

SINTESI DA TESTO MULTILINGUE CON IL SISTEMA ACTOR®

Silvia Quazza, Pier Luigi Salza

CSELT S.p.a., Via G. Reiss Romoli 274, 10148 Torino

RIASSUNTO

Nella sintesi vocale da testo le componenti linguistico-fonetiche rappresentano una parte consistente della conoscenza che è necessario fornire globalmente al sistema. Inoltre, la natura e la complessità di tali componenti di conoscenza sono fortemente specifiche e dipendenti dalla lingua considerata, tanto da poter affermare che soltanto i modelli matematici di rappresentazione e di analisi del segnale vocale e il cosiddetto “motore di sintesi” possono considerarsi in larga misura indipendenti dalla lingua. Prendendo come esempio ACTOR®, nome commerciale del sistema di sintesi realizzato in CSELT, ultima evoluzione di Eloquens® [1], si illustrano le componenti di conoscenza linguo-specifiche che è necessario acquisire e fornire al sistema nel momento in cui si intende realizzare la sintesi di una nuova lingua.

Il “motore di sintesi”, come già descritto in precedenza [2] è basato su una tecnica di sintesi nota come ‘sintesi per concatenazione di unità acustiche’, che impiega segmenti acustici estratti direttamente dalla voce naturale. ACTOR® appartiene alla nuova generazione di sistemi di sintesi per concatenazione di unità acustiche, detti *corpus based*, i cui principi fondamentali consistono nel:

- disporre di una grande base di dati vocale registrata;
- fornire all’algoritmo di sintesi una notevole mole di conoscenze esterne di tipo fonetico-prosodico,.

Il sistema è poi in grado, in tal modo, di sfruttare al meglio il materiale vocale disponibile, selezionando *run time* le unità, di lunghezza non uniforme e contestuali, che meglio si combinano per generare il messaggio da sintetizzare. Tecniche di modifica del segnale sono applicate solo quando necessario, per smussare salti prosodici residui ai confini tra unità o quando mancano unità con una determinata prosodia.

Le principali componenti di conoscenza linguo-specifiche che è necessario acquisire e fornire al sistema per sviluppare una nuova lingua sono qui schematicamente elencate sotto forma di passi operativi, ordinati secondo un criterio di priorità:

1. definire il sistema fonologico della lingua;
2. sviluppare regole di espansione di abbreviazioni, sigle e numeri;
3. sviluppare regole di scansione del testo;
4. sviluppare regole di accentazione e/o preparare liste di parole con marcatura dell’accento e categoria grammaticale per addestrare il classificatore lessicale;
5. sviluppare regole e/o preparare liste di parole per la deaccentazione;
6. sviluppare le regole di trascrizione fonetica ovvero approntare un lessico di base trascritto;
7. preparare liste di parole trascritte (liste di fonetizzazione) per eccezioni e parole straniere;

8. sviluppare eventuali regole di analisi morfologica per la determinazione della trascrizione fonetica e dell'accento
9. sviluppare regole per la determinazione della struttura sintattico-prosodica (*phrasing*, assegnazione delle pause ed eventuale *focus* di frase);
10. sviluppare regole di back-up per assegnare durata fonetica e contorno intonativo, almeno per le modalità dichiarativa e interrogativa;
11. determinare la copertura fonetico-prosodica desiderata mediante analisi statistica di un vasto corpus di testi rappresentativo della lingua (centinaia di migliaia di frasi) precedentemente raccolto;
12. realizzare progetto dei testi da registrare in modo da garantire la copertura desiderata;
13. selezionare uno speaker madrelingua;
14. registrare i testi per l'addestramento e il test del segmentatore automatico in foni e difoni per la nuova lingua (allineatore fonetico);
15. registrare i testi di copertura (basati acustica);
16. applicare le procedure automatiche di: segmentazione (in foni, difoni e *demifoni*¹), analisi del pitch, generazione del dizionario acustico;
17. controllare la corrispondenza tra il segnale e la segmentazione e trascrizione fonetica automatica dei testi registrati mediante ascolto e visualizzazione del segnale della basati acustica.

Per tutte queste operazioni, appare evidente quanto, nello sviluppo di nuove lingue, sia di fondamentale importanza l'apporto della consulenza di esperti linguisti 'nativi'. Alcuni dei passi elencati verranno sviluppati con maggiore dettaglio. Di altri verranno forniti esempi che mostrano il peso della specificità delle singole lingue nella ricerca della migliore soluzione.

[1] Balestri, M., Pacchiotti, A., Quazza, S., Salza, P.L., Sandri, S., *Choose the best to modify the least: a new generation concatenative synthesis system*, Proc. EUROSPEECH '99, Budapest, Agosto 1999, vol. 5, pp. 2291-2294.

[2] Salza P. L., Quazza S. (1999): "Una tecnica di sintesi della voce che conserva le caratteristiche acustico-prosodiche dello speaker donatore", Atti delle X Giornate di Studio del Gruppo di Fonetica Sperimentale (A.I.A.), Napoli, dicembre 1999

¹ Il *demifono* è definito come la "porzione di segnale delimitata da un confine di fono e un confine di difono". Tale definizione conserva un nesso sia con il concetto linguistico del fono e sia, nel contempo, con la ricerca del miglior punto di giunzione tipico della sintesi concatenativa.