

Ritrovato per migliorare la qualità della traduzione automatica basato sulla teoria correlazionale del pensiero

Giulio Benedetti¹

¹ Sono grato a Giorgio Marchetti per il suo aiuto nella stesura di questo articolo.

Abstract

In questo articolo l'autore descrive un ritrovato per migliorare la qualità della traduzione automatica ideato da Silvio Ceccato e basato sulla sua teoria correlazionale del pensiero. Basandosi su esso Ceccato e i suoi collaboratori si dedicarono al problema della traduzione automatica dal 1959 al 1966, compiendo ricerche di tipo quasi esclusivamente teorico. Il progetto fu interrotto quando la ricerca sulla traduzione automatica cessò negli USA in conseguenza del noto rapporto ALPAC¹, prima che fosse possibile valutare i risultati che questo ritrovato può dare.

L'autore ritiene che questo ritrovato e la teoria su cui è basato, per quanto entrambi pochissimo noti per varie ragioni, siano a tutt'oggi estremamente validi e che pertanto potrebbero permettere di ottenere, nel campo della traduzione automatica, risultati migliori rispetto a quelli ottenibili oggi con i programmi disponibili, specialmente quando la lingua di ingresso è una lingua con pochissima morfologia e con molte ambiguità, come l'inglese, e quella di uscita è una lingua con ricca o ricchissima morfologia. Questo essenzialmente perché: a) la teoria correlazionale del pensiero è (anche) una *teoria linguistica completamente nuova e molto diversa da tutte le altre*; b) il programma basato sul suddetto ritrovato *riproduce in parte ciò che effettivamente fa l'essere umano con la sua intelligenza* quando traduce. Questo sistema di traduzione automatica è inoltre probabilmente l'unico basato completamente ed esclusivamente su un'unica teoria linguistica e ideato da chi ha proposto la teoria linguistica stessa. L'unico problema che esso presenta, cioè il fatto che il lavoro umano preliminare alla realizzazione del programma aumenta in maniera esponenziale all'aumentare del numero dei termini del dizionario, non è di natura teorica ma soltanto pratica.

Avvertenza

Questo articolo non può essere compreso senza avere prima acquisito una certa conoscenza della Metodologia Operativa. Una conoscenza sufficiente può essere rapidamente acquisita leggendo il mio articolo "Una presentazione della Metodologia Operativa". Il presente articolo è stato infatti concepito come avente l'altro articolo come sua indispensabile premessa. La lettura di un altro mio articolo, "Le operazioni mentali di base costituenti le categorie mentali", non è strettamente necessaria ma può risultare utile per avere un'idea più completa della materia.

Questo articolo è stato concepito per poter essere letto anche da persone che non hanno alcuna conoscenza riguardo alla Traduzione Automatica, ma sono interessate a sapere come la Metodologia Operativa e la teoria correlazionale del pensiero possono trovare applicazione in questo campo.

Informazioni sul copyright (© Giulio Benedetti, 2005)

Questo articolo può essere copiato e diffuso liberamente, tranne che per scopi commerciali. Devono obbligatoriamente essere riportati l'autore (Giulio Benedetti) e il sito di origine www.mind-consciousness-language.com (di Giorgio Marchetti).

Introduzione

In questo articolo intendo descrivere il ritrovato per realizzare un programma di traduzione automatica cui ho accennato nel mio articolo “Una presentazione della Metodologia Operativa”. Come ho detto in questo articolo, le teorie che costituiscono la Metodologia Operativa (M.O.) cominciano ad essere sviluppate negli anni '50 e il pensiero di Ceccato raggiunge la piena maturità nel corso degli anni '60 e '70. Il progetto per la traduzione automatica di Ceccato e collaboratori viene portato avanti dal 1959 al 1966, anno in cui il ben noto rapporto ALPAC fece cessare i finanziamenti¹. Il progetto di Ceccato risale dunque alle prime fasi della storia della traduzione automatica. Può dunque sembrare strano che in un articolo odierno venga presentato nuovamente, sia pure con qualche modifica, un progetto tanto datato. A mio avviso ci sono almeno quattro ragioni valide per presentare nuovamente questo progetto. La prima è che questo progetto e le teorie che stanno alla sua base, per quanto così datate, sono pochissimo note. La seconda ragione è che queste teorie (e quindi anche il progetto stesso) sono state realmente comprese solo da pochissime persone. La terza ragione è che il progetto fu interrotto dopo pochi anni e non è stato quindi possibile vedere quali risultati avrebbe potuto dare. Infine, i progressi fatti dalla tecnologia dei calcolatori e della produzione di software da 40 anni fa ad oggi sono enormi e potrebbero consentire di raggiungere risultati non raggiungibili al tempo in cui fu concepito questo progetto. Di queste quattro ragioni la più importante è sicuramente la seconda. È naturale chiedersi perché la M.O. sia una teoria così poco diffusa e così raramente ben compresa. In un articolo di questo tipo non è possibile parlare di questo problema in maniera approfondita, tuttavia è necessario fare almeno alcuni accenni ad esso, ampliando almeno un poco quanto detto nel breve articolo “Una presentazione della Metodologia Operativa”. Le principali ragioni di questo fatto sono, a mio avviso, almeno quattro.

1) La prima è che effettivamente le teorie che costituiscono la M.O. possono risultare difficili ad essere comprese, sia per la loro stessa natura sia perché sono radicalmente innovative. Questa tuttavia è la meno importante delle quattro suddette ragioni. Infatti, impostando l'esposizione in maniera adeguata, sono convinto che questa difficoltà può essere superata. L'esperienza didattica che ho personalmente finora maturato, pur non essendo molta, ha sempre confermato questa mia convinzione.

2) La seconda ragione è proprio che nell'esposizione di queste teorie è stata usata un'impostazione didattica a mio avviso completamente errata: la M.O. è sempre stata presentata come una critica radicale alle teorie filosofiche tradizionali, ma proprio per questo essa è stata sentita come una nuova teoria

¹ Nel 1966, il rapporto della commissione governativa istituita negli USA per la valutazione dei risultati e delle prospettive della traduzione automatica (Automatic Language Processing Advisory Committee, ALPAC) giudicò inutile la prosecuzione degli investimenti per la ricerca sulla traduzione automatica, causando la virtuale cessazione di essa negli USA per oltre dieci anni.

filosofica, cosa a cui non molti sono interessati. I frequenti richiami alla filosofia hanno poi richiesto l'uso di un linguaggio spesso difficile, a cui si aggiunge il problema di un'esposizione non sempre chiara.

3) La terza ragione è che a mio avviso le teorie di Ceccato contengono gravi errori, specialmente nella loro parte più generale (soprattutto quella riguardante la percezione). Poiché questa parte è quella su cui Ceccato ha posto più l'enfasi e quella che egli espone sempre per prima, è molto probabile che molti di coloro che si sono avvicinati a queste teorie se ne siano allontanati presto proprio per questo motivo.

4) Infine, a quanto mi risulta, Ceccato aveva spesso atteggiamenti e posizioni che gli hanno alienato moltissime simpatie, specialmente nell'ambiente accademico.

Il mio lavoro è consistito in un'attenta, sistematica e approfondita revisione critica di tutto il pensiero di Ceccato e della Scuola Operativa Italiana (S.O.I.), cercando di prendere ciò che in esso è valido e di eliminare gli errori. Quest'opera ha portato a modifiche molto profonde e a molti sviluppi, cosicché *le teorie che propongo sono sensibilmente diverse dalle teorie di partenza*. Ho inoltre adottato *un'impostazione dell'esposizione completamente diversa*, che dovrebbe consentire una comprensione di esse molto più facile.

La Metodologia Operativa e la Traduzione Automatica

La M.O. non nasce come teoria linguistica né tanto meno come teoria linguistica formulata in vista del problema della traduzione automatica. La M.O. nasce come teoria riguardante l'attività mentale umana (cioè la percezione, la rappresentazione, la memoria, il pensiero, gli atteggiamenti ecc.) studiata con un metodo strettamente introspettivo. Pertanto essa può essere considerata come una teoria nell'ambito della Psicologia Cognitiva. Tuttavia, poiché il linguaggio non è altro che il mezzo attraverso il quale noi esprimiamo il nostro pensiero, la M.O. è *anche* una teoria linguistica.

È per questo che la M.O. può avere importanza nel campo della traduzione automatica. Infatti, come affermano Hutchins e Somers, “i maggiori ostacoli alla traduzione automatica sono, e sono sempre stati, non di natura informatica ma linguistica”². Tuttavia i rapporti tra ricerca nel campo della traduzione automatica e teorie linguistiche sono sempre stati problematici. Un chiaro quadro della situazione è stato tracciato dagli stessi autori:

“... ‘state-of-the-art’ MT systems are not in general based on one single linguistic theory. [...] a number of approaches to the linguistic description of natural languages have been influential in the design of MT systems. However, there have been very few systems based on a single linguistic model, and these have been predominantly experimental projects undertaking basic research using MT as a test-bed for computational linguistic theories. The great majority of MT systems are amalgams of different approaches and models, or even occasionally (particularly in the early years of MT research [...]) with no discernible theoretical basis at all. Most commonly, systems are vaguely based on a general theory, such as transformational grammar or dependency theory, greatly modified by borrowings from other theories and by the demands of computational implementation.

MT research has often been criticised for ignoring developments in linguistic theory. There would appear to be a wide communication gap between theoretical linguistics and practical MT research. Some observers believe that there are good reasons for this situation: until recently, linguistic theories had not provided adequate accounts of all aspects of language use [...]. In the past, and unfortunately it is still generally true today, much of a linguistic theory was based on phenomena observed in English, the language of the majority of the theoretical linguists. This neglect of other languages has been a further reason why linguistic theory has had less impact on MT than some observers might have expected.

[...] MT research is sometimes regarded as an ‘engineering’ task, a search for computational methods that work with the facts of language. The aims of many theoretical linguists are more abstract, concerned with investigations of human faculties in general, the nature of language itself, and the psychological foundations of language acquisition and use.

The results has been somewhat cynically described by Yorick Wilks (1989; 59):

² Hutchins W. J. (con Somers H. L.), *An introduction to machine translation*, p. 2.

“the history of MT shows, to me at least, the truth of two (barely compatible) principles that could be put crudely as ‘Virtually any theory, no matter how silly, can be the basis of some effective MT’ and ‘Successful MT systems rarely work with the theory they claim to.’”

The search of solutions that work, whatever their theoretical status and whether or not they fit the alleged principles of the project, has meant that MT systems inevitably present a confusing picture of disparate methodologies and that researchers have been obliged to take much more pragmatic attitudes to theoretical issues than their colleagues in computational linguistic and in linguistic theory.”³

[“... i sistemi di traduzione automatica che rappresentano uno stato dell’arte non sono in generale basati su una singola teoria linguistica. [...] un certo numero di approcci alla descrizione linguistica dei linguaggi naturali hanno avuto influenza sul disegno dei sistemi di traduzione automatica. Comunque, ci sono stati pochissimi sistemi basati su un singolo modello linguistico, e questi sono stati progetti prevalentemente sperimentali che intraprendevano ricerca di base usando la traduzione automatica come banco di prova per teorie di linguistica computazionale. La grande maggioranza dei sistemi di traduzione automatica sono un’amalgama di approcci e modelli differenti, o anche, occasionalmente (particolarmente nei primi anni della ricerca sulla traduzione automatica [...]) del tutto privi di una base teorica discernibile. Più comunemente, i sistemi sono vagamente basati su una teoria generale, come la grammatica trasformazionale o la teoria della dipendenza, grandemente modificata da prestiti presi da altre teorie e dalle richieste della realizzazione informatica.

Alla ricerca sulla traduzione automatica è stata spesso mossa la critica di ignorare gli sviluppi nella ricerca linguistica. Sembrerebbe che ci sia un grande vuoto comunicativo fra linguistica teorica e ricerca pratica sulla traduzione automatica. Alcuni osservatori credono che ci sono delle buone ragioni per il fatto che ci sia una situazione del genere: fino a poco tempo fa le teorie linguistiche non avevano reso conto di tutti gli aspetti dell’uso del linguaggio [...]. In passato, e sfortunatamente in genere ancora oggi, molte teorie linguistiche erano basate sui fenomeni osservati in inglese, la lingua della maggioranza dei linguisti teorici. Questo trascurare gli altri linguaggi è stata un’ulteriore ragione per cui la teoria linguistica ha avuto meno impatto sulla traduzione automatica di quanto alcuni osservatori si sarebbero potuti aspettare.

[...] La ricerca sulla traduzione automatica è spesso considerata come un compito “ingegneristico”, una ricerca di metodi computazionali che funzionino con i fatti del linguaggio. Gli scopi di molti linguisti teorici sono più astratti, riguardano investigazioni sulle facoltà umane in generale, la natura del linguaggio stesso, e i fondamenti psicologici dell’acquisizione e dell’uso del linguaggio.

Il risultato è stato descritto piuttosto cinicamente da Yorick Wilks (1989; 59):

“... la storia della traduzione automatica mostra, almeno a me, la verità di due (apertamente compatibili) principi che potrebbero essere messi giù crudamente come ‘Virtualmente qualunque teoria, non importa quanto sciocca, può essere la base di una qualche traduzione automatica efficace’ e ‘i sistemi di traduzione automatica di successo raramente operano con la teoria con la quale sostengono di operare.’”

La ricerca di soluzioni che funzionino, qualunque sia il loro stato teorico e che si adattino ai principi dichiarati del progetto oppure no, ha significato che i sistemi di traduzione automatica inevitabilmente presentano un quadro confuso di metodologie disparate e che i ricercatori sono stati obbligati ad assumere atteggiamenti molto più pragmatici riguardo alle questioni che i loro colleghi in linguistica computazionale e in teoria linguistica.”]

In una situazione come questa, è chiaro che una teoria linguistica radicalmente nuova e diversa da quelle già esistenti potrebbe avere un impatto nel campo della traduzione automatica molto superiore a quello che hanno avuto le altre teorie linguistiche.

Differenze fra le teorie linguistiche della S.O.I. e le altre teorie linguistiche. Implicazioni per la traduzione automatica

Le teorie prodotte dalla S.O.I. sono (*anche*, come abbiamo visto) teorie linguistiche *profondamente diverse* dalle altre. Le somiglianze spesso sono solo superficiali o apparenti. Le ragioni di questa profonda diversità sono tre.

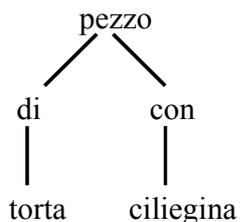
1) La prima ragione, pur essendo la più importante in senso assoluto, è quella che, almeno in tempi brevi o medi, ha meno importanza per la traduzione automatica. In un lontano futuro, invece, potrebbe avere un’importanza decisiva. Questa ragione è che la M.O., come ho cercato di evidenziare nel breve articolo suddetto, ha dato delle risposte riguardo al fondamentale problema del significato delle parole,

³ Ibidem, p. 81.

particolarmente di quelle di cui si occupa la grammatica (congiunzioni, preposizioni, pronomi, verbi fondamentali come “essere”, “avere” ecc., i principali avverbi ecc.). È possibile che le teorie della S.O.I. a questo riguardo debbano ancora essere in parte corrette, modificate e sviluppate (è quello che ho cercato e cerco di fare), ma nella loro sostanza esse sono, a mio avviso, enormemente più soddisfacenti rispetto a quelle prodotte da ogni altro tipo di approccio. La M.O. ha asserito che i significati di queste parole sono costituiti da complessi di operazioni mentali in cui quelle dell’attenzione giocano un ruolo decisivo e ha dato delle analisi di queste parole in termini di queste operazioni. Poiché queste operazioni, almeno teoricamente, sembrano riproducibili, la M.O. a mio avviso ha forse aperto la strada alla realizzazione della vera Intelligenza Artificiale. So benissimo che, anche se le vedute della S.O.I. e questa mia opinione sono corrette, questo è un compito comunque difficilissimo e che ci potrebbero essere degli ostacoli anche insormontabili (per esempio certe funzioni supportate dalle strutture nervose, quindi da strutture biologiche, potrebbero non essere riproducibili mediante dei componenti elettronici). Ma, se non è così, ed in un lontano futuro potremmo davvero avere un qualche successo nella realizzazione di una vera intelligenza artificiale, cioè di una macchina capace di comprendere un testo, questo sarebbe sicuramente importantissimo per poter produrre una traduzione automatica di alta qualità. Ad ogni modo, questo è un traguardo che non sappiamo se sia raggiungibile e comunque lontanissimo, mentre in quest’articolo mi interessa proporre ciò che può essere fatto al presente.

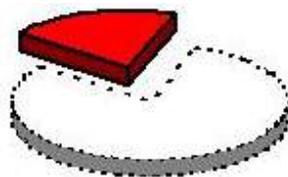
2) La seconda ragione è quella di importanza fondamentale per la traduzione automatica. La teoria correlazionale del pensiero, proposta da Ceccato, indipendentemente dal fatto che è appunto una teoria riguardante l’attività umana del pensiero e non solo una teoria linguistica, è profondamente diversa da tutte le altre teorie linguistiche perché parte da un presupposto opposto a quello da cui partono esse. Le altre teorie linguistiche, così come la grammatica tradizionale, partono dal presupposto che nella frase le parole siano riferite l’una all’altra in una maniera diretta, cioè che una certa parola sia riferita o legata ad un’altra, in maniera tale che i legami fra le parole sono rappresentabili con delle linee. La struttura delle frasi viene così rappresentata mediante i ben noti “alberi” o con altri tipi di rappresentazione che sono comunque sostanzialmente equivalenti. Il presupposto da cui parte la teoria correlazionale del pensiero è del tutto opposto. Secondo questa teoria, le parole appartengono a due categorie ben distinte, quella dei correlatori (alcune decine di parole), cioè le parole che hanno la funzione di *legare* altre parole, e quella dei correlati (tutte le altre parole), vale a dire le parole che devono *essere legate* da qualcos’altro. Due correlati non sono cioè *mai* legati in maniera diretta, ma c’è sempre un terzo elemento che li lega. Questo elemento può essere esplicito, cioè essere rappresentato da un’altra parola oppure da una marcatura delle due parole correlate, oppure può essere implicito. In quest’ultimo caso possiamo indicare quali sono le parole che esso lega mettendole l’una di seguito all’altra, ma non sempre questo è possibile e allora questa indicazione manca del tutto. È proprio questo, come vedremo, uno dei maggiori ostacoli per la realizzazione dei programmi di traduzione automatica.

Illustro questa differenza fra teoria correlazionale del pensiero e le altre teorie linguistiche con un semplicissimo esempio. Prendiamo l’espressione “pezzo di torta con ciliegina”. Secondo la grammatica tradizionale e le teorie linguistiche diverse dalla teoria correlazionale del pensiero (d’ora in avanti per semplicità le chiameremo “linguistiche convenzionali”) la struttura corrispondente a questa espressione sarebbe la seguente:



Cioè un sostantivo (“pezzo”) regge due frasi preposizionali, “di torta” e “con ciliegina”. È da notare che queste strutture sono di tipo fondamentalmente gerarchico. Questo si riflette nella terminologia tipica della grammatica e delle linguistiche convenzionali (“reggere”, “essere retto da”, “dipendere da”, “essere subordinato a”, “governare”, “governatore”, “testa”, “madre”, “figlia”, “sorella” ecc.). Ci sono cioè delle parole che sarebbero più “importanti” (in genere vengono considerati tali il nome che funge da soggetto e il verbo in forma personale, perché le frasi che produciamo contengono quasi sempre almeno questi due elementi) alle quali le altre parole della frase sarebbero legate direttamente con un rapporto di subordinazione. Ma, a livello del pensiero, è veramente così? Cerchiamo di capire bene che cosa succede nella nostra mente, vale a dire che cosa ci rappresentiamo, quando udiamo questa espressione. Io non direi che ci rappresentiamo l’astratta nozione di “pezzo” a cui subordiniamo la frase preposizionale “di torta” e poi facciamo la stessa cosa con l’altra frase preposizionale “con ciliegina”. Io direi che leghiamo mediante la categoria mentale indicata dalla preposizione “di” un’altra categoria mentale, quella indicata dalla parola “pezzo”, alla rappresentazione di un oggetto fisico, cioè una “torta”. Cioè noi immaginiamo una torta e restringiamo la nostra attenzione ad una parte di essa (in genere di forma triangolare, come sono le parti in cui siamo soliti tagliare le torte), scartando il resto ma mantenendo mentalmente presente che la parte che abbiamo selezionato con l’attenzione proviene dalla torta intera (categoria di “pezzo” applicata a “torta”). Il correlatore designato dalla preposizione “di” indica, come ogni altro correlatore, che i due elementi che esso lega sono *entrambi presenti, insieme con il correlatore stesso*, nella nostra mente e sono *sullo stesso piano* (la sola cosa che può essere considerata una “gerarchia” è il fatto che “pezzo” precede nel tempo “torta” e pertanto dobbiamo riferire “torta” a “pezzo” e non viceversa). Se vogliamo dare una rappresentazione grafica, quello che c’è nella nostra mente non è questo:

Primo livello gerarchico:



(nozione astratta di “pezzo”)

Secondo livello gerarchico:

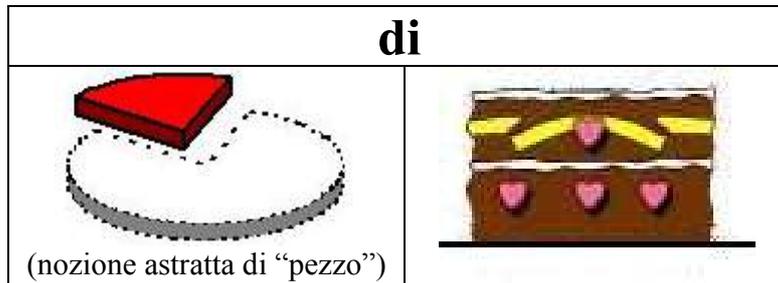
preposizione “di”

Terzo livello gerarchico:



(rappresentazione di “dolce”)

ma questo:



il cui significato è:



Poi, noi correliamo *l'intera correlazione* “pezzo di torta” (*e non solo la nozione di “pezzo”*) con la rappresentazione della “ciliegina”. Il correlatore “con” indica che due oggetti distinti sono in una relazione tale che l'attenzione è portata a focalizzarli insieme⁴. Anche in questo caso entrambi i correlati e il correlatore sono presenti insieme nella nostra mente e formano un tutto unico, e cioè questo:

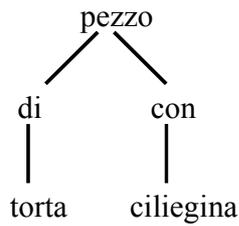


che significa questo:



Rapporti di questo tipo non possono essere rappresentati graficamente mediante degli alberi di dipendenza. Infatti, quello sopra riportato

⁴ L'analisi qui proposta è una mia modifica dell'analisi originale di Ceccato (che invece è: “due cose vengono focalizzate insieme dall'attenzione e poi vengono separate da essa”).

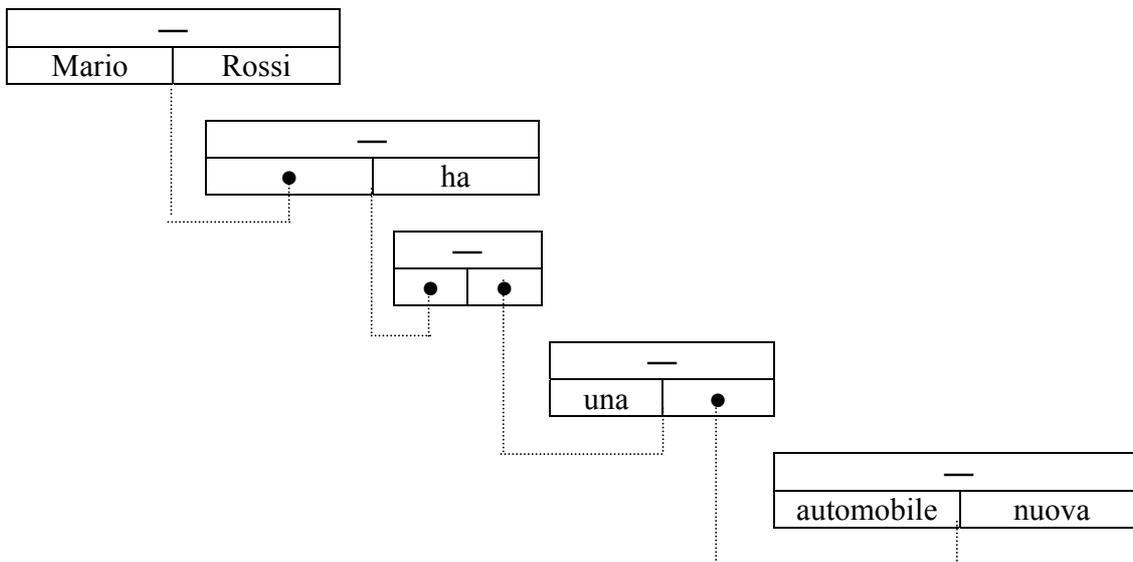


non indica, né può indicare, il fatto che, come abbiamo appena visto, *l'intera correlazione* “pezzo di torta” (e non solo la nozione di “pezzo”) è correlata con la rappresentazione della “ciliegina”.

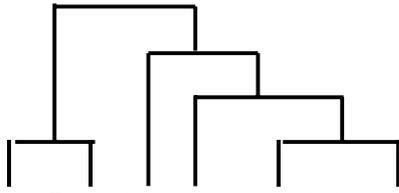
La rappresentazione grafica più adatta è quella che abbiamo visto, cioè degli elementi grafici costituiti da una casella superiore, quella del correlatore, che per così dire “abbraccia” le due caselle inferiori, quelle dei correlati.

La triade correlazionale così costituita può poi essere il correlato di un'altra triade correlazionale. Questo è indicato da una linea tratteggiata che parte dal centro dell'elemento grafico che rappresenta la prima triade (a simboleggiare che essa deve essere presa nel suo complesso) e termina con un punto al centro di una delle due caselle dei correlati della seconda triade. La rappresentazione grafica della struttura di una frase viene così ad assomigliare a una rete, ad un tessuto, e non alle ramificazioni di un albero.

Nei casi in cui in una frase non ci siano né preposizioni né congiunzioni, si può avere la falsa impressione che la rete correlazionale sia la stessa cosa che un albero di dipendenza. Per esempio la rete correlazionale corrispondente alla frase “Mario Rossi ha una automobile nuova”, cioè questa:



può sembrare la stessa cosa che l'albero di dipendenza corrispondente:

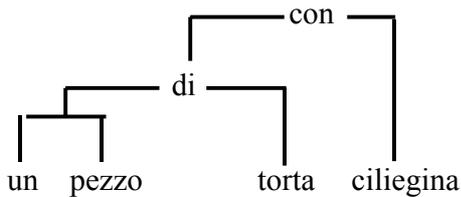


Mario Rossi ha una automobile nuova

Ma questo è dovuto solo al fatto che in questa frase non ci sono né preposizioni né congiunzioni e quindi non abbiamo dovuto collocare *delle parole* della frase in maniera tale da mostrare che esse legano altre parole.

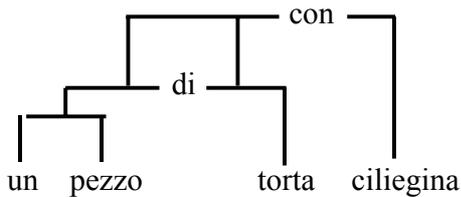
Se tentiamo di rappresentare con una struttura in qualche modo simile ad un albero la rete correlazionale del pensiero, il risultato non è buono perché questo tipo di rappresentazione è del tutto innaturale per essa. Infatti dovremmo costruire, per tutte le frasi che contengono preposizioni o congiunzioni (come, per esempio, “un pezzo di torta con ciliegina”), delle strutture come questa:

[1]

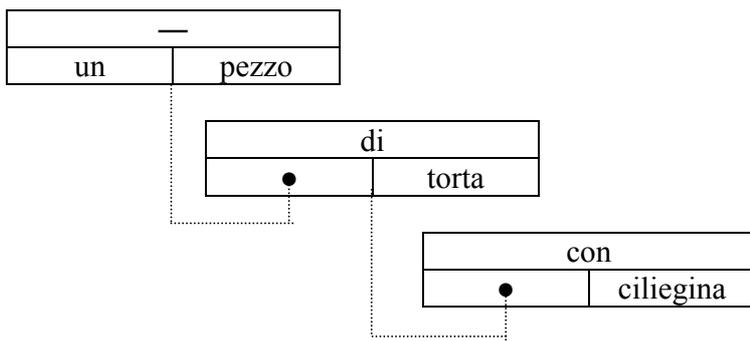


che è del tutto insoddisfacente perché la preposizione “di” sembra essere legata alla preposizione “con”; oppure come questa:

[2]



anch'essa molto insoddisfacente perché da essa non risulta affatto chiaro quello che è invece chiarissimo in questo tipo di rappresentazione:



cioè che la correlazione “un-pezzo” (il cui correlatore è il correlatore implicito) è il primo correlato della correlazione che ha come correlatore la preposizione “di” e come secondo correlato “torta” e che *tutta questa struttura nel suo insieme* è il primo correlato di un’altra correlazione il cui correlatore è la preposizione “con” e il secondo correlato “ciliegina”.

A prescindere dal fatto che strutture ad albero come la [1] o la [2] sono innaturali e insoddisfacenti come rappresentazioni della rete correlazionale del pensiero, esse sono comunque profondamente diverse, come si vede bene, dai tradizionali alberi di dipendenza, perché alcune parole della frase sono *dentro* le ramificazioni che costituiscono la struttura dell’albero e non *attaccate* ad esse.

Se vogliamo fare un paragone, grossolano ma che può servire a rendere visivamente l’idea, possiamo paragonare la rete correlazionale del pensiero ad una struttura come un telaio, un traliccio metallico, dove ci sono dei pezzi (sbarre, travi ecc.) che sono tenuti insieme da altri pezzi che hanno appunto questa specifica funzione di tenere insieme (bulloni, dadi ecc.), mentre la struttura della frase secondo le linguistiche tradizionali assomiglia ad un albero, con il suo tronco, i suoi rami e le sue foglie attaccate ad essi.

Questo paragone serve anche ad evidenziare un fatto molto importante che deve essere ben sottolineato, cioè che secondo la teoria correlazionale del pensiero, a differenza delle linguistiche convenzionali, la struttura della frase *non è gerarchica*, cioè non ci sono delle parole più “importanti” delle altre che reggono queste ultime (così come in un albero il tronco è più “importante” dei rami che sono più “importanti” delle foglie) ma tutti gli elementi sono egualmente essenziali (come in un traliccio o telaio, dove i bulloni sono essenziali come le sbarre che essi uniscono). Il fatto che il nome che funge da soggetto e il verbo personale siano considerati gli elementi fondamentali della proposizione deriva, come abbiamo accennato, semplicemente dal fatto che la maggior parte delle frasi che noi produciamo contengono almeno questi due elementi, perché una frase serve generalmente a parlare di qualche cosa che può esistere indipendentemente, cioè una cosa designata da un nome⁵, e di ciò che succede a questa cosa nel tempo, il che è designato da un verbo, che assume la forma personale perché è il secondo correlato di un nome (che diventa così il suo soggetto). Ma non sempre è così. Qualche volta, specialmente nel linguaggio parlato, noi produciamo frasi prive della struttura soggetto-verbo personale, come ad esempio “Ecco Mario” o “I miei migliori auguri a te e famiglia” o “Che persona gentile!”⁶. Secondo la teoria correlazionale del pensiero, dunque, certe parole sono più “importanti” di altre solo da un punto

⁵ L’aggettivo, invece designa una *caratteristica* di una cosa, cioè qualcosa che non può esistere indipendentemente. Per esempio, ciò che è designato dal nome “penna” può esistere indipendentemente, mentre ciò che è designato dall’aggettivo “rossa” non può esistere indipendentemente da qualcosa che è rosso. L’unica eccezione è data dai nomi cosiddetti astratti (“leggerezza”, “felicità”, “bellezza” ecc.) in cui siamo noi che consideriamo isolatamente una caratteristica di una cosa per parlare specificamente di essa. Per tali motivi nella correlazione sostantivo-aggettivo il sostantivo è sempre il primo correlato, anche se in certe lingue, come ad esempio l’inglese, l’aggettivo generalmente precede il nome a cui è riferito.

⁶ Si potrebbe pensare che in casi come questi un verbo è sottinteso. Ma bisogna notare che in questi casi i verbi che possono essere sottintesi sono in genere più di uno e anche molto diversi fra loro. Per esempio, nell’espressione “Che persona gentile!” i verbi che potrebbero essere sottintesi sono “è”, “ho incontrato”, “ho conosciuto” e ancora altri. È molto più semplice pensare che in questi casi a livello del pensiero non c’è nessun verbo.

di vista pratico, mentre a livello della struttura del pensiero, cioè a livello della “struttura profonda” corrispondente ad una frase⁷, non ci sono gerarchie ma solo coppie di correlati legati da un correlatore esplicito o implicito e così formanti delle triadi, a loro volta legate da correlatori. È certamente vero che nella frase certe parole sono subordinate a certe altre (per esempio in una frase come “Giovanni ha occhi azzurri” il verbo “ha” dipende dal soggetto “Giovanni” e l’aggettivo “azzurri” dipende dal sostantivo “occhi”, ed entrambi dipendono dal verbo). Ma, per la M.O., questo “dipendere” significa soltanto, nell’esempio che abbiamo appena fatto, che noi per prima cosa focalizziamo la nostra attenzione su “Giovanni” e, mantenendolo presente, costituiamo successivamente la categoria mentale corrispondente al verbo “ha”, poi correliamo la triade “Giovanni-correlatore implicito-ha” con la triade “occhi-correlatore implicito-azzurri” previamente costituita. È una dipendenza che origina dal fatto che noi *prima* focalizziamo con l’attenzione una certa cosa e *poi*, mantenendo presente la prima, ne focalizziamo una seconda, rendendo così la seconda cosa riferita alla prima, ma il tipo di rapporto (la triade correlazionale) che lega le parole di questa frase e di tutte le frasi è sempre lo stesso e non ci sono parole più “importanti”. La sola cosa che può essere considerata una “gerarchia” è il fatto che il primo correlato precede nel tempo il secondo correlato e pertanto è quest’ultimo che è riferito al primo e non viceversa.

3) La M.O., per quanto riguarda le parti del discorso, usa la stessa terminologia della grammatica tradizionale, parla cioè di “articolo”, “nome”, “aggettivo”, “verbo”, “avverbio” ecc.. In questo, dunque, essa sembra non aver portato alcuna novità rispetto alla grammatica tradizionale. Questo è dovuto al fatto che la teoria linguistica della M.O. non è in alcun modo contrapposta alla grammatica tradizionale. Al contrario, la M.O. riconosce che le distinzioni e le classificazioni operate dalla grammatica tradizionale sono perfettamente valide e molto utili. Ma fra grammatica tradizionale e teoria linguistica della M.O. vi sono due differenze. La prima, meno importante, è che le congiunzioni e le preposizioni vengono classificate insieme nella categoria dei “correlatori”, mentre la categoria dei “pronomi” viene inserita in una categoria più ampia avente un altro nome e in cui vi sono anche altri elementi (il perché di questo fatto non può essere spiegato in lavoro relativamente breve come questo). La seconda differenza, molto più importante, è che la M.O. non si limita ad accettare la classificazione delle parti del discorso operata dalla grammatica tradizionale, ma *definisce* le parti del discorso in maniera a mio avviso soddisfacente, cosa che non sono riuscite a fare la grammatica tradizionale e le altre linguistiche (nell’articolo suddetto abbiamo visto la definizione di “nome”, mentre in questo vedremo quella di “verbo personale”). Del fatto che la grammatica tradizionale e le altre linguistiche non siano riuscite a definire le parti del discorso in maniera soddisfacente sono ben consapevoli i linguisti stessi, come risulta chiaramente da questo passo tratto da un trattato di linguistica recente:

“Le parti del discorso possono essere perciò riconosciute in base a criteri puramente **distribuzionali**: i nomi, i verbi, ecc.. saranno definiti in base alle altre classi di parole assieme alle quali possono, oppure non possono, ricorrere. La definizione precisa delle varie parti del discorso in termini distribuzionali è un’operazione complessa, ma abbiamo visto che le definizioni tradizionali di tipo semantico sono decisamente inadeguate, e un’impostazione distribuzionale è la via più promettente per superare queste difficoltà.”⁸ [sottolineature mie]

In rapporto al problema della traduzione automatica, la differenza fondamentale fra la teoria linguistica della M.O. e le altre teorie linguistiche è, come abbiamo detto, quella esposta al punto 2). Il significato complessivo di una frase è dato da due cose:

- 1) i significati delle singole parole che la compongono;
- 2) il modo in cui le parole sono legate fra loro, cioè la struttura del pensiero corrispondente.

L’essere umano che comprende il significato complessivo di una frase, azione che precede sempre la

⁷ L’espressione “struttura profonda” è naturalmente qui usata nel senso più generale che essa ha in linguistica, cioè senza alcun riferimento alla grammatica generativa di Chomsky.

⁸ Giorgio Graffi, Sergio Scalise, *Le lingue e il linguaggio*, 2003, pag. 118.

traduzione, capisce i significati delle singole parole e ricostruisce la struttura del pensiero corrispondente. Oggi è assolutamente impossibile costruire una macchina che capisca i significati delle parole, ma è possibile cercare di costruire un programma che, partendo da una frase, cerchi di ricostruire la struttura del pensiero corrispondente. *Il fatto che ci siano certe parole che possono essere solo dei correlatori e altre possano essere solo dei correlati è un potente strumento per individuare qual è la corretta struttura della rete correlazionale corrispondente a una frase.* Questo per due motivi:

1) le parole che designano correlatori possono andare solo nelle caselle superiori delle triade, mentre quelle che designano correlati possono andare solo nelle caselle inferiori;

2) certi correlatori, come vedremo, ammettono come correlati solo certi tipi di correlati.

Anche alcuni programmi di traduzione automatica oggi esistenti cercano in qualche modo di individuare la struttura delle frasi, cioè come sono fatti i suddetti “alberi”, ma non dispongono dello strumento che abbiamo ora descritto. Per far questo essi dispongono solo:

1) delle regole grammaticali di una lingua (per esempio quelle che stabiliscono che ci deve essere una concordanza di genere e numero fra nome e aggettivo, di persona tra forma personale del verbo e suo soggetto ecc., regole che sfrutta anche il programma che descriveremo);

2) di una “sfera nozionale” che cerca di riprodurre la conoscenza del mondo reale che hanno gli esseri umani, “sfera nozionale” che, nel caso del nostro programma, ha dimensioni enormi.

I principali problemi della traduzione automatica

Come ognuno sa, per tradurre un testo occorre sostituire le parole della lingua di origine, disposte in un certo ordine, con appropriate parole della lingua di uscita, disposte anch'esse in un certo ordine, che spesso è diverso dall'altro. Se ad ogni parola della lingua di origine corrispondesse sempre una, ed una sola, parola della lingua di uscita e l'ordine delle parole fosse lo stesso, la traduzione automatica non sarebbe un problema. Sfortunatamente le cose non stanno affatto così. Una certa quantità (che varia secondo la coppia di lingue considerate) delle singole parole della lingua di origine deve essere sostituita da parole della lingua di uscita diverse a seconda dei casi. Per l'essere umano che traduce questo non costituisce un problema. Egli, partendo dal testo della lingua di origine, *comprende* il testo di origine, cioè ricostruisce *uno ed un solo* significato di esso, quindi esprime questo significato univoco usando le parole e le regole della lingua di uscita. Naturalmente, la macchina non può fare niente di tutto questo. Da questo fatto nascono molti problemi. I principali sono i seguenti due.

1) Il primo problema è che vi sono delle parole, elencate come voci nel dizionario, che hanno più di un significato. Può trattarsi di due o più significati non in relazione fra loro (per esempio “riso”, che può indicare l'atto del ridere o un alimento); oppure, molto più frequentemente, derivati l'uno dall'altro (per esempio “capo”, che può indicare una parte del nostro corpo o una persona che comanda). Inoltre può darsi il caso che la flessione del nome, del verbo ecc. porti alla formazione di parole eguali ad altre (per esempio “amo”, che può essere un nome che indica un attrezzo da pesca o una voce del verbo “amare”). Questi casi non sono un problema per l'essere umano perché egli può capire, basandosi sul contesto, qual è il significato che di volta in volta è in gioco. Se, per esempio, si sta parlando di un fucile e nel testo da tradurre figura la parola italiana “cane” l'uomo che traduce capirà subito che essa qui non designa il comune animale domestico, ma una parte meccanica di cui sono dotate certe armi da fuoco.

Strettamente connesso a questo problema è quello che nelle lingue con scarsa morfologia, come per esempio l'inglese, una stessa parola può essere nome, verbo, aggettivo ecc. e il ruolo che la parola gioca di volta in volta può essere stabilito solo in base alla struttura generale della frase e al suo significato. Nel caso che la lingua di uscita sia una lingua con ricca morfologia questo è un problema molto importante, perché di regola queste lingue hanno parole diverse per la forma nominale, verbale ecc..

2) Il secondo grosso problema è che, più o meno frequentemente a seconda della lingua, *non tutti* i rapporti che costituiscono una rete correlazionale di pensiero sono indicati nel discorso che è l'espressione di quel pensiero. Si considerino i seguenti esempi:

- *una certa quantità di acqua, anche molto piccola*
- *una certa quantità di acqua, anche molto fredda*
- *bottiglia di grappa vuota*
- *bottiglia di grappa veneta*

Nel primo esempio l'aggettivo finale "piccola" è riferito, a livello del pensiero, a "quantità", mentre nel secondo esempio l'aggettivo finale "fredda" è invece riferito a "acqua". Nel terzo esempio l'aggettivo "vuota" è riferito a "bottiglia", mentre nel quarto l'aggettivo "veneta" è riferito a "grappa". Ma, nell'espressione italiana di questi pensieri, non c'è nulla che indichi che le cose stanno così. In altre parole, in questi casi le informazioni grammaticali presenti nel discorso (cioè la morfologia e l'ordine delle parole) non permettono di risalire ad un'unica rete correlazionale del pensiero. L'essere umano comprende i significati delle parole "quantità", "acqua", "piccola", "fredda", "bottiglia", "grappa", "vuota", "veneta" e, servendosi di quella cultura di base di cui tutti sono dotati (nei testi che trattano di traduzione automatica essa è spesso chiamata "conoscenza del mondo reale"), capisce subito qual è la rete correlazionale corretta. Anche in questo caso, il problema nella traduzione automatica dipende da qual è la lingua di uscita. Quando, per esempio, si traducono le prime due frasi dall'italiano ad un'altra lingua anch'essa priva dei casi, come per esempio l'inglese, non sorge alcun problema perché in un caso come questo la semplice traduzione parola per parola dà un risultato soddisfacente, ma non è così se si traduce in una lingua dotata dei casi, come per esempio il tedesco o il russo. In questo caso solo la comprensione della frase italiana permette di tradurre i due aggettivi finali "piccola" e "fredda" in maniera corretta, e cioè flettendo il primo al caso nominativo e il secondo al caso genitivo. In casi come questi i programmi attuali traducono o sostanzialmente a caso, e quindi con una probabilità di errore elevata o, se ricorrono a qualche accorgimento, questo si dimostra insufficiente, come vedremo più avanti.

Si potrebbe pensare che casi come questi non sono troppo frequenti. Non è affatto così. La frequenza di questo fenomeno varia da lingua a lingua, ma esso è sicuramente presente in moltissime lingue e probabilmente in tutte. Nelle lingue con una ricca morfologia, cioè dotate della flessione del nome, e magari anche dell'aggettivo, secondo vari casi (nominativo, accusativo, genitivo, dativo ecc..) e della flessione del verbo a seconda della persona a cui esso è riferito, queste ambiguità sono meno frequenti che nelle lingue che hanno una morfologia molto scarsa, come l'inglese. Ma anche nel caso delle prime il fenomeno è presente. Ad ogni modo l'inglese è una lingua estremamente diffusa come madrelingua e ancor più come seconda lingua, pertanto disporre di programmi che traducano da questa lingua in maniera soddisfacente è certo altamente desiderabile.

Naturalmente, come si capisce bene, questo fenomeno causa problemi particolarmente gravi quando occorre tradurre da una lingua con scarsa morfologia ad una lingua con ricca o ricchissima morfologia, come lo sono molte lingue indoeuropee (ad esempio lo spagnolo, l'italiano, il francese, il tedesco, il russo ecc.). Se si pensa a quanto una traduzione di questo tipo è richiesta, si capisce bene che l'importanza che questo problema viene ad assumere è enorme.

Vediamo dunque degli esempi di quello che succede nella traduzione da una lingua con scarsa morfologia ad una con ricca morfologia. In inglese, per esempio, il verbo viene flesso pochissimo in rapporto alla persona a cui si riferisce: in pratica solo la terza persona dell'indicativo presente ha una desinenza diversa dalle altre persone. Inoltre anche l'infinito è uguale all'indicativo, e le forme personali del passato dei verbi regolari sono eguali al participio passato. Per giunta, il verbo spesso non differisce, come si è accennato, dal sostantivo. Per esempio, la parola "love" dovrà essere tradotta, a seconda del contesto, con una delle seguenti parole italiane: "amore"; "amo", "ami", "ama", "amiamo", "amate", "amano"; "amate", "amino". La parola "loved" dovrà essere tradotta con una delle seguenti parole (o coppia di parole) italiane: "amai", "amasti", "amò", "amammo", "amaste", "amarono"; "amavo", "amavi",

“amava”, “amavamo”, “amavate”, “amavano”; “avevo amato”, “avevi amato”, “aveva amato”, “avevamo amato”, “avevate amato”, “avevano amato”; “ebbi amato”, “avesti amato”, “ebbe amato”, “avemmo amato”, “aveste amato”, “ebbero amato”; “amassi”, “amasse”, “amassimo”, “amaste”, “amassero”; “abbia amato”, “abbiamo amato”, “abbiate amato”, “abbiano amato”; “avessi amato”, “avesse amato”, “avessimo amato”, “aveste amato”, “avessero amato”; “amato”, “amata”, “amati”, “amate”. Inoltre l’inglese non ha generi, mentre molte altre lingue hanno il maschile, il femminile e diverse anche il neutro. Anche in questo caso, un’unica parola inglese, per esempio un aggettivo, dovrà essere tradotta in due o tre modi diversi a seconda del sostantivo a cui si riferisce (più altri due o tre modi, se si pensa che di regola in queste lingue c’è flessione secondo il genere anche al plurale).

Naturalmente, anche nelle lingue con scarsa morfologia, non sempre occorre ricorrere alla comprensione dei significati delle parole e alla conoscenza del mondo reale per ricostruire le corrette reti correlazionali. Da una parte perché anche queste lingue hanno una certa morfologia (ad esempio, in inglese, l’aggiunta della desinenza “-s” al tema del verbo quando esso è riferito ad una terza persona singolare e coniugato all’indicativo presente), ma soprattutto perché ricorrono ad un ordine delle parole molto rigido ed anche perché certe relazioni sono comprensibili tenendo conto della struttura generale della frase. Ad esempio, nella frase inglese “*John hit Bob*” (“Giovanni picchiò Roberto”) il fatto che il soggetto sia “*John*” e il complemento oggetto sia “*Bob*” è indicato inequivocabilmente dall’ordine delle parole ed il fatto che *hit* sia un verbo e non un nome (il che potrebbe essere benissimo, da un punto di vista morfologico) è ricavabile dalla regola quasi generale che le frasi hanno un verbo in forma personale (possiamo pertanto fare in modo che un programma tenga conto di questo e se ne serva per fare le scelte corrette).

Come si vede bene, un programma che fosse in grado di ricostruire la giusta rete correlazionale del pensiero a partire dal testo garantirebbe la completa assenza di errori nella traduzione delle numerosissime forme non flesse presenti nelle lingue con scarsa morfologia che devono essere convertite in forme flesse di lingue con ricca morfologia.

Naturalmente, fin dall’inizio della storia della traduzione automatica si è compreso che, per operare una traduzione automatica corretta, è necessario ricostruire la struttura della frase basandosi su una qualche teoria linguistica. I programmi esistenti sono già dotati di accorgimenti che permettono in una certa misura di far questo, ma i risultati che danno non sono, come vedremo, del tutto soddisfacenti. Il programma che descriverò, invece, tenta *sempre* di operare una ricostruzione *integrale* della struttura della frase secondo *un’unica e rigorosa* teoria linguistica, basandosi su un complesso ritrovato in cui la simulazione della conoscenza del mondo reale gioca un ruolo fondamentale. Esso pertanto simula da vicino ciò che fa l’essere umano che traduce.

Per questo motivo ritengo che questo ritrovato conservi a tutt’oggi tutto il suo valore teorico e pratico e che potrebbe permettere di ottenere nel campo della traduzione automatica risultati decisamente migliori rispetto a quelli ottenibili oggi con i programmi disponibili che non sono basati su questo ritrovato e sulla teoria correlazionale del pensiero. È solo un problema di tipo pratico e non teorico il fatto che per realizzare questo ritrovato è necessario un notevole lavoro umano di tipo linguistico prima di costruire il programma, e che questo lavoro aumenta enormemente all’aumentare del numero dei termini del dizionario di cui è dotato il programma. Disponendo di risorse sufficienti questo problema potrebbe essere superato. L’impiego di una quantità di risorse elevata sarebbe ampiamente ripagata dalla grande ricaduta commerciale che sicuramente avrebbe un programma capace di eseguire una traduzione automatica di qualità sostanzialmente superiore a quella ottenibile con gli attuali programmi. A questo proposito è importante notare che la qualità della traduzione automatica non sembra essere sostanzialmente migliorata negli ultimi dieci anni circa, né sembra che gli approcci fin qui usati promettano significativi miglioramenti per il futuro. Cito al riguardo due passi molto significativi di uno studioso della massima autorevolezza come W. J. Hutchins.

The ‘real’ question is whether the *quality* of machine translation has improved. Some will say that it *must* have. Others of an older generation insist that MT has the same problems and throws up the same errors as it did in the 1960s and 1970s. Many, however, will perhaps share my intuition that while progress in quality is not evident during say the last ten years, there are discernible advances since the early 1980s and definite improvements since the ALPAC report [see note 1, editor’s note] of the mid 1960s.⁹

[La “vera” domanda è se la *qualità* della traduzione automatica è migliorata. Alcuni diranno che *deve* essere migliorata. Altri di una generazione più vecchia insistono che la traduzione automatica ha gli stessi problemi e produce gli stessi errori come faceva negli anni 60 e 70. Molti, comunque, forse condivideranno la mia intuizione che mentre il progresso nella qualità non è evidente durante, diciamo, gli ultimi dieci anni, ci sono discernibili avanzamenti dai primi anni ‘80 e miglioramenti ben definiti dai tempi del rapporto ALPAC [vedi nota 1, n.d.r.], cioè dalla metà degli anni 60.]

Despite the prospects for the future, it has to be said that the new approaches of the current decade have not yet resulted in substantial improvements in the quality of the raw output from translation systems. These improvements may come in the future, but overall it has to be admitted that at present the actual translations produced do not represent major advances on those made by the MT systems of the 1970s. We still see the same errors: wrong pronouns, wrong prepositions, garbled syntax, incorrect choice of terms, plurals instead of singulars, wrong tenses, etc., etc. – errors that no human translators would ever commit.

[...]

Unfortunately, this situation will probably not change in the near future. There is little sign that basic general purpose MT engines are going to show significant advances in translation quality for many years to come.¹⁰

[A dispetto delle prospettive per il futuro, bisogna dire che i nuovi approcci del corrente decennio non hanno ancora dato dei sostanziali miglioramenti nella qualità del prodotto grezzo dei sistemi di traduzione. Questi miglioramenti possono venire in futuro, ma complessivamente bisogna ammettere che al momento le traduzioni reali prodotte non rappresentano importanti avanzamenti rispetto a quelle fatte dai sistemi di traduzione automatica degli anni 70. Noi vediamo sempre gli stessi errori: pronomi errati, preposizioni errate, sintassi alterata, scorretta scelta dei termini, plurali invece che singolari, tempi sbagliati, ecc. ecc. – errori che nessun traduttore umano commetterebbe mai.

[...]

Sfortunatamente, questa situazione probabilmente non cambierà nel prossimo futuro. Ci sono pochi segni che i motori di traduzione automatica di base per uso generale siano sul punto di mostrare significativi avanzamenti nella qualità della traduzione per molti anni a venire.]

Vediamo ora qualche esempio pratico di errori che compiono gli attuali programmi di traduzione automatica. Il primo esempio è costituito dalle versioni inglese e francese di un testo di una semplicità elementare. Si tratta di informazioni per i turisti prese dal sito Internet di un notissimo museo. È naturale che il redattore abbia fatto ogni sforzo perché il testo risulti estremamente semplice e chiaro. Infatti quasi tutte le parole sono di uso comune e le frasi, tutte molto brevi, sono quasi tutte costituite da un’unica proposizione.

Anziché riportare tutto il brano e, dopo, tutta la traduzione, ho disposto sotto ciascun paragrafo del testo originale la sua traduzione, in modo che il lettore possa più facilmente confrontare il testo di origine con quello tradotto. La traduzione operata dal programma (uno dei più noti tra quelli oggi disponibili¹¹) è scritta con caratteri rossi. La traduzione compiuta dall’essere umano invece è scritta con caratteri blu. Gli errori presenti nel testo tradotto automaticamente sono evidenziati in giallo (il significato delle sottolineature verrà spiegato più avanti), mentre le improprietà, le inesattezze ecc. sono evidenziate in grigio. La traduzione corretta delle parti che risultano invece errate nel testo tradotto automaticamente è stata evidenziata con gli stessi due colori in modo che il lettore la possa facilmente individuare.

⁹ Hutchins W. J., ‘Has machine translation improved?’. *MT Summit IX: proceedings of the Ninth Machine Translation Summit, New Orleans, USA, September 23-27, 2003*, 181-188. [East Stroudsburg, PA: AMTA.]

¹⁰ Hutchins W. J., ‘*Retrospect and prospect in computer-based translation*’. In *Machine Translation Summit VII, 13th-17th September 1999, Kent Ridge Digital Labs, Singapore. Proceedings of MT Summit VII “MT in the great translation era”*, 30-34. [Tokyo]: Asia-Pacific Association for Machine Translation, 1999.

¹¹ Vengono riportati i risultati prodotti da un solo programma perché i risultati prodotti dai migliori programmi attualmente disponibili non differiscono in maniera sostanziale l’uno dall’altro.

Anche se il lettore conosce la lingua di origine, è bene che prima legga il testo della traduzione automatica (cioè tutto il testo in caratteri rossi), in modo da farsi un'idea della qualità globale di questa. Dopo potrà passare ad esaminare singolarmente gli errori di traduzione.

TESTO ORIGINALE SCRITTO IN FRANCESE

Aide à la visite

Plan/Information

Un plan/information gratuit, en 9 langues (français, anglais, espagnol, allemand, italien, japonais, chinois, néerlandais, arabe), est disponible à la banque d'information. Il permet au visiteur de s'orienter dans le musée et les collections.

Aiuto alla visita

Piano/informazione

Un piano/informazione gratuito, in 9 lingue inglese (francesi, spagnola, tedesco, italiano, giapponesi, cinesi, olandesi, arabo), è disponibile alla banca d'informazione. Permette all'ospite di orientarsi nel museo e le raccolte.

Aiuto alla visita

Piano/informazione

Un piano/informazione gratuito, in 9 lingue inglese (francese, spagnolo, tedesco, italiano, giapponese, cinese, olandese, arabo), è disponibile al banco informazioni. Permette all'ospite di orientarsi nel museo e nelle raccolte.

Audioguides

Un audioguide (lecteur individuel de CD ROM) est à la disposition des visiteurs du musée; il comprend des commentaires en français, anglais et espagnol sur plus de 1000 oeuvres et lieux majeurs du musée, sans parcours préétabli. Le commentaire est disponible aussi en allemand, italien et japonais pour 350 oeuvres.

L'audioguide est proposé aux 3 entrées du musée (Sully, Denon et Richelieu, niveau entresol). Son prix de location est de 5 euros, en échange d'une pièce d'identité.

Règlement uniquement en espèces, chèque bancaire ou chèque de voyage.

Audioguides

Audioguide (lettore individuale di CD ROM) è a disposizione degli ospiti del museo; comprende commenti in francese, inglese e spagnolo su più di 1000 opere e luoghi principali del museo, senza percorso prestabilito. Il commento è disponibile così in tedesco, italiano e giapponese per 350 opere.

Audioguide è proposto alle 3 entrate del museo (Sully, Denon e Richelieu, livello entresol). Il suo prezzo d'affitto è di 5 euro, in cambio di un documento d'identità.

Regolamento soltanto in specie, assegno bancario o assegno di viaggio.

Audioguide

L'Audioguida (lettore individuale di CD ROM) è a disposizione degli ospiti del museo; comprende commenti in francese, inglese e spagnolo su più di 1000 opere e luoghi principali del museo, senza percorso prestabilito. Il commento è disponibile anche in tedesco, italiano e giapponese per 350 opere.

Audioguide è proposto alle 3 entrate del museo (Sully, Denon e Richelieu, al mezzanino). Il suo prezzo d'affitto è di 5 euro, in cambio di un documento d'identità.

Pagamento soltanto in contanti, assegno bancario o traveller's cheque.

Les feuillets

Dans les salles du musée se trouvent des "feuillets", textes édités en plusieurs langues (français, anglais, allemand, espagnol, italien, japonais), sous forme de plaques plastifiées mobiles. Il s'agit de textes explicatifs des collections, portant sur une oeuvre en particulier, un artiste, un style, un site, une technique...

La version française en format réduit est en vente à la librairie, hall Napoléon.

Gli strati

Nelle sale del museo si trovano "strati", testi pubblicati in molte lingue (francesi, inglesi, tedesche, spagnolo, italiano, giapponesi), sotto forma di targhe plastificate mobili. Si tratta di testi esplicativi delle raccolte, riguardanti un'opera in particolare, un artista, uno stile, una località, una tecnica...

La versione francese in formato ridotto è in vendita alla libreria, hall Napoleone.

I foglietti

Nelle sale del museo si trovano dei "foglietti", testi pubblicati in molte lingue (francese, inglese, tedesco, spagnolo, italiano, giapponese), sotto forma di targhe plastificate mobili. Si tratta di testi esplicativi delle raccolte, riguardanti un'opera in particolare, un artista, uno stile, una località, una tecnica...

La versione francese in formato ridotto è in vendita alla libreria, hall Napoleone.

Public handicapé

Un guide d'orientation spécifique (en français et en anglais) pour les visiteurs à mobilité réduite indiquant les itinéraires avec ascenseurs est disponible à la banque d'information, ainsi que des fauteuils roulants.

Pour les visiteurs aveugles et malvoyants, le département des sculptures a ouvert un espace tactile regroupant une vingtaine d'œuvres que l'on peut découvrir par le toucher et à l'aide d'un audioguide.

Des visites-conférences pour les groupes d'handicapés peuvent être organisées sur demande.

Tel. : (33) 01 40 20 59 90

Pubblico minorato

Una guida d'orientamento specifico (in francese ed in inglese) per gli ospiti con mobilità ridotta che **indicano** gli itinerari con ascensori è disponibile **alla banca d'informazione**, **così soltanto** delle poltrone mobili.

Per gli ospiti ciechi e portatori di handicap visivo, il dipartimento delle sculture ha aperto uno spazio tattile che raccoglie una ventina di opere che si possono scoprire con il **contatto** ed **all'**aiuto di audioguide.

Visite-conferenza per i gruppi di minorati possono essere organizzate su domanda.

Tel.: (33) 01.40.20.59.90

Pubblico minorato

Una guida d'orientamento specifico (in francese ed in inglese) per gli ospiti con mobilità ridotta che **indica** gli itinerari con ascensori è disponibile **al banco informazioni**, **così come** delle poltrone mobili.

Per gli ospiti ciechi e portatori di handicap visivo, il dipartimento delle sculture ha aperto uno spazio tattile che raccoglie una ventina di opere che si possono scoprire con il **tatto** e **con l'**aiuto di audioguide.

Visite-conferenza per i gruppi di minorati possono essere organizzate su domanda.

Tel.: (33) 01.40.20.59.90

Règlement de visite

Pour le confort des visiteurs, l'usage des flashes est vivement déconseillé.

Il est interdit de fumer dans le Hall Napoléon et dans les salles du musée.

Ne sont pas acceptés au vestiaire, ni à la bagagerie, ni dans les collections: les aliments, les boissons, les animaux, les gros sacs et les valises.

Le livret résumant le règlement de visite est à disposition des visiteurs à la banque d'information sous la pyramide.

Regolamento di visita

Per la comodità degli ospiti, l'impiego **degli** flashes è vivamente sconsigliato.

È vietato fumare nell'hall Napoleone e nelle sale del museo.

Non sono accettati al guardaroba, né alla **bagagerie**, né nelle **raccolte**: i prodotti alimentari, le bevande, gli animali, le grandi borse e le **borse**.

L'opuscolo che riassume **[...]** regolamento di visita è a disposizione degli ospiti alla banca d'informazione sotto la piramide.

Regolamento di visita

Per la comodità degli ospiti, l'impiego **dei** flashes è vivamente sconsigliato.

È vietato fumare nell'hall Napoleone e nelle sale del museo.

Non sono accettati al guardaroba, né al **deposito bagagli**, né nelle **collezioni**: i prodotti alimentari, le bevande, gli animali, le grandi borse e le **valigie**.

L'opuscolo che riassume **[il]** regolamento di visita è a disposizione degli ospiti alla banca d'informazione sotto la piramide.

Pourquoi il ne faut pas toucher les oeuvres

Les oeuvres d'art sont uniques et fragiles.

Elles ont traversé les siècles et doivent être conservées pour les générations futures.

Toucher, même très légèrement, une peinture, un objet, une sculpture, un meuble l'abîme.

Surtout lorsque ce geste est répété des milliers de fois.

Aidez-nous à protéger notre patrimoine commun.

Perché non occorre toccare le opere

Le opere d'arte sono uniche e fragili.

Hanno attraversato i secoli e devono essere conservate per le generazioni future.

Toccare, anche molto leggermente, una **vernice**, un oggetto, una scultura, un mobile lo danneggia.

Soprattutto quando questo gesto è ripetuto migliaia di volte.

Aiutiamo a proteggere il nostro patrimonio comune.

Perché non bisogna toccare le opere

Le opere d'arte sono uniche e fragili.

Hanno attraversato i secoli e devono essere conservate per le generazioni future.

Toccare, anche molto leggermente, un **dipinto**, un oggetto, una scultura, un mobile lo danneggia.

Soprattutto quando questo gesto è ripetuto migliaia di volte.

Aiutateci a proteggere il nostro patrimonio comune.

TESTO ORIGINALE SCRITTO IN INGLESE

Visit Information

Handbook

A free handbook in 9 languages (French, English, Spanish, German, Italian, Japanese, Chinese, Dutch, Arabic) to help visitors find their way in the museum and the collections, is available from the Information Desk.

Le Informazioni Di Chiamata

Manuale

Un manuale libero in 9 lingue (francese, inglese, spagnolo, tedesco, italiano, giapponese, cinese, olandese, arabo) per aiutare gli ospiti a trovare il loro senso nel museo e nelle collezioni, è disponibile dallo scrittorio delle informazioni.

Informazioni di visita

Manuale

Un manuale gratuito in 9 lingue (francese, inglese, spagnolo, tedesco, italiano, giapponese, cinese, olandese, arabo) per aiutare gli ospiti a trovare la loro via nel museo e nelle collezioni, è disponibile al banco delle informazioni.

Audioguides

An audioguide is provided for visitors to the museum. It includes commentaries in English, French and Spanish on 1000 works and major locations within the museum, without the need to follow a specific route. Available as well in German, Italian and Japanese on 350 works. It can be obtained from the entrances to the three wings of the museum (Sully, Denon and Richelieu), on presentation of an identity card. Fee: 5 euros.

Payment only in cash, bank cheque or traveller's cheque.

Audioguides

Un audioguide è fornito per gli ospiti al museo. Include i commenti in inglese, francese e lo Spagnolo su 1000 impianti e le posizioni importanti all'interno del museo, senza la necessità di seguire un itinerario specifico. Disponibile pure in tedesco, italiano ed il giapponese su 350 impianti. Può essere ottenuto dalle entrate alle tre ale del museo (Sully, Denon e Richelieu), sulla presentazione di una carta di identità. Tassa: 5 euros.

Pagamento soltanto in denaro, l'assegno della serie o l'assegno del viaggiatore.

Audioguide

Una audioguida è fornita per gli ospiti al museo. Include i commenti in inglese, francese e spagnolo su 1000 opere e i luoghi importanti all'interno del museo, senza la necessità di seguire un itinerario specifico. Disponibile pure in tedesco, italiano e giapponese su 350 opere. Può essere ottenuta dalle entrate alle tre ali del museo (Sully, Denon e Richelieu), su presentazione di una carta di identità. Tassa: 5 euro.

Pagamento soltanto in denaro, assegno bancario o traveller's cheque.

Information sheets

Information sheets are available for consultation in the museum's galleries. These are hard plastic-covered pages, A3 size, available in several languages (French, English, German, Spanish, Italian, Japanese). The texts explain the collections in that room, or deal with a specific work, artist, style, location, technique etc.

The French version is on sale in reduced format at the bookshop in the Hall Napoléon.

Fogli delle informazioni

I fogli delle informazioni sono disponibili per consultazione nelle gallerie del museo. Queste sono pagine duro plastica-coperte, formato A3, disponibili in parecchie lingue (francese, inglese, tedesco, spagnolo, italiano, giapponese). I testi spiegano le collezioni in quanto stanza, o l'affare con un lavoro, un artista, uno stile, una posizione, una tecnica specifici ecc.

La versione francese è sulla vendita nella disposizione ridotta al bookshop nel Corridoio Napoléon.

Fogli delle informazioni

I fogli delle informazioni sono disponibili per consultazione nelle gallerie del museo. Queste sono pagine coperte con plastica dura, formato A3, disponibili in parecchie lingue (francese, inglese, tedesco, spagnolo, italiano, giapponese). I testi spiegano le collezioni in quella stanza, o trattano di un lavoro, un artista, uno stile, un luogo, una tecnica specifici ecc.

La versione francese è in vendita in formato ridotto alla libreria nella Hall Napoléon.

Disabled Visitors

An orientation guide (in French and English) for less mobile visitors offers a special tour, and is available at the Information Desk. Loan of wheelchairs upon request, tel.: (33) 01 40 20 53 17.

For blind and partially sighted visitors, the Department of Sculptures has opened a gallery where around twenty works can be explored by touch and with the aid of an audioguide.

Guided tours for groups of disabled persons can be organised on request, tel.: (33) 01 40 20 59 90.

Ospiti Disabled

Una guida di orientamento (in francese ed inglese) per gli ospiti meno mobili offre un giro speciale ed è disponibile allo **scrittorio** delle informazioni. Prestito delle sedie a rotelle su richiesta, tel.: (33) 01 40 20 53 17.

Per **i ciechi** e gli ospiti parzialmente **avvistati**, il reparto delle sculture ha aperto una galleria in cui **intorno** venti **impianti** possono essere esplorati tramite il **tocco** e con l'aiuto di un audioguide.

I giri guidati per i gruppi delle persone **disabled** possono essere organizzati a richiesta, tel.: (33) 01 40 20 59 90.

Ospiti Disabili

Una guida di orientamento (in francese ed inglese) per gli ospiti meno mobili offre un giro speciale ed è disponibile allo **banco** delle informazioni. Prestito delle sedie a rotelle su richiesta, tel.: (33) 01 40 20 53 17.

Per gli ospiti **ciechi** e con capacità visiva parziale, il reparto delle sculture ha aperto una galleria in cui **circa** venti **opere** possono essere esplorate tramite il **tatto** e con l'aiuto di un audioguide.

I giri guidati per i gruppi delle persone **disabili** possono essere organizzati a richiesta, tel.: (33) 01 40 20 59 90.

Regulations for Visitors

Please refrain from using flashes.

It is prohibited to smoke in the Hall Napoléon and in the galleries.

No food may be brought into the galleries or the cloakrooms.

Visitors are not allowed to bring cumbersome objects or animals into the galleries.

The Visit regulation book for visitors may be consulted at the Information desk.

Regolazioni per gli ospiti

Prego **astensione** dal **usando** i flash.

È proibita **per** fumare nel **Corridoio** Napoléon e nelle gallerie.

Nessun alimento **non** può essere introdotto nelle gallerie o nei guardaroba.

Agli ospiti **non sono permessi** introdurre gli oggetti o **gli animali ingombranti** nelle gallerie.

Visit regulation Il libro per gli ospiti può essere consultato allo **scrittorio** delle informazioni.

Regole per gli ospiti

Prego **astenersi** dall'**usare** i flash.

È proibito fumare nella **Hall** Napoléon e nelle gallerie.

Nessun alimento può essere introdotto nelle gallerie o nei guardaroba.

Agli ospiti **non è permesso** introdurre gli oggetti **ingombranti** o gli animali nelle gallerie.

Il libro del regolamento di visita per gli ospiti può essere consultato al **banco** delle informazioni.

Why works of art should not be touched

Works of art are unique and fragile.

They have survived centuries and must be preserved for future generations.

Touching, even lightly, a painting, object, sculpture or piece of furniture causes damage.

Especially when this gesture is repeated thousands of times.

Help us protect our common heritage.

Perchè le opere d'arte non dovrebbero essere toccate

Le opere d'arte sono uniche e fragili.

Hanno secoli superstiti e devono essere conservate per le generazioni future.

Toccando, persino leggermente, una **pittura**, **l'oggetto**, **la** scultura o **la parte di mobilia** causa danni.

Particolarmente quando questo **gesture** è migliaia **ripetute dei periodi**.

Aiutaci a proteggere la nostra eredità comune.

Perché le opere d'arte non dovrebbero essere toccate

Le opere d'arte sono uniche e fragili.

Sono sopravvissute per secoli e devono essere conservate per le generazioni future.

Toccando, persino leggermente, un **dipinto**, **un oggetto**, **una** scultura o **un mobile** causa danno.

Particolarmente quando questo **gesto** è **ripetuto migliaia di volte**.

Aiutaci a proteggere la nostra eredità comune.

Per quanto riguarda la traduzione in italiano del testo originale scritto in francese, l'impressione generale, direi, è che essa sia di discreto livello. Bisogna però notare che l'italiano e il francese sono due lingue vicinissime, tanto che un italiano riesce, con un po' di sforzo, a capire una buona parte di un testo scritto in francese anche se non conosce per niente questa lingua. L'italiano ed il francese hanno

una morfologia quasi eguale e pertanto molte parole che si formano in base alla flessione e alla coniugazione di temi si corrispondono reciprocamente in maniera priva di qualsiasi ambiguità. Anche la sintassi è quasi uguale e pertanto non occorre operare modifiche nell'ordine delle parole. Il lessico è composto da parole che molto spesso si corrispondono perfettamente come significato. Di conseguenza, nel caso di queste due lingue, anche la pura e semplice traduzione parola per parola dà risultati discreti.

Per quanto invece riguarda la traduzione in italiano del testo originale scritto in inglese, l'impressione generale è decisamente peggiore. Gli errori sono molti di più e non mancano parti che risultano incomprensibili o comprensibili solo con sforzo e grazie all'esperienza che tutti hanno delle cose che di solito vengono dette nelle informazioni per la visita di un museo.

È importante poi notare che il testo di partenza è veramente di livello elementare. È del tutto inusuale che chi scrive scriva così.

Proviamo allora a prendere un testo che sia più vicino al modo di scrivere più comune. Si tratta di informazioni relative a un notissimo motore di ricerca in Internet. È sempre un testo molto semplice, ma le frasi sono costituite anche da più di una proposizione, come avviene di regola nel parlare e scrivere comune. I testi originari sono ancora in francese e in inglese (la traduzione dall'inglese al francese è stata operata in maniera un po' libera, ma questo non ha influenza sul nostro discorso) e la traduzione in italiano, sia quella fatta dalla macchina che quella fatta dall'essere umano sono riportate sotto ciascun paragrafo seguendo le stesse convenzioni seguite nell'esempio precedente.

TESTO ORIGINALE SCRITTO IN FRANCESE

Introduction

Les performances et la précision des recherches Google reposent sur la qualité du matériel et des logiciels utilisés. La quasi-instantanéité des résultats est due en partie à l'efficacité de notre algorithme de recherche et en partie aux milliers (!) de PC que nous avons installés en réseau pour constituer un moteur de recherche ultrarapide.

L'élément fondamental de notre logiciel est PageRank, un système de classement des pages Web mis au point par les fondateurs de Google (Larry Page et Sergey Brin) à l'université de Stanford. Et pendant que plusieurs dizaines d'ingénieurs et de spécialistes consacrent leurs journées à améliorer les différents aspects de Google, PageRank reste la pierre angulaire de nos outils de recherche.

Introduzione

Le prestazioni e la precisione delle ricerche Google si basano sulla qualità **strumentazione** e software utilizzati. **Il quasi-instantanéité** dei risultati è dovuta in parte all'efficacia del nostro algoritmo di ricerca ed in parte alle migliaia (!) di PC che abbiamo installato in rete per costituire un motore di ricerca ultrarapido.

L'elemento fondamentale del nostro software è PageRank, un sistema di classificazione delle pagine web messo a punto dai fondatori di Google (Larry **pagina** e Sergey Brin) all'università di Stanford. E **durante che** molte decine **degli** ingegneri e di specialisti dedicano i loro giorni **da** migliorare i vari aspetti di Google, PageRank resta la pietra angolare dei nostri **attrezzi** di ricerca.

Introduzione

Le prestazioni e la precisione delle ricerche Google si basano sulla qualità **del hardware** e software utilizzati. **La quasi-instantaneità** dei risultati è dovuta in parte all'efficacia del nostro algoritmo di ricerca ed in parte alle migliaia (!) di PC che abbiamo installato in rete per costituire un motore di ricerca ultrarapido.

L'elemento fondamentale del nostro software è PageRank, un sistema di classificazione delle pagine web messo a punto dai fondatori di Google (Larry **Page** e Sergey Brin) all'università di Stanford. E **mentre** molte decine **di** ingegneri e di specialisti dedicano i loro giorni **a** migliorare i vari aspetti di Google, PageRank resta la pietra angolare dei nostri **strumenti** di ricerca.

PageRank

PageRank est un champion de la démocratie: il profite des innombrables liens du Web pour évaluer le contenu des pages Web -- et leur pertinence vis-à-vis des requêtes exprimées. Le principe de PageRank est simple : tout lien pointant de la page A à la page B est considéré comme un vote de la page A en faveur de la page B. Toutefois, Google ne limite pas son évaluation au nombre de "votes" (liens) reçus par la page; il procède également à une analyse de la page qui contient le lien. Les liens présents dans des pages jugées importantes par Google ont plus de "poids", et contribuent ainsi à "élire" d'autres pages.

Les sites qui se distinguent par leur qualité sont affectés d'une valeur PageRank plus élevée, et Google en tient compte lors de chaque recherche. Bien entendu, les pages jugées "importantes" par Google vont vous laisser indifférent si elles ne

répondent pas à vos requêtes... Aussi, pour retrouver les pages qui correspondent au mieux à votre requête, Google complète l'évaluation PageRank par des mécanismes évolués de correspondance de texte. Google ne se contente pas de compter le nombre d'occurrences d'un terme de recherche dans une page : il examine différents aspects du contenu de cette page (et du contenu des pages liées à celle-ci) afin de déterminer si elle correspond à votre requête.

PageRank

PageRank è un campione della democrazia: approfitta dei legami innumerevoli del web per valutare il contenuto delle pagine web e la loro pertinenza di fronte delle richieste espresse. Il principio di PageRank è semplice: ogni legame che indica della pagina A alla pagina B è considerato come un voto della pagina A in favore della pagina B. Tuttavia, Google non limita la sua valutazione al numero di "voti" (legami) ricevuti dalla pagina; procede anche ad un'analisi della pagina che contiene il legame. I legami presenti in pagine giudicate importanti da Google hanno più "peso", e contribuiscono così "ad eleggere" altre pagine.

Le unità che si distinguono con la loro qualità sono destinate di un valore PageRank più elevato, e Google ne tiene conto allora di ogni ricerca. Ben inteso, le pagine giudicate "importanti" da Google vi lasceranno indifferente se non rispondono alle vostre richieste... Inoltre, per trovare le pagine che corrispondono al massimo alla vostra richiesta, Google completa la valutazione PageRank con meccanismi evoluti di corrispondenza di testo. Google non si accontenta di contare il numero di verificarsi di un termine di ricerca in una pagina: esamina vari aspetti del contenuto di questa pagina (e del contenuto delle pagine legate a questa) allo scopo di determinare se corrisponde alla vostra richiesta.

PageRank

PageRank è un campione di democrazia: approfitta dei legami innumerevoli del web per valutare il contenuto delle pagine web e la loro pertinenza nei confronti delle richieste espresse. Il principio di PageRank è semplice: ogni legame che indica dalla pagina A alla pagina B è considerato come un voto della pagina A in favore della pagina B. Tuttavia, Google non limita la sua valutazione al numero di "voti" (legami) ricevuti dalla pagina; procede anche ad un'analisi della pagina che contiene il legame. I legami presenti in pagine giudicate importanti da Google hanno più "peso", e contribuiscono così "ad eleggere" altre pagine.

I siti che si distinguono per la loro qualità assumono un valore PageRank più elevato, e Google ne tiene conto al momento di ogni ricerca. Ben inteso, le pagine giudicate "importanti" da Google vi lasceranno indifferenti se non rispondono alle vostre richieste... Perciò, per trovare le pagine che corrispondono al meglio alla vostra richiesta, Google completa la valutazione PageRank con meccanismi evoluti di corrispondenza di testo. Google non si accontenta di contare il numero di volte che un termine di ricerca ricorre in una pagina: esamina vari aspetti del contenuto di questa pagina (e del contenuto delle pagine legate a questa) allo scopo di determinare se corrisponde alla vostra richiesta.

Intégrité

Les méthodes complexes et automatiques utilisées par les recherches Google rendent quasi impossible toute manipulation humaine des résultats. Comme nous l'indiquons clairement dans nos listes de résultat, certains sites peuvent être associés à une publicité "Sponsored Link". Toutefois, Google ne pratique pas la vente des positions dans ces résultats ; autrement dit, il n'est pas possible d'acheter une valeur PageRank supérieure à la réalité du Web. Avec la recherche Google, vous disposez d'une solution simple, rapide, honnête et objective pour trouver des sites Web de la plus haute qualité et dont les informations répondent parfaitement à vos besoins.

Integrità

I metodi complessi ed automatici utilizzati dalle ricerche Google rendono quasi impossibile ogni manipolazione umana dei risultati. Come lo indichiamo chiaramente nei nostri elenchi di risultato, alcune unità possono essere associate ad una pubblicità "Sponsored Link". Tuttavia, Google non pratica la vendita delle posizioni in questi risultati; in altre parole, non è possibile comperare un valore PageRank superiore alla realtà del web. Con la ricerca Google, disponete di una soluzione semplice, rapida, onesta ed oggettiva per trovare siti web della più alta qualità e le cui informazioni rispondono perfettamente alle vostre necessità.

Onestà

I metodi complessi ed automatici utilizzati dalle ricerche Google rendono quasi impossibile ogni manipolazione umana dei risultati. Come indichiamo chiaramente nei nostri elenchi di risultato, alcuni siti possono essere associati ad una pubblicità "Sponsored Link". Tuttavia, Google non pratica la vendita delle posizioni in questi risultati; in altre parole, non è possibile comperare un valore PageRank superiore alla realtà del web. Con la ricerca Google, disponete di una soluzione semplice, rapida, onesta ed oggettiva per trovare siti web della più alta qualità e le cui informazioni rispondono perfettamente alle vostre necessità.

TESTO ORIGINALE SCRITTO IN INGLESE

Introduction

Google runs on a unique combination of advanced hardware and software. The speed you experience can be attributed in part to the efficiency of our search algorithm and partly to the thousands of low cost PC's we've networked together to create a superfast search engine.

The heart of our software is PageRank™, a system for ranking web pages developed by our founders Larry Page and Sergey Brin at Stanford University. And while we have dozens of engineers working to improve every aspect of Google on a daily basis, PageRank continues to provide the basis for all of our web search tools.

Introduzione

Google funziona su una combinazione unica di hardware e di software avanzati. La velocità che sperimentate può essere attribuita in parte al risparmio di temi della nostra procedura di ricerca ed alle migliaia del we've del pc di basso costo networked per creare parzialmente insieme un motore di ricerca del superfast.

Il cuore del nostro software è PageRank™, un sistema per le pagine di Web di posto sviluppate dai nostri fondatori Pagina di Larry e Sergey Brin all'università di Stanford. E mentre abbiamo dozzine degli assistenti tecnici che lavorano per migliorare ogni funzione di Google su una base quotidiana, PageRank continua a fornire la base per tutti i nostri strumenti di ricerca di Web.

Introduzione

Google funziona in base a una combinazione unica di hardware e di software avanzati. La velocità che sperimentate può essere attribuita in parte alla efficienza del nostro algoritmo di ricerca e parzialmente alle migliaia di pc di basso costo che abbiamo connesso insieme in rete per creare un motore di ricerca superveloce.

Il cuore del nostro software è PageRank™, un sistema per classificare le pagine Web sviluppato dai nostri fondatori Larry Page e Sergey Brin all'università di Stanford. E mentre abbiamo dozzine di ingegneri che lavorano quotidianamente per migliorare ogni aspetto di Google, PageRank continua a fornire la base per tutti i nostri strumenti di ricerca Web.

PageRank Explained

PageRank relies on the uniquely democratic nature of the web by using its vast link structure as an indicator of an individual page's value. In essence, Google interprets a link from page A to page B as a vote, by page A, for page B. But, Google looks at more than the sheer volume of votes, or links a page receives; it also analyzes the page that casts the vote. Votes cast by pages that are themselves "important" weigh more heavily and help to make other pages "important."

Important, high-quality sites receive a higher PageRank, which Google remembers each time it conducts a search. Of course, important pages mean nothing to you if they don't match your query. So, Google combines PageRank with sophisticated text-matching techniques to find pages that are both important and relevant to your search. Google goes far beyond the number of times a term appears on a page and examines all aspects of the page's content (and the content of the pages linking to it) to determine if it's a good match for your query.

PageRank Ha spiegato

PageRank conta sulla natura unicamente democratica del Web usando la relativa struttura ampia di collegamento come indicatore di valore della pagina specifica. Nell'essenza, Google interpreta un collegamento dalla pagina A per paginare la B come .. voto, dalla pagina A, per la pagina B. Ma, Google guarda più del volume puro dei voti, o collega una pagina riceve; egualmente analizza la pagina che lancia il voto. I voti lanciano dalle pagine che sono essi stessi "importanti" pesano più pesante e contribuiscono a rendere altre pagine "importanti."

I luoghi importanti e di alta qualità ricevono un più alto PageRank, che Google si ricorda di ogni tempo che conduce una ricerca. Naturalmente, le pagine importanti non significano niente a voi se non abbinano la vostra domanda. Così, Google unisce PageRank con le tecniche testo-abbinanti specializzate alle pagine del ritrovamento che sono sia importanti che relative alla vostra ricerca. Google va lontano oltre il numero di volte che un termine compare ad una pagina ed esamina tutte le funzioni del soddisfare della pagina (e del contenuto delle pagine che si collegano ad esso) per determinare se esso [...] una buona corrispondenza per la vostra domanda.

PageRank Ha spiegato

PageRank conta sulla natura unicamente democratica del Web usando la sua ampia struttura di collegamento come indicatore di valore della pagina specifica. In essenza, Google interpreta un collegamento dalla pagina A alla pagina B come un voto, dalla pagina A, per la pagina B. Ma, Google guarda più del volume puro dei voti, o collegamenti che una pagina riceve; esso analizza anche la pagina che dà il voto. I voti dati dalle pagine che sono esse stesse "importanti" pesano di più e contribuiscono a rendere altre pagine "importanti."

I siti importanti e di alta qualità ricevono un più alto PageRank, che Google si ricorda di ogni volta che conduce una ricerca. Naturalmente, le pagine importanti non significano niente per voi se non rispondono alla vostra domanda. Così, Google

unisce PageRank con le sofisticate tecniche di abbinamento del testo per trovare pagine che sono sia importanti che pertinenti alla vostra ricerca. Google va molto oltre il numero di volte che un termine compare su una pagina ed esamina tutti gli aspetti del contenuto della pagina (e del contenuto delle pagine che si collegano ad essa) per determinare se esso è una buona risposta per la vostra domanda.

Integrity

Google's complex, automated methods make human tampering with our results extremely difficult. And though we do run relevant ads above and next to our results, Google does not sell placement within the results themselves (i.e., no one can buy a higher PageRank). A Google search is an easy, honest and objective way to find high-quality websites with information relevant to your search.

Integrità

Google complesso, metodi automatizzati rende l'alterazione umana i nostri risultati estremamente difficile. E benchè facciamo funzionare il ads relativo sopra e vicino ai nostri risultati, Google non vende la disposizione all'interno dei risultati essi stessi (cioè, nessuno possono comprare un più alto PageRank). Una ricerca di Google è un modo facile, onesto ed obiettivo [...] trovare i Web site di alta qualità con le informazioni relative alla vostra ricerca.

Onestà

I complessi metodi automatizzati di Google rendono l'alterazione umana dei nostri risultati estremamente difficile. E benchè facciamo funzionare un rilevante ads [sistema di annunci pubblicitari, N.d.T.] sopra e vicino ai nostri risultati, Google non vende il piazzamento all'interno dei risultati stessi (cioè, nessuno può comprare un più alto PageRank). Una ricerca di Google è un modo facile, onesto ed obiettivo per trovare siti Web di alta qualità con le informazioni rilevanti per vostra ricerca.

La struttura più complessa delle frasi presenti in questo testo non ha sostanzialmente modificato la qualità della traduzione automatica nel caso del passaggio dal francese all'italiano, com'era da aspettarsi in ragione delle forti analogie nella morfologia e nella sintassi tra queste due lingue. Le sensibili differenze sintattiche e soprattutto le grosse differenze morfologiche fra l'inglese e l'italiano portano invece, nel caso di questo testo più complesso, ad un sensibile peggioramento della qualità della traduzione automatica. Non riporto traduzioni di altri passi perché per testi di media complessità, cioè più complessi di questo (che è pur sempre un testo semplice), ovviamente c'è da aspettarsi una qualità della traduzione automatica più scadente o, al massimo, simile (questo naturalmente vale anche, e a maggior ragione, per testi complessi o molto complessi).

Esaminiamo ora attentamente il tipo di errori presenti nel testo tradotto automaticamente. Essi sono sostanzialmente di tre tipi.

1) Il primo tipo di errori sono quelli causati dal fatto che il programma alcune volte non “capisce” a quale altra parola (o gruppo di parole) una certa parola (o gruppo di parole) si riferisca e/o a quale categoria grammaticale appartenga. Questo non determina molti errori nel caso del passaggio dal francese all'italiano, per le ragioni appena viste. Ma nel passaggio da una lingua con scarsa morfologia come l'inglese ad una con una morfologia piuttosto ricca come l'italiano (e ancor più nel caso di quelle lingue che hanno una morfologia ancora più ricca perché possiedono i casi) gli errori sono molti. Nella traduzione automatica dei due brani sopra riportati vi sono esempi di verbi tradotti in maniera completamente errata nel modo, nel tempo e nella persona, di nomi scambiati per verbi e viceversa, di grossolani errori nell'ordine delle parole ecc.. Si noti che gli errori di questo genere, oltre ad essere numerosi, sono anche i più gravi. Infatti, per il lettore che non conosce per niente la lingua del testo originario, non raramente è impossibile riuscire a capire che cosa significano quelle parti della traduzione automatica che risultano palesemente insensate. Questo anche se egli dispone di un dizionario bilingue, perché esso non riporta le forme flesse. Nel testo della traduzione automatica riportata come esempio gli errori che sono *strettamente* di questo tipo sono stati evidenziati mediante sottolineatura (oltre che con l'evidenziatore giallo, comune a tutti gli errori gravi).

2) Un altro tipo di errori è quello dovuto al fatto che certe parole hanno più di un significato (come per esempio la parola italiana “cane” che può indicare sia un animale sia una parte di certe armi da fuoco) o diverse sfumature di significato. Questo genere di errori, anch'esso frequente, è però meno grave rispetto al tipo precedente, perché essi possono essere corretti da chi legge la traduzione, basandosi sul

senso generale del testo e ricorrendo ad un dizionario bilingue.

3) Il terzo tipo di errori sono le improprietà, le inesattezze, le espressioni che suonano male in una certa lingua ecc.. Questi errori tutto sommato non compromettono la comprensione del testo, perciò, pur essendo abbastanza frequenti, sono chiaramente i meno gravi.

La seguente tabella mostra la frequenza di questi tre tipi di errori nel testo della traduzione automatica riportata come esempio.

	<i>errori tipo 1 (di riferimento)</i>	<i>errori tipo 2 (di significato)</i>	<i>errori tipo 3 (improprietà ecc.)</i>
<i>brano "museo" in francese</i> (22 frasi, 403 parole)	13 0,59 — 3,22	18 0,81 — 4,46	4 0,18 — 0,99
<i>brano "museo" in inglese</i> (22 frasi, 379 parole)	31 1,40 — 8,17	22 1 — 5,80	16 0,72 — 4,22
<i>brano "Google" in francese</i> (16 frasi, 449 parole)	4 0,25 — 0,89	18 1,16 — 4,00	3 0,18 — 0,66
<i>brano "Google" in inglese</i> (15 frasi, 346 parole)	31 2,06 — 8,95	16 1,06 — 4,62	21 1,4 — 6,06

La cifra in alto nella casella indica il numero assoluto degli errori, mentre le due cifre in basso (evidenziate in carattere grassetto) indicano, quella a sinistra il numero medio degli errori per frase (in genere, è stata considerata come un errore ogni singola parola tradotta in maniera errata) e quella a destra la percentuale delle parole tradotte in maniera errata. Nel caso della traduzione dall'inglese all'italiano, la media per frase degli errori del tipo 1 e 2 (quelli che più ci interessano) anche nel caso del primo brano, che è estremamente semplice e le cui frasi sono costituite da un'unica proposizione, è pari o superiore all'unità per ciascuno di questi due tipi di errori. Come si vede, si tratta di frequenze abbastanza elevate, se si pensa che anche solo due o tre errori di questi due tipi in una frase possono rendere incomprensibile o difficilmente comprensibile il senso generale di tutta la frase.

Il ritrovato di Ceccato (con alcune mie modifiche e aggiunte)

Vediamo dunque qual è il ritrovato suggerito da Ceccato per risolvere, del tutto o perlomeno in buona parte, il problema degli errori nella traduzione automatica derivati dal fatto che la macchina non ha la capacità di ricostruire la rete correlazionale di pensiero partendo dal testo perché manca della capacità di comprendere il significato delle parole né possiede quella cultura generale che in molti casi permette all'uomo di capire in quale modo le parole sono correlate fra loro.

Il primo e fondamentale presupposto teorico che sta alla base di questo ritrovato è, naturalmente, la teoria correlazionale del pensiero, sommariamente esposta nel mio articolo "Una presentazione della Metodologia Operativa" e, un po' più dettagliatamente, in questo articolo. Il secondo presupposto è che se noi ricostruiamo in qualche modo la rete correlazionale di pensiero di cui una certa frase è l'espressione e basiamo la traduzione automatica su questo risultato otteniamo una qualità di traduzione nettamente migliore. Questo è facilmente comprensibile in base a quello che abbiamo finora detto. Tornando al nostro esempio delle due espressioni

- "una quantità di acqua, anche piccola"
- "una quantità di acqua, anche fredda"

se noi ricostruiamo prima la corrispondente rete correlazionale di pensiero sarà poi facile far sì che il programma traduca in maniera corretta i due aggettivi "piccola" e "fredda" in una lingua che flette l'aggettivo secondo il caso del nome a cui esso è riferito. Abbiamo visto anche altri esempi in cui la

mancanza di una ricostruzione della corretta rete correlazionale di pensiero porta a grossolani errori di traduzione e abbiamo anche visto che questi casi sono tutt'altro che rari.

Ricostruire la rete correlazionale di pensiero corrispondente al discorso che rappresenta l'espressione di questo pensiero prima di operare la traduzione permetterebbe dunque di migliorare grandemente la qualità di essa.

Ceccato partiva dal presupposto che questa ricostruzione dovesse essere operata dalla macchina. Questo perché le sue ricerche erano finanziate dalle Forze Armate statunitensi che negli anni '50 e '60, cioè ai tempi della cosiddetta "guerra fredda" fra gli USA e l'Unione Sovietica, erano interessate ad una soluzione più economica della traduzione operata dall'uomo per l'enorme quantità di materiale espresso in lingua russa che veniva allora quotidianamente tradotto. In questo caso chi produceva questo materiale linguistico non aveva alcuna intenzione di facilitarne la traduzione, ma non sempre le cose stanno in questo modo, anzi direi che oggi, stante la forte tendenza che c'è alla diffusione internazionale delle informazioni, sarebbe gradito agli autori di una buona parte dei testi che vengono prodotti che tali testi potessero essere facilmente tradotti in altre lingue.

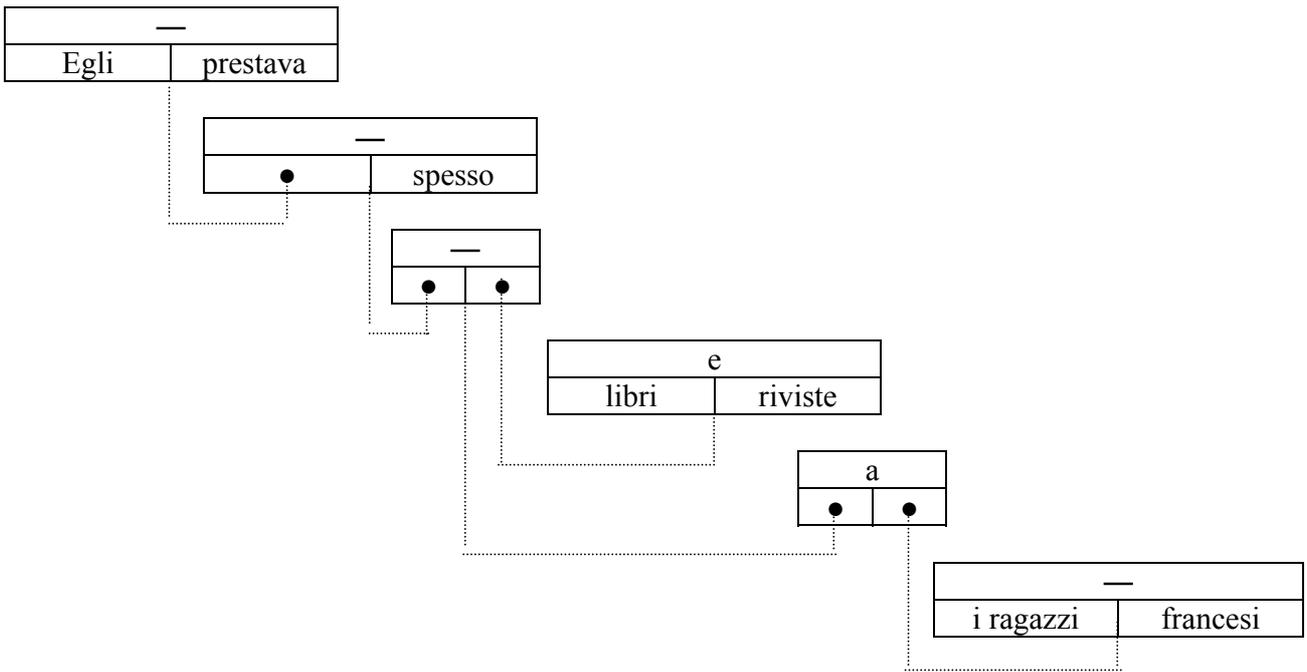
Anche se disporre di un programma che traduca in maniera soddisfacente partendo da un testo in nessun modo preparato è più desiderabile, sarebbe sicuramente molto più facile produrre un programma che traduce partendo da un testo in cui la ricostruzione della rete correlazionale di pensiero è stata fatta dall'uomo. Già in questo modo viene completamente eliminata la figura del traduttore umano, cioè dell'essere umano che conosce almeno due lingue. La preparazione del testo, scritto in una certa lingua, viene effettuata da una persona a cui non è richiesto altro che conoscere bene quella lingua (e naturalmente la teoria correlazionale del pensiero). Quindi questo lavoro può essere fatto da chiunque per qualsiasi testo scritto nella sua madrelingua.

Anche se in questo articolo parlerò prevalentemente del ritrovato di Ceccato (concepito per realizzare un programma di traduzione completamente automatica), vorrei prima accennare ad una mia proposta per realizzare un programma di traduzione semi-automatica.

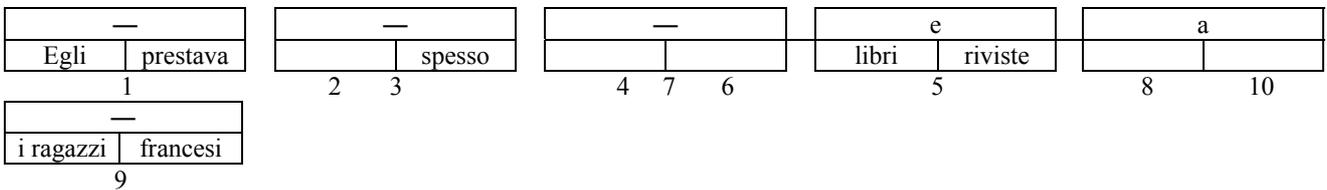
La mia proposta è quella di usare come base di partenza per la traduzione automatica una specie di "formattazione" del testo che indichi chiaramente quale è la corrispondente rete correlazionale di pensiero. Questa formattazione del testo deve essere naturalmente molto semplice da scrivere e da leggere. Il modo di rappresentare le reti correlazionali proposto da Ceccato non si presta bene a questo scopo, prima di tutto perché esso occupa molto spazio nella pagina, e poi perché la rappresentazione grafica della rete potrebbe essere più semplice. Prendiamo come esempio una frase piuttosto semplice, cioè costituita da una sola proposizione, che però contenga tutte le parti fondamentali del discorso (articolo, nome, pronome, aggettivo, verbo, avverbio, congiunzione e preposizione) e gli elementi sintattici fondamentali (soggetto, predicato, complemento diretto e complemento indiretto). Molte delle frasi che noi produciamo sono più o meno di questo tipo. Un esempio del genere può essere:

Egli prestava spesso libri e riviste ai ragazzi francesi.

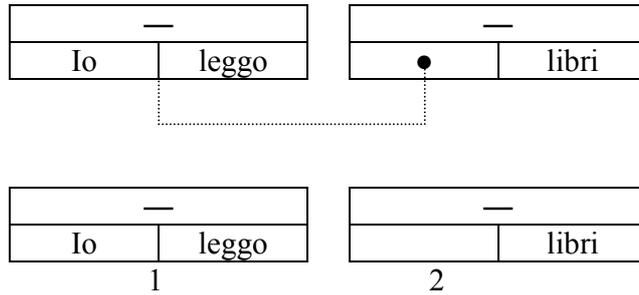
Secondo il modo di rappresentare le reti correlazionali proposto da Ceccato quella corrispondente a questa frase è questa (per semplicità la correlazione articolo-nome, il cui correlatore è il correlatore implicito, è stata indicata ponendo i due termini nella stessa casella anziché usare una triade correlazionale specifica):



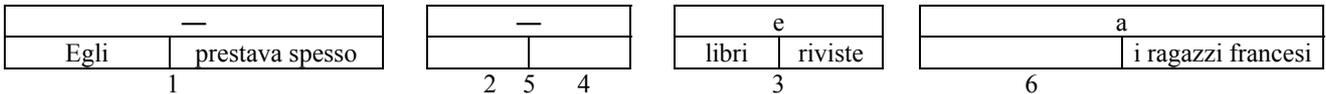
Per risolvere il problema del molto spazio occupato e della rappresentazione grafica della rete non molto semplice è sufficiente disporre le correlazioni in maniera lineare e rappresentare con dei semplici numeri le loro relazioni, in questo modo:



Il numero posto sotto la separazione fra le due caselle inferiori indica che l'intera correlazione è il correlato di un'altra correlazione; il numero posto sotto una delle caselle dei correlati, in posizione all'incirca centrata, indica quale è questa correlazione e se la prima correlazione è il primo o il secondo correlato della seconda. Cioè i due numeri, l'uno immediatamente successivo all'altro nella serie numerica, sostituiscono, il primo l'inizio della linea tratteggiata e il secondo la sua fine (rappresentata dal grosso punto nero); il percorso obbligato per andare dall'uno all'altro rispettando l'ordine che hanno i numeri sostituisce la linea tratteggiata, in questo modo:

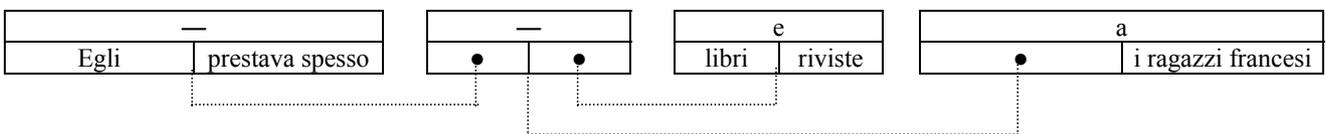


Se si vuole semplificare ulteriormente, si può applicare la convenzione applicata alla correlazione articolo-nome, quella di non rappresentare separatamente i due termini, anche ad altre correlazioni il cui correlatore è il correlatore implicito, come la correlazione verbo-avverbio, quella sostantivo-aggettivo ecc., ponendo semplicemente, anche in questo caso, i due termini nella stessa casella. In questo caso la rete correlazionale corrispondente alla fase del nostro esempio verrebbe rappresentata così:

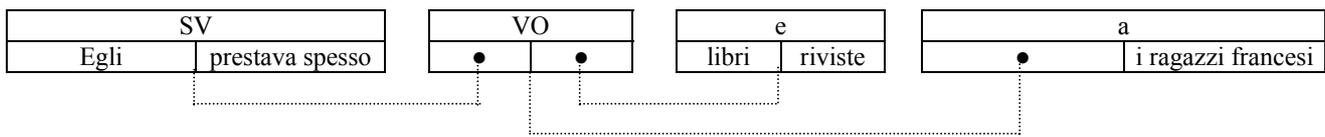


È molto facile costruire un programma che permetta all'essere umano di scrivere ed eventualmente modificare rapidamente le reti correlazionali rappresentate nel modo da me proposto (basta che la pagina sia già formattata in righe fatte da triadi di caselle con sotto le righe per i numeri; che ci sia un sistema per mettere rapidamente al posto giusto i numeri ecc.). Con un simile programma digitare un testo in forma di reti correlazionali di pensiero non è molto più lungo che digitarlo nel modo usuale. Anche l'individuazione della rete correlazionale giusta a partire dalla frase è un compito che l'essere umano impara, con l'allenamento, ad eseguire rapidamente.

Indicare le relazioni che legano le triadi correlazionali mediante dei numeri anziché mediante delle linee tratteggiate è del tutto privo di ambiguità e quindi va benissimo per un programma di calcolatore. L'essere umano invece si rende conto più rapidamente e facilmente di queste relazioni se esse sono indicate mediante delle linee tratteggiate. È molto facile costruire un programma che permetta di passare automaticamente da un modo all'altro di indicare. Le linee tratteggiate possono essere usate anche se le triadi correlazionali vengono disposte in maniera lineare, mettendole parallelamente sotto di esse, in questo modo:



Infine, è opportuno, per ragioni che vedremo più avanti, indicare alcuni tipi di relazioni sintattiche fra correlati il cui correlatore è il correlatore implicito, come il rapporto soggetto-verbo (SV) e verbo-oggetto (VO). Esse possono essere indicate in questo modo:



Disporre di un programma che effettui la traduzione automatica partendo da un testo in cui la rete correlazionale sia stata individuata dall'essere umano è cosa non priva di utilità già di per sé. Ma l'interesse di un simile tipo di programma è anche un altro. Infatti, sia che l'individuazione della rete correlazionale del pensiero nella lingua di partenza sia affidata ad un essere umano sia che si tenti di creare un programma che sia capace di questo, una volta giunti a questo risultato la tappa successiva è in entrambi i casi la stessa, cioè quella di creare un programma che partendo dalla rete correlazionale del pensiero la esprima con il lessico e le regole grammaticali della lingua di uscita. In altre parole, , cioè con aiuto umano, che parte dal testo "formattato" in modo da indicare la rete correlazionale del pensiero, coincide con la seconda delle due fasi di cui deve essere composto un programma di traduzione totalmente automatica, e pertanto questo programma di traduzione semi-automatica va in ogni caso creato. La creazione di un simile programma non dovrebbe porre problemi. In una qualsiasi lingua per passare da un certo pensiero al discorso che lo esprime dobbiamo rispettare delle regole rigide e ben note, e pertanto la realizzazione di un programma che faccia questo non dovrebbe essere difficile. Ad esempio, la correlazione sostantivo-aggettivo, come questa

—	
penna	rossa

viene generalmente espressa in inglese antepoendo l'aggettivo al sostantivo, mentre invece in italiano è l'aggettivo che segue il sostantivo, tranne casi particolari. Inserire queste regole nel programma non è certo difficile e pertanto, poiché la rete correlazionale di pensiero di partenza è nota, la traduzione sarà sempre corretta. Grossolani errori di traduzione si possono verificare, come si è visto, quando non è nota la rete correlazionale del pensiero.

A parte queste considerazioni sulla necessità e utilità di un programma di traduzione semiautomatica che parta dalla rete correlazionale del pensiero, nondimeno un programma in grado di tradurre con risultati soddisfacenti senza bisogno di alcuna preliminare preparazione del testo da parte dell'essere umano sarebbe molto più desiderabile. Vediamo dunque qual era il ritrovato proposto da Ceccato per realizzare un programma simile.

Come in altri programmi, come "unità di ingresso" vengono assunte le singole parole e, nel caso delle lingue flettenti, come dizionario di cui è dotato il programma viene usato un dizionario in cui siano presenti tutte le forme flesse possibili di ogni singola voce del vocabolario della lingua di ingresso. L'originalità della proposta di Ceccato risiede, come abbiamo detto, nel fatto che il programma cerca di ricostruire la rete correlazionale del pensiero di ogni singola frase. Questo viene fatto essenzialmente nei due seguenti modi.

1) Tutte le parole della lingua di ingresso vengono distinte in due categorie. La prima categoria è quella dei correlatori, la seconda quella dei correlati. La prima categoria è pertanto costituita dalle congiunzioni, dalle preposizioni e, nelle lingue che li possiedono, anche dai casi, mentre la seconda è costituita, in pratica, da tutte le altre parole. Come abbiamo visto, nella triade correlazionale i correlati occupano sempre le caselle inferiori. I correlatori invece occupano sempre la casella superiore, tranne rari casi che comunque sono in genere marcati in maniera particolare. Un esempio di questi rari casi è la frase

- “e” e “o” sono congiunzioni

nella quale la congiunzione “e” e la congiunzione “o” poste fra virgolette non sono dei correlatori ma dei correlati legati dall’altra congiunzione “e”. Casi come questi, a parte il fatto che sono molto rari, sono comunque, come si vede, chiaramente marcati, perché le due congiunzioni, quando svolgono il ruolo, per loro anomalo, di correlati, nello scritto sono poste fra virgolette e nel parlato sono isolate da delle pause particolari.

Vi sono anche delle parole che indicano contemporaneamente un correlato ed un correlatore. Esempi di questo genere sono tutti i verbi in forma personale nelle lingue che, come l’italiano e molte altre lingue, flettono il verbo secondo la persona a cui è riferito. Il verbo in forma personale, infatti, è una parola che indica contemporaneamente tre cose:

- un correlato
- un particolare correlatore, il correlatore implicito
- la posizione del correlato, cioè quella di secondo correlato.

Cioè, per esempio, la forma personale del verbo “ridere” presente nella frase “Giovanni ride” indica che la correlazione di pensiero corrispondente è inequivocabilmente la seguente:

—	
Giovanni	ride

o, scritta in forma più rigorosa, questa:

-e	
Giovanni	rid-

perché il secondo correlato non è propriamente “ride”, ma il significato del verbo “ridere” e cioè il tema “rid-” (con vocale tematica “-e”), mentre la desinenza “-e” indica, fra l’altro, che il significato del verbo “ridere” deve essere correlato mediante il correlatore implicito a una terza persona singolare, in questo caso Giovanni, che è il primo correlato, mentre il verbo “ridere” deve essere il secondo correlato. Pertanto, se volessimo scrivere questa correlazione in maniera del tutto rigorosa, dovremmo scriverla così:

-e	
(= correlatore implicito, che correla il secondo correlato con una terza persona singolare)	
Giovanni	rid- + -e (= presente indicativo)

Questo perché la lingua italiana, come molte altre lingue, è tendenzialmente “fusiva”, cioè tende appunto a fondere in una sola desinenza più significati. Quella semplice desinenza “-e” indica infatti che il significato del verbo “ridere” è coniugato al presente indicativo ed è correlato mediante il correlatore implicito con una terza persona singolare.

Un altro importantissimo esempio di parole che indicano contemporaneamente un correlato ed un correlatore sono, nelle lingue che posseggono i casi, tutte le parole che sono flesse secondo un caso, cioè tutti i sostantivi ed eventualmente anche tutti gli aggettivi. Nel caso del sostantivo flesso secondo un particolare caso, per esempio il sostantivo latino flesso al caso genitivo *urbis* (“della città”), il tema (*urb-*) indica il correlato, cioè “città”, mentre la desinenza (*-is*) indica il correlatore, cioè la categoria mentale di genitivo, che invece in italiano viene espressa mediante una preposizione, la preposizione “di” (nelle lingue che non possiedono i casi le corrispondenti categorie mentali vengono espresse mediante preposizioni). Quando una lingua, come il latino, flette anche l’aggettivo assegnandogli lo stesso caso del sostantivo a cui esso si riferisce, la forma flessa dell’aggettivo non indica il correlatore “caso”

(genitivo, dativo ecc..) ma il correlatore implicito ed il fatto che esso lega l'aggettivo, che funge da secondo correlato, con un sostantivo che è flesso secondo lo stesso caso.

Naturalmente i casi in cui una parola indica contemporaneamente un correlato e un correlatore, e magari anche ci dà delle informazioni su come deve essere l'altro correlato, rappresentano una notevole facilitazione alla ricostruzione della rete correlazionale di pensiero, sia per l'uomo che ascolta o legge sia per il tipo di programma per computer che stiamo descrivendo.

Abbiamo detto che il primo presupposto su cui è basato questo ritrovato per la traduzione automatica è questa distinzione di tutte le parole di una lingua in due categorie, quella dei correlatori e quella dei correlati. Il programma sarà quindi dotato di due liste di parole, quella delle parole che indicano quasi sempre (tranne i rari casi visti sopra) dei correlatori (cioè le parole che la grammatica classifica come preposizioni e congiunzioni), e quella delle parole che indicano sempre dei correlati. Nel caso delle lingue che tendono a fondere in un'unica parola un correlato e un correlatore (e magari anche altre informazioni relative all'altro correlato) la lista dei correlati comprenderà tutte le possibili forme flesse per ciascuna voce dei comuni dizionari. Accanto a ciascuna forma flessa figurerà il tema (o la forma flessa al nominativo) e ciò che la flessione (in genere una desinenza) esprime, specificato secondo la terminologia propria della grammatica (per esempio: "genitivo", "dativo", "correlazione implicita con terza persona singolare" ecc..).

Il programma collocherà sempre (tranne che nei rari casi di cui abbiamo detto sopra) le parole che fanno parte della lista dei correlatori nelle caselle superiori delle triadi correlazionali, mentre collocherà sempre nelle caselle inferiori le parole che fanno parte della lista dei correlati. Questo è il primo accorgimento che permette al programma di ricostruire la rete correlazionale del pensiero a partire dallo scritto di cui esso è espressione.

2) Il secondo accorgimento consiste nel dotare il programma di una cosiddetta "sfera nozionale" fatta in modo tale da permettere ad esso di completare l'operazione di ricostruzione della rete correlazionale di pensiero svolgendo la stessa funzione che nell'uomo che ascolta o legge ha quella cultura di base da tutti posseduta, la quale molto spesso ci permette, come abbiamo visto, di ricostruire la struttura di certe correlazioni e dei rapporti fra esse anche se nel discorso o nello scritto le relative informazioni non sono presenti. L'essere umano può far questo perché comprende il significato delle parole di una lingua che rappresentano dei correlati e ha appunto una cultura di base che gli permette di stabilire se fra due cose designate da queste parole vi può essere un rapporto oppure no. Riprodurre questa funzione in una macchina non è per ora in alcun modo possibile, ma è possibile avvicinarsi al risultato della traduzione operata dall'essere umano sostituendo la comprensione dei significati delle parole e le nozioni circa i possibili rapporti delle cose da essere designate con delle complesse liste di classificazioni che costituiscono appunto la sfera nozionale. Le liste sono essenzialmente le tre seguenti.

a) Una prima classificazione consiste nel dividere tutte le parole del vocabolario di cui è dotato il programma in classi più o meno generali. Il criterio generale per stabilire quali devono essere queste categorie è che all'interno di ogni singola classe tutte le parole che la compongono hanno delle caratteristiche che influenzano la loro possibilità o meno di essere correlate con altre parole. Per esempio tutti i componenti della classe "cibi solidi" non possono essere l'oggetto di un verbo come "bere", mentre possono benissimo essere l'oggetto di un verbo come "mangiare", ed è assai probabile che ciò avvenga se, in una frase, una di queste parole è vicina al verbo "mangiare"; tutti i componenti della classe "vegetali" non possono essere il soggetto di verbi come "muoversi", "camminare", "volare" ecc., mentre possono esserlo benissimo di verbi come "fiorire", "seccare", "crescere" ecc.. Tuttavia non è possibile stabilire esattamente a priori quali devono essere esattamente queste classi. Ceccato e i suoi collaboratori operarono queste divisioni in classi basandosi sull'esperienza maturata formando molte frasi (circa 1000) fra quelle che possono essere formate a partire da un certo dizionario (essi usavano un dizionario contenente i termini essenziali della lingua inglese formato da circa 500 parole). Essi prima introdussero delle classi molto generali (per esempio: "esseri viventi", "esseri non viventi", "esseri umani", "animali", "vegetali", "minerali", "solidi", "liquidi" ecc. ecc.), poi, andando ad esaminare ognuna delle

sudette circa 1000 frasi, verificavano se la lista delle classi compilata era sufficiente ad escludere le reti correlazionali non corrette. Se così non era, introducevano nuove classi più specifiche. Ad esempio, se la classe delle parole designanti “animali” era insufficiente per escludere che un suo elemento, per esempio la parola “cane”, presente in una delle suddette 1000 frasi, facesse parte di una correlazione possibile dal punto di vista grammaticale ma errata (per esempio quella di “cane” come soggetto del verbo “volare”) veniva introdotta una nuova classe o nuove classi che permettessero di escludere questa correlazione errata (ad esempio: “animali d’aria”, “animali d’acqua” e “animali di terra”). La parola “cane” del nostro esempio veniva dunque nuovamente classificata anche come “animale di terra” in modo da escludere che essa possa essere soggetto di un verbo come “volare”.

Riproduco qui di seguito la lista delle classificazioni usata da Ceccato e i suoi collaboratori. Come ben si vede, essa comprende delle categorie molto generali ed altre un po’ meno generali. Tutti i componenti di ciascuna di queste categorie presumibilmente possono o non possono essere dei correlati di determinate altre parole. Vi sono anche delle classificazioni ancora più specifiche (per esempio la classe “strumenti di giardinaggio”, che come ho accennato sono state introdotte per dirimere dei casi di ambiguità nel ricostruire la rete correlazionale che si venivano effettivamente a creare in qualcuna delle circa 1000 frasi suddette.

Tabella 1

Lista delle classificazioni¹²

1) esseri viventi	26) animali che strisciano
2) esseri non viventi	27) erbivori
3) esseri animati	28) carnivori
4) esseri inanimati	29) animali pericolosi
5) esseri umani	30) animali miti
6) animali	31) animali selvatici
7) vegetali	32) animali domestici
8) minerali	33) animali scavatori
9) parti di 003	34) animali predatori
10) parti di 005	35) animali da macello
11) parti di 006	36) animali da caccia
12) parti di 007	37) alberi da frutta
13) collettivi di 002, 004	38) cose opache
14) collettivi di 005	39) cose trasparenti
15) collettivi di 006	40) liquidi
16) collettivi di 007	41) aeriformi
17) 005 + professione (o attività svolta)	42) solidi
18) 005 + appartenenza geografica	43) fluidi
19) 005 + appartenenza politica	44) polveri
20) 005 + relazioni familiari	45) fluidi e polveri
21) 005 + relazioni sociali	46) aeriformi trasparenti
22) comunità politiche	47) aeriformi opachi
23) animali d’acqua	48) solidi trasparenti
24) animali d’aria	49) corpi celesti
25) animali di terra	50) agenti atmosferici

¹² Un elemento dell’elenco originale è stato eliminato perché lo ritengo troppo simile ad un altro.

- | | |
|---|--|
| 51) fenomeni atmosferici | 94) misure cubiche |
| 52) condizioni meteorologiche | 95) pesi |
| 53) punti cardinali | 96) misure di tempo |
| 54) estensioni geografiche | 97) nomi dei giorni della settimana |
| 55) estensioni geografiche di terra | 98) nomi dei mesi |
| 56) estensioni geografiche d'acqua | 99) indicazioni di tempo provenienti dall'astronomia |
| 57) estensioni geografiche caratterizzate dalla forma | 100) oggetti economici |
| 58) estensioni geografiche caratterizzate dal suolo | 101) oggetti semantici |
| 59) estensioni geografiche caratterizzate dalla flora | 102) eventi |
| 60) cibi | 103) posti pubblici |
| 61) cibi solidi | 104) posti aperti al pubblico |
| 62) cibi liquidi e bevande | 105) posti dove si svolge attività economica |
| 63) cibi in forma di polveri | 106) servizi pubblici |
| 64) frutta | 107) cose coperte |
| 65) verdura | 108) coperture |
| 66) oggetti naturali | 109) cose apribili |
| 67) artefatti | 110) cose aperte o chiuse per sottrazione o aggiunta |
| 68) posti abitati | 111) cose aperte o chiuse per la loro posizione nei confronti di altro |
| 69) edifici e costruzioni | 112) cose aperte o chiuse per la posizione delle loro parti |
| 70) parti di 069 | 113) cose che aprono o chiudono ruotando o slittando |
| 71) parti interne di 069 | 114) strumenti per aprire o chiudere |
| 72) parti esterne di 069 | 115) cose che possono essere appese |
| 73) oggetti di decorazione interna | 116) strumenti per appendere |
| 74) mobili | 117) prodotti d'arte |
| 75) tessuti | 118) cose che possa essere tenute in mano |
| 76) abbigliamento | 119) cose fisse (che possono essere tenute in mano) |
| 77) effetti personali | 120) cose mobili (che possono essere tenute in mano) |
| 78) parti di 074 | 121) cose trasportabili |
| 79) parti di 076 | 122) cose trasportabili per scorrimento |
| 80) strumenti | 123) cose trasportabili spingendole |
| 81) mezzi di trasporto | 124) cose trasportabili tirandole |
| 82) mezzi di trasporto acquatici | 125) cose appuntite |
| 83) mezzi di trasporto aerei | 126) cose taglienti |
| 84) mezzi di trasporto terrestri | 127) segni e cose usate per indicare |
| 85) utensili domestici | 128) numeri |
| 86) contenitori | 129) cose che si presentano in coppia |
| 87) strumenti musicali | 130) cose che si presentano in file |
| 88) strumenti di giardinaggio | 131) |
| 89) giocattoli | |
| 90) strumenti di misura | |
| 91) misure | |
| 92) misure lineari | |
| 93) misure quadrate | |

b) La seconda fondamentale classificazione da operare è quella che riguarda i rapporti fra le cose. Occorre individuare quali sono i più comuni rapporti che possono intercorrere fra due cose. La lista di questi rapporti compilata da Ceccato e dai suoi collaboratori è la seguente.

Tabella 2

Sfera nozionale: lista dei rapporti¹³

1) elemento		collezione
2) membro		classe
3) specie		genere
4) parte		tutto
5) componente		composto
6) caratteristica		cosa caratterizzata
7) prodotto		produttore
8) prodotto		luogo di produzione
9) cosa contenuta		contenitore
10) cosa tirata		cosa che tira
11) cosa diretta o guidata		cosa che dirige o guida
12) cosa che proviene		provenienza
13) precedente		sussequente
14) cosa coperta o chiusa		cosa che copre o chiude
15) cosa decorata		decorazione
16) cosa spinta		cosa che spinge
17) cosa principale		accessorio
18) mezzo di protezione		cosa da cui ci si protegge
19) materiale		forma
20) materiale		suo stato fisico
21) stadio di sviluppo precedente		stadio di sviluppo sussequente
22) relazione di parentela ascendente		relazione di parentela discendente
23)	relazione di parentela collaterale	
24)	associazione storica	
25)	contiguità	
26)	opposizione	
27)	relazione economica	
28)	relazione semantica	
29) soggetto		attività
30) soggetto		oggetto dell'attività
31) soggetto		risultato dell'attività
32) soggetto		materiale dell'attività
33) soggetto		strumento dell'attività
34) soggetto		luogo abituale dell'attività
35) soggetto		tempo abituale dell'attività
36) soggetto		oggetto
37) attività		risultato

¹³ Due elementi dell'elenco originale sono stati eliminati perché li ritengo troppo simili ad altri.

38) attività		materiale
39) attività		strumento
40) attività		luogo abituale
41) attività		tempo abituale
42) risultato dell'attività		materiale dell'attività
43) risultato dell'attività		strumento dell'attività
44) oggetto dell'attività		risultato dell'attività
45) oggetto dell'attività		materiale dell'attività
46) oggetto dell'attività		strumento dell'attività
47) oggetto		luogo abituale
48) oggetto		tempo abituale
49) materiale dell'attività		strumento dell'attività
50)	strumenti complementari	
51)	oggetti complementari	
52) cosa		materiale
53) funzione		organo
54) organo		macchina
55) soggetto		agente dell'attività
56)

A questo punto occorre costruire concretamente la sfera nozionale relativa al vocabolario di cui è dotato il programma. Un frammento di sfera nozionale relativa ad un vocabolario molto piccolo (quello costituito dalle parole di un racconto per bambini, *The little train* ["Il trenino"] di L Lenski) è mostrata nella figura 1.

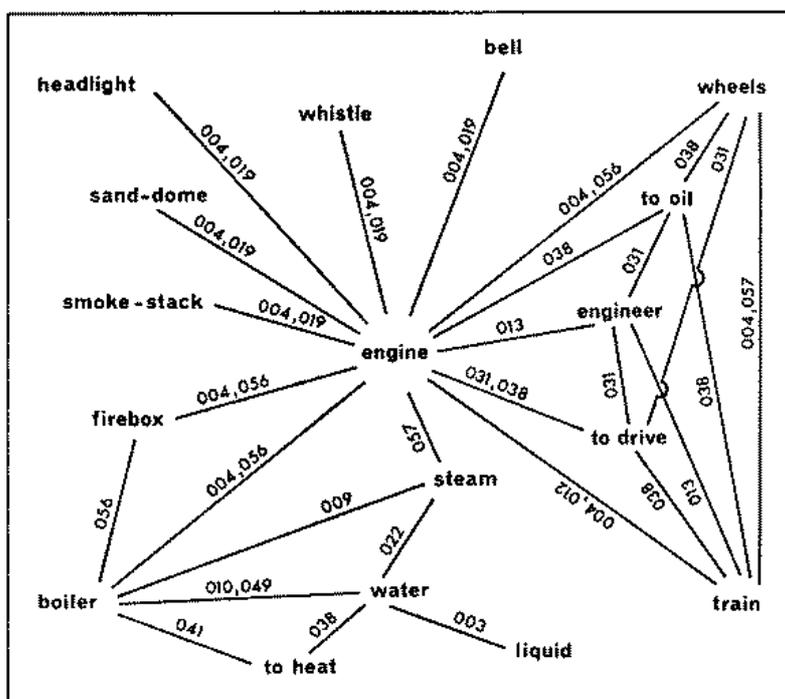


Figura 1

Le linee che uniscono le parole indicano che fra le cose designate da queste parole è stato individuato almeno un rapporto di quelli elencati nella tabella 2; i numeri disposti lungo le linee sono i numeri in codice del particolare tipo di rapporto. Ad esempio, la linea che unisce le due parole *boiler* e *engine* significa che fra le cose designate da queste due parole c'è il rapporto di "parte-tutto" (004) e "organo-macchina" (056).

Non tutti i possibili tipi di rapporto vengono indicati esplicitamente, ma alcune relazioni vengono ricavate dal programma, secondo regole di derivazione, a partire da alcune relazioni di base. Per esempio, nel caso del rapporto genere-specie, una regola permette di trasferire i rapporti della cosa classificata come genere alla cosa classificata come specie (tranne, naturalmente, il rapporto specie-genere con il verso opposto). Per esempio, se *water* è classificata come specie di *liquid*, e *liquid* come oggetto di *to pour* (versare) quest'ultimo rapporto sarà trasferito anche a *water*, che diventerà così automaticamente oggetto dell'attività di versare. Non sarà invece trasferita a *water* la relazione di genere-specie che *liquid* ha per esempio con *whisky*.

Un'altra regola permette di classificare come "specie dello stesso genere" parole come *water* e *whisky* attraverso il rapporto che esse hanno con *liquid*.

Un'altra regola stabilisce che certi rapporti, come quello di specie-genere, parte-tutto, cosa-materiale, ecc. hanno la proprietà di trasferirsi automaticamente nell'ambito dello stesso rapporto e nella stessa direzione. Se, per esempio, "pistone" è parte di "motore", che a sua volta è parte di "automobile", anche "pistone" è parte di "automobile".

Altre regole sono così descritte da Ceccato e collaboratori:

Un altro gruppo di regole riguarda i rapporti fra attività e loro complementi. Il verbo "dipingere", per esempio, è stato inserito in una relazione di attività-materiale con "colore"; di attività-risultato con "quadro"; e di attività-strumento con "pennello". Da questi rapporti è possibile dedurre che fra "colore" e "quadro" si pone il rapporto di materiale-risultato, fra "quadro" e "pennello" quello di risultato-strumento, ecc. I rapporti sono stati invece espressi caso per caso quando il dizionario non possieda un nome per l'attività. Per esempio, è esplicitamente indicato il rapporto 'soggetto-risultato della sua attività' per 'ciabattino/calzolaio' e 'scarpe', in quanto le attività designate con 'fare', 'fabbricare', 'aggiustare', ecc., hanno raggio di applicabilità troppo vasto per essere connesse con particolari soggetti o con particolari risultati/prodotti.

c) La terza classificazione consiste nell'individuare tutti quei verbi il cui significato suggerisce già quali complementi essi possono avere. Per questi verbi vengono costruite delle sfere nozionali individuali chiamate "costellazioni". Riporto parzialmente¹⁴ qui sotto come esempio la tabella-costellazione costruita da Ceccato e collaboratori per il verbo "portare".

¹⁴ La tabella originale contiene anche le colonne relative al russo e al polacco.

Tabella 3

Contenuto semantico		Caratteristiche del soggetto	Caratteristiche dell'oggetto e suoi rapporti	Modalità	Costellazione	Francese		Inglese		Tedesco	
							Costell.		Costell.		Costell.
1	Attività di trasportare una cosa (animata o inanimata) da un luogo ad un altro	Essere animato Forza naturale	Non muoventesi di moto proprio. Rispetto al soggetto è il rapporto di sostenuto-sostenente. (all'accusativo [in italiano per i pronomi])		da	porter	de	to carry	from	tragen	von
					a, a + art. dat. (pron.)	porter	à, à + art. dat. (pron.)	to carry	to, at dat.	tragen	zu
2	Reggere un peso			in	porter	en, dans	to carry	In	tragen	in	
				su + art. con	porter	sur avec	to carry	on with	tragen	auf mit	
				dentro		dedans	to carry in		hereinragen	in	
				fuori	di, da	sortir	to carry out	of	heraustragen	aus	
				su		monter	to carry up	at, on	herauftragen	auf	
				giù	di, da	descendre	to carry down	from	hereuntertragen	von	
				via	di, da	emporter	to carry away	from	wegtragen	von	
				indietro	a, dat. (pron.)	reporter	to carry back	to, dat. (pron.)	zurücktragen	von dat.	
				direzione del soggetto, o di colui che parla	a, dat. (pron.)	apporter	to bring	to, dat. (pron.)	bringen	zu, dat.	
		Essere umano	Essere animato muoventesi di moto proprio (all'accusativo)	direzione del soggetto, o di colui che parla	a, dat. (pron.)	amener	to bring	to, dat. (pron.)	bringen	zu, dat.	
3	Attività di trasferimento dell'oggetto da un luogo ad un altro tramite un mezzo di trasporto	Essere umano	Essere animato o mezzo di trasporto (all'accusativo)		a, a + art.	porter	à, à + art.	to drive	to	führen	zu, nach
						conduire	à, à + art.	to drive	to	führen	zu, nach
					a, a + art. a, a + art.	porter apporter	à, à + art. à, à + art.	to drive to bring	to to	führen bringen	zu, nach
				Contenente l'arrivo	via, da	emporter	de, de + art.	to carry away	from	wegführen	von
4	Condurre	Essere animato, strada, via			a, a + art.	conduire	à, à + art.	to lead	to	führen	zu, nach
5	Recare su sé o indosso	Essere umano	Indumenti Effetti personali (all'accusativo)			porter		to wear		tragen	

Questi sono dunque i ritrovati, basati sulla teoria correlazionale del pensiero, proposti per migliorare la qualità della traduzione automatica. Come possiamo ben vedere, il primo dei due ritrovati (dividere tutte le parole di una lingua in due categorie, quella dei correlati e quella dei correlatori, e collocare sempre, tranne che nei rari casi di cui abbiamo detto sopra, le parole che fanno parte della lista dei correlatori nelle caselle superiori delle triadi correlazionali e quelle che fanno parte della lista dei correlati nelle caselle inferiori) è semplice. Il secondo ritrovato (la sfera nozionale) è invece estremamente complesso e il lavoro umano richiesto per la realizzazione pratica di essa aumenta esponenzialmente all'aumentare del numero delle parole del dizionario di cui è dotato di programma. Per avere un'idea di quanto sia grande questo lavoro, calcoliamo quello necessario a costruire la parte della sfera nozionale riguardante i rapporti fra le cose (punto b e tabella 2), che è quello più facile da calcolare. Ogni parola del dizionario deve essere messa in rapporto con tutte le altre (cioè tutte meno se stessa) e occorre dire se, fra le parole di ciascuna delle coppie che si vengono così a formare, c'è o non c'è ciascuno dei tipi di rapporto elencati nella tabella 2. Formare tutte le coppie possibili può essere fatto rapidamente con il computer. Ma solo un essere umano può rispondere alla domanda se fra i due elementi di ogni coppia c'è o no ciascuno dei tipi di rapporto elencati. Il numero di queste domande è dato dalla seguente formula:

$$n_{\text{dom}} = (n_{\text{voc}}^2 - n_{\text{voc}}) \times n_{\text{rapp}}$$

dove n_{dom} è il numero delle domande, n_{voc} è il numero delle parole del dizionario del programma e n_{rapp} è il numero dei tipi di rapporto elencati in tabella 2. Per esempio, per un numero di parole del dizionario pari a 1000 (cioè un dizionario piuttosto piccolo), il numero delle domande è:

$$n_{\text{dom}} = (1000^2 - 1000) \times 55 = 54.945.000$$

mentre per un dizionario di sole 50 parole (probabilmente il minimo possibile per una sperimentazione di primo livello) il numero delle domande è:

$$n_{\text{dom}} = (50^2 - 50) \times 55 = 134.750$$

Anche se il numero delle domande è alto, dobbiamo considerare che il tempo necessario ad un essere umano per rispondere ad esse “sì” o “no” è molto breve, meno di due secondi in media (gruppi interi di tipi di rapporto possono spesso essere esclusi a colpo d'occhio, per esempio i gruppi comprendenti un'attività se le due cose non sono attività). Pertanto, le ore di lavoro richieste nel primo caso sono 30.525 (3.816 giorni lavorativi). Nel secondo caso invece il tempo è di circa 75 ore (circa 9 giorni lavorativi), quindi piuttosto breve (per altre considerazioni sulla quantità di lavoro umano necessario per poter fare delle sperimentazioni di primo livello di questo programma si veda alle conclusioni di questo articolo).

È importante notare che questa parte della sfera nozionale, una volta costruita, è costituita da molti meno elementi di quante sono le domande cui è stato necessario rispondere per costruirla. Ipotizziamo, per semplicità, che fra due cose vi sia solo un rapporto o nessun rapporto (in realtà, come si vede nella figura 1, i rapporti possono essere anche più di uno, qualche volta due e, raramente, più di due), in modo da calcolare il numero minimo teorico (che comunque non è molto lontano dal numero reale, come si capisce facilmente) degli elementi che compongono questa parte della sfera nozionale. Questo numero è uguale a quello delle coppie di parole che si possono formare combinando ciascuna parola del dizionario del programma con tutte le altre, cioè:

$$(n_{\text{voc}}^2 - n_{\text{voc}})$$

Ciascun elemento è formato semplicemente dalle due parole della coppia con accanto un numero: 0 (nessun rapporto) o il numero di codice di uno dei tipi di rapporto elencati nella tabella 2.

Vediamo ora le fasi del procedimento che il programma segue per attuare concretamente i due suddetti ritrovati.

Le fasi del procedimento

1) Il programma prende in considerazione una frase alla volta. Questo perché la frase, cioè l'insieme di parole comprese fra due punti fermi, rappresenta una rete correlazionale chiusa. Per ogni singola frase il programma tenta dunque di ricostruire la rete correlazionale giusta cercando di disporre correttamente le parole che la compongono.

Il programma genera tutte le possibili strutture correlazionali che, secondo la teoria correlazionale del pensiero, possono corrispondere alla frase in esame da un punto di vista teorico, cioè tenendo conto solo della distinzione fra correlati e correlatori e non prendendo in considerazione i rapporti (possibili o impossibili) fra essi. È da notare che il numero di queste combinazioni è un numero finito e non particolarmente elevato. Esso può essere calcolato matematicamente partendo dal numero delle parole che compongono la frase e dal numero dei correlatori espliciti in essa presenti. Come calcolare questo numero e come il programma genera queste combinazioni sono problemi che non possono essere affrontati in un articolo come questo, che è solo una presentazione di carattere generale. Possiamo però notare che molte combinazioni non vengono nemmeno generate perché sono incompatibili con l'ordine delle parole del testo. Per esempio, in una frase come “ci sono una bottiglia di vino e un vassoio con alcuni dolci” le regole che governano l'ordine delle parole nella lingua italiana (e probabilmente in qualsiasi altra lingua) escludono che “bottiglia” sia correlata a “dolci” o “vino” sia correlato a “vassoio”.

2) Il programma, come altri programmi di traduzione automatica, è dotato di un dizionario che comprende anche tutte le forme flesse di ogni voce: esso è pertanto formato da parole e non da voci. Per ogni singola parola sono memorizzate una serie di informazioni. La prima e fondamentale informazione è se la parola è un correlatore oppure un correlato (nel modello di programma proposto da Ceccato e collaboratori era anche prevista la possibilità che una parola che normalmente designa un correlatore sia invece un correlato, ma come abbiamo visto questi casi sono così rari che a mio parere nelle fasi iniziali di realizzazione del programma non è conveniente prevederli). Poiché i correlatori possono occupare solo le caselle superiori degli elementi grafici che usiamo per simbolizzare le correlazioni, mentre i correlatori possono occupare solo quelle inferiori, questa fondamentale classificazione riduce drasticamente le possibilità di combinazione fra le parole della frase in esame.

La lista dei correlatori italiani compilata da Ceccato e collaboratori (tabella 4, parole in corsivo¹⁵) era vasta, probabilmente completa (comprendeva 82 elementi, in grande maggioranza preposizioni e con-

¹⁵ Ho aggiunto a fianco di ciascun correlatore italiano, fra parentesi, le parole o le espressioni inglesi usate per tradurlo, in modo da evidenziare un importante problema che si presenta nella traduzione automatica dei correlatori. Infatti, va notato che in molti casi a un singolo correlatore italiano non corrisponde un singolo correlatore inglese. Questo perché l'inglese in certi casi fa distinzioni o presenta sfumature di significato che sono diverse o assenti in italiano; oppure, più spesso, perché in certi contesti usa correlatori diversi. Ho evidenziato in grassetto la parola (o l'espressione) inglese che può essere considerata l'omologo del correlatore italiano in esame (talora è stato necessario indicarne più di una). Le altre parole rappresentano gli altri modi in cui può o, in certi casi, deve essere tradotto il correlatore italiano (perché in quei casi l'inglese usa un altro correlatore). Il fatto che un correlatore debba essere tradotto in modi diversi, a seconda dei casi, quando si traduce da una lingua all'altra, rappresenta un problema importante per la traduzione automatica, del quale parleremo più avanti.

giunzioni, ed anche qualche parola considerata dalle grammatiche avverbio), perché comprendeva anche correlatori di uso raro o arcaico, come ad esempio “acciocché”, “allorché”, “allorquando” ecc.. La maggior parte di questi termini non hanno un senso diverso da termini usati molto più comunemente (per esempio: “allorquando” ha significato uguale a “quando”) per cui probabilmente conviene far sì che il programma sostituisca preliminarmente il termine raro o arcaico con quello comune.

Tabella 4

1) <i>a</i>	(to; at; in, on)	34) <i>giacché</i>	(since , as, now that)
2) <i>acciocché</i>	(so that)	35) <i>in</i>	(in , at, inside, on, to, into, through, across, forma aggettivale, by)
3) <i>affinché</i>	(so that)	36) <i>innanzi</i>	(before , in front of)
4) <i>allorché</i>	(when)	37) <i>lungo</i>	(along , during, over)
5) <i>allorquando</i>	(when)	38) <i>ma</i>	(but , still, yet, non è richiesta una parola inglese)
6) <i>ancorché</i>	(even if , though)	39) <i>malgrado</i>	(in spite of , notwithstanding, with all, for all)
7) <i>anzi</i>	(on the contrary; even better; rather)	40) <i>mediante</i>	(by , by means of, through)
8) <i>anziché</i>	(rather than , instead of)	41) <i>meno</i>	(except , but)
9) <i>appena</i>	(as soon as)	42) <i>mentre</i>	(while , whereas)
10) <i>appresso</i>	(close; near to; behind)	43) <i>né</i>	(neither , nor, either, or)
11) <i>attraverso</i>	(through; across)	44) <i>nonché</i>	(let alone , still less, as well as)
12) <i>benché</i>	(although , though, however)	45) <i>nonostante</i>	(in spite of , despite, for all)
13) <i>che</i>	(that , non è richiesta una parola inglese, than, when, since, for, only, but)	46) <i>o</i>	(or , either... or..., whether... or...)
14) <i>circa</i>	(about , as regard, concerning)	47) <i>oltre</i>	(beyond , over, more than)
15) <i>come</i>	(as , as soon as)	48) <i>onde</i>	(so that , in order that)
16) <i>con</i>	(with , by)	49) <i>oppure</i>	(or , or on the other hand, or alternatively; or else , otherwise)
17) <i>contro</i>	(against)	50) <i>ovvero</i>	(that is , in other words, or, or rather)
18) <i>cosicché</i>	(so , so that)	51) <i>per</i>	(for , in order to, to)
19) <i>da</i>	(from , away (from), to, at, through, by, for, with, since, (as) from, like, as, forme aggettivali)	52) <i>perché</i>	(because; so that , in order that, so as)
20) <i>dacché</i>	(since)	53) <i>più</i>	(plus)
21) <i>dentro</i>	(inside , in, within, into)	54) <i>poiché</i>	(since , as, for)
22) <i>di</i>	(of , genitivo Sassone, aggettivo attributivo, forma aggettivale o avverbiale, at, in, by, than, with, for, from)	55) <i>presso</i>	(near, beside , by, next to, with, in, at, among)
23) <i>dietro</i>	(behind , after)	56) <i>però</i>	(yet , but; nevertheless , though, however)
24) <i>dopo</i>	(after , past, since)	57) <i>purché</i>	(provided , as long as; if only)
25) <i>durante</i>	(during)	58) <i>qualora</i>	(if)
26) <i>e</i>	(and , but)	59) <i>quando</i>	(when , whenever, while, since, if)
27) <i>eccetto</i>	(except , excepting, but, save)	60) <i>quanto</i>	(what)
28) <i>entro</i>	(within , in, by, before)	61) <i>quantunque</i>	(although; even if)
29) <i>finché</i>	(till , until, as long as)	62) <i>quasi</i>	(as if)
30) <i>fino</i>	(till , until, up to, as far as, to)	63) <i>rasente</i>	(non esiste un omologo inglese ¹)
31) <i>fintantoché</i>	(= <i>finché</i>)		
32) <i>fuorché</i>	(except , but, other than, apart from)		
33) <i>fuori</i>	(out of , outside)		

64) <i>salvo</i>	(except , but, save, barring; apart)		
65) <i>se</i>	(if , whether, if only)		
66) <i>sebbene</i>	(though)		
67) <i>secondo</i>	(according to , in accordance with, depending on)		
68) <i>senza</i>	(without , -less, un-, in-)		
69) <i>seppure</i>	(even if)		
70) <i>sicché</i>	(so , so that)		
71) <i>siccome</i>	(as , since, because)		
72) <i>sopra</i>	(on , up, on to, onto; over ; above; after)		
73) <i>sotto</i>	(under , beneath, underneath, be-		
			low)
		74) <i>su</i>	(on , up, on to, onto; over ; above; after; about; out of)
		75) <i>talché</i>	(= <i>cosicché</i>)
		76) <i>tra</i>	(between , among , amid , within , in)
		77) <i>tramite</i>	(through)
		78) <i>tranne</i>	(except , but, save)
		79) <i>verso</i>	(toward(s) ; near, about, to)
		80) <i>via</i> ²	(through , by means of)
		81) (<i>relativo</i>)	(relative)
		82) (<i>operazione matematica</i>)	(mathematical operation)

¹ Questa preposizione ha lo stesso significato dei verbi inglesi “to graze”, “to skim”, ed è generalmente tradotta con questi verbi o un altro verbo di moto più “close to”.

² Il nome italiano “via” può essere usato come una preposizione, nel senso di “attraverso”, “per mezzo di” (per esempio: “Lo so via mio fratello”).

Oltre a questi correlatori, la lista comprendeva un'altra serie di tipi di correlazioni in cui il correlatore è sempre lo stesso, cioè il correlatore implicito, e ciò che cambia è la categoria grammaticale dei correlati. Erano cioè elencate queste correlazioni:

Tabella 5

1) soggetto-svolgimento	19) svolgimento al congiuntivo-svolgimento al condizionale
2) svolgimento-oggetto	20) svolgimento al congiuntivo-.....
3) articolo-.....	21) svolgimento al condizionale-.....
4) aggettivo-.....	22) svolgimento al participio-.....
5)-aggettivo	23) svolgimento al gerundio-.....
6) avverbio-.....	24) svolgimento al condizionale-svolgimento al congiuntivo
7) predicato-.....	25)-genitivo
8)-predicato	26) svolgimento-dativo
9) ausiliare-participio passato	27) svolgimento-separazione
10) ausiliare-gerundio presente	28)-relazione
11) ausiliare-infinito	29)-..... (ripetizione)
12) “si”-svolgimento	30) nome comune-nome comune
13) forma media-svolgimento	31) nome comune-nome proprio
14) “ci, vi”-svolgimento	32) nome proprio-nome comune
15)-svolgimento al congiuntivo	33) nome proprio-nome proprio
16)-svolgimento al condizionale	
17)-svolgimento al participio	
18) svolgimento al gerundio	

Anche se il correlatore di tutti questi tipi di correlazioni è sempre lo stesso, è necessario distinguere tutti i tipi di correlazioni a cui esso può dare origine a seconda di che cosa sono, dal punto di vista grammaticale, i due correlati. Questo per due motivi:

– il correlatore implicito è di gran lunga il più usato di tutti i correlatori e nella maggior parte delle frasi la maggioranza dei correlatori sono rappresentati da esso: pertanto distinguere le correlazioni a cui esso dà luogo in base alla natura grammaticale di ciò che esso lega è di fondamentale importanza per la ricostruzione della rete correlazionale del pensiero;

– questa classificazione è fondamentale per la successiva opera di traduzione (per esempio se la correlazione è “soggetto-svolgimento” il primo correlato, in una lingua che possiede i casi, andrà flesso al nominativo, se la correlazione è “svolgimento-oggetto” il secondo correlato andrà all’ accusativo, ecc. ecc.).

Le parole che possono essere solo dei correlati, come abbiamo detto, possono occupare solo le caselle inferiori della triade correlazionale. Su esse deve essere compiuto un lungo e paziente lavoro.

Per prima cosa ogni singola parola della lingua dalla quale si vuole che il programma traduca (lingua di entrata o L1) deve essere classificata dal punto di vista grammaticale. Occorre cioè specificare quale parte del discorso essa può rappresentare. Si noti che le possibilità sono più numerose di quelle che elencano i comuni dizionari per le singole voci. Infatti una parola come l’italiana “amo” per il dizionario può essere soltanto un nome, ma nel dizionario del programma, dove sono comprese anche tutte le parole che derivano dalla flessione, essa può essere anche la prima persona del presente indicativo del verbo “amare”; la parola inglese *bears* può essere la terza persona del presente indicativo del verbo *to bear*, “portare”, ma può essere anche il plurale del nome *bear*, “orso”; ecc. ecc.. La flessione, cioè, può portare alla formazione di due o più parole aventi grafia identica ma significato completamente diverso. Inoltre c’è il fatto, molto più importante, che quando L1 è una lingua poco flettente mentre la lingua nella quale si vuole che il programma traduca (lingua di uscita o L2) è una lingua fortemente flettente, succede spesso che a una stessa parola di L1 corrispondono in L2 molte parole diverse (abbiamo visto sopra quante possono essere le uscite italiane per parole come le inglesi *love* e *loved*). Per ogni singola parola avremo quindi un elenco di possibili significati (S1, S2, S3 ecc.).

Ciascuno di questi possibili significati, a sua volta, deve poi essere classificato dal punto di vista grammaticale. Per esempio, la parola italiana “amo” avrà questi due possibili significati, che verranno così classificati:

- **AMO**
 - S1: “nome”
 - genere: maschile
 - numero: singolare
 - S2: “verbo”
 - modo: indicativo
 - tempo: presente
 - persona: 1^a

Per ognuno dei possibili significati di ogni singola parola occorre poi indicare quali sono le sue possibilità di porsi in rapporto con i singoli correlatori. Occorre cioè specificare se ogni singolo significato può essere o no il correlato di ognuno dei correlatori che abbiamo visto sopra e, in caso affermativo, se esso può essere il primo oppure il secondo oppure entrambi. Ad esempio, il significato S1 della parola italiana “amo” (l’attrezzo da pesca, dal punto di vista grammaticale un nome) potrà essere per esempio il primo correlato della correlazione “soggetto-svolgimento” (“[l’]amo è nascosto dall’esca), il secondo correlato della correlazione “svolgimento-oggetto” (“ho [un] amo”), ma non potrà essere per esempio il

primo correlato della correlazione “-avverbio”, così come non potrà essere un correlato di molti dei tipi di correlazione in cui da correlatore funge il correlatore implicito e che abbiamo elencato nella tabella 4 (si veda la tabella). Per quanto riguarda gli altri correlatori, la parola “amo” come nome potrà essere il correlato di preposizioni come “di”, “a”, “da” ecc. (“amo di acciaio”, “abbozzare all’amo”, “amo da buttare” ecc.), ma non può essere il correlato di congiunzioni come “affinché” e “quando” che correlano fra loro sempre due proposizioni; non può essere il primo correlato della congiunzione “durante” perché esso deve essere una proposizione, e nemmeno il suo secondo correlato perché esso deve essere il nome di qualcosa che ha una durata; ecc. ecc. (se il lettore farà delle prove con i correlatori che abbiamo elencato nella tabella 4 si renderà conto che le incompatibilità sono molte).

Il significato S2 della parola italiana “amo” (la voce del verbo “amare”) potrà per esempio essere il secondo correlato della correlazione “soggetto-svolgimento” (“io amo”) o il primo correlato della correlazione “svolgimento-oggetto” (“amo lei”), ma non potrà essere il primo correlato della correlazione “.....-aggettivo” né il secondo correlato della correlazione “articolo-.....” così come non potrà essere un correlato di molti dei tipi di correlazione in cui da correlatore funge il correlatore implicito e che abbiamo elencato nella tabella 4 (si veda la tabella). Per quanto riguarda gli altri correlatori, la parola “amo” come verbo potrà per esempio essere il correlato di una congiunzione come “quando” (“quando amo sono felice”), il primo correlato della congiunzione “affinché” ma non il suo secondo correlato perché quest’ultimo deve essere un verbo al congiuntivo; potrà essere il primo correlato della preposizione “con” (“amo con passione”) ma non il suo secondo correlato, che deve essere un nome o un verbo all’infinito; ecc. ecc. (anche in questo caso, se il lettore farà delle prove con i correlatori che abbiamo elencato nella tabella 4, si renderà conto che le incompatibilità sono molte).

Come ben si vede, l’aver indicato, per ognuno dei possibili significati di ogni singola parola, quali sono le sue possibilità di porsi in rapporto con i singoli correlatori esclude molte delle triadi correlazionali che si possono teoricamente formare combinando i correlatori e i correlati di una determinata frase in tutti i modi teoricamente possibili.

3) A questo punto il programma esamina le reti di correlazioni teoricamente possibili rimaste tenendo conto di due serie di regole che sono presenti nella grammatica della lingua di entrata e cioè:

a) le regole di concordanza per genere, numero, persona, caso ecc. che, come si è detto, sono molto numerose nel caso delle lingue tendenzialmente flettenti. In italiano, per esempio, è escluso che un aggettivo al femminile si riferisca ad un nome di genere maschile, il verbo personale concorda sempre nella persona col pronome col quale è correlato ecc.;

b) le regole riguardanti l’ordine che le parole debbono avere, che, come si è detto, sono molto importanti nel caso delle lingue tendenzialmente isolanti (in inglese, per esempio, il complemento oggetto segue sempre il soggetto, l’aggettivo generalmente precede il nome a cui è riferito ecc.).

Questa fase della elaborazione porta ad un drastico calo delle reti correlazionali teoricamente possibili. Questo è particolarmente vero nel caso delle lingue fortemente flettenti come ad esempio il latino, le lingue slave ecc.. È probabile che in molti casi, specie appunto in questo tipo di lingue, a questo punto risulti possibile una sola rete correlazionale, quella corretta.

4) Se questo non si verifica, entra in gioco la “sfera nozionale” di cui abbiamo già parlato. Essa è composta, come abbiamo accennato, delle due seguenti parti.

a) Una lista di classificazioni come quella della tabella 1. A ciascuna delle parole di cui è dotato il vocabolario del programma vengono attribuite una o più di queste classificazioni. Per esempio, la parola “cane” sarà classificata come appartenente alle seguenti categorie della tabella 1:

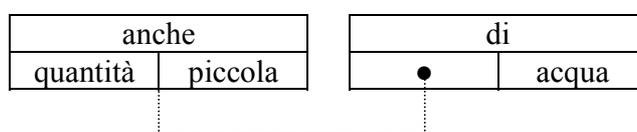
- esseri viventi (1)
- esseri animati (3)
- animali (6)
- animali di terra (25)
- carnivori (28)

- animali domestici (32)

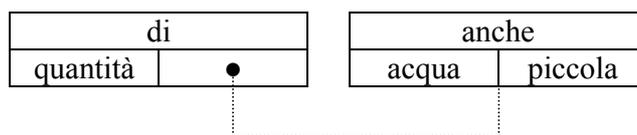
b) Una rete di relazioni fra parole come quella riportata parzialmente nella figura 1, dove una linea che congiunge due parole indica che fra i loro significati sono state trovate una o più relazioni, quelle indicate dai numeri posti lungo la linea stessa.

Le reti correlazionali che non sono state escluse dalla precedente fase vengono esaminate singolarmente tenendo conto delle numerose classificazioni che costituiscono le due suddette parti della “sfera nozionale” e tutte le reti che risultano incompatibili con una o più di queste classificazioni vengono eliminate. Per esempio, nel caso, cui ci siamo già riferiti, delle due espressioni “quantità di acqua, anche piccola” e “quantità di acqua, anche fredda”, il programma, giunto alla fine della fase 3, dovrebbe aver considerato come teoricamente possibili solo queste due reti correlazionali per ciascuna espressione:

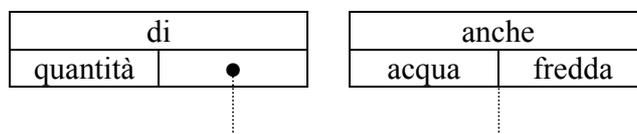
1a)



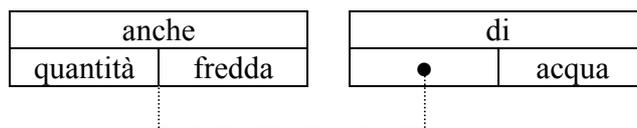
1b)



2a)



2b)



Se nella sfera nozionale la parola “quantità” è legata alla parola “piccola” da un rapporto classificato come “caratteristica costitutiva-cosa caratterizzata (n° 6 della tabella 2), la rete correlazionale 1b) verrà esclusa e rimarrà pertanto solo la rete correlazionale corretta. Se nella sfera nozionale la parola “fredda” è classificata, fra l’altro, come “caratteristica fisica” mentre la parola “quantità” è classificata come “cosa astratta” nel caso della frase “quantità di acqua, anche fredda” la rete correlazionale 2b) verrà esclusa e rimarrà pertanto solo la rete correlazionale corretta.

Un altro esempio di espressione in cui dopo l’elaborazione portata avanti fino alla fine della fase rimangono possibili due reti correlazionali è “penna a sfera rossa”. Si tratta di un caso più critico, perché l’aggettivo “rossa” è perfettamente compatibile sia con il nome “penna” sia con il nome “sfera” e pertanto la lista di classificazioni della tabella 1 non esclude nessuna delle due possibilità. Ma in questo caso una sfera nozionale, del tipo di quella che abbiamo riportato nella figura 1, che fosse ben fatta dovrebbe contenere fra “penna” e “sfera” le relazioni “tutto-parte” e “composto-componente”. In questo

caso la mia proposta è che il programma proceda non per esclusione ma per probabilità. Poiché la correlazione tra “rossa” e “sfera” è solamente ritenuta possibile mentre la correlazione di “penna” con “sfera” è sostenuta da ben due tipi di rapporto della sfera nozionale di cui alla tabella 2, il programma sceglierà quest’ultima possibilità.

Naturalmente, la probabilità che il programma arrivi a considerare possibile solo una rete correlazionale per ogni frase e che questa sia quella giusta aumenta tanto più quanto più è ricco ed accurato il complesso di classificazioni che costituisce la sfera nozionale. Nella realizzazione pratica del programma del quale sto parlando il problema è proprio questo: il lavoro umano necessario alla costruzione della sfera nozionale aumenta in maniera esponenziale all’aumentare del numero delle parole di cui è formato il vocabolario del programma. Come ben si vede, però, si tratta solo di una difficoltà di tipo pratico, non teorico.

C’è un’altra difficoltà che si potrebbe pensare che si presenti. Le espressioni o le frasi che abbiamo riportato finora come esempi di testo da tradurre o tradotto sono tutte brevi e prive di segni di interpunzione. Ricostruire le reti correlazionali corrispondenti ad esse è, per un essere umano, assai semplice. Ma le frasi che compongono un testo spesso non sono così brevi. Possono essere anche molto lunghe e complesse, dipende dallo stile dello scrivente e in parte anche dalla lingua usata. Ci si può domandare se anche in questi casi la teoria correlazionale del pensiero è valida, cioè se per ogni frase è ricostruibile una rete correlazionale del tipo di quelle viste, anche se molto più complessa. In base alla esperienza fin qui maturata dalla SOI e da me, la risposta è affermativa. In certi casi si possono avere dei dubbi, delle perplessità, dei quali non è opportuno parlare in uno scritto come questo di carattere molto generale, ma si tratta di difficoltà risolvibili. Davanti a frasi lunghe all’inizio si possono avere delle difficoltà, ma con un poco di esperienza il compito non risulta difficile. Questo vale per l’essere umano, ma varrà anche per il programma? A questa domanda è impossibile, o perlomeno estremamente difficile, dare a priori una risposta sicura. Occorre vedere come si comporta il programma all’atto pratico. Ovviamente, le prime sperimentazioni vanno fatte con dizionari molto contenuti e su frasi molto semplici. All’aumentare della lunghezza della frase aumenta, e di molto, la quantità di elaborazione necessaria, ma, vista l’enorme potenza dei processori attualmente disponibili, questo non è sicuramente un problema. Tuttavia, quando la frase è composta da più proposizioni (periodo), una principale a cui sono legate altre proposizioni coordinate e/o subordinate, reti complesse di parole fra loro correlate rappresentano uno dei correlati di una triade correlazionale ed è possibile che questo crei delle difficoltà al programma. A quanto mi risulta, questo problema non fu seriamente affrontato da Ceccato e collaboratori nei pochi anni della durata del loro programma di ricerca (dal 1959 al 1966). Non ho compiuto ricerche approfondite su questo problema, ma a prima vista mi sembra che esso possa essere affrontato in due modi.

1) Il primo modo è quello di prendere in considerazione i segni di punteggiatura. Abbiamo visto che il punto fermo chiude una rete correlazionale. Lo stesso vale per il punto interrogativo e per il punto esclamativo. Ci si può chiedere che cosa designano la virgola, il punto e virgola, e i due punti. Sulle grammatiche troviamo definizioni di questo tipo:

“I segni di punteggiatura [...] servono ad indicare, nella scrittura, le pause più o meno lunghe che tutti facciamo parlando, e a separare tra loro le proposizioni di uno stesso periodo, e anche le parti della stessa proposizione.” [Bruno Migliorini, Grammatica Italiana].

Un’affermazione di questo tipo non è sbagliata, ma non costituisce certo una definizione esauriente. Occorre capire che cosa sono queste separazioni e perché vengono messe in certe posizioni e non in altre. Non ho ancora condotto ricerche approfondite sull’argomento, ma in base ai primi risultati ottenuti posso fare alcune considerazioni. Vediamo quelli che sembrano essere gli usi principali della virgola,

facendo alcuni esempi. Occorre però premettere che l'uso della virgola è piuttosto libero, cioè in diversi casi la virgola può essere messa oppure no .

a) La virgola separa la proposizione principale da quella coordinata o subordinata in certi casi, ma non in altri:

- Non sapeva nuotare, ma si è salvato lo stesso
- Ha un'automobile, anzi due
- Penso, dunque sono
- Mi disse che era andato
- Che era andato, mi disse
- Non so come ha fatto
- Come ha fatto, non so
- Quando arriverai, troverai Giovanni alla stazione
- Troverai Giovanni alla stazione quando arriverai
- Mentre dormivi, hanno telefonato due volte
- Hanno telefonato due volte, mentre dormivi
- Se verrai, ti ospiterò molto volentieri
- Ti ospiterò molto volentieri, se verrai
- Non so se dovrò partire

b) Si mettono fra due virgole le apposizioni, se esse seguono il sostantivo a cui si riferiscono, talora sì talora no se lo precedono:

- Cristoforo Colombo, lo scopritore dell'America, nacque a Genova
- Lo scopritore dell'America, Cristoforo Colombo, nacque a Genova
- Omero, il grande poeta greco, era cieco
- Il grande poeta greco Omero era cieco

c) Si mettono fra due virgole le proposizioni e le espressioni incidentali:

- Giovanni, a mio avviso, è stato il migliore
- Il dottor Bianchi, dicono, è molto ricco
- Il dispositivo, ideato dall'ingegner Rossi, è molto valido

mentre non si mettono le virgole se si vuole che le espressioni o la proposizione non abbiano valore incidentale:

- Il dispositivo ideato dall'ingegner Rossi è molto valido

d) Si mettono fra due virgole le costruzioni cosiddette assolute:

- La donna, fatte le valigie, se ne andò

e) Si mette la virgola dopo i vocativi:

- Cari amici, vi saluto

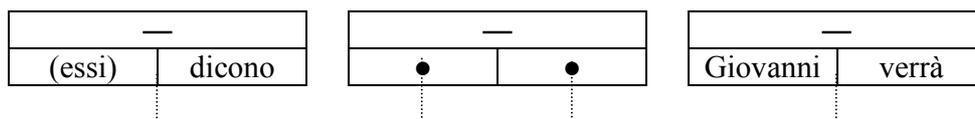
f) Si mettono delle virgole dopo i vari elementi di un elenco, tranne l'ultimo, in genere, ma talora anche dopo questo:

- Ho comprato pere, mele e uva
- Artisti, intellettuali, scienziati hanno aderito alla manifestazione

g) Si mette la virgola quando si potrebbe pensare che due parole adiacenti siano correlate fra di loro ed invece non è così:

- “Sorgi la sera e vai,
contemplando i deserti” [G Leopardi, *Alla luna*]

In base a questi esempi, sembra che la virgola indichi, nella maggior parte dei casi, che le parole comprese fra due di esse (oppure fra una virgola e il punto che la precede o la segue) formino nel loro insieme una “sezione” della rete correlazionale della frase, cioè una parte di essa che *nel suo insieme* è legata al resto della frase. Essa può essere il correlato di una parola o di un insieme di parole che fa parte del resto della frase, come nel caso di cui al soprastante punto a), ma può anche essere apparentemente più indipendente, come nel caso delle proposizioni e delle espressioni incidentali. In quest’ultimo caso si potrebbe addirittura pensare che un’unica frase è formata da due reti correlazionali indipendenti, ma direi che anche se grammaticalmente la proposizione principale e quella incidentale talvolta potrebbero sussistere anche indipendentemente, a livello del pensiero vi è un innegabile legame fra le due. Cioè una frase come “Giovanni, dicono, verrà” equivale a “Dicono che Giovanni verrà”. Pertanto nel caso delle espressioni e proposizioni incidentali direi che la struttura correlazionale del pensiero è la seguente:



Si può pertanto dire che nella maggior parte dei casi la virgola esclude che la parola che la precede immediatamente e quella che immediatamente la segue siano correlate fra loro direttamente. L’unica eccezione importante sembra il caso di cui al punto f), quello in cui la virgola separa gli elementi di un elenco. In questo caso direi però che la virgola ha il valore di un correlatore, quello della congiunzione “e”, che altrimenti verrebbe ripetuta con un effetto sgradevole.

Quindi le virgole dividono le frasi in parti relativamente indipendenti che possono essere considerate dal programma come frasi a sé, cioè sequenze di parole delle quali programma deve ricostruire la struttura correlazionale senza tener conto delle parole che formano la restante frase. In questo modo si ottiene una notevolissima riduzione della elaborazione necessaria per ricostruire le reti correlazionali di frasi lunghe e complesse. Nel caso che il testo di partenza venisse preparato in vista della traduzione automatica, sarebbe pertanto raccomandabile un uso abbondante, purché corretto, delle virgole.

Questo modo di procedere però in diversi casi può causare problemi, particolarmente quando la lingua di entrata è una lingua con poca morfologia mentre quella di uscita è una lingua con ricca morfologia. Si consideri per esempio la frase che ho appena scritto “sarebbe pertanto raccomandabile un uso abbondante, purché corretto, delle virgole”. In questo caso, nella traduzione per esempio dall’inglese all’italiano, il programma, che considera l’espressione corrispondente a “purché corretto” (*as long as correct*) a sé stante, non saprebbe se tradurre l’aggettivo *correct* con il maschile oppure con il femminile. Nel caso quindi che vi siano queste ambiguità il programma dovrebbe cercare, con le consuete modalità, quale è la parola o la correlazione con cui è correlato l’insieme delle parole compreso fra due virgole (o fra la virgola e un punto). Questo non toglie che le virgole rappresentano in ogni caso un modo per ridurre considerevolmente l’elaborazione necessaria alla traduzione di una frase.

Il caso particolare degli elenchi costituisce però un problema che va risolto. Infatti, nel caso di una frase come “ho comprato mele, pere e uva”, che il programma consideri “pere e uva” un’espressione indipendente può anche essere accettabile nel caso di una lingua che non possiede i casi, ma nel caso di una

lingua che li possiede il programma non saprebbe quale caso assegnare alle parole che devono tradurre “pere” e “uva”. Non dovrebbe però essere difficile fare in modo che il programma riconosca gli elenchi perché essi sono sempre composti da sequenze di parole (o espressioni) i cui elementi di separazione sono una o più virgole e, da ultimo, una congiunzione “e”. Inoltre queste parole, o le prime parole delle espressioni, appartengono alla stessa categoria grammaticale, cioè sono o tutti i nomi o tutti aggettivi ecc..

Per quanto riguarda gli altri segni di punteggiatura, cioè il punto e virgola, i due punti, le parentesi, le lineette, valgono, a maggior ragione, le stesse considerazioni fatte per la virgola, dal momento che essi rappresentano delle interruzioni più forti della virgola.

2) Il secondo modo per ridurre l’elaborazione necessaria per ricostruire la rete correlazionale di frasi composte da più proposizioni, è quello di tener conto del fatto che le proposizioni coordinate o subordinate alla proposizione principale sono introdotte da parole specifiche. Nel caso delle proposizioni coordinate si parla di congiunzioni coordinative (“e”, “o”, “né”, “ma”, “anzi”, “dunque”, “inoltre”, “invece” “nemmeno” ecc.); nel caso delle proposizioni subordinate si parla di congiunzioni subordinate (“che”, “come”, “quando”, “mentre”, “se”, “benché”, “finché”, “poiché”, “perché” ecc.). Molte di queste congiunzioni hanno come correlati, obbligatoriamente o almeno molto spesso, non singole parole ma intere proposizioni. Pertanto queste congiunzioni possono essere usate come marcatori di inizio di proposizioni che in un primo momento possono essere considerate indipendenti, cioè di sequenze di parole che devono essere correlate fra loro a formare una rete correlazionale senza tener conto del resto. Solo in un secondo momento si cercherà la parola (o il gruppo di parole), precedente (o, più raramente, seguente) la congiunzione, che la congiunzione stessa correla a questa rete.

Come si vede, usando questi due metodi anche le frasi lunghe e complesse si riducono ad un insieme di proposizioni che sono decisamente più brevi, per le quali dunque è più facile per il programma ricostruire la rete correlazionale. Solo in un secondo momento, se necessario, queste proposizioni verranno correlate fra loro.

Una volta che il programma ha ricostruito la rete correlazionale, si passa alla fase di traduzione vera e propria. Se la rete correlazionale è stata correttamente ricostruita, il programma non può compiere quegli errori di traduzione che derivano dal non stabilire a quale parola o gruppo di parole ogni singola parola è riferita, perché, come si è detto, le regole che governano il passaggio dalla rete correlazionale del pensiero alla sua espressione linguistica sono ben precise in ogni lingua. Saranno evitati anche molti di quegli errori di traduzione che derivano dal fatto che certe voci del dizionario danno luogo, a causa della flessione, a parole identiche ma con significato completamente diverso. Per esempio, se il programma ricostruisce la rete correlazionale del pensiero corrispondente ad una frase come “Mi ha detto che la porta è aperta” è del tutto escluso che consideri la parola “porta” come voce del verbo “portare” (producendo per esempio una traduzione inglese tipo “He has told me he is bringing it is open”, cioè “Mi disse che lo/la sta portando è aperto”).

Rimane il problema degli omonimi che non derivano dalla flessione e che quindi hanno caratteristiche grammaticali identiche, come ad esempio la parola “cane” che, come ho ricordato, rimanendo sempre un nome maschile e singolare, può indicare, oltre che il comune animale domestico, una parte di certe armi da fuoco, un attrezzo ecc.. Anche in questi casi, la ricostruzione della rete correlazionale del pensiero è estremamente utile. Nei casi più fortunati, quando una parola di questo genere è in diretta correlazione con qualcos’altro che indica chiaramente qual è il significato in gioco, come avviene per esempio in una correlazione del tipo “cane del fucile” o “cane inceppato”, sarà la sfera nazionale di cui è dotato il programma che, contenendo la classificazione del rapporto fra questi due oggetti come “parte-tutto”, “componente-composto” ecc., automaticamente farà fare la scelta giusta. Per gli altri casi si potrebbe dotare il programma di una funzione di “esplorazione” del contesto. Cioè, quando in una frase compare una di queste parole critiche, come “cane”, il programma, dopo aver ricostruito la relativa rete

correlazionale, potrebbe esaminare le parole che la compongono (e magari anche quelle che compongono le frasi immediatamente precedenti o seguenti) alla ricerca di parole che indichino se è più probabile che sia in gioco un significato piuttosto che un altro della parola in esame (per esempio, se in una frase in cui è compresa la parola “cane” e/o nelle frasi vicine ricorrono parole come “fucile”, “pistola”, “arma”, “grilletto” ecc., è molto probabile che il significato in gioco sia quello relativo alle armi da fuoco. Naturalmente per questa funzione di esplorazione del contesto ci si avvarrebbe delle classificazioni contenute nella sfera nazionale di cui nella tabella 2.

Ancora un altro problema che la traduzione automatica ha sempre presentato è quello della traduzione delle preposizioni¹⁶, quindi di correlatori. Come abbiamo già accennato nella nota 15, sono frequenti i casi in cui un certo correlatore della lingua di entrata non può essere tradotto sempre con lo stesso correlatore della lingua di uscita, ma deve essere tradotto in modi diversi secondo il contesto. Una complessa sfera nozionale come quella che abbiamo descritto, contenente le molte classificazioni che abbiamo elencato nella tabella 2, è in grado di risolvere, almeno in buona parte, questo problema. Infatti, supponiamo per esempio di dover tradurre dall’italiano all’inglese le due correlazioni “profumo di rosa” e “quadro di Raffaello”. In italiano esse hanno per correlatore entrambe la preposizione “di”, ma in inglese nel secondo caso non è possibile usare la preposizione “of”, ma è necessario usare “by”. Nella sfera nozionale, fra le parole “profumo” e “rosa” sarà sicuramente presente il rapporto 6 della tabella 2 (caratteristica-cosa caratterizzata) e fra le parole “quadro” e “Raffaello” sarà sicuramente presente il rapporto 7 (prodotto-produttore). Non sarà quindi difficile fare in modo che il programma traduca la preposizione italiana “di” con la preposizione inglese “of” nel primo caso e invece con la preposizione “by” nel secondo.

L’ultima fase delle elaborazioni è costituita dall’inserimento o dalla rimozione o dalla sostituzione di elementi secondo le regole grammaticali della lingua di uscita (per esempio: inserimento o rimozione o sostituzione di articoli a seconda che le due lingue in gioco li abbiano o no o li usino in maniera diversa ecc.; inserimento dei pronomi personali che in lingue come l’italiano possono essere sottintesi mentre in lingue come l’inglese devono essere espressi; ecc.). Anche in questa fase, naturalmente, la preventiva ricostruzione della rete correlazionale del pensiero sarebbe di grande utilità.

Si possono prevedere anche ulteriori fasi, per esempio la ricerca delle espressioni idiomatiche, che, una volta che è stata individuata la loro struttura correlazionale, sono più facili da trovare e da collocare nella posizione giusta nella loro versione tradotta.

Conclusioni

Mi sembra dunque che si possa affermare, come conclusione di questo scritto, che la realizzazione di un programma di traduzione automatica basato sulla ricostruzione della rete correlazionale del pensiero, anche se necessita di un lavoro umano molto grande e che cresce esponenzialmente all’aumentare delle parole che compongono il dizionario di cui è dotato il programma, dovrebbe consentire un considerevole miglioramento della qualità della traduzione automatica prodotta. La ragione fondamentale di questo è che la ricostruzione, mediante gli artifici illustrati, della rete correlazionale del pensiero *simula, anche se non riproduce, una parte di ciò che effettivamente fa l’essere umano con la sua intelligenza quando traduce*. Ricostruire la rete correlazionale di pensiero corretta corrispondente al testo significa compiere una delle due fondamentali operazioni che, come abbiamo visto, costituiscono la comprensione di un testo. Se il programma fa questo prima di tradurre, i gravi errori, che alcune volte rendono il testo di uscita del tutto incomprensibile, derivati dal fatto che il programma non “capisce” a quale paro-

¹⁶ Hutchins W. J., ‘*Retrospect and prospect in computer-based translation*’, p. 2. In Machine Translation Summit VII, 13th-17th September 1999, Kent Ridge Digital Labs, Singapore. Proceedings of MT Summit VII “MT in the great translation era”, 30-34. [Tokyo]: Asia-Pacific Association for Machine Translation, 1999.

la o gruppo di parole una certa parola è riferita, vengono eliminati completamente. Gli unici gravi errori che possono rimanere sono fondamentalmente quelli derivati dal fatto che il programma non ha scelto il significato giusto nel caso delle parole con più di un significato, ma anche per questo problema un programma come quello descritto dovrebbe essere, come abbiamo visto, di grande aiuto. Inoltre, quest'ultimo problema è di minore importanza perché, se una persona che legge il testo di uscita si rende conto che una certa parola non ha nulla a che vedere con il senso generale della frase, può sempre, se il programma consente di accedere al suo dizionario bilingue, andare a vedere quali sono gli altri possibili significati della parola in questione. In alternativa il programma stesso può indicare, ponendoli per esempio fra parentesi quadre, quali sono i possibili significati dei principali omonimi.

La traduzione automatica pone altri problemi, come per esempio il fatto che a volte due lingue usano costruzioni diverse per esprimere lo stesso concetto (come per esempio nel caso dell'espressione italiana "Giovanni ha 20 anni" che si traduce in inglese "John is 20 years old" cioè, letteralmente, "Giovanni è 20 anni vecchio"), il fatto che ci sono delle diversità di stile ecc.. Ma questi sono problemi minori. In genere, come avviene nell'esempio che abbiamo appena fatto, l'essere umano che legge il testo di uscita capisce il senso anche se sono presenti errori derivati dal fatto che il programma non riesce a risolvere problemi di questo genere. Ciò che è veramente importante, a mio avviso, è ottenere un testo di uscita *comprensibile e senza distorsioni*, perché un utente che non conosce affatto o conosce poco una certa lingua possa comprendere un testo scritto in quella lingua¹⁷.

Riassumendo, le particolarità che rendono il ritrovato per realizzare un programma di traduzione automatica che abbiamo descritto diverso dagli altri sono le seguenti.

1) Il ritrovato è basato sulla teoria correlazionale del pensiero, che è una teoria (anche) linguistica radicalmente nuova e profondamente diversa dalle altre teorie linguistiche.

2) Basandosi su questa teoria, il programma simula, anche se non riproduce, una parte di ciò che effettivamente fa l'essere umano con la sua intelligenza quando traduce.

3) Il ritrovato è probabilmente l'unico basato completamente ed esclusivamente su un'unica teoria linguistica.

4) Il ritrovato è probabilmente l'unico ideato da chi ha proposto la teoria linguistica su cui il ritrovato stesso è basato.

5) Il numero dei modi in cui si possono combinare, secondo la teoria correlazionale del pensiero, le parole che formano una frase è un numero finito e non particolarmente elevato (esso può essere calcolato matematicamente partendo dal numero delle parole che compongono la frase e dal numero dei correlatori espliciti in esso presenti). Il programma produce tutte le combinazioni possibili (per poi esaminarle successivamente), quindi viene sicuramente prodotta anche la combinazione corretta.

Qualora il programma, giunto alla fine della elaborazione che esso compie, non sia in grado di scegliere fra due o più strutture teoricamente possibili, esso può segnalare questo fatto all'utilizzatore in modo che possa essere egli ad operare questa scelta (con ogni probabilità è raro che si realizzi il caso che rende questa scelta impossibile, cioè che fra due o più strutture teoricamente possibili ve ne siano più di una dotata di senso).

Se la realizzazione pratica di un programma di traduzione automatica che ricostruisca autonomamente la rete correlazionale del pensiero e che sia dotato di un vocabolario abbastanza ampio comporta un la-

¹⁷ A mio avviso, l'uso dei programmi di traduzione automatica per rendere più veloce la traduzione quando è disponibile un traduttore professionista oggi è meno interessante che in passato, perché gli eccellenti risultati che danno i programmi di riconoscimento vocale permettono a un traduttore professionista di scrivere alla velocità con cui parla (gli errori sono molto pochi e quindi la correzione è molto rapida). In questo modo è possibile ottenere la traduzione di un testo in un tempo che non è molto diverso da quello che occorre per far correggere da un traduttore professionista la traduzione automatica dello stesso testo.

voro umano enorme, una sperimentazione di primo livello per verificare la sua efficacia è, al contrario, piuttosto semplice. Infatti, è possibile procedere nel seguente modo.

1) Inizialmente dotare il programma di un vocabolario molto piccolo, di poche decine di termini. Anche con così pochi termini il numero delle frasi che si possono costruire è grande ed è possibile andare a verificare se il programma non commette quei tipi di errori che invece sono comuni con i programmi attualmente disponibili.

2) Si può costruire solo il modulo di generazione, cioè la parte del programma che parte dalla rete correlazionale ricostruita per generare il testo nella lingua di uscita. Realizzare questa parte del programma invece che quella che esegue la ricostruzione della rete correlazionale è enormemente più facile e non è dipendente dalla ampiezza del vocabolario utilizzato.

3) È anche possibile eseguire delle verifiche della efficacia del programma “sulla carta”, cioè prima di aver realizzato praticamente il programma. Quando si è stabilito quali classificazioni usare per formare la sfera nazionale ed esse sono state applicate anche a poche decine di termini del vocabolario, è possibile andare a verificare per così dire “manualmente” come si comporterà il programma nella traduzione di una certa frase.

Bibliografia essenziale

- Benedetti G. (2005) *Una presentazione della Metodologia Operativa*, in www.mind-consciousness-language.com di G. Marchetti.
- Benedetti G. (2005) *Le operazioni mentali di base che costituiscono le categorie mentali*, in www.mind-consciousness-language.com di G. Marchetti.
- Ceccato S. (a cura di) (1969) *Corso di linguistica operativa*, Milano, Longanesi.
- Ceccato, S. Zonta B. (1980) *Linguaggio consapevolezza pensiero*, Feltrinelli, Milano.
- Glaserfeld, E. von, Pisani, P.P. (1970), *The multistore parser for hierarchical syntactic structures*. Communications of the ACM 13(2), 74-82.
- Hutchins W. J., (1986) *Machine translation: past, present, future*, (Ellis Horwood Series in Computers and their Applications.) Chichester, Ellis Horwood.
- Hutchins W. J. (con Somers H. L.), (1992) *An introduction to machine translation*, Academic Press, London.
- Hutchins W. J., 'Retrospect and prospect in computer-based translation'. In Machine Translation Summit VII, 13th-17th September 1999, Kent Ridge Digital Labs, Singapore. Proceedings of MT Summit VII "MT in the great translation era", 30-34. [Tokyo]: Asia-Pacific Association for Machine Translation, 1999.
- Hutchins W. J., 'Towards a new vision for MT'. Introductory speech at the 'MT Summit VIII' conference, 18-22 September 2001, Santiago de Compostela, Galicia, Spain.
- Hutchins W. J., 'Machine translation and human translation: in competition or in complementation?'. *International Journal of Translation*, vol.13, no.1-2, Jan-Dec 2001, pp. 5-20. Special theme issue on machine translation, [editor] Michael S. Blekhman. Anche in: *Machine translation theory & practice*, edited by Michael S. Blekhman. New Delhi: Bahri Publications, 2001. (BP Series in Translation Studies, 8).
- Hutchins W. J., 'Machine translation today and tomorrow'. In Computerlinguistik: was geht, was kommt? Festschrift für Winfried Lenders, hrsg. Gerd Willée, Bernhard Schröder, Hans-Christian Schmitz. Sankt Augustin: Gardez! Verlag, 2002, pp.159-162.
- Hutchins W. J., 'Has machine translation improved?'. MT Summit IX: proceedings of the Ninth Machine Translation Summit, New Orleans, USA, September 23-27, 2003, 181-188. [East Stroudsburg, PA: AMTA.]