

## Progetto CLIPS

Corpora e Lessici di Italiano Parlato e Scritto

### W0- a1

(Coordinamento Generale del Progetto)

**Title:** *Sulle tecniche di elicitazione di dialoghi di parlato semi-spontaneo*

**Document No:** CLIPS/W0-a1 /CGP/001

**Document Name:** NADL001

**Status:** **pubblico**

**Date:** 10/3/2006

**Authors:** *Loredana Cerrato*

In questa relazione vengono passate in rassegna le tecniche più diffuse per elicitare dialoghi di parlato “semi-spontaneo”, ne vengono discussi pregi e difetti e in conclusione viene proposto di utilizzare una versione “riveduta” della tecnica del map task per elicitare parlato per l’acquisizione del materiale vocale per il progetto CLIPS.

<b>INTRODUZIONE</b> .....	<b>3</b>
<b>LE TECNICHE DI ELICITAZIONE DI PARLATO SEMI-SPONTANEO DIALOGICO</b> .....	<b>3</b>
L’INTERVISTA .....	4
“INSTRUCTION GIVING DIALOGUES” .....	4
“SITUAZIONI-GIOCO” .....	5
<b>ELICITAZIONE DI MATERIALE VOCALE PER LE APPLICAZIONI FINALIZZATE ALLO SVILUPPO DI SISTEMI DI INTERAZIONE UOMO-MACCHINA</b> .....	<b>6</b>
<b>CRITERI PER LA PROGETTAZIONE DI UN CORPUS VOCALE</b> .....	<b>8</b>
<i>L’acquisizione del materiale per il progetto CLIPS</i> .....	9
PROPOSTE PER L’ ALLESTIMENTO DELLE SESSIONI DI ACQUISIZIONE DEL MATERIALE VOCALE CON LA TECNICA DEL MAP TASK E LA TECNICA DEL GIOCO DELLE DIFFERENZE .....	9
<i>Caratteristiche dei parlatori</i> .....	10
<i>Sessione di registrazione</i> .....	10
<i>Caratteristiche delle mappe</i> .....	10
<i>Caratteristiche delle immagini</i> .....	11
BIBLIOGRAFIA .....	11

## Introduzione

Per corpus vocale si intende una raccolta di registrazioni di parlato, in forma digitale, dotata di annotazioni e trascrizioni in formato elettronico, riutilizzabile per scopi scientifici e/o tecnologici. Tale definizione tende ad escludere una grande quantità di materiale registrato, disponibile in formato analogico e solitamente di qualità non sufficientemente controllata, come ad esempio le registrazioni di parlato spontaneo registrato in situazioni reali. In questo caso le registrazioni risultano essere spesso inadatte per raccolte sistematiche di materiale vocale di alta qualità perché non permettono il controllo da parte degli sperimentatori e perché presentano una serie di distorsioni sul segnale vocale, che possono dipendere sia dal canale di trasmissione, sia dalla incostante distanza del parlatore dal microfono. [Paoloni, Cerrato 1996].

Ad esempio il materiale orale raccolto per l'analisi del lessico di frequenza dell'italiano parlato [De Mauro '93] e per il *British National Corpus* dell'Oxford University Computing Service [Kennedy 1999], pur essendo una preziosissima fonte per effettuare studi sugli aspetti lessicali, sintattici e pragmatici degli scambi comunicativi, risulta inutilizzabile per effettuare accurate misurazioni spettro-acustiche, a causa della presenza di una serie di distorsioni che agiscono come elementi di disturbo sulla qualità del segnale.

Le aree di ricerca linguistica descrittiva e applicata necessitano di un'ampia disponibilità di parlato trascritto (ortograficamente, foneticamente, prosodicamente) e di registrazioni acustiche digitalizzate contenenti "parlato naturale". L'esigenza di analizzare, di registrare e di acquisire campioni di parlato dialogico quanto più possibile "naturale", senza che questo implichi la totale mancanza di controllo sul materiale raccolto, ha indotto i linguisti a escogitare delle "tecniche di elicitazione di parlato dialogico". Si tratta semplicemente di tecniche che stimolano la produzione vocale, e il parlato "elicitato" con queste tecniche può essere definito semi-spontaneo, per differenziarlo dal parlato spontaneo vero e proprio, che si riferisce alla produzione informale di un parlatore senza vincoli sull'argomento di conversazione e sul vocabolario da utilizzare.

### Le tecniche di elicitazione di parlato semi-spontaneo dialogico

Le tecniche di elicitazione di parlato possono essere classificate in due categorie: a) tecniche linguistiche in cui il testo da produrre viene proposto al parlatore in modo più o meno esplicito, ovvero sotto forma di liste di parole, frasi o brani da leggere e/o riprodurre e b) tecniche non linguistiche in cui al parlatore viene affidato un compito non linguistico, che lo distraiga dal reale scopo della registrazione, calandolo in situazioni che favoriscano la produzione di espressioni utili al dominio d'indagine [Eskenazy 1993, Gibbon et alii 1997,].

Indurre i soggetti a calarsi in determinate situazioni e a recitare determinati ruoli, tramite un compito "non linguistico", serve ad attenuare gli effetti di quello che Labov definì il "paradosso dell'osservatore" ("the researcher has to observe how people speak when they are not being observed" [Labov 1972]). Le tecniche di elicitazione consentono di controllare la produzione dei parlatori sul livello di analisi scelto: un controllo a livello lessicale per esempio, si può ottenere proponendo, in maniera più o meno esplicita, il testo da produrre (parole, frasi, brani); oppure facendo leggere ai soggetti una storia che contiene delle "parole target" e chiedendo loro pochi istanti dopo di raccontarsela: in questo modo c'è un'alta probabilità che vengano ripetute numerose parole presenti nel testo proposto; un altro modo può essere quello di far inventare una storia sulla base di fumetti o immagini in grado di evocare le parole relative all'interesse specifico.

Se si vogliono preservare la naturalezza della produzione e la variabilità sia al livello sintattico-lessicale, sia al livello semantico, sia a livello prosodico, si può richiedere ai parlatori di raccontare

un episodio particolare della loro vita, senza introdurre limitazioni di alcun tipo, o si possono creare delle situazioni-gioco per distrarre i soggetti dalla (particolare) condizione di registrazione e indurli a esprimersi con una certa naturalezza.

La scelta della “migliore” tecnica da utilizzare è determinata da una serie di considerazioni preliminari necessarie ad accertare che questa permetterà una conversazione “libera”, che i partecipanti alla conversazione saranno approssimativamente attivi allo stesso modo e che verrà usata una gamma ragionevole di vocaboli in accordo con gli scopi applicativi del materiale da raccogliere.

### ***L'intervista***

L'intervista, uno dei metodi tradizionali di elicitazione del parlato, è utilizzata soprattutto nell'ambito degli studi di dialettologia e sociolinguistica [Ammond 1988, Swedia 2000]. Il suo vantaggio è di fornire vari strumenti di controllo allo sperimentatore/intervistatore, il quale, se abile, può interrompere il parlante, chiedergli di ripetere qualcosa, cambiare l'ordine degli argomenti saltare o inserire argomenti nuovi. Di solito per ottenere materiale quanto più spontaneo possibile si propone ai soggetti un argomento predeterminato di conversazione. In questo modo si possono ottenere registrazioni in cui il vocabolario e la sintassi della produzione non risultano limitate da eccessivi vincoli, ma in cui, allo stesso tempo, lo sperimentatore riesce a mantenere un certo controllo sul contenuto delle produzioni e può stabilire la durata della conversazione.

Questa tecnica viene usata anche per la raccolta di materiale vocale da utilizzare per studi sull'effetto delle emozioni sulla produzione vocale [si vedano ad esempio i lavori di Magno Caldognetto], infatti, chiedendo ai soggetti di raccontare degli eventi particolari della propria vita, o di discutere su argomenti delicati quali divorzio aborto, droga, si riesce ad ottenere parlato “emotivo”.

Nonostante i suoi vantaggi, questa tecnica non è ottimale per elicitare parlato dialogico, perché i due partecipanti alla conversazione non sono attivi allo stesso modo: infatti l'intervistatore, avendo il compito di stimolare il parlante a parlare, produce semplici e brevi domande, mentre l'intervistato, che deve rispondere alle domande discutendo e argomentando, presenta solitamente dei turni più lunghi e delle strutture pragmatiche più varie rispetto a quelle utilizzate dall'intervistatore. Inoltre la buona riuscita di una intervista dipende fortemente dall'abilità dell'intervistatore: un bravo intervistatore può ottenere con un'intervista la stessa quantità e qualità di materiale parlato ottenibile tramite registrazioni casuali o spontanee, col vantaggio che il materiale registrato sotto forma di intervista è caratterizzato da una qualità tecnica di registrazione molto migliore e molto più facile da gestire.

### ***“Instruction giving dialogues”.***

Esistono una serie di tecniche, molto simili tra loro, che inducono due parlanti a interagire comunicativamente al fine di portare a termine un compito o “task”.

Queste tecniche si somigliano molto nei caratteri generali e nella struttura utilizzata per elicitare i dialoghi, e per questa ragione possiamo racchiuderle in una classe che chiameremo “instruction giving dialogues”. La tecnica più diffusa è quella del “map task”, che serve ad elicitare scambi comunicativi finalizzati a dare e a ricevere indicazioni utili per disegnare un itinerario su una mappa schematica. Altri compiti consistono nella messa in opera di qualcosa o nella costruzione di un oggetto.

La tecnica del map task, originariamente ideata da Brown [1983] e poi sviluppata dal gruppo di ricerca del HCRC di Edimburgo per l'acquisizione del corpus HCRC Map Task [Anderson et alii 1991], ha preso sempre più piede per l'acquisizione di corpora di parlato dialogico, perché prevede una conversazione a due abbastanza naturale e allo stesso tempo permette un controllo da parte

dell'operatore. Con questa tecnica si inducono due parlanti a interagire comunicativamente affidando loro il compito di disegnare un itinerario su una mappa schematica.

In ogni mappa sono segnati circa un dozzina di elementi indicativi, rappresentati da icone e dai loro nomi (semplici o composti) con funzione di toponimi (per esempio "casetta", "albero di albicocche", "baia verde", "cinema Totò" ecc.), che possono essere parole o sintagmi target.

Le due mappe non sono completamente identiche, perché il numero, la natura e la posizione degli oggetti differiscono, in modo da rendere possibili eventuali fraintendimenti e momenti di difficoltà o di sorpresa nello scambio comunicativo e garantire, quindi, il requisito della spontaneità e di una relativa varietà elocutiva.

Solo su una delle due mappe è tracciato un itinerario e il compito del partecipante senza l'itinerario è di disegnarne uno sulla propria mappa grazie alle indicazioni fornitegli dal partecipante che ha l'itinerario. I ruoli dei due interlocutori sono quindi stabiliti a priori: c'è un "instruction giver" (colui che dà indicazioni-istruzioni) e un "instruction follower" (colui che segue le indicazioni); anche lo scopo dell'interazione, ossia la destinazione da raggiungere, è definito a priori. La responsabilità maggiore della riuscita dello scambio comunicativo ricade per lo più sul giver, che ha il compito di fornire al follower le indicazioni più appropriate per condurlo/a alla meta sul tracciato della mappa. I due parlatori siedono in un laboratorio di registrazione e sono separati da un pannello che impedisce loro di comunicare visivamente, al fine di massimizzare l'interazione verbale.

I dialoghi elicitati con la tecnica del map task risultano altamente strutturati ed esibiscono delle fasi ben individuabili. In ognuna di queste fasi sono compiute mansioni interazionali ben identificabili e lo scopo finale (ovvero portare a termine il task) modella l'alternanza dei turni, il disegno dei turni e le scelte linguistiche. Per questo motivo tutti i dialoghi elicitati con la tecnica del map task presentano caratteristiche pragmatiche simili e non presentano eccessive sovrapposizioni di turni. Questo dipende dal fatto che i partecipanti a questo tipo di interazioni "prestabilite" effettuano determinate mansioni in modo che gli scopi (generali o specifici) siano ottenuti in maniera ordinata e riconoscibile.

Che questi dialoghi siano altamente strutturati dal punto di vista pragmatico, può rappresentare un vantaggio per chi li utilizza ai fini dell'implementazione di sistemi di dialogo basati sul riconoscimento del parlato, ma può anche rappresentare una limitazione per chi desidera studiare le caratteristiche pragmatiche degli scambi comunicativi.

Un'altra limitazione di questo metodo deriva dal fatto che uno dei due interlocutori, precisamente il giver, essendo investito di maggiore "responsabilità" durante le interazioni, presenta dei turni più estesi, rispetto al follower, che si limita a produrre una serie di risposte minime per comunicare che sta recependo il messaggio, che riesce a seguire le indicazioni, che capisce quanto gli viene spiegato. [Cerrato 1999].

### ***"Situazioni-gioco"***

Esiste un'altra serie di tecniche che si basano sull'idea di creare un contesto conversazionale utilizzando procedure sotto forma di gioco. Le situazioni-gioco, pur vincolando l'argomento di conversazione, hanno il pregio di coinvolgere il parlatore per distrarlo dalla condizione di registrazione e condurlo ad esprimersi con molta naturalezza.

Queste tecniche sono state messe a punto per acquisire parlato in diversi stili di produzione. Infatti l'acquisizione avviene in due fasi successive: la prima in cui si registra il parlato "semi-spontaneo" elicitato tramite la situazione gioco, e la seconda in cui si chiede ai soggetti di leggere la trascrizione della loro conversazione.

A seconda del grado di libertà nella conversazione che si vuol lasciare ai soggetti, si possono scegliere i seguenti giochi:

“Indovina chi”: consiste nell’indovinare il nome di un personaggio o di un oggetto attraverso una serie di domande si/no. Trattandosi di un gioco domanda-risposta effettuato secondo un protocollo fisso, la conversazione che si riesce ad ottenere è molto limitata, ma il materiale che si ottiene può essere utile per studi sull’andamento di  $f_0$  nelle frasi interrogative si/no.

“Ricostruzione di frasi”: consiste nell’indurre due soggetti a ricostruire una serie di frasi, definite a priori e di solito contenenti parole target.

Ogni frase viene divisa dallo sperimentatore nelle parole che la compongono e ciascuna di queste parti viene trascritta su un cartoncino. I cartoncini relativi a ciascuna frase vengono poi divisi casualmente in due gruppi e a ciascun elemento della coppia viene consegnato un gruppo di cartellini e viene richiesto loro di costruire, utilizzando anche gli altri elementi del partner, una frase di senso compiuto.

I soggetti vengono separati da un elemento divisorio che esclude la reciproca vista e massimizza la comunicazione verbale.

Al termine della costruzione delle frasi viene chiesto ai parlanti di leggere la lista delle frasi e la lista delle parole. In questo modo si ottiene parlato su due livelli stilistici differenti: parlato semi-spontaneo (prodotto durante il gioco) e parlato letto (prodotto nella fase di lettura della lista di frasi e di parole). Questo metodo ha il vantaggio di permettere l’inserimento di parole target nella conversazione, distrarre i soggetti dalla consapevolezza della registrazione e consente di ottenere un’alta qualità di registrazione, infatti il materiale raccolto con questa tecnica è stato utilizzato per specifiche analisi segmentali, come ad esempio l’analisi del *burst* nelle occlusive dell’italiano [Cerrato, Falcone 1997].

“Racconti a confronto”: in questo giochino i due soggetti devono scoprire un certo numero di differenze fra due versioni di una storia. Prima della prova, ognuno dei partecipanti legge e memorizza una versione di una breve storia, successivamente i due devono raccontarsela e scoprirne le differenze.

“Gioco delle differenze”: due soggetti devono scoprire una serie di differenze presenti in due immagini apparentemente uguali e descriverle. Solitamente gli oggetti differenti rinviano alle parole target. La produzione orale viene poi trascritta e fatta leggere agli stessi parlanti in modo da ottenere due livelli stilistici differenti: parlato spontaneo e parlato letto.

Questa tecnica è stata utilizzata per l’acquisizione del corpus ICY: “Corpus for the study of intraspeaker variability and the characteristics of speaking styles” [Pean et alii 1993].

Queste tecniche possono avere la limitazione di elicitare conversazioni molto brevi. A causa del task specifico che richiedono, infatti, i giochi a volte vengono conclusi velocemente e senza troppe interazioni. Per di più la ricchezza del dialogo dipende fortemente dalla loquacità, dal coinvolgimento e dalla “sveltezza” dei partecipanti al gioco. Inoltre, dato che i parlatori non hanno dei ruoli prestabiliti, capita spesso che i loro turni si accavallino.

### **Elicitazione di materiale vocale per le applicazioni finalizzate allo sviluppo di sistemi di interazione uomo-macchina.**

L’esigenza di elicitare parlato semi-spontaneo è sorta anche nell’ambito dello sviluppo di applicazioni uomo-macchina. Per poter studiare e modellizzare le strutture di dialogo utilizzate dagli utenti di un determinato sistema per interagire con esso è necessario avere a disposizione corpora di parlato in domini limitati. Questi corpora devono riflettere i differenti livelli del parlato di discorsi mirati all’utilizzo di specifiche applicazioni (ad esempio richiesta di informazioni di volo, prenotazioni alberghiere ecc.) [Sullivan, Tyler].

Per acquisire un materiale che abbia le caratteristiche del parlato naturale, ma sia allo stesso tempo controllato rispetto al contenuto è stata escogitata una particolare tecnica, definita “Wizard of Oz” - Mago di Oz [Fraser, '91]. Questa tecnica consiste nella simulazione, da parte di un operatore umano (il mago, appunto), del comportamento del calcolatore in un'interazione uomo-macchina simulata; il compito dell'operatore è di scegliere tra una serie di risposte predefinite quella più adatta alla domanda di un utente a cui non è nota la condizione di simulazione.

Occorrono due pre-condizioni affinché la simulazione risulti credibile: le specifiche del sistema simulato devono essere compatibili con le limitazioni umane (si pensi ai tempi di risposta di un uomo piuttosto che di una macchina in un'interrogazione di database) e devono essere definiti dettagliatamente i comportamenti dell'operatore in presenza di interrogazioni ambigue o situazioni di errore al fine di garantire la coerenza della simulazione.

Per indurre gli utenti a interagire con questi sistemi si utilizzano degli scenari; uno scenario può essere definito come la descrizione di una delle possibili situazioni previste nel dominio di applicazione e ogni parlatore deve interpretare vocalmente tale situazione. Ad esempio nel dominio di applicazione della prenotazione di voli aerei, un possibile scenario testuale sarà:

*Maria Bianchi e Paolo Rossi vivono a Milano. Devono partecipare ad una riunione a Stoccolma giovedì della settimana prossima. Il meeting inizia alle 11 am e finisce alle 4 pm. Il volo dura circa 2 ore e sono necessari 45 minuti per andare dall'aeroporto al luogo in cui si tiene il meeting. Perciò essi vorrebbero la partenza alle 7 e il ritorno alle 17:30.*

A partire da questo testo, i potenziali utenti produrranno richieste di informazioni relative ai voli tra Milano e Stoccolma.

Uno dei vantaggi di questi scenari è che permettono un alto controllo sulla produzione dei soggetti; con uno scenario si può indurre un soggetto a produrre parole o frasi target, come ad esempio: date, numeri, prezzi, determinati nomi propri, nomi geografici e via dicendo.

Questo vantaggio allo stesso tempo rappresenta uno dei maggiori limiti di questi scenari, perché il testo influenza il parlatore nella scelta del lessico usato per esprimere il concetto; questo fenomeno è stato definito *polarizzazione linguistica* e per evitarlo sono stati utilizzati degli scenari visivi, ovvero scenari capaci di stimolare gli utenti a richiedere una serie di informazioni quali date, orari, disponibilità di servizi ecc, attraverso simboli grafici o icone. Nel progetto europeo Eutrans [ Aiello et alii 1999], ad esempio, allo scenario testuale:

<chiamare la reception dell'Hotel Hitlon per chiedere il costo del servizio di colazione in camera > corrispondeva lo scenario visivo in fig.1.

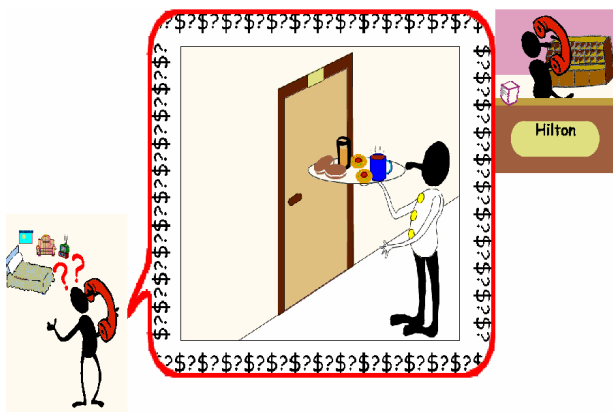


Fig. 1 Esempio di scenario visivo utilizzato per la raccolta di materiale vocale nel progetto europeo Eutrans.

A partire da questi due tipi di scenari gli utenti hanno prodotto due differenti tipologie di frasi, ben rappresentate dai seguenti esempi:

Scenario testuale: buongiorno vorrei sapere il costo del servizio in camera

Scenario visivo: buongiorno, chiamo dalla stanza duecentotrenta, volevo sapere se è possibile avere il servizio di colazione in camera e quanto viene a costare.

Come possiamo notare l'utilizzo degli scenari visivi induce gli utenti a produrre frasi caratterizzati da una maggiore differenziazione del lessico rispetto alle frasi prodotte in base agli scenari testuali, ma allo stesso tempo gli scenari visivi, se non vengono interpretati correttamente dagli utenti, stimolano la produzione di frasi inesatte.

I dialoghi che si ottengono con questo tipo di tecnica sono fortemente condizionati dalla applicazione finale del sistema. La tecnica del mago di Oz infatti, è particolarmente apprezzabile perché permette di studiare le caratteristiche del sistema ed il comportamento dell'utente in circostanze che non sono obbligate dalle limitazioni delle possibilità tecnologiche correnti, inoltre permette la simulazione delle applicazioni in cui una grande popolazione degli utenti potenziali può essere messa alla prova [Bly et alii 1990 ] e ha il vantaggio di permettere di ricreare varie situazioni, per esempio richieste di informazioni ferroviarie, richieste ad agenzie di viaggi, servizio di consegne a domicilio, prenotazione di cinema e teatro, appuntamenti medici, noleggio di automobili eccetera. Per questo è stata utilizzata durante gli ultimi anni in molti laboratori e progetti per lo sviluppo di sistemi automatici di parlato: ha ottenuto un ruolo rilevante nel programma DARPA Spoken Language Systems negli Stati Uniti, ove è stata usata per simulare l' applicazione di ATIS (Air Travel Information System) per la raccolta di dati per addestramento e sviluppo del sistema e per la valutazione dell' utente e del sistema [ATIS] ed è stata utilizzata in vari progetti europei, ad esempio in Esprit Sundial per la simulazione di servizi di inchiesta e di prenotazione di volo e del servizio di prenotazione degli alberghi [Choukri, 1990], nella seconda fase di raccolta del corpus per il progetto Vermobil (finalizzato allo sviluppo di un sistema di traduzione di dialoghi di parlato spontaneo in domini limitati quali: programmazione di appuntamenti, pianificazione di viaggi, prenotazioni alberghiere). Un esempio di scenario utilizzato in Vermobil :

*a foreign visitor could use a GSM cellular phone to have a conversation in English with a German-speaking partner about plans for a trip to Hannover (travel connections, hotel reservations, tickets).*  
[Hess et al. 1995].

Lo svantaggio di questo tipo di interazioni è causato dalla loro struttura fissa, infatti, dato il dominio limitato dell'applicazione, i dialoghi prodotti hanno generalmente la stessa struttura pragmatica che possiamo schematizzare come segue:

Saluto

Apertura

Negoziazione

Chiusura

Saluto

Di conseguenza i dialoghi elicitati con questa tecnica possono risultare significativamente più brevi rispetto ad interazioni comunicative prodotte durante prove colloquiali più tradizionali [Pätzols M. et alii 1995].

### **Criteri per la progettazione di un corpus vocale.**

Nonostante l'elaborazione del linguaggio parlato costituisca un'area di ricerca di estremo interesse, solo recentemente sono state definite e diffuse le specifiche per la progettazione e l'acquisizione dei corpora linguistici nonché le tecniche di misura e valutazione delle prestazioni dei sistemi di trattamento automatico del linguaggio.



Il contributo migliore è rappresentato dall'Handbook of Standards and Resources for Spoken Language Systems [Gibbon,'97] che fornisce dei criteri per la progettazione, la raccolta e la gestione di banche dati finalizzate all'applicazione del trattamento automatico del linguaggio.

Generalmente la progettazione di un corpus deve soddisfare due requisiti antitetici: fornire una base di dati così ampia da contenere numerosi esempi di tutti i fenomeni d'interesse, ma allo stesso tempo di dimensioni limitate tali da consentirne l'etichettatura in tempi brevi o comunque compatibili con la programmazione progettuale.

Assumendo come irrealizzabile la progettazione e la realizzazione di un corpus di piena generalità che fornisca una rappresentazione completa e totalizzante di tutti i fenomeni linguistici ed extra-linguistici occorrenti nelle varie situazioni comunicative, una possibile risposta ai due requisiti sopra illustrati è la specializzazione della base di dati. Una specializzazione che sia determinata dagli scopi di ricerca. Per cui un pre-requisito fondamentale per la raccolta di un corpus è la definizione del dominio di applicazione. Definire un dominio di applicazione è necessario se si vuole che il database ricopra tutte le realizzazioni fonetiche, lessicali e sintattiche di interesse, ma allo stesso tempo abbia una dimensione tale da consentire la sua etichettatura e quindi il suo uso, in tempi brevi o comunque in tempo utile per i progetti di ricerca programmati

**L'acquisizione del materiale per il progetto CLIPS.**

Il progetto CLIPS si prefigge di realizzare un archivio fonico basato su parlato articolato su più livelli stilistici, atto a garantire una sufficiente rappresentatività della variabilità diatopica, utile per le principali attività sia di ricerca linguistica di base sia per i molteplici indirizzi applicativi individuabili nel settore industriale del trattamento automatico dei segnali vocali, per cui il suo scopo è abbastanza vasto e, data la genericità delle applicazioni previste per la sua utilizzazione, anche il dominio è molto generico, quindi comprensivo di un vocabolario non specifico e adatto per un'ampia gamma di applicazioni. Di conseguenza trovare una tecnica di elicitazione ottimale risulta compito abbastanza arduo. Valutando pregi e difetti delle tecniche passate in rassegna, la scelta ricade nuovamente sulla tecnica del map task.

Le tecniche gioco, come abbiamo visto, presentano più svantaggi rispetto alla tecnica del map task, in quanto la riuscita dell'interazione dipende fortemente dalla personalità dei parlatori (dalla loro loquacità, sveltezza, coinvolgimento nel task), inoltre non essendoci una suddivisione precisa dei ruoli, può risultare difficile controllare l'alternanza dei turni.

La tecnica del mago di Oz è indicata per acquisizione di interazioni uomo-macchina e l'intervista non è raccomandabile perché con essa si ottengono dialoghi poco bilanciati, in cui i due partecipanti non sono attivi allo stesso modo.

***Proposte per l'allestimento delle sessioni di acquisizione del materiale vocale con la tecnica del map task e la tecnica del gioco delle differenze.***

Per le sessioni di registrazione di dialoghi all'interno del progetto CLIPS, allo scopo di porre rimedio alle due principali limitazioni riscontrate nel materiale elicitato precedentemente con la tecnica del map task in precedenti progetti, vale a dire la prevedibilità della struttura pragmatica degli scambi interazionali e la produzione di dialoghi sbilanciati dal punto di vista degli interventi, si propone dunque di operare con due differenti modalità di elicitazione dei dialoghi:

1) utilizzare una versione "riveduta" della tecnica del map task che preveda semplicemente la creazione di una serie di mappe "incrociate", ovvero divise in due parti, una con e una senza itinerario, in modo tale che entrambi i parlatori possano assumere, nello stesso dialogo, sia ruolo di giver sia il ruolo di follower al fine di completare la stessa mappa. In questo modo si eviterebbe di

ottenere dialoghi sbilanciati dal punto di vista degli interventi e si continuerebbero ad avere tutti i vantaggi offerti da questa tecnica.

2) integrare il compito del map task con un ulteriore compito, simile a quello descritto sotto il titolo di “gioco delle differenze” nel paragrafo “Situazioni-gioco”. Nel compito del “gioco delle differenze” i due partecipanti (che non possono vedersi) hanno due versioni della stessa immagine che differiscono tra loro in un numero finito di dettagli grafici. Il loro compito è di fornire una dettagliata descrizione della propria immagine e cercare di scoprire quali siano le differenze rispetto all'altra.

Questa tecnica è spesso usata come metodo di elicitazione di parlato nella didattica delle lingue straniere.

L'aggiunta di questo task interromperebbe la struttura interazionale tipica dell' *instruction giving dialogue* e potrebbe permettere di elicitare parlato meno strutturato dal punto di vista pragmatico.

Infatti in questo task i due interlocutori non hanno un ruolo prestabilito e essendo investiti dallo stesso compito, ovvero descrivere l'immagine e scoprire le differenze con l'immagine dell'altro partecipante, possono produrre dialoghi meno sbilanciati dal punto di vista degli interventi.

### Caratteristiche dei parlatori.

Nella raccolta del materiale per l'HRC corpus, le coppie di parlatori erano classificate in due categorie: parlatori che si conoscevano e che quindi avevano familiarità tra di loro, e parlatori che non si conoscevano.

Selezionare parlatori che si conoscono garantisce una maggiore naturalezza della produzione, ma probabilmente favorisce una serie di riferimenti “impliciti” a fonti di conoscenze in comune, che potrebbero rendere i dialoghi meno espliciti.

Ciò nonostante, l'esperienza maturata nell'ambito del progetto AVIP è tale da far preferire la prima delle due opzioni, inducendo a raccomandare di impiegare coppie di persone che si conoscono bene.

Un criterio indispensabile per la buona riuscita dei dialoghi è che i parlatori siano loquaci.

Altri criteri relativi alle caratteristiche sociali e linguistiche dei parlatori saranno definite in altra sede.

### Sessione di registrazione

I due parlatori siederanno in un ambiente silente e saranno separati da un pannello che impedisca loro di comunicare visivamente, al fine di massimizzare l'interazione verbale.

Prima di iniziare le sessioni di registrazione bisogna accertarsi che i soggetti si sentano a proprio agio. Per dare loro il tempo di adattarsi allo studio e per allentare l'eventuale tensione, si può chiacchierare in maniera informale con loro per alcuni minuti, spiegando, in maniera informale, anche le generali procedure di registrazione. I parlatori vanno istruiti a evitare i rumori di fondo (per esempio spostamenti di sedie, colpi di tosse, rumori fatti con le mani, i piedi, ecc.).

E' consigliabile inoltre iniziare le sessioni di registrazione con un dialogo di addestramento per permettere ai soggetti di adattarsi.

### Caratteristiche delle mappe

In ogni mappa vanno segnati un numero pari di elementi indicativi, non inferiori alla dozzina, rappresentati da icone molto particolareggiate (per esempio: un ufficio postale con una particolare architettura).

I criteri per la scelta dei nomi delle icone e per la selezione delle parole target saranno forniti in altra sede.

Le due mappe non devono essere completamente identiche, ma il numero, la natura, e la posizione delle icone devono differire, in modo da rendere possibili eventuali fraintendimenti, momenti di difficoltà o di sorpresa, necessità di chiarimenti ecc. nello scambio comunicativo.

Le due mappe devono avere lo stesso punto di partenza, uno stesso punto intermedio di scambio ruoli e lo stesso punto di arrivo.

Sono necessarie tante coppie di mappe per quanti dialoghi diversi si vogliono elicitare.

### Caratteristiche delle immagini per il compito di “gioco delle differenze”.

Per il compito del “gioco delle differenze” si forniscono ai due partecipanti due immagini (o due vignette) che differiscono tra loro in un numero finito di dettagli grafici, e si chiede loro di:

- a) fornire una dettagliata descrizione della propria vignetta,
- b) scoprire i particolari che differenziano le due versioni della vignetta.

Le immagini possono essere facilmente ricavate da libri di glottodidattica o da rubriche di enigmistica.

Per stimolare descrizioni ricche e particolareggiate e per evitare il rischio che i dialoghi elicitati con questa tecnica risultino troppo brevi, è necessario disporre di immagini molto dettagliate e possibilmente a colori.

Cercare di ottenere parlato quanto più naturale possibile, effettuando però un controllo sui parlatori e sul materiale prodotto è un altro dei paradossi dell’osservatore! Infatti, tutte le interazioni “prestabilite”, avendo uno scopo preciso, sono effettuate in maniera ordinata e riconoscibile. Se dal punto di vista strettamente pragmatico questo può essere visto come un limite, dal punto di vista delle applicazioni tecnologiche questo rappresenta un vantaggio, perché una macchina può essere istruita a riconoscere un dialogo se questo ha una struttura abbastanza prevedibile.

### *Bibliografia*

Aiello, D., Delogu, C., De Mori, R., Di Carlo, A., Nisi, M., Tumeacciu, S., 1999, Comparative evaluation of spoken corpora acquired by presentation of visual scenarios and textual descriptions, *Proc. ICASSP’99, Phoenix, March*.

Eutrans <http://hermes.zeres.de/Eutrans>.

Ammon U, Dittman R. Mattheier K.J., 1988, *Sociolinguistics Vol.2, Chapt.2 “Elicitation Methods”*.

Anderson A.H., Bader M., Bard E.G., Boyle G., Doherty G., Garrod S., Isard S., Kowtko, J. McAllister J., Miller J. Sotillo C., Thompson H., Weinert R., 1991, The HCRC Map Task Corpus, *Language and Speech*, 34, (4):351-366.

ATIS <http://www.ai.sri.com/natural-language/projects/arpa-sls/atis.html>

Bly, B., Price, P., Park, S., Tepper, S., Jackson, E., Abrash, V., 1990, Designing the Human Machine Interface in the ATIS Domain, *Proc. Third DARPA Speech and Language Workshop*, Hidden Valley, PA, pp. 136-140.

Brown G., Anderson A., Yule G. e R. Shillcock, 1983, *Teaching talk*, Cambridge, CUP.

- Labov W., 1977, Some principles of Linguistic methodology, *Language and Society*, 1:97-120.
- Choukri, K., 1997 Simulation Studies, *SUNDIAL (ESPRIT Project P2218), Work Package 3, deliverable 1*.
- Cerrato L., Falcone M., 1997, Il burst nelle occlusive in sequenze VCV e VC:V dell'italiano: un'analisi acustica, in: P.M. Bertinetto, L. Cioni (a c. di), Unità fonetiche e fonologiche: produzione e percezione, *Atti delle VIII Giornate del Gruppo di Fonetica Sperimentale*, pp. 29-40.
- Cerrato L., 1999, Il feedback verbale in dialoghi elicitati con la tecnica del map-task, *Atti delle Giornate del Gruppo di Fonetica Sperimentale IUO Napoli dic.1999*, in stampa.
- De Mauro T., Mancini F., Vedovelli M., Voghera M., 1993, Lessico di frequenza dell'Italiano Parlato, Etas libri Milano.
- Eskenazy M., 1993, Trends in speaking styles research, *Proceedings of Eurospeech 1993*, pp. 501-509.
- Fraser, N., Gilbert, G.N., 1991, Simulating speech systems, *Computer Speech and Language, Vol 5(1)*, 81-99.
- Grice, H.P. 1957, Meaning, *The philosophical review*, 66:377-388.
- Gibbon D, Moore R., Winski R.(eds.) 1997, Handbook of Standards and Resources for Spoken Language Systems, Mouton de Gruyter, Berlin - New York.
- Kennedy G. 1999, An Introduction to corpus Linguistics, Longman London.
- Hess W., Kohler K., Tillmann H., 1995, The PhonDat/Verbmobil Speech Corpus. In *Proceedings of Eurospeech95*, Madrid P. 863-870.
- Labov W. 1972, Sociolinguistic patterns, University of Pennsylvania press.
- Magno Caldognetto E., Zmarich C., Ferrero F., 1998, Indici acustici macroprosodici dello stato emotivo del parlante, *Atti del XXVI convegno naz. Ass. Ital. di Acustica*, Torino, p. 263-266.
- Norman, D.A., Draper, S.W. ,1986, User centred system design, Lawrence Erlbaum Associates: Hillsdale, NJ.
- Paoloni A., Cerrato L.,1996, Sulla trascrizione delle intercettazioni ambientali, *Rivista Italiana di Acustica* Vol.20 n.4, p.159-164.
- Peau V., Williams S., Eskenazy M., 1993, the design and recording of ICY, a corpus for the study of intraspeaker variability and the characterisation of speaking styles, *Eurospeech 1993*, pp.627-630.
- Pätzold M. Scheffers M. Simpson A Controlled Elicitation and processing of Spontaneous speech in Vermobil ICPHS 1995 pp.312-318.
- Swedia 2000 [http://www.ling.umu.se/~anderse/SWEDIA/publ\\_sv.html](http://www.ling.umu.se/~anderse/SWEDIA/publ_sv.html)

Sullivan E.S., Tyler W. (eds.) Intelligent user interfaces New York ACM Press.