L'esperanto e la traduzione automatica: Storia, risultati e prospettive esperantologiche dell'approccio statistico

L'esperanto e la traduzione automatica: storia, risultati e prospettive esperantologiche dell'approccio statistico

Federico Gobbo
Università degli Studi dell'Insubria, Varese

1. Una lunga storia di rapporti a volte difficili

L'uso dell'esperanto nell'ambito della traduzione automatica è tutt'altro che una novità. Il primo tentativo di applicazione avvenne in Unione Sovietica, addirittura prima della costruzione del primo calcolatore moderno e prima della nascita dell'informatica. Nel 1933 infatti Petr Petrovich Trojanskij pubblicò un brevetto sovietico che descriveva una macchina per tradurre. Questa macchina prendeva in input un testo le cui parole erano state previamente trascritte in una notazione pseudo-formale, fatta di una combinazione di lessemi e "simboli logici", in termini moderni diremmo "morfemi derivazionali". I simboli logici erano ricalcati direttamente sull'esperanto: la macchina restituiva in output un testo nella lingua d'arrivo, in cui i simboli logici rimanevano invariati, mentre i lessemi erano stati sostituiti opportunamente in maniera meccanica, mediante una tavola di corrispondenze (Hutchins 1993).

Dopo quella pionieristica esperienza, bisogna aspettare cinquant'anni prima di vedere un progetto di traduzione automatica che facesse uso dell'esperanto. Nel periodo 1983-1992, infatti, è stato portato avanti il progetto ancor oggi più importante in questo campo: il sistema DLT (*Distributed Language Translation*). Partito con uno studio di fattibilità condotto nel biennio 1982-1983 dal centro di ricerca della compagnia di software BSO di Utrecht (Olanda) e sostenuto dalla Comunità Europea, ricevette un finanziamento per produrre un sistema commerciale che fosse in grado di tradurre tra le principali lingue ufficiali in uso nella Comunità Europea (francese, inglese, italiano e tedesco), entro il 1993 (Witkam 1983). DLT intendeva fornire un servizio di traduzione di testi commerciali e manualistica, e, come si evince dal nome, doveva essere fruito mediante computer collegati in rete (si noti che non c'era ancora il web come lo conosciamo oggi). Il sistema DLT era basato sul paradigma dell'interlingua. Molto schematicamente, un'interlingua è indipendente dalle lingue in uso perché rappresenta la conoscenza, semantica e pragmatica, in maniera astratta e generale, di solito mediante un insieme predefinito di ruoli tematici, vale a dire elementi che descrivono la funzione assunta dai componenti all'interno della frase e i reciproci collegamenti (Jurafsky-Martin 1999: 812). Nel

prototipo presentato nel 1987, il testo di input veniva scritto in una forma semplificata di inglese, e lo strato interlingua, sostanzialmente una formalizzazione dell'esperanto scritta in Prolog, generava l'output in francese (Schubert 1992). I tentativi di sviluppare la versione commerciale si discostarono dall'architettura del prototipo e non ebbero successo (per una valutazione complessiva: Hutchins-Somers 1997, cap. 17).

Negli anni Novanta, grazie alla diffusione sempre più capillare di internet, ci sono stati altri tentativi di applicazione dell'esperanto alla traduzione automatica. In particolare, nel 1999 quattro ricercatori del Dipartimento di Informatica di La Coruña (Spagna) lanciano il progetto UTL (*Universal Translation Language*) che viene mantenuto per quattro anni su base volontaria. UTL è un linguaggio formale molto simile all'esperanto, dove le ambiguità semantiche sono state ridotte mediante l'introduzione di elementi a priori – per esempio mediante l'inserimento di preposizioni per marcare i diversi usi dell'accusativo. A differenza del progetto olandese, UTL intendeva essere usato come lingua di partenza della traduzione e non come interlingua (per una presentazione di UTL, si veda Sabarís *et al.* 2001). Nel periodo 2001-2003 UTL è stato adattato con successo a UNL (*Universal Network Language*), il progetto più vasto di traduzione automatica su paradigma interlingua mai avviato. Lanciato dall'Università dell'Onu (UNU) di Tokio nel 1996, UNL è un'interlingua che rappresenta la conoscenza della frase mediante una rete semantica (per una descrizione: UNL Center 2004). Attualmente il progetto è applicato su 15 lingue, tra cui l'esperanto non figura in nessuna forma. L'obiettivo a lungo termine è di rendere UNL disponibile per la traduzione automatica di testi scritti sul web (Cardeñosa *et al.* 2005). Al di fuori del paradigma interlingua, esistono altri sistemi di traduzione automatica che usano l'esperanto, invariabilmente basati sul paradigma transfer. In un sistema transfer le grammatiche delle lingue in oggetto vengono descritte individualmente mediante regole di produzione, di solito dipendenti dalla freccia di traduzione (per dettagli sul paradigma transfer, si veda Arnold *et al.* 1994, cap. 4). Il più importante di questi è Ergane, il quale però viene usato per scopi diversi dall'intento originale, in particolare la ricerca su interfacce innovative per il dialogo uomo-macchina, in particolare il reperimento di informazioni mediante interrogazioni in linguaggio storico-naturale multilingue (per es. si veda Brand-Brünner 2003). In ogni caso, nessun sistema transfer facente uso dell'esperanto è mai decollato come servizio di traduzione automatica.

2. L'esperanto per la traduzione automatica: risultati ottenuti

Molte caratteristiche strutturali dell'esperanto lo rendono una lingua particolarmente adatta al trattamento computazionale, e sono evidenti a chi conosca la lingua. In primo luogo, per i suoi tratti morfologici, l'esperanto ha caratteri di buona segmentabilità e invariabilità dei morfemi (nessuno cumulativo), e un grado di allomorfia tendente a zero. È significativo che una delle poche modifiche compiute dal gruppo

di sviluppo di DLT per il trattamento automatico abbia riguardato proprio l'allomorfia dovuta a composti, in casi come *sen-dat-a* ('senza data') vs. *send-ata* ('spedito'). In DLT infatti non era stata prevista un'analisi morfologica o riconoscimento delle parti del discorso in automatico mediante parsing, e ciò può essere visto come una caratteristica saliente della lingua sottovalutata dai progettisti di DLT (Hutchins-Somers 1997, cap. 17). E in effetti un etichettatore (tagger) per le parti del discorso è stato realizzato in maniera indipendente a DLT, e con sforzo molto minore rispetto a qualsiasi analogo per altre lingue (Minnaja-Paccagnella 2000). I progettisti di DLT si sono concentrati su una descrizione formale, attenta e puntuale della sintassi, con un formalismo basato sulle grammatiche della dipendenza, e su regole di trasformazione – dette *metataxis* – tra l'albero di dipendenza dell'esperanto e i corrispondenti alberi delle altre due lingue in gioco nel prototipo, vale a dire inglese e francese (Schubert 1987). A parere di chi scrive, queste regole di trasformazione sono il contributo più importante di quel progetto.

DLT ha dunque inteso sfruttare appieno l'altra caratteristica che rende l'esperanto particolarmente adatto al trattamento computazionale: la marcatura esplicita dei ruoli sintattici in maniera generale e univoca, mediante soli quattordici morfemi. È stato proprio il nostro maestro Fabrizio Pennacchietti a compiere un'attenta disamina delle caratteristiche morfologiche e sintattiche – così regolari e trasparenti! – dell'esperanto, che la rendono "la meno indo-europea delle lingue occidentali" (Pennacchietti 1987). E fu sempre lui, durante la stesura della mia tesi di laurea (perdonate la parentesi aneddotica!), a segnalarmi il legame profondo tra il fondatore teorico delle grammatiche della dipendenza, Lucien Tesnière, e l'esperanto. I quattro morfemi di marcatura sintattica basilare servirono infatti a Tesnière come mnemotecnica per le categorie fondamentali: nomi (O), nominanti (A), verbanti (E) e verbi (I), per usare la terminologia del suo allievo Claude Hagège.

Conseguentemente alla sua morfologia, l'esperanto è altamente produttivo da un punto di vista lessicale: di norma preferisce parole composte anziché importare lessemi nuovi, almeno per i registri linguistici di uso quotidiano (un solo esempio: *poŝt-ofic-ej-o* "ufficio postale", *lern-ej-o* "scuola"; si noti la produttività del morfema *ej-* per 'luogo'). Questo porta un altro grosso vantaggio da un punto di vista computazionale: bastano relativamente pochi lemmi per avere un quadro ampio della lingua. Da sempre è proprio la costruzione del lessico il compito più lungo e gravoso per costruire un qualsiasi sistema di trattamento automatico del linguaggio storico-naturale. Per dare qualche numero, si noti il fatto che con oltre 16.000 lemmi si è costruito il dizionario monolingue esperanto più ampio (e criticato), il PIV (*Plena Ilustrita Vortaro*, ed. 2002) mentre l'ottimo monolingue italiano del De Mauro ha circa 160.000 lemmi, dieci volte tanto. Ciò non implica, è importante sottolinearlo, che l'esperanto sia meno espressivo dell'italiano o di altra lingua non pianificata, in linea di principio – è vero che i contesti d'uso dell'italiano sono più ampi di quelli dell'esperanto, ma si tratta di un fattore contingente, non di un limite a priori. La storia di questa lingua, usata ininterrottamente per oltre un secolo, sopravvissuta a due

guerre mondiali, ha dimostrato la sua espressività mediante manufatti letterari e culturali, intendendo questi ultimi due termini in senso ampio, antropologico.

È proprio questo il fattore 'nascosto' dell'esperanto e non ancora utilizzato nel campo della traduzione automatica, ad oggi: l'uso vivo della lingua come base da cui estrarre la conoscenza per effettuare la traduzione. Questo riutilizzo dei dati di corpora linguistici divide l'approccio tradizionale alla traduzione automatica, basato su regole a priori, da quello attuale, a posteriori. Attualmente lo stato dell'arte fa uso di tecniche statistiche che prendono in input i dati linguistici da corpora linguistici paralleli, cioè insiemi di documenti o testi coerenti, opportunamente codificati in un formato leggibile da un elaboratore, e tradotti parallelamente secondo i paragrafi o meglio ancora le frasi. I corpora paralleli più usati sono gli atti dei Parlamenti di istituzioni multilingue, come l'Unione Europea o il Canada. L'ampia diffusione di internet tra i membri della comunità esperantista e la produzione di testi in lingua permette oggi l'applicazione di questo paradigma di ricerca. Chi scrive ritiene possibile costruire un corpus parallelo con l'esperanto dalle edizioni on-line di *Le Monde Diplomatique* accessibili da chiunque sul web e da altri repertori liberi, come il dizionario collaborativo ReVo (*Reta Vortaro*).

3. L'approccio statistico alla traduzione automatica in breve

Curiosamente, le prime teorizzazioni di motori di traduzione automatica su basi statistiche non sono recenti, anzi risalgono agli albori dell'informatica, vale a dire gli anni Quaranta. Com'è noto, i primi prototipi di elaboratori furono costruiti durante la Seconda Guerra mondiale negli Stati Uniti per decifrare i messaggi dei nemici. Fu proprio Warren Weaver, in un memorandum divenuto celebre, a rappresentare il problema della traduzione come un problema di decrittazione, in analogia a quei risultati: Weaver considerava un testo in cinese o in russo come fosse un testo scritto in inglese e criptato in "codice" cinese o russo, e quindi tutto ciò che bisognava fare era eliminare il "rumore" e recuperare l'informazione nel testo. In sostanza, si trattava di applicare il modello del canale della teoria matematica di Shannon e Weaver (1949). Si è dovuto attendere gli anni Novanta (mezzo secolo!) prima di poter effettuare i primi tentativi di applicazione dell'intuizione di Weaver. In quegli anni, furono molti i successi ottenuti nel campo del riconoscimento vocale e ottico dei metodi statistici, in termini sia di incremento di prestazioni sia di riduzione dei tempi di sviluppo (per un sistema basato su regole robusto ci vogliono anni, per uno statistico mesi). Questo ha portato alcuni ricercatori a riprendere l'idea di Weaver, grazie anche alla possibilità di avere elaboratori potenti a costi ragionevoli (Jurafsky-Martin 2000, cap. 21).

Presentiamo ora molto schematicamente i fondamenti del paradigma statistico, così come esposti negli articoli fondanti di Brown *et. al.* (1990, 1993), indispensabili per poter proseguire. Possiamo rappresentare formalmente la traduzione come il problema di trovare la frase più probabile nella lingua di arrivo e, data una frase

ignota nella lingua di partenza `f`, cioè `P(e|f)`. Possiamo dividere questo problema in due parti distinte. Applicando la regola di Bayes otteniamo l'equazione fondamentale della traduzione automatica statistica:

```
argmax P(e|f) = argmax P(e) * P(f|e)
 e          e
```

Perché c'è bisogno di dividere il problema della traduzione in due parti distinte? Da un lato, ciò è conveniente da un punto di vista computazionale, ma c'è anche una motivazione di principio, che può essere spiegata mediante un'analogia. Se la traduzione fosse la ricerca del colpevole di una scena del delitto, `e` sarebbe il colpevole, `f` sarebbe la scena del delitto, `P(e)` verrebbe a rappresentare i moventi, e `P(f|e)` i fattori legati alla scena del delitto. Si possono avere motivazioni delittuose ma non avere l'occasione adeguata per compiere il delitto, e analogamente nel caso della traduzione: si può avere un'ottima descrizione della lingua di arrivo, ma se non si ha una buona relazione con la lingua di partenza non si può tradurre.

Vediamo ora come sono fatte le due parti. La prima parte consiste nel calcolo del *modello di linguaggio*, che assegna una probabilità a priori `P(e)` ad ogni parola (stringa) appartenente alla lingua d'arrivo `e`. Questa probabilità dà conto del fatto che la conoscenza della lingua d'arrivo `e` è più importante della conoscenza della lingua di partenza `f` in una traduzione. Si noti inoltre che il modello di linguaggio può essere calcolato nelle maniere più disparate. Nella teoria di Shannon e Weaver, il modello di linguaggio corrisponde al modello dell'emittente. La seconda parte consiste invece nel calcolo del *modello di traduzione*, che assegna una probabilità condizionata `P(f|e)` a ciascuna coppia di unità `(f,e)`. Una volta calcolata, si dice che i dati del corpus sono stati *allineati*. Nella teoria di Shannon e Weaver, il modello di traduzione corrisponde al modello del canale.

I due modelli vengono messi in relazione dal *decoder*, che effettua la traduzione vera e propria. Data una frase nuova, ignota nella lingua di partenza `f`, il decoder calcola la frase corrispondente, cioè quella più probabile, nella lingua d'arrivo `e` massimizzando `P(f|e)`. Un buon algoritmo di decodifica è la chiave di un sistema di traduzione automatica efficiente: da un lato il decoder dev'essere preciso, dall'altro deve dare risultati accettabili in tempi ragionevoli.

Dopo quasi quindici anni di applicazione di questo approccio, abbiamo formalizzato da Brown *et al.* (1993), i modelli di traduzione e i decoder hanno raggiunto delle buone prestazioni. Inoltre, esiste una metrica di valutazione automatica della qualità della traduzione, detta BLUE, utile in fase di sviluppo del motore (Papineni *et al.* 2002). D'altro canto, l'esperienza sul campo ha evidenziato alcuni limiti di questo approccio, tracciando le linee per ricerche ulteriori. Uno degli ostacoli più grossi alla traduzione automatica, indipendentemente dall'approccio, è costituito dall'ordine dei costituenti. Nei sistemi statistici, la diversa dislocazione delle parti del discorso nella frase tra una lingua e l'altra viene catturata dall'allineamento delle unità nel corpus parallelo (per esempio, poniamo che l'occorrenza 'Unione Europea'

venga allineata dal motore a 'European Union' nelle sezioni italiana e inglese, catturando così la diversa dislocazione nome-aggettivo). La ricerca ha dimostrato di recente che è più efficiente prendere come unità di misura non singole parole ma pezzi di frase probabilisticamente significativi, vale a dire senza strutturazione sintagmatica: in questo modo, oltre alle dislocazioni a breve distanza (nomi, aggettivi, determinanti, verbi, avverbi) si riesce a catturare anche dislocazioni a medio raggio, per esempio l'ordine dei complementi. Attualmente i modelli migliori sono basati su probabilità logaritmiche lineari e prendono come unità di misura pezzi di frase invece di singole parole. Ma anche procedendo in questo modo, dislocazioni distanti non vengono catturate: si pensi ad esempio ai verbi composti separabili in tedesco, dove la preposizione va in fondo di frase, e può essere molto distante dal verbo (come introduzione al problema, si veda Collins *et al.* 2005).

Un altro grosso ostacolo riguarda la fluidità del testo generato in automatico: un sistema basato su regole in linea di principio può evitare frasi asintattiche e dunque inintellegibili perché impone dei vincoli a priori, un sistema statistico no. Infine, l'ultimo limite noto in letteratura è legato paradossalmente al suo vantaggio principale, vale a dire i corpora paralleli. Da un lato va notato che per lingue poco diffuse corpora paralleli sono meno disponibili di quanto si creda. Dall'altro, da un punto di vista linguistico il sistema è limitato dal corpus stesso. Gli atti dei parlamenti, per esempio, sono un registro linguistico abbastanza formale ma certamente molto legato al parlato: i testi nuovi in input, se molto distanti da questo registro, possono dare risultati inferiori alle aspettative.

Ci sono due strade, diverse ma complementari, per superare questi ostacoli e limiti e migliorare le prestazioni dei sistemi di traduzione automatica. Da un lato, si può e si deve fare ricerca dal lato computazionale, cercando cioè tecniche statistiche sempre più raffinate per ottimizzare i motori. In particolare, la quasi totalità dei sistemi di traduzione automatica statistica correnti sono rigidi rispetto ai dati, in altre parole non sanno adattarsi all'inserimento di dati nuovi, per esempio da parte di correzioni in post-editing da parte degli utenti del sistema, se non azzerando la conoscenza e riaddestrando il motore da capo, il che è gravoso (ci vogliono circa due settimane di computazione continua per addestrare un motore di traduzione automatica di media grandezza). Dall'altro lato, si può e si deve fare ricerca dal lato linguistico. Invece di considerare testi e frasi come agglomerati grezzi di stringhe, la direzione attuale della ricerca va verso l'inserimento di conoscenza sintattica. In altri termini, ad ogni frase viene associato un albero sintattico, e vengono calcolate le probabilità tra i rispettivi alberi, e infine generata la lingua d'arrivo. Finora però i risultati ottenuti negli esperimenti compiuti per tradurre il cinese e l'arabo in inglese, addestrando il sistema con corpora di frasi annotate con grammatiche a costituenti (alla Chomsky, per intendersi) sono scoraggianti. Si è ottenuto un decadimento delle prestazioni, anziché un miglioramento, e oltretutto pagando un costo computazionale elevato, perché il grado di complessità computazionale richiesto per allineare alberi è ben più elevato di quello relativo all'allineamento di stringhe.

Un'altra strada, non ancora percorsa, è quella di inserire come conoscenza linguistica a priori anziché alberi sintattici conoscenze morfologiche mediante etichettatura delle parti del discorso (PoS-tagging). Il problema è il seguente: perché il decoder deve calcolare *tutte* le combinazioni per l'allineamento, anche se linguisticamente insensate? In altre parole, è possibile trovare un criterio per ridurre lo spazio di ricerca dell'algoritmo, aumentando contemporaneamente il grado di intelligibilità. Nell'esempio riportato da Brown *et al.* (1993) su un caso reale, i valori di allineamento per *farmers* in un motore di traduzione dal francese sono, in quest'ordine: *agriculteurs*, *les*, *cultivateurs*. È chiaro che il determinante *les* viene allineato al sostantivo *farmers* a causa della vicinanza dei determinanti ai sostantivi nell'ordine dei costituenti. Si prenda allora come ragionevole l'ipotesi che ogni parola venga probabilmente allineata a una analoga parte del discorso, vale a dire verbi con verbi, verbanti con verbanti, nominanti con nominanti, nomi con nomi (ipotesi della consistenza grammaticale). Quello di cui c'è bisogno è dunque un insieme di etichette (tagset) valido per più lingue e calcolare di conseguenza il modello di traduzione. A questo punto può entrare in gioco l'esperanto, per valutare se questa ipotesi di percorso di ricerca è valido.

4. L'esperanto come *tertium comparationis* della traduzione automatica

Già Eugen Wüster considerava l'esperanto come *tertium comparationis* tra due lingue storico-naturali nel suo lavoro, e questa intuizione può essere portata alla traduzione automatica. Il progetto di ricerca che intende intraprendere l'autore è volto proprio a verificare la consistenza delle considerazioni fatte finora nel caso speciale dell'esperanto.

Innanzitutto, bisogna esplicitare il corpus parallelo dalle edizioni in linea di *Le Monde Diplomatique* riportando i testi in un formato dei dati adeguato, analogamente a *ReVo*. L'eterogeneità dei formati di digitalizzazione dei corpora paralleli è un limite noto tra gli specialisti, che può essere superato mediante una variante dello standard per la codifica dei testi letterari chiamato TEI, che viene considerato la base per fare filologia nei testi elettronici (per una discussione: Fiormonte 2003). Sulla base del TEI, è in via di definizione una variante per i corpora paralleli, detta XCES, che permette di elaborare il testo mediante tutti gli strumenti di trattamento del linguaggio naturale noti e mediante quelli futuri, basati su basi di dati native in XML.

Una volta sistematizzato il corpus parallelo, si può iniziare a costruire i modelli di linguaggio e i modelli di traduzione con gli strumenti resi disponibili per usi non militari dalle Università di tutto il mondo che fanno ricerca in questo campo. Il lavoro di ricerca si concentra sulla definizione del modello di linguaggio. In particolare, quello che si vuole dimostrare è l'ipotesi della consistenza grammaticale, e ciò può essere fatto agevolmente mediante l'esperanto. Come si è visto, il parser morfologico e un modello di grammatica della dipendenza sono già disponibili: si tratta

di riprendere le esperienze del passato e adattarle al contesto attuale, vale a dire rendere fruibile il motore via web. In questo modo, se il motore di ricerca automatica è robusto rispetto all'inserimento di dati nuovi, è possibile ricavare esempi significativi dal post-editing dei parlanti, vale a dire dalle frasi corrette sulla lingua d'arrivo (nel gergo degli specialisti, queste frasi vengono chiamate "memorie di traduzione"). La differenza di prestazioni tra il corpus originale e il corpus arricchito delle memorie di traduzioni dà la misura dell'incremento del sistema globale.

Dall'altra parte, è possibile misurare l'impatto della sintassi sulla traduzione lanciando tre istanze dello stesso motore di traduzione sullo stesso corpus: la prima non ha alcun tipo di conoscenza sintattica; nella seconda istanza il corpus è stato etichettato dal parser multilingue, e serve a verificare l'ipotesi della consistenza sintattica; nella terza istanza, una parte del corpus in esperanto è stato annotato mediante la grammatica della dipendenza ricavata dall'esperienza di DLT, ed estesa su tutto il corpus mediante tecniche di apprendimento automatico. Alle tre istanze vengono presentati in input gli stessi testi, nuovi. La differenza di prestazioni tra le tre istanze dà la misura dell'impatto della sintassi sulla traduzione automatica.

5. Conclusioni

I tempi sono maturi per verificare se i nostri modelli di sintassi, e in particolare le grammatiche della dipendenza, sono formalmente validi, e questo è possibile verificarlo computazionalmente mediante la traduzione automatica. Naturalmente, sono possibili diverse obiezioni valide a quanto descritto. La prima obiezione, la più ovvia e immediata, riguarda il parser: nessun parser è preciso in tutti i casi, nemmeno per l'esperanto, e dunque si inserisce una probabilità di errore in più nelle stime globali. Una seconda obiezione riguarda i costi: è conveniente, da un punto di vista dei tempi di computazione e delle prestazioni finali, aggiungere conoscenza morfologica e sintattica al corpus? Una terza obiezione riguarda infine l'esperanto: quanto le prestazioni dipendono dalla struttura speciale dell'esperanto? In altri termini, quanto è generalizzabile questo modello? Ovviamente non è possibile rispondere a queste obiezioni in questa fase preliminare, progettuale, perché mancano i dati sperimentali. Quel che è certo, è che queste tre obiezioni sono fondamentali e dovranno essere prese in seria considerazione per valutare il lavoro, una volta svolto.

Bibliografia

Arnold, D. - Balkan, L. - Meijer, S. - Humphreys, R. - Sadler, L. (1994). *Machine Translation: an Introductory Guide*. London.

Brand, Roel & Brünner, Marvin (2003). 'Océ'. In: Cross Language Evalutaion Forum 2002, *Lecture Notes in Computer Science* 2785 (January): 59-65.

Brown, P.F. - Cocke, J. - Della Pietra, S.A. - Della Pietra, V.J. - Jelinek, F. - Lafferty, J.D. - Mercer, R.L. - Roossin, P.S. (1990). 'A Statistical Approach to Machine Translation'. *Computational Linguistics* 16 (2): 79-85.

Brown, P.F. - Della Pietra, S.A. - Della Pietra, V.J. - Mercer, R.L. (1993). 'The Mathematics of Statistical Machine Translation: Parameter Estimation'. *Computational Linguistics* 19 (2): 263-311.

Cardeñosa, J. - Gelbukh, A. - Tovar, E. (eds.) (2005). 'Universal Networking Language: Advances in Theory and Applications'. Special issue of *Research on Computing Science*. IPN.

Collins, M. - Koehn, Ph. - Kučerová, I. (2005). 'Clause Restructuring for Statistical Machine Translation'. *Proceedings of the 43rd Annual Meeting of the ACL*, 531-540.

Fiormonte, D. (2003). *Scrittura e filologia nell'era digitale*. Torino.

Hutchins, J.W. (1993). 'The first MT patents'. *MT News International* 5: 14-16.

Hutchins, J.W. (1997). 'First steps in mechanical translation'. In: V. Teller - B. Sundheim (eds.), *MT Summit VI: past, present, future. Proceedings, 29 October - 1 November 1997, San Diego, California*. Washington: 14-23.

Hutchins, J.W. - Somers, H.L. (1997). *An Introduction to Machine Translation*. London.

Jurafsky, D. - Martin, J.H. (2000). 'Speech and Language Processing'. In: *An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. New York.

Minnaja, C. - Paccagnella, L.G. (2000). 'A Part-of-Speech Tagger for Esperanto Oriented to MT'. International Conference "MT 2000 – Machine Translation and multilingual Application in the new Millennium", Exeter. 13.1-13.5.

Papineni, K. - Soukos, S. - Ward, T. - Henderson, J. - Reeder, F. (2002). 'Corpus-based Comprehensive and Diagnostic MT Evaluation: Intial Arabic, Chinese, French and Spanish Results'. In: *Proceedings of the Human Language Technology Conference*.

Pennacchietti, F.A. (1987). 'L'internazionalità dell'esperanto e il carattere degli elementi indoeuropei in esso'. In: A. Chiti-Batelli (a cura di), *La comunicazione internazionale tra politica e glottodidattica*. Milano: 130-142.

Rosenfeld, R. (2000). 'Two Decades of Statistical Language Modeling: Where do we go from here'. In: *Proceedings of the IEEE* 88(8).

Shannon, C. - Weaver, W. (1949). *The Mathematical Theory of Communication*. Urbana, Ill.

Schubert, K. (1987). *Metataxis: Contrastive Dependency Syntax for Machine Translation*. Dordrecht.

Schubert, K. (1992). 'Esperanto as an Intermediate Language for Machine Translation'. In: J. Newton (ed.). *Computers in Translation*. London: 78-95.

UNL Center (2004). *The Universal Networking Language. Specifications. Version 3. Edition 3*. UNDL Foundation (http://www.unlc.undl.org/unlsys/unl/UNL%20Specifications.htm).

Witkam Toon, A.P.M. (1983). 'Distributed Language Translation. Feasibility Study of a Multilingual Facility for Videotex Information Networks'. BSO, documento interno. Utrecht.