

# *La ricerca in Linguistica Computazionale tra modelli formali ed analisi empirica*

Giacomo Ferrari  
Università del Piemonte Orientale  
Dipartimento di Studi Umanistici  
e-mail: [ferrari@lett.unipmn.it](mailto:ferrari@lett.unipmn.it)

Apparirà negli "Atti del Convegno in onore di T.Bolelli", Pisa, ETS

## **1. Ma che cosa è la linguistica computazionale?**

Il termine Linguistica Computazionale è una traduzione non eccessivamente raffinata, piuttosto una trasposizione, dall'inglese "Computational Linguistics". Con esso si designa una branca di studi e di applicazioni che, nonostante si sia sviluppata su un arco di più di 50 anni<sup>1</sup>, allineando numerosi successi sia teorici che pratici, suscita ancora oggi incertezze, perplessità, curiosità, fondate sostanzialmente sull'ignoranza dei suoi obiettivi, delle sue metodologie e, in una parola, della sua specificità.

In un certo senso si tratta di un'ignoranza di secondo ordine. Se, infatti, la stessa linguistica non gode di un'immagine pubblica prominente tra i comuni cittadini<sup>2</sup>, per la linguistica computazionale il problema d'informazione si pone fino in ambito accademico, dove si oscilla tra chi vede nei linguisti computazionali degli esperti di programmi che abbiano qualcosa a che fare con la scrittura di testi (una sorta di super-utenti di Word), e chi invece, specialmente informatici ed ingegneri, vede in essi dei fastidiosi filologi che tentano di piegare il software disponibile sul mercato ad utilizzazioni bizzarre, se non addirittura distorte e, soprattutto, inutili.

Non posso pretendere di avere la soluzione per ridurre e prevenire l'eccesso di semplicismo e la disinformazione che circonda questa disciplina, ma

---

<sup>1</sup> Non ho informazioni sulla prima attestazione del termine "computational linguistics", tuttavia si può ricordare la fondazione della prima Association for Machine Translation and Computational Linguistics (AMTCL), nel 1962, divenuta nel 1968 Association for Computational Linguistics (ACL). Tuttavia nel 1954 era stato fondato il giornale "Mechanical Translation and Computational Linguistics", assorbito nel 1965 dall'AMTCL.

<sup>2</sup> Tanto da indurre M.Berretta a chiedersi, nella circolare 166 apparsa sul bollettino SLI del 1999, "che rispondete voi a un'anima semplice che vi chieda a che cosa serve la linguistica?".

voglio tentare di dare alcune risposte che smontino almeno in parte i principali preconcetti che ancora oggi sono ampiamente diffusi, e, ripeto, non solo tra anime semplici, a proposito di una materia che ha ormai la maturità necessaria per entrare a far parte del patrimonio comune di conoscenze di quanti si occupano in modo scientifico del linguaggio umano. In particolare vorrei mettere a fuoco due punti che ritengo fondamentali; la linguistica computazionale ha fondamenti epistemologici e metodologici illustri e costituisce una branca della linguistica a pieno titolo.

E' certamente vero che essa istituisce un collegamento tra *una qualche conoscenza linguistica* e il computer. Il problema è proprio "quale tipo di rapporto". E' proprio la complessità del rapporto tra studio del linguaggio e computers, tra teoria linguistica e calcolatori che segna la storia della disciplina dal suo nascere ai giorni nostri. L'Association for Computational Linguistics<sup>3</sup> definisce oggi questo rapporto nei termini seguenti

Simply put, computational linguistics is the scientific study of language from a computational perspective. Computational linguists are interested in providing computational models of various kinds of linguistic phenomena. These models may be "knowledge-based" ("hand-crafted") or "data-driven" ("statistical" or "empirical").

Work in computational linguistics is in some cases motivated from a scientific perspective in that one is trying to provide a computational explanation for a particular linguistic or psycholinguistic phenomenon; and in other cases the motivation may be more purely technological in that one wants to provide a working component of a speech or natural language system. Indeed, the work of computational linguists is incorporated into many working systems today, including speech recognition systems, text-to-speech synthesizers, automated voice response systems, web search engines, text editors, language instruction materials, to name just a few.<sup>4</sup>

Questa definizione mette ben in evidenza sia i principi metodologici della disciplina che la sua doppia anima teorica ed applicativa. Ma se cerchiamo i fondamenti epistemologici di questa affermazione, troveremo che affondano le radici in discipline formali tradizionali, piuttosto che nella semplice pratica del computer.

Quando si parla di teorie linguistiche si possono fare spesso richiami ad altre discipline. La teoria della genealogia delle lingue, proposta da August Schleicher, è stata certamente influenzata dalle scienze naturali e dall'evoluzionismo diffuso, spesso con proprio rischio, da Charles Darwin, come prova l'opera che lo stesso Schleicher pubblica nel 1863, *Die*

---

<sup>3</sup> Prestigiosa associazione scientifica, divenuta nel 1983 l'organismo internazionale che raccoglie tutti quanti praticano questa disciplina; cfr. nota 1.

<sup>4</sup> La definizione è data come "manifesto" dell'associazione nel sito [www.aclweb.org](http://www.aclweb.org), sotto l'intestazione "What is Computational Linguistics?".

*Darwinsche Theorie und die Sprachwissenschaft*. Non si può escludere che lo stesso pensiero di Saussure si sia formato nel clima intellettuale da cui trasse le mosse il pensiero del sociologo Durkheim.

Non ha senso escludere, quindi, che anche la matematica, un certo tipo di matematica, abbia avuto un suo peso ed una sua influenza sul pensiero linguistico, specialmente in un periodo che vede diffondersi l'interesse per l'uso di codici e per i meccanismi formali della comunicazione. Ci accorgeremo, allora, che l'evoluzione computazionale della linguistica non è il risultato di occasionali ingegnerizzazioni dovute all'invenzione dei calcolatori, ma eredita un'elaborazione di numerosi decenni di interazione tra matematici, biologi e linguisti, di cui l'uso del calcolatore è solo un coronamento.

L'ingegneria linguistica esiste, ed è molto di moda, ma la sua collocazione è ben riflessa nella distinzione che, anni fa, compiva lo stesso "garante" della disciplina, l'Association for Computational Linguistics, tra Computational Linguistics e Applied Computational Linguistics. Essa sta alla Linguistica Computazionale, come l'ingegneria civile sta ai principi fisici che regolano l'equilibrio statico dei corpi.

Voglio, quindi, tentare di illustrare questo aspetto poco noto delle interazioni epistemologiche della linguistica con alcuni rami della matematica che finiscono col collegare la linguistica computazionale con i fondamenti stessi dell'informatica, tanto da farne una scienza parallela, piuttosto che una delle sue applicazioni.

## **2. Linguistica e raccolta di dati empirici**

### **2.1. Herdan e il paradigma statistico**

La linguistica ha, per tradizione, rivendicato la sua assoluta e totale dipendenza dai dati empirici, che ne costituiscono l'unica sorgente di informazione valida.

Che questi dati siano riducibili a strutture statistiche è messo in evidenza da uno dei fondatori della Statistica Linguistica, Gustav Herdan, all'inizio del suo libro *Language as Choice and Chance*, 1956, non soltanto un manuale di statistica per linguisti, ma una specie di paradigma epistemologico statistico per la linguistica. Herdan scrive che

..linguistics is not possible without some degree of statistical classification  
(Herdan, 1965:1)

proseguendo

How else than by patiently recording and classifying linguistic forms according to their function can we imagine that the grammar of a language was extracted from the actual speech occurrences? (Herdan, 1965:1)

Secondo Herdan la stessa legge di Grimm

Could have revealed itself to the investigator's mind only after a long and patient collection of data, and thus by a method which we usually describe as statistical (Herdan, 1965:1)

L'uso della statistica nello studio della lingua era stata materia di esercizio occasionale fin dalla metà dell'800, come mostrano gli studi di Förstemann sulle frequenze delle vocali in greco, latino, tedesco e sanscrito pubblicati tra il 1852 e il 1853, o il libro di Mariotti del 1880 su Dante. Ancora nel 1954, però, Guiraud nella sua *Bibliographie critique de la statistique linguistique* considera che la statistica linguistica sia

...une discipline récente et encore en train de se définir... (Guiraud, 1954:1)

e considera che tra i linguisti

Si la plupart...reconnaissent que c'est à partir de laborieuses compilations qu'on a pu reconstituer l'essentiel de la métrique ou de la syntaxe classiques, ils y voient en même temps un travail purement descriptif, sujet de dissertation ou de diplômes...la méthode quantitative a échoué devant des synthèses plus ambitieuses (Guiraud, 1954:3)

Al contrario, Herdan lo considera un vero e proprio paradigma di ricerca, creando un parallelismo teorico tra l'approccio statistico e la linguistica Saussuriana

If we regard language to be the total of word-engrams ('la langue' in de Saussure's terminology) *plus* their probability of occurrence in individual speech, and thus as the various ways in which the event 'engram' could happen, together with the relative frequencies of the different engrams in actual use, the conception fulfils all the requirements of what is known as the statistical population of such events, or their statistical univers. Any individual speech utterance ('la parole' in de Saussure's terminology) plays then the role of a sample from that population (Herdan, 1956:3)

e finisce poi per gettare un collegamento con la teoria dell'informazione

The statistical conception leads to a deeper understanding of established relations in linguistics, for instance of de Saussure's dichotomy of 'la langue' and 'la parole', and in general, of language as a coding system for ideas and

all that this implies, and thus prepares the ground for what is known to-day as Information Theory (Herdan, 1956:3).

Dunque, quello statistico non è soltanto un metodo di elaborazione dei dati linguistici, ma si trasforma nell'interpretazione di Herdan, in un vero paradigma scientifico che si affianca ed integra gli stessi fondamenti della linguistica.

Il problema che resta aperto è come si reperiscono, si predispongono e si accedono i materiali empirici. Anche su questo punto troviamo precedenti illustri nelle varie tradizioni lessicografiche.

## 2.2 Lessici e concordanze

I dizionari sono un patrimonio della conoscenza linguistica degli uomini fin dai tempi in cui i mercanti Sumeri, Assiri e Babilonesi decisero di dotarsi di repertori bilingui che servissero di aiuto, probabilmente, nelle trattative commerciali. Attraverso i millenni la tradizione lessicografica si è evoluta fino a raggiungere la fisionomia che ha adesso. Uno dei punti cardine di questa tradizione è l'identificazione delle unità lessicali e dei loro contesti di uso, che sono la fonte da cui scaturiscono poi le spiegazioni assegnate dal dizionario ad ogni parola, i contesti particolari di uso, le locuzioni ed ogni altra informazione che forma il cosiddetto articolo. Lo strumento base è, quindi, la *scheda lessicografica*, la cui realizzazione computazionale è la *concordanza*, cioè la lista dei contesti in cui una data parola appare.

Ma anche le concordanze giungono da lontano; l'idea di costruire indici di parole collegandole ai luoghi di citazione nasce in collegamento con la tradizione biblica.

Si parla di questo genere di strumenti di consultazione già con le tabelle dei Masoreti elaborate nel X secolo in Tiberiade, sia pure solo per scopi di corretta traduzione. Il lavoro continuò attraverso i secoli, segnando tappe importanti come l'*index locorum* prodotto nel 1230 da Hugo de Sancto Claro, o le concordanze della Bibbia ebraica di Isaac Nathan finite nel 1448, fino all'enorme "Complete Concordance to the Old & New Testaments" del 1738, di Alexander Cruden.

Non è forse un caso se il primo a proporre di utilizzare il calcolatore per fare un lavoro in tutto comparabile con le concordanze bibliche, l'*Index Thomisticus*, sia stato proprio un religioso, Padre Roberto Busa, un gesuita che inventò le concordanze elettroniche nel 1949. In un articolo del 1974 su *Civiltà Cattolica*, motiva così la scelta del computer:

Il *computer*, alla stregua di ogni altra macchina, può diminuire drasticamente l'impiego di lavoro umano non qualificato; ma aumenta l'esigenza sia del

lavoro organizzato sia del lavoro qualificato. Impegna l'uomo sia a lavorare in équipes più disciplinate che mai, sia a condensare in minor tempo una maggior quantità di lavoro ai livelli superiori dell'intelligenza, quali sono l'organizzare e il programmare: allargandone la presa su zone di realtà molto più estese. Con il *computer* l'uomo si è attrezzato a dominare più cose in minor tempo (Busa, 1974: 255).

Aggiungendo, come chiosa, il fatto che

il *computer* permette di introdurre ed elaborare più informazioni, con più severa sistematicità e con più profondità e varietà di calcoli (Busa, 1974: 255).

E' interessante osservare che Padre Busa si riconnette con la statistica linguistica, riconoscendole un valore di scienza umana:

... la scoperta dei rapporti quantitativi interni al linguaggio, ha tanto valore in sede di scienza quanto ne ha qualsiasi studio sull'uomo. E la linguistica - è certamente non l'ultima tra le discipline umanistiche. Perciò "contare le parole" non è proprio sempre e in assoluto un perditempo .... una panoramica quantitativa di una situazione linguistica ha sempre i contorni sfumati e slabbrati dei dati calcolati probabilisticamente, come li ha una carta delle correnti marine o un diagramma meteorologico (Busa, 1974: 253).

Da qui si diffonde per il mondo una quantità di iniziative di costruzione di concordanze come base di riferimento per una serie di studi empirici, da cui la statistica non è del tutto aliena. I dizionari fondamentali degli anni '60 e '70 (Bortolini, Tagliavini & Zampolli 1971, Juilland 1964, 1965, 1973) rappresentano il primo conseguimento di questa nuova sintesi tra la tradizione lessicografica e il paradigma statistico, in cui il computer, come ben sottolinea Padre Busa, è soltanto un utensile.

Naturalmente anche questo, come tutti gli utensili tecnologici, si evolve con il tempo ed offre sempre più servizi. Le concordanze non sono più necessarie.

Le informazioni che vogliamo associare a ciascuna unità linguistica, parola, sintagma o segmento testuale che sia, possono essere inserite nel testo stesso e memorizzate grazie all'uso di linguaggi universali di rappresentazione, i cosiddetti linguaggi di mark-up come SGML e XML<sup>5</sup>. I testi possono essere

---

<sup>5</sup> I linguaggi di mark-up servono ad inserire in un testo indicazioni relative alla modalità secondo cui un browser (es. Internet Explorer) deve interpretare il segmento di testo annotato. Così, se marco la parola "cane" come <italics>cane</italics>, questa verrà interpretata come *cane*. Estendendo l'uso di queste note (tags) si possono aggiungere annotazioni di tipo linguistico. Ad es.: <w type="nome" font="italics">cane</w>, ci

acquisiti con tecnologie più veloci (scanner, riutilizzo del formato elettronico utilizzato dagli editori ecc.). Si passa, quindi, dai tradizionali lessici accompagnati dalle concordanze, alla costituzione delle cosiddette *risorse linguistiche*. Si tratta di "depositi" di informazione linguistica utilizzabili per scopi realmente ingegneristici: grandi dizionari in formato elettronico, banche di dati acustici, fonetici e fonologici, e, tra questi, i *corpora*, predisposti in modo da poterne estrarre l'informazione desiderata "on demand", cioè quando se ne ha bisogno. L'idea è che questi repertori siano accessibili per offrire conoscenze e informazioni da utilizzarsi nella costruzione di prodotti industriali che abbiano alla base il trattamento del linguaggio naturale<sup>6</sup>.

Naturalmente l'evoluzione delle metodologie informatiche ha comportato anche l'evoluzione delle metodologie di elaborazione. Agli albori lo sforzo necessario per acquisire i campioni (attraverso l'uso di schede perforate), definire i formati di memorizzazione ed i programmi di elaborazione era enorme, soprattutto perché la fase di codifica (dei caratteri, dei riferimenti interni al testo, dell'unità "parola" ecc.) doveva essere svolta preliminarmente. Oggi l'acquisizione dei campioni è relativamente più semplice e, in ogni caso, non occorre dedicare imponenti risorse, dal momento che l'attività di codifica avviene successivamente all'acquisizione. L'attenzione è maggiormente diretta verso gli standard di codifica<sup>7</sup>.

### **3. I modelli del pensiero**

#### **3.1 Pensiero e computazione**

Nel 1936, il matematico inglese Alan Turing propose il modello di automa universale, noto come "automa" o "macchina di Turing". E' interessante che inizi la presentazione di questa astrazione teorica, universalmente considerata l'archetipo di tutti i computer, con le parole

---

informa che la sequenza "cane" è una parola di tipo "nome" e carattere corsivo. Su questo argomento si veda Ferrari et al. (2002) e Ferrari (2003).

<sup>6</sup> Per avere un'idea di un corpus in formato elettronico e dell'uso che se ne può fare si possono vedere alcuni siti come quello del British National Corpus ([www.natcorp.ox.ac.uk](http://www.natcorp.ox.ac.uk)), del Cobuild ([www.cobuild.collins.co.uk](http://www.cobuild.collins.co.uk)), il corpus del dizionario Collins e della Text Encoding Initiative ([www.tei-c.org](http://www.tei-c.org)), nonché numerosi altri siti connessi.

<sup>7</sup> La comparabilità dei corpora è sempre stata affidata all'uso di formati standard. Alle origini era già problematico affrontare le prime fasi di acquisizione del testo, per cui la discussione sugli standard era spesso la semplice espressione di un desiderio lontano dall'essere realizzato. Oggi, invece, l'attenzione per gli standard coinvolge anche l'uso di particolari annotazioni linguistiche. Si veda per questo tutto il materiale EAGLES (Expert Advisory Group on Language Engineering Standards, cfr. [www.ilc.cnr.it/EAGLES96/home.html](http://www.ilc.cnr.it/EAGLES96/home.html)).

We may compare a man in the process of computing a real number to a machine which is only capable of a finite number of conditions  $q_1, q_2, \dots, q_R$  which will be called " $m$ -configurations" (Turing, 1936: 231)

Più tardi, nel 1950, inizierà un suo altrettanto fondamentale articolo con le parole

I PROPOSE to consider the question, 'Can machines think?' This should begin with definitions of the meaning of the terms 'machine' and 'think'... I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words. The new form of the problem can be described in terms of a game which we call the 'imitation game'. It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. ...We now ask the question, 'What will happen when a machine takes the part of A in this game?' Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, 'Can machines think?' (Turing, 1950: 433-434).

Nasce con queste parole il famosissimo "test di Turing", considerato la discriminante per valutare il grado di "intelligenza" di un certo programma. A noi interessa soprattutto il metodo, tipicamente matematico, di sostituire un problema con un altro di ugual complessità, ma articolato in maniera più formale. Ad ogni modo, al di là degli sviluppi matematici del pensiero di Turing è importante osservare che l'idea stessa di computazione è derivata da un parallelismo con l'intelligenza umana. Il calcolatore è il risultato di un tentativo di costruire un modello formale del ragionamento umano. Non è certo un'idea nuova neanche questa, ma risale agli illuministi.

L'ipotesi di Turing si appoggia ad un altro supporto teorico detto tesi di Church-Turing (Turing 1936, Church 1936). I due matematici formularono una tesi sul processo di discriminazione dei numeri in due insiemi che è divenuta il fondamento dell'ottimismo razionalista dell'epoca computazionale. La si può formulare in termini semplici nel modo seguente:

“Supponiamo che ci sia un metodo che un essere intelligente segue per dividere i numeri in due classi. Supponiamo che questo metodo produca sempre una risposta entro una quantità finita di tempo e che per un dato numero dia sempre una risposta. Allora: esiste una funzione generalmente ricorsiva che dà esattamente la stessa risposta che dà il metodo dell'essere intelligente”

Questa tesi stabilisce il fondamento dell'equivalenza tra pensiero umano e funzioni ricorsive, che sono l'essenza della computazione.

I calcolatori, in quanto oggetti elettronici, derivano da un'altra tradizione, quella che risale alla "macchina delle differenze", concepita da Ch. Babbage nel 1819, che calcolava le tavole logaritmiche, all'invenzione da parte di Lady Ada Byron, moglie di Ch. Babbage, dei fondamenti della programmazione, ed infine, all'interpretazione che il matematico americano di origine ungherese John von Neumann<sup>8</sup> seppe dare del primo mostruoso calcolatore elettronico, l'ENIAC, disegnato per fare i complessi calcoli necessari all'elaborazione della prima bomba atomica<sup>9</sup>.

Dunque, l'idea di utilizzare i calcolatori per simulare in qualche modo i meccanismi del pensiero umano non nasce con i calcolatori in sé, ma è il risultato del confluire tra l'ipotesi di Church-Turing e il nascere di macchine adatte a compiere complessi calcoli in poco tempo.

### 3.2. Linguistica e Computazione

Se trasferiamo questa impostazione alla linguistica dovremmo formulare qualche principio che, di fatto, sta alla base di un'altra branca della linguistica computazionale, quella che in Ferrari (2000: 16) abbiamo definito modellistico.

Il principio fondamentale è dato da una ripresa ed estensione della tesi di Church-Turing, che potremmo formulare come segue:

“Se l'attività linguistica umana è riducibile in termini di una serie di processi di discriminazione; Se questi processi sono simili a quelli che gli stessi esseri parlanti/intelligenti usano per distinguere i numeri in due classi (cioè: se questi processi sono ricorsivamente enumerabili e ricorsivi) allora esiste una funzione ricorsiva che dà esattamente la stessa risposta che dà il metodo dell' (adotta lo stesso comportamento dell') essere intelligente”.

Il primo a ridurre un problema linguistico ad un problema di discriminazione fu Noam Chomsky. Al di là delle interpretazioni che si possono dare della grammatica generativa, essa è un meccanismo di discriminazione tra frasi appartenenti o non appartenenti ad una lingua. Al di là del modello di rappresentazione delle grammatiche *context-free*, resta vero ancora oggi che

---

<sup>8</sup> Cfr. Heims (1980).

<sup>9</sup> Un'interessante ricostruzione delle ideologie e delle impostazioni teoriche di questo periodo si trova in Castelfranchi & Stock (2000).

Parsing algorithms for context-free grammars...play a large role in the implementation of compilers and interpreters for programming languages and of programs which "understand" or translate natural languages (Earley, 1969: 94).

Questo spiega perché l'analisi sintattica automatica sia stata il primo prodotto della linguistica computazionale e quello che, forse, ha dato i risultati migliori

Questa non è e non può essere un argomento a sostegno della linguistica chomskiana. Infatti, quando si scatenarono le critiche più aspre al modello trasformazionale, i computazionalisti fecero anch'essi la loro parte.

Dopo l'uscita di *Aspects of a Theory of Syntax* (Chomsky, 1965) i cosiddetti lessicalisti, mostrarono le difficoltà della "lexical insertion transformation" e l'influenza del lessico sulla sintassi<sup>10</sup>, mentre i matematici mostrarono che la grammatica trasformazionale produce un linguaggio ricorsivo, ma indecidibile e quindi poco utile a caratterizzare il linguaggio naturale<sup>11</sup>. Kaplan (1972) evidenziò che un modello computazionale (l'Augmented Transition Network<sup>12</sup>), trasformava le frasi in input direttamente in termini di struttura profonda, senza passare attraverso le trasformazioni, e si rivelava, su base sperimentale, perfettamente adeguato psicologicamente, se non migliore del modello generativo trasformazionale.

Anzi, fu dalla linguistica computazionale che nacque quel movimento "anti-chomskiano" che dette origine ad una famiglia intera di grammatiche formali alternative come la Lexical-Functional Grammar<sup>13</sup>, la Generalized Phrase-Structure Grammar<sup>14</sup> e molte altre.

Chomsky costituisce soltanto un richiamo allo schema "operativo" del metodo computazionale. Per costruire un modello computazionale occorre disporre di un intermediario che trasforma, al minimo, un modello linguistico in un modello di discriminazione alla Church-Turing. La teoria dei linguaggi formali di Chomsky funzionò come primo intermediario, anche se successivamente gli studi portarono ad identificarne i limiti e i possibili miglioramenti sul piano computazionale e formale.

Per alleggerire il peso dell'argomentazione matematica, useremo i termini di un altro padre dell'Intelligenza Artificiale, però di matrice cognitiva, Terry

---

<sup>10</sup> Cfr. ad es. Gruber (1976).

<sup>11</sup> Cfr. Peters & Ritchie (1973).

<sup>12</sup> Ideato da W.Woods, cfr. Woods (1970).

<sup>13</sup> Proposta da Kaplan & Bresnan, (1982).

<sup>14</sup> Cfr. Gazdar et al. (1985).

Winograd<sup>15</sup>. Egli scrive, nella prima pagina del suo *Language as a Cognitive process* del 1983:

In looking at language as a cognitive process, we deal with issues that have been the focus of linguistic study for many years....But we look at language from a different perspective. The emphasis is less on the structure of the sounds or sentences, and more on the structure of the processes by which they are produced and understood.....In the forty years since digital computers were first developed, people have programmed them to perform many activities that we think of as requiring some form of intelligence. In doing this, they have developed new ways of talking about knowledge - what it is, and how it can be stored, modified, and used (Winograd, 1983: 1).

Più avanti fornisce una spiegazione più dettagliata della relazione che può esservi tra un programma e la teoria linguistica:

The computer shares with the human mind the ability to manipulate symbols and carry out complex processes that include making decisions on the basis of stored knowledge. Unlike the human mind, the computer's workings are completely open to inspection and study, and we can experiment by building programs and knowledge bases to our specifications. Theoretical concepts of program and data can form the basis for building precise computational models of mental processing. We can try to explain the regularities among linguistic structures as a consequence of the computations underlying them (Winograd, 1983: 13).

Terry Winograd getta un ponte tra l'impostazione algoritmica ed i modelli cognitivi, conferendo validità di modello teorico ai programmi che implementano qualche funzione della mente umana. Per decenni la storia della linguistica computazionale ha visto molti contributi in questo senso.

In questa prospettiva, analizzare sintatticamente una frase non esaurisce il compito della linguistica computazionale; alle frasi occorre anche assegnare un'interpretazione semantica ed una pragmatica.

Tuttavia, chi ha sviluppato programmi di interpretazione semantica, che dovevano tradurre le frasi in un formato "comprensibile" per il calcolatore, non ha trovata nella linguistica i modelli adeguati e si è rivolta alla psicologia, derivandone il sistema di rappresentazione detto reti semantiche<sup>16</sup> e quello detto dei frames<sup>17</sup>, o alla logica quantificata del primo

---

<sup>15</sup> Terry Winograd ha adottato un approccio meno matematico e più psicologico-cognitivo. Nel 1970 implementò un sistema che permetteva di muovere oggetti su un video mediante comandi espressi in linguaggio naturale. Il modello sottostante era quello della pianificazione (cfr. nota 46). Un particolare sorprendente è che l'analisi sintattica veniva compiuta secondo un modello funzionale alla Halliday, unico caso in un mondo dominato dalle grammatiche context-free.

<sup>16</sup> Nella versione proposta in Ross-Quillian (1968).

ordine, attraverso numerose sue ipostasi, fino a giungere alla Discourse Representation Theory<sup>18</sup>, oggi giorno una specie di standard. Nella pragmatica dell'interazione ci si è rivolti alla teoria degli atti linguistici di Searle, a sua volta interpretandoli secondo il modello formale della pianificazione, una tecnica di soluzione automatica di problemi<sup>19</sup>.

#### 4. Teoria o ingegneria?

Perché, con fondamenti teorici così illustri, si continua a considerare la Linguistica Computazionale come un qualcosa di intermedio tra vaghe conoscenze linguistiche, applicazioni informatiche ed ingegneristiche? Per certi aspetti è colpa dei linguisti computazionali stessi.

Se ci riconduciamo all'*ottimismo computazionale* degli anni 60 e 70, i più grandi programmi che riflettono modelli del linguaggio umano sono stati portati a termine presso dipartimenti di informatica, proprio a causa di questa felice sinergia tra teorie formali del linguaggio e inizi del calcolo simbolico, primordi dell'informatica moderna. Ma l'informatica stessa, in quanto sviluppatrice di strumenti di calcolo, era a sua volta in cerca di motivazioni applicative. La capacità di calcolo non bastava più, soprattutto se si doveva giustificare l'uso di calcolo simbolico.

Già l'idea dell'utilizzo per tradurre da una lingua all'altra era nata alla fine degli anni 40. Nel 1949, uno dei padri della Teoria della Comunicazione, W. Weaver, coinvolto nel "code cracking", un'attività su cui si è fondato gran parte del lavoro dei servizi segreti durante la seconda guerra mondiale, scriveva in un suo appunto:

I have a text in front of me which is written in Russian but I am going to pretend that it is really written in English and that it has been coded in some strange symbols. All I need do is strip off the code in order to retrieve the information contained in the text<sup>20</sup>

Weaver propone, per questa attività di "trasduzione", l'uso di tecniche statistiche. Ben presto però ci si accorse che queste non erano sufficienti, ma occorreva mettere in gioco serie conoscenze linguistiche (formali).

Gli anni 60 vedono, poi, il nascere di un altro ottimismo: un espandersi illimitato dell'uso del computer su tutti i domini dell'attività umana, un computer reso accessibile anche ai più refrattari grazie alla possibilità di usare il linguaggio naturale. I realizzatori della prima interfaccia sperimentale (BASEBALL) scrivono

---

<sup>17</sup> Cfr. Minsky (1975).

<sup>18</sup> Nella formulazione di Kamp & Reyle (1993).

<sup>19</sup> Secondo le proposte di Cohen & Perrault (1979), Allen & Perrault (1980) e Allen (1983).

<sup>20</sup> L'appunto è stato poi pubblicato nel 1955 da Locke & Booth.

For convenience and speed, many future computer-centered systems will require men to communicate with computers in natural language. The business executive, the military commander, and the scientist need to ask questions of the computer in ordinary English, and to have the computer answer the questions directly (Green et al., 1961: 219).

Ed è a questo punto che i linguisti computazionali offrono il loro aiuto, ponendosi come depositari del modello, o dei modelli, per attuare questi disegni così ambiziosi, dai quali doveva dipendere il futuro dell'informatica stessa.

I linguisti computazionali hanno dovuto, però, “stare al passo” con i nuovi obiettivi applicativi, sostituendosi spesso a quegli ingegneri che continuano a fare programmi senza molta competenza e sensibilità per il linguaggio.

Questo ha avuto il benefico effetto che i linguisti si sono trovati schierati con una parte che non era loro propria, anche se potremmo dire, brutalmente, redditizia in termini di finanziamenti<sup>21</sup>. E' stato, quindi, necessario rivestire le ricerche di base di un ampio mantello di ottimismo pratico ed applicativo. In una tavola rotonda tenutasi nel 1980 a cura della Association of Computational Linguistics, Jerry Hobbs affermava

Progress on natural language interfaces can perhaps be stimulated or directed by imagining the ideal natural language system of the future.<sup>22</sup>

Orientamenti di questo genere, assai frequenti tra gli anni 70 ed 80, si fondano su un principio del quale è meglio non cercare di dimostrare la fondatezza, cioè che, considerando che l'uomo è il miglior "processore di linguaggio" che conosciamo, le soluzioni antropomorfe sono necessariamente le migliori. Si tratta di un argomento molto utilizzato, appunto negli anni 70, negli ambienti di Intelligenza Artificiale, argomento che non ha mancato di suscitare numerose ironie. Secondo questo ragionamento, infatti, le migliori automobili dovrebbero aver gambe e le migliori imbarcazioni dovrebbe essere spinte da pale.

Winograd stesso aveva ben sottolineato che

---

<sup>21</sup> Uno dei padri della Linguistica Computazionale, Aravind Joshi, disse, informalmente, verso la fine degli anni 80 che non esisteva nessuna teoria linguistica dietro la quale non si celava un costruttore di computer; alludeva a Lexical Functional Grammar, sviluppata a Xerox, Head Driven Phrase Structure Grammar, sviluppata presso Hewlett Packard, e, secondo lui, Government and Binding, cresciuto all'ombra di IBM.

<sup>22</sup> Intervento in una tavola rotonda tenutasi durante il convegno ACL nel 1980 (Proceedings p. 64).

The desire for more effective human-machine interaction has been one of the major motivations for extensive government support of research on natural language (Winograd, 1983: 24).

La situazione è magistralmente riassunta da W.Wahlster<sup>23</sup> durante una tavola rotonda tenutasi nel 1986 a Saarbrücken:

A commercial success of NL interfaces could have a great value to our field. When thousands of users appreciate the capabilities of NL interfaces, computational linguistics have something to be proud of. The economic interest might push forward the needed basic research and secure funding<sup>24</sup>

L'evoluzione verso il "prodotto" ha preso, però, un'altra strada, come mette in evidenza, durante la stessa tavola rotonda, Gary Hendrix<sup>25</sup>:

Note that Q&A is not sold as a natural language interface per se.....The key is to realize that users do not care about NL interfaces --they only want to get a job done...I think NLP will be widely used to help do sophisticated spelling and grammar checking in support of word processors. In the short term, this application of NLP may be more important commercially than NL interfaces.<sup>26</sup>

sottolineando, in parole povere, che il futuro dei prodotti di trattamento del linguaggio naturale è dei prodotti piccoli e poco complessi, piuttosto che di prodotti sofisticati come le interfacce in linguaggio naturale.

Certamente, senza lo sforzo necessario per costruire i primi modelli teorici ed i prototipi, oggi ci sarebbe ben poco da ingegnerizzare. Un esempio per tutti è dato dai sistemi di dialogo con cui ci abituiamo sempre di più ad interagire telefonicamente (servizi Telecom, Ferrovie, ecc.). Nel 1980, James Allen implementò un modello di dialogo fondato su un'interpretazione computazionale della teoria degli atti linguistici di Searle<sup>27</sup>. L'idea generale è che un atto illocutivo come il CHIEDERE è un'azione come qualunque altra azione fisica, analizzabile in termini di

---

<sup>23</sup> W.Wahlster è stato il fondatore ed è ancora oggi il Direttore del Deutsche Forschungszentrum für Künstliche Intelligenz di Saarbrücken, un enorme istituto nazionale per la ricerca in Intelligenza Artificiale, che dedica vastissime risorse al trattamento del linguaggio naturale, tra cui le interfacce multimediali e la traduzione automatica.

<sup>24</sup> Si tratta dell'intervento introduttivo alla tavola rotonda sulle interfacce in linguaggio naturale tenuta nell'ambito del Coling (International Conference on Computational Linguistics (Proceedings p. 161).

<sup>25</sup> G.Hendrix è noto per aver sviluppato il più "ingegneristico" dei sistemi di interfaccia in linguaggio naturale (LIFER cfr. Hendrix et al. 1978) e per aver fondato la prima azienda per la commercializzazione dell'interfaccia Q&A, un derivato di LIFER.

<sup>26</sup> In Coling86-Proceedings, p. 164-166.

<sup>27</sup> Vedi sopra § 3 e nota 35.

precondizioni (condizioni che devono essere vere perché l'azione possa aver luogo), corpo dell'azione ed effetti (condizioni che sono vere dopo l'esecuzione dell'azione). Questo modello è l'intermediario perfetto tra la teoria di Searle, a patto di interpretare come precondizioni le condizioni di felicità, o almeno una parte di esse, e come effetti, gli effetti perlocutivi, e la rappresentazione delle azioni in termini di planning, una tecnica di soluzione automatica di problemi applicata anche in robotica<sup>28</sup>. Ogni ascoltatore che interpreta una richiesta, la ricollega all'azione corrispondente e la interpreta come parte di una sequenza di azioni che costituiscono il piano del parlante. Le risposte vengono costruite a partire da questa interpretazione, piuttosto che dalla comprensione letterale della richiesta<sup>29</sup>.

A partire da questo complesso modello, si sono sviluppate implementazioni di complessità crescente, che facevano appello perfino a logiche di secondo ordine, notoriamente non computabili. Tuttavia, il modello fu adottato da chi ne fece implementazioni pratiche, con semplificazioni importanti, come rinominare gli atti linguistici utilizzando categorie fondate sull'utilizzo pratico che se ne doveva fare, riducendone il numero, e riducendo al minimo i processi logici di ragionamento.

I sistemi di dialogo attuali hanno ormai poco a che fare con quello di Allen, ma senza quell'elaborazione teorica non sarebbe stato possibile giungere alle realizzazioni attuali.

D'altra parte, lo stesso Allen insiste recentemente sull'utilità di proseguire le ricerche teoriche, proprio in un'ottica applicativa:

The belief that humans will be able to interact with computers in conversational speech has long been a favorite subject in science fiction....With recent improvements in computer technology and in speech and language processing, such systems are starting to appear feasible (Allen et al., 2001: 27).

respingendo anche altre pressioni:

Some people argue that spoken language interfaces will never be as effective as graphic user interfaces (GUIs) except in limited special case situations.....speech provides a worthwhile and natural additional modality....because it adds considerable flexibility and reduces the amount of training required....Dialogue-based interfaces allow the possibility of extended mixed-initiative interaction (Allen et al., 2001: 27).

---

<sup>28</sup> Per la quale vedi Sacerdoti (1977) e Chapman (1987).

<sup>29</sup> Per una spiegazione più dettagliata vedi Ferrari (1991: 144-159).

## 5. I compiti per il futuro prossimo e remoto

L'intelligenza umana non è fatta solo di procedure discriminatorie sicure e non ambigue come quelle previste da Church e Turing. In molti casi la discriminazione presenta scelte di buon senso, e il buon senso non è la stessa cosa che l'intelligenza.

Il buon senso, per quel che ne sappiamo, è la capacità di operare scelte in situazioni di incertezza, sulla base dell'esperienza, ricercando il male minore. Questo tipo di comportamento si chiama, in una visione algoritmica alla Church-Turing, *euristica*. L'euristica, come l'esperienza, si basa sull'acquisizione statistica di dati empirici, anche se qualche volta si tratta di una statistica accumulata nel patrimonio genetico.

Se poi guardiamo all'aspetto più squisitamente chomskiano, vale proprio l'osservazione iniziale di Herdan, che nessuna grammatica può essere estrapolata senza un'osservazione quantitativa dei fatti.

Queste osservazioni, che scaturiscono dalla pratica, hanno portato in tempi recenti ad un rinascere e rafforzarsi del paradigma statistico ed empirico propugnato da Padre Busa, sia pure con paradigmi più sofisticati, anche se la sostanza non cambia. Cambiano i mezzi informatici, si introduce l'uso di formati universali di scambio di documenti, l'acquisizione dei campioni è facilitata dal modo in cui è organizzata l'informazione in rete, i programmi che una volta andavano scritti espressamente per ogni singolo problema, sono ormai dotazione dei più semplici *word processors*.

Tra tutti, emerge un elemento nuovo che costituisce un compito per il futuro. Se usassimo l'enorme quantità di dati che ormai possiamo acquisire in poco tempo, rimarremmo, paradossalmente, indietro rispetto al processo meccanico. Ed è per questo che il prossimo obiettivo di ricerca è l'applicazione al linguaggio naturale dei metodi, già noti come sistemi di apprendimento automatico, in modo da produrre programmi che siano in grado di acquisire i campioni, elaborarli ed estrarre automaticamente le conoscenze linguistiche utili per i processi successivi. Si tratta di un compito per il futuro prossimo, perché esiste già molta elaborazione in materia; e certamente le ricadute sulla linguistica saranno importanti.

Ricadute fruttuose si possono avere anche nel caso della comunicazione multimodale, l'ultima evoluzione dei sistemi di dialogo. La comunicazione umana è di per sé multimodale, in quanto in essa è coinvolta la parola, il gesto, l'espressione, la postura e l'intero contesto percepibile. Le ricerche per raggiungere un buon livello di simulazione di questi canali e, soprattutto, di sincronizzazione procedono vigorosamente, ma porrei l'obiettivo nel lontano futuro.<sup>30</sup>

---

<sup>30</sup> Sugli ultimi sviluppi della linguistica computazionale si veda Ferrari (2004).

## 6. ...e allora?

Come definiremo il linguista computazionale? Certo, spero che dopo questa "galoppata" nella storia della disciplina e dei suoi fondamenti epistemologici, si possano correggere almeno tre delle "proposizioni" che si formulano abitualmente sulla sua professionalità:

1. "i linguisti computazionali maneggiano le loro macchinette per ordinare dati linguistici (contare le parole), ma la vera interpretazione non è possibile senza la sensibilità dello studioso"; questo è certamente vero, ma il solo ordinare dati linguistici meccanicamente implica lo sviluppo di approcci teorici formali che avrebbero un impatto importante sulla metodologia della linguistica *tout court*.
2. "sono più informatici che linguisti"; direi di più, hanno anticipato gli informatici stessi su molti punti e si avvalgono, in ogni caso, dell'applicazione del paradigma formale della computazione allo studio della lingua
3. "sono un po' ingegneri che trafficano con la lingua per produrre i pezzi necessari a sviluppare servizi informatici come i correttori automatici o i dizionari elettronici"; questo è falso, sono piuttosto coloro che producono le formalizzazioni necessarie a che gli ingegneri facciano quello che è descritto sopra.

Ma possiamo anche porci un'altra domanda: nella ricerca computazionale qual è, quindi, il ruolo della linguistica e che cosa ne ricava in cambio? La linguistica dovrebbe fornire dei paradigmi alle ricerche che vengono condotte in maniera troppo focalizzata. Ai primordi della disciplina la linguistica era in posizione di maggior avanzamento rispetto a quella computazionale, per cui il compito del computazionalista era quello di tradurre modelli già sufficientemente completi e sofisticati per renderli computabili. Oggi i bisogni dell'applicazione hanno preso il sopravvento, cosicché a fianco di risultati apparentemente validi, si ha una grande scarsità di paradigmi teorici, e cui la linguistica, che si è trovata in posizione di priorità alle origini della disciplina, si trova ora a non avere risposte a numerose domande come "quali sono le principali differenze tra scritto e parlato?", "dato un testo quali sono i tratti strutturali principali e quali quelli dipendenti dal genere?", "quali sono i parametri che ci permetterebbero di identificare la struttura di un dialogo?" ecc. La linguistica affronta da molto tempo queste tematiche, ma non ha ricercato risposte sistematiche. E questo non fa bene né alla linguistica né all'ingegneria.

Dobbiamo cambiare qualcosa nei nostri paradigmi? Ritengo di no, ritengo piuttosto che dobbiamo ritrovare i nostri paradigmi di ricerca per produrre nuovi modelli traducibili nelle procedure di Church-Turing.

Mi rendo conto che la mia è una concezione molto razionalista, ma trovo difficile immaginare una scienza che non lo sia, una scienza che non sia un ridurre ad ordine quello che è solo apparentemente disordinato ed ambiguo. E la linguistica, computazionale o no, non fa eccezione.

## **Bibliografia**

- Allen J.F., (1983), *Recognizing intentions from natural language utterances*, in Brady M. & Berwick, R.C., (1983, a cura di), pp. 107-166.
- Allen J. F. & Perrault C. R., (1980), *Analysing intention in utterances*, "Artificial Intelligence" XV, 3, pp. 143-178.
- Allen J.F., Byron D.K., Dzikovska M., Ferguson G., Galescu L., Stent A., (2001), *Toward Conversational Human-Computer Interaction*, in "AI Magazine", XXII ,4, pp. 27-37.
- Bortolini U., Tagliavini C. & Zampolli A., (1971), *Lessico di frequenza della lingua italiana contemporanea*, IBM Italia.
- Brady M. & Berwick, R.C., (1983, a cura di), *Computational Models of Discourse*, MIT Press, Cambridge MA.
- Bresnan J., (1982, a cura di), *The mental representation of grammatical relations*, MIT Press, Cambridge MA
- Busa R., (1974), *L'Index Thomisticus, contenuto, finalità, prospettive*, "Civiltà Cattolica", 2967, pp. 250-257.
- Castelfranchi Y. & O.Stock (2000), *Macchine come noi, la scommessa dell'intelligenza artificiale*, Laterza, Bari.
- Chapman J., (1987), *Planning for Conjunctive Goals*, in "Artificial Intelligence", XXXII, pp. 33-57.
- Chomsky N., (1965), *Aspects of the Theory of Syntax*, MIT Press, Cambridge MA.
- Church, A., (1936a), *An Unsolvability Problem of Elementary Number Theory*, in "American Journal of Mathematics", LVIII, pp. 345-363.

- Church, A., (1936b), *A Note on the Entscheidungsproblem*, in "Journal of Symbolic Logic", I, pp. 40-41.
- Cohen, P. R. & Perrault C. R., (1979), *Elements of a plan-based theory of speech acts*, in "Cognitive Science", III, pp. 117-212.
- Coling86: Proceedings of the International Conference on Computational Linguistics*,
- Earley J., (1970), *An efficient context-free parsing algorithm*, in "Communications of the ACM", XIII ,2, pp. 94 - 102.
- Ferrari G., (1991), *Introduzione al Natural Language Processing*, Calderini, Bologna, 1991.
- Ferrari G., (2000), *Livelli di Analisi del testo. Due approcci a confronto*, in Rossini Favretti R. (2000, a cura di), pp. 15-27.
- Ferrari G., Cutugno F., Savy R., (2002), *ALMMA: Annotazione Linguistica Multilivello Manuale e Automatica; Strumenti e metodologie per l'analisi e la codifica di corpora di parlato*, in "La comunicazione, Note Recensioni & Notizie, Pubblicazione dell'Istituto superiore C.T.I. - Umero Speciale TIPI: Tecnologie Informatiche nella Promozione della Lingua Italiana", LI, pp. 107-113.
- Ferrari G., (2003), *TAL, TIC, TCI ed altre sigle: saperi di base e formazione ad alcuni sbocchi professionali*, in Giacalone Ramat A., Rigotti E., Rocci A., (2003, a cura di), pp. 95-109.
- Ferrari G.,(2004), *A State of the Art in Computational Linguistics*, in "17<sup>th</sup> Congress of Linguists Proceedings", John Benjamins, Amsterdam.
- Förstemann E., (1852), *Numerische Lautverhältnisse im Griechischen, Lateinischen und Deutschen*, in "Zeitschrift für Vergleichende Sprachforschung (Kuhn Zeitschrift)", Berlino, pp. 163-179.
- Förstemann E., (1853a), *Numerische Lautbeziehungen des Griechischen, Lateinischen und Deutschen zum Sanskrit*, in "Zeitschrift für Vergleichende Sprachforschung (Kuhn Zeitschrift)", 1853, pp. 35-44.
- Förstemann E., (1853b), *Numerische Lautverhältnisse in griechischen Dialekten*, in "Zeitschrift für Vergleichende Sprachforschung (Kuhn Zeitschrift)", 1853, pp. 401-414.

- Gazdar G., Klein E., Pullum G.K., Sag I., (1987), *Generalised Phrase Structure Grammar*, Basil Blackwell, Oxford.
- Giacalone Ramat A., Rigotti E., Rocci A., (2003, a cura di) *Linguistica e Nuove Professioni*, Franco Angeli, Milano.
- Green, B., Wolf, A., Chomsky, C., and Laugherty, K., (1961), *BASEBALL: An automatic question answerer*. in "Proceedings of the Western Joint Computer Conference", pp. 219–224.
- Gruber, G. (1976), *Lexical structures in syntax and semantics*. North-Holland, New York.
- Guiraud P. (1954a), *Bibliographie critique de la statistique linguistique*, Spectrum, Utrecht.
- Guiraud P. (1954b), *Les caractères statistiques du vocabulaire*, Presses Universitaires de France, Parigi.
- Heims S.J. (1980), *John von Neumann and Norbert Wiener*, MIT Press, Cambridge MA.
- Hendrix G.G., Sacerdoti E.D., Sagalowicz D., Slocum J., (1978), *Developing a natural language interface to complex data*, in "ACM Transactions on Database Systems", III, pp. 105-147.
- Herdan G. (1956), *Language as Choice and Chance*, Noordhoff, Groningen.
- Juilland A. (1964), *Frequency Dictionary of Spanish Words*, De Gruyter, Berlino.
- Juilland A. (1965), *Frequency Dictionary of Rumanian Words*, De Gruyter, Berlino.
- Juilland A. (1973), *Frequency Dictionry of Italian Words*, De Gruyter, Berlino.
- Kaplan R. M. (1972), *Augmented transition networks as psychological models of sentence comprehension*, in "Artificial Intelligence" III, pp. 77-100.

- Kaplan R.M. & Bresnan J., (1982), *Lexical-functional grammar: A formal system for grammatical representation*, in Bresnan J., (1982, a cura di).
- Kamp H. & Reyle U., (1993), *From Discourse to Logic*, Kluwer, Dordrecht.
- Locke W.N.& Booth A.D., (1955, a cura di), *Machine Translation of Languages: Fourteen Essays*, The Technology Press of the MIT/John Wiley/Clapham Hall pp. 15-23.
- Mariotti F., (1880), *Dante e la statistica delle lingue*, Firenze, 1880.
- Minsky M., (1968, a cura di) *Semantic Information Processing*, MIT Press, Cambridge MA, 1968.
- Minsky M. (1975), *A framework for representing knowledge*, in Winston P.E., (1975, a cura di), pp. 211-277.
- Peters, S. & Ritchie R., (1973), *On the Generative Power of Transformational Grammars*, in "Information Sciences" VI, pp.49-83.
- Ross Quillian M. (1968), *Semantic Memory*, in Minsky M., (1968, a cura di), pp. 227-270.
- Rossini Favretti R. (2000, a cura di), *Linguistica e Informatica. Corpora, multimedialità e percorsi di apprendimento*, Bulzoni, Milano 2000
- Sacerdoti E. (1977), *A Structure for Plans and Behavior*, Elsevier North-Holland, New York.
- Schleicher A. (1863), *Die Darwinsche Theorie und die Sprachwissenschaft*, Böhlau, Weimar.
- Tov E. (2001), *Textual Criticism of the Hebrew Bible*, Fortress Press, Minneapolis / Royal Van Gorcum, Assen.
- Turing A.M. (1936), *On computable numbers, with an application to the Entscheidungsproblem*, in "Proceedings of the London Mathematical Society", Series 2, Vol.42, pp. 230-265.
- Turing A.M. (1950), *Computing Machinery and intelligence*, in "Mind – A quarterly review in Psychology and Philosophy", LIX., 236, pp. 433-460.

Winograd T. (1983), *Language as a Cognitive Process: Syntax*, Addison-Wesley.

Winston P.E., (1975, a cura di), *The Psychology of Computer vision*, McGrawHill, New York.

Woods W. (1970), *Transition Network Grammars for Natural Language Analysis*, in "Comm. ACM", XIII, No. 10, pp. 591-606.