

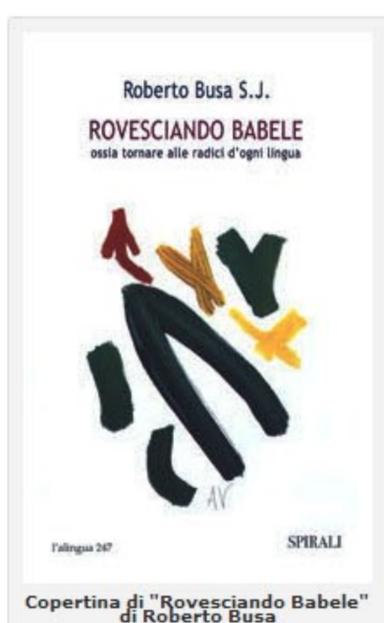
## Padre Busa l'uomo che insegnava le lingue ai computer

Pino Bruno | agosto 10th, 2011 - 08:01

Mi piace 21 +1 1 Tweet 5 Share 2

(Tempo di lettura stimato per questo post 372 secondi)

Padre Roberto Busa non c'è più e la scienza, l'informatica, perdono una delle menti più geniali e produttive del secolo scorso. Il gesuita - scienziato aveva ideato l'ipertesto quando nessuno pensava lontanamente che avrebbe cambiato le sorti della scrittura. Aveva 98 anni ma non aveva mai smesso di pensare al domani. Un futuro in cui i computer avrebbero aiutato l'umanità a superare Babele, i confini delle lingue. E' stato lui a inventare l'informatica linguistica, un mix di umanismo e tecnologia che ha poi portato alla nascita di Google Translate e degli altri traduttori automatici. Che balbettano ancora, è vero, ma presto saranno interpreti forbiti degli idiomi del mondo.

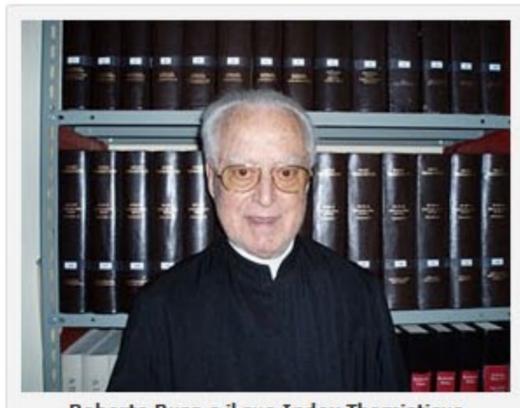


Una delle ultime interviste (se non l'ultima) è dell'aprile 2008. Gino Banterla, giornalista del mensile **Green**, gli aveva chiesto se "riuscirà davvero il computer a farci parlare un'unica lingua". Padre Busa, che aveva appena pubblicato per **Spirali** il saggio sulla linguistica computazionale "**Rovesciando Babele**", gli aveva risposto così: "Direi di no se intendiamo che in prospettiva tutti gli uomini possano parlare una sola lingua.

Ma rispondo senz'altro di sì se consideriamo il computer come un intermediario tra persone che parlano lingue diverse, come portatore di una lingua ontologica e ideografica nella quale ogni singola parola ha un solo significato. Un cinese e un italiano grazie al computer potranno un giorno dialogare tra di loro senza conoscere l'uno la lingua dell'altro e neppure l'inglese.

In pratica accadrà questo: il cinese invia un testo nella sua lingua al computer, il quale lo traduce e lo trasmette all'italiano; quest'ultimo scrive in italiano e il computer traduce in cinese. Prima che si arrivi a questi risultati gli esperti italiani e cinesi dovranno ovviamente elaborare i necessari programmi, risultato del matrimonio tra la fisica dell'informatica e la microanalisi delle lingue".

L'intervista prosegue così: "Chiamiamo viva qualsiasi lingua venga usata istintivamente da chiunque su qualunque argomento. Essa punta alla ridondanza e all'ornamento: quando uno si esprime spesso cerca di abbellire il suo discorso come fa con la sua persona. La lingua viva non è domabile dal computer, come ha dimostrato il **Rapporto ALPAC del 1966** negli Stati Uniti. La ragione sta nel fatto che le variabili della lingua viva - parlata e scritta - sono tante quanti gli individui umani, moltiplicate per gli anni della loro vita e per il numero degli ambienti in cui essi si sono mossi: paese, scuola, professioni, famiglia, religione, letteratura, arte, sport e via dicendo.



Chiamo invece disciplinata una lingua che sia stata potata, cioè ridotta all'essenziale, per finalità particolari. Esempi di potatura dei testi sono gli indici, gli abstracts di articoli scientifici, i telegrammi, gli sms. La lingua disciplinata, a differenza di quella viva, può essere domata dal computer"

Con quale procedura? Gli chiede Gino Banterla:

"È necessaria una premessa. Nel lessico di chiunque parli o scriva vi sono due emisferi: poche voci di altissima frequenza (le cosiddette parti grammaticali, variabili e invariabili) presenti in ogni argomento ed esprimenti la logica; moltissime voci che specificano ciò di cui si tratta e quindi esprimono la cultura o il contesto (alimentazione, trasporti, sport, musica, arte...). Prendiamo ora un determinato testo scritto in una lingua viva, per esempio un manuale di diritto commerciale o di produzione industriale, per sottoporlo a microanalisi e a microsintesi.

Va fatto anzitutto un censimento di tutti gli elementi linguistici, di tutte le parole, classificandole senza eccezioni, con le loro frequenze sia assolute sia percentuali. Forse è utile, per far capire questa fase, richiamare quanto ho fatto nell'**Index Thomisticus**. Gli 11 milioni di parole che formano l'opera di san Tommaso d'Aquino sono riuniti in 147.088 forme diverse di parola. Messe in ordine di frequenza decrescente, risulta che la più alta, la congiunzione et, ne ha 295.593, mentre sono ben 29.637 le voci rigorosamente hapax, cioè presenti una sola volta. Le 80 parole più frequenti rappresentano insieme il 41 per cento di tutte le parole scritte da san Tommaso, le 800 più frequenti il 66 per cento.

Sulla base del censimento del nostro testo comincia la potatura: una équipe di esperti fa una cernita delle parole ritenute indispensabili sia del primo che del secondo emisfero e attraverso un complesso processo di confronti e di decantazione del sistema lessicologico si arriva a una lingua semplificata, standardizzata, potata fino al suo nucleo essenziale..."

Gli chiede ancora Banterla: Ma che necessità c'è oggi della traduzione automatica quando in tutto il mondo si sta affermando sempre di più la lingua inglese?

È vero, tramontata l'utopia dell'esperanto e di altri circa duecento tentativi di lingua artificiale molti considerano l'inglese come una sorta di lingua franca planetaria. In realtà l'inglese è già correntemente usato in tutti i Paesi del mondo nel campo della medicina, della chimica, della fisica, della matematica, in genere in tutte le scienze esatte e naturali.

Il mio progetto di traduzione automatica si rivolge invece a quella infinità di persone - penso per esempio ai piccoli imprenditori - che hanno necessità di un mediatore linguistico preciso e versatile in un'epoca di grande mobilità in cui le porte del mondo sono costituite da un biglietto d'aereo.

Spero che qualche organismo internazionale si renda conto dell'importanza del **Progetto Lingue disciplinate** e voglia sostenerlo, anche perché esso sarebbe un investimento sicuro, in continua crescita e duraturo. Esso rappresenta una straordinaria opportunità, un'infrastruttura pubblica di comunicazione sociale a livello mondiale capace di produrre vantaggi duraturi e di dare un contributo per l'abbattimento delle disuguaglianze tra i popoli".

La biografia di padre Roberto Busa è qui, su Wikipedia.

**Qui c'è l'intervista completa** dello scienziato rilasciata a Gino Banterla.

**Qui c'è un mio post** del 2008

**Qui c'è un bell'articolo** di Stefano Lorenzetto per l'**Osservatore Romano**: "Se navighi in Internet, lo devi a lui. Se usi il pc per scrivere mail e documenti di testo, lo devi a lui. Se puoi leggere questo articolo, lo devi, lo dobbiamo a lui".

Ne rubo un pezzettino: "... Nel 1949 "il gesuita s'era messo in testa di analizzare l'opera omnia di san Tommaso: un milione e mezzo di righe, nove milioni di parole (contro le appena centomila della Divina Commedia). Aveva già compilato a mano diecimila schede solo per inventariare la preposizione 'in', che egli giudicava portante dal punto di vista filosofico.

Cercava, senza trovarlo, un modo per mettere in connessione i singoli frammenti del pensiero dell'Aquinate e per confrontarli con altre fonti. In viaggio negli Stati Uniti, padre Busa chiese udienza a Thomas Watson, fondatore dell'Ibm. Il magnate lo ricevette nel suo ufficio di New York.

Nell'ascoltare la richiesta del sacerdote italiano, scosse la testa: "Non è possibile far eseguire alle macchine quello che mi sta chiedendo. Lei pretende d'essere più americano di noi". Padre Busa allora estrasse dalla tasca un cartellino trovato su una scrivania, recante il motto della multinazionale coniato dal boss - Think, pensa - e la frase "Il difficile lo facciamo subito, l'impossibile richiede un po' più di tempo".

Lo restituì a Watson con un moto di delusione. **Il presidente dell'Ibm, punto sul vivo, ribatte': 'E va bene, padre. Ci proveremo. Ma a una condizione: mi prometta che lei non cambierà Ibm, acronimo di International business machines, in International Busa machines'.** "E' da questa sfida fra due geni - ricorda l'Osservatore romano - che nacque l'ipertesto..."