

# METODOLOGIE E SISTEMI PER L'ANNOTAZIONE LINGUISTICA

**Piero Cosi**

IFD-CNR

Istituto di Fonetica e Dialettologia – Consiglio Nazionale delle Ricerche

e-mail: [cosi@csrf.pd.cnr.it](mailto:cosi@csrf.pd.cnr.it)

www: <http://www.csrf.pd.cnr.it>

<http://nts.csrf.pd.cnr.it/Ifd>

## 1. SOMMARIO

Senza avere la pretesa di coprire esaustivamente tutte le possibili alternative, vengono descritte le principali metodologie e le principali architetture software per l'annotazione linguistica di corpora vocali. Più che un confronto comparativo, in questo lavoro introduttivo, si vogliono solamente elencare le informazioni necessarie per una semplicistica descrizione dei vari sistemi e per un loro facile reperimento nella rete.

## 2. INTRODUZIONE

Con il termine "Annotazione Linguistica" ci si riferisce ad ogni notazione descrittiva o analitica che possa essere applicata a dati linguistici raccolti sotto forma di dati testuali o di dati espressi come funzioni temporali (registrazioni audio, video e/o di dati articolatori o fisiologici). Le notazioni possono includere le trascrizioni di ogni tipo (dalle caratteristiche fonetiche o prosodiche alle strutture della frase, del discorso o del dialogo), "part-of-speech" e altre specifiche annotazioni, analisi sintattiche, annotazioni incrociate con multi-riferimenti, e così via. Assieme alle metodologie, vengono illustrati i pregi e i difetti di alcuni fra i più comuni sistemi hardware/software per l'annotazione linguistica.

## 3. METODOLOGIE DI ANNOTAZIONE

Esistono essenzialmente due metodologie per l'annotazione di corpus di dati vocali: una procedura automatica, spesso poco utilizzata, ed una procedura manuale, largamente più diffusa.

### 3.1 PROCEDURA MANUALE

Al fine di eseguire l'etichettatura e/o la segmentazione manuale di corpora vocali, si deve poter disporre essenzialmente di un buon sistema di visualizzazione e di ascolto del segnale vocale in grado di fornire visivamente ed uditiveamente all'utilizzatore tutte le informazioni necessarie per questo scopo. In altre parole, per poter eseguire una corretta etichettatura, un qualsivoglia sistema software deve essere in grado di visualizzare la forma d'onda, lo spettrogramma, la curva dell'intensità, la curva della frequenza fondamentale ed eventualmente altre grandezze, oltre ovviamente a fornire un buon sistema di ascolto selettivo di porzioni qualsiasi del segnale vocale. In Figura 1 è illustrata visivamente una sessione standard di etichettatura del segnale comune a tutti i sistemi elencati di seguito in questo lavoro.

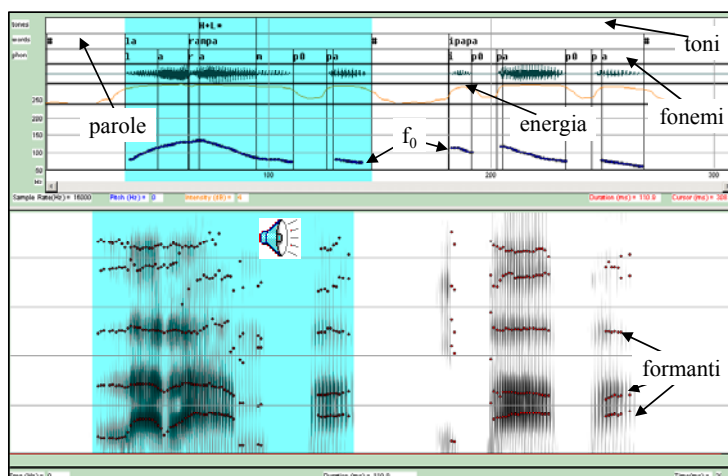


Figura 1. Illustrazione di una sessione standard di etichettatura del segnale.

### 3.1 PROCEDURA AUTOMATICA

Al fine di eseguire l'etichettatura e/o la segmentazione di corpora vocali in modo automatico, si deve poter disporre essenzialmente di un buon sistema di riconoscimento in grado di allineare la trascrizione fonetica del testo da trascrivere con il corrispondente segnale audio, come illustrato nello schema a blocchi di Figura 2. Qualora la realizzazione audio del testo da trascrivere non corrisponda perfettamente alla sua "corretta" trascrizione fonetica il sistema deve essere anche in grado di riconoscere eventuali cancellazioni, sostituzioni e introduzioni di fonemi non previsti.

Questi sistemi sono spesso disponibili solo ai laboratori specializzati sul riconoscimento automatico del segnale vocale ma si stanno sempre più diffondendo anche presso utilizzatori inesperti. Il loro utilizzo è consentito ovviamente solo se le percentuali di riconoscimento fonetico dei sistemi raggiungono valori elevati.

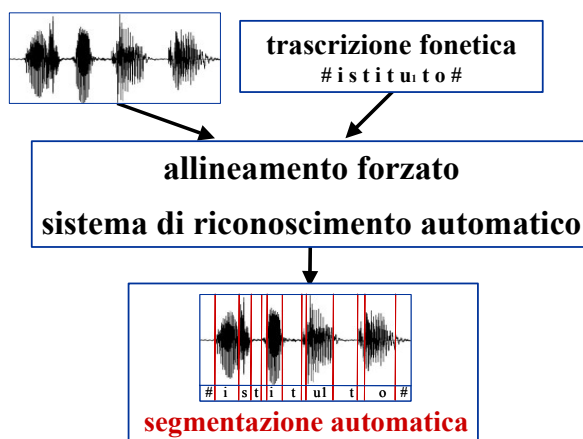


Figura 2. Schema a blocchi di un sistema di etichettatura automatica

#### 4. PRINCIPALI SISTEMI DI ANNOTAZIONE (Manuale)

Di seguito, senza avere la pretesa di coprire esaustivamente tutte le possibili alternative, sono elencati i principali sistemi software l'annotazione linguistica di copora vocali facilmente recuperabili dalla rete.

##### 4.1 Emu

EMU è una collezione di strumenti software, liberamente disponibile, per la creazione, la manipolazione e l'analisi di database vocali. Il cuore di EMU è rappresentato da un sistema di ricerca in un database di segnale vocale, definibile ed impostabile dall'utente, che consente di trovare vari segmenti vocali a seconda della loro struttura sequenziale e/o gerarchica all'interno della frase in cui occorrono. EMU include un sistema interattivo per l'etichettatura su vari livelli (parola, fonemi, toni...) che può visualizzare, come illustrato in Figura 3, oltre alla forma d'onda, anche spettrogrammi, formanti e andamenti di f0, e consente la creazione di etichette sequenziali e/o gerarchicamente organizzate.

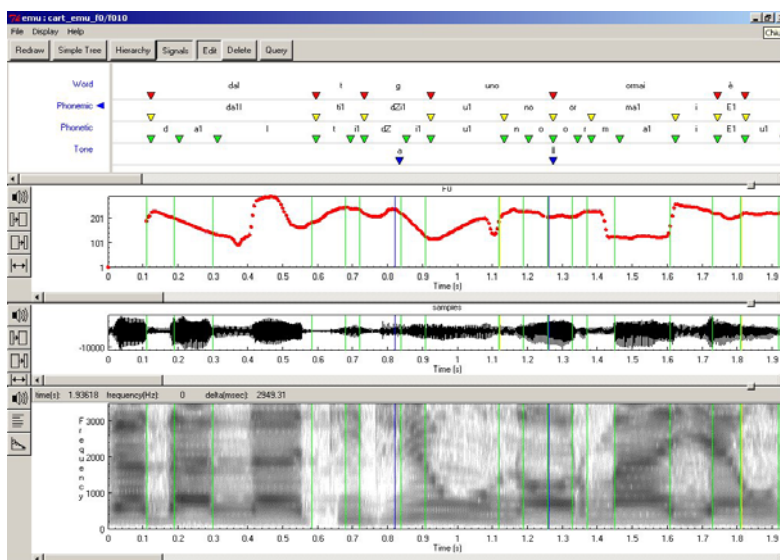


Figura 3. Semplice videata tipo di EMU

Emu - SourceForge free service to Open Source

<http://www.shlrc.mq.edu.au/emu/>

Steve Cassidy ([Steve.Cassidy@mq.edu.au](mailto:Steve.Cassidy@mq.edu.au))

SHLRC Speech Hearing and Language Research Centre, Macquarie University, Sydney, Australia

##### 4.2 WaveSurfer

WaveSurfer è uno strumento *Open Source* per la visualizzazione e la manipolazione del segnale. E' adatto a qualsiasi tipo di utenti specializzati o meno, infatti, ha un'interfaccia semplicissima e funzionale. Può essere usato in modalità stand-alone ma può anche essere integrato in altre applicazioni ed è estendibile mediante opportuni plug-ins. Consente l'etichettatura e la segmentazione su multilivelli. In Figura 4 è visibile una videata tipo di Wave Surfer.

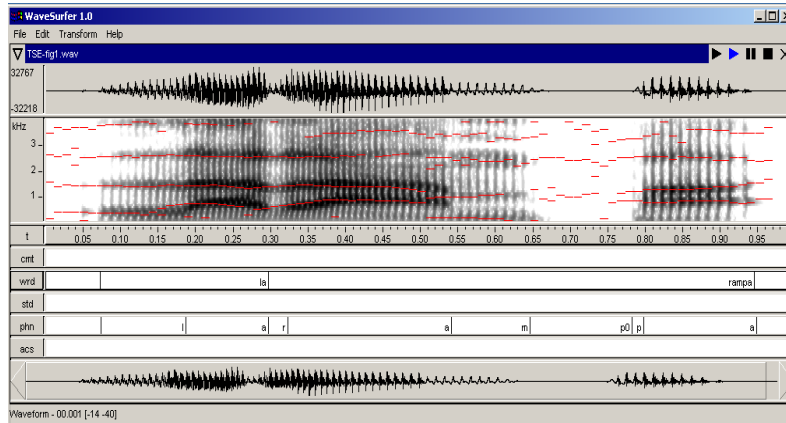


Figura 4. Semplice videata tipo di WaveSurfer

WaveSurfer: <http://www.speech.kth.se/wavesurfer/>  
 Jonas Beskow ([beskow@speech.kth.se](mailto:beskow@speech.kth.se)), Kåre Sjölander ([kare@speech.kth.se](mailto:kare@speech.kth.se))  
 Centre for Speech Technology (CTT) at KTH in Stockholm, Sweden

WaveSurfer per CLIPS: <http://augusto.cirass.unina.it/clips/>  
 Massimo Petrillo ([petrillo@cirass.unina.it](mailto:petrillo@cirass.unina.it)), Franco Cutugno ([cutugno@cirass.unina.it](mailto:cutugno@cirass.unina.it))  
 CIRASS

### 4.3 Gipos

Gipos come i sistemi precedenti è uno strumento per la visualizzazione e la manipolazione del segnale. E' adatto a qualsiasi tipo di utenti specializzati o meno, infatti, ha un'interfaccia semplicissima e funzionale come illustrato in Figura 5. Può essere usato in modalità stand-alone ma può anche essere utilizzato mediante batch script. Consente l'etichettatura e la segmentazione. GIPOS (versione 2.3) funziona come essenzialmente tutti i sistemi illustrati in questo lavoro, su Windows 9x e NT ed è disponibile liberamente su richiesta.

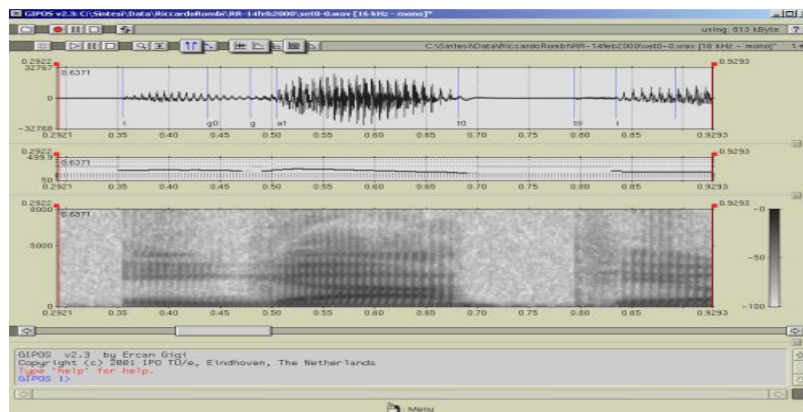


Figura 5. Semplice videata tipo di Gipos

Gipos: <http://www.ipo.tue.nl/ipo/gipos/>  
 L.L.M. Vogten ([L.L.M.Vogten@tue.nl](mailto:L.L.M.Vogten@tue.nl)) - Signal Processing Systems group, room EH 3.23  
 Eindhoven University of Technology - PO Box 513, NL 5600 MB Eindhoven, The Netherlands  
 Phone: +31 40 247 5261, secretary: +31 40 247 3288



#### 4.6 *SpeechAnalyzer*

SpeechAnalyzer è realizzato dal *Summer Institute for Linguistics* ed è utilizzabile solo in modalità stand-alone. E' possibile utilizzarlo assieme a Speech Manager, un programma per la gestione di database vocali, specificatamente progettato per consentire ai linguisti l'analisi di più linguaggi parlati. Consente l'etichettatura (anche IPA da tastiera) e la segmentazione su più livelli (vedi Figura 8).

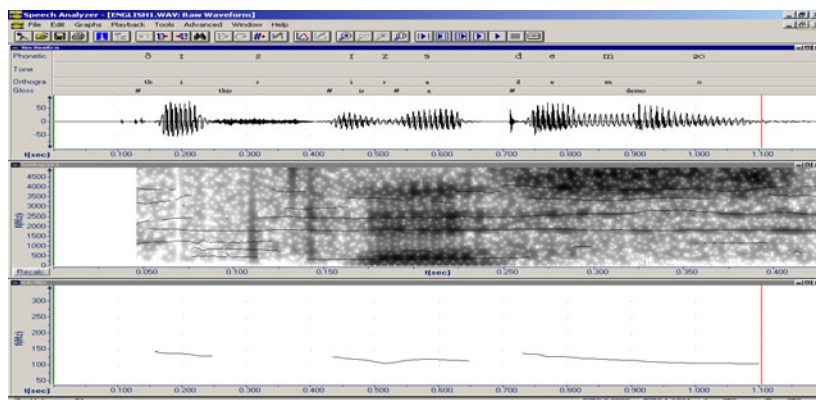


Figura 8. Semplice videata tipo di SpeechAnalyzer.

SpeechAnalyzer: <http://www.sil.org/computing/speechtools/speechanalyzer.htm>  
SIL International (Summer Institute of Linguistics): <http://www.sil.org/>

#### 4.7 *SpeechViewer*

SpeechViewer (Figura 9) è uno strumento Open Source integrato nel software Cslu Speech Toolkit per la visualizzazione e la manipolazione del segnale. Come altri sistemi è adatto a qualsiasi tipo di utenti specializzati o meno, infatti, ha un'interfaccia semplicissima e funzionale. E' disponibile il sorgente, quindi, oltre a poter essere usato in modalità stand-alone può anche essere integrato in altre applicazioni. Consente l'etichettatura e la segmentazione su multilivelli.

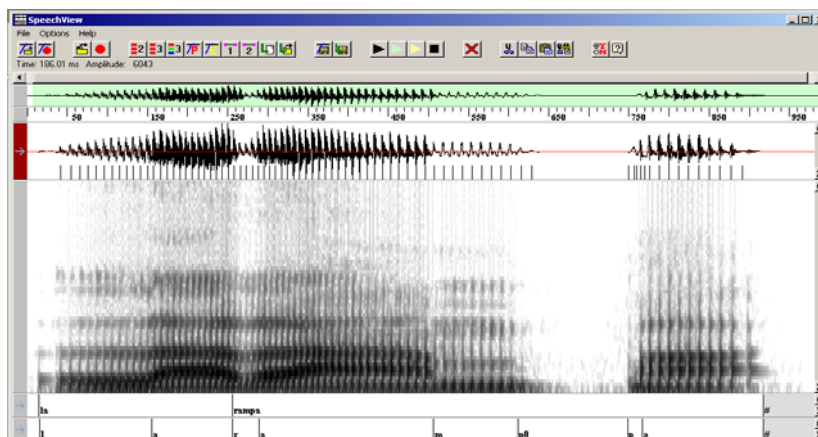


Figura 9. Semplice videata tipo di SpeechViewer.

SpeechViewer, Cslu Speech Toolkit: <http://cslu.cse.ogi.edu/>  
Centre for Spoken Language Understanding @ OGI, Portland OR, USA

#### 4.8 Praat

PRAAT è un potentissimo strumento per l'analisi, la sintesi, la visualizzazione e la manipolazione del segnale. E' ottimo soprattutto per l'analisi e la sintesi del segnale, ma è principalmente adatto ad utenti specializzati, infatti, l'interfaccia non è semplicissima. Non è disponibile il sorgente, quindi, è destinato ad essere utilizzato in modalità stand-alone, però si possono creare utilissimi script di lavoro per personalizzare ed automatizzare le procedure di analisi. Consente l'etichettatura e la segmentazione su multilivelli (Figura 10).

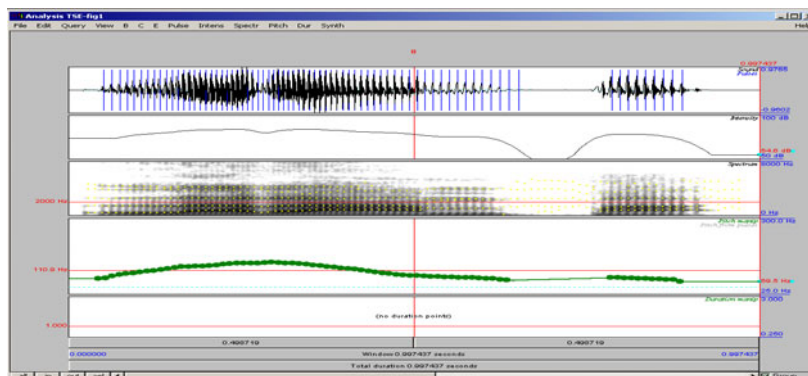


Figura 10. Semplice videata tipo di Praat.

Praat: <http://www.fon.hum.uva.nl/praat/>

Paul Boersma (Paul.Boersma@hum.uva.nl) Institute of Phonetic Sciences of the University of Amsterdam.  
<http://www.fon.hum.uva.nl/>

#### 4.9 SegWin e SegView

SegWin è uno software di ausilio alla costruzione della base di dati ed alla segmentazione del corpus AVIP. La sequenza di etichettatura è particolarmente rigida essendo le dipendenze dei livelli di etichettatura automaticamente determinate. Consente l'etichettatura e la segmentazione su multilivelli. PHN: "fonetico stretto", PHB: "fonetico largo" o "fonologico della varietà", WRD: ortografico, PHM: "citation form" (conversione automatica grafema-fonema), TON: "intonativo" (Figura 11). Attualmente è specifico al corpus AVIP ma sarebbe (?) adattabile anche ad altre basi di dati.

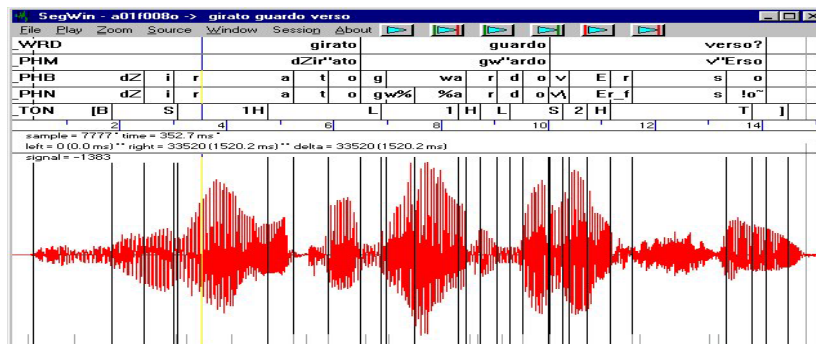


Figura 11. Semplice videata tipo di SegWin.

SegWin & SegView

M. Refice (refice@poliba.it), M. Savino, M. Altieri, R. Altieri e A. Enriquez - DEE - Politecnico di Bari

#### 4.10 Transcriber

Transcriber (Figura 12) è uno strumento distribuito come “free software”, su licenza GNU General Public License, che consente la creazione, l’organizzazione ed il mantenimento di corpora vocali. Transcriber è uno strumento rivolto principalmente alla trascrizione di registrazioni di lunga durata, come quelle, ad esempio, relative alle notizie radio televisive. Consente l’etichettatura di parlanti diversi e l’indicazione degli argomenti del dialogo. E’ stato progettato per essere il più semplice possibile ed è facilmente configurabile dall’utente. Transcriber è sviluppato in Tcl/Tk con estensioni in C++ e si basa sul software Snack/WaveSurfer (KTH).

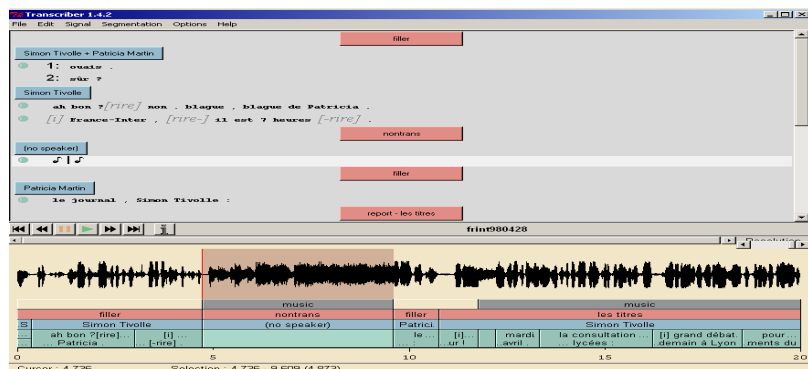


Figura 12. Semplice videata tipo di Transcriber.

Transcriber - Copyright (C) 1998-2000, DGA

<http://www.etc.fr/CTA/gip/Projets/Transcriber/> Claude Barras (Claude.Barras@limsi.fr) LIMSI - CNRSEdouard Geoffrois (Edouard.Geoffrois@etc.fr), Steve Bird (sb@unagi.cis.upenn.edu) & Zhibiao Wu, LDC

#### 4.11 MultiSpeech

Multi-Speech (Model 3700), è un programma di analisi del segnale *low-cost* per Windows®. Utilizza gli hardware multimediali standard (Sound Blaster™) per acquisire, analizzare e riprodurre il segnale. Validissimo per l’analisi, ma non altrettanto per l’etichettatura e la segmentazione.



Figura 12. Semplice videata tipo di Transcriber.

Kay Elemetrics

<http://www.kayelemetrics.com/multispe.htm>

#### 4.12 Scicon: Pc/MacQuirer, PitchWorks

Programmi di analisi (Pc/MacQuirer) ed etichettatura (PitchWorks) del segnale *low-cost* per Windows®. Utilizza gli hardware multimediali standard (Sound Blaster™) per acquisire, analizzare e riprodurre il segnale. Utilizza le ultime convenzioni in fatto di etichettatura su multilivelli, soprattutto per quanto riguarda l'intonazione (TOBI) e dispone di un'ottimo algoritmo per l'estrazione di f0 (Figura 13).

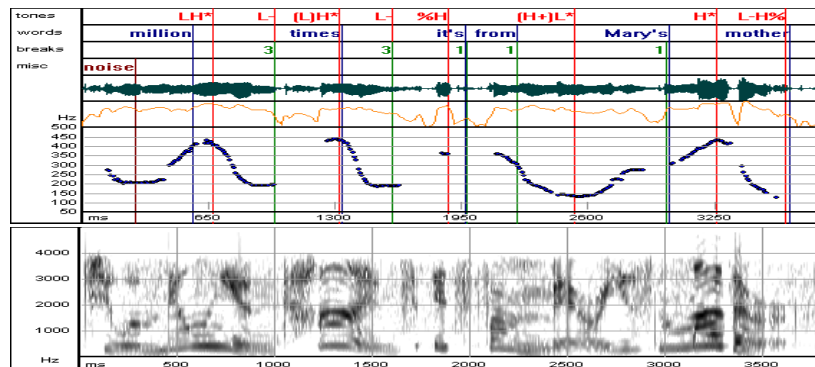


Figura 13. Semplice videata tipo di PCQuirer.

Scicon

<http://www.sciconrd.com/>

#### 4.13 ANVIL: Annotation of Video and Language Data

Anvil è uno strumento per l'annotazione di dati audio/video. Anvil (Figura 14) consente un'annotazione gerarchica multi-livello con "oggetti" definibili dall'utente e, di conseguenza, di natura arbitraria. La visualizzazione sincronizzata delle immagini video con le annotazioni rendono la codifica assai intuitiva. Sono consentiti legami fra livelli ed è stato realizzato un "project manager" di ausilio alle procedure di ricerca. Anvil è scritto in Java e utilizza il linguaggio XML per la memorizzazione dei dati.

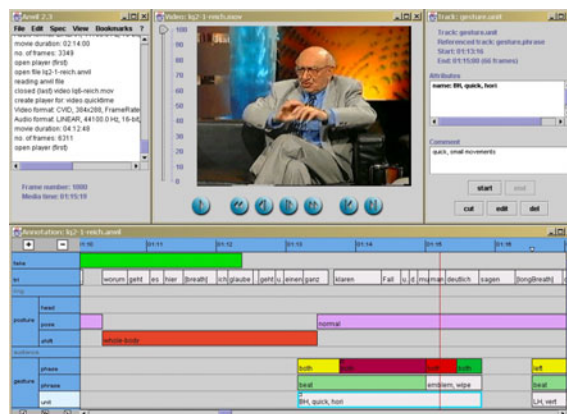


Figura 13. Semplice videata tipo di PCQuirer.

Anvil: <http://www.dfki.uni-sb.de/~kipp/anvil/>

Michael Oliver Kipp ([kipp@dfki.de](mailto:kipp@dfki.de)) German Research Center for Artificial Intelligence GmbHDFKI GmbH, Stuhlsatzenhausweg 3, D-66123 Saarbrücken - Phone: +49 (0) 681 302 5232, Fax: +49 (0) 681 302 5341, Email: [info@dfki.de](mailto:info@dfki.de)

## **5. CONSIDERAZIONI FINALI**

Sono ormai disponibili molti “buoni” sistemi per l’annotazione linguistica e vi sono, per ognuno, pro e contro. Visto che, non sempre rispondono a tutte le esigenze, sono da preferirsi “sistemi aperti”, (basati su Tcl/tk, Java, ecc.) per consentire una facile interazione qualora sia necessaria una qualche modifica. All’Istituto di Fonetica e Dialettologia attualmente utilizziamo EMU per l’etichettatura manuale oppure un sistema per l’allineamento automatico mediante ASR allenato su APASCI. Per l’analisi del segnale è utilizzato PRAAT. Buona annotazione!

## **BIBLIOGRAFIA**

- [1] Annotazione Linguistica  
WWW page: <http://nts.csrf.pd.cnr.it/biblos/annotazione-linguistica.htm>, versione italiana (P. Cosi) <http://nts.csrf.pd.cnr.it/biblos/annotazione-linguistica.htm>.