

Progetto CLIPS

Corpora e Lessici di Italiano Parlato e Scritto

W₁ – a₂
(Corpus telefonico)

Title: **Definizione del contenuto del corpus telefonico e linee guida per la raccolta**

Document No: CLIPS/W1-a2/DCS/003

Document Name: FUBDL002.rtf

Status: **pubblico**

Date: 10/3/2006

Authors: *Andrea Di Carlo, Leandro D'Anna*
FONDAZIONE UGO BORDONI
Via B. Castiglione, 59 - 00142 Roma - Italy

INTRODUZIONE.....	3
DEFINIZIONE DEL DOMINIO	3
TASK ANALYSIS.....	3
LINEE GUIDA PER LA RACCOLTA	5

Introduzione

Scopo di questo corpus è quello di fornire materiale acustico di qualità telefonica di tipo "naturale" nell'ambito di un dominio applicativo realistico.

Per quanto attiene alla qualità, il materiale viene acquisito attraverso un sistema telefonico reale impiantato con hardware convenzionale e commerciale (linee ISDN, terminazione analogica, scheda telefonica Dialogic,...). L'acquisizione avviene alla frequenza di campionamento di 8000Hz e la codifica è nel formato mu-law (compressione logaritmica).

Il corpus telefonico, pur considerato in questo progetto, distinto dai corpora speciali (oggetto del W1_a3), è per sua stessa natura da riferirsi ad un contesto applicativo. E' stato oggetto di studio e di dibattito, durante gli scorsi anni, l'individuazione di un'applicazione telefonica "particolarmente significativa" a fini scientifici; i risultati di questo dibattito sono tutt'altro che evidenti ma c'è stata una certa convergenza a livello internazionale sull'interesse del dominio turistico perché promettente di prospettive economiche (accesso via telefono mobile), affrontabile con la tecnologia di riconoscimento e di elaborazione del linguaggio (traduzione automatica, interrogazione,...) attualmente disponibile e, al tempo stesso, di complessità media (vocabolari di qualche migliaia di parole) tale da consentire una qualche naturalezza di espressione

Definizione del dominio

Nell'ambito turistico, abbiamo immaginato un servizio alberghiero automatico che permette all'utente di fare richieste al ricevimento di un albergo via telefono e che può essere utilizzato sia da fuori che dall'interno dell'albergo.

Lo schema semplificato di interazione dovrebbe essere:

- il turista chiama dall'esterno (dall'aeroporto, dalla stazione ferroviaria, da casa, dall'ufficio, da un altro albergo) or dall'interno (dalla propria stanza d'albergo) il sistema per richiedere informazioni e/o servizi;
- il sistema elabora la richiesta dell'utente, accede ad un sistema di back-end (database, scheduler,...) e quindi manda all'utente una risposta adeguata e/o predispone l'attuazione del servizio.

Nell'ambito di questo contesto, abbiamo fatto la cosiddetta "task analisi" del dominio alberghiero.

Task Analysis

Sono stati individuati 5 tasks principali che gli utente possono svolgere; per ogni task, sono stati individuati due o più sotto-task:

Tasks -

1. chiedere informazioni
2. fare una richiesta
3. cancellare una richiesta
4. lamentarsi per qualcosa
5. richiedere un cambiamento

Subtasks -

1. chiedere informazioni dall'interno o dall'esterno:
 - sul costo di una stanza con certe caratteristiche (vedi I.)
 - sulla disponibilità di una stanza con certe caratteristiche per un dato periodo
 - sulla disponibilità di un servizio nell'albergo (vedi II.) e all'esterno (III)
 - sulla localizzazione dell'albergo
 - sui tempi di un servizio
 - sui costi di un servizio

- 2.a fare una richiesta dell'interno o dall'esterno:
 - fare una prenotazione per una stanza con certe caratteristiche e per un certo periodo
- 2.b fare una richiesta dall'interno:
 - richiedere un servizio
 - richiedere il conto

- 3.a cancellare una richiesta dall'esterno:
 - cancellare una prenotazione
- 3.b cancellare una richiesta dall'interno:
 - cancellare la richiesta di un servizio

- 4.a lamentarsi di qualcosa dall'esterno:
 - lamentarsi per la navetta dall'aeroporto all'albergo ancora non arriva
- 4.b lamentarsi di qualcosa dall'interno:
 - lamentarsi per una stanza
 - lamentarsi per un servizio

- 5.a fare un cambiamento dall'esterno:
 - cambiare una prenotazione:
 - cambiare la stanza
 - cambiare il periodo
 - cambiare l'orario di arrivo
- 5.b fare un cambiamento dall'interno:
 - cambiare una stanza
 - cambiare la richiesta di un servizio

I. *Caratteristiche delle stanza:*

- singola/doppia/matrimoniale
- con cassaforte
- per fumatori/non fumatori
- con TV
- con aria condizionata
- con il frigobar
- con il bagno

- con la doccia nel bagno
- con la vasca nel bagno
- con l'asciugacapelli nel bagno

II. Servizi: servizi interni:

- piscina
- palestra
- ristorante
- bar
- parcheggio
- navetta da/per l'aeroporto/la stazione
- servizio sveglia
- colazione
- babysitting
- lavanderia
- ammissione animali
- servizio fotocopie
- servizio di posta
- servizio di messaggeria

III. Servizi: servizi esterni:

- taxi

Come ulteriore requisito è stato posto che il corpus contenesse anche esempi di stringhe alfa-numeriche che sono costantemente presenti nei corpora applicativi di tipo telefonico; sulla base dell'analisi del dominio, si è ritenuto che questo tipo di materiale possa essere raccolto sotto forma di "numero di prenotazione" (necessario in task di modifica e/o di cancellazione per riferirsi a prenotazioni già fatte) e sotto forma di "numero di carta di credito".

Linee guida per la raccolta

Per questo progetto, ci si propone la raccolta di materiale telefonico prodotto da circa 300 parlatori per un totale di circa 10 ore di parlato. I parlatori verranno reclutati attraverso un'agenzia, tipo Doxa, che ne curerà la disponibilità e la localizzazione nelle aree indicate in coerenza con le relative specifiche fissate dal partner UNILE. [vedi il documento LEDL001]

I parlatori riceveranno una guida con 10 "scenari". Per "scenario" intendiamo una descrizione testuale di un compito che il singolo parlatore dovrà svolgere usando il sistema. Ogni parlatore (utente) chiamerà un centro di acquisizione e interagirà con esso. I parlatori chiameranno dalle località predefinite mentre la stazione di acquisizione sarà fissa a NA; quindi il materiale raccolto sarà in massima parte telefonico di tipo non locale. Gli utenti dovranno completare tutti i compiti descritti negli scenari consegnati.

L'acquisizione avverrà durante l'interazione con una versione simulata di un'applicazione telefonica. Una stazione di lavoro dovrà essere destinata a questa simulazione; poiché la simulazione potrebbe

richiedere l'intervento di un operatore chiamerò la stazione di lavoro l'Operator Console (OC).

L'operatore dovrà agire secondo un protocollo predefinito per controllare le risposte del sistema in modo da simulare un comportamento automatico; si delineano per l'operatore i compiti che, in letteratura, sono propri della tecnica del Mago di Oz; pertanto l'operatore verrà equivalentemente indicato anche come Mago di Oz o WoZ (Wizard of Oz).

La linea telefonica sarà connessa all'OC mediante una scheda telefonica. La linea e le chiamate saranno gestite da un pacchetto software basato sulla libreria API della scheda. Sarà compito di questo software di gestione:

1. aprire e chiudere la linea telefonica;
2. attendere chiamate;
3. rispondere a e terminare chiamata
4. riprodurre ed registrare archivi di segnale vocale sulla linea telefonica: l'acquisizione avverrà ad una frequenza di campionamento di 8000 Hz e il segnale verrà codificato in mu-law.
5. riconoscere i toni DTMF
6. controllare il livello di segnale sulla linea telefonica per riconoscere silenzi, cadute, chiusure, ecc.

Per ragioni di affidabilità si è preferito duplicare l'OC in modo da

1. assicurare la disponibilità del servizio in caso di crash di una delle due OC
2. sopportare su due linee il flusso di chiamati entranti (vedi Traffico previsto)

Le linee telefoniche sono costituite dai due canali di una linea ISDN configurata in modalità mononumero, dotata di interfaccia NTPlus; su tale linea ISDN è stato attivato un numero verde.

Le due linee vengono attestate a due OC fisicamente separate e identiche quanto ad Hw di interfaccia, a configurazione del SW di controllo, al SW applicativo e alla configurazione delle directories; la sola differenza delle due macchine è nella dotazione HW/SW di base (Pentium II e Athlon con rispettivamente i sistemi operativi NT4 e Windows 2000). Sarà verificata l'assenza di essenziali differenze di comportamento delle due OC.

AI fini del monitoraggio, le due linee analogiche in ingresso alle OC sono inviate anche in ingresso alla scheda audio del computer dove opera il mago di Oz utilizzando per ognuna di esse uno dei due canali audio (sinistro e destro) di un cavo stereo. Dalla scheda audio il segnale è inviato in uscita ad una coppia di altoparlanti amplificati e a cui può essere collegata una cuffia.

Sempre ai fini dell'affidabilità della raccolta, è stata predisposta una procedura di salvataggio dei dati raccolti. Tale procedura prevede, per ogni stazione, la creazione di una copia giornaliera zippata dell'intero directories acquisite nella giornata del salvataggio e nel giorno precedente e l'aggiornamento di una copia incrementale zippata contenente le intere directories che sono state acquisite fino al giorno precedente il salvataggio. Entrambi i files vengono salvati su un HD supplementare e su un CD riscrivibile. Inoltre ogni settimana viene creata una copia integrale su CD registrabile di quanto è stato acquisito da ognuna delle due stazioni dall'inizio dell'acquisizione.

Sulla base del software di gestione andrà costruito il software di interfaccia per l'operatore.

La OC può lavorare in due modalità differenti: automatica o WoZ: passare da una modalità all'altra è facile e sempre possibile. Ciò consente di estendere il tempo utile per le acquisizioni alle ore fuori orario di ufficio o ai momenti di pausa in cui la stazione non è presidiata dal WoZ. Naturalmente l'interazione in modalità automatica è notevolmente semplificata non essendo la gestione del dialogo stata automatizzata.

In modalità WoZ, il sistema funziona nel seguente modo. Ogni parlatore, chiamato a giocare il ruolo di utente del sistema, è fornito di una descrizione testuale (lo scenario) del compito che deve svolgere. Deve chiamare il numero verde al quale gli risponde un sistema automatico con cui dovrà interagire. Il sistema gli chiede di inserire, via tastiera, il codice stampato sul suo scenario. In questo modo il sistema ritrova le informazioni relative a quello scenario e le mostra sullo schermo al WoZ: in particolare, viene mostrato un modulo che il WoZ dovrà riempire. Il WoZ sarà quindi aiutato e indirizzato nello svolgere il suo compito dal riempimento di questo modulo. Per ogni compito dell'utente, esiste un modulo che indica al WoZ quali informazioni è necessario raccogliere dall'utente e quindi è necessario fargli produrre vocalmente. Queste informazioni sono classificate in tre categorie: identità, obbligatorie e opzionali. Il compito del WoZ (e indirettamente quello dell'utente) è compiuto se le informazioni di identità e quelle obbligatorie sono state raccolte.

Consideriamo per esempio, il compito "prenotazione di una stanza"; il corrispondente scenario per il parlatore potrà essere

"Lei si trova nella propria abitazione. Sta telefonando alla reception dell'albergo Excelsior di Parigi per prenotare una camera tripla, con vista, con TV, con doccia, per non fumatori, con cassaforte, per la settimana di Natale a nome di tre amici. Il suo numero di carta di credito è 7497 3792 1801 9340"

il corrispondente modulo per il WoZ sarà del tipo:

Identità

Nome utente:

Info obbligatorie

Tipo di stanza

Data di arrivo

Data di partenza

Numero di carta di credito

Info facoltative

Richiesta di vista panoramica

Tipo di bagno

Dotazione della cassaforte

Stanza per fumatori

Orario di arrivo

Messaggi preregistrati sono associati ad ogni slot del modulo in modo che il WoZ possa eventualmente selezionarli e attivarli per mandare messaggi vocali all'utente. Pur essendo negli obiettivi di questo corpus raccogliere materiale spontaneo, non vogliamo che messaggi troppo naturali del sistema possano indurre il parlatore ad un comportamento linguistico troppo spontaneo (non consono alle possibilità di riconoscimento e di comprensione dell'attuale tecnologia di riconoscimento vocale e di elaborazione del linguaggio naturale) e quindi abbiamo deciso che i messaggi preregistrati siano prodotti da un sistema di sintesi (TTS) in modo da ricordare indirettamente ai parlatori che stanno interagendo con una macchina.

Riempendo il modulo e ascoltando in cuffia ciò che viene detto dall'utente (e registrato dal sistema) il WoZ può esercitare un controllo sul materiale acquisito:

- se l'espressione dell'utente non è chiara (rumorosa, troppo involuta nella forma,...), può attivare messaggi per richiedere la ripetizione o la riformulazione dell'espressione;
- se l'espressione è chiara ma non contiene tutte le informazioni rilevanti per concludere il compito, può richiedere ulteriori dettagli e le informazioni mancanti;
- se l'espressione è chiara e completa, il WoZ può inviare il messaggio finale di commiato più adatto per quello scenario.

Ogni azione del sistema e/o del WoZ è memorizzata in logfile.

Quando il sistema è posto in modalità automatica, esso automaticamente sceglie un messaggio finale adatto allo scenario e lo invia all'utente per chiudere la conversazione. Questo è possibile perché ad ogni scenario (individuato mediante il codice inserito dall'utente) è associato un insieme di risposte pertinenti tra le quali viene selezionata casualmente quella attivata. In modalità automatica, non è raccolta alcuna turnazione (o meglio i dialoghi sono tutti ad un solo turno dell'utente umano) e le interazioni sono costituite da 3 turni (prompt-del-sistema, enunciato-utente, messaggio-di-commiato-del-sistema).

La prima OC è stata permanentemente presidiata da un operatore addetto alla funzione del WoZ compatibilmente con gli orari (l'acquisizione è rimasta disponibile 24h/24 grazie alla modalità automatica); la seconda stazione è stata invece predisposta in modalità automatica.

Traffico previsto

Il reclutamento dei parlatori è affidato ad una società di indagine statistica che provvederà a coordinare il flusso delle chiamate al sistema di acquisizione.

Il traffico è stato così predisposto: ogni giorno verrà attivato un parlatore in ognuna delle città campione che provvederà ad effettuare tutte le chiamate del suo compito nell'arco dell'orario lavorativo; sarà quindi assicurato un flusso di 15X10 telefonate al giorno per una durata di 20 giorni.

In realtà, pur rispettando il flusso previsto e la durata complessiva dell'acquisizione, le chiamate sono state organizzate in maniera differente e meno sistematica: tendenzialmente ogni giorno è stato tentato di esaurire il compito di una città provvedendo al massimo 20X10 telefonate. Il flusso massimo non è comunque stato mai raggiunto.

Numeri di carta di credito

Nell'ambito di SpeechDat, è stata proposta una lista di numeri di carta di credito rispondenti ad alcuni requisiti di formato. Questa lista è stata già parzialmente utilizzata per una precedente raccolta. Per questa raccolta, è stata utilizzata la lista dei rimanenti; i singoli numeri sono stati integrati negli scenari di prenotazione di una stanza.

Il numero di carta di credito è formattato come un numero di 4 blocchi di 4 cifre, cioè xxxx xxxx xxxx xxxx. I numeri sono composti in modo da fornire una buona copertura della distribuzione di trifoni e, nello stesso tempo, un realistico controllo di parità attraverso l'ultima cifra.

Orari

Sul totale di 3000 scenari da raccogliere, ci sono 316 scenari che prevedono l'indicazione di un orario (ad esempio la prenotazione o la disdetta di servizi interni quali la sveglia o la baby-sitter e servizi esterni quali il ristorante o il taxi) e 265 scenari che prevedono l'indicazione di due orari (ad esempio la variazione di orario di servizi interni quali la sveglia o la baby-sitter, di servizi esterni quali il ristorante o il taxi o l'orario di arrivo di un cliente).

Per quanto riguarda la copertura degli orari (dalle 0.00 alle 24:00), ogni intervallo di un'ora è stato suddiviso in frazioni di dieci minuti (es. 14:00 – 14:10 – 14:20 – 14:30 – 14:40 – 14:50) per un totale di $24*6=144$ orari distinti da pronunciare. Ognuno di questi orari, poi, verrà pronunciato 5-6 volte da parlanti diversi per un totale di 846 orari pronunciati.