I quaderni di Telèma
A cura di Alberto Mucci

Una sfida dell'Europa a 25: la molteplicità delle traduzioni

L'editoria multimediale a Francoforte

Europa "allargata": 25 paesi e 25 lingue. La necessità di comprendersi, di dialogare, di scambiare messaggi, dati, informazioni diventa sempre più urgente. Una prospettiva che sta facendo compiere significativi passi in avanti alle tecniche che portano a realizzare (non è più un sogno!) la traduzione automatica, la comprensione di una notizia, in qualunque lingua venga scritta o detta.

La multimedialità è da tempo oggetto di sperimentazioni e di realizzazioni significative e importanti. In quest'ambito uno spazio di primo piano sta conquistando la traduzione automatica, come dimostra il "libro bianco" sul TAL (trattamento automatico del linguaggio) che la Fondazione Ugo Bordoni presenta alla Fiera del Libro a Francoforte (6-10 ottobre 2004), l'appuntamento più atteso e significativo dell'editoria mondiale.

Da anni la traduzione automatica è sul tavolo della sperimentazione. Progressi marcati si stanno registrando da quando si sono diffusi anche a livello di singoli cittadini (e non soltanto di imprese) i servizi on line gratuiti di TA (traduzione simultanea). In pratica: un testo dettato in italiano può essere oggi tradotto immediatamente e automaticamente in altre lingue (una quindicina sulla base delle tecnologie finora messe a punto). E l'apparecchio è in grado di leggere il testo nelle singole lingue prescelte.

L'editoria del futuro è multimediale e nel contempo multilingue. Una sfida è politica (oltre che tecnologica): non si tratta di imporre una lingua comune a tutti, una "globalizzazione del parlare", oltretutto impossibile, ma di creare le condizioni tecniche perché i popoli possano comprendersi e avanzare lungo la strada della comunicazione e della conoscenza. Che è la strada del progresso e della pace.

Supplemento al numero 219 di settembre 2004
di MEDIA DUEMILA

INDICE

Il Quaderno è stato realizzato dalla Fondazione Ugo Bordoni (Presidente il Prof. Giordano Bruno Guerri, Direttore Generale il Consigliere Guido Salerno, Direttore delle Ricerche l'Ing. Mario Frullone). Coordinatore del Quaderno l'ing. Andrea Paoloni. Hanno collaborato: Andrea Di Carlo, Fondazione Ugo Bordoni; Johanna Monti, Università degli Studi di Napoli "L'Orientale"; Claudio Cirilli, SYNTHEMA s.r.l.; Gianni Lazzari, Itc-irst; Elisa Ranucci-Fischer, DGT.

Sono usciti:

I satelliti nella società multimediale	dicembre-gennaio **2003**
Telefonia mobile e emissioni elettromagnetiche	febbraio **2003**
Le reti di telecomunicazioni diventano intelligenti	marzo **2003**
Mentre viaggi lavori con Internet	aprile **2003**
Come garantire sicurezza con lo sviluppo di Internet	maggio **2003**
Le macchine che parlano	giugno **2003**
Le macchine che capiscono	luglio-agosto **2003**
Il progresso tecnologico fra brevetti e standard	settembre **2003**
La rendicontazione? Automatica, ma…	ottobre **2003**
Le nuove tecnologie fotoniche	novembre **2003**
Il progetto Galileo sta diventando realtà	dicembre-gennaio **2004**
Non confondere la biometrica con il "grande fratello	febbraio **2004**
Dal call center al contact center	marzo **2004**
La larga banda si diffonde cambia la vita della gente	aprile **2004**
I campi elettromagnetici non sono più "sconosciuti"	maggio **2004**
Anche l'Italia si dota di un organismo che certifica la sicurezza informatica	giugno **2004**
Il digitale terrestre accende i motori	luglio-agosto **2004**

Il "libro bianco" sul TAL

Il ForumTAL sul trattamento automatico del linguaggio, costituito presso il ministero delle comunicazioni con l'obiettivo di supportare la ricerca e lo sviluppo nelle tematiche della comunicazione parlata e scritta (Trattamento Automatico del Linguaggio - TAL), ha individuato come suo primo obiettivo la stesura di un libro bianco che fornisca, nel medesimo tempo, un riferimento sullo stato dell'arte di questa tecnologia e una guida del mercato italiano delle tecnologie del linguaggio contenente l'elenco completo degli Enti interessati al TAL nell'ambito della ricerca, della formazione e dell'industria.

Il Libro Bianco sul TAL vuole pertanto essere uno strumento di lavoro per tutti coloro che svolgono un'attività nel campo della linguistica computazionale e soprattutto una guida per coloro che sono interessanti ad introdurre queste tecnologie nella loro attività per migliorare l'impatto della loro offerta commerciale.

Il Libro Bianco parte dalla definizione di cosa sia il TAL, e il compito non è facile perché il trattamento del linguaggio pervade molte applicazioni senza identificarsi nella relativa tecnologia di riferimento. Si pensi ai telefonini, che contengono numerose tecnologie TAL, relative sia all'elaborazione del parlato, come la codifica della voce alla base della tecnologia "GSM", sia il riconoscimento del parlato, per consentire di effettuare una chiamata senza comporre il numero di telefono ma semplicemente pronunciando un nome, sia l'elaborazione del testo, si pensi alle tecniche di scrittura degli SMS; nonostante ciò il cellulare non viene identificato come prodotto TAL. Un altro esempio di uso inconsapevole della tecnologia TAL è il correttore di testi (in genere disponibile per varie lingue) che tutti abbiamo sul computer, che tutti usiamo, soprattutto per evitare i refusi; un altro esempio è costituito dai sistemi di traduzione automatica, che permettono anche a chi non conosce una lingua di avere almeno un'idea del contenuto di un messaggio o di una pagina web.

Il testo presenta le principali tecnologie riconducibili al TAL, come la gestione dei contenuti, l'indicizzazione e la ricerca d documenti testuali disponibili in rete, l'indicizzazione e l'archiviazione di materiale vocale in basi di dati multimediali, la traduzione automatica dei testi scritti e della comunicazione orale, l'interfaccia vocale per il comando vocale agli elettrodomestici o dell'automobile o per applicazioni più complesse tra le quali l'aiuto ai disabili. Alla rassegna delle applicazioni e degli algoritmi che sono alla base dei sistemi oggi disponibili, segue un capitolo che illustra lo Stato dell'arte del TAL in Italia; vengono descritti i risultati dell'analisi di 71 questionari riempiti da soggetti del mondo accademico e del mondo dell'industria e dei servizi. In particolare, l'attenzione è stata rivolta all'attività e alla produzione realizzate da strutture nazionali; nella rassegna pertanto non compaiono soggetti stranieri o multinazionali che, pur presenti sul mercato italiano con prodotti di TAL, non hanno in Italia strutture di produzione o di ricerca e sviluppo.

Sono stati descritti ben 206 progetti di ricerca e sviluppo, l'offerta commerciale relativa a 229 prodotti, e anche l'attività di formazione universitaria.

Da questa analisi sono risultati alcuni interessanti spunti per ulteriori discussioni, ad esempio l'esigenza di far conoscere e meglio promuovere le potenzialità del TAL nelle Pubbliche Amministrazioni che sembrano essere scarsamente presenti nella risposta ai questionari e nelle collaborazioni ivi evidenziate, ma d'altro canto mostrano un significativo interes-

se partecipando ai livelli più alti alla vita del ForumTAL.

Dall'analisi emergono difficoltà nel quantizzare economicamente il dominio e una scarsa diffusione della cultura della condivisione dei prodotti per scopi di ricerca. Sembra inoltre opportuno promuovere una maggiore sinergia tra le varie componenti della comunità nazionale.

Sul piano della formazione si segnala un progressivo disimpegno dal dominio delle componenti scientifiche fisico-matematiche, a favore dell'ingegneria e delle cosiddette "scienze umane".

Segue poi un capitolo che contiene dieci interviste ad esperti di differente qualifica i quali illustrano il loro punto di vista sulla tecnologia del TAL.

Agli intervistati, che sono utenti di questa tecnologia, sviluppatori della stessa o addetti alla ricerca, sono state poste domande relative alla diffusione della conoscenza della tecnologie, all'importanza che riveste nell'ambito della cultura, alle strade da battere per promuoverla, su quali siano le ricerche da privilegiare, e su come stimolare lo sviluppo della crescita di questo mercato. Non intendiamo qui proporre le risposte, talvolta contraddittorie, che sono state fornite alle precedenti domande, ci limitiamo a citare l'accordo unanime circa la necessità di una formazione multidisciplinare attualmente carente e la scarsa informazione sulle potenzialità di questa tecnologia che, conseguentemente, è poco diffusa negli ambienti di lavoro.

Alle interviste fa seguito un capitolo con le presentazioni di tutti gli enti che hanno risposto al questionario loro sottoposto e sulla base del quale sono state tratte le notizie del capitolo secondo. Infine i capitoli successivi contengono l'elenco dei prodotti disponibili in Italia e l'indirizzario delle ditte e degli enti impegnati nel TAL.

Il Libro Bianco verrà presentato alla Fiera del Libro di Francoforte, che costituisce l'evento più significativo per l'editoria mondiale; presso la mostra saranno disponibili sistemi dimostrativi della tecnologia TAL, volti a rendere visibile le potenzialità di questa tecnologia.

Andrea Paoloni e Andrea Di Carlo
Fondazione Ugo Bordoni

Dal sogno meccanico alla e-translation: la traduzione automatica è realtà?

La traduzione automatica ripropone in termini moderni uno dei più antichi sogni dell'uomo: la possibilità di costruire una macchina in grado di pensare e agire come un essere umano. L'ipotesi di partenza è che i complessi meccanismi mentali che governano l'attività umana di traduzione da una lingua naturale ad un'altra siano riducibili, almeno in parte, ad un insieme di procedure eseguibili da un programma di calcolatore elettronico.

La storia della traduzione automatica/assistita, ovvero il tentativo di automatizzare tutto o parte del processo di traduzione, è, da un lato, caratterizzata dagli entusiasmi, talvolta eccessivi, da parte dei ricercatori impegnati nella progettazione e nello sviluppo dei sistemi (che soprattutto all'inizio speravano di ottenere risultati se non uguali, almeno comparabili alla traduzione effettuata da traduttori professionisti) ma, dall'altra, dalla diffidenza del grande pubblico, e dai timori

da parte dei traduttori nei confronti delle nuove tecnologie.

LA PRIMA IDEA

La prima idea di un dizionario basato su codici numerici per tradurre da e verso altre lingue risale addirittura all'Illuminismo europeo e fu sviluppata da Descartes e Leibniz. Tale concetto si ispirava al movimento che teorizzava il "linguaggio universale", ovvero un linguaggio basato su principi logici e su simboli iconici comprensibili universalmente. Questa idea venne sviluppata solo con l'avvento del calcolatore e con i progressi dell'informatica nel secolo scorso: gli universali linguistici, ovvero regole comuni a tutte le lingue naturali, sembravano poter essere la base ideale per la realizzazione di un software in grado di tradurre da una lingua naturale ad un'altra senza alcun intervento da parte dell'uomo.

Negli anni '60, sull'onda della grammatica generativo-trasformazionale elaborata dal linguista americano Noam Chomsky, la ricerca nel campo della TA si orientò alla realizzazione di grammatiche formali, concentrandosi esclusivamente sugli aspetti sintattici del linguaggio. Ma questo approccio approdò ad un punto di non ritorno verso la metà degli anni '60 quando si realizzò che la sintassi da sola rappresentava una base insufficiente per la traduzione automatica: i risultati prodotti dai sistemi di traduzione automatica di tipo sintattico erano infatti assolutamente scoraggianti. L'idea del linguaggio universale basato su meri principi sintattici si scontrò con il problema della polisemia, dell'ambiguità e della complessità del linguaggio naturale. Superare le barriere semantiche divenne il vero problema per i ricercatori impegnati in questo settore. Essendo oramai divenuto chiaro che la sola analisi sintattica consentiva di disambiguare i testi solo a livello morfo-sintattico, ma non era di alcun aiuto nella comprensione ed interpretazione di un testo, la ricerca si orientò verso lo sviluppo di sistemi basati su modelli semantici, che si ritenevano potessero operare in maniera più efficace rispetto ai modelli sintattici. Ma anche in questo caso i risultati furono alquanto deludenti.

L'idea di un metalinguaggio basato su universali linguistici venne quindi abbandonato in favore di un approccio meno ambizioso, più pragmatico, ma che produsse risultati qualitativamente migliori, ovvero l'approccio a transfer, utilizzato da molti sistemi commerciali attualmente sul mercato, come ad esempio Systran, su cui si basa il servizio di traduzione automatica on-line BabelFish. I sistemi a transfer sono basati su una struttura a tre fasi.

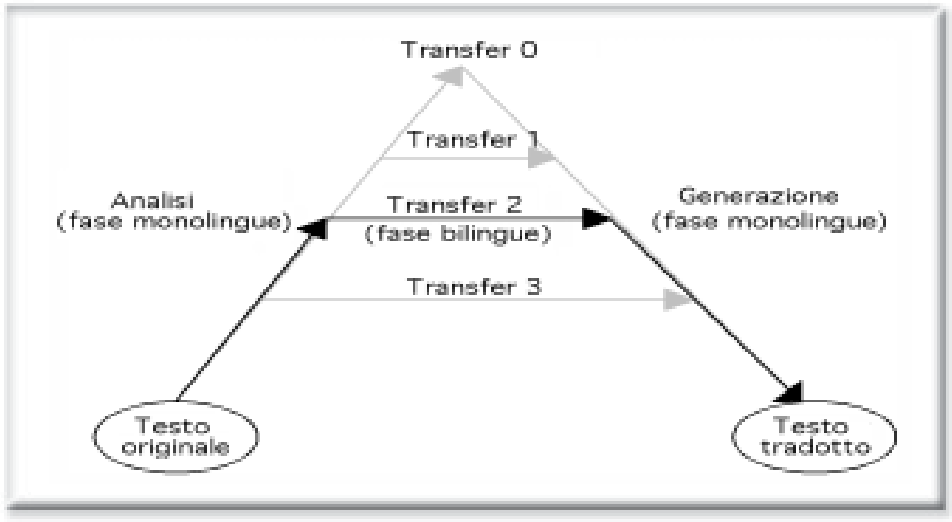

Figura 1. Schema logico dei sistemi di traduzione.

La prima fase è rappresentata dall'analisi della lingua di partenza che ha come risultato il passaggio dalla lingua naturale ad una rappresentazione astratta della lingua stessa, sia dal punto di vista lessicale che grammaticale (ed, in alcuni casi, anche dal punto di vista semantico): questa rappresentazione astratta intermedia rappresenta la base per la successiva fase di transfer in cui si trasforma, ovvero si converte, tale rappresentazione della lingua di partenza nella corrispondente rappresentazione astratta della lingua di arrivo. L'ultima fase di generazione converte nuovamente la rappresentazione astratta intermedia della lingua d'arrivo nella corrispondente lingua naturale.

Ad approcci di tipo linguistico al problema della traduzione automatica, come quelli fin'ora descritti, si sono affiancate, nel corso degli ultimi anni, linee di ricerca diverse che riguardano sostanzialmente lo sviluppo di basi conoscenza, di modelli statistici e di ampi corpora testuali bilingui e multilingui.

La tendenza attuale è quella di un approccio integrato, ovvero l'integrazione di tecnologie tradizionali con approcci diversi, come quelli appena menzionati, anche di tipo non linguistico, avvantaggiandosi dei benefici offerti dalle diverse tecnologie, unificate all'interno di un unico sistema.

L'obiettivo perseguito

Nonostante i progressi dell'informatica, negli ultimi anni la ricerca nel campo della TA non ha avuto grosse evoluzioni e non ha prodotto grossi miglioramenti relativamente alla qualità della traduzione e soprattutto non è stata ancora sviluppata una macchina in grado di riprodurre il ragionamento ed il linguaggio di un essere umano. Obiettivo questo difficilmente perseguibile, almeno allo stato attuale, da parte di un computer soprattutto se si considera che il linguaggio naturale è caratterizzato dalla massima libertà di espressione e dalla massima flessibilità, è aperto al cambiamento, non è basato su regolarità bensì su irregolarità, è frutto della creatività dell'essere umano. Il linguaggio utilizzato dal computer, invece, è un linguaggio artificiale, caratterizzato da una estrema rigidità, basato su regolarità e ricorsività. Nel passaggio dal linguaggio naturale al linguaggio del computer si perde sempre qualcosa.

Così, benché l'obiettivo della traduzione automatica sia la traduzione da una lingua naturale ad un'altra senza l'intervento umano, il risultato del processo di traduzione automatica non è una lingua naturale, bensì la lingua che è il sistema è stato programmato a produrre. Il linguaggio prodotto da un sistema di TA è di fatto il risultato dell'esecuzione di algoritmi, basati su un insieme limitato di dati (lessicali, sintattici e talvolta semantici) ed è dunque una lingua artificiale. Il linguaggio umano è il risultato della creatività umana, che non è guidata da processi di tipo algoritmico, bensì da processi complessi non pienamente rappresentabili in un codice macchina in quanto guidati da intuizione e flessibilità.

Nel primo capitolo del suo libro Dire quasi la stessa cosa: esperienze di traduzione, (Bompiani 2003) U. Eco chiarisce molto bene questo contrasto tra lingua naturale e lingua utilizzata dal computer, sulla base di una serie di traduzioni effettuate utilizzando appunto BabelFish, il servizio gratuito di traduzione automatica offerto da Altavista.

Storia della traduzione automatica

Il primo vero impulso allo sviluppo di sistemi di traduzione automatica si ebbe nel secolo scorso. I primi tentativi degni di nota si ebbero infatti negli anni '30 quando Pëtr Smirnov-Troyanskii presentò in Russia un primo prototipo di traduttore automatico, e nel 1933 un ingegnere francese di origini armene, Georges Artsrouni, brevettò una macchina per tradurre dal nome "Mechanical Brain".

Gli storici della materia fanno risalire il vero inizio della traduzione automatica alle conversazioni ed alla corrispondenza che ebbero

luogo nel 1947 tra Andrew D. Booth, un cristallografo inglese, e Warren Weaver, direttore della divisione di scienze naturali della fondazione Rockfeller e più precisamente ad un memorandum scritto da Weaver nel 1949 per la fondazione Rockfeller, in cui, per la prima volta, vengono delineate le prospettive della traduzione automatica. Questo documento diede l'avvio alla ricerca sulla traduzione automatica: agli inizi degli anni '50, negli Stati Uniti e in Europa si formarono i primi gruppi di ricerca, che poterono beneficiare di finanziamenti di non poco conto da parte dei governi locali, e, con il progresso dell'informatica, si ebbero i primi risultati concreti e aumentò l'interesse sull'argomento. Nel 1952 si tenne la prima conferenza sulla traduzione automatica e nel 1954 ebbe luogo la prima dimostrazione pubblica di un sistema di traduzione automatica. Il sistema sviluppato da IBM e dalla Georgetown University negli Stati Uniti si basava su un vocabolario di 250 parole e solo sei regole sintattiche ed era in grado di tradurre in inglese un insieme selezionato di 49 frasi russe. Questa dimostrazione diede l'impulso ad un finanziamento su vasta scala alla ricerca sulla traduzione automatica negli Stati Uniti.

Da quel momento in poi si verificarono importanti progressi nella progettazione e nel funzionamento dei programmi di traduzione. Tra gli anni '50 e '60 assistiamo ad un gran fermento soprattutto a livello accademico: gruppi di ricerca negli Stati Uniti (Georgetown University, MIT, Università di Harward, Università del Texas e di Berkley), in Unione Sovietica (Istituto di linguistica di Mosca e Leningrado) e nel Regno Unito (Cambridge Research Unit) lavoravano alla realizzazione di prototipi per dimostrare la fattibilità della traduzione automatica. Tuttavia la qualità delle traduzioni lasciava ancora molto a desiderare e l'iniziale euforia cedette il passo a giudizi molto negativi sul futuro della traduzione automatica. Nel 1966 il Rapporto dell'Automatic Language Processing Advisory Committee (ALPAC) sulle prospettive della Traduzione Automatica non intravide nessuna utilità immediata e non ritenne necessari ulteriori investimenti e finanziamenti in questo campo.

Fortunatamente, la comunità scientifica in altri paesi non condivise questo giudizio così negativo e così si formarono altri gruppi di ricerca principalmente in Canada ed in Europa, dove il problema del multilinguismo era particolarmente sentito.

Nel 1976 in Canada venne sviluppato il sistema Meteo per tradurre i bollettini metereologici delle trasmissioni televisive. Contemporaneamente, la Comunità Europea divenne uno sponsor attivo della traduzione automatica, innanzitutto con adozione del sistema SYSTRAN per la traduzione di documentazione scientifica, tecnica, amministrativa e legale. Successivamente la Comunità Europea decise di finanziare l'ambizioso progetto EUROTRA, il cui scopo principale era la realizzazione di un sistema pre-industriale multilingue avanzato per la traduzione tra tutte le lingue della comunità europea basato su una struttura ad interlingua. Tuttavia il progetto non riuscì a produrre un vero sistema operativo cosicché si concluse alla fine degli anni '80. Anche se EUROTRA fallì nel suo scopo principale, ovvero la creazione di un sistema di traduzione multilingue da e in tutte le lingue europee, ebbe però il merito di stimolare la ricerca transnazionale nel campo della linguistica computazionale.

LO SVILUPPO NEGLI ANNI '80

Lo sviluppo più significativo degli anni '80 fu la diffusione dei primi prodotti commerciali di traduzione automatica come ad esempio il sistema LOGOS negli Stati Uniti, SYSTRAN in Francia, e METAL in Germania.

Alla fine degli anni '80 in Francia si ebbe anche la prima realizzazione di un servizio di traduzione automatica in rete ad uso del grande pubblico. Infatti nel 1988 il Servizio Postale Francese offrì su Minitel un servizio di traduzione automatica basato su Systran per diverse

coppie di lingue, disponibile agli utenti su terminali collegati alla rete delle Poste. Il servizio presentava però una serie di svantaggi: era infatti caro, relativamente lento e non integrato in ambiente PC.

Gli anni '90 furono caratterizzati da un pluralismo di orientamenti nella ricerca. Questo pluralismo si rifletté in una varietà di approcci che videro coesistere sistemi basati sulla semplice traduzione parola per parola o sull'approccio a transfer con sistemi che sperimentavano nuove teorie. Dalla metà degli anni '90 inoltre si ebbe un rapido incremento di numero e di tipologie di sistemi di traduzione disponibili: traduzione automatica, sistemi di traduzione assistita, ambienti di lavoro per il traduttore, memorie di traduzione, sistemi on-line forniti su Internet.

Da allora si è affermato un approccio più pragmatico e soprattutto più attento alle reali necessità degli utenti: è stato abbandonato il mito di una macchina capace di tradurre come un essere umano, per fornire al pubblico strumenti realmente utilizzabili e soprattutto realmente di ausilio al processo di traduzione.

La vera svolta per la traduzione automatica arriva con la diffusione di Internet e con la realizzazione di servizi di traduzione automatica on-line ad opera di alcuni produttori di software che avevano intuito quanto Internet potesse rappresentare un potente mezzo di pubblicità per i loro prodotti e servizi.

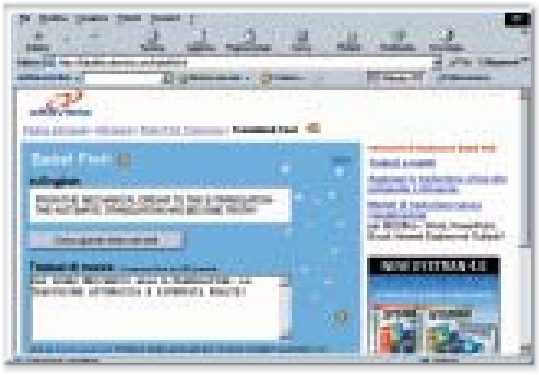

Figura 2. Interfaccia del sistema di traduzione BabelFish.

Nel 1996 venne pubblicato The Language Engineering Directory - A resource guide to Language Engineering Organisations, products and services, che conteneva i risultati di una ricerca condotta per delineare lo stato dell'arte nell'ambito dell'Industria delle Lingue. Nella sezione dedicata ai servizi di ingegneria linguistica nella categoria "Machine Translation via Modem/Minitel" sono enumerate ben sei società che già a quel tempo offrivano servizi di questo tipo: Compuserve Inc., Globalink Inc., Language Engineering Corporation, Nec Corporation - C&C IT Research Laboratorie, Smart Communications Inc. e infine Systran SA.

È proprio quest'ultima società che diede un impulso decisivo alla diffusione dei servizi di traduzione automatica on-line alleandosi con Altavista, il famoso motore di ricerca ed offrendo al grande pubblico il primo servizio gratuito di traduzione automatica on-line in tempo reale sul dominio noto ancora oggi come Babelfish, un concetto ripreso dal libro "The Hitchhiker's Guide to Galaxy" dell'autore di fantascienza Douglas Adams, in cui degli autostoppisti galattici riuscivano a capire ogni lingua, semplicemente attivando un piccolo pesce giallo nelle loro orecchie.

Questo primo esperimento divenne in brevissimo tempo un grande successo; come ci riferiscono Jin Yang ed Elke D. Lange, in due articoli sui servizi online offerti in BabelFish: il numero di richieste di Traduzione automatica aumentò da 500.000 nel maggio 1998 a 1,3 milioni di richieste al giorno nel 2000.

A partire dal 1997 si diffusero velocemente altri servizi gratuiti di TA on-line allo scopo di pubblicizzare i prodotti che erano alla base di tali servizi e quindi di creare un mercato richiamando l'attenzione del vasto popolo di internauti, affinché usassero e sperimentassero questo tipo di servizio.

In Hutchins J.& W. Hartmann (2003) Compendium of Translation Software Commercial machine translation systems and computer-aided translation support tools (European Association for Machine Translation

Sixth edition March 2003) sono enumerati ca. 55 servizi on-line di traduzione automatica, di cui ben 25 offrono coppie di lingue con l'italiano. Questi dati indicano chiaramente come Internet abbia costituito per la traduzione automatica un valido trampolino di lancio nella Società dell'Informazione, creando una sempre più ampia e diffusa richiesta di mercato per i servizi di Traduzione Automatica on-line. Paradossalmente però tale richiesta di mercato si orienta verso una tipologia di traduzioni che i sistemi attualmente in uso e commercializzati non sono preparati ad affrontare, dando risultati non sempre di buona qualità.

Internet ha cambiato il modo in cui il grande pubblico percepisce questa tecnologia ed il modo in cui la utilizza, aprendo alla traduzione automatica delle prospettive inaspettate, e probabilmente contribuirà in futuro anche a migliorarne la qualità.

I DIVERSI USI DELLA TRADUZIONE AUTOMATICA

Se dal punto di vista della ricerca non sono stati fatti grandi progressi negli ultimi anni e la qualità delle traduzioni prodotte dai sistemi di Traduzione Automatica non è, nella maggior parte dei casi, comparabile a quella delle traduzioni effettuate da traduttori professionisti, tuttavia è innegabile che il prodotto del processo di traduzione automatica, la traduzione "grezza" (raw translation) prodotta in tempi rapidissimi, è ormai largamente utilizzata per scopi diversi: come base per una traduzione finita, per facilitare l'accesso alle informazioni, per la comprensione di un testo in una lingua sconosciuta, per il rapido interscambio di informazioni.

La TA come "base" per una traduzione finita (Dissemination Tool). Si tratta dello scopo principale per cui vengono utilizzati i sistemi di traduzione automatica, inseriti all'interno di un ciclo completo di traduzione che ha come risultato la produzione di traduzioni pubblicabili. La traduzione automatica è in questo senso solo una fase di un più ampio processo (o progetto) di traduzione, suddiviso in più fasi: analisi della traduzione e reperimento del materiale di riferimento, aggiornamento del sistema di traduzione automatica, preparazione del testo da sottoporre a traduzione automatica (o pre-editing), la traduzione automatica vera e propria, revisione della traduzione automatica da parte di traduttori professionisti ed esperti nel settore di appartenenza del testo, controlli di qualità. I sistemi di TA commerciali (tra i quali ad esempio LOGOS, Systran, ecc.) sono stati progettati principalmente per rispondere a questo scopo, soprattutto nel settore delle traduzioni di testi di tipo tecnico-scientifico, ovvero di tipo non-letterario. Si tratta dunque della traduzione di testi "chiusi", per mutuare una definizione dal Lector in fabula di U. Eco (1979), ovvero di testi il cui scopo principale è comunicare al lettore delle informazioni che siano meno ambigue possibili e più precise possibili, come è il caso dei manuali di istruzione di un software, oppure dei manuali operativi di un aereo, caratterizzati da una sintassi semplice, da una alta ripetitività dei contenuti e semmai da una notevole complessità e densità dal punto di vista terminologico. Se opportunamente "addestrati" con la terminologia specifica di settore, i sistemi di traduzione automatica possono dare dei risultati qualitativamente non trascurabili ed economicamente vantaggiosi. Rispetto a questo scopo la traduzione automatica/assistita presenta diversi vantaggi relativamente, ad esempio, alla gestione della terminologia specializzata (linguaggi settoriali) ed alla gestione di grossi volumi di traduzione.

Nella traduzione tecnico-scientifica è necessario utilizzare in modo coerente e rigoroso la terminologia specifica all'interno dei testi, che a volte possono essere costituiti anche da migliaia di pagine, come nel caso dei manuali di un aereo. I sistemi di traduzione automatica garantiscono che tale terminologia, una volta immessa nel sistema, venga usata in modo coerente in

tutto il testo. Le potenzialità dei sistemi di TA, negli ultimi anni, sono state inoltre aumentate dal fatto che questi sono stati integrati all'interno di "postazioni di lavoro del traduttore" che includono altri strumenti di ausilio al traduttore quali ad esempio dizionari elettronici, memorie di traduzione, strumenti per il controllo della qualità della traduzione, ecc.

La TA per facilitare l'accesso alle informazioni (Information Access Tool): è la possibilità di utilizzare la traduzione automatica integrata in sistemi di Information Retrieval (ricerca e recupero di informazioni) su database testuali, o in sistemi di interrogazione di database strutturati.

La TA come strumento immediato di comprensione di un testo straniero (Assimilation Tool) di cui non si conosce la lingua. Questa applicazione della Traduzione automatica è sempre stata considerata storicamente un effetto derivato o secondario rispetto allo scopo principale che era quello appunto di produrre traduzioni grezze che fornissero la base per traduzioni pubblicabili di tipo tecnico-scientifico. È interessante notare come invece negli ultimi anni la diffusione di servizi di traduzione automatica on-line per la traduzione di pagine WEB abbia in qualche modo enfatizzato questa funzione della traduzione automatica e abbia contribuito a far conoscere questi sistemi al grande pubblico. Quotidianamente vengono effettuate milioni di richieste a servizi di traduzione automatica in tempo reale offerti ad esempio da siti come Babelfish allo scopo di conoscere i contenuti dei testi che circolano nelle più svariate lingue su Internet. La disponibilità di questo tipo di servizi ha consentito in qualche modo di abbattere le barriere linguistiche: la comunicazione su Internet è multilingue e se l'Inglese rappresenta ancora la lingua per eccellenza nel campo delle comunicazioni transnazionali, questa tendenza è destinata ad attenuarsi per cedere il passo all'uso delle lingue nazionali: gli utenti non anglofoni costituiscono infatti ben l'80% della popolazione mondiale e man mano che le nuove tecnologie di comunicazione saranno accessibili a questi utenti, il predominio dell'inglese è destinato a diminuire. Secondo le stime dell'IDC (International Data Corporation), gli utenti di Internet non anglofoni raggiungeranno entro il 2004 la quota del 70% della totalità della popolazione che si collega a Internet. Sono quindi necessari strumenti che facilitano la comunicazione multilingue, consentendo agli utenti di accedere alle informazioni presenti su Internet, utilizzando la propria lingua: in tal senso gli strumenti di TA rappresentano un valido aiuto.

La TA per il rapido interscambio di informazioni (Interchange Tool): in contesti in cui lo scambio di informazioni tra persone che parlano lingue diverse deve avvenire in tempo reale, come ad esempio nelle discussioni delle chatroom, la traduzione intesa in senso tradizionale ovvero il ricorso a traduttori professionisti è assolutamente fuori discussione. La traduzione automatica rappresenta in questi contesti l'unica soluzione possibile, riuscendo ad offrire traduzioni in tempo reale in svariate lingue.

Oltre a questi usi della traduzione automatica, che ormai potremmo definire tradizionali, Jin Yang ed Elke D. Lange in un loro articolo sui servizi di Traduzione Automatica offerti su BabelFish da Systran, riscontrano due ulteriori usi da parte degli utenti:

La TA come strumento per di trattenimento (Entertainment Tool): l'uso più diffuso della TA su Internet sembra essere la traduzione da una lingua di partenza (ad esempio l'italiano) ad una lingua di arrivo (ad esempio l'inglese) e la successiva ritraduzione nella lingua di partenza (di nuovo l'italiano) o la traduzione in un'altra lingua ancora. Si tratta di un uso assolutamente sconsigliabile dei servizi di TA, soprattutto se in questo modo si intende valutare la qualità delle traduzioni. Il sito Lost in tran-

slation (http://www.tashian.com/multibabel/) utilizzando il dominio di Babelfish propone un gioco di traduzione, che riecheggia il più antico "telefono senza fili", a partire da un testo inglese, che una volta tradotto in una lingua viene ritradotto in un'altra e così via fino a ritornare alla lingua di partenza: il risultato è a dir poco fantasmagorico e sicuramente ci ricorda la massima di Antoine de Saint-Exupéry: "Language is the source of misunderstandings".

La TA come strumento per l'apprendimento di una lingua straniera (Learning Tool): gli strumenti di traduzione automatica online vengono spesso utilizzati dagli studenti, principalmente di scuola superiore, per svolgere i loro compiti a casa di lingua straniera.

CONCLUSIONI

Nell'era digitale la Traduzione Automatica ha trovato la sua ragion d'essere, avendo abbandonato le velleità di un sogno impossibile da realizzare, e cioè di essere "macchina pensante" e di produrre dei risultati comparabili al pensiero umano, per diventare strumento, pur con tutti i suoi limiti, al servizio di una efficace comunicazione multilingue. Internet e la traduzione automatica hanno stretto una alleanza indissolubile, diventando indispensabili l'uno per l'altra: Internet ha infatti contribuito a far conoscere la traduzione automatica e l'ha resa, al di là di ogni ragionevole aspettativa, utilizzabile da parte del grande pubblico e utile ai fini dell'abbattimento delle barriere linguistiche nel villaggio globale. La traduzione automatica ha invece reso Internet uno strumento di comunicazione e di reperimento delle informazioni più che mai efficace.

JOHANNA MONTI
Università degli Studi di Napoli "L'Orientale"

Traduzione automatica: storia, situazione e prospettive

Il TAL (Trattamento Automatico del Linguaggio) è molto più di una sigla: significa un impegno, a nostro avviso molto importante, per coordinare e rilanciare le molteplici potenzialità (e necessità) di contribuire con ricerche e sviluppi informatici al superamento delle barriere linguistiche, e più in generale al miglioramento della comunicazione e della conoscenza.

La Traduzione Automatica, cioè la possibilità di affidare a un computer la traduzione completa di un testo, riducendo al minimo la successiva e comunque necessaria revisione da parte dell'utente, è una parte importante del TAL, anche se è stata per molti anni un'utopia, e una sfida per le attività di ricerca nelle discipline coinvolte (Informatica, Linguistica Computazionale, Intelligenza Artificiale, Scienze Cognitive).

La **"storia"** della Traduzione Automatica inizia negli anni '50, e quei tentativi iniziali furono caratterizzati da un eccessivo ottimismo nelle possibilità degli elaboratori di allora, e da una grave sottovalutazione delle complessità connesse con i problemi di traduzione, complessità che ancora oggi devono essere sempre ben presenti per un approccio serio e realistico.

Verso la fine degli anni '60 una valutazione dei risultati ottenuti, praticamente inutilizzabi-

li nonostante il rilevante impegno di risorse su scala mondiale, portò ad una situazione di pessimismo generalizzato sulle possibilità della Traduzione Automatica, e alla conseguente interruzione delle attività in quest'area.

Dopo un black-out durato più di venti anni, e partendo da una tecnologia informatica che nel frattempo aveva visto progressi sorprendenti, la ricerca e sviluppo per la Traduzione Automatica ha ripreso un ruolo significativo. Ma questa "rinascita" non è dovuta soltanto alle nuove possibilità fornite dall'hardware e dal software: da parte di tutti i gruppi di ricerca seri è cambiato radicalmente l'approccio, diventato interdisciplinare, consapevole delle complessità e orientato verso prodotti effettivamente usabili.

Questo vuol dire fra l'altro impiegare architetture e tecniche linguistiche adeguate, per certi aspetti molto complesse, se si vogliono realizzare sistemi di vera Traduzione Automatica, e non solo strumenti di aiuto basati su dizionari o con traduzione parola per parola.

D'altra parte è anche essenziale chiarire agli utilizzatori i "limiti", attuali ma anche futuri, di questi sistemi: la traduzione completamente automatica di qualsiasi tipo di testo è, e resterà, un'utopia.

La validità di un sistema di Traduzione Automatica è sempre associata a una successiva revisione umana, e dipende da quanto il sistema riesce a tradurre in modo completamente corretto o con necessità di correzioni semplici, garantendo in aggiunta una correttezza e coerenza delle terminologie tecniche.

Il mercato mondiale

Il mercato mondiale, e le comunicazioni internazionali sempre più diffuse, richiedono strumenti di questo tipo, sia per traduzioni completamente automatiche non perfette ma comprensibili di comunicazioni (ad esempio di posta elettronica), sia per aumentare la produttività e la qualità delle traduzioni tecniche, permettendo al traduttore professionista di concentrarsi solo sulla revisione degli aspetti più complessi che il sistema automatico non è stato in grado di risolvere completamente.

Questi sistemi integrano complessi componenti linguistici e adeguate funzionalità per l'interfaccia utente, in modo da permettere la traduzione di comunicazioni e di documentazioni tecnico/commerciali, ottenendo un testo tradotto automaticamente, con opportune funzioni di aiuto per la successiva revisione e con la ricostruzione automatica di tutte le caratteristiche di struttura del testo originale (font, tabelle, figure, ecc.).

L'utilizzatore tipico può essere sia chi, in ambiente individuale o di ufficio, ha necessità di tradurre documenti, sia il traduttore professionista che opera in settori tecnico/commerciali (informatica, meccanica, medicina, finanza, ecc.), che con l'aggiunta di opportuni dizionari specialistici può ottenere elevati aumenti di produttività e qualità.

Più in generale, per affrontare correttamente i problemi della Traduzione Automatica con strumenti informatici è necessario partire dalla consapevolezza che si tratta di un processo spesso molto complesso, e che per avere risultati accettabili il sistema informatico deve avere una "architettura" opportuna. Nella Figura 1 è rappresentato uno schema concettuale di riferimento, che permette di identificare le modalità e le architetture per la traduzione.

L'architettura di tipo "diretto" prevede appunto il passaggio diretto dalla frase nella lingua di partenza a quella di destinazione, con l'utilizzo di dizionari mono e bilingue, con traduzione in pratica "parola per parola". Questa architettura era usata dai primi sistemi "storici" (ed ancora oggi in sistemi molto semplificati), e si è rivelata assolutamente inadatta per ottenere risultati di vera e propria Traduzione Automatica.

L'architettura di tipo "transfer" (descritta più in dettaglio nel seguito) affronta invece i principali aspetti linguistici necessari, in modo

modulare, e prevede l'utilizzo di "grammatiche automatiche" per le singole lingue, "grammatiche contrastive" per la coppia di lingue, e dizionari mono e bilingue opportunamente arricchiti con informazioni semantiche e di contesto, per permettere la disambiguazione. Questa architettura è quella ad oggi più usata dai sistemi di Traduzione Automatica realmente usabili.

L'architettura di tipo "interlingua" prevede invece un unico linguaggio intermedio interno (una specie di "esperanto informatico"), sufficientemente formale e con capacità di rappresentazione dei significati, a cui ridurre le frasi da tradurre e da cui partire per ottenere le corrispondenti frasi tradotte. Questa architettura è interessante dal punto di vista della ricerca, ma di difficile realizzazione per sistemi praticamente usabili.

I principali componenti di un sistema di vera Traduzione Automatica, basato sull'architettura di tipo "transfer", sono:

- **Analisi del testo da tradurre**. In base ad un'opportuna grammatica automatica della lingua di partenza, e al relativo dizionario monolingua, un analizzatore morfo/sintattico ("parser") prende in considerazione il testo da tradurre e produce in modo formale tutte caratteristiche linguistiche necessarie per la successiva traduzione. In pratica ogni frase viene "compresa", per quanto possibile, nei suoi aspetti morfologici, sintattici e semantici.
- **Transfer strutturale**. La struttura sintattica di ogni frase da tradurre, risultato della fase di analisi, non ha in genere una identica struttura per la lingua in cui deve essere tradotta. Il sistema deve quindi contenere una "grammatica contrastiva" automatica, e le relative funzioni di trasformazione degli alberi sintattici.
- **Transfer lessicale**. Oltre alla trasformazione delle strutture sintattiche, è necessaria la traduzione appropriata di tutte le parole che compongono la frase. In generale una parola in una lingua può avere diverse traduzioni a seconda della collocazione sintattica e del contesto. Questo aspetto è uno dei più complessi dal punto di vista della scelta automatica, e richiede che il sistema sia dotato di un opportuno dizionario bilingue per poter effettuare automaticamente di volta in volta la traduzione giusta.
- **Generazione (sintesi) del testo tradotto**. Per ottenere la traduzione completa della frase è necessario che il sistema sia in grado di co-

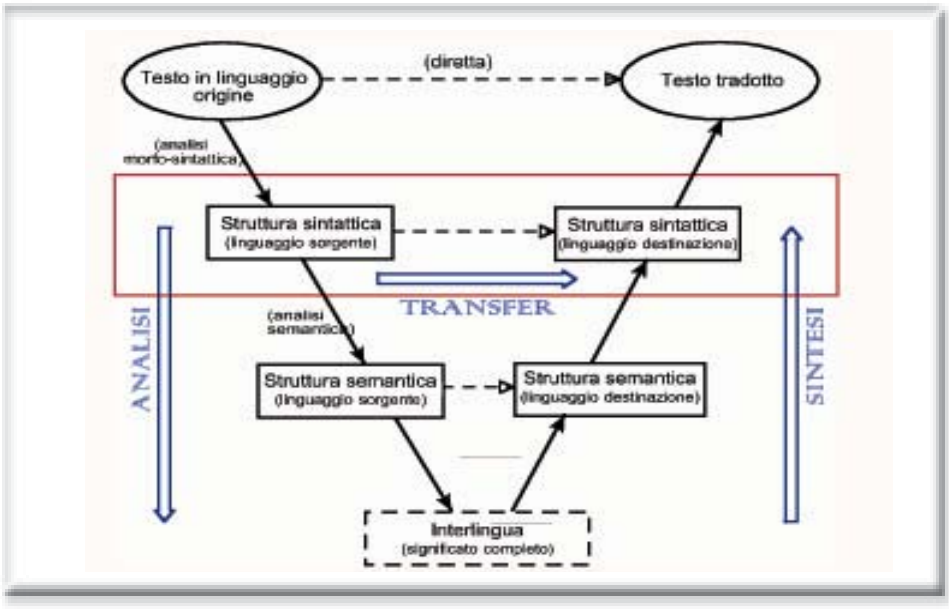

Figura 1. Traduzione Automatica: schema concettuale.

struire tutte le forme flesse appropriate (genere/numero, tempo/modo/persona, ecc.) insieme a tutti gli aspetti caratteristici della lingua di destinazione (ad esempio, per l'italiano, concordanze, troncamenti, preposizioni articolate, pronomi enclitici, ecc.). Per far questo il sistema deve avere un opportuno dizionario monolingua, e la relativa grammatica generativa automatica.

- **Interfaccia utente**. In aggiunta ai componenti precedenti (il cosiddetto "motore linguistico"), un buon sistema di traduzione automatica deve fornire all'utente, in modo semplice ed amichevole, tutte le funzioni operative necessarie, sia per le varie modalità di traduzione sia per le attività specifiche di consultazione, creazione e aggiornamento dei dizionari aggiuntivi di utente.

Un'architettura di questo tipo, e una appropriata attenzione alla "copertura" dei componenti linguistici, sono requisiti essenziali per sistemi di vera Traduzione Automatica.

Per evitare comunque aspettative non realistiche è importante ribadire che per avere una traduzione completamente corretta è quasi sempre necessaria una revisione umana. La validità di un sistema automatico si misura appunto sull'entità di questa revisione: se il testo tradotto automaticamente ha un'alta percentuale di frasi corrette o che richiedono semplici modifiche, allora si può definire efficace e usabile in modo produttivo.

L'APPROCCIO STATISTICO

Un approccio completamente diverso da quelli descritti in precedenza è il cosiddetto "approccio statistico", che da alcuni anni è oggetto di ricerca (e anche di polemiche fra gli esperti di traduzione automatica). In termini molto semplificati, si parte dal concetto che invece di "insegnare" al computer gli aspetti lessicali, sintattici e semantici di una coppia di lingue (attività molto complessa e molto costosa), si fornisce al computer un corpus (molto grande) di testi "paralleli", tradotti correttamente, e tramite opportuni software statistici il computer "impara" a tradurre per quella coppia di lingue.

Con questo approccio si ha un risultato certamente di minor qualità e accuratezza, ma con software ben fatto si può ottenere una traduzione che fa capire il significato del testo originale, pur con una correttezza grammaticale molto bassa. I vantaggi sono nel tempo/costo di realizzazione molto più bassi dei sistemi con approccio linguistico, purché i corpus paralleli di partenza siano opportunamente vasti e corretti.

Attualmente è opinione abbastanza diffusa che l'approccio statistico, anziché come alternativa, possa essere utile per integrare quello linguistico.

Per concludere, alcune riflessioni sulla situazione in Italia. Purtroppo esistono pochissime aziende italiane in grado di sviluppare prodotti di vera traduzione automatica, sia perché le complessità e i costi di sviluppo sono molto elevati, sia perché a nostro avviso è ancora carente l'iniziativa per lanciare questo settore con opportune strategie di politica economica.

In questo ristretto panorama, chi scrive ha fondato 10 anni fa (come "spin off" del Centro di Ricerca IBM) un'azienda informatica (SYNTHEMA) che appunto sviluppa prodotti di vera Traduzione Automatica. L'approccio fin qui adottato è quello linguistico, in particolare con architettura di tipo transfer sintattico/semantico, ma sono in corso esperimenti per una integrazione con l'approccio statistico. Esistono da tempo sul mercato nostri prodotti "consumer", di basso costo ma di qualità elevata, come anche "soluzioni" aziendali per intranet, che SYNTHEMA può personalizzare sia dal punto di vista terminologico che linguistico, in modo da raggiungere risultati qualitativi molto elevati: la personalizzazione, aziendale o di settore, è a nostro avviso una delle chiavi determinanti per il successo di questi sistemi.

CLAUDIO CIRILLI *SYNTHEMA s.r.l.*

Comunicare con le tecnologie del linguaggio nell'era globale

"Impara il tedesco!", "Non puoi fare a meno dell'inglese!", "È giunta l'ora del cinese!", "Ma perché non impari una lingua araba, troverai sicuramente un lavoro!" Tutte queste sollecitazioni, questi slogan, che leggiamo sui giornali e sentiamo nelle nostre discussioni quotidiane esprimono bisogni, necessità ed aspirazioni, che negli ultimi anni, si sono accentuate soprattutto in seguito ai cambiamenti avvenuti nell'Est Europeo ed in Asia.

L'internazionalizzazione dei rapporti dovuta all'aumento impressionante sia della circolazione delle merci che delle persone, alle grandi migrazioni e alla diffusione del Web, stanno creando una nuova necessità di comunicazione interpersonale e di accesso ad informazioni e documenti che richiede una diffusa conoscenza delle lingue. Dal primo di maggio del 2004 dieci nuovi Stati ed altrettante lingue si sono aggiunte all'Unione Europea con tutte le relative incombenze: predisposizione di tutta la modulistica europea, preparazione dei resoconti del parlamento europeo, sottotitoli e/o traduzione dei programmi televisivi, ecc.

Questo nuovo contesto all'interno del quale intere popolazioni trovano nuove opportunità di sviluppo ed altre cercano di difendere posizioni acquisite, impone da parte degli stati e dei governi delle scelte molto forti sia per far fronte all'immediato sia per giocare un ruolo strategico nel futuro. Analizzando in modo sommario alcune tendenze in atto si notano due fenomeni, da una parte un grande sforzo formativo di apprendimento delle lingue: in particolare l'apprendimento diffuso dell'inglese in Europa ed in Asia e l'apprendimento del cinese e delle lingue arabe in occidente in modo molto più selezionato. Il secondo fenomeno è quello dell'investimento in ricerca per lo sviluppo di sistemi di supporto all'apprendimento del linguaggio e di sofisticati sistemi di elaborazione linguistica dell'informazione contenuta in forma elettronica in documenti, giornali, notizie televisive, in generale nel Web.

Relativamente alla prima linea di tendenza, in Europa la scuola si fa carico di cambiare i propri programmi indicando nuove priorità formative in funzione delle lingue europee e cercando di migliorare il metodo di apprendimento per aumentare l'efficacia dei risultati nella comunicazione interpersonale; in Asia, e principalmente in Cina, si stanno attuando dei programmi di massa per l'apprendimento dell'inglese basato sull'utilizzo dei nuovi strumenti informatici di supporto all'apprendimento delle lingue. In India, invece, la popolazione non ha il problema di conoscere l'Inglese sia per il suo trascorso coloniale che per l'impostazione del suo sistema formativo, ma ha delle difficoltà a rendere visibile nel Web le proprie lingue.

Relativamente alla seconda linea di tendenza, gli Stati Uniti hanno come priorità la conoscenza delle lingue arabe e del cinese allo scopo di conoscere per fini economici e politici tutte le informazioni ed i documenti prodotti in queste lingue. L'Europa, al contrario, ha come esigenza primaria la salvaguardia di tutte le sue lingue e quindi la necessità di renderne sostenibile il costo, in quanto gestire con pari dignità una trentina di lingue potrebbe diventare un'impresa impossibile se interamente basata sul puro lavoro umano.

Se rimaniamo sul terreno dell'informazione, e non consideriamo l'aspetto sociale della comunicazione, oggi notiamo che gran parte

dell'informazione è elettronica e questo ci permette di elaborarne anche enormi quantità in modo molto efficiente. L'informazione è codificata nelle varie lingue e possiamo pensare quindi di elaborare il linguaggio in modo automatico, per estrarre informazioni, per trovare documenti, per tradurre anche in modo sommario delle notizie o dei notiziari televisivi, per poter comunicare, anche in modo primitivo e semplice con altre persone che non parlano la nostra lingua, allo scopo di ottenere informazione.

LA RICERCA SULLA TRADUZIONE E LA COMUNICAZIONE MULTILINGUA

La traduzione del linguaggio parlato (SLT) rappresenta un'area di ricerca piuttosto recente nell'ambito delle tecnologie sul linguaggio umano. Le date più importanti nella storia recente della SLT sono:
– il 1986: data di inizio in Giappone del progetto di ATR sulla traduzione del linguaggio parlato;
– il 1992: data di nascita del consorzio C-STAR <http://www.c-star.org>;
– il 1993: il governo Tedesco avvia il progetto nazionale VERBMOBIL <http://verbmobil.dfki.de/overview-us.html>;
– il 2000: DARPA lancia TIDES<http://www.darpa.mil/ipto/Programs/ tides/index.htm>;
– il 2004: anno in cui l'Unione Europea ha iniziato a finanziare il progetto TC-STAR <http://tc-star.org>.

Analizzando questi ultimi venti anni, si nota che la ricerca della traduzione del linguaggio parlato ha seguito il percorso tipico di una nuova area di ricerca nel settore delle tecnologie dell'informazione e della comunicazione. Nei primi dieci anni, infatti, sono stati realizzati progetti che hanno dimostrato la fattibilità di alcune applicazioni: interprete telefonico, e-commerce, comunicazione faccia a faccia e solo recentemente, dopo un ulteriore decennio, si è giunti alla definizione di una vera e propria agenda di ricerca, caratterizzata da obiettivi scientifici, domini applicativi, metodi ed approcci e criteri di valutazione. Un ruolo chiave in questo percorso è stato svolto dal consorzio C-STAR e dal Progetto Verbmobil. Il primo per aver dimostrato la fattibilità delle tecnologie SLT attraverso due principali dimostrazioni su scala planetaria nel 1995 e nel 1999. Il secondo, per aver proposto lo scenario applicativo della comunicazione multilingua faccia a faccia, e per aver iniziato ad esplorare diversi approcci alla soluzione del problema della traduzione: alcuni

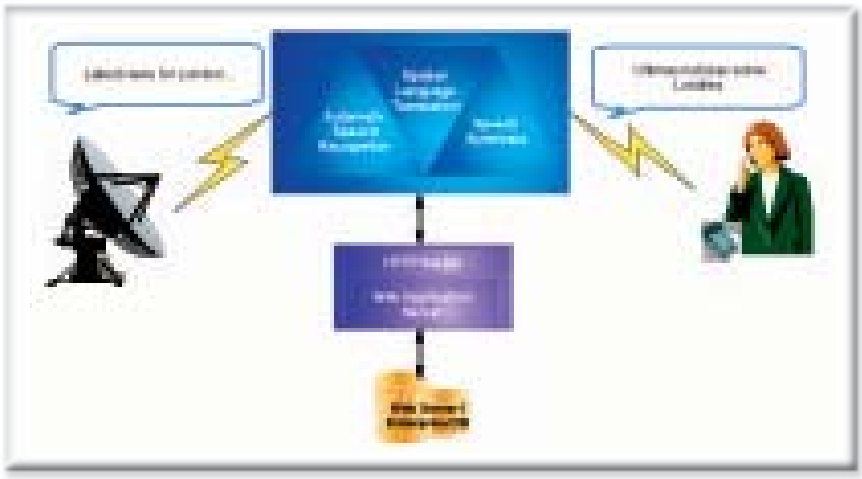

Figura 1. Uno scenario applicativo della traduzione, poter ascoltare le notizie ovunque nella propria lingua.

basati su metodi di intelligenza artificiale, quali il trasferimento semantico e l'interlingua, altri basati su metodi di apprendimento dai dati "data-driven", quali la traduzione basata su esempi e la traduzione statistica.

È importante notare che lo spostamento da attività basate sulla dimostrazione di fattibilità ad attività basate sulla valutazione tecnologica è stato il percorso seguito dalla comunità scientifica del riconoscimento automatico del parlato (ASR). È abbastanza naturale aspettarsi che la traduzione da parlato a parlato (SLT) segua un percorso analogo. Programmi come TIDES, il nuovo progetto europeo integrato TC-STAR e le attività portate avanti all'interno del consorzio C-STAR III <http://www.slt.atr.co.jp/ IWSLT2004/>, dimostrano che questo settore di ricerca sta affrontando una nuova fase di maturità caratterizzata da:

- un consenso diffuso, su un approccio "data-driven", dove i metodi statistici, combinazione di modelli di linguaggio e di teorie della decisione statistica, sembrano essere i più promettenti;
- l'adozione di una organizzazione della ricerca basata sul paradigma di competizione e di cooperazione. Ciò significa mettere a fuoco principalmente obiettivi comuni, utilizzando dati comuni, ma significa anche mettersi in competizione sui metodi e sugli algoritmi. Questo processo ha l'obiettivo di accelerare lo sviluppo della ricerca, come è già stato sperimentato all'interno della comunità ASR, soprattutto se la competizione viene accompagnata da una forte cooperazione incentrata sull'adozione di una metodologia open source (un'esperienza interessante è portata avanti dalla Comunità della bioinformatica);
- una cooperazione scientifica globale che coinvolga tutti i principali centri al mondo che partecipano a questa grande sfida. Non ci sarà avanzamento nel campo senza il coinvolgimento di una massa critica di ricercatori e senza la disponibilità di una grande quantità di dati allineati in molte lingue.

I programmi di ricerca al mondo sulla traduzione

Attività di SLT in Asia

In Asia, la ricerca in SLT è stata per molti anni un settore importante. I laboratori ITL del Advanced Telecommunications Research Institute International (ATR) <http://www.atr.jp> di Kyoto in Giappone sono stati costituiti 1986 e sono stati il primo laboratorio che ha lavorato in Asia sulla traduzione del parlato. L'ATR inoltre è stato uno dei principali sponsor del consorzio C-STAR. Nella prima fase della ricerca (dal 1986 al 1992), l'obiettivo era finalizzato a studiare la fattibilità di tecnologie di traduzione del linguaggio. Il primo esperimento internazionale congiunto di interpretazione della telefonia tra Giappone, USA, e Germania è stato condotto con successo nel gennaio del 1989. Il linguaggio usato era molto semplice e permetteva all'utente di usare solo frasi grammaticalmente corrette e pronunciate in modo controllato, in un ambiente privo di rumore significativo.

Nella seconda fase, dal 1993 al 1999, l'obiettivo di ricerca si è spostato verso lo studio del dialogo naturale multilingua. Il sistema doveva trattare frasi tipiche di parlato, anche non grammaticalmente corrette e pronunciate in modo naturale. A partire dal 1997 sono stati sviluppati sistemi sperimentali e prototipi che traducono dal giapponese in coreano, in tedesco, in inglese e cinese. Nel 1998 è stata creata la prima piattaforma bidirezionale giapponese-inglese: ATR-MATRIX. Un secondo esperimento congiunto internazionale di comunicazione multilingue è stato eseguito con successo nel luglio del 1999. Nello stesso anno ATR ha realizzato anche una piattaforma semplificata per l'uso di un pc portatile o di un cellulare. A partire dal 2000 è iniziata la terza fase di ricerca, il cui obiettivo primario è l'estensione della tecnologia alla comunicazione multilingua senza vincoli di pronuncia e di contesti per l'utente ed in grado di ridurre il rumore ambientale.

Il primo progetto di SLT finanziato dal governo cinese, <http://www.nsfc.gov.cn/>, è stato sviluppato dall'Istituto di Automazione dell'Accademia delle scienze cinese (CAS-IA, < http://www.ia.ac.cn/ > dal gennaio 1999 a dicembre 2002. In questo progetto sono state avviate le attività di base per la realizzazione di sistemi SLT.

Nel 2002, il China High-Technical Program (programma 863) <http://www.863.org.cn> ha finanziato un importante progetto denominato "Digital Olympic". I responsabili del progetto sono Beijing Capital Information (Cap Info) Co. Ltd. <http://www.capinfo.com.cn> e CAS-IA. Questo progetto intende sviluppare le tecnologie di base per realizzare un sistema intelligente di gestione dell'informazione per i giochi olimpici 2008 a Pechino. La ricerca include tecnologie del linguaggio, tra cui la traduzione automatica delle pagine Web, i sistemi SLT, i sistemi di reperimento delle informazioni, e servizi d'informazione basati su terminali mobili.

Per la traduzione multilingua SLT, CAS-IA, sta raccogliendo i dati e sviluppando i componenti utilizzando un approccio interlingua. La prima versione del sistema SLT, sviluppato in collaborazione con l'Università Carnegie Mellon di Pittsburgh, è stata presentata con successo nell'esposizione delle scienze e tecnologie tenutasi dal 22 al 26 Maggio 2004 a Pechino. Nel luglio dello stesso anno, il sistema è stato presentato a Barcellona al forum mondiale di cultura. CAS-IA ha anche ottenuto significativi risultati per applicazioni della tecnologia SLT che usano computer palmari.

In Corea, le attività di ricerca in questo settore, sono sviluppate principalmente dall'ETRI <www.etri.re.kr>, il Centro Ricerche della Compagnia Nazionale di Telecomunicazione, all'interno del programma nazionale "Tecnologie per l'elaborazione delle informazioni linguistiche". ETRI ha già dimostrato, nell'ambito del consorzio C-STAR la fattibilità delle tecnologie di traduzione dal coreano alle lingue asiatiche ed alle lingue occidentali. L'obiettivo attuale è quello di riuscire a portare sul mercato semplici sistemi di traduzione che aiutano i viaggiatori all'estero. ETRI è inoltre impegnata nel settore della trascrizione e della traduzione delle notizie televisive dal coreano al cinese.

Orientati allo sviluppo di sistemi SLT per i giochi olimpici di Pechino del 2008, CAS-IA, ATR ed ETRI hanno firmato recentemente un accordo di cooperazione per lo sviluppo di sistemi SLT in grado di tradurre richieste e conversazioni turistiche in giapponese, cinese e coreano.

Le attività negli Stati Uniti

Nonostante la ricerca sulla traduzione del parlato sia iniziata negli Stati Uniti negli anni novanta, il riconoscimento della sua importanza strategica è stato molto graduale. Ciò è sicuramente legato al dominio dell'inglese come la lingua usata per gli affari internazionali, la scienza ed il commercio in tutto il mondo. Di conseguenza i sistemi SLT sono stati considerati con una priorità secondaria per lo sviluppo commerciale e scientifico. I recenti eventi di politica internazionale, con il coinvolgimento degli Stati Uniti a livello globale, gli scambi e gli affari internazionali, e per ultimi ma non meno importanti i cambiamenti demografici hanno cambiato drammaticamente questa percezione. Mentre per i parlatori nativi inglesi è generalmente più difficile studiare altre lingue (per mancanza di opportunità in un'ampia società monolingue), la necessità di conoscere le lingue straniere è diventata sempre più stringente.

Il bisogno di un facile e veloce accesso a documenti ed informazioni multimediali (che non sono necessariamente in inglese) appartenenti a lingue straniere è stato riconosciuto sia dal mondo degli affari che dal Governo Federale. Inoltre la possibilità di comunicare con parlatori non-inglesi in situazioni di in-

terventi umanitari e militari viene considerata oggi estremamente importante. Questo bisogno sta diffondendosi anche all'interno degli Stati Uniti per rispondere in modo più efficace alle emergenze e ai bisogni delle persone in un paese, in cui la composizione linguistica sta rapidamente cambiando e diventando più varia.

In vista di questi cambiamenti, alcune nuove iniziative di ricerca sono state lanciate negli Stati Uniti. Tre progetti in corso di svolgimento mirano ad affrontare questi problemi.

DARPA TIDES-TIDES ha l'obiettivo realizzare sistemi intelligenti per l'accesso, l'estrazione e la produzione di riassunti a partire da testi multilingua. I sistemi di traduzione da testo (MT) sono una delle più grandi sfide all'interno di TIDES, con una grande aspettativa di produrre nuovi importanti avanzamenti dopo due decadi di progresso limitato in questo settore. TIDES non include il linguaggio parlato, ma le tecnologie di base di MT sviluppate sono molto importanti anche per la ricerca nella traduzione del linguaggio parlato (SLT). L'obiettivo finale di TIDES è l'estrazione di contenuto di interesse civile e/o militare da testi che possono essere scritti in diverse lingue. Le lingue affrontate sono inglese, cinese ed arabo. Vengono inoltre sviluppati esperimenti su nuove lingue al fine di esplorare la portabilità della tecnologia. I laboratori di ricerca che partecipano alla valutazione annuale in MT sono: CMU, IBM, ISI, JHU, RWTH, ITC-irst.

DARPA BABYLON - I progetti BABYLON e la sua prosecuzione, CASTE, affrontano il problema della comunicazione multilingua tra parlatori inglesi e stranieri. Gli scenari sviluppati includono il trattamento dei rifugiati nelle aree di intervento, interviste mediche, e situazioni di controllo di polizia. Le lingue includono: arabo, cinese, Pashtu, Farsi e Thai. I sistemi che si stanno sviluppando affrontano un dominio linguistico limitato, sono bidirezionali ed utilizzano un approccio basato sull'Interlingua (vedi figura 2). Alcune applicazioni possono essere anche solo monodirezionali (esempio un medico americano che vuol comunicare con dei rifugiati, oppure un medico che fa le stesse interviste sulla salute a molte persone, in cui la traduzione dall'inglese avviene attraverso frasi e concetti standard). I sistemi di Babylon sono anche disponibili su computer palmari, al fine di permetterne l'uso in situazioni di mobilità. Un esempio è dato dal Sistema Phraselator, rappresentato in figura 3.

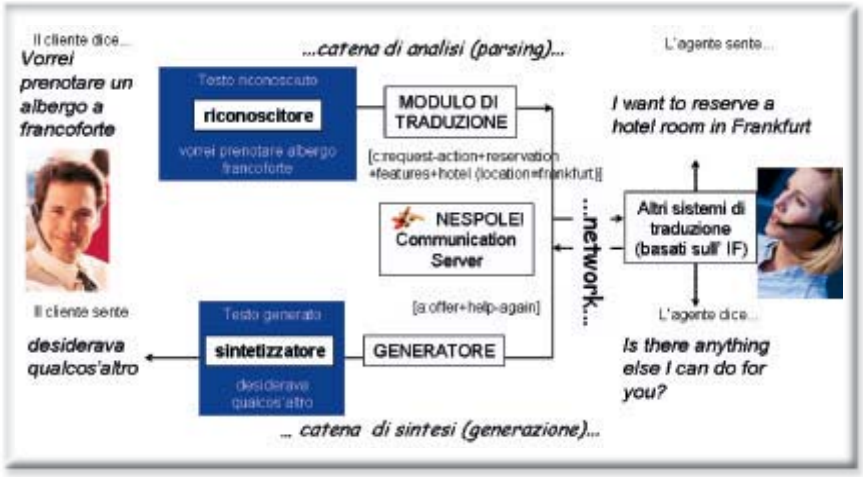

Figura 2. L'architettura del progetto NESPOLE! basata su un approccio interlingua.

I laboratori che partecipano nella traduzione del linguaggio parlato in Babylon sono: BBN, CMU, IBM, ISI/HRL, SRI.

NSF STR-DUST - STR-DUST è un'iniziativa supportata dalla National Science Foundation (NSF) della durata di 5 anni che tenta di andare oltre la traduzione del parlato in domini del discorso limitati e di indagare la traduzione del parlato senza vincoli linguistici. Il progetto si svolge all'Università Carnegie Mellon e le lingue coinvolte sono il cinese, l'arabo e l'inglese. Il progetto è appena iniziato ed è stato preparato in collaborazione con quello europeo TC-STAR.

In aggiunta a queste iniziative di ricerca finanziate dalle agenzie governative, sono in fase di sviluppo alcuni prodotti commerciali per turisti e per personale militare, basati su computer palmari (Ectaco, Marine Acoustics). Questi dispositivi non forniscono una traduzione bidirezionale completa, ma consentono l'accesso vocale a liste di frasi tipiche, come quelle che si trovano sulle guide turistiche, frasi che vengono poi lette da un sintetizzatore nella lingua dell'interlocutore.

Questi strumenti, seppure molto semplici, rappresentano il primo passo verso il superamento delle barriere linguistiche in mobilità.

LE ATTIVITÀ IN EUROPA

La lingua costituisce un tema cruciale nella costruzione dell'Unione Europea, considerate le forti implicazioni di natura sociale, politica ed economica che essa comporta. L'Unione Europea si fonda infatti sul riconoscimento e la valorizzazione della diversità linguistica. Lo sforzo necessario per affrontare questo problema è troppo grande per essere sostenuto solamente dalla Commissione e viene quindi condiviso tra la Commissione stessa, che ha il dovere di assicurare una buona comunicazione con gli stati membri, e gli stati membri che devono preservare e promuovere le loro lingue e, attraverso le loro lingue, la loro cultura. Per questa ragione la Commissione, nel V e nel VI Programma Quadro, ha identificato le Tecnologie del Linguaggio Umano (HLT) come un obiettivo strategico.

In questo momento sono attivi tre progetti: LC-STAR <www.lc-star.com>, PF-STAR <http://pfstar.itc.it>e TC-STAR <www.tc-star.org>. Precedentemente i progetti più importanti portati avanti in Europa sono stati, Verbmobil I e II finanziati dal governo tedesco, EUTRANS e NESPOLE! finanziati dalla Commissione Europea. NESPOLE!, http://nespole.itc.it, conclusosi nel 2002 è stato cofinanziato dall'UE e dalla NSF degli Stati Uniti.

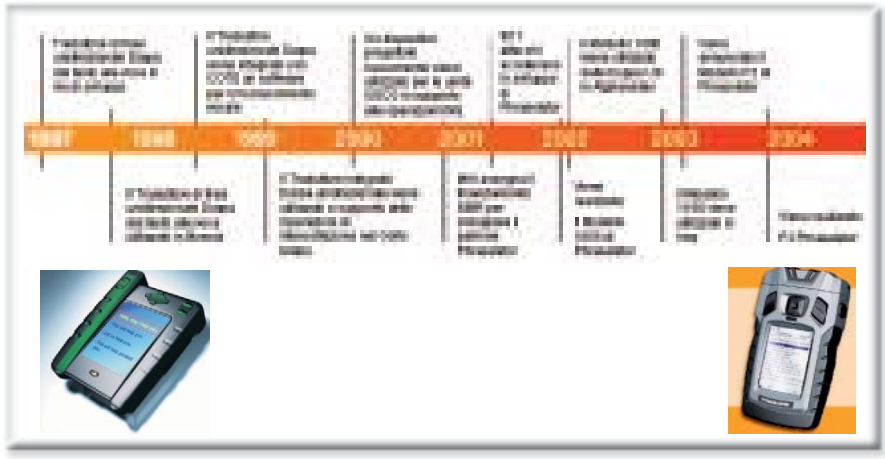

Figura 3. PHRASELATOR, uno dei risultati del progetto Babylon.

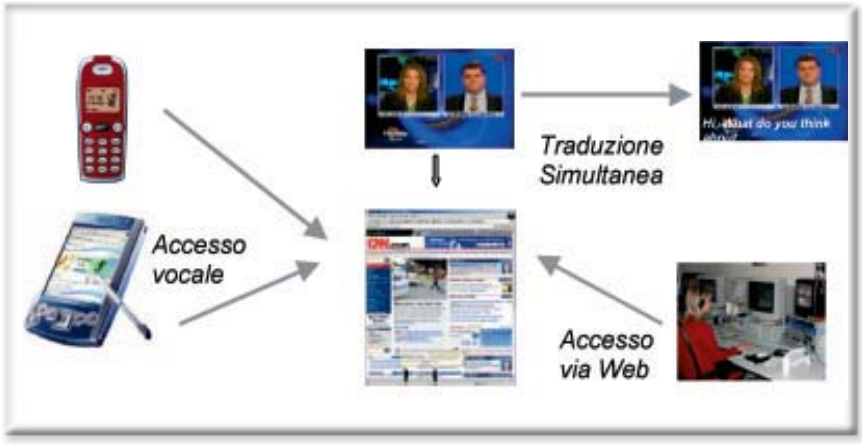

Figura 4. TC-STAR, traduzione di notiziari televisivi e di discorsi.

LC-STAR, è iniziato ufficialmente il primo febbraio del 2002 ed è un progetto che si focalizza sulla creazione di risorse linguistiche (LR) necessarie alla creazione dei componenti SLT nelle applicazioni di comunicazione da persona a persona ed uomo-macchina in ambienti multilingua. L'obiettivo del progetto è di creare lessici per 13 lingue e corpora di testo per 3 lingue e di preparare una dimostrazione di traduzione in tre lingue. I lessici copriranno 13 lingue: ogni lessico avrà almeno 100 000 entrate consistenti in 50 000 parole comuni, 45 000 nomi propri e 5 000 parole ad uso specifico. Le liste di parole corrispondenti sono estratte da grandi corpora testuali che contengono almeno 10 milioni di parole. Le lingue coperte sono l'italiano, il greco, il russo, il turco, lo spagnolo, il catalano, il tedesco, l'arabo classico, l'ebraico, l'inglese americano, il finlandese, il cinese mandarino e lo sloveno. Nella seconda fase saranno prodotti corpora testuali allineati e lessici monolingue con entrate morfosintattiche. Queste LR copriranno il dominio linguistico turistico. Le coppie di lingue considerate sono: catalano/inglese americano, spagnolo/catalano e spagnolo/ inglese americano.

PF-STAR, che si concluderà in ottobre 2004, ha l'obiettivo ambizioso di fornire piattaforme tecnologiche avanzate e valutazioni tecnologiche comparative in tre aree chiave delle interfacce multimodali: traduzione del parlato (SLT), la rilevazione di stati emozionali e le tecnologie del parlato per i bambini. PF-STAR parte dai risultati sviluppati in anni di ricerca da diversi progetti di ricerca nazionali e internazionali, quali NESPOLE!, C-STAR, Verbmobil, SmartKom. Le lingue considerate sono l'inglese, il tedesco, l'italiano e lo spagnolo.

I partners di progetto sono: ITC-IRST, RWTH, UKA, CNR ISTC-SPFD, EURLN, KTH, and UB.

IL PROGETTO TC-STAR

Trascrivere e tradurre notiziari e discorsi è l'obiettivo del progetto integrato europeo TC-STAR, coordinato dall'ITC-irst, e che mira a far progredire notevolmente le prestazioni dei sistemi di traduzione vocale, cercando di ridurre la distanza fra la traduzione umana e quella della macchina. Nella prima fase triennale di ricerca verranno affrontate le traduzioni di notiziari televisivi e di discorsi pubblici (conferenze e sessioni del parlamento europeo), vedi figura 4, mentre nel secondo triennio verrà affrontata la traduzione di conversazioni libere. Le lingue considerate nella prima fase saranno l'inglese, lo spagnolo ed il cinese.

Il progetto TC-STAR, è iniziato il primo aprile del 2004 ed è un progetto integrato con un finanziamento di 11 milioni di euro e

12 partner: ITC-IRST, RWTH, LIMSI, UPC, UKA, IBM, SIEMENS MONACO e LANNION, NOKIA, SONY STUTTGART, ELDA e SPEX. TC-STAR è pensato come uno sforzo a lungo termine focalizzato sulla ricerca avanzata in tutte le tecnologie chiave per la traduzione: riconoscimento del parlato, traduzione del parlato e sintesi del parlato. Il focus del progetto sarà sullo sviluppo di nuovi, possibilmente rivoluzionari, algoritmi e metodi che integrano tutta la conoscenza linguistica disponibile con metodi e modelli statistici.

TC-STAR è pianificato per una durata di sei anni, che rappresenta il tempo necessario per esplorare e valutare i nuovi approcci alle SLT e per creare l'infrastruttura necessaria per accelerare il tasso di progresso nel settore. Le azioni chiave per raggiungere gli obiettivi e affrontare queste grandi sfide sono:

– l'implementazione e la valutazione di una infrastruttura basata su valutazioni competitive, al fine di raggiungere il punto di svolta desiderato, come rappresentato in figura 5;
– la creazione di un'infrastruttura tecnologica che mira ad alimentare un'efficace distribuzione e valutazione dei risultati scientifici;
– il supporto della disseminazione della conoscenza dei risultati scientifici all'interno del consorzio e della comunità scientifica.

Uno sguardo al futuro

Tra una decina di anni alcuni di questi strumenti saranno con buona probabilità di uso comune. Oggi possiamo fare una ricerca sul sito Rai cercando le notizie dei telegiornali locali e con una grande precisione accediamo proprio allo spezzone di telegiornale che ci interessa. Uno dei componenti chiave di questo motore di ricerca è un sofisticato trascrittore automatico del parlato, che pur facendo degli errori, ci azzecca a sufficienza per lo scopo. Quando 8 anni fa decidemmo, dopo l'esperienza positiva della dettatura dei referti medici, di affrontare la nuova sfida della trascrizione di radio e telegiornali ci sembrava un obiettivo fuori portata, impossibile.

Guardando al futuro, le tendenze in atto suggeriscono che dovremmo essere più abili nelle lingue per aumentare la nostra capacità di comunicazione e nel contempo più dotati di strumenti automatici di analisi del linguaggio finalizzati all'elaborazione di informazione e conoscenza.

GIANNI LAZZARI *Itc-irst*

Figura 5. Il Processo di valutazione nel Progetto TC-STAR.

L'informatica a sostegno del multilinguismo: il servizio di traduzione della Commissione europea

A differenza di altre organizzazioni internazionali l'Unione europea emana atti che sono direttamente vincolanti per i cittadini europei e che debbono quindi risultare immediatamente comprensibili ai destinatari. Un regolamento della Commissione europea è direttamente applicabile in tutti gli Stati membri e perché questo sia possibile è necessario tradurlo in tutte le lingue ufficiali dei paesi membri dell'Unione. Questo obbligo, che risponde ad evidenti esigenze democratiche, è sancito da un regolamento, il primo in assoluto adottato dalla CEE e dall'EURATOM nel lontano 1958, modificato diverse volte, ma tuttora in vigore. Esso stabilisce che le lingue ufficiali degli Stati membri sono lingue ufficiali dell'Unione europea, che tutti i regolamenti e gli altri testi di portata generale delle istituzioni comunitarie devono essere redatti e pubblicati in tutte le lingue ufficiali e che i testi diretti dalle istituzioni ad uno Stato membro, o ad una persona appartenente alla giurisdizione di uno Stato membro, devono essere redatti nella lingua di tale Stato.

Per ottemperare a questi obblighi le diverse istituzioni europee hanno dovuto dotarsi di un servizio di traduzione interno, a cui affidare l'elaborazione delle diverse versioni linguistiche dei dispositivi comunitari.

La Commissione europea, che dispone di un proprio potere di decisione e di un potere di iniziativa, è l'istituzione più fortemente implicata in questo processo di redazione ed è assistita da un servizio di traduzione che può vantarsi di essere il più grande del mondo, non solo per il numero di pagine tradotte (624 499 nel 2003) e di traduttori (1092) ma anche e soprattutto per le combinazioni linguistiche attivate. Dalle quattro lingue utilizzate nel 1958 – francese, tedesco, neerlandese e italiano – il primo maggio 2004 si è passati a venti lingue (ceco, danese, estone, finlandese, francese, greco, inglese, italiano, lettone, lituano, maltese, neerlandese, polacco, portoghese, slovacco, sloveno, spagnolo, svedese, tedesco e ungherese), che diventeranno presto ventitre, con l'arrivo del bulgaro, del croato e del rumeno.

Fin dalle origini della Comunità europea l'italiano è utilizzato in modo assai limitato per la redazione di testi ed è presente nelle istituzioni comunitarie soprattutto come lingua di traduzione. Per la redazione si ricorre infatti in larghissima misura alle due lingue di lavoro delle istituzioni: l'inglese e il francese.

Nel 2003 la percentuale dei testi redatti in italiano era solo dell'1,9%, contro il 59,46% per l'inglese, il 28,79 per il francese e il 3,62% per il tedesco. Rispetto al 1999, la situazione dell'italiano appare invariata, mentre l'inglese ha guadagnato 7 punti (52%), il tedesco ha perso un punto (4,5) e il francese quasi 6 (34,5%). In numero di pagine la presenza dell'italiano non è comunque indifferente: 25 991 pagine, provenienti in massima parte dalle istituzioni italiane, di cui è stata assicurata la traduzione almeno in inglese e francese. Nel primo semestre 2004 la percentuale dei testi redatti in italiano è salita al 2,19%, e si conferma il progresso dell'inglese (59,91%), accanto al calo del francese (28,26%) e del tedesco (3,1%) *(grafico 1)*.

I traduttori italiani della Commissione traducono virtualmente da tutte le lingue ufficiali dell'Unione europea, ma di fatto quasi esclusivamente dall'inglese e dal francese, lingue in

cui è redatto il 95,8% dei testi di cui è richiesta una versione in italiano. Seguono in percentuale il tedesco (1,79), lo spagnolo (1,16), il finlandese (0,32), il neerlandese (0,31), il portoghese (0,23), il greco (0,16) il danese (0,10) e lo svedese (0,07) *(grafico 2)*. Oltre il 90% dei testi tradotti in italiano sono destinati alla pubblicazione: nella Gazzetta ufficiale dell'Unione europea (serie L per la legislazione, serie C per le proposte, le decisioni, le comunicazioni, ecc.), sotto forma di relazioni o bollettini o sui siti web della Commissione, per l'informazione dei cittadini. La percentuale restante è costituita dalla corrispondenza con le istituzioni italiane e da testi ad uso interno (*grafico 3*).

La produzione è aumentata in modo costante: da 87 451 pagine nel 1982 si è passati a 101 363 nel 1999 e a 116 836 nel 2003. Il numero dei traduttori italiani nello stesso periodo si è invece ridotto, scendendo da 112 a 91. C'è quindi stato un sensibile aumento della produttività, reso possibile soprattutto da un ricorso sempre più sistematico all'informatica.

Il primo strumento informatico utilizzato dal servizio di traduzione è stata la banca terminologica multilingue **Eurodicautom**, creata dalla Commissione europea nel 1973. Alle quattro lingue originarie – francese, italiano, neerlandese e tedesco (più il latino) – si sono via via aggiunte le lingue dei nuovi paesi aderenti all'Unione europea fino ad arrivare ad 11 lingue nel 1995. Attualmente la banca contiene oltre sei milioni e mezzo di termini e 300 000 abbreviazioni, classificate in 48 settori, con particolare attenzione alle tematiche connesse con le politiche comunitarie: agricoltura, telecomunicazioni, trasporti, finanze ecc.

Destinata originariamente ad uso interno, da diversi anni è liberamente accessibile su Internet (**http://www.cc.cec/eurodicautom/Controller**) e presenta un altissimo tasso di consultazione: in media 120 000 richieste al giorno. L'interfaccia di ricerca, disponibile solo in inglese, è molto semplice da usare: basta inserire nell'apposito campo il termine che si vuole controllare e selezionare la lingua d'origine e la lingua in cui si desidera la traduzione. Il sistema proporrà diverse traduzioni del termine ricercato, a seconda dei contesti d'uso, indicandone la fonte e nel caso di concetti particolarmente complessi fornendo anche una breve nota esplicativa. È possibile personalizzare le ricerche in base alle proprie esigenze specifi-

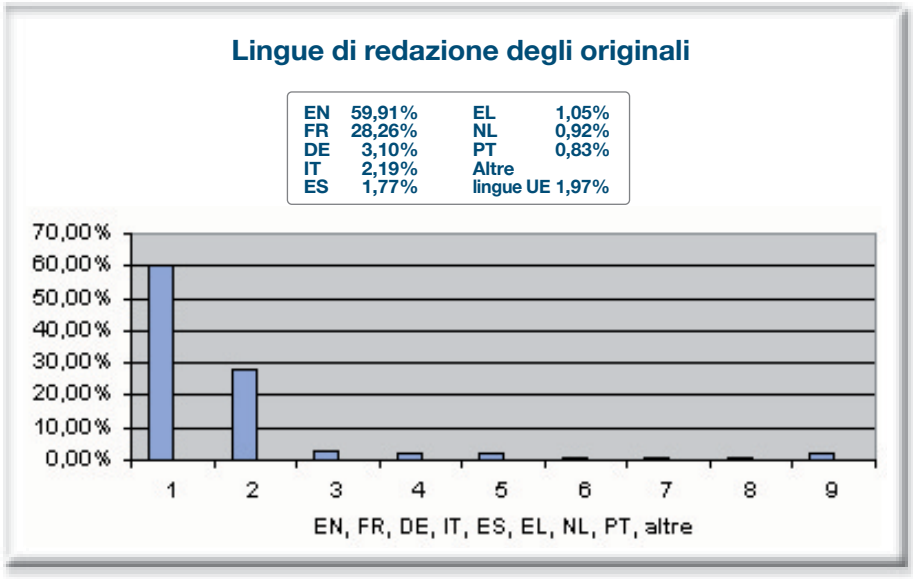

Grafico 1. Lingue di redazione dell'originale (primo semestre 2004).

che nonché lasciare una nota o un commento come feedback per i gestori del servizio.

Nei prossimi mesi, con il progetto **IATE** (Inter-Agency Terminology Exchange) la banca Eurodicautom sarà integrata in una nuova banca dati terminologica, in cui confluiranno le banche terminologiche di altre istituzioni comunitarie e che verrà sviluppata e alimentata nelle venti lingue dell'Unione.

Nel 1976, con una risoluzione del Consiglio delle Comunità europee, è stata istituita la banca dati **CELEX** (acronimo per Comunitatis Europae Lex). Si tratta di un sistema interistituzionale di documentazione automatizzata per il diritto comunitario, a cui affluiscono gli atti emanati dai principali organi comunitari che possono essere resi pubblici. Attualmente contiene circa 250 000 documenti in undici lingue comunitarie (l'inserimento delle versioni nelle nuove lingue al momento non è ancora pienamente operativo), riconducibili essenzialmente a quattro grandi categorie: legislazione, lavori preparatori, giurisprudenza, misure nazionali di esecuzione delle direttive e interrogazioni parlamentari. La legislazione comprende i trattati istitutivi dell'Unione, i trattati di adesione, gli accordi conclusi dall'Unione europea con altri paesi o organizzazioni internazionali, il diritto derivato (regolamenti, direttive, decisioni, raccomandazioni ecc.) e gli accordi conclusi tra gli Stati membri. I lavori preparatori includono essenzialmente le proposte legislative della Commissione nonché i programmi, i rapporti e le comunicazioni emananti da questa istituzione, le risoluzioni del Parlamento europeo, i pareri e le risoluzioni del Comitato economico e sociale e del Comitato delle regioni, le posizioni comuni del Consiglio ecc. Nella giurisprudenza figurano in primo luogo le sentenze e i pareri della Corte di giustizia delle Comunità europee e del Tribunale di primo grado. Nel settore "misure nazionali di esecuzione" si trovano i riferimenti relativi alle disposizioni nazionali con cui le direttive vengono recepite nell'ordinamento giuridico interno di ciascuno Stato membro.

La base CELEX poteva essere consultata anche da utenti esterni, a pagamento, fin dagli anni '80. Dal 1° luglio l'accesso è diventato libero e gratuito (**http://europa.eu.int/celex/htm/celex_it.htm**). La consultazione, basata su un'interfaccia di grande semplicità, è possibile in italiano come in ognuna delle altre dieci lingue ufficiali dell'Unione prima dell'allargamento.

Nel 1976 la Commissione ha cominciato a mettere a punto un sistema di traduzione automatica noto come **EC Systran**, basato sulla tecnologia Systran (*System Translation*), ma diverso dalla versione reperibile in commercio. Attual-

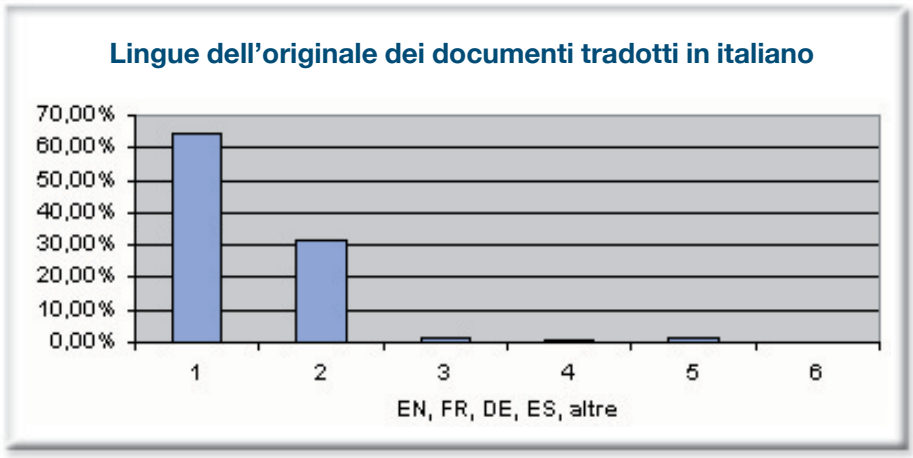

Grafico 2. Lingue dell'originale dei documenti tradotti in italiano (primo semestre 2004).

mente Systran CE copre diciotto combinazioni linguistiche :
– dall'inglese in francese, greco, italiano, neerlandese, portoghese, spagnolo e tedesco,
– dal francese in inglese, italiano, neerlandese, portoghese, spagnolo e tedesco,
– dallo spagnolo in francese e inglese,
– dal tedesco in francese e inglese,
– dal greco in francese.

La traduzione automatica di testi italiani in francese e in inglese è teoricamente ancora in fase sperimentale, ma di fatto già operativa.

Il sistema, che permette di tradurre fino a 2000 pagine all'ora, è a disposizione sia dei traduttori che degli amministratori che lavorano nelle diverse direzioni generali della Commissione. È utilizzato per la traduzione di testi, ma anche per la redazione in una lingua diversa dalla lingua principale dello scrivente. In effetti esso si è rivelato uno strumento prezioso per l'attività amministrativa in un'istituzione multilingue come la Commissione europea e il suo tasso di utilizzazione risulta molto più elevato tra gli amministratori (60%) che fra i traduttori (40%). Alcuni funzionari preferiscono infatti scrivere un testo nella propria lingua, chiederne la traduzione automatica e successivamente correggere o far correggere il risultato. La Commissione ha creato un servizio di editing "rapido", appaltato all'esterno, che offre agli utenti del sistema una versione grammaticalmente corretta, ma di mediocre qualità linguistica, utile per testi urgenti, ma a diffusione limitata. La quantità di correzioni necessarie varia in funzione della lingua, della materia e del tipo di testo, nonché dalla qualità linguistica del testo originale: se il testo di partenza è impreciso, contiene errori di ortografia o è sintatticamente troppo complesso il risultato lascerà necessariamente a desiderare. La qualità della traduzione dipende naturalmente anche dall'affinità tra la lingua di partenza e la lingua di arrivo e dagli investimenti effettuati per alimentare i dizionari specializzati di ciascuna lingua. Il sistema si basa infatti su una serie di dizionari specializzati per settore (attualmente i settori sono 36) e può essere specificamente configurato per rendere lo stile di determinati tipi di testi : verbali di riunioni, lettere, manuali di istruzioni ecc. Esso è inoltre integrato con la base terminologica Eurodicautom e con la base documentaria Celex.

Dal 1° gennaio 1994 i traduttori della Commissione dispongono di un sistema di archiviazione elettronica, denominato **SdTVista**, che contiene tutti gli originali da tradurre inviati dalle direzioni generali e tutte le traduzioni elaborate dal servizio. Un'interfaccia di facile utilizzazione permette di ritrovare quasi istantaneamente qualsiasi documento anche sulla base

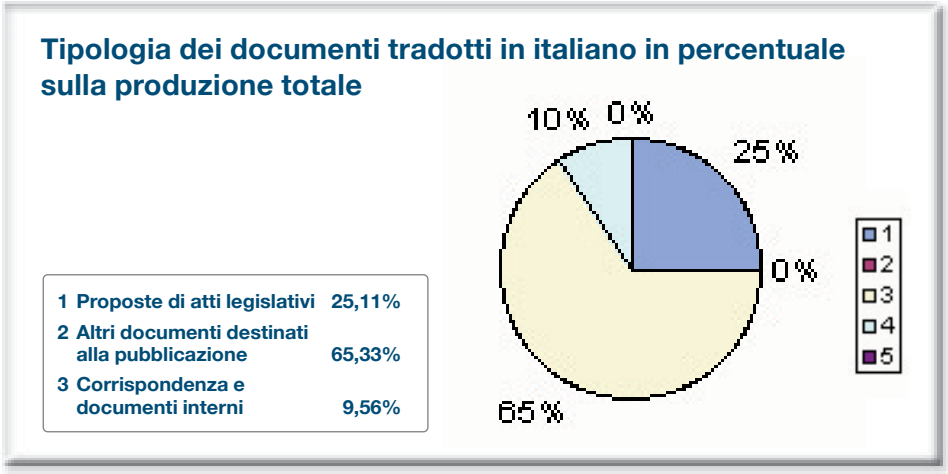

Grafico 3. Tipologia dei documenti tradotti in italiano (primo semestre 2004).

del contenuto del testo. In questo modo è possibile controllare le precedenti traduzioni di un determinato termine a partire da qualsiasi lingua comunitaria e in qualsiasi lingua comunitaria. Per molti documenti del 2004 questa possibilità di controllo è ormai estesa a tutte le 400 combinazioni linguistiche esistenti. In questo modo in pochi secondi per un determinato segmento di testo in italiano è possibile visualizzare i segmenti corrispondenti in ognuna delle altre diciannove lingue.

Attualmente lo strumento informatico più importante per il servizio di traduzione della Commissione è il sistema **EURAMIS** (European Advanced Multilingual Information System), un insieme di applicazioni che offre una serie integrata di servizi nel campo del trattamento del linguaggio naturale. EURAMIS è nato nel 1994, dalla collaborazione tra la Direzione generale Società dell'informazione e la Direzione della traduzione, concretatasi nell'indizione di una gara per la fornitura di strumenti multilingue e il loro inserimento in servizi multilingue. Tra le applicazioni offerte figura il TWB (Translator's Workbench), un sistema di traduzione assistita sviluppato dalla ditta Trados che consente la gestione locale e il trattamento interattivo di dati ottenuti mediante l'allineamento di due versioni linguistiche di uno stesso documento: nel corso dello stesso processo di traduzione ma anche e soprattutto attraverso la consultazione di una memoria di traduzione centrale. Euramis si basa infatti su uno stoccaggio centrale delle risorse linguistiche, che permette la condivisione globale dei dati (memorie di traduzione). Il sistema integra tutti gli strumenti di aiuto alla traduzione esistenti e offre la possibilità di combinare più servizi: dall'allineamento del testo con integrazione dei dati CELEX, all'estrazione di allineamenti con sostituzione automatica di porzioni di frasi contenuti nella memoria centrale fino alla combinazione con la traduzione automatica. In quest'ultimo caso all'utente viene proposta una traduzione completa del testo, in cui l'impiego di diversi colori permette di distinguere le equivalenze assolute (segmenti di testo che corrispondono perfettamente a segmenti già tradotti e convalidati), le equivalenze parziali e i risultati della traduzione automatica.

La memoria centrale è alimentata costantemente con l'invio dei segmenti tradotti dai traduttori della Commissione che utilizzano il sistema ovvero con postallineamenti di testi già pubblicati e considerati di particolare rilevanza e/o utilità. La sua ricchezza è dunque direttamente legata alla sua frequenza di impiego: alcuni settori, come ad esempio quello dell'agricoltura, in cui vi è una forte percentuale di testi ripetitivi che si prestano particolarmente a questo tipo di trattamento, dispongono di un rilevante patrimonio di allineamenti, mentre altri sono molto meno sviluppati. Nel 2003 le pagine tradotte in italiano utilizzando le applicazioni EURAMIS sfioravano il 23% della produzione totale.

La struttura del sistema permette di tradurre da qualsiasi lingua in qualsiasi lingua: una volta che per uno stesso testo esistono allineamenti dall'ungherese in inglese e dall'inglese in italiano è possibile ricavare allineamenti ungherese-italiano come supporto per una eventuale traduzione dall'ungherese in italiano. La qualità dei risultati dipende, in questo come in tutti gli altri casi, dalla qualità dei segmenti tradotti e convalidati in passato. I documenti comunitari devono essere per definizione equivalenti in tutte le loro versioni linguistiche: se tutte le versioni linguistiche di un originale inglese sono buone, la corrispondenza tra le diverse coppie di versioni dovrebbe essere teoricamente perfetta.

Attualmente la procedura che deve seguire un traduttore che utilizza gli strumenti di traduzione assistita per tradurre un determinato documento è la seguente:

– aprire il documento sul suo computer; selezionare la lingua di partenza e la lingua di arrivo; controllare l'esistenza di allineamenti immediatamente utilizzabili ; chiedere altri allineamenti utili per quel testo (testi di riferimento,

modelli ecc.); creare una memoria di traduzione importandovi gli allineamenti già proposti automaticamente dal sistema e quelli da lui eventualmente richiesti; se del caso, chiedere una traduzione automatica di supporto, con verifica dei dati CELEX; completare, controllare e convalidare il prodotto così ottenuto e inviarne copia alla memoria centrale.

In questo processo è evidente che il ruolo del traduttore diventa particolarmente cruciale nelle fasi a valle e a monte della traduzione: deve sapere quali sono i testi di riferimento da utilizzare per quel testo e deve saper valutare correttamente le proposte di traduzione del sistema, adeguandole al nuovo contesto, in modo da fornire una traduzione esatta e coerente dell'originale.

La stragrande maggioranza dei traduttori scrive direttamente il proprio testo sul computer. Il personale di segreteria, in media una persona ogni sei traduttori, un tempo addetto alla battitura dei testi dettati dal traduttore, svolge ora compiti amministrativi, collabora al pretrattamento dei testi, controlla il rispetto dei formati previsti per la presentazione e interviene su problemi specifici, ad esempio nell'elaborazione di tabelle e grafici. Ai traduttori che non amano scrivere i propri testi o che non possono farlo per motivi contingenti dal 2002 viene data la possibilità di utilizzare un sistema di riconoscimento vocale, il Dragon Naturally Speaking, che permette di dettare un testo direttamente al computer. Il sistema è attualmente disponibile in francese, inglese, italiano, neerlandese, spagnolo e tedesco, ha un tasso di precisione massima del 98%, una velocità massima di 160 parole al minuto e può essere combinato con gli altri strumenti di aiuto alla traduzione.

Conclusioni

Il ricorso a strumenti informatici sempre più sofisticati ed integrati hanno consentito al servizio di traduzione della Commissione di assicurare la traduzione di una quantità sempre maggiore di documenti in un numero sempre maggiore di lingue riducendo sensibilmente l'organico dei traduttori di ciascuna lingua e mantenendo un buon livello qualitativo. Questa evoluzione ha cambiato il lavoro del traduttore, e non solo perché sono necessarie nuove competenze per potersi avvalere efficacemente delle risorse esistenti. Il traduttore lavora oggi in una rete ed è costantemente chiamato ad interagire con essa: se da un lato le buone soluzioni del singolo diventano il patrimonio di molti, dall'altro l'errore di un traduttore rischia di moltiplicarsi a catena, inquinando la memoria centrale. È quindi essenziale assicurare il controllo di qualità del sistema e riservare, all'interno del processo di traduzione, una quota non irrilevante del tempo di lavoro del traduttore umano a compiti di verifica e revisione del proprio prodotto.

Gli strumenti informatici contribuiscono inoltre in misura rilevante alla salvaguardia del multilinguismo nelle istituzione comunitarie, in quanto ne riducono sensibilmente i costi e aprono canali di passaggio da una lingua all'altra non necessariamente unidirezionali e centripeti: le quattrocento combinazioni linguistiche attivabili nell'Unione a 25 sarebbero una sfida impossibile per un servizio non informatizzato, mentre possono diventare una realtà operativa nell'Europa di domani. Grazie ad essi, senza incorrere in un aumento esponenziale dei costi, sarebbe attuabile una comunicazione interlinguistica più ricca e articolata, che contribuisca a rafforzare l'identità europea, forte delle sue diversità anche a livello linguistico.

Elisa Ranucci-Fisher *DGT*

Elisa Ranucci è capo di una delle quattro unità italiane di traduzione presso la Direzione generale Traduzione della Commissione europea a Bruxelles. Le opinioni espresse in questo articolo sono tuttavia quelle personali dell'autrice e non riflettono necessariamente il punto di vista della Commissione europea.