



UNIVERSITÀ
CA' FOSCARI
VENEZIA

DIPARTIMENTO
DI SCIENZE
DEL LINGUAGGIO

SEGRETERIA DIDATTICA
tel. 041 234 5704-5705
fax 041 234 5706
filnia@unive.it
vanessa@unive.it

SEGRETERIA AMMINISTRATIVA
tel. 041 234 5780
fax 041 234 5703
mcduse@unive.it

CA' BEMBO
Dorsoduro, 1075
30123 VENEZIA

COD. FISC. 80007720271

STILVEN 2009: Un Traduttore INGLESE-VENETO e viceversa RELAZIONE CONCLUSIVA – Dicembre 2009

1. Introduzione

Il progetto preliminare finanziato con 20 mila euro aveva prodotto i primi risultati, regolarmente consegnati alla fine dello scorso anno su CD. In questo secondo anno di progetto abbiamo prodotto di nuovo una grande mole di materiali in forma digitale che hanno anche contribuito in maniera determinante a far avanzare il progetto del traduttore automatico ora disponibile su sito web, anche se in una forma ancora molto rudimentale in quanto a precisione della traduzione.

I risultati in nuce sono i seguenti:

1. creazione di un formario di veneto comprendente tutte le forme estratte dalle radici dei verbi regolari e dalla gran parte di nomi e aggettivi;
2. creazione di un lessico di traduzione con lemmi specificati dal punto di vista morfosintattico per un totale di 40mila entrate;
3. l'attività di una ricercatrice esperta in traduzione dall'inglese al veneto è stata dedicata alla predisposizione di un testo parallelo estratto dalla storia breve intitolata "The Little Prince" (El Principe Picinin);
4. abbiamo dedicato l'attività di più ricercatori alla creazione del prototipo di traduttore automatico fondato sui sistemi MOSES e GIZA e alla sua messa a disposizione su sito web.
 - Presentiamo il Language Model ottenuto sulla base di 120mila tokens di testi allineati.
 - Le prime traduzioni ottenute automaticamente
5. infine, abbiamo dedicato molto tempo a pubblicizzare il lavoro compiuto, presentando una relazione a conferenze internazionali e interagendo in questo modo con ricercatori che operano nello stesso campo, magari sulla base di lingue come l'arabo che si presentano dal punto di vista scientifico in maniera alquanto simile al veneto.

2. Problemi affrontati

Il problema fondamentale affrontato in questo secondo periodo di ricerca, è stato quello di risolvere lo scoglio della morfologia veneta e di produrre un algoritmo che applicasse regole e vincoli a una grammatica delle flessioni verbali, nominali e aggettivali regolari. Questo algoritmo aveva poi bisogno di un radiciario dei verbi regolari e dei nomi e aggettivi regolari, cosa questa resa possibile dai dizionari messi a disposizione da Pizzati. Su questi dizionari è stato dapprima applicato un algoritmo di derivazione automatico e successivamente il risultato è stato verificato manualmente.

Altro problema affrontato è stato quello dell'inquadramento teorico della sintassi del veneto dal punto di vista computazionale. Questo lavoro è stato reso necessario per la presentazione del progetto in sede internazionale, dove era indispensabile confrontare il veneto con le altre lingue maggiori e minori. In particolare, dal punto di vista strettamente funzionale al discorso del traduttore automatico, la presenza di 5/7 varianti ortografiche/morfologiche/sintattiche e lessicali rende il Veneto molto interessante nel contesto della traduzione automatica.

Contemporaneamente è iniziata una verifica manuale del livello di ambiguità della lingua veneta ai fini della sua analisi automatica per mezzo di strumenti di tipo linguistico computazionale. Mi sto riferendo ai tagger, che servono per annotare con etichette grammaticali le parole di una frase o un testo; ai parser, che sulla base della etichettatura prodotta automaticamente da un tagger, possono costruire una analisi

sintattica e logica a livello di frase. Entrambi questi strumenti sono essenziali per migliorare il funzionamento del traduttore automatico in quanto permetterebbero di ridurre l'ambiguità e quindi allineando testi inglese ugualmente taggati e analizzati sintatticamente sarebbe possibile pensare a una traduzione per interlingua o ibrida transfer/interlingua che possa indirizzare pezzi di frase sulla base del loro significato.

Gli studi compiuti, che saranno travasati in risultati concreti nelle prossime attività, hanno già mostrato che l'ambiguità si concentra su un numero molto limitato di parole e che quindi sarà possibile creare un tagger e un parser del veneto molto puntuale. Un esempio per tutti è costituito da “i” che può essere interpretato,

i = art

i = clit

sia come articolo che come pronome clítico. Un altro caso tipico è

me = poss

me = clit

“me” un aggettivo possessivo ma anche un clítico. La lista include una trentina di forme rispetto alle quali sarà indispensabile intervenire anche con riferimenti ortografici certi per gli utenti: ad esempio nel caso di

‘l = clit

l’ = art

“l” che può svolgere due funzioni, una di articolo e l'altra di clítico. Ma nel primo caso si tratta di una contrazione, come in “no’l vol far silensio”; invece nel secondo caso si tratta di una forma di elisione simile all'italiano. Non ci sono “l” solitarie e quindi l'utente del traduttore dovrà imparare a distinguere.

Vedremo come risolvere quindi il problema dell'interfaccia, visto che il sistema non si presta ad essere utilizzato a utenti completamente naive, o non disponibili a una breve fase di “addestramento” con la lingua veneta.