

BETWIXT

Studies in Linguistics and Communication

SERIES EDITOR:

Giuseppe **BALIRANO**

Università degli Studi di Napoli *L'Orientale* (IT)

ADVISORY BOARD:

Paul **BAKER**

Lancaster University (UK)

Susan **BASSNETT**

University of Warwick (UK)

Vijay Kumar **BHATIA**

Macquarie University (Australia)

Giuditta **CALIENDO**

Université de Lille (FR)

Rudy **LOOCK**

Université de Lille (FR)

Catalina **FUENTES RODRÍGUEZ**

Universidad de Sevilla (ES)

Bettina **MIGGE**

University College Dublin (IE)

Tommaso **MILANI**

Göteborgs Universitet (SE)

Kay **O'HALLORAN**

Curtin University, Perth (Australia)

Corinne **OSTER**

Université de Lille (FR)

Oriana **PALUSCI**

Università di Napoli *L'Orientale*

Paul **SAMBRE**

KU Leuven (BE)

Maria Grazia **SINDONI**

Università di Messina (IT)

Girolamo **TESSUTO**

Università della Campania *Luigi Vanvitelli* (IT)

JOHANNA MONTI

DALLA ZAIRJA
ALLA TRADUZIONE
AUTOMATICA

*Riflessioni sulla traduzione
nell'era digitale*

PAOLO 
LOFFREDO

Volume pubblicato con il contributo dell'Ufficio Progetti Europei e Internazionali dell'Università degli Studi di Napoli L'Orientale.

Proprietà letteraria riservata

Immagine di copertina:

Erich Kettelhut (1927), *City from Above with Tower of Babel*.

Finito di stampare nel mese di settembre 2019

ISBN 978-88-32193-17-6

ISSN 2611-1349 (collana)

PAOLO
LOFFREDO



© 2019 **Paolo Loffredo** Editore s.r.l.

Via Ugo Palermo, 6 - 80128 Napoli

www.loffredoeditore.com

paoloffredoeditore@gmail.com

INDICE

Lista delle figure	7
<i>Introduzione</i>	9
CAPITOLO I	
LE TECNOLOGIE PER LA TRADUZIONE: DEFINIZIONI, USI E MODELLI <i>CLOUD</i>	15
1.1 Alcune definizioni e classificazioni	16
1.2 Usi della traduzione automatica	20
1.3 Modelli <i>cloud</i> di traduzione automatica	25
CAPITOLO II	
TRADUZIONE AUTOMATICA E TEORIE DELLA TRADUZIONE	31
2.1 Traduzione automatica e le teorie linguistiche della traduzione	31
2.2 Traduzione Automatica e gli studi teorici e applicati della traduzione	33
2.3 La svolta tecnologica dei <i>Translation Studies</i>	42
2.4 Approcci cognitivi	44
2.5 Approcci sociologici	45
CAPITOLO III	
LA PROSPETTIVA STORICA: DALLA ZAIRJA A GOOGLE TRADUTTORE	51
3.1 Il sogno della macchina intelligente	51
3.2 I precursori	52
3.3 <i>Cerveau mécanique</i> di Georges B. Artsrouni e la macchina di Pëtr Petrovič Smirnov-Trojanskij	54
3.4 Gli inizi	55
3.5 Gli anni '50 e '60: l'euforia iniziale e la disillusione	58
3.6 Gli anni '70 e '80: la rinascita	61
3.7 Gli anni '90: la tecnologia al servizio del traduttore	65
3.8 I servizi di traduzione automatica su Internet	66
CAPITOLO IV	
L'EVOLUZIONE TECNOLOGICA	73
4.1 Approcci linguistici	74
4.1.1 Sistemi diretti o di prima generazione	75

INDICE

4.1.2 Sistemi indiretti o di seconda generazione	78
4.2 Approcci empirici	87
4.2.1 Sistemi basati su esempi	87
4.2.2 Sistemi statistici	89
4.2.3 Sistemi ibridi	96
4.2.4 Sistemi neurali	99
CAPITOLO V	
IL PROCESSO DI TRADUZIONE AUTOMATICA E LE COMPETENZE DEL TRADUTTORE NELL'ERA DIGITALE	105
5.1 <i>Pre-editing</i> e uso dei linguaggi controllati	107
5.2 <i>Post-editing</i>	110
5.3 Competenze del traduttore nell'era digitale	118
CAPITOLO VI	
QUALITÀ DELLA TRADUZIONE AUTOMATICA E LE METRICHE DI VALU- TAZIONE	127
6.1 Dal rapporto ALPAC al <i>Framework for the Evaluation of Machine Translation in ISLE</i>	127
6.2 Le metriche di valutazione basate sul giudizio umano	134
6.3 Le metriche automatiche di valutazione	137
6.4 Gli approcci recenti	142
CAPITOLO VII	
LA TRADUZIONE AUTOMATICA PARTECIPATA: IL CONTRIBUTO DEL CROWDSOURCING	149
7.1 Definizioni	149
7.2 <i>Crowdsourcing</i> e traduzione	151
7.3 <i>Crowdsourcing</i> e tecnologie per la traduzione	154
7.3.1 <i>Crowdsourcing</i> esplicito	156
7.3.2 <i>Crowdsourcing</i> implicito	162
7.4 Questioni di qualità	162
CAPITOLO VIII	
OSSERVAZIONI CONCLUSIVE	165
BIBLIOGRAFIA	169
Indice dei nomi	195

LISTA DELLE FIGURE

Figura 1.1	Traduzione umana e traduzione automatica	17
Figura 1.2	Classificazione della traduzione automatica in base all'approccio	20
Figura 1.3	Esempio di risultato della funzione di CLIR in Google	22
Figura 1.4	Esempio di CLIR su eBay	23
Figura 1.5	Skype Translator	24
Figura 1.6	Il modello di Google Translator Toolkit	28
Figura 2.1	Il processo di traduzione (Nida & Taber 1969: 33)	32
Figura 2.2	Mappa dei <i>Translation Studies</i> (Toury 1995:10)	34
Figura 2.3	Schema degli <i>Applied Translation Studies</i> (Quah 2006)	42
Figura 2.4	<i>Institutional analysis and development framework</i> (Moorkens & Lewis 2019)	48
Figura 3.1	<i>Ars brevis</i> XVIII Century	53
Figura 3.2	<i>Cerveau mécanique</i> di G. Artsrouni (da Corbé 1960)	54
Figura 3.3	Schema di flusso di una parte delle procedure di consultazione del dizionario del sistema di traduzione automatica di IBM	58
Figura 3.4	Esempio di grammatica formale	59
Figura 3.5	Schermata di Babel Fish	68
Figura 4.1	<i>Vauquois triangle</i> (Vauquois 1968)	74
Figura 4.2	Schema di flusso di un sistema diretto	76
Figura 4.3	Schema di flusso di un sistema a <i>transfer</i>	79
Figura 4.4	<i>Semanto-Syntactic Abstraction Language</i> (SAL): i nomi, con un dettaglio sui nomi aspettivi (Scott 2003)	81
Figura 4.5	L'architettura del sistema LOGOS	82
Figura 4.6	Esempio di regola TRAN scritta in FORTRAN	84
Figura 4.7	Sistema di traduzione automatica a interlingua	85
Figura 4.8	Sistema di traduzione automatica basato su esempi	88
Figura 4.9	Traduzione Automatica di tipo statistico	90
Figura 4.10	Esempio di allineamento errato in GIZA++ tra la polirematica inglese <i>I can't help</i> e la sua traduzione italiana <i>non posso fare a meno di</i>	92

LISTA DELLE FIGURE

Figura 4.11	Esempio di <i>Phrase table</i> di Moses	93
Figura 4.12	Rete neurale semplice	100
Figura 4.13	Schema del primo sistema di traduzione automatica neurale	102
Figura 5.1	Processo di traduzione basato sulla traduzione automatica	106
Figura 5.2	<i>Light e full post-editing</i> - Common Sense Advisory (CSA)	111
Figura 5.3	Competenze e abilità per il post-editing (Rico & Torrejón 2012: 169)	121
Figura 5.4	<i>European Master's in Translation (EMT) Competence Framework 2017</i>	123
Figura 6.1	Schema generale della valutazione di un sistema di traduzione automatica (Lehrberger & Bourbeau 1988:133)	132
Figura 6.2	<i>Precision, recall e F-measure</i>	138
Figura 6.3	Esempio di metrica BLEU applicata a una traduzione di una frase inglese in italiano	139
Figura 6.4	Schema della metrica MQM-DQF	147
Figura 7.1	L'applicazione Translate Facebook	152
Figura 7.2	Google Traduttore	158
Figura 7.3	Google Traduttore: commenti	158
Figura 7.4	Finestra di convalida delle traduzioni nella <i>community</i> di Google Traduttore	159
Figura 7.5	Uso del <i>crowdsourcing</i> in un ciclo di traduzione (Carson-Berndsen <i>et al.</i> 2010)	160
Figura 7.6	Esempio di <i>traduzione in linea</i> di Facebook	161
Figura 8.1	Giudizio degli studenti del laboratorio di Traduzione Automatica ed Assistita dell'Orientale di Napoli sull'utilità delle tecnologie	167

INTRODUZIONE

The Babel fish is small, yellow and leechlike, and probably the oddest thing in the Universe. It feeds on brainwave energy received not from its own carrier but from those around it. It absorbs all unconscious mental frequencies from this brainwave energy to nourish itself with. It then excretes into the mind of its carrier a telepathic matrix formed by combining the conscious thought frequencies with nerve signals picked up from the speech centres of the brain which has supplied them. The practical upshot of all this is that if you stick a Babel fish in your ear you can instantly understand anything said to you in any form of language. The speech patterns you actually hear decode the brainwave matrix which has been fed into your mind by your Babel fish.

(Douglas Adams,
The Hitchhiker's Guide to the Galaxy)

Douglas Adams nel suo romanzo *The Hitchhiker's Guide to the Galaxy* (1979) narra di *Babel Fish*, un piccolo dispositivo a forma di pesce e dal colore giallo che inserito nell'orecchio rende possibile abbattere le barriere linguistiche, consentendo la comunicazione in qualsiasi lingua. Se ai tempi della scrittura di quest'opera l'idea poteva sembrare fantascientifica, siamo ora in un'epoca dove la tecnologia ha davvero reso possibile realizzare un traduttore universale.

Le tecnologie per la traduzione, infatti, sono ormai un elemento imprescindibile della pratica lavorativa dei traduttori e della comunicazione su Internet, e riguardano da un lato applicazioni che automatizzano completamente il processo, come la traduzione automatica, e dall'altro strumenti di ausilio ai traduttori come, ad esempio, memorie di traduzione, corpora informatizzati e database terminologici, che lasciano al traduttore il compito di governare il processo dandogli però la possibilità di velocizzare i tempi di realizzazione e di migliorare la qualità del prodotto finale.

INTRODUZIONE

In particolare la traduzione automatica, a partire dal primo servizio online gratuito, dal nome appunto di Babel Fish, ad oggi ha fatto passi da gigante grazie all'evoluzione della tecnologia, all'adozione dei modelli neurali, alla disponibilità di enormi quantità di dati multilingui su Internet necessari per l'addestramento dei sistemi, tanto da far proclamare trionfalmente ad alcuni studiosi di aver finalmente realizzato il sogno dei primi ricercatori in questo campo, ovvero di aver raggiunto la *Fully High Quality Automatic Translation*, una traduzione automatica di qualità pari a quella umana.

La storia della traduzione automatica è da sempre caratterizzata dagli entusiasmi, talvolta eccessivi, da parte dei ricercatori impegnati nella progettazione e nello sviluppo dei sistemi, a cui però fanno da contraltare i timori dei traduttori nei confronti di nuove tecnologie. C'è da chiedersi se di fronte a progressi tecnologici così repentini e a cambiamenti così profondi nel mondo della traduzione queste preoccupazioni non siano del tutto infondate.

Partendo dalla convinzione che il progresso non si può fermare, ma che piuttosto è necessario conoscerne tutti i diversi risvolti e le implicazioni per poter adeguatamente affrontare la svolta tecnologica in questo settore, il presente volume intende riflettere su alcuni temi attuali nel campo della ricerca, come ad esempio i sistemi di traduzione automatica di tipo neurale, l'uso del *crowdsourcing* per l'addestramento e l'aggiornamento dei sistemi, nuovi processi di traduzione resi possibili dalle tecnologie *cloud* applicate alla traduzione e richiesti dal mercato dell'industria delle lingue, la qualità dei sistemi allo stato dell'arte e la comparabilità con le traduzioni umane.

L'idea di questo libro si ricollega alle mie esperienze di ricerca passate, avendo lavorato prima come linguista computazionale ad un progetto di traduzione automatica per la realizzazione delle coppie linguistiche inglese-italiano e tedesco-italiano del sistema sviluppato dalla LOGOS Corporation (USA) e successivamente come responsabile di un servizio di traduzioni basato sull'utilizzo di sistemi di ausilio alla traduzione. Ho affrontato in prima persona la sfida, per me davvero entusiasmante, dell'utilizzo di nuove tecnologie per il compito della traduzione, per altro in un periodo in cui la traduzione automatica era una tecnologia ancora molto di frontiera.

Ricogliendomi a questo periodo della mia vita lavorativa¹ mi piace ricordare anche la storia narrata da Socrate a Fedro nell'omonimo dialogo

¹ Ero responsabile del settore delle ricerche in Linguistica Computazionale presso il Consorzio Thamus a Salerno (Gruppo Finmeccanica).

INTRODUZIONE

platonico (*Fedro*, 274c-76a in Platone 1967: 790–792) del faraone Thamus, che ricevette in dono dal dio Theuth l’invenzione della scrittura affinché la diffondesse tra gli uomini: “Questa conoscenza o faraone renderà gli Egizi più sapienti e più capaci di ricordare: è stata infatti inventata come medicina per la memoria e la sapienza”. Ma il faraone della città di Tebe, non convinto da questa affermazione rispose al dio:

Ingegnosissimo Theuth, c’è chi è capace di dar vita alle arti, e chi invece di giudicare quale danno e quale vantaggio comportano per chi se ne avvarrà. E ora tu, padre della scrittura, per benevolenza hai detto il contrario di ciò che essa è in grado di fare. Questa infatti produrrà dimenticanza nelle anime di chi l’avrà appresa, perché non fa esercitare la memoria. Infatti, facendo affidamento sulla scrittura, essi trarranno i ricordi dall’esterno, da segni estranei, e non dall’interno, da se stessi.

La preoccupazione del faraone era che l’invenzione tecnologica, rappresentata dalla scrittura, potesse solo indebolire la memoria degli uomini e non li aiutasse a ricordare, come invece sosteneva l’inventore dio Theuth, ma al contrario impoverisse la mente umana costretta ad affidarsi a qualcosa di esterno.

La paura del faraone è rappresentativa ed esemplificativa di tutte le paure che gli esseri umani hanno da sempre nutrito nei confronti delle innovazioni tecnologiche. La sindrome di “Theuth”, se così possiamo definire il timore che la tecnologia possa modificare radicalmente il modo di pensare e di operare dell’uomo, si riflette in questo caso nella paura del “computer che sa tradurre” in grado di sostituire l’essere umano. In realtà i progressi tecnologici nel campo della traduzione, benché inarrestabili, e nonostante i proclami entusiastici di qualche ricercatore, non sono ancora tali da aver soppiantato l’essere umano, che è ancora attore indiscusso del processo traduttivo, ancorché supportato dalle tecnologie.

Vero è che però le tecnologie della traduzione fanno parte del nostro quotidiano rapporto con Internet in un mondo ormai globalizzato, agevolando l’abbattimento delle barriere linguistiche, e nel contempo ai professionisti del settore si richiede sempre di più di revisionare testi pre-tradotti automaticamente invece di tradurre ex novo.

Per lungo tempo la teoria della traduzione ha sottovalutato la portata delle nuove tecnologie per la traduzione e solo recentemente studiosi, principalmente di area anglosassone, hanno iniziato a riflettere sul fenomeno con studi sia di impianto teorico, sia di tipo applicato, con particolare attenzione alle implicazioni nella pratica lavorativa e nella formazione dei traduttori.

INTRODUZIONE

In ambito italiano i contributi su questo argomento sono relativamente pochi (Osimo 2001; Berselli *et al.* 2011; Naldi 2014; Tavosanis 2018) e dunque questo libro nasce dall'esigenza di proporre un testo che tratti un argomento di centrale rilevanza in un settore segnato da profonde trasformazioni tecnologiche al pari, se non addirittura di più, di altri settori lavorativi: la traduzione automatica appunto.

L'intento è quello di fornire uno strumento utile ad approfondire il tema delle tecnologie attualmente in uso nei processi di traduzione offrendo un percorso di lettura articolato su diversi punti di vista, fornendo informazioni su cosa sia la traduzione automatica, sulla sua evoluzione nel tempo e sulle sue possibili applicazioni al giorno d'oggi. Allo stesso tempo il lettore troverà argomenti utili a una comprensione più approfondita del tema: gli argomenti selezionati, infatti, forniscono una panoramica critica, anche se certamente non esaustiva, sulle ricerche in questo settore. Benché una comprensione approfondita del tema richiederebbe delle competenze informatiche, in realtà la prospettiva offerta al lettore è maggiormente orientata agli studi sulla traduzione.

Un certo numero di temi e di osservazioni collezionate in questo volume sono il risultato di una serie di studi da me condotti nel tempo e presentati in forma preliminare in altri articoli, durante le mie lezioni all'università degli studi di Napoli "L'Orientale" e altrove.

Il volume, composto di otto capitoli, inizia con *Le tecnologie per la traduzione: definizioni, usi e modelli cloud*, che introduce i concetti principali relativi alla traduzione automatica, ne illustra i principali usi e i modelli disponibili attualmente sul web.

Il capitolo *Traduzione automatica e teorie della traduzione* indaga il rapporto tra due ambiti di studio che per lungo tempo hanno seguito strade diverse. Gli apporti teorici nel campo della traduzione automatica si sono principalmente concentrati sui modelli più efficaci per la comprensione e la generazione del linguaggio naturale, mentre le teorie della traduzione succedutesi nel tempo hanno dedicato poco spazio alle tecnologie almeno fino al cosiddetto *Technological turn* nei *Translation Studies*, che ha preso atto dei cambiamenti significativi del modo in cui la traduzione viene effettuata nell'era digitale.

Nel capitolo intitolato *La prospettiva storica: da Zairja a Google traduttore* si ripercorre la storia del sogno di automatizzare la traduzione, uno dei processi più complessi della mente umana. Partendo dai primi tentativi di riprodurre il ragionamento umano, che si possono far risalire al addirittura al XIII secolo, il percorso storico individua le tappe fondamentali nella storia

INTRODUZIONE

della traduzione automatica: dall'ascesa e il fallimento del *sogno meccanico* di sistemi in grado di tradurre come un essere umano, per arrivare ai giorni d'oggi in cui la tecnologia non ha sicuramente sostituito il traduttore bensì fornisce una serie di strumenti di lavoro utili a velocizzare e ottimizzare il processo traduttivo.

Il terzo capitolo *L'evoluzione tecnologica* presenta una panoramica generale delle diverse metodologie utilizzate per realizzare i sistemi di traduzione automatica: dai sistemi di prima generazione, basati su approcci linguistici, ai sistemi neurali, che progressivamente stanno rimpiazzando il paradigma statistico che ha dominato gli ultimi venti anni della storia di questa tecnologia. I diversi approcci vengono esemplificati attraverso una presentazione dettagliata da un punto di vista teorico di alcuni casi di successo, ovvero alcuni sistemi scelti per le loro particolari caratteristiche.

Il capitolo intitolato *Il processo di traduzione automatica e le competenze del traduttore nell'era digitale* illustra il processo basato su sistemi di traduzione automatica, soffermandosi in particolare sulle fasi che richiedono l'intervento umano, ovvero il *pre-editing*, con cui si cerca di migliorare un testo di partenza per renderlo più facilmente interpretabile e traducibile da parte di una macchina, e il *post-editing*, finalizzato a revisionare la traduzione grezza con diversi gradi di intervento a seconda della qualità finale concordata con i clienti. Il capitolo si conclude con una discussione delle competenze e delle abilità necessarie ai traduttori per operare in ambienti di lavoro sempre più tecnologici.

Qualità e valutazione della traduzione automatica sposta l'enfasi su un aspetto molto importante in traduzione, ovvero sulla qualità dei risultati. Anche in questo caso si propongono all'attenzione del lettore le varie metodologie adottate, trattando nel dettaglio sia le metriche quantitative ampiamente utilizzate dagli sviluppatori delle tecnologie perché di facile esecuzione e a basso costo e metriche qualitative che si basano sul giudizio umano. Il discorso si focalizza sui vari criteri di qualità utilizzati, ed evidenzia come differiscano nella valutazione della traduzione automatica e quella umana.

Il capitolo *La traduzione automatica partecipata: il contributo del crowdsourcing* affronta il reperimento di risorse linguistiche necessarie al funzionamento e al miglioramento dei sistemi, attraverso l'uso del *crowdsourcing*, termine usato per la prima volta da Jeff Howe nel suo articolo sulla rivista *Wired* (Howe 2006), *The rise of Crowdsourcing* per indicare un'attività che si basa sul contributo di una moltitudine indistinta (*crowd*) su base volontaristica o a pagamento. Nel caso della traduzione automatica il

INTRODUZIONE

crowdsourcing viene usato per lo sviluppo di corpora testuali e altre risorse linguistiche, ovvero di quelle basi di dati utili per il miglioramento delle tecnologie per la traduzione. Il *crowdsourcing* ricopre un ruolo fondamentale da questo punto di vista, in quanto consente la creazione collaborativa di risorse e informazioni su larga scala a costi contenuti e in breve tempo. Nel capitolo vengono evidenziati aspetti negativi e positivi di questa pratica ampiamente utilizzata dagli sviluppatori, ma che assume modalità diverse.

Il volume è completato dalle osservazioni conclusive e da una bibliografia che raccoglie una selezione delle fonti di riferimento in questo settore.

Per concludere questa breve introduzione desidero ringraziare tutti coloro che mi hanno aiutato a portare a termine questo volume: in primo luogo la mia famiglia, senza il cui supporto e incoraggiamento in un lungo periodo di intenso lavoro, questa impresa, direi quasi titanica tra mille altri impegni di didattica e ricerca, non sarebbe andata a buon fine; Maria Pia di Buono e Gloria Gagliardi, ricercatrici appena assunte all'Orientale, che hanno letto interamente la prima bozza del libro fornendomi consigli preziosi per migliorarla, così come le mie amiche Stefania Taddeo, che si è offerta volontaria per la correzione di bozze, e Daniela Chavarria Devia che mi ha aiutato a controllare l'intera bibliografia del volume. Sono grata inoltre a Marco Turchi dell'FBK per avermi dato dei suggerimenti utili relativamente agli approcci statistici e neurali; al direttore della collana e collega dell'Orientale, Giuseppe Balirano e infine all'editore Paolo Loffredo per aver creduto in questo progetto editoriale. I miei ringraziamenti finali vanno al revisore anonimo che si è assunto l'onore di valutare l'opera e ne ha dato un giudizio positivo.