

La Traduzione Automatica si è evoluta insieme alle diverse tipologie di applicazioni di Traduzione Assistita e sono stati raggiunti notevoli progressi nel miglioramento della qualità delle traduzioni prodotte da questi sistemi.

Tuttavia, nonostante i recenti sviluppi positivi nell'ambito delle tecnologie per la traduzione, non tutti i problemi sono stati risolti ed in particolare l'identificazione, interpretazione e traduzione delle cosiddette polirematiche, ovvero di quegli elementi lessicali costituiti da più di una parola come ad esempio anima gemella, carta di credito, acqua e sapone, che hanno una particolare coesione strutturale e semantica interna, rappresenta ancora una sfida aperta, sia da un punto di vista teorico che pratico.

La scadente qualità dell'analisi e traduzione di queste unità lessicali nell'ambito delle tecnologie per la traduzione ed in particolare della traduzione automatica indica che c'è ancora la necessità di investire in ulteriore ricerca allo scopo di migliorare le prestazioni delle diverse applicazioni per la traduzione.

Le polirematiche rappresentano un fenomeno linguistico complesso, che spazia da unità lessicali con una relativa variabilità di co-occorrenza delle parole a espressioni fisse o semi-fisse. Tali unità sono molto frequenti sia nel linguaggio di tutti i giorni che nelle lingue per scopi speciali. La loro interpretazione e traduzione presenta talvolta ostacoli inaspettati anche per i traduttori umani, soprattutto a causa di intrinseche ambiguità, di asimmetrie strutturali e lessicali tra lingue ed infine di differenze culturali.

Un approccio efficace al problema deve tener conto dei seguenti aspetti: (i) le polirematiche hanno diversi gradi di composizionalità e, in diversi casi, significati opachi; (ii) la traduzione delle polirematiche è talvolta imprevedibile e una traduzione parola-per-parola può produrre gravi errori; infine, (iii) le loro proprietà morfosintattiche consentono, in alcuni casi, un certo numero di variazioni formali con la possibilità di dipendenze di elementi anche se distanti tra loro all'interno di una frase.

Le attuali tendenze teoriche su questo argomento riguardano tecniche e formalismi diversi, rilevanti per il trattamento delle polirematiche in traduzione automatica, così come anche per altre applicazioni per la traduzione, come ad esempio: il riconoscimento automatico delle polirematiche in contesti monolingui e bilingui, metodologie di allineamento e parafrasi, sviluppo e usabilità di risorse linguistiche monolingui e bilingui e grammatiche sviluppate manualmente; uso delle polirematiche nella traduzione automatica di tipo statistico per scopi di adattamento al dominio, così come ricerche di tipo empirico che riguardano l'accuratezza del modello e l'adeguatezza descrittiva tra varie lingue.

A livello pratico, la questione delle polirematiche è stata affrontata nell'ambito dei diversi approcci alla traduzione automatica: si tratta infatti di una questione di cruciale importanza sia per i sistemi basati su conoscenze, sia per quelli di tipo statistico (word-based, phrase-based o factored-based) nonché per i nuovi sistemi ibridi.

Benché la traduzione delle polirematiche sia un problema noto fin dagli albori della traduzione automatica, rimane ancora irrisolto e dunque la ricerca su questo argomento è suscettibile ancora di possibili significativi miglioramenti.

Recentemente si registra una crescente attenzione verso il trattamento delle polirematiche nell'ambito della traduzione automatica e delle tecnologie per la traduzione, essendo stato riconosciuto che non è possibile sviluppare applicazioni su vasta scala senza affrontare in maniera adeguata questo problema.

La presente dissertazione, basata sui principi teorici e metodologici della teoria del Lessico-Grammatica, si propone di analizzare quest'area critica della traduzione automatica e presenta un lavoro di ricerca fondato su un'analisi linguistica contrastiva inglese-italiano relativa ai diversi tipi di polirematiche, confrontando i diversi approcci utilizzati per risolvere le difficoltà poste dal trattamento di questo particolare fenomeno lessicale in traduzione automatica.

Il risultato di questa ricerca è rappresentato dallo sviluppo di una strategia di trattamento computazionale delle diverse forme di polirematiche che utilizza fondamentalmente due diversi tipi di risorse: un dizionario bilingue Inglese-Italiano delle polirematiche e un insieme di grammatiche locali per l'identificazione e la traduzione delle stesse.

Tutte le informazioni linguistiche sono state sviluppate con l'ambiente per il Trattamento Automatico del Linguaggio (TAL) NooJ NLP.