



TECNOLOGIE DEL LINGUAGGIO PER L'EUROPA

Work commissioned by



Trento, Italy

Work carried out by

ACCIPIO
CONSULTING

Aachen, Germany

This work has been funded by the European Union under the integrated project TC-STAR – Technology and Corpora for Speech-to-Speech Translation. Project Manager: Gianni Lazzari – IST-2002-FP6-506738, Web Site: <http://www.tc-star.org>

Aprile 2006

Nota legale: Il contenuto di questo documento riflette solamente il punto di vista dell'autore. La Commissione Europea e il Consorzio TC-STAR non sono responsabili per l'utilizzo che può essere fatto delle informazioni contenute di seguito.

Prefazione

La ricchezza delle lingue è considerata da molti il coronamento dell'evoluzione umana. Il linguaggio ci fornisce il mezzo per comunicare idee, emozioni e conoscenze e per esprimere l'identità culturale. Tutte le realizzazioni umane – nel campo della scienza e della tecnologia, della filosofia, dell'arte e della cultura – sono rese possibili dal linguaggio.

In Europa la diversità linguistica è un dato di fatto. L'Unione europea la considera una componente inalienabile del nostro patrimonio culturale; il principio dell'uguaglianza fra le lingue è quindi presente nei trattati istitutivi della Comunità europea. La tutela della diversità linguistica è da sempre al centro della politica comunitaria. Tuttavia, la diversità linguistica richiede un investimento continuo e consistente e le istituzioni europee destinano una quota rilevante dei bilanci operativi ai servizi di traduzione e interpretazione.

Per l'industria e le aziende europee, la diversità linguistica è una sfida e un punto di forza, come ha fatto recentemente notare la Commissione nella sua prima comunicazione riguardante un nuovo quadro strategico per il multilinguismo¹. È indubbio che investimenti tempestivi nelle tecnologie della comunicazione multilingue possono fornire un rapido accesso a nuovi mercati emergenti in ogni parte del mondo, il che costituisce un fattore cruciale per il successo a lungo termine dell'Europa.

La capacità di accedere alle informazioni e di utilizzarle nelle varie lingue è di vitale importanza per i cittadini, i governi e le attività commerciali e le tecnologie del linguaggio umano possono quindi avere un ruolo rilevante nel facilitare la comunicazione tra le persone, le amministrazioni e le aziende. Negli ultimi vent'anni l'Unione europea, in collaborazione con gli Stati membri, ha sponsorizzato diverse azioni di ricerca e sviluppo che hanno contribuito a creare competenze, risorse e un'infrastruttura linguistica paneuropea.

Attualmente l'Europa è uno dei mercati più avanzati per le tecnologie del linguaggio e la traduzione automatica. L'Unione europea si impegna a garantire che

¹ http://europa.eu.int/comm/education/policies/lang/key/legislation_en.html

siano messi a disposizione gli strumenti e le risorse necessari per tutte le lingue dell'UE e per le principali lingue utilizzate negli scambi internazionali, aprendo la strada a una società dell'informazione multilingue diffusa in tutta Europa. Fornendo prodotti e servizi multilingue, come la ricerca di informazione interlinguistica e i sistemi di traduzione automatica, la Commissione europea si prefigge di raggiungere l'ambizioso obiettivo di generalizzare l'accesso alle informazioni a tutti i cittadini europei, che è uno degli obiettivi chiave dell'iniziativa i2010.

Il multilinguismo è diventato ormai una politica di livello europeo e sarà promosso mediante una serie di azioni nell'ambito dei programmi di istruzione, formazione e ricerca, quali i programmi per l'apprendimento delle lingue, la ricerca sulla diversità linguistica e sulle tecnologie del linguaggio umano e i programmi sui contenuti digitali.

Il presente documento espone chiaramente lo stato dell'arte, le sfide e le opportunità che si profilano per l'Europa in questo importante campo di ricerca; rappresenta inoltre una fonte di ispirazione per i ricercatori, l'industria e i responsabili politici e contribuirà certamente a rendere l'Europa sempre più multilingue.



Viviane Reding
Membro della Commissione Europea,
Responsabile per la Società
dell'Informazione e Media



Ján Figel'
Membro della Commissione Europea,
Responsabile per l'Istruzione, la Formazione,
la Cultura e il Multilinguismo

Sommario

Per l'Europa multilingue, la comunicazione e lo scambio di informazioni tra lingue diverse è di fondamentale importanza. Venti lingue europee ufficiali, cioè 190 coppie di lingue o 380 direzioni di traduzione, hanno un costo notevole su ogni attività che coinvolga contenuti linguistici, nel governo, nelle attività economiche e nelle comunità. Anche se questo sforzo è relativamente piccolo per alcuni tipi di transazioni e comunicazioni, rappresenta comunque una barriera insormontabile per la grande parte delle comunicazioni e transazioni. In futuro questo sforzo potrà diminuire notevolmente a causa della capacità di traduzione fornita da sistemi automatici, che non saranno perfetti come i traduttori professionali umani, ma saranno più economici, più veloci, disponibili e sufficientemente utili per molti scopi. Mentre ci si aspetta che questi sistemi permettano di migliorare l'efficienza nelle tradizionali traduzioni umane, il maggiore impatto di queste tecnologie sarà nelle applicazioni automatiche. La traduzione automatica del linguaggio parlato e scritto comincerà ad essere efficace in nicchie di mercato, ma si espanderà rapidamente in altri ambiti, in modo del tutto indipendente dai servizi commerciali di traduzione attualmente esistenti. Queste tecnologie agiranno come tecnologie abilitanti stimolando il commercio e l'economia dell'Europa. Per l'Europa infatti è una necessità strategica avere la disponibilità di tecnologie del linguaggio, così da facilitare il più possibile la comunicazione e lo scambio di informazioni tra lingue diverse.

Questo documento inizia presentando l'importanza delle tecnologie del linguaggio, in particolare per l'Europa, e descrivendone lo stato attuale. Esamina quindi la prospettiva europea in un contesto globale con riferimento specifico agli Stati Uniti, all'India e all'Asia Orientale. Viene descritto lo stato dell'arte nella ricerca e nelle attività economiche e si delineano le aspettative per i futuri sviluppi di mercato. Infine alcune interviste con *decision makers* e specialisti sia nel campo della ricerca che dell'economia permettono di approfondire l'argomento e di ampliarne la prospettiva.

Indice

1. Cosa sono le tecnologie del linguaggio?	9
2. L'importanza delle tecnologie del linguaggio per l'Europa	11
2.1. Un ostacolo decisivo per il Mercato Europeo Interno	11
2.2. Le lingue nel mondo	12
2.2.1. Le principali lingue del mondo	12
2.2.2. Lingue in via di estinzione	14
2.2.3. L'importanza della diffusione: lingue primarie, secondarie e terziarie e forze di mercato	15
2.3. Uno sguardo da vicino all'Unione Europea	15
<i>Intervista con Karl-Johan Lönnroth</i>	17
2.4. La comunicazione	19
2.5. Il prossimo passo nell'industrializzazione: macchine che elaborano lo scritto o il parlato	20
2.6. Il mercato	21
2.6.1. La traduzione come un fattore di costo	21
2.6.2. I mercati consolidati della localizzazione e della traduzione	22
<i>Intervista con Renato Beninato</i>	24
2.6.3. Dalla traduzione umana alla traduzione automatica: straordinaria riduzione dei costi e miglioramento dell'accessibilità	26
<i>Intervista con Michael Anobile</i>	27
2.6.4. Innovazioni tecnologiche incrementali e di rottura	30
3. Asia, Europa e Stati Uniti: analogie e differenze	32
3.1. Unione Europea	32
3.2. Stati Uniti d'America	33
3.2.1. Ruolo strategico delle HLT	33
3.2.2. Programmi di ricerca	33
<i>Intervista con Joseph Olive</i>	35
3.3. Asia Orientale	38
3.3.1. L'inglese come lingua franca in Asia Orientale	38
3.3.2. Le coppie linguistiche asiatiche e l'importanza crescente del cinese	38
3.3.3. Programmi di ricerca	39
<i>Intervista con Jun-ichi Tsujii</i>	40
3.4. India	43
3.5. Le condizioni economiche di contesto	43

3.6. Una strategia per l'Europa	44
<i>Intervista con Joseph Mariani</i>	47
4. La situazione attuale	48
4.1. Il lavoro di traduzione oggi	48
4.2. Tecnologie utilizzate nella traduzione professionale	48
<i>Intervista con Kevin Bolen</i>	50
4.3. La ricerca sulle tecnologie su cui si basa la traduzione da parlato a parlato	53
4.4. Il progetto TC-STAR	55
5. Il potere di una tecnologia abilitante	57
5.1. Bisogni umani insaziabili	57
<i>Intervista con Deimitris Sabatakakis</i>	60
6. Conclusioni	62
7. Bibliografia e altre informazioni supplementari	63
7.1. Bibliografia	63
7.2. Ulteriori letture	64
7.3. Tabelle	64
7.4. Figure	65
7.5. Abbreviazioni	66

1. Cosa sono le tecnologie del linguaggio?

Sappiamo da lungo tempo che la tecnologia può cambiare in maniera profonda il mondo intorno a noi e il nostro modo di vivere. Guardando al passato, il mondo ha subito cambiamenti straordinari grazie all'informatica ed alle telecomunicazioni. E poiché ognuna di queste rivoluzioni si è imposta gradualmente nel corso degli ultimi anni, il cambiamento, sebbene continuo, non è stato particolarmente traumatico.

Ora siamo consapevoli dell'esistenza di un'altra tecnologia informatica che ha un enorme potenziale, e bisogna riconoscere che siamo in presenza di un'ulteriore rivoluzione. Il suo progresso è stato piuttosto lento, e molti dei temi di ricerca che la caratterizzano sono stati esplorati per così tanti anni che qualcuno ha abbandonato la speranza. È molto difficile insegnare ai computer a trattare il linguaggio umano – sia nella forma orale che in quella scritta – nei modi in cui riescono a farlo le persone: parlare in modo naturale, capire cosa è stato detto (e il suo significato), sintetizzare un documento o una conversazione, trovare una registrazione audio dato il suo contenuto, tradurre da una lingua all'altra. Noi vorremmo essere in grado di interfacciarci con le macchine attraverso la voce ed il linguaggio, perchè questi sono i mezzi di

comunicazione che noi usiamo, e vorremmo che i computer fossero in grado di trattare le informazioni in tutti i modi che riteniamo utili. L'insieme delle tecnologie che si occupano di questi argomenti sono conosciute come tecnologie del linguaggio (in inglese “*Human Language Technologies*”, HLT). Il riconoscimento automatico della voce, la traduzione automatica e la sintesi vocale rappresentano le tecnologie di rilievo, ma ce ne sono molte altre. Come è già successo per l'informatica e le telecomunicazioni, le HLT possono potenzialmente modificare in maniera radicale il modo di rapportarci e di lavorare con le informazioni, perchè esse accedono ed elaborano le informazioni codificate nel linguaggio in modi essenzialmente nuovi. Questo documento si focalizza su un singolo aspetto delle HLT, la capacità di superare la barriera imposta da una lingua diversa, sia nella comunicazione tra persone che nell'elaborazione di informazioni contenute in testi scritti non strutturati.

Un importante insieme di applicazioni basate sulle tecnologie per la traduzione automatica e per la traduzione del linguaggio parlato¹ è costituito dalla cosiddetta “*cross-lingual information retrieval, summarization, and data assimilation*” (recupero, sintesi ed assimilazione di informazione tra lingue diverse). Un se-

¹ Per maggiori dettagli su queste tecnologie cfr. paragrafo 4.3.

condo insieme altrettanto importante comprende la comunicazione tra persone che si esprimono in lingue diverse. In questo caso – in base al tipo di input e output, voce o testo scritto – si parla di traduzione automatica del testo scritto, traduzione del linguaggio parlato o

traduzione da parlato a parlato. Queste tecnologie sono talmente nuove che non è stato ancora coniato un termine per definirle – potrebbe essere *cross-lingual information and communication technology* (*Tecnologie dell'informazione e della comunicazione tra lingue diverse*)?

2. L'importanza delle tecnologie del linguaggio per l'Europa

2.1. Un ostacolo decisivo per il Mercato Europeo Interno

Quattro tipi fondamentali di libertà sono previste dal trattato della Comunità Europea², il *libero movimento di beni, di persone, di servizi e di capitali*. Il mercato interno, il principale risultato dell'integrazione europea, fu realizzato alla fine del 1992 come un'area senza frontiere interne per beni e servizi. Per il commercio interno ai Paesi dell'Unione Europea non c'è alcuna tassa di importazione, e i sistemi di tassazione nazionali devono rispettare queste quattro libertà fondamentali.

Relativamente ai processi di armonizzazione dei regolamenti e delle esigenze giuridiche nazionali, molto è stato fatto per sostenere il mercato interno. La libera circolazione di beni e servizi è garantita dal *principio di mutuo riconoscimento* nel singolo mercato, che elimina il bisogno di un'armonizzazione completa della legislazione nazionale nei singoli Stati Membri: la vendita di beni che sono legalmente prodotti in uno degli stati membri dell'UE non può es-

sere vietata sul territorio di un altro stato membro, anche se quest'ultimo realizza prodotti con specifiche tecniche o con qualità diverse³. Lo stesso principio si applica ai servizi.

Le barriere linguistiche sono l'ultimo ostacolo del commercio per i servizi informativi in Europa.

Con la regolamentazione della tassazione e l'armonizzazione delle leggi, due importanti barriere commerciali sono state essenzialmente superate. Sul versante logistico, la distribuzione dei prodotti è ancora un problema. Per l'informazione l'avvento di Internet ha portato ad una rivoluzione: la sua distribuzione è diventata così veloce ed economica che viene percepita come gratuita e disponibile immediatamente su richiesta. Quindi, mentre le precedenti tre barriere del commercio tra i Paesi sono state abbattute, ne rimane una quarta: la barriera del linguaggio⁴. Lo sforzo richiesto per presentare un prodotto nella lingua del cliente finale può variare in modo notevole e dipende dal tipo di prodotto considerato. Generalmente, questo sforzo sarà elevato in re-

² Articolo 14 del Trattato della CE.

³ Unica eccezione consentita – scavalcando un interesse generale come la protezione della salute, del consumatore o dell'ambiente – è soggetta a condizioni rigorose.

⁴ Questa semplificazione ignora che ci sono, oltre ai problemi legati alla Lingua, anche differenze culturali che devono essere prese in considerazione quando si introduce un bene o un servizio in un mercato locale. Comunque, esso rappresenta in molti casi un problema minore se confrontato con lo sforzo richiesto dalla traduzione.

lazione ai servizi informativi richiesti. Queste considerazioni portano alla seguente tabella:

	Beni	Informazione
Tassazione	😊	😊
Conformità	😊	😊
Distribuzione	😐	😊
Lingua	😐	😞

Tabella 1: Condizioni commerciali per il commercio internazionale all'interno dell'EU, per i beni e per l'informazione. Per i servizi informativi, la barriera della lingua rappresenta un ostacolo rilevante da superare.

Delle quattro barriere considerate, la lingua rimane il maggiore ostacolo, specialmente nel fornire servizi informativi in diversi Paesi. Utilizzando le tecnologie del linguaggio oggetto di questo documento, questo ostacolo rilevante, che rappresenta l'ultima e maggiore differenza tra il nostro mercato comune europeo e un grande mercato interno come quello degli Stati Uniti, sarà eliminato – con significativi benefici economici.

2.2. Le lingue nel mondo

Su scala globale, ci sono 6.912 lingue conosciute e viventi⁵. Molte di queste lingue sono situate in Asia/Pacifico e in Africa (vedi fig. 2).

2.2.1. Le principali lingue del mondo

L'inglese è probabilmente la lingua più importante, ma in termini di persone madrelingua, è superata dal cinese. Comunque, se si sommano alle persone di madrelingua inglesi (375 milioni) le persone che parlano l'inglese come seconda lingua (altri 375 milioni) e quelle che la parlano come lingua straniera (750 milioni), il risultato è la cifra significativa di 1,5 miliardi di persone⁶ in grado di parlare inglese.

La *tabella 2*, ricavata dalla letteratura esistente, mostra la lista completa

Milioni di persone in base a			
Posiz.	Madrelingua	fonte A	fonte B
1	Cinese	1.113	1.123
2	Inglese	372	322
3	Indiano/Pakistano	316	236
4	Spagnolo	304	266
5	Arabo	201	202
6	Portoghese	165	170
7	Russo	155	288
8	Bengalese	125	189
9	Giapponese	123	125
10	Tedesco	102	98
11	Francese	70	72
12	Italiano	57	63
13	Malese	47	47

Tabella 2: Le maggiori lingue del mondo in milioni di persone madrelingua secondo due differenti fonti, (A) The English Company's engco model [Gra] e (B) cifre comparative tratte da Ethnologue ([Gri]; vedi [Gra]).

⁵ Fonte: [Gor]. – Dev'essere almeno menzionato il fatto che c'è un problema di identificazione della lingua. – Una *lingua vivente* è tale se esiste almeno un persona madrelingua.

⁶ [Cry] e altre fonti. Si deve notare che, comunque, le cifre riportate in letteratura sono inconsistenti, e che dipendono molto dal livello di competenza presunta nella lingua straniera.

delle prime lingue ordinate in termini di parlanti madrelingua. È interessante notare come la tabella presenti le prime 13 lingue⁷, non le prime 10 o

	Lingua	Influenza
1	Inglese	100
2	Tedesco	42
3	Francese	33
4	Giapponese	32
5	Spagnolo	31
6	Cinese	22
7	Arabo	8
8	Portoghese	5
9	Malese	4
10	Russo	3
11	Indiano/Pakistano	0,4
12	Bengalese	0,09

Tabella 3: 'Influenza globale' delle 12 maggiori lingue secondo il modello engco (vedi tabella 2). Il punteggio 100 rappresenta la posizione dell'inglese nel 1995 [Gra].

20; un numero scelto, almeno all'apparenza, per poter includere il francese e l'italiano. Perfino nell'arida statistica, quando è coinvolta la lingua, entrano in gioco i sentimenti nazionali, le questioni culturali e le prospettive nazionali (o europee). La lingua è molto legata alla cultura e al senso di appartenenza ad una comunità.

È piuttosto ovvio che l'importanza che intuitivamente attribuiamo a una lingua non è in diretta relazione con il numero delle persone che la parlano. Ad esempio, il francese è percepito come una lingua importante ed è la seconda lingua più insegnata nelle scuole europee dopo l'inglese. L'importanza (percepita) di una lingua dipende anche da altri fattori, come l'importanza economica e politica (oltre all'eredità culturale). La *tabella 3* indica l'importanza globale di alcune lingue rispetto all'Inglese. È inte-

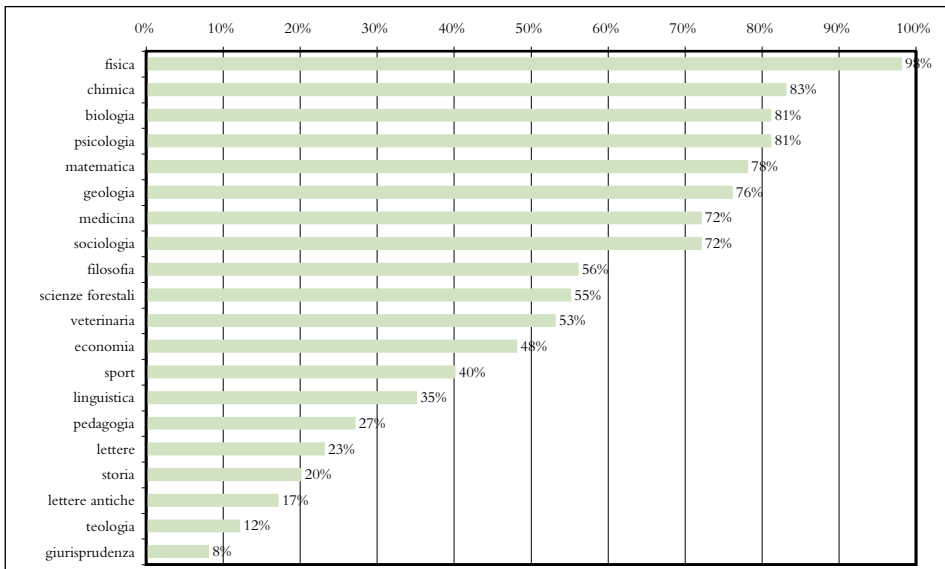


Tabella 4: Discipline in cui gli accademici tedeschi dichiarano che l'inglese è la loro lingua di lavoro [Gra].

⁷ La tabella è tratta da [Gra] che mostra le prime 13 lingue.

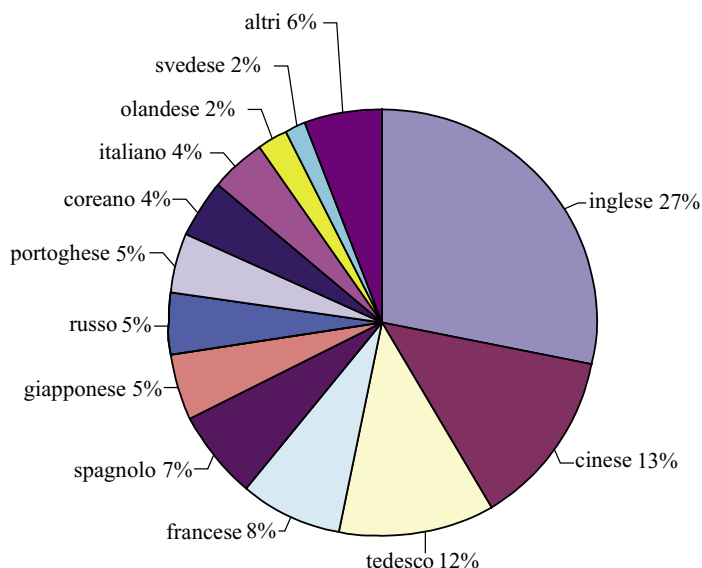


Fig. 1: Proporzione dei libri annualmente pubblicati nel mondo in ogni lingua. L'inglese è la lingua più utilizzata per la pubblicazione di libri: oltre 60 Paesi pubblicano i titoli in inglese. [Gra]

ressante notare che l'importanza di una lingua varia a seconda della disciplina o dell'area di utilizzo. Un esempio specifico è presentato nella *tabella 4*.

2.2.2. Lingue in via di estinzione

Il pericolo di estinzione delle lingue è legato a due fattori: il numero di

parlanti e ed il numero di funzioni per cui la lingua è utilizzata. Tendenzialmente, le persone bilingui iniziano ad utilizzare solo la seconda lingua con i loro bambini, oppure utilizzano la loro lingua madre sempre meno frequentemente. Circa 500 lingue sono considerate quasi estinte⁸, secondo Ethno-

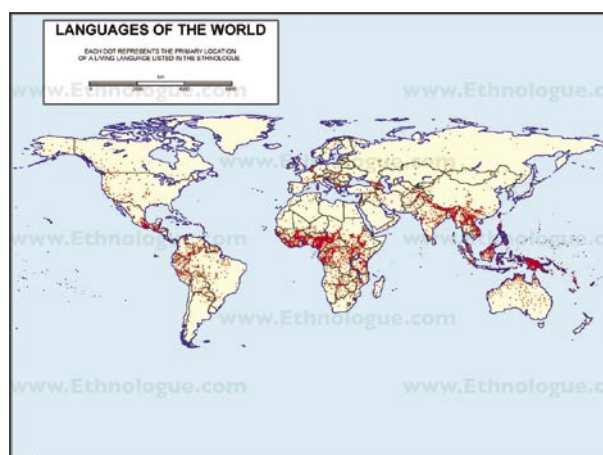


Fig. 2: Lingue del mondo. Ogni punto rappresenta la localizzazione principale delle lingue viventi elencate in Ethnologue [Gor], un catalogo con più di 6.700 lingue parlate in 228 Paesi.

⁸ Quasi estinte: sono definite le lingue per cui la popolazione che parla la lingua è costituita da un numero inferiore a 50 o da una frazione molto piccolo del gruppo etnico.

logue⁹. C'è una certa preoccupazione per questo rischio di estinzione poiché la lingua è strettamente collegata alla cultura; questo legame è tale che l'estinzione della lingua è quasi sempre accompagnata dalla disgregazione sociale e culturale. (Con sorpresa dell'autore di questo rapporto, un'altra preoccupazione è costituita dalla perdita di importanza della comunità accademica che studia tali lingue!).

2.2.3. L'importanza della diffusione: lingue primarie, secondarie e terziarie e forze di mercato

Una lingua non ha bisogno di essere a rischio di estinzione per essere svantaggiata. Considerate una società che ha intenzione di estendere i suoi affari dal mercato locale ad un mercato internazionale. Dato che il costo della localizzazione per una nuova lingua non è correlato al numero di parlanti, ma è più o meno fisso, ci sono sicuramente lingue primarie – quelle per cui è necessario avere un'offerta di mercato – lingue secondarie e, forse, anche lingue terziarie di minore importanza commerciale. Dipende dalle circostanze *quale* lingua sia considerata secondaria o terziaria, ma si può affermare che generalmente le forze di mercato penalizzano alcune lingue, di solito quelle con una popolazione numericamente poco significativa che parla quella lingua; o quelle associate ad economie deboli.

Una situazione simile si ha anche quando le persone iniziano a studiare una lingua straniera. Un tedesco studierebbe l'olandese, che è simile al tedesco e richiede quindi uno sforzo comparativamente minore, o lo spagnolo, considerando che la popolazione mondiale che parla spagnolo è estremamente più numerosa rispetto a quella che parla olandese? (Vedi la *tabella 2* per le cifre tipiche.) E perché studiare l'olandese quando la maggior parte degli olandesi parlano inglese¹⁰? Molte considerazioni influenzano una scelta. La lingua è parlata in un Paese vicino o molto lontano? Esiste in quel Paese una seconda lingua che può essere utilizzata per comunicare?

Molte lingue sono sotto pressione a causa delle forze di mercato.

L'attrattiva di una lingua aumenta in relazione alla probabilità che si abbia bisogno di parlarla. Naturalmente, le lingue parlate da una piccola comunità sono ulteriormente svantaggiate.

2.3. Uno sguardo da vicino all'Unione Europea

La lingua è strettamente legata alla cultura. In alcuni casi, è una componente vitale dell'identità nazionale. Non stupisce il fatto che noi europei, creando l'Unione Europea, abbiamo preso la decisione consapevole di non introdurre

⁹ Ethnologue [Gor] è un catalogo di più di 6.700 lingue parlate in 228 Paesi.

¹⁰ 91% della popolazione olandese gestisce una conversazione in almeno un'altra lingua [EB5]. È importante rilevare anche che il parlamento olandese ha recentemente approvato una legge per cui la conoscenza dell'olandese è un requisito per ottenere la cittadinanza. Ciò è indicativo dell'importanza della lingua per la cultura e l'identità.

Cecoslovacco	CS	Čeština
Danese	DA	Dansk
Olandese	NL	Nederlands
Inglese	EN	English
Estone	ET	Eesti
Finlandese	FI	Suomi
Francese	FR	Français
Tedesco	DE	Deutsch
Greco	EL	Elinika
Uungherese	HU	Magyar
Italiano	IT	Italiano
Lettone	LV	Latviesu valoda
Lituano	LT	Lietuviu kalba
Maltese	MT	Malti
Polacco	PL	Polski
Portoghese	PT	Português
Slovacco	SK	Slovenščina
Sloveno	SL	Slovenščina
Spagnolo	ES	Español
Svedese	SV	Svenska

Tabella 5: Le 20 lingue ufficiali dell'Unione Europea e le loro abbreviazioni [ELP]. L'Irlandese diventerà la 21esima lingua ufficiale dell'UE dal primo gennaio 2007.

re una lingua primaria ma di mantenere le diverse lingue e di attribuire loro pari diritti. Sebbene la grande importanza dell'inglese come lingua franca non sia da sottovalutare, si può sicuramente dire che viviamo in un mondo multilingue in cui la lingua è importante. L'Unione Europea è nata multilingue: al punto che è stato nominato un Commissario per l'Educazione, la Formazione, la Cultura e il Multilinguismo.

Le diverse lingue hanno pari importanza, e alcuni documenti, in particolare le leggi o i dibattiti parlamentari, sono disponibili (cioè sono stati tradotti) in tutte le lingue ufficiali. D'altra parte, per motivi di efficienza e di economicità nell'operatività quotidiana vengono utilizzate tre lingue: l'inglese, il francese e il tedesco.

Riassumendo, abbiamo buone ragioni per proteggere la nostra eredità culturale, ma il mantenimento di molte lingue ha un costo¹¹, sia in termini economici (ad esempio i costi localizzazione di un prodotto) sia in termini di impegno (lo sforzo richiesto ad un individuo per imparare un'altra lingua). Qualsiasi tecnologia in grado di ridurre questi costi rappresenta un valido sostegno alla nostra eredità culturale.

¹¹ Non fornire traduzioni non rappresenta un'opzione: i cittadini hanno bisogno di capire la legge, i prodotti devono essere localizzati nelle diverse lingue per essere venduti. La nostra convinzione è che qualsiasi costo, anche se basso, costituisce un ostacolo, e che qualsiasi riduzione in termini di costi e miglioramento nell'accessibilità avrà un impatto positivo.

Intervista con Karl-Johan Lönnroth, Direttore Generale, Direttorato Generale per la Traduzione (DGT)

Dal Gennaio 2004, Karl-Johan Lönnroth è Direttore Generale del DGT. Precedentemente ha assunto questi incarichi:

2000-2003, Vice Direttore Generale, Direttorato Generale per l'Occupazione e gli Affari Sociali (General for Employment and Social Affairs, CEC, European Commission)

1996-2000, Direttore della Strategia per l'Occupazione e dei Fondi Sociali Europei, Direttorato Generale per l'Occupazione e gli Affari Sociali (CEC)

1991-1996, Direttore del Dipartimento per l'Occupazione, International Labour Office, Ginevra

1971-1991, differenti cariche al ministero finlandese del lavoro come ricercatore, responsabile della pianificazione, vice direttore dei servizi per il mercato del lavoro, e special advisor.

1973-1977, Segretariato del Consiglio dei Ministri Nordico, funzionario responsabile per l'occupazione e per gli affari dell'emigrazione e per la co-operazione tripartita. Consulenze per l'OECD, il Consiglio Nordico dei Ministri e l'Ambasciata Finlandese in Stoccolma.

Formazione: Master of Political Sciences, University of Helsinki (1970), Master of Arts, University of Wisconsin, USA (1972), Ecole Nationale d'Administration (ENA) (1983).

Abilità linguistiche: Finlandese, Svedese, Inglese, Francese e Tedesco; conoscenza elementare del Russo e dello Spagnolo.

Karl-Johan Lönnroth ha un'ampia esperienza nella cooperazione: Paesi Nordici, OECD, UN, ILO; e nella cooperazione bilaterale e.g. Europa orientale. Ha pubblicato più di 40 articoli, pubblicazioni etc. relativi al lavoro, all'occupazione, all'emigrazione, a questioni sociali e politiche, e a tematiche sociali.



Karl-Johan Lönnroth
Direttore Generale
Direttorato Generale per la Traduzione
Commissione Europea
Lussemburgo e Bruxelles

Il DGT è un'organizzazione piuttosto grande, è corretto?

Suppongo che siamo la più grande nel mondo. Lavoriamo con 21 lingue. La cifra globale per l'intero servizio linguistico per l'Unione Europea è di 1,1 miliardi di € all'anno o circa l'1% dell'intero budget dell'UE. Questo include l'interpretazione e la traduzione, e coinvolge non solo la Commissione ma anche tutte le istituzioni come il Parlamento Europeo. Sembra molto ma in tutto sono 2,55 € per cittadino.

Il grande numero di lingue non costituisce un peso per la società e l'economia europee?

Questa domanda ha già una risposta implicita. Presuppone che sia un peso e un costo piuttosto che un vantaggio. Il multilinguismo è realmente parte di un modello sociale europeo; noi abbiamo questa diversità culturale. Si deve considerare come un punto di forza e un elemento dinamico piuttosto che un vincolo, e direi che naturalmente l'aver molte lingue apre i mercati. La politica del multilinguismo beneficia la nostra società dinamica. In Europa, noi rispettiamo le libertà fondamentali e la diversità culturale, e attraverso questa diversità, si possono ottenere nuove idee e dare nuovo impulso all'economia. È una limitazione conoscere solamente una lingua all'interno di un singolo mercato del lavoro europeo. Questo è uno dei motivi per cui l'Unione cerca di promuovere la conoscenza e l'apprendimento delle lingue.

Come si è relazionata la DGT alle sfide connesse all'allargamento dell'UE?

Abbiamo assunto più di 500 nuove persone nel nostro staff durante l'ultimo anno. Penso che stiamo affrontando abbastanza bene la nuova sfida. La difficoltà che abbiamo è di sviluppare e coniare nuovi termini e migliorare la qualità delle traduzioni, perché alcuni dei termini di politica europea non esistono nelle nuove lingue. Questa è la sfida principale. Noi traduciamo prevalentemente un testo legislativo partendo da zero e per la prima volta, e questo talvolta è abbastanza complicato. L'integrazione di queste nuove lingue è andata piuttosto bene.

Come si aspetta che si svilupperà la traduzione automatica?

La traduzione automatica come strumento di traduzione assieme alla translation memory è qualcosa che si sta ancora sviluppando e, dal mio punto di vista, avrà un futuro brillante. Comunque questo non significa che l'elemento umano diminuirà. Lo sviluppo va più nella direzione di una combinazione che io chiamerei traduzione intelligente, il traduttore assicura quindi la qualità, la coerenza, la terminologia corretta, e la traduzione automatica aiuta a capire e a migliorare la produttività.

Come vede la professione dei traduttori?

L'industria delle tecnologie del linguaggio rappresenta l'industria che sta crescendo più velocemente nel mondo. La globalizzazione e la moltitudine di lingue aumenta il bisogno di servizi multilingue, il che significa anche che la professione deve essere riconosciuta meglio, perché svolge un ruolo importante. Date le grandi sfide, cambierà anche la figura del traduttore che non sarà più solamente un mero traduttore di testi, ma diventerà un consulente linguistico, un editor. Così la diversificazione della professione continuerà.

Siamo in contatto con Università e loro associazioni per sviluppare i curricula e la formazione, e stiamo cercando di realizzare quello che chiamiamo l'European master of translation, che dovrebbe offrire una qualificazione standard, riconosciuta da tutti, facilitando le assunzioni. Il tutto dovrebbe portare ad una migliore professionalizzazione.

2.4. La comunicazione

Molti europei parlano due o più lingue. Ma circa metà dei cittadini dell'Unione Europea non parla lingue diverse dalla propria.

Non sarebbe un progresso significativo se gli Europei che non parlano la stessa lingua potessero parlare facilmente gli uni con gli altri? Qualsiasi forma di comunicazione, perfino ad un livello ridotto, sarebbe un progresso rispetto alla situazione attuale. La traduzione non ha necessariamente bisogno di essere perfetta. Questi requisiti sono molto differenti da quelli richiesti per una traduzione professionale. Come vedremo meglio più avanti, questo ha un effetto fondamentale sulle forze di mercato e specialmente sulle prospettive dei nuovi mercati.

Possiamo immaginare oggi cosa significherebbe per noi poter utilizzare in modo naturale e a basso costo dei sistemi di traduzione? Noi abbiamo sperimentato quanto la tecnologia possa fare

per migliorare la comunicazione tra le persone riducendo le distanze di tempo e di spazio. Una delle prime e più significative conquiste nella cultura umana è stata lo sviluppo della scrittura, che ha permesso di trasmettere la conoscenza alle generazioni future. Per quanto riguarda lo spazio si pensi alla trasmissione di informazioni lungo i confini degli imperi Romano o Cinese utilizzando segnali visivi. Altri passi da gigante sono stati l'introduzione del servizio postale, del telefono, della linea di telecomunicazione transatlantica, del telefono cellulare e di Internet.

Stiamo affrontando oggi un'altra grande sfida: la disponibilità istantanea ed economica della traduzione, che permette la comunicazione tra le culture al di là del tempo e dello spazio. Avrà successo questa sfida nei prossimi vent'anni? Per noi è difficile immaginare un futuro in cui ognuno potrà comunicare, a costi accettabili, con persone che parlano lingue diverse usando sistemi di traduzione istantanea del parlato e dello scritto, perchè in

Lingua	Parlata come madrelingua	Parlata non come madrelingua (lingua straniera)	Totale parlatori
Inglese	13%	34%	47%
Tedesco	18%	12%	30%
Francese	12%	11%	23%
Italiano	13%	2%	15%
Spagnolo	9%	5%	14%
Polacco	9%	1%	10%
Olandese	5%	1%	6%
Russo	1%	5%	6%

Tabella 6: Abilità di parlare lingue straniere nell'UE: Percentuali di cittadini europei che parlano le rispettive lingue nell'UE (come madrelingua / come seconda lingua o lingua straniera / entrambe). L'indagine si è svolta nell'Unione Europea dei 25 Stati Membri e, in aggiunta, nei Paesi entranti (Bulgaria and Romania), nei Paesi candidati (Croazia e Turchia) e tra la comunità turco cipriota. - Fonte: [EB5]

questo momento questo tipo di comunicazione semplicemente non esiste. È comunque un'assunzione ragionevole pensare che la disponibilità diffusa di uno strumento come questo abbia un impatto altrettanto significativo di quello conseguente all'introduzione del telefono o di Internet.

Dopo aver riavvicinato il tempo e lo spazio, la prossima grande sfida è il collegamento tra lingue e culture.

2.5. Il prossimo passo nell'industrializzazione: macchine che elaborano lo scritto o il parlato

Come abbiamo appena osservato, la disponibilità della traduzione a basso costo in ogni luogo condurrà a cambiamenti fondamentali. Siamo di fronte ad una vera e propria rivoluzione: un processo che precedentemente poteva essere svolto solamente da persone ora può essere svolto da macchine e in modo molto efficace. Negli anni '60 la tecnologia dell'informazione a partire dalle applicazioni governative, scientifiche e militari si è diffusa all'ambito commerciale, in primo luogo nel campo assicurativo e bancario e più tardi praticamente in tutti i settori della vita umana. La maggior parte del lavoro che precedentemente era svolto da persone competenti ed esperte, ora viene svolto da macchine più veloci, economiche ed efficienti. La capacità di elaborare informazione era però limitata e l'elaborazione del linguaggio, con l'eccezione dei dati strutturati come nomi e indirizzi, era pressoché inesistente. Mentre la produzione di testi

scritti, come quotidiani, libri e materiale stampato in generale, fu fortemente sostenuta dall'IT, l'analisi e l'elaborazione di informazione è rimasta molto limitata.

Con l'avvento di Internet, ci sono stati molti cambiamenti. È cresciuta enormemente la necessità di elaborare il linguaggio allo scopo di fare ricerche, sintesi, traduzioni e classificazioni. L'introduzione su vasta scala di servizi self-service su Internet ha condotto ad un notevole aumento nella comunicazione con gli utenti finali sia per le imprese che per la pubblica amministrazione. Allo stesso tempo, le aspettative degli utenti sono cresciute, e la maggior parte di loro danno per scontato che una e-mail riceverà risposta entro un giorno. La ricerca di documenti su Internet basata su parole chiave, effettuata principalmente attraverso Google, è riuscita ad entrare praticamente in tutte le case. Sono però sorti subito due problemi. In primo luogo, la ricerca di parole è solamente un sostituto della ricerca di informazioni, e si preferirebbe avere a disposizione un web *organizzato* semanticamente (*semantic web*) e la possibilità di cercare concetti piuttosto che parole. In secondo luogo, dato che Internet è usato da una larga parte della popolazione, c'è un bisogno crescente di localizzazione, specialmente per quanto riguarda le lingue. Questo crea un'enorme richiesta di traduzione su larga scala e/o in tempo reale, in modo particolare per la ricerca di documenti in lingue diverse (IBM, SAP, Google e Yahoo) e per la produzione di siti web multilingue, che sono diventati

uno dei maggiori volani dell'industria mondiale della traduzione.

Su larga scala, la traduzione è guidata dalla globalizzazione e da internet.

2.6. Il mercato

Considerando la traduzione dal punto di vista economico, non è sufficiente guardare al mercato attuale della traduzione, ma si devono considerare altri aspetti rilevanti. Per l'Europa, il costo della traduzione e il suo effetto sull'economia è di estrema importanza se si considera che, oltre ai costi diretti di traduzione, le barriere linguistiche possono ostacolare lo sfruttamento di nuove opportunità.

2.6.1. La traduzione come un fattore di costo

Se da una parte abbiamo a cuore la ricchezza culturale dell'Europa, riflessa dalle sue molteplici lingue, dall'altra le barriere linguistiche pesano sulla nostra economia, dato che aggiungono costi di transazione a qualunque attività che attraversi i confini linguistici. In molti casi tipici, questo costo è solamente una piccola frazione del costo del prodotto, indicativamente potrebbe essere compreso tra un valore minimo di 0,25% e un massimo del 2%. Alcuni di questi costi sono diretti (sebbene non sia semplice determinarli) e altri indiretti nel senso che impediscono la creazione di valore.

È assai difficile stimare in modo attendibile i costi diretti. In una gran-

de società, per esempio, ci può essere un budget per la realizzazione di un sito web. Però risulta difficile contabilizzare tutte le piccole attività di traduzione condotte nei diversi settori che collaborano alla sua manutenzione, attività non coperte da cifre specifiche e che potrebbero facilmente far lievitare il budget stimato. D'altra parte, neanche conteggiare i salari di tutti i traduttori funzionerebbe adeguatamente, dato che molti di loro lavorano part-time o come liberi professionisti. Difficoltà di quantificazione emergono se un'organizzazione paga una società di localizzazione che utilizza un sotto contraente che a sua volta paga un libero professionista: bisognerebbe conteggiarne il costo in tutte le fasi o solo nella prima fase? Di conseguenza, le cifre disponibili cambiano in relazione alla metodologia utilizzata¹². Sembra comunque che si possa affermare che il mercato globale della traduzione oscilla tra gli 8 e i 30 miliardi di euro. Di seguito due dati sulle spese di governo dell'UE:

- Ogni anno il Parlamento Europeo spende 300 milioni di euro, o il 30% del suo budget, per la traduzione di tutti i dibattiti parlamentari e dei documenti dell'UE nelle 20 lingue europee ufficiali.
- L'Unione Europea spende 1,1 miliardi di euro all'anno, cioè 1% del suo budget, per tutti i servizi di traduzione e di interpretariato.

L'UE spende 1,1 miliardi di euro all'anno per i servizi di traduzione e di interpretariato.

¹² Potete leggere di più su questo argomento nell'intervista a Renato Beninato.

2.6.2. I mercati consolidati della localizzazione e della traduzione

L'opinione pubblica ha almeno un'idea approssimativa sul lavoro svolto dai traduttori, che traducono libri o documenti, ma generalmente non si sa molto dell'industria che cura la localizzazione del software e delle pagine Internet, che rappresenta gran parte dell'industria delle tecnologie del linguaggio. È importante capire che le società e le istituzioni interessate hanno realmente bisogno della "localizzazione" e non solamente della traduzione. A parte l'aspetto della traduzione, la lo-

calizzazione del software assicura anche che il software funzioni correttamente nella nuova lingua: questo significa intervenire anche sugli aiuti e sulla documentazione online, sul fatto che il materiale scritto sia formattato in modo appropriato. Devono essere considerate ad esempio le diverse lunghezze dei testi tradotti, sia in termini di numero di parole che in numero di caratteri, e non bisogna sottovalutare gli stili di scrittura. Oltre a lingue che si scrivono da sinistra a destra o da destra a sinistra, c'è da considerare il caso di testi bidirezionali: se si includono parole in la-

Posiz.	Azienda	Sede principale	Fatt. in Mill. US\$	Impiegati	Uffici	Status
1	Lionbridge Technologies	US	377,1	4.000	50	pubbl.
2	Titan Corp.	US	285,4	---	---	pubbl.
3	SDL International	UK	146,0	1.400	36	pubbl.
4	STAR AG	CH	96,0	750	33	priv.
5	RWS Group	UK	63,4	350	7	pubbl.
6	SDI Media Group	US	60,3	200	20	priv.
7	Xerox Global Services	UK	60,0	200	4	pubbl.
8	Euroscript S.à.r.l.	LU	54,5	600	9	priv.
9	Transperfect/Translations	US	50,2	325	29	priv.
10	CLS Communication	CH	36,0	260	11	priv.
11	Logos Group	IT	36,0	150	17	priv.
12	LCJ EEIG	DE/IT/BE/SP	21,6	140	9	priv.
13	Thebigword	UK	20,0	122	7	priv.
14	Hewlett-Packard ACG	FR	20,0	65	6	pubbl.
15	Moravia	CZ	19,0	350	11	priv.
16	TOIN	JP	19,0	105	5	priv.
17	Merrill Brink International	US	18,5	120	4	priv.
18	VistaTEC	IE	18,2	123	3	priv.
19	Transware	IE	18,0	160	8	priv.
20	McNeil Multilingual	US	17,2	105	9	priv.

Tabella 7: Posizionamento dei primi 20 fornitori di servizi linguistici – reddito del 2004. (Nelle cifre riportate si sono considerate le due importanti acquisizioni del 2005, l'acquisizione di Bowne Global Solutions da parte di Lionbridge e l'acquisizione di TRADOS da parte di SDL.)

tino in testi arabi, si devono supportare entrambe le direzioni di scrittura! Ci sono molti insiemi di caratteri che devono essere supportati, e lingue con caratteri rappresentati con due byte come il giapponese, il cinese e il coreano che richiedono specifico supporto software. (Il sistema Unicode è una soluzione a questi problemi, ma potrebbe non essere applicabile a vecchi documenti elettronici.) In termini generali, il software che si intende localizzare dovrebbe essere progettato come tale fin dall'inizio, per evitare che sorgano costi non necessari in un secondo momento. Infine un ruolo importante nella localizzazione è assunto dalle differenze culturali. Per trasferire un concetto in una diversa lingua o cultura in modo che risulti ancora comprensibile, sono necessarie creatività ed empatia.

Il metodo usuale per la localizzazione consiste nel separare il prodotto nelle sue componenti testuali e in quelle di interfaccia utente. Il testo viene tradotto, e l'interfaccia utente e la documentazione vengono modificate per poter funzionare in modo appropriato nella lingua tradotta.

I principali segmenti di mercato della localizzazione sono la localizzazione del software e dei siti web.

Secondo una fonte¹³, l'industria delle tecnologie del linguaggio ha genera-

to un reddito di 8,8 miliardi di dollari statunitensi nel 2005. Questo comprende sia la traduzione umana che quella fatta attraverso l'utilizzo di strumenti. Due importanti segmenti di mercato attualmente guidano la crescita: la gestione dei siti multilingue e la localizzazione del software. Sono 5.000 le società che in tutto il mondo, contando solamente quelle con cinque o più dipendenti, contribuiscono a questo mercato.

Le maggiori tecnologie utilizzate attualmente per la traduzione sono: le memorie di traduzione (TM), le risorse terminologiche, i software e gli strumenti per trattare i siti web multilingue e per fare localizzazione software. Le TM sono una tecnologia di supporto ben affermata (anche se non universalmente utilizzata) che abbate i costi e migliora la qualità, specialmente la coerenza della traduzione, importante per esempio nei settori tecnici e legali. In modo simile ci si può aspettare che la traduzione automatica o la traduzione del linguaggio parlato, utilizzate come uno strumento, abbiano un effetto benefico sia sui prezzi che sulla qualità della traduzione umana. Come illustrato nel capitolo 2.6.4., che definisce anche la terminologia utilizzata, la traduzione automatica, in questo scenario di utilizzazione, è un'innovazione di supporto (*sustained*).

¹³ Common Sense Advisory [CSA]. Altre fonti stimano che le cifre globali sono significativamente più alte, dell'ordine dei 30 miliardi di euro. Per questo documento non sono le cifre esatte che contano ma il fatto che l'industria del linguaggio è un'industria avviata e in espansione, inferiore come dimensione rispetto ad altre industrie ma che agisce indirettamente su attività commerciali con volumi molto più significativi.

Intervista con Renato Beninato, COO, Common Sense Advisory

Beninato ha più di 20 anni di esperienza a livello direttivo nell'industria della localizzazione. Ha fatto parte dei team direzionali di alcune delle più importanti società, più recentemente rispettivamente come Vice Presidente di AlpNet Inc e Direttore di Berlitz GlobalNET. Egli si è concentrato sulle strategie che guidano la crescita su una scala globale. E' specializzato nel portare al successo una società nei mercati globali e nell'avviare business a livello internazionale. Attualmente è partner e analista responsabile della ricerca al Common Sense Advisory, Inc. una società di consulenza e ricerca di mercato specializzata nella traduzione e nell'industria della localizzazione, con clienti in tutti i continenti.



Renato Beninato
COO e Vicepresidente di
Consulting Practice
Common Sense Advisory, Inc.
Boston, USA

Ho visto cifre sul volume del mercato globale della traduzione che non forniscono un quadro coerente. Perché è così difficile arrivare a queste cifre?

Individuare le dimensioni di un mercato è un esercizio complesso che coinvolge uguali dosi di logica e di scetticismo. Il fine ultimo è di raggiungere un'approssimazione credibile del mercato. Quello che ha condotto la Common Sense Advisory a misurare l'industria della traduzione è stato guidato in parte dalla mia precedente esperienza nella vendita di servizi di traduzione per due società pubbliche e in parte da contatti personali con i più grandi utenti di traduzione nel mercato. Quando si parla di mercato della traduzione agli outsiders, questi vedono immancabilmente enormi opportunità, dal momento che vedono il gap tra ciò che viene pubblicato in tutte le lingue e quello che è pubblicato nella loro lingua. Ma quando si arriva al dunque, cioè si cerca di vendere quei servizi, la realtà è molto più dura. La traduzione non è strategica e, quindi, non rende. Abbiamo proposto di paragonare il prezzo della traduzione con quello della carta igienica per descrivere quanto la traduzione sia presente nei budget delle società. Il peggior difetto delle metodologie di dimensionamento del mercato condotte dagli *outsiders* è che si basano sulle informazioni degli attori di mercato. Una delle caratteristiche chiave del mercato della traduzione è il processo di sotto-contraenza fino a quattro livelli. Per esempio, un venditore multilingue affida le lingue dell'Europa dell'Est di cui ha bisogno ad una società in Ungheria, che a sua volta le gira ad altre società nella Repubblica Ceca, in Polonia, e in Bulgaria, che a loro volta si affidano a traduttori liberi professionisti. Alla Common Sense Advisory, noi contiamo solo il primo *outsourcing*, che rappresenta i soldi realmente sborsati dall'utente finale per il lavoro.

Che metodologia ha seguito?

In “Beggars at the Globalization Banquet,” un rapporto che abbiamo pubblicato nel Novembre del 2002, abbiamo trovato che le società, a seconda dell’industria e della dimensione – spendono in traduzione tra lo 0,25% e il 2% delle loro entrate provenienti dall’estero*. Questo rappresenta molto meno del 3% delle entrate *totali* che, di solito, le organizzazioni assumono come ipotesi.

Per arrivare ai nostri numeri, abbiamo controllato altri approcci al dimensionamento del mercato per vedere se i nostri numeri fossero coerenti. Così abbiamo considerato il numero di società di traduzione sul mercato e i loro ricavi, il numero di traduttori nel mondo, e le spese documentate dei governi. Il confronto incrociato di questi dati ha confermato i nostri numeri.

Che cosa è emerso di importante?

Stiamo costantemente facendo ricerca sull’intera industria delle tecnologie del linguaggio in modo analitico. E’ emerso che la traduzione è qualcosa che viene fatta ai livelli di *mid-management* e raggiunge la stanza del Consiglio di Amministrazione solamente quando qualcosa va molto storto.

Abbiamo inoltre rilevato che con i 9,5 miliardi di US\$ nel 2006, il mercato è quasi della stessa dimensione del mercato della bicicletta mondiale, e che sebbene si parli molto della tecnologia della traduzione, il reddito aggregato di tutti gli attori è solamente di circa 100 milioni di US\$.

Quindi sta dicendo che la traduzione e la localizzazione hanno solamente una piccola importanza commerciale?

No. Al contrario. Perfino se la traduzione è molto economica, rende possibile alle società di penetrare nuovi mercati e di moltiplicare i loro profitti. Ridurre il suo costo promuoverà la reale globalizzazione dei mercati.

Può fornire delle cifre di esempio per i grandi utenti della traduzione?

Come sapete, la DGT ha speso 1,1 miliardi di euro nel 2004 con l’aggiunta dei 10 nuovi Paesi, ma i loro numeri ora sono scesi intorno agli 800 milioni di euro. L’ufficio di traduzione del Canada spende meno di 200 milioni di US\$. Fra le società private, la Microsoft e la Oracle sono quelle che spendono di più. La Microsoft spende circa 300 milioni US\$ all’anno e la Oracle circa 200 milioni di US\$ all’anno. Le società del settore automobilistico spendono tra i 10 e i 35 milioni di US\$ all’anno. Sono veramente poche le società che spendono più di 3 milioni di US\$ all’anno nel mercato, e sono le più ricercate.

* Redditi provenienti dall’esterno rispetto al Paese di origine.

2.6.3. Dalla traduzione umana alla traduzione automatica: straordinaria riduzione dei costi e miglioramento dell'accessibilità

Mettendo da parte per un attimo il problema della qualità, due tra le maggiori variabili che guidano il mercato incentivano l'utilizzo della traduzione automatica: il costo e l'accessibilità. Abbassare i *costi* di un ordine di grandezza alimenta certamente l'utilizzo della traduzione, dato che la domanda è latente, e per ora lontana dall'essere soddisfatta. Una forte crescita nella domanda è altamente probabile e plausibile ed è stata osservata in altri casi, come nei voli aerei a basso costo e nell'industria delle telecomunicazioni con l'avvento del VOIP. Si deve sottolineare che non solamente la traduzione in quanto tale contribuisce all'effetto economico della tecnologia: nuovi tipi di transazioni resi possibili dallo sviluppo tecnologico ne rappresentano una porzione significativa. La seconda variabile è *l'accessibilità*. La maggior parte della traduzione oggi non è prodotta in tempo reale ma su supporto cartaceo in un ufficio distante. Il materiale che deve essere tradotto è generalmente spedito in un altro luogo, e la traduzione viene fornita dopo ore o giorni. La traduzione in tempo reale

esiste, ma è costosa. Se confrontata con tempi di consegna di ore, la disponibilità istantanea di un testo tradotto apre una serie di possibilità per nuove applicazioni, altrimenti inaccessibili.

Due fattori di mercato principali sostengono la traduzione automatica: costo e accessibilità.

In termini di qualità, la traduzione automatica rimarrà inferiore alla traduzione umana per molti anni. Di conseguenza, i diversi segmenti di mercato saranno dominati o dalla traduzione umana o dalla traduzione automatica. Quale prodotto possa prevalere in un particolare segmento di mercato dipende dalla peculiarità delle caratteristiche e dalle specifiche richieste di quel segmento. La traduzione umana prevarrà in tutte le aree dove l'alta qualità rappresenta una assoluta necessità. Invece, la traduzione automatica avrà la meglio sui segmenti più bassi di mercato¹⁴, e dominerà i nuovi mercati o i nuovi segmenti di mercato che emergeranno come conseguenza della disponibilità di tecnologia di traduzione a basso costo. Con miglioramenti di qualità e di prestazioni, nel tempo, la traduzione automatica aumenterà la sua quota di mercato.

¹⁴ Basso si riferisce alla qualità della traduzione. In termini di tempo richiesto e di accesso la traduzione automatica è chiaramente nel segmento alto di mercato. Ci si aspetta che produca volumi alti.

Intervista con Michael Anobile, Managing Director di LISA

Michael Anobile ha oltre 25 anni di esperienza nel settore IT a livello internazionale, si è laureato in Scienze della comunicazione all'Università di Syracuse, ed ha un fatto parte del programma di Master in political communications dell'Università del Maryland. Dopo essersi trasferito con la famiglia in Svizzera nel 1980 è diventato European Training Manager per l'Exxon Office Systems, e successivamente ha ricoperto una serie di ruoli come senior management europeo e svizzero nel settore dell'IT e nelle industrie della tecnologia del linguaggio, focalizzandosi sullo sviluppo di un business globale e sul marketing.



Michael Anobile
Managing Director
The Localization Industry Standards
Association (LISA)
Romainmôtier, Schweiz

Membro fondatore di LISA (*The Localization Industry Standards Association*), e Managing Director fin dall'inizio, è responsabile della gestione ordinaria dell'associazione, inclusa la consulenza alle organizzazioni che sviluppano standard (e.g., ISO, Unicode, Open18N, W3C, OASIS, etc.) e alle agenzie governative (ad esempio., US Department of Commerce, the DoD, FBI, NVTC, NSA insieme a diverse istituzioni asiatiche, canadesi ed europee che si occupano di standard nelle tecnologie del linguaggio). E' anche responsabile di forum internazionali, programmi di training e progetti di marketing industriale & pubbliche relazioni.

Di cosa si occupa LISA?

Siamo un'organizzazione guidata dai membri che si focalizza su società, governi e NGOs. Oltre al settore delle SME, lavoriamo a livello istituzionale – gruppi come World Bank, McDonald's, Coca-Cola, IBM, European Directorate-General for Translation, Industry Canada o il Canadian Bureau of Translation. Alcuni di questi gruppi comprendono più di 1.000 traduttori. Noi li aiutiamo a capire la prospettiva globale che la localizzazione offre in termini di business, tecnologia, e flussi di lavoro. Questo comprende l'internazionalizzazione, la traduzione, e come progettare i prodotti e i servizi per il mercato globale.

Qual'è la sua principale raccomandazione per fare localizzazione in modo corretto?

La *best practice* richiede che un prodotto sia progettato per un utilizzo internazionale. Perciò sono temi cruciali lo sforzo di internazionalizzazione, il ciclo di vita del prodotto o del servizio, i mercati e la distribuzione. Questo approccio "olistico" presuppone che l'informazione sia tradotta. Quindi, il modo in cui il processo di traduzione viene automatizzato è fondamentale per la qualità, per il consenso dell'utente finale e per il costo.

Potrebbe stimare la domanda di traduzione, se fosse fornita gratuitamente?

Non ne ho idea. Comunque, l'esperienza dimostra che se si chiede ad un cliente se acquisterebbe o meno una traduzione fatta in modo automatico, è molto probabile che risponda: no. Se si rende disponibile una traduzione che è fatta da un sistema automatico, e poi la si offre al cliente come scelta gratuita, è più probabile che scarichi e utilizzi quel documento. Il dilemma è evidente: C'è sempre una certa resistenza verso una "Tecnologia di rottura".

Che contributo ha dato la tecnologia alla traduzione?

La Translation Memory è uno dei principali contributi che l'industria ha dato alla traduzione. Tra l'altro, questa è una tecnologia che risale a vent'anni fa. Guardate quanto ha impiegato a farsi strada. Non posso che confermare l'efficacia di TRADOS, SDL e altri sviluppatori di TM come Atril e Logos, perché hanno lavorato veramente duro per implementare le Translation Memory su vasta scala, in molti settori commerciali. TM fa suo il concetto fondamentale dell'editing: se avete digitato qualcosa una volta, poi non dovrete digitarlo più. Così quando traducete qualcosa che sapete che sarà utilizzato ancora, poi non avete più bisogno di tradurlo ancora. Naturalmente, dovete controllare in quale contesto state operando, e assicurarvi della riusabilità attraverso varie piattaforme e strumenti. Questo è il modo in cui la TM standard di LISA, TMX[®], contribuisce all'industria.

Qual è il suo punto di vista nei confronti della traduzione automatica?

La traduzione automatica (MT) risulta molto efficace ed efficiente in termini di costi nelle applicazioni di traduzione ben definite come i database della conoscenza, i call centers, e la documentazione tecnica. La *best practice* dell'industria è di controllare i processi di gestione della terminologia e di *authoring*. Capire come l'utente finale interagirà con l'applicazione, capire il livello di informazione e la qualità richiesta, renderà possibile la razionalizzazione del processo di traduzione, definendo, costruendo e mantenendo terminologie specifiche. Di conseguenza si avranno una migliore accuratezza di traduzione e una maggiore diffusione. MT lavora molto bene quando si soddisfano le aspettative e si impiegano le risorse adeguate.

Una delle intenzioni di questo report è di far aumentare la consapevolezza per le tecnologie di traduzione automatica ma anche per la traduzione stessa. Qual'è il suo messaggio?

È molto importante che i nostri leader politici ed economici capiscano l'importanza delle lingue. La lingua facilitata, non ostacola, aumenta la comprensione e la cooperazione tra le persone e le culture. La lingua fornisce un ampio accesso alle informazioni sociali e politiche più importanti. Sono d'accordo con l'obiettivo di aumentare la consapevolezza pubblica, poiché si tratta di

qualcosa di più rispetto alla semplice traduzione. Si tratta di comunicare e di dare accesso a dati rilevanti che possono aiutare le persone ad essere più protagoniste dal punto di vista economico, sociale e politico. Un buon esempio è costituito dalla politica sul linguaggio dell'*European Union Directorate of Translation's* per sostenere lo sviluppo dei mercati nell'Europa Orientale e per i nuovi Paesi membri.

2.6.4. *Innovazioni tecnologiche incrementali e di rottura*

Non è strano come grandi imprese, leader di mercato e ben gestite, non riescano a sfruttare le innovazioni emergenti? Bene, ora capiamo le dinamiche dell'innovazione molto meglio di dieci anni fa. Sono alcune caratteristiche innovative che determinano in gran parte il successo sul mercato e la possibilità di creare un leader di mercato totalmente nuovo; queste caratteristiche sono utilizzate per classificare le innovazioni *in innovazioni di supporto* e *innovazioni di rottura*, come definite di seguito¹⁵.

La maggior parte delle innovazioni tecnologiche migliorano le prestazioni di un prodotto. Questo tipo di innovazioni sono chiamate *di supporto* (*sustained*). Siano esse incrementali o discontinue, hanno in comune il miglioramento delle prestazioni di determinati prodot-

ti secondo criteri storicamente consolidati nei maggiori mercati.

Alcune volte, invece, le innovazioni peggiorano le prestazioni di un prodotto, almeno nel breve periodo, ma possiedono altre caratteristiche che hanno valore per nuovi clienti. Queste innovazioni si definiscono *innovazioni di rottura* (*disruptive*).

Le innovazioni di supporto tendenzialmente non cambiano il panorama di mercato, cioè il leader di mercato le sviluppa (o le compra) e mantiene la propria leadership. Poiché invece le *innovazioni di rottura* offrono prestazioni inferiori rispetto allo stato dell'arte o all'offerta di prodotti esistenti, esse non hanno valore per il leader di mercato in quel particolare momento.

Per alcuni meccanismi plausibili e quantitativamente verificabili, questo conduce ad una situazione dove nuovi attori scendono in campo ed estendono

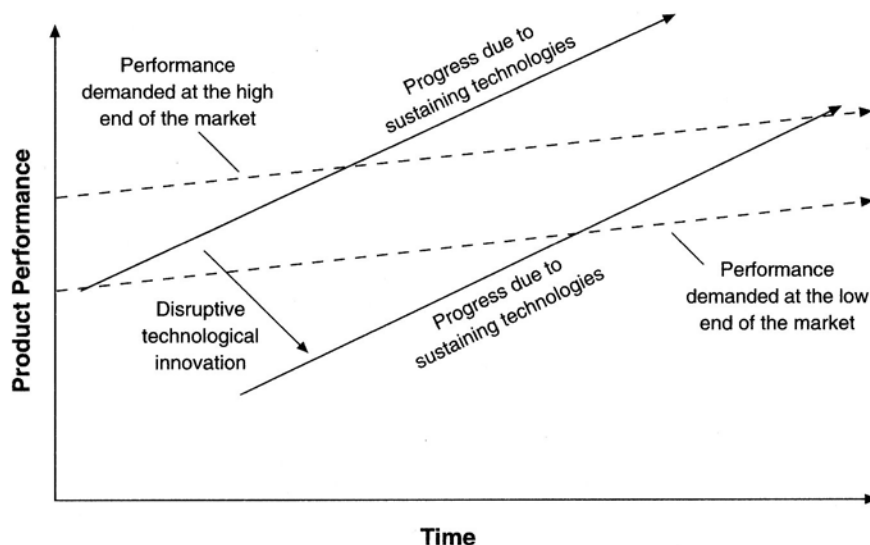


Fig. 3: Il cambiamento tecnologico indotto dalle innovazioni di sostegno e di rottura. – Fonte: [CRa].

¹⁵ I due paragrafi successivi sono vicini a [CRa].

Tecnologie consolidate	Tecnologie di rottura
Minicomputer	PC
Computer portatili	PDA
Fotografia con alogenuro d'argento	Fotografia digitale
Telefonia a rete fissa	Telefonia mobile
Microsoft OS e Office	Linux e Open Office
Hard Disks drives	Flash Memory

Tabella 8: Esempi di innovazioni di rottura, insieme con le tecnologie corrispondenti consolidate. – Fonte: [Chr], [CRa].

una piccola nicchia di mercato, che è di scarso interesse per il leader di mercato, finché essa non cresce ed assume una dimensione rilevante.

È interessante notare come il leader di mercato, seguendo indicazioni di buon management, decida di non perseguire la nuova opzione.

Con le innovazioni di rottura, storicamente, il leader di mercato è di solito incapace di conservare la propria posizione e viene rimpiazzato da un altro attore di mercato. Questo pone un'interessante domanda per gli attori di mercato affermati ed entranti, ma anche per i politici: la traduzione automatica e la traduzione del parlato sono *innovazioni di sostegno o di rottura?*

Nel caso della traduzione del linguaggio parlato, l'innovazione è di rottura: non è facile vedere come questa tecnologia possa supportare gli interpreti umani, ed inoltre se confrontata con la traduzione umana, le sue prestazioni, sebbene stiano migliorando, rimangono tuttavia molto inferiori. Ci si aspetta quindi che questa tecnologia prenderà piede in mercati molto diversi da quelli esistenti ed estenderà la sua quota di mercato nel tempo. Questi mercati target, d'altro canto, non sono coperti da traduttori umani (pensiamo alla traduzione dal vivo delle trasmissioni televisive).

La traduzione automatica può essere considerata un'innovazione sia di sostegno che di rottura, in relazione al tipo di applicazione coinvolta e al mercato in cui viene introdotta. Le tecniche della *translation memory*, per esempio, stanno diventando uno strumento di preparazione e di supporto alla traduzione umana, sia facilitando la traduzione che migliorandone la qualità. In questo contesto la traduzione automatica può essere vista come un'innovazione di sostegno: migliora quello che già esiste.

In netto contrasto, la maggior parte delle altre applicazioni della traduzione automatica possono essere considerate di rottura. Così come nel caso della traduzione del linguaggio parlato, anche i servizi di traduzione on line di siti web, delle *chat on line* ed i servizi *self service* per cercare documenti in varie lingue danno prestazioni inferiori alla traduzione umana in termini di qualità. Il vantaggio in termini di costo, però, consente la crescita di questi mercati, che rimangono poco attraenti per i servizi di traduzione umana a causa dei bassi margini di guadagno.

MT e SLT sono essenzialmente innovazioni di rottura come il PC o la fotografia digitale: Iniziano da una nicchia e si rafforzano in seguito.

3. Asia, Europa e Stati Uniti: analogie e differenze

I due eventi che hanno guidato la tecnologia della traduzione negli ultimi anni sono stati l'11 settembre e l'allargamento dell'UE.

3.1. Unione Europea

La situazione relativa alle lingue e il bisogno di tecnologie del linguaggio è stato presentato ampiamente nel paragrafo 2.3.: *Uno Sguardo da Vicino all'Unione Europea*. Ricapitolando i punti principali:

- L'Europa ha una società multilingue ed è multilingue per definizione.
- Tutte le lingue ufficiali europee hanno gli stessi diritti.
- La traduzione tra queste venti lingue ufficiali, formando 190 coppie linguistiche, richiede un notevole sforzo.
- Qualsiasi misura atta a facilitare la traduzione alimenterebbe la comunicazione tra i cittadini europei e amplificherebbe il commercio tra i Paesi dell'UE.
- Qualsiasi società che tratta con il mercato dell'UE deve localizzare i propri prodotti rispetto ai mercati

regionali presentandoli nelle lingue locali.

La Commissione Europea è impegnata nella ricerca sulle HLT e ha finanziato nel tempo alcuni progetti di ricerca sulla traduzione automatica. Si può accedere velocemente a queste informazioni mediante il sito Web dell'IST (Information Society Technologies) tramite la sezione di *project search*¹⁶. Una lista non esaustiva comprende LC-STAR¹⁷, MATCHPAD¹⁸, METIS e METIS-II¹⁹, NESPOLE!²⁰, TC-STAR_p e TC-STAR²¹, TQPRO²² e TransType2²³.

La Commissione Europea ha giocato un ruolo decisivo nella produzione delle risorse linguistiche necessarie, finanziando molti progetti in quest'area.

Nell'ambito del VI Programma Quadro, l'Unione Europea ha speso 135 milioni di euro sulle interfacce multimodali e sulle tecnologie del linguaggio, i.e. approssimativamente 15 milioni di euro all'anno sulle tecnologie del linguaggio.

La comunità di ricerca europea nell'area del linguaggio ha tutte le caratteristiche per poter affrontare un com-

¹⁶ IST project search: under <http://www.cordis.lu/ist/projects/projects.htm>

¹⁷ LC-STAR: <http://www.lc-star.com/>

¹⁸ MATCHPAD: <http://www.systransoft.com/R&D/Matchpad/index.html>

¹⁹ METIS-II: <http://www.ilsp.gr/metis/>

²⁰ NESPOLE! : <http://nespole.itc.it>

²¹ TC-STAR: <http://www.tc-star.org/>

²² via the IST page: under <http://www.cordis.lu/ist/projects/projects.htm>

²³ TransType2: <http://tt2.atosorigin.es/>

pito così complesso. L'Europa ha molti istituti di ricerca pubblici di qualità riconosciuta a livello internazionale, che permettono di affrontare diversi settori di ricerca e di intraprendere una varietà di approcci. I concetti di multilinguismo e di multiculturalismo sono generalmente ben compresi e condivisi sia a livello accademico che a livello economico e sono fortemente sostenuti a livello politico. Inoltre è operativa da dieci anni un'infrastruttura per la preparazione e distribuzione delle risorse linguistiche, l'European Language Resources Association (ELRA).

3.2. Stati Uniti d'America

3.2.1. Ruolo strategico delle HLT

Il ruolo strategico delle HLT negli USA si differenzia in modo sostanziale da quello europeo. Gli Stati Uniti hanno un vasto mercato domestico con sostanzialmente una sola lingua base. Questo rimane vero nonostante la presenza di una considerevole popolazione ispanica che parla Spagnolo Americano. Comunque, non c'è una forte necessità economica di servire questo gruppo economicamente debole, e non ci sono requisiti legali per dover supportare questa lingua. Gli Stati Uniti sono essenzialmente un mercato basato su una sola lingua.

Il predominio internazionale dell'inglese come lingua straniera è un vantaggio per gli Stati Uniti, sotto molti aspetti, ma porta un grave svantaggio in relazione alla sicurezza nazionale. Questo fatto attribuisce alla traduzione, e in modo particolare dalle altre lingue

verso l'inglese, un ruolo cruciale per la raccolta di informazioni dal mondo e riguardanti il mondo, sia di interesse generale che in relazione ad operazioni di sorveglianza.

L'interesse degli US nelle HLT è guidato in gran parte da aspetti di sicurezza nazionale.

3.2.2. Programmi di ricerca

Dopo gli attacchi dell'11 settembre 2001 al World Trade Center, è risultata evidente la necessità di avere a disposizione dei traduttori competenti in *qualsiasi* lingua e con breve preavviso. Questa è stata la ragione per cui gli Stati Uniti hanno lanciato un programma di ricerca sulla raccolta di informazioni in lingue diverse e da fonti multiple, inclusa la traduzione automatica. Il finanziamento ammonta ad oltre 50 milioni di dollari all'anno.

Questi finanziamenti – che sono un multiplo dell'ammontare speso dalla Commissione Europea – e il programma stesso avranno un impatto sulla comunità della ricerca. Mentre alcuni dei progressi tecnologici e scientifici saranno in linea con i bisogni specifici dell'Europa, altri non lo saranno – la raccolta di informazione per scopi militari è, per diversi aspetti, differente dal compito di tradurre per una comunità multilingue. Inoltre, la struttura essenzialmente monolingue del mercato interno statunitense fa nascere ragionevoli dubbi sul fatto che il mercato della tecnologia del linguaggio sviluppata negli USA possa offrire ciò di cui l'Europa ha bisogno.

Senza entrare nei dettagli del programma di ricerca, si può affermare

in modo generale che gli USA tendono a finanziare un numero piccolo di grandi progetti mentre l'UE finanzia grandi programmi che comprendono piccoli progetti. L'attuale fiore all'occhiello è costituito dal progetto GALE (Global Autonomous Language Exploitation) che si focalizza sul-

l'estrazione di informazioni da testo multilingue e da documenti audio all'interno di un dominio illimitato (il progetto affronta anche la traduzione del parlato). Ci sono altri progetti in corso che si occupano di traduzione come TransTac, STR-DUST e ACTD.

UNCLASSIFIED

RDT&E BUDGET ITEM JUSTIFICATION SHEET (R-2 Exhibit)							DATE		
APPROPRIATION/BUDGET ACTIVITY RDT&E, Defense-wide BA2 Applied Research							R-1 ITEM NOMENCLATURE Information and Communications Technology PE 0602303E, Project IT-04		
COST (In Millions)	FY 2004	FY 2005	FY 2006	FY 2007	FY 2008	FY 2009	FY 2010	FY 2011	
Language Translation IT-04	0.000	57.389	65.744	69.687	75.221	75.593	65.593	60.593	

(U) **Mission Description:**

(U) This project will develop and test powerful new technology for processing human languages that will provide critical capabilities for a wide range of national security needs. This technology will enable systems to (a) automatically exploit large volumes of speech and text in multiple languages; (b) revolutionize human-computer interaction via spoken and written English and foreign languages; (c) perform computing and decision-making tasks in stressful, time-sensitive situations; and (d) autonomously collate, filter, synthesize and present relevant information in timely and relevant forms. This program element and project were created in accordance with congressional intent in the FY 2005 DoD appropriations bill. Prior year funding was budgeted in PE 0602301E, Project ST-29, and is noted as a memo entry in each program below.

(U) **Program Accomplishments/Planned Programs:**

	FY 2004	FY 2005	FY 2006	FY 2007
Situation Presentation and Interaction	(10.870)	11.500	11.616	14.387

(U) There are two programs involving direct speech-to-speech translation:

- The Compact Aids for Speech Translation (CAST) program is providing the tactical warfighter with real-time, face-to-face speech translation during combat and humanitarian operations in foreign territories. The program addresses domain-specific translation accuracy and response time. Early CAST prototypes relied on simple dictionaries and phrases. The CAST program resulted primarily in quickly making one-way translation systems (from English to multiple foreign languages) available to warfighters in the field. The DARPA Phrasclator is the key prototype system in use today. The system was deployed in Operation Iraqi Freedom and Operation Enduring Freedom. Future versions will offer a more sophisticated, flexible and fluid translation and paraphrasing capability that is robust and conducive to normal human conversations.

Fig. 4: Budget del DoD (Department of Defense) degli Stati Uniti per le tecnologie del linguaggio per la traduzione. Stime di budget per l'anno fiscale 2005 per la ricerca, lo sviluppo, il test e la valutazione (RDT&E), difesa globale, presso DARPA (Defense Advanced Research Projects Agency). – Fonte: DARPA.

Intervista con Joseph Olive, Program Manager, DARPA

Joseph Olive è il program manager dell'Information Processing Technology Office di DARPA. Il suo attuale portfolio comprende un grande programma, GALE (Global Autonomous Language Exploitation).

Ha più di trent'anni di esperienza in ricerca e sviluppo ai Bell-Labs. È stato il leader mondiale nella ricerca della sintesi da testo a parlato e ha guidato un gruppo di livello internazionale nei sistemi di dialogo con il computer e comunicazione uomo macchina. Nel suo ruolo di Direttore della ricerca sul parlato e di CTO della Lucent Speech Solutions, ha supervisionato la produzione delle tecnologie del parlato dei Bell-Labs: il riconoscimento, la sintesi e la verifica del parlatore.

Joseph Olive si è laureato all'Università di Chicago con una laurea in Fisica (fisica atomica computazionale) e un M.A. in composizione musicale. Dopo aver finito l'Università, ha combinato il suo interesse per la computazione e per la musica e ha iniziato la ricerca in acustica e nell'elaborazione del segnale.

Ha vinto un premio della Fondazione Nazionale per le Arti nel 1974 per aver scritto un'opera sul computer. Ha ricevuto anche un riconoscimento dai Bell-Labs come Distinguished Member del Technical Staff nel 1984.



Joseph Olive

Program Manager
IPTO (Information Processing Technology Office)
DARPA - Arlington, VA, USA

Quali attività di ricerca e obiettivi sono perseguiti nel programma GALE?

Global Autonomous Language Exploitation (GALE) è un programma che fornirà informazioni, concise e utilizzabili dai nostri militari. Dal momento che la fonte può essere in una lingua straniera, GALE richiede una traduzione e una sintesi. L'input può essere testo o parlato e l'output può essere una traduzione completa o una risposta filtrata ad una domanda (non necessariamente in linguaggio naturale). Lo scopo di GALE è di raggiungere un'accuratezza di traduzione e di selezione tale da rendere utilizzabile il sistema dai militari. GALE fa seguito a due precedenti programmi di DARPA in HLT: EARS – un programma di trascrizione e TIDES – un programma per la traduzione, la ricerca, l'estrazione ed la sintesi di informazione. I programmi precedenti non erano stati pensati per diventare un unico sistema di traduzione, ma hanno permesso di ottenere grandi progressi nelle HLT.

Qual è la sua posizione relativamente alla cooperazione transatlantica?

Ci sono alcuni gruppi europei coinvolti in GALE. Sono sottocontraenti dei grandi team creati dai "Principle Investigators" (PI) di GALE e stanno collaboran-

do all'interno dei loro team. In generale, credo che le collaborazioni tra gli scienziati degli Stati Uniti e gli scienziati Europei/Asiatici siano buone. Queste relazioni sono a diversi livelli, all'interno del governo, degli enti e della ricerca universitaria.

Quali sono i finanziamenti pubblici negli Stati Uniti per le tecnologie del linguaggio come la traduzione automatica?

Il programma totale delle HLT in DARPA ammonta approssimativamente a 50 milioni di US\$ per il 2006. Non credo che dovrebbe essere ridotto perchè le HLT sono (o dovrebbero essere) collegate e integrate per raggiungere il successo in questo campo. Non ho informazioni riguardo alla spesa degli altri enti governativi sulle HLT, ma sono sicuro che DARPA non è l'unica organizzazione interessata a questo settore.

Quanto sono importanti per gli US le HLT, in modo specifico MT e le tecnologie collegate?

Personalmente sento che le HLT sono estremamente importanti per il governo degli US e per il settore militare. È necessario capire e comunicare con i nostri alleati e nemici, e molti di loro non parlano inglese. Anche a causa dell'enorme quantità di informazione, è cruciale cercare l'informazione importante come si cerca "l'ago nel pagliaio".

Come è organizzato il finanziamento pubblico alla ricerca nelle HLT negli Stati Uniti?

Non ho visto un piano generale per la ricerca nelle HLT negli Stati Uniti. C'è un investimento da parte di diverse agenzie nazionali e da parte dell'industria. C'è molta collaborazione tra la ricerca sponsorizzata dal governo, gli enti privati e la ricerca universitaria, ma è più informale. Per quanto riguarda DARPA, noi abbiamo una visione e degli obiettivi molto ambiziosi per le HLT. Organizziamo delle valutazioni periodiche per assicurare il progresso. Lavoriamo sia attraverso la cooperazione che la competizione. I nostri team hanno l'obiettivo di vincere le valutazioni. Inoltre, noi abbiamo legato i nostri obiettivi e le valutazioni all'utilità della tecnologia piuttosto che solamente alla misurazione dell'accuratezza. Le HLT hanno a che fare con le lingue ed il primo e principale obiettivo è di preservare il significato della lingua nei documenti (in forma scritta o parlata).

Qual è lo stato dell'arte e quali sono le prossime sfide scientifiche che devono essere affrontate nelle tecnologie del linguaggio, in particolare nell'ambito della traduzione?

La tecnologia della traduzione automatica è migliorata molto nei due o tre anni passati. Gran parte di questo miglioramento è dovuto a paradigmi statistici associati a procedure di ottimizzazione che utilizzano BLEU. Comunque, temo che questo paradigma stia raggiungendo velocemente il suo limite, se già non lo ha

raggiunto. Penso che la traduzione automatica è solo a metà strada rispetto a dove dovrebbe essere. Vorrei vedere nuovi approcci al problema, anche se questo significa avere risultati non competitivi all'inizio. È necessario per le tecnologie sia di MT che di ASR adottare un approccio che incorpori altre tecniche NLP come IR, analisi linguistiche, estrazione di informazioni, etc. Sebbene anche queste siano tecniche statistiche, non lavorano allo stesso modo e quindi un sistema combinato utilizzando tutte queste tecnologie, potrebbe migliorare i risultati ed offrire una soluzione al problema. È estremamente importante integrare in modo più stretto l'ASR e la MT, non utilizzando solo il risultato 1-best dell'ASR per guidare la componente di MT.

3.3. Asia Orientale

3.3.1. L'inglese come 'lingua franca' in Asia Orientale

Considerata la sua varietà di popoli, culture e lingue, l'Asia assomiglia più all'Europa che agli Stati Uniti. Ma mentre l'Europa ha iniziato a considerare se stessa come un'entità unica, in Asia non c'è ancora una politica coerente e un'identità di unità autonoma. I Paesi, per comunicare tra loro utilizzano essenzialmente l'inglese, e l'ampia diffusione dell'uso dell'inglese come *lingua franca* in Asia implica che c'è una domanda significativamente minore per la traduzione diretta tra le coppie linguistiche come il tailandese – giapponese. Come vantaggio, ogni Paese deve assicurare in primo luogo che sia garantita la traduzione della propria lingua da e verso l'inglese, e la restrizione ad un'unica coppia di lingue per lingua riduce la complessità.

3.3.2. Le coppie linguistiche asiatiche e l'importanza crescente del cinese

C'è anche uno svantaggio legato alla dipendenza dell'Asia dall'inglese come *lingua franca*. Si consideri, per esempio, la traduzione tra giapponese e cinese: date le radici comuni di queste lingue, l'etimologia delle parole è simile. Sarebbe naturale trarre vantaggio da questo fatto, così come dalle somiglianze culturali. L'utilizzo dell'inglese come mediatore è indiretto e conduce a complicazioni non necessarie. Seguendo un nuovo trend, le società giapponesi si stanno interessando in modo crescente alla traduzione diretta tra il giapponese e il cinese. Questo

è collegato alla fiorente economia cinese e al fatto che molte società giapponesi trasferiscono la loro produzione manifatturiera in Cina. L'interesse nella traduzione da e verso il cinese sta aumentando sia a livello globale che in Giappone, e le relazioni tra la Cina e il Giappone sono più strette rispetto al passato.

L'inglese è una lingua franca in Asia. Il cinese sta acquistando importanza.

Economicamente forte e con una lingua che ha le proprie radici nel cinese antico, la Corea è un Paese con forti relazioni sia con la Cina che con il Giappone. Sia le parole giapponesi che quelle coreane hanno spesso origine cinese, cioè condividono un'etimologia simile, anche se il coreano non usa caratteri cinesi. I termini tecnologici sono spesso usati inizialmente in giapponese, utilizzando caratteri cinesi, e successivamente sono assorbiti dal cinese. Questa somiglianza lessicale aiuta molto nella traduzione. Tuttavia, le tre lingue hanno diverse pronunce e sono abbastanza diverse da un punto di vista linguistico. Considerando questo bagaglio culturale comune e il fatto che il commercio e l'interazione tra i tre Paesi è più forte che con altri Paesi asiatici, si può assumere che per le tre lingue, cinese, giapponese e coreano verranno probabilmente sviluppati traduttori diretti nel prossimo futuro, mentre per le altre lingue asiatiche i traduttori sfrutteranno l'inglese come lingua intermedia.

In un prossimo futuro per il cinese, il giapponese e il coreano saranno sviluppati traduttori diretti.

3.3.3. Programmi di ricerca

In Giappone, la ricerca nelle tecnologie del linguaggio umano aveva raggiunto l'apice negli anni '80 grazie ad un elevato livello di finanziamenti pubblici e all'interesse delle società verso progetti di traduzione automatica come il progetto *EDR, Electronic Dictionary Project*. Poichè le grandi aspettative per la traduzione automatica non furono soddisfatte, come per il progetto europeo *EUROTRA*, l'interesse e i finanziamenti pubblici cominciarono a ridursi. Oggi ci sono segnali di un'inversione di tendenza.

Dato il ruolo predominante dell'inglese, l'attenzione della ricerca sulla traduzione è relativa alla traduzione dall'inglese alla lingua madre e viceversa, e ciò sostanzialmente conduce a programmi di ricerca condotti a livello nazionale.

Sarebbe difficile eseguire progetti asiatici transnazionali, dal momento che non esistono appropriate organizzazioni asiatiche. Comunque, alcuni aspetti del lavoro di ricerca richiedono cooperazione internazionale, che auspicabilmente è fornita e sostenuta dalle associazioni asiatiche come l'*Asian Federation of Natural Language Processing*²⁴ che è stata recentemente costituita.

Al fine di alimentare la ricerca sulla traduzione tra le lingue asiatiche, sarebbe piuttosto importante costruire risorse linguistiche per queste coppie linguistiche. Dato lo sforzo richiesto e la necessità di condividere risorse, il modo naturale per raggiungere questi obiettivi è attraverso la cooperazione internazionale. Un ulteriore passo di grande importanza sarebbe la costituzione di un'agenzia di valutazione.

²⁴ <http://afnlp.org>

Intervista con Jun-ichi Tsujii, Direttore del National Centre for Text Mining a Manchester e Professore a Manchester e a Tokyo

Jun-ichi Tsujii è dal Luglio 2005, direttore del *National Centre for Text Mining* e Professore in Text Mining alla *School of Informatics*, Università di Manchester, Regno Unito. È anche Professore in Natural Language Processing al *Department of Computer Science*, Università di Tokyo, Giappone. Ha lavorato nel campo del natural language processing dal 1976. Ha iniziato con la traduzione automatica, ha ampliato la sua ricerca ai formalismi grammaticali per applicazioni pratiche NLP, HPSG-based parsing, information extraction e intelligent question answering. Il suo team di ricerca su NLP all'Università di Tokyo recentemente è riuscito ad applicare con successo un analizzatore linguistico accurato per produrre rappresentazioni semantiche di tutti gli abstracts di Medline (1,4 miliardi di parole).



Jun-ichi Tsujii

Direttore del National Centre for Text Mining, Manchester, UK
Professore presso la School of Informatics, University of Manchester, UK
e presso il Department of Computer Science, University of Tokyo, Japan

È riconosciuto come una delle figure leader nel campo del bio-text mining, della traduzione automatica e del NLP multilingua, ed è stato un appassionato promotore della cooperazione attraverso tutta l'Asia. È stato invitato per *tutorial*, *talk* e *keynote speeches* a numerose importanti conferenze, sia in bio-informatica che NLP. È membro permanente dell'ICCL (International Committee of Computational Linguistics, dal 1992), vice-presidente (2005) e presidente (2006) dell'ACL (Association of Computational Linguistics), presidente (2003-2005) dell'IAMT (International Association of Machine Translation), e vice-Presidente dell'AFNLP (Asian Federation for Natural Language Processing).

La traduzione automatica è stata l'argomento di grandi programmi di ricerca negli anni '80, ma i risultati non hanno incontrato le aspettative. Come vede questi sforzi oggi?

È certamente vero che gli sforzi degli anni '80 non hanno soddisfatto le aspettative di un largo mercato potenziale. Per una persona come me che è stata coinvolta in alcuni di questi progetti, direi che è stato deludente. Comunque, grazie a questi sforzi, c'è stata predisposta una solida base per la ricerca e lo sviluppo di sistemi di MT. In Giappone, almeno sei o sette produttori di MT sono ancora attivi sul mercato. Inoltre, penso che, le visioni che avevamo allora precorressero i tempi, in quanto le tecnologie necessarie a raggiungere i nostri obiettivi non erano disponibili. Credo che, attualmente, dal momento che oggi abbiamo le tecnologie che mancavano allora, siamo in grado di rivitalizzare il settore. Per esempio, lo stesso task di eseguire l'analisi sintattica di una frase attraverso sofisticati formalismi grammaticali che richiedeva ore di calcolo, adesso richiede meno di un secondo con il nostro programma all'Università di Tokyo.

Le tecnologie del linguaggio, compresa la traduzione automatica, dovrebbero essere importanti per l'Asia. Questo è percepito a livello nazionale?

Sfortunatamente no. A differenza di quanto succede in Europa, l'inglese è stato considerato a lungo come l'unica lingua internazionale per la comunicazione e, a causa di questo, le persone non pensano che sia importante considerare le "lingue locali" come il cinese, il coreano, il giapponese etc. Ovviamente, questo è sbagliato, e le persone cominciano a capirlo. A causa di Internet, ci siamo resi conto che c'è un'enorme richiesta di elaborare lingue "locali", o meglio, che le nostre lingue locali non sono realmente locali.

Quali sono le maggiori differenze delle lingue asiatiche rispetto a quelle europee? Quanto sono simili le principali lingue asiatiche tra loro?

Dipende tutto dalle definizioni di lingue europee ed asiatiche. Gli Europei possono dichiarare che le lingue in Europa sono diverse, e io sono d'accordo. Comunque, penso che la diversità tra le lingue asiatiche è enorme e va ben oltre quella tra le lingue europee. Molte lingue in India, nel Medio Oriente, in Malesia, etc. sono completamente scollegate dalle lingue dell'Estremo Oriente, il cinese, il coreano e il giapponese. Inoltre, il cinese, e il giapponese/coreano, mentre condividono vocabolari comuni grazie ad una lunga storia di scambi culturali, appartengono a famiglie linguistiche completamente diverse.

Come descriverebbe la situazione in tutti i Paesi asiatici in generale?

Non sono la persona giusta per rispondere a questa domanda, ma la situazione nei paesi asiatici è piuttosto diversificata. Alcune lingue non hanno nemmeno codificato un sistema comune di trascrizione, e tantomeno codici standard per rappresentare caratteri. Comunque, dal momento che i progressi tecnologici in molti Paesi, come l'India, la Thailandia, la Cina, etc. sono stati accelerati, i livelli tecnologici di questi Paesi sono più o meno simili. Vediamo molti articoli di ricerca interessanti pubblicati dai ricercatori in questi Paesi.

Ci sono attività internazionali in Asia che coordinano la ricerca in questo settore o addirittura una collaborazione a livello globale?

Abbiamo fondato due anni fa un'associazione accademica, AFNLP (Asian Federation for Natural Language Processing) al fine di promuovere la cooperazione e di coordinare le attività. Ci sono altre iniziative naturalmente. Comunque, in confronto al livello di coordinamento europeo, noi siamo ancora lontani.

Quali sono i prossimi passi che devono essere fatti al fine di promuovere le HLT e la MT in Asia?

Noi abbiamo bisogno di un maggiore coinvolgimento governativo. L'UE ha giocato un ruolo significativo nella promozione della ricerca e nello sviluppo delle HLT e della MT. Fino alla fine degli anni 80', il Giappone è stato all'avvan-

guardia non solo nella tecnologia ma anche nei finanziamenti alla cooperazione regionale nel settore. La situazione è cambiata in modo straordinario da allora. Vediamo alcuni Paesi che possono contribuire finanziariamente e intellettualmente al settore. Comunque, non abbiamo organizzazioni governative panasiatiche per coordinare le attività. La comunità accademica è pronta a cooperare, ma abbiamo bisogno, per esempio, di enti di finanziamento che supportino i grandi progetti in Asia.

3.4. India

È possibile pensare ad un Paese con così tanta diversità linguistica come l'Europa unita? Bene, l'India²⁵ e i suoi 35 stati hanno 22 lingue ufficiali, cioè approvate dalla costituzione. E come se non ci fosse sufficiente varietà, ogni lingua ha in media una ventina di dialetti. Le lingue appartengono a tre differenti famiglie linguistiche, e ci sono anche diversi sistemi di scrittura. La lingua nazionale è l'hindi, ma la costituzione Indiana afferma che anche l'inglese può essere utilizzato per scopi ufficiali. Nelle città, le persone spesso conoscono tre lingue: la lingua nazionale (e.g. hindi, bengalese, etc.), l'hindi (la lingua ufficiale dell'India), e l'inglese. Approssimativamente il 30% della popolazione parla l'hindi, e circa il 5% della popolazione si trova a proprio agio con l'inglese.

Tutti i documenti ufficiali nelle capitali di Stato devono essere in tre lingue (in inglese, in hindi e nella lingua dello Stato). Le direzioni di traduzione impiegate più frequentemente sono dall'inglese all'hindi e dall'inglese o hindi nelle rispettive lingue di stato. Come in altri Paesi, comunque, la maggior parte delle traduzioni è effettuata da persone e non ci sono sufficienti traduttori per rispondere alla domanda. Relativamente agli strumenti tecnici, la *translation memory* non è molto utilizza-

ta, ma la traduzione è spesso supportata da dizionari elettronici e da strumenti come analizzatori morfologici.

La ricerca sulla traduzione è finanziata sia dal governo che da società internazionali come IBM, Microsoft, Google e Yahoo benché esse non trattino ancora tutte le lingue indiane. Dal momento che queste società di norma stanno utilizzando l'approccio statistico, molti corpora paralleli²⁶ esistono ma non sono pubblicamente disponibili. A partire dal 2006, questo gap sarà colmato dall'agenzia LDC-IL²⁷ (*Linguistic Data Consortium for Indian Languages*), sponsorizzata dal nuovo governo.

3.5. Le condizioni economiche di contesto

In Europa le condizioni economiche di contesto relativamente alle tecnologie del linguaggio sono diverse da quelle degli USA. Di fatto, le condizioni di mercato per il riconoscimento automatico del parlato e per la traduzione automatica negli USA e nell'UE sono diametralmente opposte.

Diamo un'occhiata prima di tutto al riconoscimento automatico della voce. In questo momento, l'utilizzo economicamente più interessante del riconoscimento automatico e delle tecnologie di dialogo è relativo all'automazione dei servizi di *call center*.

²⁵ Si ringrazia il Professor Dr. Pushpak Bhattacharyya (Indian Institute of Technology, Mumbai) per il briefing sulla situazione in India.

²⁶ Un *corpus parallelo* è una collezione di testi in due versioni di lingue, insieme all'informazione che le frasi sono associate l'una all'altra. (E.g., le prime due frasi nella lingua A possono corrispondere alle prime tre frasi nella lingua B, etc.)

²⁷ <http://www.ciilcorpora.net/ldcil.htm>

Riguardo a questo mercato e a questa applicazione, l'Europa differisce dagli Stati Uniti per due aspetti. Per prima cosa, negli USA i sistemi chiamati IVR²⁸ hanno avuto un'ampia diffusione, mentre questo non è successo in Europa. Il pubblico americano ha utilizzato spesso noiosi sistemi IVR per un periodo abbastanza lungo, ed è rimasto generalmente molto contento di passare dai toni DTMF (Dual Tone Multi-Frequency) e dai menu rigidi ai dialoghi parlati²⁹. In Europa, dove in alcuni Paesi i servizi di *call center* sono stati disponibili gratuitamente e hanno fornito un servizio di livello piuttosto elevato, l'introduzione dei sistemi automatici è stato spesso percepito dagli utenti come un passo indietro. Inoltre, gli Stati Uniti rappresentano un mercato vasto e abbastanza omogeneo con una sola lingua, molto più ampio in dimensione rispetto a qualsiasi mercato monolingue europeo. Di conseguenza, c'è una migliore economia di scala negli USA, visto che lo sviluppo di un'applicazione di dialogo parlato richiede uno sforzo significativo. Per queste ragioni, non c'è sorprendersi nello scoprire che il riconoscimento del parlato è stato adottato in modo più incisivo dagli USA rispetto all'UE.

Le condizioni economiche per la traduzione automatica sono esattamente

te all'opposto: difficili per gli USA, ma molto favorevoli per l'Europa sia per la traduzione del linguaggio parlato sia per la traduzione automatica. Questo è dovuto al fatto che ci sono molte lingue parlate in Europa, mentre qualsiasi sistema utilizzato per il vasto mercato americano coprirebbe solamente una piccola frazione di tutte le attività. In Europa, fare affari significa essere multilingue.

Condizioni economiche favorevoli per le società europee come fornitori di servizi di MT.

Ci sono anche importanti forze economiche che sostengono l'industria della traduzione in Asia, ma a causa del ruolo dominante dell'inglese, il mercato e le prospettive per un'industria delle tecnologie del linguaggio sono minori di quelli europei. Data la situazione attuale e la situazione delle società, si può supporre che, in normali circostanze, l'Europa assumerà un ruolo guida.

3.6. Una strategia per l'Europa

Riassumendo, il ruolo strategico delle HLT e la situazione nelle tre grandi regioni possono essere sintetizzati come segue:

Europa: La tecnologia del linguag-

²⁸ IVR sta per *interactive voice response*. I sistemi classici IVR rispondono alla pressione dei tasti con messaggi preregistrati.

²⁹ Un report citato frequentemente – Nuance Communications – Market Research: “Nuance Speech User Scorecard”, May 2000 – afferma che per il riconoscimento del parlato la soddisfazione generale era alta (circa l'87% dei rispondenti) e significativamente più alta rispetto a quella evidenziata per i sistemi DTMF. Questo risultato è ragionevole e coerente rispetto all'esperienza dell'autore, ma è necessaria qualche cautela, poiché il report originale non è più disponibile sul sito della Nuance o in internet cercando tramite Google.

gio è una necessità economica, politica e culturale. Superando la barriera del linguaggio saranno incentivate sia l'economia che la comunicazione. Le HLT sono già il focus di un considerevole sforzo di ricerca europea, ma l'importanza strategica di questa tecnologia per l'Europa giustifica una priorità molto più alta nell'agenda della ricerca.

USA: L'uso delle HLT è dominato da considerazioni militari e dalla lotta al terrorismo. Fondi molto significativi sono stati attualmente allocati alla ricerca e alla tecnologia relativa alle HLT.

Asia: La traduzione da e verso l'in-

glese ha attualmente la priorità. La mancanza di un'identità politica e di un'infrastruttura comune rende difficile l'assunzione di un ruolo guida da parte dell'Asia. C'è una notevole domanda di traduzione in India.

L'Europa, relativamente allo sviluppo e allo sfruttamento commerciale, si trova allo stesso tempo in una situazione di urgente bisogno e in una posizione privilegiata. Data la natura generica della tecnologia, essa offre l'opzione di essere commercializzata anche in altre regioni del mondo. Dall'altro lato, non ci si può aspettare che i nostri bisogni siano soddisfatti da altri.

Intervista con Joseph Mariani, Direttore del Dipartimento ICT Ministero Francese della Ricerca

Joseph Mariani si occupa di ricerca nei settori delle tecnologie del linguaggio, della comunicazione multimodale uomo-macchina, del riconoscimento del parlato, delle risorse linguistiche e della valutazione.

Egli è stato presidente dell'European Language Resources Association (ELRA), presidente dell'European (ora International) Speech Communication Association (ISCA), membro del consiglio dell'European Network on Language & Speech (ELSNET), e il coordinatore della rete francofona FRANCIL network.

Joseph Mariani è stato direttore del LIMSI e il capo del dipartimento Human-Machine Communication (1989- 2001), membro del Consiglio Scientifico del CNRS, chair dell'Information Science and Technology Advisory Committee del CNRS e membro dell'Evaluation Committee di INRIA.

Dal 2001 è direttore del dipartimento di ICT al Ministero Francese della Ricerca, dove è responsabile per i programmi di ricerca nelle telecomunicazioni, nelle tecnologie software, nelle nanotecnologie e nei multimedia, incluse le tecnologie del linguaggio.



Joseph Mariani
Direttore del Dipartimento ICT
Ministero Francese della Ricerca
Ricercatore senior LIMSI-CNRS
Parigi, Francia

Rispetto all'Europa, agli Stati Uniti e all'Asia: dove vede le somiglianze, e dove le differenze?

Gli Stati Uniti con il loro vasto mercato domestico vedono il multilinguismo specialmente da un punto di vista militare e della sicurezza: tutti comprendono l'inglese ma fanno fatica a capire le lingue straniere! L'Asia, come l'Europa, utilizza molte lingue, c'è quindi un bisogno di natura commerciale per il multilinguismo, anche se non è un mercato comune come l'UE. Per l'Europa, il multilinguismo è di estrema importanza con più di 20 lingue parlate nei 25 Stati membri dell'UE. Oltre alla dimensione economica, che fa del multilinguismo una pura necessità, ci sono anche dimensioni politiche, culturali e sociali. Questo rende la nostra situazione molto speciale.

Quale è la sua propensione verso la cooperazione con gli Stati Uniti o l'Asia?

La cooperazione internazionale ed una sana competizione fanno bene al progresso tecnologico e scientifico, e io apprezzo e supporto quando i nostri Paesi uniscono le loro forze nella scienza e nelle infrastrutture per affrontare le sfide nelle HLT. La cooperazione internazionale, però, non può sostituire la nostra agenda e rispondere ai nostri specifici bisogni. Relativamente al multilinguismo europeo, l'UE deve assumere un ruolo guida poiché, nessun altro può o vuole farlo al suo posto.

Quali sviluppi sono necessari per migliorare la posizione dell'Europa?

Nonostante i considerevoli sforzi fatti nel passato, credo che il livello di finanziamento della ricerca e della tecnologia non sia ancora compatibile con la dimensione della sfida. Nonostante il Commissario per l'istruzione, la cultura e il multilinguismo, Ján Figel, abbia recentemente sottolineato l'importanza del multilinguismo per l'Europa, le tecnologie del linguaggio appaiono come una piccola parte nel programma preliminare del VII Programma Quadro, citate solo alla fine del componente tecnologico "Simulation, visualisation, interaction and mixed reality". Ed i temi selezionati attualmente per le azioni possibili nell'ambito dell'Articolo 169 sono "Ricerca nel Mare Baltico", "Assistenza agli anziani" o "Metrologia", che sono temi rispettabilissimi, ma, secondo la mia opinione, meno strategici per l'Europa rispetto al problema delle lingue. Le tecnologie del linguaggio sono probabilmente l'argomento che si adatta meglio all'idea di un'azione europea coordinata, dato che lo sforzo per coprire le diverse tecnologie e le varie lingue europee è troppo oneroso per essere portato avanti solamente dalla Commissione, e può facilmente essere condiviso con gli Stati membri dell'UE, con le loro lingue e i loro programmi. Mentre gli Stati membri dovrebbero in primo luogo occuparsi delle specificità della loro lingua, come le risorse linguistiche (corpora di testo e parlato, dizionari), gli adattamenti della tecnologia specifici per quella lingua, l'EC potrebbe in primo luogo affrontarne gli aspetti generali indipendentemente da una lingua specifica: il coordinamento generale, lo sviluppo della tecnologia di base, la valutazione, gli standards.

L'Europa ha ancora bisogno di mettere a punto un'infrastruttura per valutare le tecnologie del linguaggio comparabile con quella esistente negli Stati Uniti con NIST, e le agenzie di finanziamento dovrebbero prendere in considerazione la valutazione delle prestazioni delle tecnologie di base, quando selezionano i progetti, al fine di evitare di finanziare lo sviluppo di applicazioni che si basano su una tecnologia non adeguata

Questo completerebbe l'European Language Resources Association, ELRA, che ha appena celebrato il suo decimo anniversario.

In effetti sì. La situazione sugli aspetti delle risorse del linguaggio e sulla valutazione delle tecnologie del linguaggio è migliorata di molto grazie all'esistenza di entità permanenti come il NIST e LDC negli Stati Uniti, o ELRA in Europa, ma la sfida per l'Europa richiede molto di più. Io sostengo la creazione di un'Agenzia per le Tecnologie del Linguaggio a livello di EC, una struttura permanente che coordina gli sforzi dell'EC e degli Stati membri verso le tecnologie del linguaggio necessarie per un'Europa multilingue, e pone l'Europa alla frontiera delle HLT nel mondo, avvantaggiandosi dalla sua sfida linguistica.

4. La situazione attuale

4.1. Il lavoro di traduzione oggi

Traduzione e localizzazione sono forniture di servizi anche se supportati dalla tecnologia.

Il lavoro di traduzione è molto diversificato. Esso spazia dalla traduzione del diploma di studio di uno studente ospite, alla gestione dell'enorme sito web della società Fortune 500, da coppie di lingue simili come l'olandese e il tedesco a lingue totalmente diverse come l'italiano e il cinese, e può coprire linguaggi standard, prosa o linguaggi altamente specializzati.

In parte, la traduzione si svolge all'interno degli istituti o delle società che ne hanno bisogno, e in parte viene delegata all'esterno a fornitori di servizi. I primi 20 fornitori di servizi coprono solamente il 16,3% del mercato³⁰, di conseguenza esso risulta decisamente molto frammentato. Contando le società che hanno almeno cinque impiegati, nel mondo ci sono 5.000 offerte di servizi di traduzione. Molti traduttori lavorano come liberi professionisti, e perfino i più grandi fornitori di servizi di traduzione e i più grandi utenti dei servizi di traduzione utilizzano liberi professionisti per il proprio lavoro di traduzione. Se si prende come esempio il DGT (Directorate-General for Translation of the European Commission),

la proporzione del lavoro fatto dai liberi professionisti è aumentata dall'11,8% nel 1992 al 23,0% nel 2004.

4.2. Tecnologie utilizzate nella traduzione professionale

Le principali tecnologie utilizzate nei servizi di traduzione oggi sono le *translation memory* e le risorse terminologiche.

La sofisticatezza tecnica del lavoro di traduttore dipende molto sia dalla società per cui lavora che dalla natura del materiale che deve essere tradotto. La situazione è molto differente se le traduzioni variano e i testi sono corti, o se si vogliono localizzare grandi pacchetti software in nuove lingue, a partire da una versione già esistente. In questo caso è importante considerare la precedente traduzione. Per questa ragione, la *translation memory* è ampiamente usata in determinati settori dell'industria, e ci sono metodi di controllo automatico che assicurano che le parole siano tradotte in modo coerente. L'utilizzo di terminologia coerente e di alta qualità può essere supportato da risorse *terminologiche*.

La traduzione automatica non è ancora utilizzata su vasta scala, ma è impiegata in alcune applicazioni. La traduzione automatica entra in gioco

³⁰ Fonte: Common Sense Advisory [CSA].

quando sono necessarie traduzioni grezze da eseguire molto rapidamente. Nell'Unione Europea, ad esempio, traduzioni grezze sono utilizzate per facilitare le comunicazioni interne. Quando però il materiale è presentato esternamente, la traduzione o è fatta completamente a mano o il risultato della traduzione automatica è rivisto da un traduttore.

Le Tecnologie come le TM e le risorse terminologiche non solo riducono i costi ma migliorano anche la qualità e la coerenza della traduzione.

È importante capire che le *translation memory* e le risorse *terminologiche* non solo aumentano l'efficacia di

un traduttore ma anche la qualità e la coerenza del lavoro. Si è visto che la traduzione automatica è utilizzata per fornire una traduzione grezza da usare come punto di partenza per la traduzione umana, allo scopo di migliorarne l'efficienza.

Il vantaggio principale della traduzione automatica risiede però nella traduzione rapida di testi che devono essere compresi con urgenza per motivi di lavoro ed i cui costi e tempi di traduzione umana sono improponibili. La quantità di questi *testi online dell'ultimo minuto* che devono essere tradotti con breve preavviso e di dimensioni piuttosto limitate (per esempio il *newswire*) sta crescendo rapidamente.

Intervista con Kevin Bolen, Chief Marketing Officer, Lionbridge

Kevin Bolen è responsabile del marketing globale per la Lionbridge e definisce le direzioni del mercato, le soluzioni di management, il posizionamento competitivo, e la promozione del marchio. Mr. Bolen si è unito alla Lionbridge nel 2005 a seguito dell'acquisizione di Bowne Global Solutions di cui era stato VP del Marketing dal 2002. Oltre a guidare il marketing globale, Bolen è stato responsabile delle vendite per la costa est. Prima di entrare in Bowne, ha gestito il marketing alla LexiQuest, Inc. un'azienda di sviluppo software specializzata nella tecnologia del linguaggio naturale. Prima di far parte della LexiQuest, ha trascorso sei anni con IBM Global Services dove ha assunto molti ruoli: dalla consulenza, alla crescita di aziende, al marketing. Nella sua ultima posizione ha svolto il ruolo di Senior Marketing Manager responsabile per lo sviluppo di soluzioni mondiali e di pianificazione strategica nei settori della vendita al dettaglio, del packaging, del trasporto, e della produzione.

Kevin Bolen ha un Bachelor in Business Administration in International Business della Pace University e un Master in Business Administration in Marketing and Management dalla Stern School of Business della New York University.



Kevin Bolen
Chief Marketing Officer
Lionbridge
Waltham, MA, USA

Potrebbe descrivere in poche parole il mercato della localizzazione?

Il mercato della localizzazione è altamente frammentato con migliaia di piccoli competitori che lavorano a livello regionale, verticale o funzionale. Alcuni offrono l'intera gamma di supporto per la localizzazione, mentre altri offrono il semplice servizio di traduzione, spesso come sotto-contraenza di fornitori più grandi. Alcune delle più grandi ditte offrono anche servizi complementari come authoring, design creativo, e interpretazioni o software di supporto per gestire il processo di localizzazione e i patrimoni.

Quali sono i principali fattori che guidano il mercato, anche relativamente alla sua crescita?

Le società continuano ad espandersi globalmente per raggiungere nuovi consumatori o per ridurre i loro costi operativi spostando il lavoro in regioni a basso costo. Entrare in questi nuovi mercati richiede la localizzazione del prodotto, il marketing, il packaging, la documentazione di supporto conforme alla legislazione locale, i materiali di addestramento per le vendite e per i clienti, e le comunicazioni ai dipendenti. I mercati di consumo come quello della Cina o dell'India stanno spingendo la localizzazione per un ampio numero di prodotti e di servizi in lingue non tradizionalmente servite da imprese consolidate, quindi alimentando

nuove opportunità per questi fornitori di servizi. Dall'altra parte, i cicli dei prodotti stanno accelerando e questo significa che più aumentano le specifiche dei prodotti, maggiori informazioni saranno distribuite con frequenza maggiore. Questo genera nuove opportunità di localizzazione ma richiede che i fornitori di servizi rivalutino i loro modelli di produzione come ad esempio il tempo totale di consegna del materiale tradotto ed il trasferimento di dati, fattori questi cruciali nella competizione.

Quali sono le tecnologie più importanti utilizzate nella traduzione delle lingue?

Le *Translation Memory* (TM) rimane la tecnologia leader in termini di efficienza produttiva. Comunque, i modelli tradizionali di *licensing* e le applicazioni a livello di *desktop* stanno diventando obsolete nell'era dei modelli *software on-demand, web-enabled*. Le TM attraverso il web permettono a più traduttori di lavorare simultaneamente su un singolo progetto, sfruttando il lavoro fatto da ognuno in ogni istante per migliorare la prestazione globale.

Anche i dizionari terminologici rappresentano un elemento chiave, dal momento che rendono possibile una maggiore coerenza all'interno di un gran numero di contenuti e di traduttori. L'abilità di includere il controllo della lingua nel processo di produzione assicura accuratezza e riduzione del tempo di produzione, un punto cruciale in un mondo *on-demand*.

Qual è la sua esperienza con la traduzione automatica?

Lionbridge, attraverso la acquisizione della Bowne Global Solutions, possiede uno dei motori di traduzione automatica basata su regole più usati dall'industria, "Barcelona". Sebbene i sistemi statistici e basati sugli esempi si siano dimostrati promettenti nella ricerca, l'insieme di lingue disponibili e di domini rimane troppo limitato per essere di uso pratico commerciale nel mercato odierno. Lionbridge ha utilizzato con successo il suo motore Barcelona su una varietà di progetti riducendo tempi e costi associati all'elemento traduzione di un intero progetto di localizzazione. Questi progetti devono essere attentamente esaminati dal momento che il sistema richiede di personalizzare i dizionari e le regole prima di ottenere una qualità sufficiente per permettere che il lavoro sia rivisto da un post-editor piuttosto che da un traduttore. C'è bisogno di un volume sufficiente di parole per assicurare un ROI positivo sulla spesa iniziale di personalizzazione.

La localizzazione di alta-qualità e la fornitura di traduzione automatica a basso costo mirano a diversi mercati. Intende anche coprire le "tecnologie di rottura"?

Lionbridge offre già un sito web gratuito per la traduzione basata su MT e ha abilitato i propri clienti ad utilizzare MT per tradurre contenuti che a causa dei costi associati con HT, non sarebbero stati tradotti. In futuro, via via che MT diventerà un'alternativa più accessibile, i clienti avranno una varietà di livelli di soluzione tra cui scegliere, visto che la qualità, comunque, è molto soggettiva. Siamo convin-

ti che i clienti stiano cercando una soluzione più integrata che faccia leva su web-based TMs, MT, e HT/post-editing per distribuire HT di qualità più velocemente e a costi inferiori rispetto all'attuale unico modello TM/HT. Questo permetterebbe un'espansione in più mercati, in tempi più brevi e con un migliore ritorno dall'investimento.

4.3. La ricerca sulle tecnologie alla base della traduzione da parlato a parlato

Negli ultimi decenni il linguaggio, sia nella forma scritta che parlata, è stato oggetto di ricerca scientifica da parte di migliaia di ricercatori in tutto il mondo. Consideriamo le diverse tecnologie coinvolte.

Il riconoscimento automatico della voce, cioè la creazione di una sequenza scritta di parole a partire da un flusso di parlato si è dimostrato un problema scientifico molto difficile, certamente più arduo che giocare a scacchi al livello delle performance umane. La comunità della ricerca ha avuto successo in primo luogo nel fronteggiare inizialmente problemi più circoscritti e poi via via più difficili. La tecnologia inizialmente trattava vocabolari ristretti, sistemi dipendenti dal parlatore in ambienti tranquilli e parole isolate, ma ora è capace di gestire vocabolari molto grandi, e di riconoscere il parlato continuo e indipendentemente dal parlatore. Mentre il pubblico pensa che questa tecnologia sia ancora ad un livello basso, in realtà ora il compito di dettare del testo in contesti professionali è spesso supportato da riconoscitori automatici del parlato, anche se rimangono ancora errori di riconoscimento. La comunità scientifica si è concentrata sulla sfida successiva, cioè il parlato proveniente da persone che non cooperano. Il parlatore non parla cioè con l'intenzione di essere riconosciuto da un sistema auto-

matico. Tipici scenari di ricerca riguardano il riconoscimento di conversazioni tra persone che parlano in maniera naturale e non vincolata. Il riconoscimento del parlato è ora spesso una delle componenti di un'interfaccia multimodale³¹.

Text-to-speech (TTS), chiamata anche sintesi del parlato (**speech synthesis**), genera il parlato dal testo scritto. Mentre i sistemi di TTS sono intelligibili già da diversi anni, l'ultimo significativo miglioramento, che sfrutta la concatenazione di unità a lunghezza variabile, ha portato a sistemi che producono una voce sintetica molto più naturale e fluida rispetto a pochi anni fa. I sistemi di sintesi giocano un ruolo importante nella comunicazione parlata uomo-macchina. Mentre frasi pre-registrate – o loro concatenazioni – possono essere utilizzate essenzialmente per contenuto di tipo statico, il TTS diventa una necessità nel caso di contesti variabili. Gli sforzi attuali nello sviluppo della sintesi del parlato si focalizzano su aspetti come la generazione veloce di voci differenti e la possibilità di arricchire la voce sintetica con emozioni e intonazioni contrastanti.

La traduzione automatica traduce un testo scritto da una lingua di partenza ad una lingua obiettivo. Negli ultimi decenni si sono investiti molti sforzi in questo difficile problema. Come per il riconoscimento del parlato, dove attorno al 1990 è cambiato il paradigma metodologico e i metodi basati sulle regole sono stati rimpiazzati da un

³¹ Ccome nei progetti europei CHIL (Computers in the Human Interaction Loop, <http://chil.server.de/>) e AMI (Augmented Multi-party Interaction, <http://www.amiproject.org/>).

approccio statistico, anche per la traduzione automatica ci sono due approcci fondamentalmente diversi. Il primo approccio, storicamente, si basa essenzialmente su metodi linguistici e sulla codifica esplicita della conoscenza umana, mentre il più recente è molto più guidato dai dati e sfrutta varianti di metodi statistici che si sono dimostrati efficaci nel campo del riconoscimento del parlato. Per ognuno di questi approcci, i sistemi allo stato dell'arte producono traduzioni di testo – per domini non circoscritti – che rimane di bassa qualità se confrontato con le traduzioni umane e può essere utilizzato solamente per traduzioni grezze che forniscono il significato basilare, frammisto ad

errori. Mentre i metodi linguistici tendono a tradurre bene alcune frasi ma falliscono miseramente su altre, i metodi statistici diffondono gli errori più uniformemente, tipicamente traducendo qualsiasi cosa senza completi fallimenti ma con pochi casi di traduzione esemplare. Detto questo, si deve notare che il progresso in questo settore è veloce, e che competizioni internazionali e misure obiettive su insiemi di dati di test mostrano notevoli progressi, in modo particolare nei sistemi che utilizzano approcci basati sui dati³².

Traduzione automatica del parlato, o Spoken language translation (SLT) traduce una lingua dalla forma

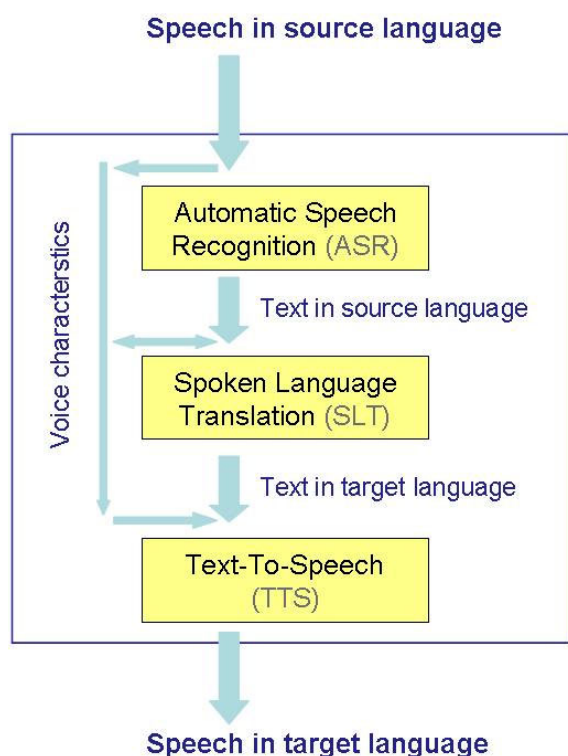


Fig. 5: Diagramma funzionale di un sistema di traduzione da parlato a parlato (SST). Il parlato della lingua di partenza è convertito in testo da un modulo di riconoscimento automatico (ASR). Questo testo viene poi passato ad un componente di traduzione automatica (SLT) che è specifico perché il testo contiene errori di riconoscimento e sfrutta le caratteristiche del parlato rispetto allo scritto – frasi che non seguono la grammatica, esitazioni, false partenze. Il suo output, un testo nella lingua di arrivo, costituisce l'input per il modulo di sintesi (TTS) che genera il parlato nella lingua di arrivo. Al fine di preservare le caratteristiche della voce originale, informazioni aggiuntive sono trasferite al modulo di sintesi (TTS).

³² Cf. [Ney].

parlata alla forma scritta in un'altra lingua. Come si può immaginare, i sistemi SLT essenzialmente combinano un riconoscitore vocale e un traduttore automatico, ma c'è di più: il testo contiene degli errori di riconoscimento che devono essere presi in considerazione. Un'altra difficoltà è che gli scenari in cui lavorano i ricercatori considerano il parlato che non è direttamente destinato a essere riconosciuto da un sistema, cioè il parlato che si sente in televisione, durante un meeting o in una conversazione da persona a persona.

Speech-to-speech translation (SST) combina tutti i metodi menzionati sopra.

Qui, il parlato nella lingua di partenza viene tradotto in parlato nella lingua di arrivo, in modo che le persone possano parlare tra di loro in diverse lingue, usando il computer come interprete. Per raggiungere questo scopo, un sistema di SLT è messo in cascata con un sistema TTS, e si adottano alcune misure aggiuntive. In primo luogo l'identità del parlatore dovrebbe essere preservata nel parlato nella lingua di arrivo, assieme alle caratteristiche prosodiche che appaiono nella lingua di partenza. Queste caratteristiche non sono naturalmente visibili in alcuna forma scritta intermedia. Questo problema molto difficoltoso è oggetto del progetto europeo TC-STAR, che lavora sulla traduzione da parlato a parlato dei dibattiti del parlamento europeo. Anche

se il compito è ambizioso, è interessante notare che la qualità delle traduzioni è già abbastanza buona per comprendere il senso dei discorsi parlamentari in una lingua non familiare.

4.4. Il Progetto TC-STAR

La traduzione da parlato a parlato è un compito di ricerca ambizioso. Per semplificare il compito e renderlo più fattibile, i primi progetti di ricerca che hanno affrontato l'argomento negli anni '90 (già con il parlato spontaneo) hanno lavorato su domini limitati come la pianificazione di appuntamenti³³.

TC-STAR è il primo progetto di ricerca congiunta che affronta la traduzione da parlato a parlato in un dominio illimitato.

È difficile descrivere la differenza in termini di complessità necessaria per passare da un dominio limitato ad un dominio illimitato, ma se vi è capitato di utilizzare una guida turistica tascabile per ordinare una colazione in Portogallo e poi avete provato ad ascoltare un dibattito parlamentare relativo ad un argomento qualunque, potreste esservi fatti un'idea. TC-STAR³⁴ (*Technology and Corpora for Speech to Speech Translation*) è il primo progetto di ricerca congiunta che affronta la traduzione *da parlato a parlato* in un dominio *illimitato*. A parte la traduzione di trasmissioni tele-

³³ Il progetto tedesco Verbmobil, 1993-2000, con un volume totale di finanziamento pari a 53M€ ha intrapreso un grande sforzo in questa direzione. L'iniziativa internazionale C-STAR, un grande consorzio dedicato alla traduzione da parlato a parlato con venti partners, e' iniziato nel 1991 ed e' tuttora in corso.

³⁴ TC-STAR (<http://www.tc-star.org>) è un progetto integrato (IP), i.e. un grande progetto, all'interno del VI programma quadro.

visive dal cinese all'inglese, che è stata scelta come riferimento a indicatori di *misura* internazionali, TC-STAR lavora sui discorsi del parlamento europeo in inglese e in spagnolo, sia di madrelingua che non. I temi includono qualsiasi argomento di cui si occupa il Parlamento europeo.

Questo sforzo si focalizza sulla ricerca avanzata in tutte le tecnologie chiave per la traduzione da parlato a parlato (SST): il riconoscimento del parlato, la traduzione automatica e la sintesi del parlato. Gli obiettivi del progetto sono molto ambiziosi: raggiungere un punto di svolta nella ricerca sulla traduzione da parlato a parlato per ridurre in modo significativo la differenza in prestazioni tra persone e macchine. La prima metà dei tre anni del progetto ha mostrato risultati molto soddisfacenti in questo difficile compito, ma c'è ancora molta strada da fare.

Oltre al lavoro scientifico, TC-STAR ha creato l'infrastruttura necessaria per accelerare la velocità del progresso nel campo. Ha raccolto i dati necessari per la metodologia statistica perseguita dal progetto, e ha messo a punto un'infrastruttura di valutazione, basata su valutazione competitiva. Questo approccio assicura sia che i progressi di ricerca siano riconosciuti come tali, sia che i metodi sviluppati dai vari partners di progetto possano essere confrontati e validati in modo appropriato. Combinando opportunamente cooperazione e una salutare competizione, questo approccio dovrebbe massimizzare il progresso scientifico. Allo stesso tempo, il consorzio si misura all'interno della comunità scientifica esterna, e i partners di progetto hanno raggiunto le migliori posizioni nelle recenti valutazioni internazionali IWSLT³⁵.

³⁵ International Workshop on Spoken Language Translation. IWSLT 2005 in Pittsburgh, USA, <http://www.is.cs.cmu.edu/iwslt2005/> . IWSLT 2004 in Kyoto, Japan, <http://www.slt.atr.jp/IWSLT2004/> .

5. Il potere di una tecnologia abilitante

5.1. Bisogni umani insaziabili

La traduzione automatica ha appena cominciato ad essere utilizzata in pratica, la traduzione da parlato a parlato è ancora ricerca di frontiera così come molte tecnologie legate a quest'ultima. Dall'altra parte c'è un'enorme domanda latente che deve essere soddisfatta, e sappiamo che la tecnologia un giorno potrà soddisfarla. Possiamo prevedere il futuro anche se in modo vago?

Per non perdersi in dettagliate analisi di mercato che dovrebbero poi essere riviste continuamente, cerchiamo di identificare le forze rilevanti che determinano la domanda e l'offerta di un futuro mercato. Ci sono alcune tendenze che sono ampiamente accettate sia da molti specialisti che da non specialisti, come quella verso la mobilità, il *networking* ed il calcolo distribuito. Se comprendiamo le variabili chiave, tecnologiche e commerciali, le catene del valore (o, più precisamente, le reti del valore) e le dinamiche del mercato, possiamo essere quasi certi che la nostra visione sul futuro non è poi così poco accurata. Esiste uno stratagemma che rende la previsione ancora più attendibile: *Il modo migliore per predire il futuro è inventarlo*³⁶.

Se si guarda alle vere storie di successo delle nuove tecnologie entrate nel

mercato, si troveranno due ingredienti che appaiono regolarmente: una nuova tecnologia e un vecchio bisogno umano. La telefonia mobile è un esempio tipico, che soddisfa l'insaziabile bisogno umano di comunicare con altre persone e il desiderio di avere anche un impatto nonostante la distanza. Fin da piccoli si utilizzava la comunicazione quasi in tempo reale, ma non isocrona, con brevi messaggi: un tempo si scrivevano piccole note sulla carta a scuola – i bambini ora usano gli SMS. Sarebbe forse più preciso confrontare gli SMS di oggi con i telegrammi del passato? Ci sono delle somiglianze: in entrambi i casi, un breve testo è consegnato attraverso un canale elettronico a dei destinatari. Comunque, a causa di un'enorme differenza di prezzo e a scenari di utilizzo molto diversi, non ha senso considerare l'SMS come una forma moderna di telegramma. Ogni tentativo di prevedere l'ammontare di traffico di SMS di oggi basato sul numero di telegrammi spediti venti anni fa fallirebbe fatalmente. In modo simile, sarebbe stato ingannevole e altamente impreciso stimare l'attuale uso pubblico di Google basandosi sulle cifre del "desktop paper research" o cercando nei database di venti anni fa. Comunque, perfino nel passato era chiaro che sarebbe stata adottata qualsiasi tecnologia che incontrava i bisogni

³⁶ Citazione attribuita ad Alan Kay, pioniere del PC e specialista dell'interfaccia utente.

umani basilari ed essenzialmente insaziabili: le persone hanno bisogno di comunicare con altre persone (telefono, cellulare, e-mail, SMS, chat), le persone hanno bisogno di dominare gli spazi (automobile, aeroplano, e anche la comunicazione), le persone sono assetate di informazioni e di divertimento (internet, televisione).

**I bisogni umani insaziabili:
la Comunicazione con le persone,
il dominio dello spazio e del tempo,
la sete di informazione e di divertimento.**

Il mercato esistente per la traduzione e per la localizzazione ci dice molto poco riguardo a che cosa succederebbe se la traduzione fosse disponibile in tempo reale e ad un costo molto ragionevole. In molte situazioni, pagare un euro per ogni linea di testo tradotto è semplicemente irragionevole se confrontato con il valore aggiunto ottenuto. Far scendere il costo di uno o due ordini di grandezza consentirebbe un enorme incremento nell'utilizzo della traduzione, perfino se la qualità non fosse perfetta. C'è un forte desiderio umano di comunicare e di ottenere informazioni. È vero che la traduzione è già possibile con traduttori professionisti, ma proprio come vent'anni fa non avremmo potuto immaginare l'incredibile successo degli SMS basandoci sull'uso dei telegrammi, è ora difficile immaginare come sarà la vita dopo l'eliminazione universale delle barriere linguistiche.

I mercati stanno cambiando in modo costante. Il valore esce da alcune catene del valore per entrare in altre. Le vecchie domande incontrano nuove

possibilità, e nascono nuovi prodotti e nuovi mercati.

Cosa succederebbe se la traduzione in tempo reale fosse disponibile ad un costo ragionevole? Quali sarebbero gli effetti della disponibilità di testo e parlato, di documenti, di siti web, di video, di trasmissione di contenuti? Cosa succederebbe se non ci fossero più barriere legate alle lingue?

Non dovremmo basare le nostre stime sul modello di business della traduzione com'è oggi. La quantità attuale di traduzioni è appena la punta dell'iceberg, in confronto alla domanda latente di traduzione. Lasciate che vi illustri cosa *io* vorrei fare.

Vorrei guardare Al Jazeera – sottotitolata nella mia lingua – per avere un'idea di cosa pensa il mondo arabo. Parlando inglese con i miei pari italiani e spagnoli potremmo talvolta aver bisogno di un aiuto per migliorare i nostri canali di comunicazione. È difficile per me leggere un documento francese, specialmente quando non conosco la terminologia – una traduzione approssimativa in tedesco o inglese renderebbe sostanzialmente più veloce la lettura. Durante la mia vacanza in Portogallo, vorrei capire i menu ed essere in grado di scambiare alcune parole con gli abitanti locali.

**Immaginate un mondo con la traduzione
in tempo reale disponibile
in modo economico ed immediato.**

Guardereste un sito web in cinese? Cucinereste quella meravigliosa ricetta proveniente dalla vostra vacanza in Spagna che, sfortunatamente, è scritta in spagnolo? Quali sono le vostre richieste?

Noi Europei comunicheremmo più

facilmente l'uno con l'altro e potremmo sentirci più vicini.

L'utilizzo delle tecnologie del linguaggio fornirà infine un significativo vantaggio economico. È in atto un processo di automatizzazione del contatto con i clienti al fine di gestire in modo multilingue le domande più frequenti (FAQ). La maggior parte dei beni non tangibili delle società sono nelle menti e nei file dei loro impiegati, in molte lingue. Avere facile accesso all'informazione indipendentemente dalla lingua è una pura necessità per ogni società che opera in modo globale. Questo sarà un mercato interessante, e le società come l'IBM e la SAP, Yahoo e Google si stanno preparando.

La traduzione in tempo reale di qualsiasi fonte di dati – *newswire*, contenuti web dinamici, video; la ricerca di documenti in lingue diverse; l'accesso semplificato ai mercati stranieri, in particolare per i professionisti e le SMEs, costituiscono opportunità illimitate per l'Europa.

L'avvento della traduzione automatica sarà sostenuto da tre forti fattori di mercato trainanti:

- i bassi costi
- le operazioni in tempo reale
- le proprietà dell'elaborazione automatica

Ciascuno di questi tre fattori di mercato dovrebbe essere sufficientemente forte per sostenere da solo un business.

Intervista con Dimitris Sabatakakis, CEO, SYSTRAN

SYSTRAN è leader di mercato per la fornitura di prodotti e soluzioni software per la traduzione nelle diverse modalità: "desktop", "enterprise" ed "Internet". Le soluzioni offerte facilitano la comunicazione in 40 direzioni di traduzione (20 coppie linguistiche) e 20 contesti. Con oltre tre decenni di esperienza di ricerca e sviluppo, il software della SYSTRAN è usato da società leader globali, di portali e di agenzie pubbliche. L'utilizzo dei prodotti e delle soluzioni SYSTRAN migliorano la comunicazione e incrementano la produttività dell'utente e fanno risparmiare tempo ai mercati B2E, B2B e B2C, dato che essi distribuiscono soluzioni in tempo reale per la ricerca, la gestione del contenuto, il supporto on line al cliente, le comunicazioni all'interno delle società e l'e-commerce.



Dimitris Sabatakakis
CEO of SYSTRAN
Paris, France

Dimitris Sabatakakis è nato nel 1962 ad Atene, Grecia. Laureato all'Università di Strasburgo in Scienze Economiche, ha cominciato la sua carriera nella finanza, per poi proseguire nell'industria. Unitosi agli investitori, ha rilevato e gestito il recupero della società Gachot, che è stata venduta al KEYSTONE/TYCO Group nel 1995. Dimitris Sabatakakis è CEO di SYSTRAN da Febbraio 1997.

Come si relaziona il mercato della traduzione automatica con quello della traduzione tradizionale?

Il mercato tradizionale della traduzione è stabile e rappresenta un processo umano come lo scrivere. Il mercato della MT (traduzione automatica) è molto differente. L'offerta comprende servizi e applicazioni di traduzione in tempo reale e basati sul web. Risulta pratico adottare la MT per pubblicazioni multilingue in situazioni in cui si possono combinare sia la richiesta di tradurre grandi volumi di contenuti sia il controllo dell'utente sul testo originale. Il 99% del mercato attuale delle applicazioni di traduzione automatica riguarda la ricerca di informazioni importanti nelle migliaia di pagine di portali come Yahoo! e Google che vengono tradotti giornalmente.

MT o i software di traduzione del linguaggio sono dei processi automatici che permettono all'utente di:

- Comprendere il contenuto espresso in lingue straniere nella propria lingua madre in tempo reale e senza costi;
- Pubblicare il contenuto in diverse lingue in tempo reale, naturalmente "controllando" il testo di origine.

Le persone comprendono le capacità dei sistemi automatici? Tendono a sopravvalutarli o a sottovalutarli?

Il valore della MT utilizzata per capire il significato generale di un testo è ovvio. La dimostrazione di questo sono le migliaia di pagine che vengono tradotte ogni giorno.

Anche il valore dell'utilizzo della MT per la pubblicazione in molte lingue è ovvia ma l'investimento significativo richiesto per "controllare" o per strutturare il testo d'origine è spesso sottovalutato o trascurato. Per valutarlo correttamente, l'intero flusso di lavoro di pubblicazione deve considerare i problemi della produzione multilingue sin dal principio. In realtà, le organizzazioni producono contenuto (come schede tecniche di prodotti, materiale di marketing, informazioni di supporto tecnico, basi di conoscenza, etc.) in una lingua di partenza che di solito è l'inglese e considerano la localizzazione del contenuto solo come un passo aggiuntivo, in questo modo a causa di restrizioni di budget, l'ammontare di testo localizzato è molto minore rispetto al testo originale.

Infine, il mantenimento di versioni di contenuto localizzato è una sfida dal momento che è costoso, lento (è un processo umano) e non risolve del tutto i problemi di inconsistenza della terminologia.

Vede la MT più come uno strumento di razionalizzazione per la traduzione umana o più come una nuova modalità di traduzione?

MT è sia uno strumento per la traduzione umana sia una traduzione di tipo nuovo. I traduttori umani dovrebbero adottare la MT dal momento che aumenta la produttività dell'utente e riduce il tempo di traduzione. Questa soluzione non ha ancora sfondato perchè l'ergonomia e l'interfaccia utente non sono state create esplicitamente per i traduttori umani. SYSTRAN ha fatto grandi sforzi per fornire questi strumenti in modo che la MT possa essere utilizzata all'interno di servizi di traduzione tradizionali al fine di aumentarne la produttività.

6. Conclusioni

Strumenti veloci, affidabili ed economici per comunicare, trasportare ed elaborare dati rappresentano la spina dorsale di una società dell'informazione moderna. Quello che è vero per una normale tecnologia dell'informazione e della comunicazione lo è anche per i sistemi che traducono da una lingua all'altra, sia per la comunicazione che per l'accesso all'informazione. Per l'Unione Europea, con le sue venti lingue ufficiali e molte più lingue parlate, la disponibilità di traduzioni veloci, affidabili ed economiche è una necessità, e la tecnologia di traduzione dovrebbe essere considerata strategicamente importante.

Altre parti del mondo hanno bisogno di servizi automatici di traduzione, ma in contesti diversi. È evidente che Stati Uniti e Paesi asiatici non potranno risolvere i nostri problemi, per i quali noi stessi siamo ora in grado di trovare adeguate risposte.

In Europa, a causa del bisogno di traduzione automatica e di traduzione del linguaggio parlato, ci sono condizioni di mercato favorevoli per le società che pensano di offrire tecnologia e servizi di traduzione. Alcuni forti grup-

pi di ricerca europei possono svolgere il ruolo di fornitori di tecnologia di traduzione. Oltre al mercato europeo, la domanda di traduzione automatica sarà elevata anche in Giappone, in Cina e in Corea così come in India.

L'industria delle tecnologie del linguaggio sta crescendo rapidamente, ma il pieno potenziale della traduzione automatica andrà oltre un miglioramento dell'efficienza della traduzione umana attuale. Si aprirà ai segmenti di mercato che ora esistono appena, e che hanno margini di profitto talmente ridotti che non saranno mai attrattivi per i servizi di traduzione umana. Così come sarebbe stato impossibile estrapolare l'utilizzo di SMS dall'utilizzo di telegrammi, è difficile predire il volume di traduzioni del futuro. Sarà molto grande, e aiuterà ad ampliare quella parte del nostro potenziale economico che è ancora limitata dalle barriere linguistiche.

Tutte queste innovazioni richiedono tempo, ma stanno arrivando, e avranno un impatto enorme. Nel frattempo, cerchiamo di delineare il futuro in modo vantaggioso per noi. Sta a noi dargli una forma.

7. Bibliografia e altre informazioni supplementari

7.1. Bibliografia

- [Chr] C. Christensen: *The Innovator's Dilemma. When New Technologies Cause Great Firms to Fail*. Harvard Business School Press, Boston, Mass., 1977.
- [CRa] C. Christensen, M. Raynor: *The Innovator's Solution. Creating and Sustaining Successful Growth*. Harvard Business School Press, Boston, Mass., 2003.
- [Cry] D. Crystal: *English as a Global Language*. Cambridge University Press. 2003² (1997¹).
- [CSA] R. Beninatto, D. DePalma: *Ranking of Top 20 Translation Companies*. Common Sense Advisory, Inc., June 2005. Può essere scaricato dal sito <http://www.commonseadvisory.com> alla sottovoce http://www.commonseadvisory.com/members/res.cgi.php/050701_QT_top_20.php
- [DGT] The Directorate-General for Translation of the European Commission (DGT). http://europa.eu.int/comm/dgs/translation/index_en.htm
- [EB1] Europeans and languages. Eurobarometer 54 Special. INRA Report, 60 p., February 2001. Il documento può essere scaricato dal sito http://europa.eu.int/comm/education/policies/lang/languages/barolang_en.pdf e viene richiamato in http://europa.eu.int/comm/education/policies/lang/languages/index_en.html. I dati si riferiscono all'Europa dei 15.
- [EB5] Europeans and languages. Eurobarometer 63.4, September 2005. Una rilevazione che interessa i 25 Stati membri dell'EU, gli Stati che hanno presentato domanda di adesione (Bulgaria e Romania), gli Stati candidati (Croazia e Turchia) e nella comunità turca cipriota. Il documento può essere scaricato dal sito http://europa.eu.int/comm/public_opinion/archives/ebs/ebs_237.en.pdf e viene richiamato in <http://europa.eu.int/languages/en/document/80/20>
- [ELP] Languages of Europe. Nel sito ufficiale della Commissione Europea relativo alle lingue parlate nell'UE. http://europa.eu.int/comm/education/policies/lang/languages/index_en.html
- [Gra] D. Graddol: *The Future of English? A guide to forecasting the popularity of the English language in the 21st century*. (A report commissioned by The British Council.) The English Company (UK), 64 p., 2000 (1997¹), ISBN 0-86355-356-7.
- [Gri] B. Grimes: *Ethnologue Language Database*. <http://www.sil.org/ethnologue/> Ethnologue è un catalogo che comprende più di 6.700 lingue parlate in 228 Paesi.

- [Gor] R.G. Gordon (ed.): *Ethnologue: Languages of the World*. Dallas, Tex., 2005¹⁵, SIL International. Versione online: <http://www.ethnologue.com/>
- [Ney] H. Ney: *One Decade of Statistical Machine Translation: 1996-2005*. In: *Proceedings of the MT Summit X*, pp. i-12 - i-17, Phuket, Thailand, September 2005. Il documento può essere scaricato dal sito <http://www-i6.informatik.rwth-aachen.de/web/Publications/index.html>
- [TMC] *Translating for a multilingual community*. Directorate-General for Translation of the European Commission (DGT), April 2005, 19 p. Il documento pdf può essere scaricato dal sito http://europa.eu.int/comm/dgs/translation/bookshelf/brochure_en.pdf
- [TTW] *Translation tools and workflow*. Directorate-General for Translation of the European Commission (DGT), April 2005, 25 p. Il documento pdf può essere scaricato dal sito http://europa.eu.int/comm/dgs/translation/bookshelf/tools_and_workflow_en.pdf

7.2. Ulteriori letture

In italiano:

Portale: Le lingue e l'Europa <http://europa.eu/languages/it/home>

In inglese:

Common Sense Advisory <http://www.commonsenseadvisory.com/>

EAMT – European Association for Machine Translation <http://www.eamt.org/>

ELRA – European Language Resources Association <http://www.elra.info/>

ELSNET – European Network in Language and Speech <http://www.elsnet.org/>

GALA – The Globalization and Localization Association <http://www.gala-global.org/>

LISA – Localization Industry Standards Organization <http://www.lisa.org/>

Multilingual Computing, Inc. <http://www.multilingual.com/>

7.3. Tabelle

Tabella 1: Condizioni commerciali per il commercio internazionale all'interno dell'EU, per i beni e per l'informazione. Per i servizi informativi, la barriera della lingua rappresenta un ostacolo rilevante da superare.

Tabella 2: Le maggiori lingue del mondo in milioni di persone madrelingua secondo due differenti fonti, (A) The English Company's engco model [Gra] e (B) cifre comparative tratte da Ethnologue ([Gri]; vedi [Gra]).

Tabella 3: 'Influenza globale' delle 12 maggiori lingue secondo il modello engco (vedi tabella 2). Il punteggio 100 rappresenta la posizione dell'inglese nel 1995 [Gra].

Tabella 4: Discipline in cui gli accademici tedeschi dichiarano che l'inglese è la loro lingua di lavoro [Gra].

Tabella 5: Le 20 lingue ufficiali dell'Unione Europea e le loro abbreviazioni [ELP]. L'Irlandese diventerà la 21esima lingua ufficiale dell'UE dal primo gennaio 2007.

Tabella 6: Abilità di parlare lingue straniere nell'UE: Percentuali di cittadini europei che parlano le rispettive lingue nell'UE (come madrelingua /come seconda lingua o lingua straniera / un'altra lingua). L'indagine si è svolta nell'Unione Europea dei 25 Stati Membri e, in aggiunta, nei Paesi entranti (Bulgaria and Romania), nei Paesi candidati (Croazia e Turchia) e tra la comunità turca cipriota. – Fonte: [EB5]

Tabella 7: Posizionamento dei primi 20 fornitori di servizi linguistici – reddito del 2004. (Nelle cifre riportate si sono considerate le due importanti acquisizioni del 2005, l'acquisizione di Bowne Global Solutions da parte di Lionbridge e l'acquisizione di TRADOS da parte di SDL.)

Tabella 8: Esempi di innovazioni di rottura, insieme con le corrispondenti tecnologie consolidate. – Fonte: [Chr], [CRa].

7.4. Figure

Fig. 1: Percentuale dei libri mondiali annualmente pubblicati in ogni lingua. L'inglese è la lingua più utilizzata per la pubblicazione di libri: oltre 60 Paesi pubblicano i titoli in inglese [Gra].

Fig. 2: Lingue del mondo. Ogni punto rappresenta la localizzazione principale delle lingue viventi elencate in Ethnologue.

Fig. 3: Il cambiamento tecnologico indotto dalle sustaining e disruptive innovations. – Fonte: [CRa].

Fig. 4: Budget del DoD (Department of Defense) degli Stati Uniti per le tecnologie del linguaggio per la traduzione. Stime di budget per l'anno fiscale 2005 per la ricerca, lo sviluppo, il test e la valutazione (RDT&E), a DARPA (Defense Advanced Research Projects Agency). – Fonte: DARPA.

Fig. 5: Diagramma funzionale di un sistema di traduzione da parlato a parlato.

7.5. Abbreviazioni

AFNLP	Asian Federation of Natural Language Processing
ASR	Automatic Speech Recognition (riconoscimento automatico del parlato)
BLEU	(Misura statistica per valutare la qualità della traduzione correlata con il giudizio umano; più alto è il punteggio ottenuto migliore è la qualità della traduzione.)
CEC	Commission of the European Community (Commissione dell'Unione Europea)
DARPA	Defense Advanced Research Projects Agency
DGT	Directorate General for Translation
DTMF	Dual Tone Multiple Frequency
EU	European Union (UE)
GALE	Global Autonomous Language Exploitation
HLT	Human Language Technologies (tecnologie del linguaggio umano)
HT	Human Translation (traduzione umana)
IP	qui: integrated project (in altri contesti: Internet Protocol; Intellectual Property)
IR	Information Retrieval
IT	Information Technology
IVR	Interactive Voice Response
LISA	Localization Industry and Standards Association
MT	Machine Translation (traduzione automatica)
NLP	Natural Language Processing
ROI	Return On Investment
SMS	Short Message Service
TC-STAR	Technology and Corpora for Speech-to-Speech Translation
TM	Translation Memory
TTS	Text-To-Speech (dal testo al parlato)

Finito di stampare
nel mese di novembre 2006
dalla Tipolitografia TEMI - Trento