

Progetto CLIPS

Corpora e Lessici di Italiano Parlato e Scritto

W0- a1

(Coordinamento Generale del Progetto)

Title: *Criteria per la digitalizzazione del materiale audio CLIPS*

Document No: CLIPS/W0-a1/CGP/005

Document Name: NADL005

Status: **pubblico**

Date: 10/3/2006

Authors: *Francesco Cutugno*

CRITERI PER LA DIGITALIZZAZIONE DEL MATERIALE AUDIO CLIPS

0. Premessa:

Questo documento è destinato a persone che siano a conoscenza delle procedure di configurazione di schede di ingresso/uscita digitale dei suoni in ambiente Windows e che abbiano una pratica anche minima con procedure di campionamento e digitalizzazione di segnali audio. Non verranno quindi fornite le informazioni di base necessarie alla predisposizione preliminare dell'hardware e del software necessari.

Nell'ambito delle azioni che ricadono sotto il workpackage w2 è previsto il trasferimento su supporto informatico di tutto il materiale precedentemente acquisito nel workpackage w1. Qui di seguito si forniscono le indicazioni re sulle procedure relative al materiale dialogico incluse le letture delle tre liste (parole nelle mappe, parole nelle vignette del test delle differenze, liste di frasi lette) e al materiale radiotelevisivo.

1. Panorama delle attrezzature necessarie:

Come precedentemente indicato nel documento relativo alle specifiche tecniche, ogni sede dovrebbe essere dotata di una scheda audio marca Creative modello "Live Platinum" o superiori, dotata di pannello di collegamenti "Live drive II" con ingressi per fibre ottiche.


I registratori Sony TCD100 e TCD6 che in precedenza sono stati suggeriti come standard per l'acquisizione del parlato, si collegano al pannello di contatti della scheda sonora attraverso il sistema di connessione a fibre ottiche POC-DA12P(-MP) (-SP) fornito di serie con il TCD6 e da acquistare separatamente se si è in possesso del TCD100 .

Per il trasferimento dei segnali su disco rigido e la loro trasformazione nel formato definito per CLIPS può essere utilizzato qualsiasi programma di registrazione digitale dei suoni che sia in grado di individuare, fra i vari canali di ingresso della scheda audio, quello denominato SPDIFin/Aux2, sono ad esempio accettabili: 1) il registratore di suoni tipicamente fornito con il sistema operativo, 2) il programma di registrazione fornito "in bundle" con la scheda Sound Blaster, 3) alcuni shareware come CoolEdit o **Goldwave**.

Proprio l'ultimo dei programmi sopra elencati sarà utilizzato per esemplificare la procedura, se non si ha a disposizione uno specifico prodotto se ne consiglia fortemente l'impiego¹.

2. Dialoghi map task e test delle differenze

2.1 Acquisizione

In Goldwave, prima della acquisizione del segnale, attraverso il comando **new**²  è necessario predisporre una finestra di destinazione del segnale acquisito, le caratteristiche di questa finestra devono prevedere:

file di destinazione in formato stereo, 16 bit signed, campionamento a 22050 Hz, durata fissata intorno ad un valore leggermente maggiore di quello della durata effettiva della porzione di dialogo da acquisire (recuperabile dalle schede che accompagnano le registrazioni).

Una volta acquisita la traccia stereo sono consigliate le due seguenti operazioni:

1) eliminazione delle code iniziali e finali di silenzio: attraverso l'uso dei tasti destro e sinistro del mouse selezionare la porzione di segnale di interesse riducendo al minimo i silenzi alle estremità.

Successivamente cliccare sul tasto **trim** 

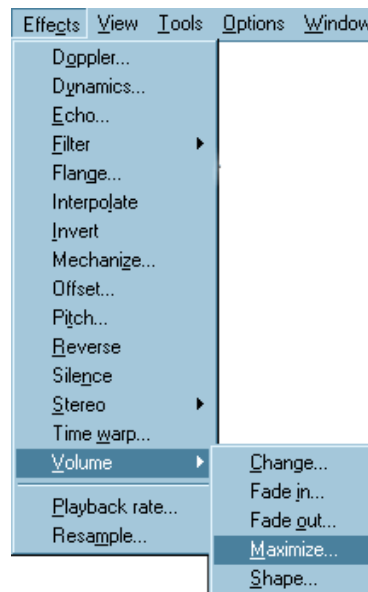
2) controllo della dinamica: le schermate in cui è mostrato il segnale in Goldwave riportano in ordinata l'ampiezza del segnale normalizzata rispetto ai livelli di quantizzazione. L'asse normalmente va da -1 a 1. La dinamica non deve essere saturata durante l'acquisizione, cioè il grafico del segnale non deve se non per pochissimi istanti (ma sarebbe meglio se non succedesse mai) superare gli estremi del grafico pena l'introduzione di una distorsione, dannosa per i successivi livelli di indagine e, più in generale, per la qualità acustica del segnale. Allo stesso tempo il segnale non deve apparire troppo debole: nel caso in cui ciò accada si deve riacquisire il segnale intervenendo per quanto possibile sui livelli di registrazione. La condizione ideale è quella in cui il grafico del segnale, in media, copre una fascia di ampiezze che si aggiri fra -0.6/-0.7 e 0.6/0.7.

¹ Goldwave è un potente programma di acquisizione ed elaborazione dei segnali audio. E' scaricabile gratuitamente in una versione completamente utilizzabile dal sito www.golwave.com La sola limitazione, che viene rimossa all'atto della registrazione del programma unitamente al pagamento di 30\$ circa, è il numero operazioni (in corrispondenza di 150 pressioni dei tasti del mouse) che possono essere fatte.

Fra i programmi di uso più professionale e che, ovviamente, costano decisamente di più, consigliamo l'impiego del "Sound Forge".

² Le icone che mostreremo in questo documento sono normalmente disponibili all'avvio di Goldwave sulla barra degli strumenti in alto. Tuttavia qualche icona può non essere presente sullo schermo, per aggiungerla selezionare la voce di menu "options-> toolbar" e seguire le istruzioni per la personalizzazione della barra strumenti.

Prima di salvare il segnale acquisito secondo le procedure sopra descritte e procedere alle acquisizioni successive, si consiglia di operare una normalizzazione a +/- 1 del segnale attraverso il comando indicato nella voce di menu **effects->volume->maximize**:



2.2 *Suddivisione in turni*


Il segnale acquisito e salvato (cfr. più avanti per il problema dei nomi dei file) dovrà essere trascritto ortograficamente secondo norme che saranno oggetto di specifica documentazione. Alla fine di questo lavoro sarà disponibile la suddivisione in turni di tutto il materiale selezionato per il progetto.


Il materiale selezionato per l'analisi fonetica dovrà dunque essere salvato in singoli file di segnale ciascuno corrispondente con precisione ad un singolo turno del dialogo.

Sempre con riferimento alle procedure possibili in Goldwave, diamo di seguito le indicazioni per la generazione di detti file:

- 1) si apra il file stereo contenente l'intero dialogo da cui estrarre i turni;
- 2) tramite i tasti destro e sinistro del mouse si selezionino con precisione le porzioni da trasferire in nuovi file corrispondenti a singoli turni. Bisogna ancora una volta fare attenzione a ridurre al minimo la presenza di silenzi in testa ed in coda ad ogni selezione;
- 3) si annoti separatamente la posizione relativa (in millisecondi) della selezione impiegando le informazioni disponibili nella barra di stato di Goldwave. In altre parole si deve preparare un

prospetto nel quale tutti i turni acquisiti devono essere catalogati indicando nome del file (cfr. più avanti) istante di inizio del turno all'interno del dialogo, istante di fine. Il prospetto può essere tenuto in un file (word, excel) aperto in parallelo a Goldwave o annotato a mano e trascritto successivamente.

4) selezione del canale: attraverso l'operatore **chan**  in Goldwave è possibile selezionare uno solo dei due canali stereofonici del dialogo. Eseguire il comando per selezionare fra i due parlanti quello che deve essere ricopiato nel file di singolo turno. A questo punto dovrebbe risultare

selezionato uno solo dei due segnali: cliccare su **copy**  e successivamente su **paste new**




per generare una nuova finestra contenente un segnale mono con le stesse caratteristiche tecniche dei precedenti e contenente il turno. Salvare questo nuovo file con il nome appropriato.

NOTA BENE: il dialogo di partenza non deve essere manomesso durante questa operazione. Si consiglia quindi di marcarlo come “sola lettura” durante queste operazioni. Non è strettamente necessario togliere in seguito questa opzione.

3. Liste di parole (nomi degli oggetti sulle mappe e nelle vignette del test differenze) e di frasi.

In fase di registrazione alcuni materiali di questo gruppo sono stati letti due volte, altri una sola. Inoltre in ogni lettura possono verificarsi delle disfluenze o errori che richiedono la ripetizione dell'item pronunciato. Il prodotto finale di progetto, però, in questo caso prevede una sola ripetizione di tutte le parole e frasi, tassativamente montate nell'ordine previsto dai documenti originali di progetto. Il materiale registrato deve quindi essere ascoltato, se la lista è stata letta due volte si parte dalla prima replica e la si integra eventualmente impiegando ripetizione di items provenienti dalla successiva replica. In caso di una sola ripetizione della lista, l'operatore che ha condotto le registrazioni dovrebbe avere avuto l'accortezza di far ripetere eventuali items letti in maniera non appropriata. Se possibile, vale a dire in tutti i casi in cui il disturbo non sia sovrapposto al segnale di interesse (ma speriamo che non sia accaduto mai) tutte le esitazioni, gli errori, i messaggi di tipo “ripeto”, i rumori di fondo, devono essere eliminati dal file finale. Il montaggio di un pezzo proveniente da un altro punto del segnale nella posizione finale può essere condotto, in Goldwave, selezionando il pezzo da spostare con i tasti destro e sinistro del mouse, cliccando su **cut**




la porzione così rimossa può essere “incollata” in una nuova posizione scelta cliccando con il tasto sinistro nel punto di inserimento e scegliendo il comando **paste**  Le porzioni da

eliminare possono essere selezionate con la solita procedura e cancellate con il comando **delete**



Pur non essendo prevista una spaziatura temporale regolare fra un item ed il successivo sarebbe auspicabile che, mediamente, non intercorra un silenzio di durata sensibilmente variabile fra una voce della lista e la successiva, a titolo indicativo si suggerisce di non lasciare trascorrere, in media, più (ma neanche meno) di due secondi. In Goldwave esistono comandi specifici per inserire

silenzi  per allungare un intervallo troppo corto, mentre per accorciarne uno più lungo la procedura è quella sopra indicata per la cancellazione di porzioni di segnale.

4. Parlato radiofonico e televisivo:

La porzione di corpus radiofonico e televisivo è stata acquisita direttamente sul supporto digitale. I file radiofonici dovrebbero essere direttamente disponibili in formato .wav, sarà dunque solo necessario osservare le seguenti indicazioni:

- 1) verificare che il formato finale sia .wav, mono, frequenza di campionamento 22050 Hz, 16 bit, signed. In caso contrario è possibile ricorrere agli strumenti disponibili all'interno di Goldwave per la conversione del segnale nel formato richiesto;
- 2) selezionare il materiale in modo che sia coerente per quantità e tipologie con le richieste di progetto. All'interno del materiale selezionato in maniera definitiva non vanno rimosse le porzioni non adatte (perché disturbate, per la presenza di musica di sottofondo, eccetera), in fase di trascrizione ortografica si procederà all'annotazione dei fenomeni.

Il materiale di provenienza televisiva è stato salvato in formato .AVI. Goldwave (nella sua versione più recente, la 4.21) apre direttamente i file di questo tipo visualizzando il segnale audio ed ignorando il materiale video. Una volta ottenuto il segnale si procede come sopra indicato in relazione al materiale radiofonico. L'archivio dei file .AVI verrà conservato senza ulteriori elaborazioni e fornito a fine progetto insieme al corpus acustico. (NOTA valuteremo insieme se convertirlo in mpeg, dipende dalla qualità del segnale che è già povera così).

5. Norme per i nomi dei file:

In questo paragrafo vengono fornite indicazioni per i nomi dei file per i dialoghi interi, le letture di liste e per le macro porzioni di parlato radiofonico e televisivo:

I nomi dei file relativi ai singoli turni non possono essere definiti se non dopo la trascrizione ortografica, per queste specifiche si rimanda al documento per la trascrizione ortografica.

5.1 Legenda delle sigle:

prime due posizioni (XX):

DG = dialogico

RD = radiofonico;

TV = televisivo;

LF = lettura frasi;

LM = lettura liste oggetti nelle mappe;

LT = lettura liste oggetti nelle vignette del test differenze;

terza e quarta posizione (YY):

mt = map task;

td = test differenze;

dc = divulgazione e cultura;

it = intrattenimento

is = informazione e servizio

pb = pubblicità

oppure:

p1 = parlante uno (canale destro)

p2 = parlante due (canale sinistro),

quinta sesta e settima (ZZZ):

A = mappa e test differenze di tipo A

B = mappa e test differenze di tipo B

_ (underscore) = posizione vuota per il radiotelevisivo.

01...nn = numerazioni progressive per i dialoghi o per le porzioni di parlato radiotelevisivo

ottava posizione (T):

A = Asti;

B = Bari

C = Cagliari

D = Bergamo

E = Bologna

F = Firenze

G = Genova

H = Catanzaro

L = Lecce

M = Milano

N = Napoli

O = Perugia

P = Palermo

R = Roma

V = Venezia

5.2 Esempi

Un nome di file si presenterà quindi nella forma

XXYYZZZT.wav. quindi:

DGmtA01N.wav è il file stereo contenente l'intero dialogo map task numero 1 della mappa di tipo A registrato a Napoli.

RDdc_04V.wav è il file mono contenente la quarta subporzione di parlato radiofonico di tipo divulgazione e cultura registrato a Venezia;

LTp1B06P.wav è il file mono contenente la lettura delle parole corrispondenti agli oggetti nelle vignette del test differenze di tipo B, sesto dialogo, registrato a Palermo.

Schermata Goldwave

