

LE NOSTRE LINGUE A RISCHIO: UNA SFIDA PER LE TECNOLOGIE DEL LINGUAGGIO

Siamo testimoni di una rivoluzione digitale che sta avendo un impatto radicale sulla comunicazione e sulla società. I recenti sviluppi nella tecnologia dell'informazione digitale e della comunicazione vengono talvolta paragonati all'invenzione della stampa da parte di Gutenberg. Ma cosa può dirci questa analogia sul futuro della società dell'informazione europea e, in particolare, delle nostre lingue?

La rivoluzione digitale è paragonabile
all'invenzione della stampa da parte di
Gutenberg.

In seguito all'invenzione di Gutenberg, furono compiuti grandi progressi nella comunicazione e nello scambio di conoscenza attraverso opere quali la traduzione della Bibbia in una lingua volgare da parte di Lutero. Nel corso dei secoli successivi, sono state sviluppate tecniche per gestire meglio l'elaborazione del linguaggio e lo scambio di conoscenza:

- la standardizzazione ortografica e grammaticale delle lingue principali ha permesso di disseminare nuove idee scientifiche e intellettuali in modo rapido;
- lo sviluppo delle lingue ufficiali ha reso possibile ai cittadini la comunicazione all'interno di determinati confini (spesso politici);

- l'insegnamento delle lingue e la traduzione ha reso possibili gli scambi tra persone che parlavano lingue diverse;
- la creazione di linee guida editoriali e bibliografiche ha assicurato la qualità e la disponibilità di materiale stampato;
- la creazione di diversi mezzi di comunicazione, come i giornali, la radio, la televisione e i libri, ha permesso di soddisfare bisogni di comunicazione di natura diversa.

Negli ultimi vent'anni, la tecnologia dell'informazione ha aiutato ad automatizzare e facilitare molti processi:

- i software per il *desktop publishing* hanno sostituito la dattilografia e la composizione tipografica;
- PowerPoint di Microsoft ha sostituito i lucidi;
- con la posta elettronica si spediscono e si ricevono documenti più velocemente che utilizzando un fax;
- Skype offre la possibilità di fare chiamate telefoniche su Internet in modo economico e permette di organizzare incontri virtuali;
- grazie a formati di codifica audio e video è possibile scambiarsi in maniera semplice contenuti multimediali;
- i motori di ricerca forniscono un accesso alle pagine web basato su parole chiave;

- servizi online come Google Translate producono veloci traduzioni approssimate;
- le piattaforme di social media come Facebook, Twitter, e Google+ facilitano la comunicazione, la collaborazione e la condivisione dell'informazione.

Sebbene queste applicazioni e questi strumenti siano utili, essi non sono ancora in grado di supportare pienamente una società europea multilingue in cui l'informazione e le merci possano circolare liberamente.

2.1 I CONFINI LINGUISTICI FRENANO LA SOCIETÀ EUROPEA DELL'INFORMAZIONE

Non siamo in grado di prevedere esattamente come sarà la società dell'informazione del futuro. Tuttavia, esiste un'elevata probabilità che la rivoluzione nelle tecnologie della comunicazione avvicinerà persone che parlano lingue diverse in nuovi modi. Questa tendenza induce gli individui a imparare nuove lingue e gli sviluppatori, in particolare, a creare nuove applicazioni tecnologiche per assicurare la comprensione reciproca e l'accesso alla conoscenza condivisa.

In uno spazio economico e di informazione globale, una maggiore quantità di lingue, di parlanti e di contenuti interagiscono più velocemente con nuovi tipi di mezzi di comunicazione. L'attuale popolarità dei social media (Wikipedia, Facebook, Twitter, YouTube e, recentemente, Google+) rappresenta soltanto la punta dell'iceberg.

L'economia e lo spazio d'informazione globali ci mettono di fronte a lingue, parlanti e contenuti diversi.

Oggi possiamo trasmettere gigabyte di testo in tutto il mondo in pochi secondi prima di accorgerci che si trat-

ta di una lingua che non comprendiamo. Secondo un recente rapporto della Commissione Europea, il 57% degli utenti di Internet in Europa acquista merci e servizi in lingue diverse dalla loro lingua nativa; l'inglese è la lingua straniera più comune, seguito dal francese, dal tedesco e dallo spagnolo. Il 55% degli utenti legge contenuti in una lingua straniera mentre il 35% usa un'altra lingua per scrivere e-mail o per spedire commenti sul Web [3]. Alcuni anni fa, l'inglese poteva essere considerato la lingua franca del Web – la grande maggioranza dei contenuti sul Web era in inglese – ma la situazione ora è cambiata sensibilmente. La quantità di contenuti online in altre lingue europee (così come per quelle asiatiche e medio-orientali) si è moltiplicata.

Sorprendentemente, questo onnipresente divario digitale dovuto ai confini linguistici non ha ricevuto molta attenzione pubblica; eppure, esso solleva una domanda molto pressante: quali lingue europee prospereranno nella società dell'informazione e della conoscenza in rete, e quali sono destinate a scomparire?

2.2 LE NOSTRE LINGUE A RISCHIO

Se da un lato l'invenzione della stampa contribuì certamente ad intensificare lo scambio di informazioni in Europa, essa al contempo portò anche all'estinzione di molte lingue europee. Le lingue regionali e minoritarie venivano stampate raramente e lingue come il cornico e il dalmatico vennero ridotte a forme di trasmissione orale, il che a sua volta restrinse gli ambiti d'uso di queste lingue. Internet avrà lo stesso impatto sulle nostre lingue?

L'ampia varietà di lingue esistenti in Europa rappresenta una delle sue ricchezze più importanti.

Le circa 80 lingue dell'Europa costituiscono uno dei più ricchi e più importanti patrimoni culturali dell'Europa, e una parte vitale del suo modello sociale unico [4]. Mentre lingue come l'inglese e lo spagnolo probabilmente sopravviveranno nel mercato digitale emergente, molte altre lingue Europee potrebbero diventare irrilevanti all'interno di una società in rete. Questo porterebbe ad un indebolimento dello stato globale dell'Europa e andrebbe contro l'obiettivo strategico di assicurare un'uguale partecipazione a tutti i cittadini europei indipendentemente dalla lingua.

Secondo un rapporto dell'UNESCO sul multilinguismo, le lingue rappresentano un mezzo essenziale per poter godere di diritti fondamentali come il diritto di espressione politica, il diritto all'educazione e alla partecipazione nella società [5].

2.3 LA TECNOLOGIA DEL LINGUAGGIO È UNA TECNOLOGIA FONDAMENTALE

In passato, gli sforzi di investimento nell'ambito della conservazione delle lingue si sono focalizzati sull'insegnamento delle lingue e sulla traduzione. Secondo una stima, il mercato europeo per la traduzione, l'interpretariato, la localizzazione di software e di siti web è stato di 8,4 miliardi di euro nel 2008 e per il futuro è attesa una crescita del 10% all'anno [6]. Eppure questa cifra copre solo una piccola parte dei bisogni attuali e futuri per quanto riguarda la comunicazione tra lingue diverse. La soluzione più convincente per assicurare in futuro ampiezza e profondità nell'uso delle lingue in Europa consiste nell'uso di una tecnologia appropriata, allo stesso modo in cui usiamo la tecnologia per risolvere le nostre esigenze di trasporto e di energia.

Le tecnologie linguistiche (rivolte a tutte le forme di testi scritti e discorsi orali) aiutano le persone a collaborare, a fare affari, a condividere la conoscenza e a parte-

cipare al dibattito sociale e politico a prescindere dalle barriere linguistiche e dall'abilità nell'uso del computer. Spesso operano in maniera invisibile all'interno di sistemi informatici complessi, per aiutarci a:

- trovare informazioni mediante un motore di ricerca su Internet;
- controllare errori di ortografia e di grammatica all'interno di un programma per l'elaborazione di testi;
- vedere, in un negozio online, le opinioni sui prodotti espresse da altri clienti;
- seguire, in automobile, le istruzioni vocali di un sistema di navigazione;
- tradurre pagine web attraverso un servizio in rete.

La tecnologia del linguaggio consiste in un certo numero di applicazioni di base che rendono possibili processi all'interno di un più ampio quadro applicativo. I Libri Bianchi di META-NET si prefiggono l'obiettivo di verificare che livello abbiano raggiunto queste tecnologie di base per ciascuna lingua europea.

L'Europa ha bisogno di tecnologie linguistiche robuste ed economicamente accessibili per tutte le lingue europee.

Al fine di mantenere la propria posizione in prima linea nell'innovazione globale l'Europa avrà bisogno, per tutte le lingue europee, di tecnologie linguistiche robuste, economicamente accessibili e saldamente integrate all'interno degli ambienti software principali. Senza le tecnologie del linguaggio, non saremo in grado di raggiungere in un prossimo futuro un'esperienza utente interattiva, multimediale e multilingue realmente efficace.

2.4 LE OPPORTUNITÀ PER LE TECNOLOGIE LINGUISTICHE

La rivoluzione tecnologica nel mondo della carta stampata fu la possibilità di duplicare rapidamente un'immagine di un testo usando una macchina da stampa sufficientemente potente. Il duro lavoro di ricerca, lettura, traduzione e sintesi della conoscenza era appannaggio degli uomini. Per registrare la lingua parlata si è dovuto aspettare fino ad Edison e di nuovo la sua tecnologia produceva semplicemente delle copie analogiche. Le tecnologie linguistiche possono ora semplificare e automatizzare i processi stessi di traduzione, produzione di contenuto e gestione della conoscenza per tutte le lingue europee. Possono anche arricchire interfacce intuitive a base vocale per elettrodomestici, macchinari, veicoli, computer e robot. Delle applicazioni commerciali ed industriali reali sono ancora agli stadi iniziali di sviluppo, ma i progressi di R&S stanno creando una vera finestra di opportunità. Per esempio, la traduzione automatica è già ragionevolmente accurata in settori specifici, ed alcune applicazioni sperimentali consentono la gestione multilingue dell'informazione e della conoscenza e la produzione di contenuto in molte lingue europee.

Come accade per la maggioranza delle tecnologie, le prime applicazioni linguistiche come le interfacce basate sulla voce e i sistemi di dialogo erano sviluppate per settori altamente specialistici, e spesso avevano prestazioni limitate. Ma l'integrazione delle tecnologie linguistiche nei giochi, nei siti legati al patrimonio culturale, nei pacchetti di *edutainment*, nelle biblioteche, negli ambienti di simulazione e nei programmi di training offre opportunità di mercato enormi nell'industria dell'educazione e dell'intrattenimento. I servizi mobili di informazione, il software per l'apprendimento delle lingue assistito da computer, gli ambienti di *eLearning*, gli strumenti di auto-valutazione e il software di rilevamento del plagio sono solo alcune delle aree applicative in cui le tec-

nologie linguistiche possono avere un ruolo importante. La popolarità delle applicazioni *social media* come Twitter e Facebook suggeriscono un ulteriore bisogno di tecnologie linguistiche sofisticate che consentano di monitorare i messaggi, sintetizzare le discussioni, suggerire andamenti di opinione, individuare risposte emotive, identificare violazioni di copyright o rintracciare usi impropri. Le tecnologie linguistiche rappresentano un'opportunità straordinaria per l'Unione Europea, in quanto possono aiutare ad affrontare il complesso problema del multilinguismo in Europa – il fatto che lingue diverse coesistono naturalmente nel mondo degli affari, delle amministrazioni e delle scuole. I cittadini, tuttavia, hanno bisogno di comunicare al di là di questi confini linguistici che attraversano il Mercato Comune Europeo, e le tecnologie linguistiche possono aiutare a superare quest'ultima barriera pur continuando a supportare l'uso libero e aperto delle singole lingue.

Le tecnologie linguistiche aiutano a superare quella forma di disabilità rappresentata dalla diversità linguistica.

Guardando ancora più avanti, le tecnologie linguistiche multilingui innovative rappresenteranno un punto di riferimento per i nostri partner globali quando le comunità multilingui cominceranno a dotarsene. Le tecnologie linguistiche possono essere viste come una tecnologia assistiva che aiuta a superare quella forma di disabilità rappresentata dalla diversità linguistica, rendendo le comunità linguistiche ancora più accessibili le une verso le altre. Infine, un campo di ricerca attivo è l'uso delle tecnologie linguistiche per operazioni di soccorso in aree colpite da emergenze, dove le prestazioni possono essere una questione di vita o di morte: i robot intelligenti del futuro con capacità trans-linguistiche hanno il potenziale di salvare vite umane.

2.5 LE SFIDE DELLE TECNOLOGIE LINGUISTICHE

Nonostante i considerevoli passi avanti compiuti dalle tecnologie linguistiche negli ultimi anni, il ritmo del progresso tecnologico e dell'innovazione produttiva è troppo lento. Tecnologie ampiamente usate come i correttori ortografici e grammaticali degli editori di testo sono in genere monolingui, e sono disponibili per poche lingue. I servizi di traduzione automatica on-line, sebbene utili per generare rapidamente una ragionevole approssimazione del contenuto di un documento, sono irti di difficoltà quando siano richieste delle traduzioni complete e molto accurate. A causa della complessità del linguaggio umano, modellare le nostre lingue per mezzo di un software che sia poi testato in applicazioni reali è un processo troppo lungo e costoso che richiede un impegno finanziario costante. L'Europa, quindi, deve mantenere il suo ruolo pionieristico nell'affrontare le sfide tecnologiche di una comunità multilingue, inventando nuovi metodi – tanto il progresso computazionale quanto tecniche come il *crowdsourcing* – per accelerare lo sviluppo a tutto campo.

Il ritmo del progresso tecnologico
deve essere accelerato.

2.6 L'ACQUISIZIONE DEL LINGUAGGIO NEGLI UMANI E NELLE MACCHINE

Per illustrare il modo in cui i computer gestiscono il linguaggio e il perché sia difficile programmarli ad usarlo, diamo un rapido sguardo al modo in cui gli umani acquisiscono le lingue, e vediamo poi come lavorano le tecnologie linguistiche.

Gli esseri umani acquisiscono le competenze linguistiche in due modi diversi. I bambini acquisiscono una lingua ascoltando delle interazioni reali che avvengono tra genitori, fratelli o membri della famiglia. A partire da circa due anni, i bambini producono le loro prime parole e delle brevi frasi. Questo è possibile solo perché gli esseri umani hanno una predisposizione genetica ad imitare e poi razionalizzare i suoni che sentono.

L'apprendimento di una seconda lingua ad un'età maggiore richiede più sforzo, in gran parte perché il bambino non è immerso in una comunità linguistica di parlanti nativi. A scuola, le lingue straniere di solito sono acquisite studiando la struttura grammaticale, il vocabolario e l'ortografia con esercizi che descrivono la conoscenza linguistica in termini di regole astratte, tabelle ed esempi.

Gli esseri umani acquisiscono il linguaggio
in due modi diversi: apprendendo dagli
esempi e apprendendo le regole
linguistiche che li governano.

I due tipi principali di sistemi di tecnologie linguistiche 'acquisiscono' delle capacità linguistiche in modo simile. Gli approcci statistici (o 'data driven') ricavano la conoscenza linguistica da vaste raccolte di esempi testuali concreti. Mentre è sufficiente usare del testo in una sola lingua per addestrare un correttore ortografico, per addestrare un sistema di traduzione automatica sono necessari dei testi paralleli in due (o più) lingue. L'algoritmo di *machine learning* poi "impara" dei modelli di come sono tradotte le parole, i gruppi di parole e le frasi complete.

Questo approccio statistico può richiedere milioni di frasi e la qualità delle prestazioni aumenta con la quantità di testo analizzato. Questo è uno dei motivi per cui i fornitori di motori di ricerca vogliono raccogliere il maggior numero possibile di materiale scritto. La correzione ortografica negli editori di testo, e servizi come

Google Search e Google Translate si basano tutti su approcci statistici. Il grande vantaggio della statistica è che la macchina impara velocemente in serie continue di cicli di apprendimento, anche se la qualità può variare arbitrariamente.

Il secondo approccio alle tecnologie linguistiche – e alla traduzione automatica in particolare – è quello di costruire sistemi basati su regole. Esperti di linguistica, linguistica computazionale e informatica devono prima di tutto codificare delle analisi grammaticali (regole di traduzione) e compilare liste di vocaboli (lessici). Questo lavoro è molto lungo e laborioso. Alcuni dei sistemi leader di traduzione automatica basati su regole sono stati in costante sviluppo da più di venti anni. Il grande vantaggio dei sistemi basati su regole è che gli esperti hanno un controllo più dettagliato sulla elaborazione del linguaggio. In questo modo è possibile correggere sistematicamente gli errori nel software e fornire all'utente un feedback dettagliato, soprattutto quando i sistemi basati su regole vengono utilizzati per l'apprendimento delle lingue. Ma a causa del costo elevato di questo lavoro, le tecnologie linguistiche basate su regole finora sono state sviluppate solo per le lingue principali.

Dal momento che i punti di forza e di debolezza dei sistemi statistici e di quelli basati su regole tendono ad essere complementari, la ricerca attuale si concentra sugli approcci ibridi che combinano le due metodologie. Tuttavia, questi approcci finora hanno avuto più successo nei laboratori di ricerca che in applicazioni industriali.

I due tipi principali dei sistemi di tecnologie linguistiche acquisiscono il linguaggio in modo simile.

Come abbiamo visto in questo capitolo, molte applicazioni ampiamente usate nella società dell'informazione di oggi si basano molto sulla tecnologia linguistica. Grazie alla sua comunità multilingue, questo è vero in particolar modo per lo spazio economico e di informazione europeo. Sebbene le tecnologie linguistiche abbiano fatto progressi notevoli negli ultimi anni, c'è ancora uno spazio di miglioramento enorme per la qualità dei sistemi di tecnologie linguistiche. Nei prossimi capitoli descriveremo il ruolo della lingua italiana nella società dell'informazione europea e valuteremo lo stato attuale delle tecnologie linguistiche per la lingua italiana.

The Italian Language in the Digital Age

Rehm, G.; Uszkoreit, H. (Eds.)

2012, VI, 69 p. 24 illus. in color., Softcover

ISBN: 978-3-642-30775-1