

Rovesciando



*Pieter Bruegel
il Vecchio
(1525-1569),
La Torre di
Babele.*

Tutta la terra aveva una sola lingua e le stesse parole. Emigrando dall'oriente gli uomini capitarono in una pianura nel paese di Sennaar e vi si stabilirono. Si dissero l'un l'altro: "Venite, facciamoci mattoni e cuociamoli al fuoco". Il mattone servì loro da pietra e il bitume da cemento. Poi dissero: "Venite, costruiamoci una città e una torre, la cui cima tocchi il cielo e facciamoci un nome, per non disperderci su tutta la terra". Ma il Signore scese a vedere la città e la torre che gli uomini stavano costruendo. Il Signore disse: "Ecco, essi sono un solo popolo e hanno tutti una lingua sola; questo è l'inizio della loro opera e ora quanto avranno in progetto di fare non sarà loro impossibile. Scendiamo dunque e confondiamo la loro lingua, perché non comprendano più l'uno la lingua dell'altro". Il Signore li dispersé di là su tutta la terra ed essi cessarono di costruire la città. Per questo la si chiamò Babele, perché là il Signore confuse la lingua di tutta la terra e di là il Signore li dispersé su tutta la terra.

Babele

di Gino Banterla



Padre Roberto Busa, rileggendo il celebre brano della Genesi (11, 1-9) viene spontaneo interrogarsi sul senso di Babele oggi. Nei primi anni del terzo millennio dopo Cristo viviamo tante "babele": della comunicazione, dei valori, dell'economia, oltre che delle lingue. La scienza nell'ultimo secolo ha fatto progressi giganteschi, eppure siamo nella stessa situazione descritta dalla Bibbia. Gli uomini con le loro invenzioni vogliono toccare il cielo, ma si trovano "dispersi sulla terra", disorientati e spesso incapaci di capirsi e di comunicare tra loro. Quale è il ruolo della lingua e delle lingue nella società globalizzata?



Disegno
di Marina
Molino
Ronza

Padre Roberto Busa, pioniere della linguistica informatica

O

Dai calcolatori a schede perforate ai bit



Padre Busa davanti allo scaffale nel quale sono raccolti i 56 volumi del suo *Index Thomisticus*.

San Tommaso d'Aquino diceva che la lingua, cioè il parlare, è un atto d'amore per il prossimo e che la menzogna è per questo un peccato contro la natura. Senza sincerità e senza trasparenza non si è galantuomini. Parlare, comunicare deve essere prima di tutto e sopra tutto un atto d'amore. Mettiamocelo bene in testa...

■ Lei è il pioniere della linguistica computazionale, alla quale ha dedicato sessant'anni di ricerche. E il 28 gennaio scorso – guarda caso, proprio nel giorno che la Chiesa dedica a San Tommaso – ha presentato pubblicamente a Milano il Progetto LD (Lingue disciplinate), che si propone di vincere una delle sfide imposte dalla

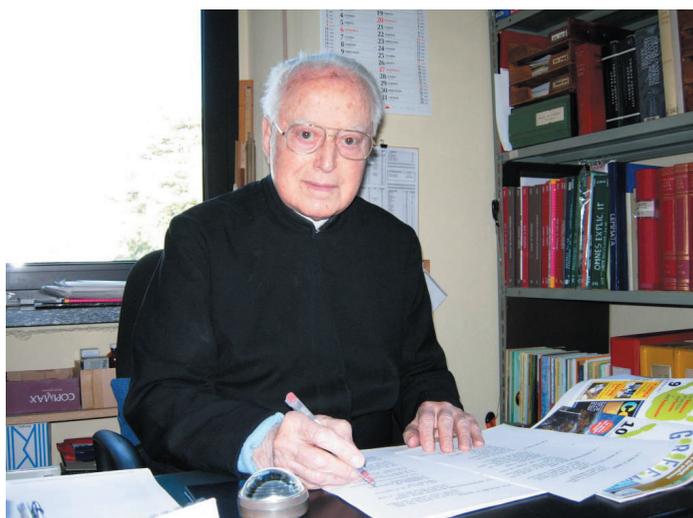
Sessant'anni di ricerche sulla lingua

Ha 95 anni e proprio non li dimostra. Padre Roberto Busa, classe 1913 (è nato il 28 novembre a Vicenza) sembra essere, tutt'al più, un "giovane" settantacinquenne. È stato compagno di scuola di un papa, Giovanni Paolo I, al secolo Albino Luciani, e nella sua lunga carriera accademica ha avuto centinaia di allievi, alcuni diventati grandi personalità: uno per tutti, il cardinale Carlo Maria Martini, già arcivescovo di Milano. Alto di statura, portamento ieratico e professorale ma battuta sempre pronta ("Mi sto perfezionando in vecchiaia", scherza accogliendoci nel suo studio), padre Busa deve alla mente sempre in allenamento la sua sbalorditiva vitalità. Sin da quando, nel 1948, andò a New York all'IBM dove incominciò le sue ricerche linguistiche sui testi di san Tommaso d'Aquino affidandosi agli enormi calcolatori a schede perforate. Nessuno ci credeva allora, ma i gesuiti in campo scientifico quando ci si mettono ottengono il meglio, l'eccellenza diremmo oggi. E

difatti padre Busa S.J. (*Societatis Jesu*, è la sigla con la quale si firmano gli appartenenti alla Compagnia di Gesù fondata nel 1540 da Ignazio di Loyola) ha dato vita a una nuova disciplina, la linguistica informatica. Dai "dinosauri" – come lui chiama i primi computer a valvole – ai più recenti strumenti informatici, dalle schede perforate e dai nastri magnetici ai bit e ai byte, padre Busa si è fatto missionario del "computer in humanities" insegnando alla Pontificia Università Gregoriana, all'Università Cattolica del Sacro Cuore, al Politecnico di Milano e partecipando a qualche centinaio di congressi e seminari in ogni parte del globo. La sua bibliografia vanta oltre 440 pubblicazioni tra libri e saggi, tra le quali spicca il ciclico *Index Thomisticus*. Il suo "regno" è il Centro da lui fondato, il Cael (Computerizzazione delle analisi ermeneutiche e lessicologiche) che ha sede all'Aloisianum di Gallarate (Varese), l'Istituto dei gesuiti dove lo abbiamo incontrato.

globalizzazione. Riuscirà davvero il computer a farci parlare un'unica lingua, "rovesciando Babele" come afferma nel suo ultimo libro?

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 lin-



Un'utopia che può diventare realtà.

Riuscirà il computer a farci parlare un'unica lingua?



L'illusione americana durante la "guerra fredda"

In piena guerra fredda, dopo la seconda guerra mondiale, il Governo americano finanziò vari gruppi di ricercatori perché mettessero a punto un sistema di traduzione automatica dal russo all'inglese attraverso il computer.

Erano gli anni del cosiddetto equilibrio del terrore, che si basava sugli arsenali militari e sulla rincorsa agli armamenti, e i due giganti che allora do-

minavano la scena mondiale – l'Unione Sovietica e gli Stati Uniti – si osservavano a distanza attraverso i rispettivi servizi segreti, il Kgb e la Cia. Ma l'illusione americana rimase tale: il progetto si arenò di fronte alle difficoltà rappresentate dalla lingua viva, non pienamente leggibile dal computer – oltretutto in quegli anni si era ancora nella preistoria degli elaboratori

Padre Roberto Busa nel suo studio all'Aloisianum di Gallarate (Varese). È nato a Vicenza il 28 novembre 1913.

gua 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.

■ **La rete veicola miliardi di dati e informazioni di fronte ai quali è sempre più difficile raccapezzarsi. Come trovare una chiave di lettura di fonte alla Babele informativa?**

Nelle forme attuali di trasmissione di dati e di informazioni via rete c'è un grande deficit di precisazioni di linguaggio al quale tutte le grandi aziende dell'ICT, Information and Communication Technology, cercano di ovviare con investimenti in ricerca, per approdare a nuovi codici linguistici con cui far lavorare le macchine. Il mio tentativo si muove proprio in questa direzione: un lessico in grado di sviluppare, standardizzare e globalizzare le attuali forme di comunicazione via rete.

■ **Questo è appunto l'obiettivo del suo Progetto Lingue disciplinate. Ce lo potrebbe illustrare in parole povere?**

Per farmi capire rispondo con un'equazione. Una lingua disciplinata sta a una lingua viva come un frutteto sta a una foresta pluviale.

Roma, 20 dicembre 2005. Il Capo dello Stato, Carlo Azeglio Ciampi, conferisce a padre Busa la massima onorificenza della Repubblica, il Cavaliato di Gran Croce.

L'Index Thomisticus: tutto il sapere del mondo antico in 11 milioni di parole

L'opera di san Tommaso d'Aquino (1221-1274), una intera vita dedicata allo studio e all'insegnamento, è stata "scandagliata" con il computer, nel corso di una ricerca durata tre decenni, da padre Busa. Il risultato di questo lavoro colossale è pubblicato nell'*Index Thomisticus* in 56 volumi di circa mille pagine ciascuno, condensati poi agli inizi degli anni Novanta su cd-rom. Attraverso l'*Index* è possibile rintracciare in un istante qualsiasi parola latina contenuta nei 118 scritti di Tommaso d'Aquino, i quali rappresentano una sorta di "enciclopedia" *ante litteram* del sapere del mondo antico. L'Aquinato condensò 40 secoli di pensieri e di civiltà, da quella assiro-babilonese (nella quale nacque-

ro le Scritture del Vecchio Testamento) ai Greci, dai Romani agli Arabi. "Questa ingente mole di scritti", spiega padre Busa, "non è suddivisa in voci, bensì sintetizzata in un sistema logico-scientifico rappresentante la realtà del cosmo, dalla materia prima su su fino all'*Amor che move il Sole e l'altre stelle* (Dante). È un esempio formidabile di sintesi, da tenere ben presente oggi, in un momento in cui lo sviluppo gigantesco degli strumenti scientifici sta frantumando e parcellizzando i saperi". Undici milioni di parole – *I Promessi Sposi*, tanto per fare un esempio, arrivano "appena" a 230 mila, la *Divina Commedia* a 100 mila – sono state schedate nell'*Index Thomisticus* attraverso un sistema che permette, tra l'altro, di distinguere i termini omografi, quelli, cioè, che si scrivono nello stesso modo ma hanno un significato diverso.

Intervista a padre Roberto Busa

elettronici – in quanto creativa e personalizzata. Infatti l'intelligibilità distribuita delle traduzioni raggiunte al massimo l'80 per cento, percentuale certamente considerevole, ma il 20 per cento restante era diffuso qua e là nel testo, che così non risultava comprensibile nel suo insieme. Nel 1966 uscì il Rapporto ALPAC, Automatic language processing advisory committee, opera di sette ricercatori americani, che decretava la fine degli esperimenti di tradu-

zione automatica. In seguito a quella ricerca infatti il Pentagono – il ministero della Difesa Usa – sospese i finanziamenti. L'idea di uno strumento capace di tradurre automaticamente testi da una lingua all'altra non venne però abbandonata. Con la diffusione dei personal computer e con il vertiginoso sviluppo di Internet, negli ultimi decenni sono stati messi a punto sofisticati programmi di traduzione assistita, che hanno portato alla definizione di



standard sempre più avanzati. Questi software operano però ancora su segmenti di testi, non su un testo organico. Non è stato infatti realizzato fino ad oggi uno strumento per la traduzione automatica e simultanea di un testo articolato. Il Progetto Lingue disciplinate elaborato da padre Roberto Busa suggerisce come colmare questa lacuna.

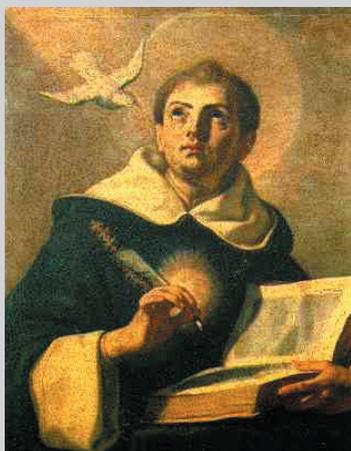
Il Checkpoint Charlie, unico punto di passaggio tra Berlino Est-Ovest, simbolo della guerra fredda.

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 (vedi box). 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?

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 es-



San Tommaso d'Aquino (1221-1274) ritratto con un volume in evidenza da Francesco Solimena noto come l'Abate Ciccio (Canale di Serino 1657 – Napoli 1747) considerato uno degli artisti che meglio incarnarono la cultura tardo-barocca in Italia.

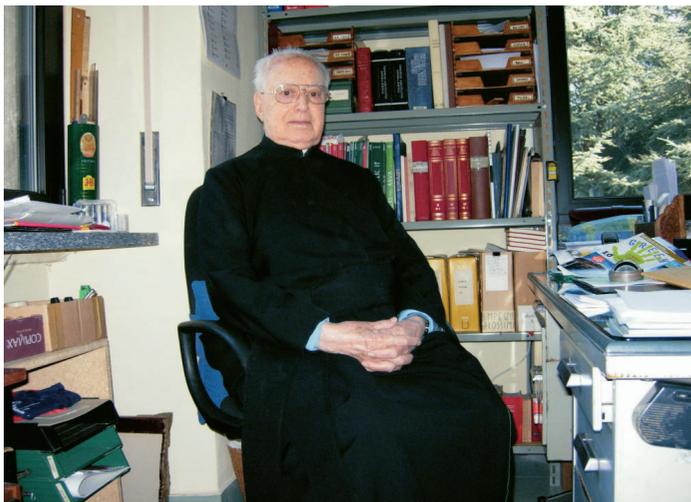
Le parti del discorso variabili e invariabili

Ve le ricordate le parti variabili e invariabili del discorso? Niente paura, ripassiamo la grammatica. Le parti del discorso sono, secondo la definizione del Devoto-Oli, "le categorie in cui sono distribuite le parole di una lingua, secondo il significato, la funzione che assolvono nella frase, le caratteristiche di formazione e flessione".

In italiano sono nove

Parti variabili: articolo, sostantivo, pronome, aggettivo, verbo.

Parti invariabili: avverbio, preposizione, congiunzione, interiezione.



Bit, byte, file, il “linguaggio” del computer

Tutta la materia è riconducibile ai numeri, afferma padre Roberto Busa nell'intervista. Mentre questa è una considerazione personale di carattere filosofico, frutto di una lunga vita tutta dedicata alla ricerca, è un dato di fatto sotto gli occhi di tutti che ai numeri è riconducibile l'immenso lavoro dello studioso gesuita sulle parole e sulle lingue. Bit, byte e files altro non sono infatti che combinazioni di numeri che vanno a formare il linguaggio informatico. Vediamo, in sintesi, come esso si configura.

Il linguaggio informatico si fonda sulla logica binaria, basata su due soli numeri 0 e 1, che altro non sono che la “codificazione” di una realtà fisica ben precisa: lo stato di

senziale ed elementare. A ognuna delle parole selezionate si fa corrispondere un insieme di numeri binari.

■ **E il computer a questo punto compie la magia della traduzione automatica: lo studente di Pechino capisce così all'istante ciò che gli scrive lo studente di Roma e viceversa...**

Sia ben inteso, il computer non fa le magie di Harry Potter. Il lavoro degli esperti, dopo la potatura, è ancora lungo e impegnativo, perché i sistemi lessicologici variano tra lingue, argomenti e scrittori diversi. Prima di arrivare all'input-output italiano-cinese e cinese-italiano dobbiamo disporre di algoritmi che ci dicano, per esempio, quale è nelle due lingue la collocazione di un aggettivo rispetto a un sostantivo o di un verbo rispetto al soggetto. Solo allora i due files saranno come un Giano bifronte e il computer diventerà parlatore di lingue diverse. Lo stesso testo potrà essere immediatamente leggibile in tutte le altre lingue disciplinate.

■ **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.

■ **Non c'è il rischio che l'intelligenza dell'uomo venga schiacciata dall'invadenza del computer?**

Sì, questo rischio effettivamente esiste, ma la colpa è di chi usa il computer. Oggi c'è in atto una pericolosa inflazione dell'informazione che certo non aiuta a pensare. Mi piace fare un altro esempio: mentre per quanto riguarda il cibo ciò che sappiamo sui vari si-

Summary

Babel reversed

Jesuit Father Roberto Busa is the pioneer of computational linguistics. He began his research on the scripts of Thomas Aquinas in 1948 at IBM in New York, entrusting himself on the enormous punched card computers. From the “dinosaurs” – as he called the first valve computers – to the more recent informatic tools, from punched cards and magnetic tapes to bits and bytes, Father Busa has been the missionary of the “computer in humanities” teaching at the Pontificia Università Gregoriana (Pontifical Gregorian University), at the Università Cattolica del Sacro Cuore (Catholic University of the

Holy Heart), at the Politecnico di Milano (Polytechnic of Milan) and participating in hundreds of congresses and seminars all over the world. And even today, as a 95 year old man, he continues his work in the centre he founded, the Cael (Computerization of the hermeneutical and lexicological analyses) which is located at the Aloisianum in Gallarate (Varese), within the Institute of Jesuits. In this interview he talks about the results of sixty years of research and about his new adventure; the Progetto di traduzione automatica Lingue disciplinate (Project of automatic translation of disciplined Languages), a possible answer to the challenges imposed by the globalization and by the Babel of communications.

Intervista a padre Roberto Busa

magnetizzazione di un materiale magnetizzabile, o la presenza/assenza di carica elettrica in un condensatore. Per capire come funziona tale linguaggio pensiamo ad una lampadina che può essere accesa o spenta, o ad una qualsiasi coppia di stati fisici autoescludentisi. Questi due stati, nel computer, corrispondono – come del resto nel caso della lampadina – al numero 1 (accesa) e 0 (spenta). Le cifre binarie sono chiamate bit, termine che deriva dall'inglese BInary digiT. Il bit, in pratica, è l'unità elementare per la misura dell'informazione. In un computer le cifre binarie vengono raggruppate in stringhe: esse contengono un numero costante di bit che servono per rappresentare una lettera, un numero, un simbolo speciale. Un gruppo di otto bit forma un byte ed è la quantità minima con la quale un computer è in grado di lavorare. Con un byte si possono ottenere 256 combinazioni di 0 e 1, attraverso le quali si possono rappresentare le lettere, i numeri, i caratteri speciali e i simboli.

stemi di alimentazione e sulle specialità culinarie non ci porta a provare cento piatti in un giorno, per via degli ovvi freni imposti dall'apparato digestivo, la curiosità del pensiero è senza limiti e là dove ci fossero dei freni, essi agirebbero come un invito e una sfida.

Nel Medioevo si diceva: non cercare di sapere ciò che non ti è lecito fare. È, questa, la legge della temperanza, una legge per così dire di igiene del comportamento umano. In altre parole: occorre essere padroni di se stessi. Anche di fronte all'inflazione dell'informazione. Solo così potremo dire: io cerco tra i miliardi di uomini che zampettano sulla crosta terrestre; cerco di essere la formichina che porta qualcosa di buono agli altri.

■ Se dovesse fare una sintesi del suo lunghissimo percorso di ricerca e di studio, che cosa metterebbe in evidenza?

Quello che mi sorprende è che tutta la materia è riconducibile ai numeri. Direi di più, è fatta da una bufera di frequenze elementari di numeri che sembrano consistenti solo di energia. Comprendo soltanto ora l'importanza della definizione di Aristotele: la materia prima non ha nessuna delle qualità delle cose reali. Comincio a pensare – non so su quale fondamento della fisica – che la struttura atomica o intra-atomica di tutto ciò che chiamiamo materia e corpo e dimensione sia fatta di bolle di energia piene di materia inesi-

Le informazioni sono codificate con i multipli del byte come segue:

kb = kilobyte (1024 byte) = 2^{10} byte

Mb = megabyte (1024 kilobyte)

Gb = gigabyte (1024 megabyte)

Tb = terabyte (1024 gigabyte)

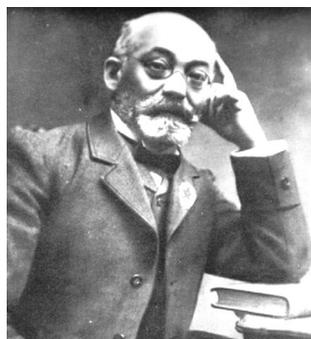
I dati digitali sono rappresentati dal numero di byte che li compongono e dalla loro sequenza e vengono chiamati campi; il record è un insieme di campi; un insieme di record va a formare un file, che può essere un testo, un video, un brano musicale.

stente e che quindi lo spazio e gli spazi siano mari di briciole d'energia.

■ Padre Busa, lei alla soglia dei suoi 95 anni è ancora in contatto con i giovani. Gli studenti, oggi, si sentono spesso disorientati: in famiglia, nella scuola, nella società. Quale consiglio darebbe a un ragazzo o a una ragazza che venissero a bussare alla sua porta per esprimere il proprio disagio di fronte a una società sempre più povera di valori?

Il disorientamento dei giovani è proprio una conseguenza della parcellizzazione delle conoscenze, ossia della mancanza di sintesi conseguente al bombardamento confuso dell'informazione che ci piove addosso da tutti i punti dell'orizzonte e da tutti i continenti. Ricordo un signore conosciuto a New York: mi diceva che per sentirsi solo gli bastava soffermarsi nella Times Square. A un giovane che bussasse alla mia porta per chiedere consiglio direi: cerca di scegliere qualche cosa di buono in qualunque settore al quale tu ti senta portato (scienze, musica, teatro, filosofia, storia, pittura...) e di dedicarvi con impegno il tuo tempo. Devi farlo come un tuo percorso di crescita personale, per questo non aspettarti e non esigere nulla dagli altri. Non farlo per ricavarne gloria, ma semplicemente perché tu senta in modo pieno la tua dignità di uomo di fronte a te stesso.

Gino Banterla



Ludwik Lejzer Zamenhof
(1859-1917)

Gli irriducibili dell'esperanto

Non esistono statistiche che documentino quante persone parlino l'esperanto nel mondo: qualcuno dice che sono appena 50.000, altri azzardano due milioni. Questa lingua "artificiale" senza popolo né nazione risale alla seconda metà dell'Ottocento. A mettere a punto le basi dell'esperanto fu un oculista polacco, Ludwik Lejzer Zamenhof

(1859-1917), figlio di un insegnante di lingue, che amava firmarsi con lo pseudonimo Doktoro Esperanto, "colui che spera".

La prima grammatica italiana di esperanto venne pubblicata nel 1890. Nel 1954 l'UNESCO, United Nations educational, scientific and cultural Organization, riconobbe "i risultati raggiunti dall'esperanto

nel campo degli interscambi intellettuali internazionali e per l'avvicinamento dei popoli del mondo", ma nonostante il sostegno dell'UNESCO e l'impegno delle associazioni esperantiste questa lingua non ha mai avuto grande diffusione. Per saperne di più (e per chi vuole cimentarsi direttamente on line con l'esperanto): <http://it.lernu.net/>