

PROCESSI
TRADUTTIVI:
TEORIE
ED APPLICAZIONI

Estratti

EDITRICE LA SCUOLA

TRADUZIONE E LINGUISTICA COMPUTAZIONALE

di Antonio Zampolli

Articoli, seminari, convegni e in genere discussioni nel settore della cosiddetta 'traduzione automatica' si incentrano di solito su uno dei seguenti argomenti:

- Esigenze del mercato della traduzione; risorse disponibili; costi della traduzione umana; necessità, dal punto di vista organizzativo ed economico, di ricorrere all'impiego di sistemi di elaborazione elettronica dei dati; valutazione dei risultati e dei costi dei sistemi attualmente disponibili o dei quali si propone la costruzione;
- Fattibilità della traduzione automatica (d'ora in poi TA) in confronto con la traduzione assistita dal calcolatore (*Computer Assisted Translation: CAT*).

Nella TA si cerca di far eseguire al calcolatore l'intero processo di traduzione, eventualmente con l'intervento umano in fase di revisione. Nella prospettiva della CAT la traduzione propriamente detta viene compiuta dal traduttore umano, mentre il calcolatore fornisce tutta una serie di supporti, che vanno dal dizionario automatico all'uso del terminale come macchina per la redazione della traduzione e delle sue successive stesure (*editing*).

- Aspetti metodologici tecnico-scientifici dei componenti che costituiscono o dovrebbero costituire i sistemi di TA e/o di CAT, e loro relazioni con le aree di ricerca alle quali sono, a vario titolo, collegati: linguistica, linguistica computazionale, linguistica applicata, informatica, ecc.

È sembrato piú consono al carattere scientifico di questo convegno limitare il discorso al terzo punto, ed in particolare ad una rapida esposizione informativa delle caratteristiche metodologiche e tecniche dei sistemi di TA fino ad oggi suggeriti o proposti, ed accennare solo di sfuggita ai primi due punti.

1. Si esaminerà brevemente il primo periodo della TA, dall'immediato dopoguerra sino al 1966, anno nel quale apparve il cosiddetto "rapporto ALPAC" (Automatic Language Processing Advisory Committee) che, con il suo pessimistico giudizio, determinò la sospensione della maggior parte dei finanziamenti che avevano permesso il moltiplicarsi, soprattutto negli USA, dei progetti di TA. Dal punto di vista dei rapporti con la linguistica, in questo periodo è da rilevare da un lato un primo ricorso all'idea degli universali linguistici, presentati come elemento a sostegno della fattibilità della TA; dall'altro l'esigenza, agli inizi confusa ma poi espressa sempre piú esplicitamente, di disporre di un apparato formale che permettesse di rappresentare adeguatamente, collegare in un insieme coerente, e rendere 'programmabili' per algoritmi di analisi e di sintesi (all'inizio in particolare per la disambiguazione degli omografi), le regolarità grammaticali.

Mentre in Europa ci si rivolse, in principio soprattutto, ai lavori di Tesnière, in America queste esigenze avrebbero trovato, piuttosto tardivamente, una risposta nelle formalizzazioni proposte dalla scuola generativista. I cultori della TA parteciparono attivamente soprattutto al dibattito sui meriti rispettivi delle diverse classi di grammatiche formali (di dipendenza, context-free, ecc.).

Dal punto di vista computazionale, è da sottolineare l'apprezzabile sviluppo cui furono condotti i sistemi per la memorizzazione e la consultazione dei 'dizionari automatici', nonché le ricerche sui parsers sintattici relativi a grammatiche di tipo context-free.

È da rilevare invece come siano sostanzialmente mancati i rapporti e gli interscambi con il settore degli spogli elettronici di testi che, negli stessi anni (a partire dal progetto dell'Index Thomisticus), si era venuto sviluppando in vari paesi europei, e in particolar modo in Francia, in Belgio, in Inghilterra, come supporto a ricerche di tipo lessicografico, filologico, letterario, ecc.

Questa mancanza è sorprendente se si considerano non solo le dichiarazioni programmatiche sulla necessità di una reciproca collaborazione, ma anche la sostanziale comunanza delle operazioni da compiere, degli strumenti da approntare, dei dati da acquisire (per esempio dizionari, parsers, sistemi di input/output sono componenti essenziali delle procedure sia di TA sia di spoglio).

Questa mancanza di scambi può, a mio avviso, trovare una spiegazione soprattutto nella diversa formazione di base dei ricercatori dei due settori.

2. La crisi della TA, successiva al rapporto ALPAC, non fu tuttavia senza conseguenze per quanto concerne lo sviluppo della linguistica computazionale in genere. Infatti le ricerche nel settore dei parsers sintattici e semantici, svincolate dalle pressanti scadenze realizzative che condizionavano contrattualmente i progetti di TA, assunsero caratteristiche di maggiore rilevanza teorica e generale, e la linguistica computazionale cominciò a configurarsi come vero e proprio settore disciplinare.

Mentre i pochi progetti di TA sopravvissuti continuavano, soprattutto negli USA, a svilupparsi quasi esclusivamente lungo le linee impostate a metà degli anni '50, con orientamenti decisamente applicativi e verso la ricerca di soluzioni ad hoc, la linguistica computazionale si sviluppava in tre filoni principali.

Il settore degli spogli lessicografici, letterari, filologici, quantitativi, ecc. conobbe uno straordinario sviluppo, grazie soprattutto alla introduzione della elaborazione elettronica nelle grandi imprese lessicografiche nazionali.

Nel settore delle ricerche teoriche e dei modelli linguistico computazionali, iniziò lo sviluppo, nei primi anni '70, di sistemi e metodi computazionali capaci di utilizzare, nell'analisi automatica del linguaggio, la ricchezza e la complessità di nozioni e di strumenti formali prodotti dagli studi grammaticali sotto la spinta della scuola generativo-trasformativa. Al limite si tende a costituire dei veri e propri modelli computazionali delle teorie grammaticali e dei processi psicolinguistici correlati.

D'altro lato (e il peso riconosciuto, nel determinare gli insuccessi

della TA, alle difficoltà di trattare adeguatamente le componenti semantiche e pragmatiche è una delle cause concomitanti), si è assistito al sorgere (si pensi ai lavori di T. Winograd e Y. Wilks, quest'ultimo direttamente operante nel contesto di un progetto-pilota di TA) di un nuovo filone di ricerche, che dai suoi stessi promotori viene — non so se ambiziosamente — definito come un nuovo 'paradigma scientifico' per lo studio del linguaggio, basato sulla assunzione che, dal momento che le proprietà essenziali del linguaggio riflettono la struttura cognitiva generale dell'uomo, (proprietà della struttura della memoria, strategie e limiti dei processi, ecc.) il 'focus' primario delle ricerche debba portare sui processi sottostanti la produzione e la comprensione degli enunciati nel contesto linguistico e pragmatico.

Questo 'paradigma' non rifiuta tuttavia l'assunzione di concetti e strumenti delle correnti contemporanee della linguistica generale: si pensi all'impiego dei primitivi semantici, della teoria dei casi, della text-theory, ecc.

Si formano anche alcuni gruppi misti di ricerca, soprattutto in California, favoriti tra l'altro dallo svolgimento delle Scuole Estive di Pisa del '74 e del '77.

3. È in questo contesto che possono essere collocati i progetti di TA che sono stati riproposti in questi ultimissimi anni.

Per quanto essi siano stati determinati essenzialmente da esigenze 'pratiche' e 'politiche', essi possono indubbiamente beneficiare di alcune importanti nuove favorevoli 'condizioni', create dagli sviluppi della linguistica, della linguistica computazionale, dell'informatica. Ne segnaliamo alcune a titolo di esempio:

- Il grado di sviluppo raggiunto dalle metodologie di parsing, soprattutto a livello sintattico.
- Nuove acquisizioni metodologiche nel settore della rappresentazione (anche computazionale) delle strutture linguistiche, anche — sia pure limitatamente — a livello semantico.
- L'esplosione recente dell'interesse per le tecniche di 'word-processing' e di 'text-processing', per es. nell'ambito dell'office-automation, della telematica, ecc.
- Lo sviluppo delle metodologie e tecniche nel settore dei databases,

e le loro possibili applicazioni nella gestione di dizionari, banche terminologiche e, in genere, di corpora di informazioni e regole linguistiche.

- L'evoluzione delle metodologie nel settore della produzione del software, che facilitano la messa a punto e la sperimentazione dei modelli computazionali in via di sviluppo e consentono — a chi lo desidera — di far evolvere separatamente la messa a punto del componente linguistico e di quello algoritmico.
- Il notevole accumulo di conoscenze grammaticali — in varie lingue — promosso (sia pure in maniera non sistematica ed organizzata) dagli studi condotti nell'ambito generativo-trasformativo. (Come è noto, alcuni linguisti ritengono che il contributo dell'approccio generativo-trasformativo vada ricercato non tanto sul piano della validità intrinseca dei modelli proposti, quanto piuttosto nell'aver fornito strumenti operativi che si sono rivelati straordinariamente potenti al fine di mettere in evidenza fatti e fenomeni grammaticali).

Sembra inoltre che in un prossimo futuro dovrebbero entrare in questo quadro anche le recenti acquisizioni della linguistica contrastiva.