

Traduzione automatica e traduzione assistita

Un argomento di attualità nella società dell'informazione

Carmelo Notaro

Istituto superiore di sanità
Settore documentazione
notaro@iss.it

L'Unione Europea è giunta al sesto allargamento: 27 stati membri. Internet è diffuso su tutta la terra e siamo tutti dentro un "villaggio globale" sempre più grande. In ambito bibliotecario il progetto Minerva mira a rendere il Web "accessibile" e di qualità. Il progetto Michael, finanziato dalla Commissione Europea, promuove un Web multilingue per il superamento delle barriere linguistiche, almeno nell'ambito comunitario, che pubblicizzi il patrimonio culturale europeo, le collezioni digitali, i patrimoni dei musei, delle biblioteche e degli archivi. Il tema della traduzione dei contenuti dei siti Web nelle varie lingue è dunque di grande attualità.

Vediamo un po' di storia. Una delle prime applicazioni non numeriche dei calcolatori negli anni Cinquanta è stata la traduzione automatica. Questa consiste nel passare da un linguaggio di origine a un linguaggio di destinazione mantenendo il più possibile inalterato il significato del testo e lo stile.

Le parole delle frasi del linguaggio di origine sono tradotte in quelle del linguaggio oggetto, con l'aiuto di un dizionario elettronico. Il vocabolario è l'applicazione fondamentale di tutti gli applicativi di traduzione. Le parole tradotte sono riorganizzate e sistemate tramite un semplice modello delle strutture delle due lingue. Quest'approccio dà come risultato una "traduzione grezza", cioè spesso un groviglio di parole e verbi. È famoso il caso di una frase della Bibbia, "Lo spiri-

to è forte ma la carne debole", che il traduttore automatico dall'inglese al russo rese con "La vodka è forte ma la carne guasta".

A metà degli anni Sessanta, la valutazione dei risultati ottenuti, praticamente inutilizzabili nonostante l'impegno profuso a livello mondiale, ha portato ad un atteggiamento generale di rinuncia.

Nel 1966 l'ALPAC (Automatic Language Processing Advisory Committee) pubblicò un rapporto in cui si affermava che la traduzione professionale, ovvero quella umana, era più veloce, più precisa e costava circa la metà della traduzione automatica. Si arrivò in breve al taglio dei fondi per questo settore di ricerca.

Per riuscire a fare un buon lavoro di traduzione è necessario prima comprendere il significato del testo. Una comprensione profonda. La traduzione non è un processo di codifica e decodifica passiva che implica il semplice passaggio da un codice a un altro, ma al contrario comporta un processo di comprensione strutturale approfondita e di riformulazione del messaggio nei termini della lingua oggetto. Questa considerazione ha portato a sviluppare un modello di traduzione automatica, chiamato *interlingua*, dove si passa per una sorta di "esperanto informatico", una lingua di mezzo, e da questa al linguaggio destinazione. A differenza del modello iniziale, detto *transfer* o *transformer*, quest'architettura, interessante dal punto di vista della

ricerca, non ha fornito applicazioni usabili praticamente.

Del resto basta andare su uno dei molti siti di traduzione in Internet (strumenti per le lingue di Google, Freetranslation, Systran ecc.) per farsi un'idea personale, con lingue ben conosciute, delle possibilità attuali dei traduttori automatici.

Nei giochi come scacchi, dama, othello, backgammon o scarabeo i progressi del settore sono stati continuativi e fantastici. Nel 1979 un programma chiamato Mighty Bee ha sconfitto il campione mondiale di backgammon. Un programma di dama, CHINOOK, realizzato da Jonathan Schaeffer, è stato in grado di tenere testa ai giocatori più forti del mondo. Per far questo Schaeffer ha memorizzato su calcolatore migliaia di partite giocate dai grandi campioni e ha immagazzinato 500 trilioni (5 con 20 zeri...) di posizioni. Iscritto nel 1990-92 a un torneo con giocatori umani, ha avuto esiti incerti, ma dopo vari aggiustamenti nel 1994 ha trionfato, umiliando l'uno dietro l'altro tutti i campioni in gara, diventando il vincitore di quell'anno. Nel campionato mondiale del 1996 è stato chiaro che il giocatore virtuale era più forte di qualsiasi umano.

Nel febbraio 1996 un supercomputer dell'IBM progettato appositamente per il gioco degli scacchi, Deep Blue, ha sfidato in torneo il campione mondiale in carica Garry Kasparov. Il computer vinse la prima partita, ma nel seguito Kaspa-

rov si impose vincendo con un risultato di 4-2.

L'anno dopo, nel maggio 1997, il team di programmazione IBM ha aumentato la profondità e la potenza di calcolo e Deep Blue ha sconfitto Garry Kasparov per 3.5-2.5. Kasparov disse che alcune volte gli era parso di notare un'intelligenza e una creatività così profonde nelle mosse della macchina, che non riusciva a comprenderle. Avanzò anche il sospetto che la macchina avesse avuto un "aiuto" umano durante la partita. Chiese la rivincita, ma l'IBM rifiutò e ritirò Deep Blue. Con la vittoria di Deep Blue su

Kasparov l'interesse per i programmi di scacchi è diminuito e sostituito dalla ricerca in programmi per altri giochi e altre applicazioni.

Successi analoghi si sono avuti anche nel gioco dello scarabeo e dell'othello. Non si sono riscontrati invece nel caso della traduzione automatica dove, tuttora, il migliore dei programmi non raggiunge la competenza linguistica di un bambino di otto anni. Ci si chiederà: i programmi che giocano, "capiscono" il gioco? La risposta è articolata: un programma di scacchi (o di dama ecc.) assegna un

valore numerico ai pezzi e alle posizioni, quindi la scelta della mossa nel cuore del gioco è ricondotta ad un calcolo numerico. Deep Blue è in grado di calcolare più di un miliardo di mosse al secondo, ed è dotato per giunta di un numero enorme di aperture e di finali di partite, giocate nei più importanti tornei di ogni epoca e luogo. Nei finali la memoria elettronica dell'elaboratore, a differenza di quella umana, è "esatta" e il

computer risulta praticamente imbattibile. I successi dei computer nei giochi si possono ricondurre alla superiorità di calcolo rispetto a qualunque essere umano – è come sfidare una calcolatrice in un calcolo a dieci decimali – e dalla memoria "formidabile e infallibile" degli stessi.

Nel frattempo, con lo sviluppo di Internet, la domanda di traduzioni si è impennata al punto da non potere essere soddisfatta da traduttori in carne ed ossa. Uno sviluppo interessante, un certo nuovo impulso ai traduttori automatici è avvenuto negli anni 2000, con l'introduzione

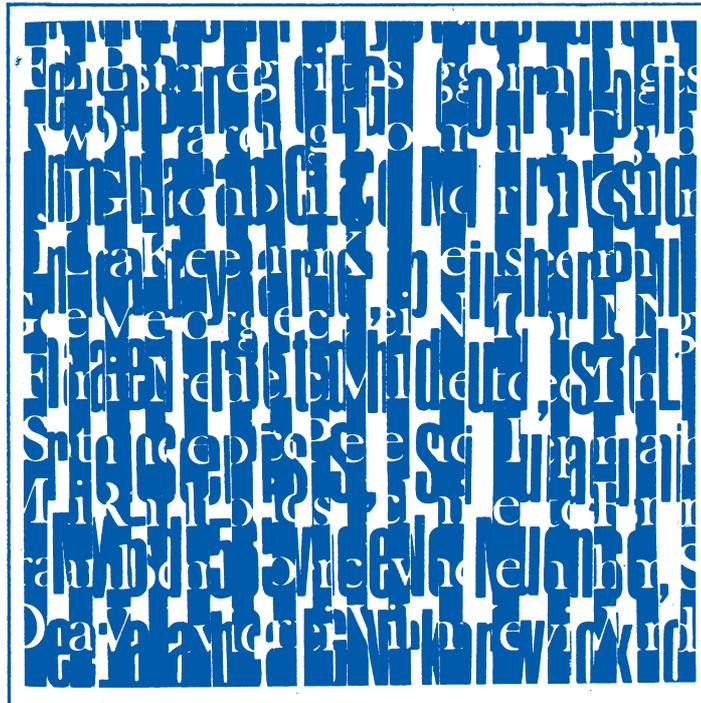
di testi piani. Se poi il testo ha idiomi o giochi di parole, è la fine; per esempio "to drop a brick", fare una gaffe, diventa automaticamente "lasciar cadere un mattone"; sembra facile inserire una simile conoscenza in un programma, ma non è così. Lo stesso per la letteratura, la poesia, i giochi di parole che mettono in difficoltà anche i traduttori professionisti. Umberto Eco riporta l'esempio della semplice frase "the fans of the Pope" trasformato dal traduttore automatico in "i ventilatori del Papa".

Del tutto diversa dalla traduzione automatica è la cosiddetta CAT, ac-

cronimo di Computer-Assisted Translation o Computer-Aided Translation o, più italianamente, Traduzione Assistita (TA). Nella TA la traduzione è sviluppata da un traduttore in carne ed ossa e il software costituisce un sistema avanzato di videoscrittura che lo assiste. L'assistenza si basa sulla costruzione progressiva di una banca dati di termini specialistici, in continua espansione, e sulla comparazione dei testi già tradotti. Questo meccanismo gioca soprattutto sull'elevata ricorrenza della terminologia set-

toriale che caratterizza i testi tecnici; di conseguenza la TA ottiene ottimi risultati nel settore della traduzione tecnica. Le applicazioni di TA oggi sono più diffuse sono TRADOS (TRADOS Corp.), Transit (Star Group) e WordFast (Champollion Ltd.).

Attualmente (2008), le lingue ufficiali dell'Unione Europea sono 23, e precisamente: ceco (Repubblica Ceca), danese (Danimarca), estone (Estonia), finlandese (Finlandia), francese (Francia, Belgio, Lussemburgo), greco (Grecia, Cipro), inglese



e l'uso di metodi statistici. Avere un archivio di testi già tradotti aiuta a ripescare frasi utili per una buona traduzione. La necessità di migliorare l'efficienza del software e ottenere una buona traduzione è enorme: si deve inserire l'equivalente di un milione di libri. Ciò nonostante i programmi traducono diversamente dai traduttori professionisti, non hanno "comprensione" del significato del testo e il meccanismo di traduzione resta "superficiale" e la traduzione "grezza": anche in caso

(Regno Unito, Irlanda), italiano (Italia), lettone (Lettonia), lituano (Lituania), maltese (Malta), nederlandese, o neerlandese (Paesi Bassi e Fiandre, in Belgio), polacco (Polonia), portoghese (Portogallo), slovacco (Slovacchia), sloveno (Slovenia), spagnolo (Spagna), svedese (Svezia), tedesco (Germania, Austria, Lussemburgo, Alto Adige/Südtirol, in Italia, Belgio), ungherese (Ungheria), gaelico irlandese (Irlanda), bulgaro (Bulgaria), rumeno (Romania).

È comprensibile quindi che il rispetto di tutte le lingue costituisca un serio problema di comunicazione. È in atto la tendenza a considerare quali lingue di redazione degli atti formali solo tre principali lingue europee (l'inglese, il francese e il tedesco) pubblicandone poi la traduzione nelle restanti. Anche il progetto Michael in una prima fase prevedeva solo tre lingue: francese, italiano e inglese. Nella letteratura tecnica e per la traduzione anche grossolana e semplicemente orientativa del senso dei documenti il DGT (Directorate-General for Translation of the European Commission) propone come traduttore automatico l'ECMT (European Commission Machine Translation) che offre la traduzione per 18 coppie di linguaggi, con tutti i limiti sopra menzionati, e come assistente traduttore lo SDL TRADOS (TWB). Propone anche altri ausili linguistici a livello centrale e locale.

Nel 2007 l'ammontare dei costi di traduzione alla Commissione Europea è stato di circa 300 milioni di euro. Nel sostenere il multilinguismo la Commissione Europea ha pubblicato un database unico nel suo genere, che rende gratuitamente accessibile al pubblico 1 milione di frasi tradotte in 22 delle 23 lingue ufficiali dell'UE (manca soltanto il gaelico). Si tratta di una *translation memory* composta di brevi frasi (vedi <[http://langtech.jrc.](http://langtech.jrc.it/DGT-TM.html)

[it/DGT-TM.html](http://langtech.jrc.it/DGT-TM.html)>) che possono essere usate per diversi scopi; il DGT le usa nell'ottica della traduzione assistita, gestita centralmente e accessibile tramite Euramis (European Advanced Multilingual Information System). È un complesso di applicazioni client-server cui si accede tramite posta elettronica. Il sistema ha un'interfaccia Web che consente di selezionare le richieste (istruzioni e file da elaborare) che vengono spedite via e-mail; analizzate ed elaborate a livello centrale, sono evase rispondendo al mittente, sempre via e-mail, il risultato finale. Dà un accesso indiretto più specifico a tutta una serie di applicazioni e servizi linguistici. Si tratta di un progetto avviato nel 1995 basato sull'idea di affidare meno lavoro ripetitivo ai traduttori e di aumentare al tempo stesso la coerenza linguistica e metodologica, contribuendo a una migliore qualità delle traduzioni. I traduttori della commissione usano un software di riconoscimento vocale, il Dragon Naturally Speaking, che permette di dettare direttamente al computer in modo naturale con elevato grado di precisione (fino al 98%) e grande efficacia. Il programma, disponibile per il francese, inglese, italiano, olandese, spagnolo e tedesco, permette ai traduttori di risparmiare molto tempo, dato che non devono più dattilografare una parte rilevante del loro lavoro, né farlo dattilografare.

Lo IATE (Inter Active Terminology for Europe) è un database interistituzionale di terminologia che è diventato pienamente operativo all'interno della Commissione Europea agli inizi del 2005; offre un servizio di terminologia per tutte le istituzioni europee. Quest (one-stop access to a series of general-interest terminology databases) è un'applicazione basata su Internet sviluppata dal DGT; non si tratta di un database terminologico ma

piuttosto di un'interfaccia di ricerca che interroga simultaneamente più database.

TRADOS è un programma di traduzione assistita che aiuta a risparmiare tempo e velocizzare il processo di traduzione. Per fare ciò usa un database di memoria le cui dimensioni aumentano di pari passo alla quantità di traduzioni svolte. Ciò significa che, specialmente nel caso di testi tecnici, è possibile recuperare frasi, termini, segmenti già tradotti in precedenza. TRADOS è ovviamente molto meno utile in caso di testi non tecnici, in cui la percentuale di ripetizioni è scarsa o nulla.

Il pacchetto di TRADOS si compone di altre applicazioni, ognuna delle quali serve a una funzionalità specifica. Le principali sono: Translator's Workbench (TWB), WinAlign, MultiTerm, Tag Editor. Translator's Workbench è l'applicazione principale di TRADOS, cioè un ambiente di lavoro che consente di operare con Microsoft Word, creare memorie e altro. È un ambiente di traduzione interattiva, in cui vengono proposte automaticamente frasi simili a quella da tradurre, pescando da quelle già tradotte.

Sulla terra ci sono oltre 7.000 lingue parlate. Nell'Unione Europea, oltre alle 23 lingue ufficiali, considerando anche i territori vicini, ci sono altre 60 lingue, come l'arabo, il cinese, il vietnamita ecc. È facile capire che già soltanto riconoscere di che lingua si tratti, anche solo all'interno dell'UE, sia un vero problema. Per esempio, come distinguere il danese dal norvegese o avere un'idea di cosa dica un testo in queste o altre lingue per chi non le conosce affatto? Ecco che, uscita dalla porta, ritorna dalla finestra non solo la necessità di un traduttore automatico, ma anche di un riconoscitore d'idiomi, cioè un programma che a partire da un testo ci sappia dire di che lingua si tratti (vedi <[http://www.yourdictionary.](http://www.yourdictionary.it)

com/morph.html#guessers>). Questo naturalmente anche ai fini dell'archiviazione e indicizzazione dei testi stessi. Non pensiamo solo all'inglese, francese, spagnolo o tedesco, molto diffuse all'interno dell'UE, ma a lingue come il danese o il norvegese, il greco e il turco: allora, per quanto elementare possa essere, avere inserito nel navigatore satellitare, in palmo di mano, un programma per "tradurre" e "pronunciare" le frasi più semplici, in una lingua sconosciuta, non è comunque impagabile?

Bibliografia

- CLAUDIO CIRILLI, *La traduzione automatica tra utopia e realtà*, "AIDA informazioni", 1 (1998), p. 3-5.
- JOHN HUTCHINS, *The History of Machine Translation in a Nutshell*, <<http://www.hutchinsweb.me.uk/Nutshell-2005.pdf>>.
- ALESSANDRO LONGO, *Se lo spirito si tra-*
- sforma in vodka. Quante follie nei traduttori web*, <http://www.repubblica.it/2008/01/sezioni/scuola_e_universita/servizi/traduttori-ue/traduttori-ue/traduttori-ue.html>.
- MINISTERO PER I BENI E LE ATTIVITÀ CULTURALI. PROGETTO MINERVA, *Manuale per la qualità dei siti web pubblici culturali*, a cura di Fedora Filippi, 2. ed. it. aggiornata, Roma, 2005.
- Michael (Multilingual Inventory of Cultural Heritage in Europe)*, brochure in italiano, <<http://www.michael-culture.org>>.
- LINDA ROSSI GOLDEN, *Un milione di frasi nel database web. È il regalo della commissione UE*, <http://www.repubblica.it/2008/01/sezioni/scuola_e_universita/servizi/traduttori-ue/frasi-ue/frasi-ue.html>.
- STUART J. RUSSELL – PETER NORVIG, *Intelligenza artificiale: un approccio moderno*, Torino, UTET Libreria, 1998.
- GARY STIX, *Il sogno sfuggente della traduzione automatica*, "Le Scienze", 486 (2006), agosto, p. 106-110.
- DAVID L. WALTZ, *Intelligenza artificiale*, "Le Scienze", 172 (1982), dicembre, p. 62-78.

WOLFRAM WILSS, *Translation Studies – The state of the Art*, "META", 49(2004), 4, p. 777-785.

Traslation tools and Workflow, Luxembourg, Office for Official Publications of the European Communities, 2007, <http://ec.europa.eu/dgs/translation/index_en.htm>.

Abstract

The computer has transformed many aspects of our life. The progress, reached in the field of games, where computers have beaten the best human champions in dama, chess, backgammon, etc., has not been achieved yet in the field of automatic translation. In fact at the moment good automatic translator doesn't still exist. However, today there are many tools, such as dictionaries, translation memories, speech synthesizer, etc., which are a precious help for all of us.