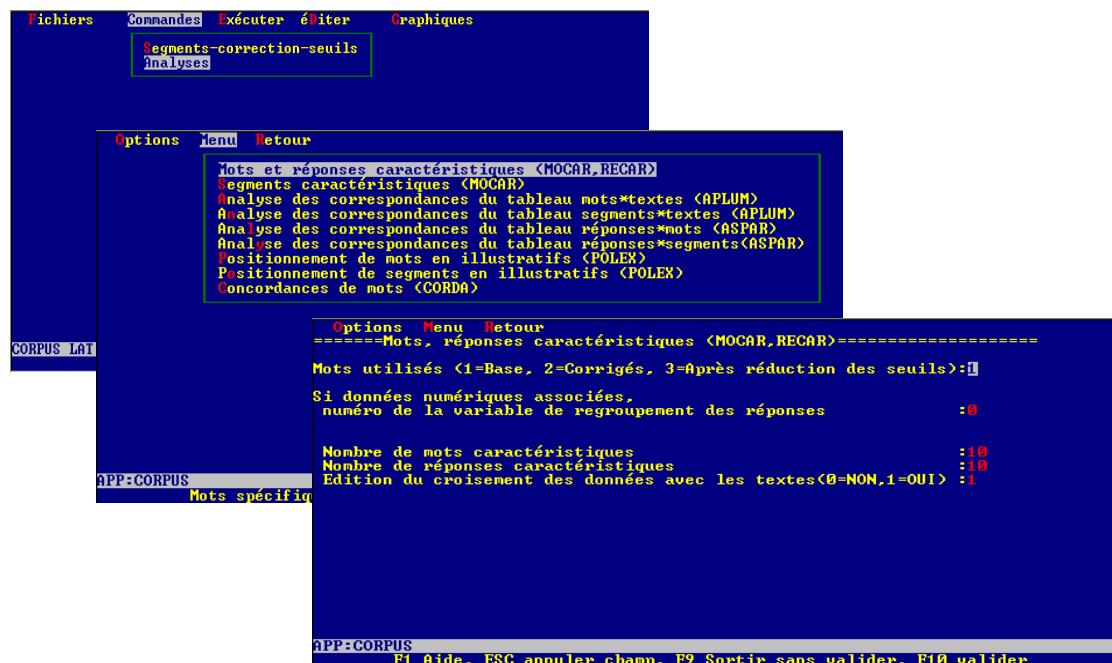
Alla base della misura di *specificità* c'è il concetto di assorbimento: una forma/frammento è caratteristica di un sub-testo quanto più il sub-testo è capace di assorbire (attrarre) la frequenza totale della forma/frammento nell'intero *corpus*. D'altra parte anche l'assenza di una forma/frammento in un sub-testo rispetto all'intero *corpus* è indice di un suo sotto utilizzo significativo.

Lo Spad-T® arriva ad una misura di specificità (*valore-test*²¹) attraverso dei test statistici partendo da ipotesi circa la natura distributiva delle parole all'interno del *corpus*.

In particolare, la procedura **MOCAR** genera le parole caratteristiche dei testi contenuti nel *corpus*, mentre la **RECAR** genera le risposte caratteristiche.

Si accede all'impostazione dei parametri della **MOCAR RECAR** attraverso il menù “Analyses” scegliendo poi “MOCAR RECAR” dalle opzioni di “Menu”.

Figura 10: Procedura MOCAR RECAR



Le opzioni consentono di impostare i parametri riguardo:

- il file sul quale si vuole applicare la MOCAR RECAR (su quello di base, su quello corretto dalla CORTE, oppure su quello trattato dalla SETEX1);

²¹ Se il *valore-test* è superiore a 2 in valore assoluto la presenza della forma/frammento è significativamente diversa da quella attesa (Bolasco, 1999: 227).

- il numero della variabile di raggruppamento delle risposte (solo nel caso di dati numerici associati);
- il numero delle parole caratteristiche;
- il numero delle risposte caratteristiche;
- la possibilità di avere nei risultati una tabella che incrocia i dati con i testi.

Validati i parametri con il tasto F10, in un file con nome a scelta (nell'esempio mocarm.pat), ecco come appare il listato del file stesso con estensione ".pat" (nell'esempio mocarm.pat)

```

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

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

PROC MOCAR
:=====M C=====
NOMOT=10 NOREP=10

PROC RECAR
:=====R l C=====
NOREP=10
STOP
:=====

```

I nuovi parametri sono:

NTEXM: dati testuali utili.

PROC MOTEX: procedura **MOTEX** che crea la tabella di contingenza parole x testi.

NVSEL: numero della variabile di raggruppamento delle risposte [Valori >0 o <0 a seconda dell'archivio. Valore 0 se non vi sono dati associati. Valore di *default* 0].

PROC MOCAR: procedura **MOCAR** per la generazione delle parole caratteristiche.

NOMOT: numero delle parole caratteristiche visualizzate per ciascun testo [Valori >0. Valore di *default* 10].

NOREP: numero delle risposte (individui Cfr. Figura 2 e Figura 3) caratteristiche visualizzate per ciascun testo [Valori >0. Valore di *default* 10].

Di seguito si riportano alcuni estratti del file dei risultati con estensione “.lst”. Nell'esempio mocarm.lst (il nome del file può essere scelto a piacere nel momento della validazione dei parametri).

GROUPEMENT DES REPONSES EN 2 TEXTES UTILISANT LA VARIABLE PRIVILEGIEE			
→Ripartizione degli individui e delle risposte nei due testi			
NUMERO DU TEXTE	IDENTIFICATEUR	NOMBRE D'INDIVIDUS	NOMBRE DE REPONSES
1	Test o1	2	2
2	Test o2	1	1
TOTAL		3	3

REPARTITION DES TERMES DANS LES TEXTES →Ripartizione dei termini all'interni dei testi (tabella generale)								
NUMERO DU TEXTE	IDENTIFICATEUR	Media			Media			* NOMBRE * DE MOTS *
		* MOTS	TOTAL	REPONSE	* DISTINCTS	TEXTE	* RETENUS *	
1 = Test o1		* 24	6000.0	12.0	* 10	416.7	* 24 *	
2 = Test o2		* 16	4000.0	16.0	* 9	562.5	* 16 *	
GLOBAL		* 40	10000.0	13.3	*		* 40 *	

TABLEAU DES NOMBRES DE MOTS DANS LES TEXTES →Ripartizione delle parole all'interno dei testi (tabella particolare)		
Test Test		
Questa	I	3.
del	I	3.
due	I	0.
la	I	3.
prima	I	1.
riga	I	3.
seconda	I	1.
terza	I	1.
testo	I	3.
uno	I	3.
è	I	3.
Test Test		

SELECTION DES FORMES LEXICALES CARACTERISTIQUES → Forme lessicali caratteristiche

LIBELLE DE LA FORME GRAPHIQUE		-- POURCENTAGE --		FREQUENCE		V.TEST	PROBA
		INTERNE	GLOBAL	INTERNE	GLOBALE		
1	uno	12.50	7.50	3.	3.	.824	.205
1	due	.00	5.00	0.	2.	-1.020	.154
TEXTE NUMERO 2 Test = o2							
LIBELLE DE LA FORME GRAPHIQUE		-- POURCENTAGE --		FREQUENCE		V.TEST	PROBA
		INTERNE	GLOBAL	INTERNE	GLOBALE		
1	due	12.50	5.00	2.	2.	1.020	.154
1	uno	.00	7.50	0.	3.	-.824	.205

SELECTION DES INDIVIDUS OU REPONSES CARACTERISTIQUES (CRITERE DU CHI-2)

→ Risposte caratteristiche (criterio del chi-quadro)

TEXTE NUMERO 1 Test = o1	
CRITERE DE CLASSIFICATION	REPONSE OU INDIVIDU CARACTERISTIQUE
.167 --	1 Questa è la prima riga del testo uno
--	1 Questa è la seconda riga del testo uno
.267 --	2 Questa è la terza riga del testo uno
TEXTE NUMERO 2 Test = o2	
CRITERE DE CLASSIFICATION	REPONSE OU INDIVIDU CARACTERISTIQUE
.000 --	1 Questa è la prima riga del testo due
--	1 Questa è la seconda riga del testo due
=====	
FIN DE LA PROCEDURE ** RECAR **	
=====R l c=====	

APPENDICE A

- 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	O
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:	€	□	/	f	„	…	†	‡	^	„	„	<	Œ	□	□	□
90:	□	„	„	„	„	„	—	—	—	—	—	—	—	—	—	—
A0:	;	¢	£	¤	¥	¦	§	¨	®	™	«	¬	—	®	—	—
B0:	°	±	²	³	‘	µ	¶	·	¸	¹	»	¾	¾	¾	¾	¾
C0:	À	Á	Â	Ã	Ä	Å	È	É	Ê	Ë	Ì	Í	Î	Ï	Ï	Ï
D0:	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	Þ
E0:	à	á	â	ã	ä	å	è	é	ê	ë	ì	í	î	ï	í	í
F0:	ë	ñ	ò	ó	ô	õ	ö	÷	ø	ú	û	ü	ý	þ	ÿ	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:	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
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:	ç	ü	é	â	ä	à	â	â	ç	ê	ë	è	í	í	í	í
90:	É	æ	Æ	ô	ö	ò	ú	ù	ÿ	Ö	Ü	ø	£	ø	×	f
A0:	á	í	ó	ú	ñ	Ñ	—	—	—	—	—	—	—	—	—	—
B0:	„	„	„	„	„	„	—	—	—	—	—	—	—	—	—	—
C0:	ł	ł	ł	ł	ł	ł	—	—	—	—	—	—	—	—	—	—
D0:	ð	ð	ð	ð	ð	ð	—	—	—	—	—	—	—	—	—	—
E0:	ó	ó	ó	ó	ó	ó	—	—	—	—	—	—	—	—	—	—
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
0	①	②	♥	♦	♣	♠	ؐ	ؑ	ؒ	ؓ	ؔ	ؕ	ؖ	ؗ	ؘ
1	▶	◀	↑	↓	↶	↷	↑	↓	→	←	↶	↷	↶	↷	?
2	!	"	#	\$	%	&	,	<	>	*	+	,	=	,	/
3	Ø	1	2	3	4	5	6	7	8	9	:	;	<	>	?
4	€	A	B	C	D	E	F	G	H	I	J	K	L	M	N
5	P	Q	R	S	T	U	W	X	Y	Z	[]	\	^	-
6	,	a	b	c	d	e	f	g	h	i	j	k	l	m	n
7	p	q	r	s	t	u	v	w	x	y	z	ؐ	ؑ	ؒ	ؓ
8	ç	ü	é	â	ä	å	ç	ê	ë	è	ï	î	î	å	æ
9	É	æ	ß	ô	ö	ö	ö	ö	ö	ö	ç	ç	ç	ø	ƒ
A	á	í	ó	ú	ñ	ñ	ñ	ñ	ñ	ñ	ñ	ñ	ñ	ñ	ñ
B	ß	ß	ß	ß	ß	ß	ß	ß	ß	ß	ß	ß	ß	ß	ß
C	ł	ł	ł	ł	ł	ł	ł	ł	ł	ł	ł	ł	ł	ł	ł
D	ñ	ñ	ñ	ñ	ñ	ñ	ñ	ñ	ñ	ñ	ñ	ñ	ñ	ñ	ñ
E	œ	œ	œ	œ	œ	œ	œ	œ	œ	œ	œ	œ	œ	œ	œ
F	߱	߲	߳	ߴ	ߵ	߶	߷	߸	߹	߻	߻	߻	߻	߻	߻

- **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

BIBLIOGRAFIA

- Aureli E.- Gennari A. (1996) *Guida all'uso dello SPAD•T*, Dip.to di Statistica
Probabilità e Statistiche Applicate. Università degli studi
di Roma “La Sapienza”, Serie D- Materiale Informatico
- Lebart L. Salem A. (1988) *Analyse statisque des données textuelles*, Dunod, Paris
- Bolasco S. (1999) *Analisi Multidimensionale dei dati*, Carocci, Roma
- Microsoft® MS-DOS® (1991) *Manuale dell'utente Sistema operativo versione 5.0*,
Microsoft, Ireland
- Lebart L. Morineau A. Bécue M., Haeusler L. (1993) *Introduction à SPAD-T intégré Version 1.5 P.C*, CISIA,
Saint – Mandé (France)
- Scuola della Società Italiana di Statistica sui metodi per l’analisi dei dati testuali (2000) *Dispense del corso S.I.S.* tenuto a Roma il 3-7 luglio del 2000