

Traduzione automatica

di ISABELLA CHIARI

L'idea di usare delle macchine per fare traduzioni è molto antica: risale ai secoli Seicento e Settecento, secoli di fiducia quasi incondizionata nel potere del calcolo, delle applicazioni matematiche e infine della ragione. Allora si pensava a dizionari più che a traduzioni vere e proprie; certamente non si sarebbe mai immaginato quanto rapide e accessibili le traduzioni automatiche sarebbero diventate alcuni secoli dopo

Fino a pochi anni fa (sì, perché la tecnologia si muove a una velocità molto più rapida delle scienze e della ricerca di base) le speranze erano considerate piuttosto illusorie. La storia della traduzione automatica (TA, o nella versione anglosassone MT, *machine translation*) è fatta di alti e bassi, di grandi aspirazioni e di fortissime disillusioni, dolenti insuccessi e amare frustrazioni.

Il primo ad avere avuto l'idea di considerare il problema della traduzione come un problema di codifica, e dunque un problema automatizzabile, è stato il matematico Warren Weaver nel 1947. Dopo i primi tentativi realizzati negli anni Cinquanta si è arrivati persino, nel 1967, a dichiarare ufficialmente fallito il programma di ricerca in traduzione automatica in una comunicazione dell'ALPAC (Automatic Language Processing Advisory Committee) rimasta ormai famosa. Nella seconda metà degli anni Settanta i progressi dell'Intelligenza artificiale e soprattutto del *Natural Language Processing* (Trattamento Automatico del Linguaggio) danno nuova linfa ai progetti di traduzione automatica.

Ma, come spesso accade, è la pressante richiesta del mercato a determinare la concentrazione degli sforzi verso nuove direzioni. Lo scambio di documenti e informazioni è oggi sempre più diffuso, anche tra comunità distanti geograficamente e linguisticamente. I problemi della traduzione e dell'interpretariato assumono dunque un carattere centrale per la vita collettiva, e comportano altissimi costi anche in termini finanziari. Sono i grandi organismi internazionali come l'Unione Europea e la Nato, oltre alle grandi multinazionali come Xerox, IBM, General Motors, a richiedere una quantità prima inimmaginabile di traduzioni in un numero enorme di coppie di lingue. Sono dunque la necessità di traduzione di documentazione istituzionale e la necessità di adattamento e traduzione di materiali quali istruzioni per l'uso, etichette e *software* (quella che si chiama oggi *localizzazione*) a richiedere uno sforzo finanziario imponente per lo sviluppo di *software*, commerciali e non, che facilitino o sostituiscano gli enormi costi del ricorso a traduttori professionisti.

Si può dire oggi che lo sforzo è stato ampiamente ripagato



dai contributi recenti della linguistica computazionale che hanno prodotto uno scenario totalmente nuovo nel settore della traduzione e hanno indotto un profondo ripensamento dei meccanismi di produzione e di mercato in questa area.

Esistono, semplificando un po', due principali tipologie di traduzione automatica: la prima, la cosiddetta traduzione automatica basata su regole, si fonda sull'idea che uno strumento, per poter svolgere un compito di traduzione, deve possedere un dizionario e una grammatica delle lingue da trattare i più ampi possibile e una serie di regole di conversione che permettano di passare da una frase in una lingua di partenza (*source language*) a una frase nella lingua di arrivo corrispondente (*target language*).

La seconda tipologia, quella che ha rivoluzionato il panorama delle traduzioni negli ultimissimi anni, è la traduzione automatica statistica (*statistical machine translation*). Questo tipo di sistema usa pochissima



Warren Weaver (1894-1978)



Uno dei traduttori online più famosi e utilizzati: **Google Translate**

informazione sulla grammatica e spesso non possiede nemmeno un dizionario. L'idea su cui si fonda è piuttosto semplice, ma incredibilmente produttiva: se noi disponiamo di una grande raccolta di testi in originale e in traduzione per una coppia di lingue (il cosiddetto *corpus parallelo*) possiamo estrarre in modo automatico le corrispondenze più frequenti tra 'pezzettini' o segmenti di frasi. Insieme alle corrispondenze o equivalenze estrarremo anche la probabilità che un dato segmento in certi contesti venga tradotto in un modo piuttosto che un altro.

Questa fase di estrazione delle regolarità statistiche (e di esempi) da un *corpus* parallelo si chiama *fase di addestramento* del motore di traduzione.

La seconda fase sarà il trattamento di materiali nuovi: al motore di traduzione saranno presentati nuovi testi nella lingua di partenza e, mediante un confronto con i dati estratti dal *corpus parallelo*, il sistema proporrà una possibile traduzione di nuovi segmenti di frase, confrontandoli per forme e per strutture a quelli trovati nel *corpus*.

Le equivalenze proposte da un modello statistico di traduzione generalmente sono più plausibili di quelle prodotte da sistemi basati su regole, perché sono basate su traduzioni reali e di qualità prodotte da traduttori professionisti e non su costruzioni di corrispondenze astratte. Ciò avviene perché, in sostanza, un sistema di traduzione automatica statistica, invece di usare una grammatica che deve essere descritta e formalizzata completamente da gruppi di linguisti, fa in modo di "scoprire" le regolarità della grammatica direttamente dai testi. Così, non solo il processo è più rapido e più completo, ma richiede molto meno lavoro da parte del linguista.

Quasi tutti noi, prima o poi, abbiamo provato a usare il sistema di traduzione del famoso motore di ricerca Google (*Google translate*) che traduce tra coppie qualunque di ben 57 lingue diverse in pochi secondi.

Prima del 2007, Google utilizzava un sistema di traduzione automatica basato su regole chiamato Systran, mentre oggi si è "convertito" a un sistema statistico che ne ha fatto un leader nelle traduzioni gratuite *online*.

La qualità però dipende dalle lingue prese singolarmente e nella coppia di traduzione. Ciò perché i testi paralleli che sono usati per addestrare il sistema statistico dipendono dalla disponibilità di testi digitali in tali lingue. Se per l'inglese abbinato a lingue importanti come il francese, il tedesco, lo spagnolo, l'italiano si trovano molti testi, meno se ne trovano per accoppiare l'inglese o l'italiano al persiano, all'armeno, al finlandese, al basco, al macedone o al latino (!).

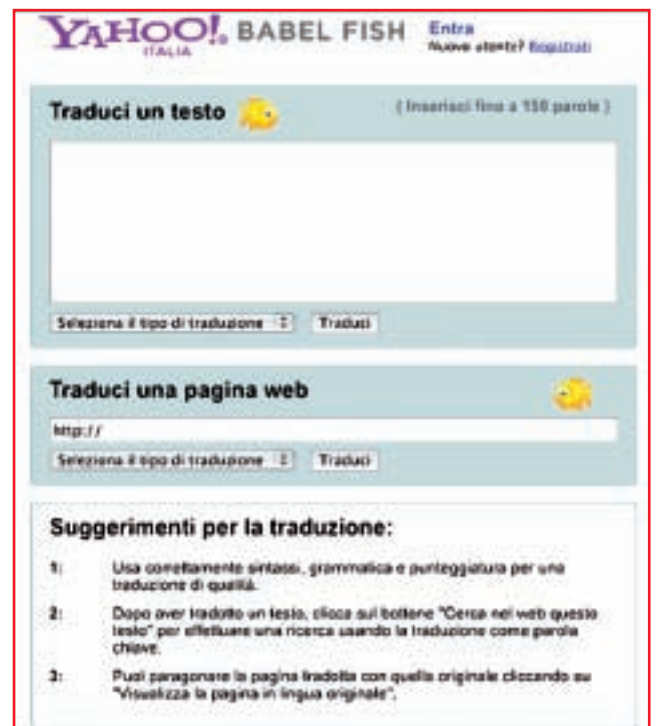
Come potete vedere più avanti, anche il tipo di testo è una variabile rilevante. Mentre testi di tipo tecnico o scientifico, descrittivo o giornalistico sono trattati con

una certa accuratezza, i testi più creativi, narrativi e letterari, costituiscono ancora un grosso ostacolo per gli strumenti automatici. E probabilmente costituiranno sempre un limite.

È comunque impossibile non riconoscere oggi il successo delle applicazioni automatiche per usi generali: capire cosa c'è scritto in una pagina *web*, leggere il messaggio di un forum, leggere istruzioni e altri documenti. Tale riconoscimento è confermato dalla rivoluzione che di fatto ha cambiato il mercato della traduzione e soprattutto il lavoro quotidiano dei traduttori professionisti, che devono offrire qualità e professionalità diverse rispetto a quelle garantite dai motori automatici.

DUE ESPERIMENTI DI TRADUZIONE AUTOMATICA ONLINE

Per vedere il funzionamento, buono o cattivo che sia, dei sistemi basati su regole e dei sistemi statistici possiamo fare una piccola prova usando il sito BabelFish (<http://babelfish.yahoo.com/>) che usa Systran che è un sistema a regole e usando Google Translate (<http://translate.google.it/>) che usa la statistica.



Un altro celebre traduttore online: **Babel Fish**

Diamo in *input* a entrambi un testo italiano e proviamo a osservare la traduzione inglese e a partire da quella una riconversione in italiano dello stesso testo. A voi il giudizio!

	Testo italiano	Traduzione in inglese	Traduzione dell'inglese in italiano
YAHOO BABELFISH	La traduzione automatica è un'area della linguistica computazionale che studia la traduzione di testi da un linguaggio naturale ad un altro mediante programmi informatici.	The automatic translation is un'area of the computazionale linguistica that studies the translation of witnesses from a natural language to another by means of programs informed to us .	La traduzione automatica è un'zona del linguistica del computazionale che studia la traduzione di testimoni da un del linguaggio naturale ad un altra per mezzo di programmi informati a noi .
GOOGLE TRADUTTORE	La traduzione automatica è un'area della linguistica computazionale che studia la traduzione di testi da un linguaggio naturale ad un altro mediante programmi informatici.	Machine translation is an area of computational linguistics that studies the translation of texts from one natural language to another by computer programs.	La traduzione automatica è un'area della linguistica computazionale che studia la traduzione di testi da una lingua naturale a un altro da programmi informatici.

Forse aveva ragione Jorge Luis Borges quando diceva che "L'originale è infedele nei confronti della traduzione".

CREATIVITÀ E AMBIGUITÀ ALLA FRONTIERA

Ma dove vanno a finire i traduttori in carne e ossa? Il mestiere del traduttore è profondamente cambiato negli ultimi dieci anni. Oggi un traduttore non può fare a meno dell'uso di strumenti informatici che lo assistano nella traduzione. Rimane sua prerogativa affrontare testi di tipo soprattutto letterario.

Gli strumenti di traduzione automatica infatti incontrano molti ostacoli derivanti da alcune caratteristiche ineliminabili delle lingue: la presenza di omografi¹ testuali e assoluti la presenza di locuzioni o espressioni complesse; collocazioni, usi ambigui o vaghi delle parole, costruzioni sintattiche intrinsecamente ambigue, usi metaforici e metonimici² delle parole, ecc.

Il poeta americano Robert Frost scriveva: "La poesia è ciò che viene perso nella traduzione". In realtà sappiamo che esistono traduzioni che riescono a rievocare la musicalità e l'espressività dell'originale. Certamente però questa operazione richiede doti interpretative e creative di altissimo livello. Proprio per questo non è immaginabile applicare strumenti come Google Translate a un testo letterario.

E come omaggio proprio a Robert Frost vediamo il risultato della traduzione automatica del finale della sua famosa composizione "The road not taken" (1916):

*I shall be telling this with a sigh
Somewhere ages and ages hence:
Two roads diverged in a wood, and I-
I took the one less traveled by,
And that has made all the difference.*

E la traduzione italiana di Google Translate?

*Mi sono detto con un sospiro
Da qualche parte le età ed età quindi:
Due strade divergevano in un bosco, e io -
Io presi la meno battuta,
E che ha fatto la differenza.*

È impossibile trovare la musicalità, l'andamento e l'evocatività dell'originale inglese in questa traduzione, che tra l'altro contiene alcune 'scelte' discutibili. Una traduzione letteraria deve rispettare o, se necessario, reinventare, l'atmosfera del testo originale come si vede leggendo la traduzione di Giovanni Giudici (R. Frost, *Conoscenza della notte e altre poesie*, trad. di Giovanni Giudici, Torino: Einaudi 1965):

*Io dovrò dire questo con un sospiro
in qualche posto fra molto molto tempo:
divergevano due strade in un bosco, ed io...
io presi la meno battuta,
e di qui tutta la differenza è venuta.*

Note

1. Gli **omografi** sono parole che si scrivono quasi o del tutto allo stesso modo, ma possiedono significati diversi, come ad esempio in italiano "principi", plurale di "principe" o "principi", plurale di principio, oppure "pesca", nome del frutto e terza persona singolare del verbo pescare.
2. La **metonimia** è una figura retorica che consiste nel sostituire una parola con un'altra che abbia con la prima una relazione di contiguità logica o materiale.

Isabella Chiari

Laureatasi in Filosofia presso l'Università "La Sapienza" di Roma, dopo aver conseguito il dottorato di ricerca in Filosofia del linguaggio presso l'Università di Palermo, è ora ricercatrice in Glottologia e Linguistica presso "La Sapienza". I suoi interessi di ricerca vanno dall'intersezione tra linguistica e filosofia del linguaggio all'approfondimento delle basi scientifiche, metodologiche e teoriche della linguistica quantitativa e dei fenomeni di comprensione e incomprensione e delle loro conseguenze sull'insegnamento e sull'apprendimento della lingua. Si occupa inoltre, dal punto di vista metodologico, di strumenti computazionali e statistici per l'analisi testuale. È autrice di numerosi saggi e articoli.

