



SCUOLA SUPERIORE PER MEDIATORI LINGUISTICI
(Decreto Ministero dell'Università 31/07/2003)

Via P. S. Mancini, 2 – 00196 - Roma

**TESI DI DIPLOMA
DI
MEDIATORE LINGUISTICO**

(Curriculum Interprete e Traduttore)

Equipollente ai Diplomi di Laurea rilasciati dalle Università al termine dei Corsi afferenti alla classe delle

**LAUREE UNIVERSITARIE
IN
SCIENZE DELLA MEDIAZIONE LINGUISTICA**

*“MACHINE TRANSLATION, CAT-TOOLS E IA:
I LIMITI DELLA TRADUZIONE AUTOMATICA”*

RELATORI:
Prof.ssa Adriana Bisirri

CORRELATORI:
Prof. Fabio Matassa
Prof.ssa Luciana Banegas
Prof.ssa Maggie Papparusso

CANDIDATA
Arianna Parente
3157

ANNO ACCADEMICO 2022/2023

Indice

Introduzione	1
CAPITOLO UNO – LA TRADUZIONE AUTOMATICA	3
1.1 Il traduttore come figura professionale	3
1.2 La nascita della traduzione automatica	5
1.3 Rule-based, Statistical e Neural Machine Translation	13
CAPITOLO DUE – I LIMITI DELLA TRADUZIONE AUTOMATICA.....	21
2.1 I principali sistemi di traduzione automatica su Internet	21
2.1.1 Google Translate, Microsoft Translator e DeepL	23
2.1.2 I CAT Tools	31
2.1.3 Dispositivi tecnologici per l’interpretariato	38
2.2 Pre-editing, post-editing e linguaggio controllato	40
2.3 Le metriche di valutazione qualitativa della traduzione automatica: BLEU, DQF, MQM	47
2.4 I limiti dell’intelligenza artificiale nel campo della traduzione	54
CAPITOLO TRE – L’ESEMPIO DI <i>STRAPPARE LUNGO I BORDI DI</i>	
ZEROCALCARE	60
3.1 L’adattamento del dialetto e dell’ironia	60
3.2 Analisi di parti selezionate dalla serie televisiva	63
Conclusioni	72
Introduction	74
CHAPTER ONE – MACHINE TRANSLATION	76
1.1 The birth of machine translation.....	76
1.2 Rule-based, Statistical and Neural Machine Translation	79

CHAPTER TWO – THE LIMITS OF MACHINE TRANSLATION	82
2.1 The main systems of Machine Translation on Internet	82
2.2 Pre-editing, post-editing and controlled language	87
2.3 The limits of AI and the importance of human translators	90
2.4 The adaptation of dialect and humor	94
Conclusion	99
Introducción	101
CAPÍTULO UNO – LA TRADUCCIÓN AUTOMÁTICA	103
1.1 El nacimiento de la traducción automática	103
1.2 Rule-based, Statistical y Neural Machine Translation	106
CAPÍTULO DOS – LOS LÍMITES DE LA TRADUCCIÓN AUTOMÁTICA	111
2.1 Los principales servicios de traducción automática	111
2.2 Pre-editing, post-editing y lenguaje controlado	119
2.3 Los límites de la traducción automática	122
2.4 La adaptación del dialecto y de la ironía	125
Conclusión	129
BIBLIOGRAFIA	131
SITOGRAFIA	132

Introduzione

La pratica della traduzione ha da sempre rivestito un ruolo chiave alla base del progresso e della comunicazione e, con oltre 6700 lingue parlate oggi nel mondo, diventa imprescindibile abbattere le barriere linguistiche attraverso questa disciplina. Nell'era digitale in cui viviamo, a causa della rapida evoluzione tecnologica degli ultimi anni, anche la traduzione ha subito una trasformazione significativa: se in passato era considerata una pratica "primitiva", in cui il traduttore si affidava esclusivamente a carta, penna e dizionario per tradurre testi, oggi questa si sta trasformando in una disciplina sempre più avanzata dal punto di vista tecnologico, la quale si propone di fornire una traduzione in tempi record con un semplice *click*.

La traduzione automatica si presenta come una risorsa potenzialmente rivoluzionaria per abbattere le barriere linguistiche e facilitare lo scambio di informazioni tra individui e culture diverse, ma, nonostante i suoi significativi progressi, continua a mostrare limiti evidenti che ne minano l'efficacia.

Il presente progetto di tesi si propone di analizzare i limiti della traduzione automatica e sottolineare l'importanza insostituibile del traduttore umano nella comunicazione interculturale e interlinguistica, soprattutto alla luce di esperienze personali maturate durante il percorso di studi. Verrà analizzata nel dettaglio la nascita della traduzione automatica, la cui storia affonda le radici negli anni Quaranta del Novecento, per poi descrivere tutta la sua evoluzione fino ad oggi. Il progresso a cui è andata incontro ha seguito un percorso che ha abbracciato diversi approcci tecnologici: dai primi sistemi basati su regole, che facevano uso di algoritmi e strutture linguistiche predefinite, si è passati a modelli statistici e, più recentemente, a sistemi neurali, i quali sfruttano reti artificiali per apprendere dati e migliorare le prestazioni, ricreando il lavoro svolto dal cervello umano. Oggi i numerosi servizi online, ampiamente diffusi, si avvalgono proprio di quest'ultimo approccio alla traduzione nell'elaborazione dei dati.

Dal momento che non tutti i traduttori automatici sono uguali, si dedicherà una sezione della tesi a un confronto dettagliato tra di essi, al fine di evidenziarne le diversità e di offrire un paragone significativo delle loro potenzialità e prestazioni. Per valutare la loro efficacia, inoltre, sono state sviluppate varie tecniche di valutazione che consentono di misurare la qualità e l'accuratezza delle traduzioni e di cui si parlerà nel dettaglio nel secondo capitolo.

Successivamente, una volta esposte le varie motivazioni che rendono la traduzione automatica ancora oggi molto limitata, verranno forniti vari esempi pratici a dimostrazione di tali affermazioni: per fare ciò, verranno selezionate alcune scene dalla serie televisiva Netflix “Strappare lungo i bordi” del fumettista Zerocalcare, distribuita in streaming nel 2021, scritta in dialetto romano e ricca di elementi comici. La scelta non è casuale: l’obiettivo sarà quello di dimostrare quanto la traduzione di testi di questo genere sia una sfida complessa, che richiede non solo competenze linguistiche, ma anche un profondo background culturale e una conoscenza approfondita delle sfumature culturali che, talvolta, risulta complessa anche per una mente umana. Proporre una traduzione fedele che non solo trasmetta il significato letterale, ma che preservi anche l’umorismo e il contesto culturale intrinseco al testo originale è un compito arduo, che una traduzione automatica non sarebbe in grado di attuare, non avendo le conoscenze necessarie per poter offrire una resa soddisfacente.

Con il presente progetto di tesi, si vuole, quindi, dimostrare quanto la traduzione automatica sia ancora oggi molto limitata, nonostante la grande evoluzione subita nell’ultimo periodo, ma, allo stesso tempo, verranno individuate alcune aree in cui la tecnologia può integrarsi positivamente nel processo traduttivo, per esempio tramite l’uso di strumenti di traduzione assistita (CAT Tools), i quali sfruttano le loro potenzialità per migliorare l’efficienza e la produttività del traduttore umano. Tramite il loro utilizzo, infatti, il compito dei professionisti del settore viene facilitato, vengono ottimizzati i tempi di trasposizione e si riesce ad avere una base da cui partire.

Ad ogni modo, i sistemi che si trovano oggi online non sono ancora in grado di cogliere quelle sfumature di significato che solamente un essere umano, con i necessari studi, è capace di comprendere ed è bene, quindi, sottolineare l’importanza insostituibile del traduttore umano.

Grazie all’analisi delle diverse esperienze personali e delle sfide incontrate nel campo della traduzione, si cercherà di fornire un contributo significativo alla comprensione di questo fenomeno, sempre più rilevante nel contesto della globalizzazione linguistica e culturale.

CAPITOLO UNO

LA TRADUZIONE AUTOMATICA

1.1 Il traduttore come figura professionale

La necessità di comprendere le lingue straniere risale a migliaia di anni fa, quando ancora le lingue parlate erano davvero poche e per questo la pratica della traduzione si configura da sempre come una delle attività umane alla base del progresso, degli scambi commerciali e culturali.

Partendo da un concetto base, la traduzione è quella pratica che consiste nell'interpretazione di un testo di partenza e nella conseguente produzione di un nuovo testo in un'altra lingua, che sia il più possibile equivalente, in termini di senso e significato, al testo originale. L'obiettivo del traduttore è, quindi, quello di portare il testo dalla lingua di origine alla lingua di destinazione cercando di preservare il significato e lo stile del testo ed eseguendo un processo di adattamento, se necessario. Certamente, non potrà mai esistere una traduzione perfetta, dal momento che un traduttore è libero di riformulare il testo secondo il suo stile adottando un proprio *modus operandi*, ma allo stesso tempo deve rimanere fedele al pensiero e al senso del testo base, cercando di trasporre in maniera leale tutto ciò che l'autore ha voluto comunicare. Per questo motivo, il traduttore tenta sempre di ampliare al massimo le proprie conoscenze e di aggiornare i propri mezzi, per utilizzare al meglio la lingua durante l'attività di adattamento e traduzione. Essere un buon traduttore non significa semplicemente comprendere a fondo il messaggio espresso dall'autore, ma vuol dire interpretare al meglio la cultura con cui si entra in contatto. In una determinata lingua, per esempio, una parola può avere varie sfumature ed è fondamentale che si riesca a trovare un corrispettivo esatto in un'altra lingua per non alterare il concetto espresso. Da un punto di vista storico, la traduzione ha origine dai tempi della mitica "Torre di Babele" che con la sua distruzione sancì la fine dell'unità linguistica. A quel punto Dio fece in modo che i discendenti di Noè, costruttori della storica Torre che volevano sfidare l'autorità del Creatore, iniziassero a parlare lingue diverse in modo tale da rendere impossibile la comprensione a vicenda e, successivamente, li disperse ai quattro angoli della Terra. Da qui nacque la necessità di comprendere e interpretare i diversi linguaggi per favorire la comunicazione tra esseri umani. Una delle primissime

testimonianze di traduzione scritta è quella risalente alla “Stele di Rosetta”. Si tratta di una stele di origine egiziana contenente un decreto del re Tolomeo V, emesso a Menfi nel 196 a.C., scritto in tre lingue: la parte superiore è scritta in egiziano antico, la parte centrale in egiziano demotico e quella inferiore, invece, in greco antico.

Secondo gli studiosi, una prima teoria della traduzione risale ai tempi degli antichi Romani, quando Cicerone ed Orazio (I secolo a.C.) fecero una prima distinzione tra traduzione letterale e traduzione per significato. Nel suo *De optimo genere oratorum*, scritto nel 46 a.C., Cicerone presenta ad Eschine e Demostene la sua personale traduzione dei discorsi spiegando che: *“non li ho tradotti come un interprete ma come un oratore, mantenendo le stesse idee e le stesse forme, o come qualcuno potrebbe definirle, le stesse espressioni del pensiero, ma in una lingua conforme all’uso quotidiano. E così facendo, non ho ritenuto necessario rendere il testo con una traduzione parola per parola ma ho mantenuto lo stile e la potenza del linguaggio.”*¹

I loro studi sulla traduzione hanno certamente influenzato le generazioni successive di traduttori fino al XX secolo. Le teorie traduttive più moderne fanno riferimento ai primi anni del Seicento, quando teorici del campo, come Sir John Denhom, Alexander Pope e John Dryden, diedero il loro contributo a questa pratica distinguendo tre diversi tipi di traduzione: la metafrasi, un meccanismo più rigido nel quale ogni parola da tradurre viene sostituita con un preciso corrispettivo nella lingua di arrivo; la parafrasi, che a livello terminologico concede una maggiore libertà al traduttore, il quale deve però sempre rimanere fedele al contenuto originale; l’imitazione, un concetto secondo cui il traduttore viene paragonato quasi ad un poeta che rielabora il testo pensando a come lo avrebbe fatto l’autore e a ciò che quest’ultimo voleva comunicare con la sua scrittura. Da quel momento in poi, dal Seicento fino ad oggi, tanti e importantissimi sono stati i progressi e le evoluzioni che hanno perfezionato sempre di più questa pratica, facendo sì che arrivasse ad alti livelli. Nell’Ottocento la figura del traduttore non si era ancora sviluppata a livello professionale proprio perché quest’ultimo era visto semplicemente come una persona bilingue con una particolare inclinazione a tradurre un testo in una lingua diversa dalla propria. Generalmente, i traduttori traducevano unicamente componimenti che facevano riferimento alla loro professione e di conseguenza ognuno di loro era specializzato in un preciso settore; questo modello

¹ Marco Tullio Cicerone, *De optimo genere oratorum*, 46 a.C. versi 11-15

rimase immutato fino al XX secolo. Una volta compresa la difficoltà del lavoro di traduzione, si iniziarono a formare le prime figure professionali di traduttori che, a differenza dei loro predecessori, erano capaci di fare propri anche un lessico e un linguaggio settoriale totalmente diverso da quello in cui erano specializzati. Nel XIX secolo la traduzione diventò una vera e propria disciplina. All'inizio faceva parte solamente dell'insegnamento delle lingue straniere, mentre successivamente veniva utilizzata anche come esercizio per migliorare l'apprendimento delle altre discipline.

1.2 La nascita della traduzione automatica

Con il termine “Traduzione Automatica” (TA), nota in lingua inglese come “Machine Translation” (MT), si intende il procedimento tramite il quale viene impiegata l'intelligenza artificiale per tradurre automaticamente un testo di partenza da una lingua ad un'altra, senza che ci sia nessun tipo di intervento umano. Si tratta di un'area della linguistica computazionale e della scienza della traduzione, la quale si occupa dell'analisi e dell'implementazione di programmi informatici per la traduzione automatica di testi.

L'idea che una macchina potesse tradurre il linguaggio umano è stata considerata per molto tempo una fantasia irraggiungibile. Tuttavia, nel XX secolo, le prime ricerche concrete in questo campo iniziarono a prendere forma. I primi tentativi risalgono agli anni '30, nello specifico al 1933, quando lo studioso Pëtr Petrovič Smirnov-Trojanskij, ingegnere e inventore russo, teorizzò un sistema per la creazione di un dizionario bilingue automatico basato sull'interazione tra l'uomo e la macchina. Il suo progetto rivoluzionario fu da lui descritto come *“una macchina per la selezione e la scrittura delle parole nella traduzione da una lingua verso una o più altre lingue”*. Fu brevettata a Mosca, ma in realtà non fu mai concretamente costruita. Nell'epoca in cui visse Trojanskij, in Russia si parlavano decine e decine di lingue e proprio per questo motivo nacque in lui l'idea di voler tradurre lingue diverse dal russo in maniera più veloce. Il suo progetto era quello di trovare due persone monolingue che traducessero da una lingua all'altra. La teoria di Trojanskij, davvero innovativa a quel tempo, era quella che tutte le lingue avessero una grammatica comune, denominata “grammatica

universale” da Chomsky², e che quindi sarebbe stato più facile per la macchina di traduzione, che il testo di partenza subisse un processo di trasformazione da un periodo complesso ad una forma base. Nello specifico, il progetto prevedeva tre passaggi: un’analisi logico-sintattica delle parole per scomporle nella loro forma base, che veniva effettuata da un esperto nella lingua di partenza del testo; la traduzione di queste forme base e funzioni sintattiche nella lingua di arrivo attraverso una macchina automatica; infine, la trasformazione di queste forme base in periodi complessi tramite il processo di post-editing eseguito da un esperto della lingua di arrivo.

Sebbene egli diede il suggerimento che la traduzione umana potesse effettivamente essere meccanizzata, le idee di Trojanskij non ottennero il giusto riconoscimento e il suo progetto fallì, anche a causa della sua morte prematura. Il suo ruolo fu però fondamentale per i successivi studi nel campo della traduzione automatica e la sua idea gettò le basi per lo sviluppo di ulteriori teorie e strumenti in questo ambito. La vera nascita della traduzione si ebbe solo con la comparsa del calcolatore e con il diffondersi delle idee sugli universali linguistici, ovvero regole che erano comuni a tutte le lingue naturali e che facevano sperare nella nascita di una traduzione automatica che potesse fare a meno dell’intervento umano da parte di un traduttore specializzato.

È all’inizio degli anni ‘50, e più precisamente nel 1949, che si verifica una vera e propria svolta nell’ambito della Traduzione Automatica con la pubblicazione del *memorandum* intitolato “Translation”, redatto dallo scienziato statunitense Warren Weaver, direttore della Divisione di Scienze Naturali della Fondazione Rockefeller. Inizialmente Weaver aveva diffuso il suo testo solamente a 30 conoscenti, per chiedere un loro parere sulla possibilità di utilizzare il sistema digitale in evoluzione per tradurre dei testi da una lingua ad un’altra. Egli fu considerato il padre della Traduzione Automatica proprio perché fu il primo a considerare le potenzialità che i computer avrebbero potuto avere in questo campo e a porre degli obiettivi traduttivi prima che qualcun altro potesse averne idea. Questo fu, infatti, uno stimolo per la ricerca della disciplina prima negli Stati Uniti e poi, successivamente, nel resto del mondo. La prima volta in cui Weaver menzionò la possibilità di usare gli strumenti digitali per tradurre, fu in una lettera scritta al cibernetico Norbert Wiener nel marzo del 1947 e in una

² Avram Noam Chomsky (Philadelphia, 7 dicembre 1928) è considerato il fondatore della grammatica generativo-trasformativa, uno dei contributi più rilevanti della linguistica del XX secolo.

conversazione che ebbe, dopo, con Andrew Booth, un cristallografo specializzato in raggi-x. Grazie a loro e ai dipendenti della Rockefeller che lo spronarono, negli anni successivi, a dare voce e vita alle sue idee innovative, nacque il suo *memorandum* che scrisse nel luglio del 1949 a Carlsbad, nel New Mexico. Il suo compito alla Fondazione Rockefeller era quello di approvare progetti di ingegneria molecolare e genetica, oltre a quelli di agricoltura e di ricerca medica. Era un matematico e aveva un interesse particolare per la statistica, tant'è vero che durante la Seconda Guerra Mondiale si allontanò da quelli che erano i suoi incarichi nella Fondazione e diresse, invece, il lavoro di moltissimi matematici. Proprio grazie all'esperienza che ottenne, conosceva perfettamente lo sviluppo delle macchine calcolatrici elettroniche ed era quindi consapevole del successo che le tecniche matematiche avrebbero potuto avere sulla decodificazione di messaggi estranei. Il *memorandum* del 1949 era così strutturato: si apriva con una breve introduzione sugli obiettivi che si andavano ad affrontare, seguita da quattro parti in cui si spiegava passo dopo passo come poter effettivamente tradurre un testo con l'aiuto della tecnologia. Prima di tutto vennero riportati alcuni esperimenti fatti dai suoi colleghi Richard H. Richens e Andrew D. Booth in cui si andavano a tradurre i testi parola per parola. Weaver pubblicò questi esempi per sottolineare i grandi limiti che una traduzione letterale poteva porre: il suo rapporto si proponeva di suggerire metodi di analisi più proficui.

Per prima cosa, infatti, si concentrò sul problema del contesto di un testo, specificando che una parola, presa da sola senza considerare tutto l'insieme, potrebbe avere un significato errato rispetto ad uno ottenuto guardando la frase per intero:

If one examines the words in a book, one at a time through an opaque mask with a hole in it one word wide, then it is obviously impossible to determine, one at a time, the meaning of words. [...] But, if one lengthens the slit in the opaque mask, until one can see not only the central word in question but also say N words on either side, then, if N is large enough one can unambiguously decide the meaning of the central word.

(Weaver 1949: 21)

Il problema che si pose riguardò anche la quantità di contesto che era necessaria per la traduzione, ma poi arrivò alla conclusione che questo dipendesse da persona a persona

e concluse dicendo che “relativamente alcuni nomi, verbi e aggettivi” erano sufficienti per la comprensione.

La sua seconda proposta prevedeva invece di superare il problema della comprensione logica del linguaggio da parte di un calcolatore elettronico, e per fare ciò si servì del teorema di McCulloch e Pitts (1943), sviluppato nel contesto della ricerca sulla modellazione matematica della struttura neurale del cervello umano secondo cui:

"...a robot (or computer) constructed with regenerative loops of a certain formal character is capable of deducing any legitimate conclusion from a finite set of premises."

Da questo assunto, tuttavia, Weaver esclude la traduzione letteraria perché piena di elementi senza una logica formale come stili personali, espressioni, metafore, detti, giochi di parole che non potevano essere riassunte in regole formali.

Il terzo assunto riguardava la possibile applicabilità di metodi crittografici. Weaver era rimasto colpito dalla crittografia basata su, come disse lui, “frequenze di lettere, combinazioni di lettere, intervalli tra lettere e combinazioni di lettere, modelli di lettere, ecc. che sono in qualche modo indipendenti dalla lingua usata”. L’autore porta nel suo *memorandum* un esempio concreto di decifrazione di un testo turco durante la guerra. Il testo fu dato a un matematico che, senza conoscere la lingua originale, era riuscito a ricreare il testo base turco. Weaver ammise che la validità dell’approccio crittografico era difficile da valutare, ma attirava lo stesso la sua attenzione:

It is very tempting to say that a book written in Chinese is simply a book written in English which was coded into the "Chinese code." If we have useful methods for solving almost any cryptographic problem, may it not be that with proper interpretation we already have useful methods for translation? (Weaver 1949: 22)

Presto i ricercatori nel campo della traduzione automatica riconobbero l’errore nell’argomentazione di questo terzo punto che risiedeva nella confusione che l’autore fece tra le attività di decifrazione e di traduzione, spesso presenti quando è la stessa persona a svolgere entrambe le attività, come accade molte volte nella crittoanalisi.

Per la sua quarta proposta, Warren Weaver divenne più utopico. Si basò sulla convinzione che ci potessero essere elementi logici comuni a tutte le lingue che,

approfonditamente analizzati, potevano aiutare a individuare schemi e strutture linguistiche assolute utili a sviluppare una meccanizzazione della traduzione. Alla fine del *memorandum*, affermava questa sua convinzione tramite una delle metafore più note nella letteratura sulla traduzione automatica:

Think, by analogy, of individuals living in a series of tall closed towers, all erected over a common foundation. When they try to communicate with one another, they shout back and forth, each from his own closed tower. It is difficult to make the sound penetrate even the nearest towers, and communication proceeds very poorly indeed. But, when an individual goes down his tower, he finds himself in a great open basement, common to all the towers. Here he establishes easy and useful communication with the persons who have also descended from their towers. Thus it may be true that the way to translate from Chinese to Arabic, or from Russian to Portuguese, is not to attempt the direct route, shouting from tower to tower. Perhaps the way is to descend, from each language, down to the common base of human communication—the real but as yet undiscovered universal language—and—then re-emerge by whatever particular route is convenient. (Weaver 1949: 23)

Dopo la pubblicazione del *memorandum* ci fu una duplice reazione: da un lato molti rifiutarono l'idea che potesse mai essere possibile una meccanizzazione della traduzione, per la sua complessità e per le molteplici sfumature da cogliere; del resto, questo è il pensiero che ancora oggi molti traduttori professionisti hanno. Dall'altro lato ci furono invece persone che la pensarono in maniera più positiva e furono meno ostili verso gli obiettivi proposti dallo scienziato statunitense.

Una prima critica arrivò dal matematico, filosofo e linguista israeliano Yehoshua Bar-Hillel, che spense subito i sogni di Warren Weaver e di tutti gli scienziati che perseguivano il suo stesso obiettivo: secondo lui una traduzione affidata completamente ad una macchina, (FAHQ MT), non poteva esistere. A dimostrazione di ciò elaborò un saggio dal titolo: *“A Demonstration of the Non-Feasibility of Fully-Automatic High Quality Machine Translation”*. A sostegno della sua tesi, in questo elaborato propose la traduzione della parola inglese “pen” all'interno di una frase. Bar-Hillel pone il problema della polisemia, ovvero della vastità di significati possibili, sostenendo che un sistema, anche se ben addestrato, non poteva mai avere le capacità necessarie per poter individuare il significato della parola conforme al contesto della

frase. Egli credeva dunque che una traduzione automatica non potesse essere in alcun modo equiparabile a quella di un traduttore professionista, che fosse qualcosa di irrealizzabile e che la ricerca doveva orientarsi verso obiettivi più applicabili come per esempio la *mixed MT*, ovvero una traduzione sì fatta da un computer, ma con l'aiuto di una figura specializzata nel campo traduttivo che intervenisse o prima o subito dopo il processo automatico, per far sì che potesse arrivare ad una trasposizione ben curata. Nel 1952 organizzò la prima conferenza sulla traduzione automatica a cui parteciparono i principali espositori del campo che discussero sulla creazione di strumenti che potessero ridurre il divario tra la traduzione umana e quella automatica. Grazie ai finanziamenti ottenuti tramite l'intervento di Léon Dostert³ della Georgetown University che volle dimostrare la fattibilità della traduzione automatica, fu possibile creare dei gruppi di ricerca, sparsi tra l'Europa e gli Stati Uniti, che ottennero dei risultati concreti man mano che la tecnologia andava affermandosi nel mondo.

Il 7 gennaio del 1954, presso la Georgetown University di Washington, D.C., ci fu la dimostrazione di un primo esperimento, organizzato in collaborazione con l'IBM (*International Business Machines Corporation*), di traduzione automatica di un testo scritto. Questo spinse il governo statunitense a investire nella linguistica computazionale.

Questa iniziale euforia degli anni '50 lasciò presto il posto allo sconforto, a causa degli scarsi risultati ottenuti dalla ricerca. I progetti finanziati precedentemente furono aspramente criticati specialmente da coloro che ci avevano creduto per primi e che li avevano sostenuti economicamente. Lo stesso Bar-Hillel nel 1960 affermò che una traduzione automatica ben fatta non potesse avere alcuna speranza a quell'epoca e nemmeno in un futuro prossimo. Nel 1966 il governo degli Stati Uniti, chiese un rapporto all'*Automatic Language Processing Advisory Committee* (ALPAC) per studiare la realizzabilità della traduzione automatica. La qualità, i costi e le prospettive future analizzati non portarono a un buon risultato: il rapporto ALPAC affermava che non vi era l'utilità di investire nei sistemi di traduzione automatica poiché la qualità era carente, i costi erano molto elevati (quasi il doppio rispetto alle traduzioni umane) e i tempi erano eccessivamente lunghi. Piuttosto si propose di investire in sistemi che

³ Léon Dostert (1904-1971) studioso di lingue e sostenitore fondamentale della traduzione automatica.

potessero venire in aiuto dei traduttori specializzati e potessero accorciare i tempi di trasposizione. Era ovvio che questo rapporto finì di screditare ancora di più l'utilità che la traduzione automatica stava già perdendo e spinse molti finanziatori del settore a tagliare i fondi alle ricerche.

Molte comunità scientifiche non concordarono, però, con il rapporto negativo dell'ALPAC e così si formarono molti gruppi di ricerca specialmente in Canada e in Europa, dato anche il fatto che cresceva sempre di più la domanda di traduzioni dall'inglese al francese in Canada. Pertanto, nel 1970 venne creato il progetto TAUM (*Traduction Automatique de l'Université de Montréal*) che sviluppò un sistema chiamato "Météo", con lo scopo di tradurre i bollettini metereologici. Venne utilizzato con successo a partire dal 1976 e prevedeva un uso ristretto di parole semplici. Lo stesso tentativo di tradurre dei manuali aeronautici fallì e il progetto TAUM si concluse nel 1981. Una delle poche aziende che sopravvisse al rapporto ALPAC fu la SYSTRAN, *System Translation*, ideata da Peter Toma, ricercatore della Georgetown University, nel 1968. Fu una delle prime aziende che si occuparono di traduzione automatica e nello specifico durante la Guerra Fredda le furono affidate le trasposizioni di numerosi documenti di carattere scientifico, tecnico, amministrativo e legale che, seppur di qualità approssimativa, erano sufficienti a far capire il contenuto dei testi. Ma si capì subito che il sistema non era davvero capace di tradurre da e verso tutte le lingue europee e così la Comunità Europea spostò la sua attenzione sull'ambizioso progetto EUROTRA, che decise di finanziare dal 1982. Il programma si basava su una struttura ad interlingua per la traduzione di tutte le lingue europee. Nonostante il progetto avesse grandi potenzialità, l'azienda non riuscì mai a creare un vero e proprio sistema operativo e la Comunità Europea smise di finanziarlo nel 1992. Altri sistemi di traduzioni iniziarono a diffondersi in varie parti d'Europa come ad esempio il gruppo di Grenoble GETA in Francia, quello di Saarbrücken in Germania e la società BSO in Olanda. Anche fuori dal continente europeo, in Giappone, Corea, Taiwan, Cina, in Unione Sovietica e Malesia si diffusero altri gruppi di ricerca nel campo della MT. Lo sviluppo più interessante degli anni '80 fu la diffusione dei primi prodotti commerciali di traduzione automatica: SYSTRAN e LOGOS. Avendo goduto, negli anni precedenti, dei notevoli finanziamenti economici da parte degli Stati Uniti, i due sistemi erano adesso pronti a presentarsi sul mercato. Il programma SYSTRAN venne

adottato non solo dalla Comunità Europea, ma da anche da grandi società come la NATO, la General Motors, Dornier, Aérospatiale e Rank Xerox. Al tempo stesso la LOGOS Development Corporation, fondata nel 1969, ampliava il numero di clienti e lavorava su coppie linguistiche sempre più numerose, tra le quali anche il tedesco-italiano e l'inglese-italiano. Mentre SYSTRAN è ancora oggi presente sul mercato, la LOGOS, in seguito ad alcune difficoltà economiche, chiuse definitivamente nel 2000. Negli anni '80 la maggior attività di tipo commerciale riguardante la TA si svolse prevalentemente in Giappone, dove grandi società informatiche svilupparono sistemi di traduzione dal giapponese all'inglese e viceversa e per lo più riguardanti il settore informatico.

Il passo più importante compiuto in questi anni, fu quando la commercializzazione dei sistemi di traduzione assistita fu progettata per essere utilizzata sui computer e non solo tramite grandi calcolatori che, peraltro, avevano un costo molto elevato.

Negli anni '90 iniziarono a circolare le prime ricerche nell'ambito della traduzione automatica del parlato tramite nuove tecnologie per il riconoscimento vocale e la generazione del parlato. In questo periodo si affermò un approccio più realistico volto ad ascoltare maggiormente le necessità degli utenti: per prima cosa si abbandonò l'idea di poter creare uno strumento tecnologico in grado di tradurre come un vero e proprio essere umano, capace di cogliere anche le più sottili sfumature di significato. L'idea era quella di fornire al pubblico un aiuto concreto per il processo di traduzione e per questo nacquero sistemi di traduzione automatica, assistita, ambienti di lavoro per il traduttore, memorie di traduzione e sistemi online forniti su Internet. Gli sviluppi del settore erano, però, principalmente rivolti ai traduttori professionisti che in questo modo potevano accorciare i tempi lavorativi e facilitare il lavoro che, in alcuni casi, richiedeva una grande ricerca e uno studio approfondito degli argomenti trattati. La prima piattaforma creata per i traduttori fu la *Translator Workstation*. Il sistema fu descritto per la prima volta nel 1990. Per la sua progettazione si tenne in considerazione il fatto che i professionisti del settore avessero bisogno di strumenti veloci e facili da utilizzare, che fossero relativamente economici e, se possibile, anche portatili, dato che gran parte del loro lavoro veniva svolto a casa. Lo scopo era quello di fornire loro un accesso online ai dizionari e agli archivi di testi bilingue, dandogli al tempo stesso l'opportunità di aggiornare e personalizzare i dati del sistema con

materiale proprio. La Workstation era costituita da due componenti principali: un sistema di database lessicale bilingue e un sistema che crea e gestisce archivi di testo bilingue. Tutto era guidato da un menu per facilitarne l'utilizzo e il traduttore era guidato nell'uso di ogni componente da un ulteriore menu. Al professionista era consentito accedere ad informazioni quali esempi di utilizzo, equivalenti di traduzione, informazioni sulla sintassi ecc. e al tempo stesso poteva personalizzare il dizionario. Questo era particolarmente utile per un traduttore che lavorava in un solo settore specifico e che quindi, tramite la sua esperienza, accumulava informazioni su termini ed espressioni particolari che non erano presenti in nessun altro dizionario. Per quanto riguarda il processo traduttivo vero e proprio, ci si muoveva in questo modo: per prima cosa si caricava il testo di partenza sul sistema e si selezionava la lingua del testo di arrivo; a questo punto la workstation cercava le equivalenze tra le due lingue e, quando le trovava, creava un link con le parole o frasi corrispondenti nella lingua scelta. Quando, invece, il dizionario utilizzato dal sistema non riusciva ad individuare le equivalenze, era possibile che la parola presa in esame fosse un nome proprio oppure una parola di un vocabolario altamente specializzato. L'abbinamento procedeva ancora parola per parola fino ad arrivare alla fine del testo. Al termine del processo, il risultato veniva mostrato all'utente in termini di numero di "corrispondenze" trovate tra le due lingue. L'operazione si considerava fallita se il numero di equivalenza era inferiore a una data percentuale del testo totale di partenza. Una volta elaborata la traduzione, questa veniva registrata nel sistema che, così, poteva riutilizzarla per i prossimi lavori di trasposizione.

È evidente che dagli esordi della traduzione automatica negli anni '40 ad oggi sono stati compiuti tanti passi in avanti grazie alle capacità di calcolo dei computer, passando da sistemi sperimentali di bassa qualità e di costi elevati a sistemi di alta qualità disponibili anche sul web.

1.3 Rule-based, Statistical e Neural Machine Translation

La tecnologia sulla quale si basava la traduzione automatica degli anni '90 era un po' diversa rispetto a quella che è disponibile oggi: questa faceva ricorso principalmente a collezioni di traduzione già svolte in archivio, a dizionari e grammatiche che

permettevano di analizzare il testo di partenza e di creare il corrispettivo nella lingua di arrivo. I sistemi di *Machine Translation* di quei tempi utilizzavano quasi tutti il modello *Rule-based MT*, conosciuto anche come approccio a regole che operava tramite tre diverse strategie: la *Direct Translation* (traduzione diretta), la *Transfer-based Machine Translation* e la *Interlingua-based Machine Translation*. Una rappresentazione di questi tre approcci tramite i quali lavora la RBMT è quella del Triangolo di Vauquois (Figura 1.1), in cui si vede come, salendo verso il vertice, la profondità di analisi del testo cresce, mentre decresce la quantità di conoscenza richiesta.

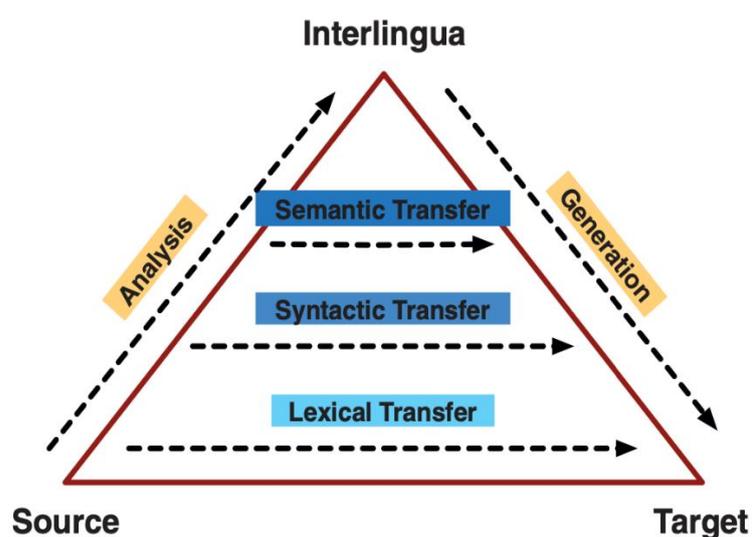


Figura 1. *Triangolo di Vauquois*

(Neural Machine Translation – P. Koehn, 2020, p.11)

Il metodo *Direct Translation* traduce il testo di partenza nella lingua di arrivo tramite tecniche molto semplici, tanto che viene definito sistema di “prima generazione”; è chiamato sistema diretto poiché la traduzione viene effettuata direttamente dalla lingua di partenza a quella di arrivo senza passaggi intermedi, come avviene per gli altri due approcci. Questa tipologia di sistema produce una traduzione parola per parola grazie all’uso di un dizionario bilingue e a un’analisi sintattica del testo iniziale. Il processo in questo caso si divide in tre fasi: la prima è un’analisi morfologica, in cui si vanno a individuare tutte le forme flesse presenti nel testo che vengono poi lemmatizzate, ovvero riportate alla loro forma base (infinito per i verbi, singolare per i nomi, ecc.). In secondo luogo, si procede alla fase di traduzione mediante l’uso di un dizionario

bilingue senza svolgere un'ulteriore analisi dal punto di vista semantico o sintattico. Si ricerca quindi la corrispondenza delle parole nella lingua di arrivo e si arriva così al terzo procedimento che consiste nella riorganizzazione di quest'ultime tramite semplici regole: per esempio nella lingua inglese si spostano gli aggettivi prima dei sostantivi. Questo era ritenuto un metodo molto approssimativo e soprattutto limitato, dato che si traduceva solo tramite corrispettivi tra le due lingue senza andare a svolgere un'ulteriore analisi di tipo semantico o sintattico. Nonostante ciò, sopravvisse non poco nel campo della MT grazie alla sua capacità di produrre risultati abbastanza comprensibili laddove, però, i due testi presi in considerazione avessero caratteristiche e strutture sintattiche simili. Ad esempio il sistema *Météo* rappresentò un caso di successo tra le traduzioni a modalità diretta, dimostrando di poter tradurre efficacemente i bollettini meteorologici dall'inglese al francese.

A differenza di quest'ultimo, gli approcci *Transfer-based MT* e *Interlingua-based MT* sono sistemi basati su traduzione indiretta ed erano definiti di "seconda generazione". La maggior parte dei sistemi di traduzione automatica risalente agli anni '80 utilizzava l'approccio a *transfer*, costituito anch'esso da tre fasi. Nella prima si svolgeva un'analisi della lingua di partenza in cui si arrivava alla sua forma naturale. Questa fase gettava le basi per il secondo passaggio, il *transfer*, in cui si creava una rappresentazione astratta intermedia corrispondente della lingua di arrivo. A questo punto c'era la cosiddetta generazione che convertiva nuovamente la rappresentazione astratta della lingua di arrivo nella sua forma naturale.

L'ultimo approccio della RBMT è l'*Interlingua-based MT*. I sistemi a interlingua sono per definizione sistemi multilingui, cioè nei quali si crea una rappresentazione astratta comune a tutte le lingue, l'interlingua, dopo aver svolto un obbligatorio lavoro di analisi lessicale, semantica e sintattica. Nonostante la rappresentazione sia neutra rispetto al testo originale, essa contiene tutte le informazioni necessarie a generare il testo nella lingua d'arrivo. L'idea era quella di creare un'interlingua realmente universale valida per tutte le combinazioni linguistiche e comune a tutte le lingue, essendo essa indipendente da quelle naturali.

Molti studiosi dell'epoca iniziarono a valutare le problematiche che presentava il modello *Rule-based Machine Translation* e capirono che questo non poteva costituire una valida alternativa alla traduzione umana. I punti deboli che portarono al suo

definitivo abbandono furono in primo luogo i grandi lavoro di post-editing che esso richiedeva, dovuti al numero limitato di regole sintattiche e grammaticali che potevano essere inserite all' interno del sistema. In secondo luogo l'ovvia incapacità di tradurre metafore, modi di dire, forme idiomatiche e sfumature di significato poiché allenata solamente a tradurre in base a regole grammaticali e, per ultimo, l'impossibilità di tradurre due lingue che avessero strutture completamente differenti tra loro, come ad esempio una lingua europea e una orientale. Per tutti questi motivi si abbandonò il tipo di approccio basato su regole in favore di un approccio di tipo statistico.

Il sistema di *Statistical Machine Translation* venne presentato per la prima volta dal matematico britannico Philip J. Brown durante la 12esima Conferenza Internazionale sulla Linguistica Computazionale di Budapest nel 1988. Dopo di lui, molti altri laboratori di ricerca accademici e commerciali hanno sviluppato questo approccio. Language Weaver è stata la prima azienda, fondata nel 2002, che abbracciava il sistema statistico, sviluppato, più avanti, da grandi leader del settore come IBM, Microsoft e Google. Questo nuovo metodo nacque come sostituto della *Rule-based Machine Translation*, ma venne presto criticato poiché sembrava assomigliargli molto. L'approccio statistico si basa su dati (*data-driven*) e tecniche di apprendimento automatico (*Machine Learning*) in cui non viene utilizzata alcun tipo di conoscenza linguistica esplicita. Invece di inserire all'interno del sistema grandi quantità di informazioni grammaticali, lessicali e sintattiche, processo molto complesso e costoso, si fornisce alla macchina un archivio di traduzione già svolte in entrambe le lingue prese in considerazione in modo tale che il computer, tramite un calcolo statistico, possa individuare quale parola o locuzione abbia la maggiore probabilità di essere quella corretta nella lingua di arrivo. La fonte principale di informazioni linguistiche è fornita da corpora che si dividono in corpora paralleli e corpora monolingui. I primi hanno la funzione di modello traduttivo e vengono utilizzati per trovare le traduzioni nella lingua di arrivo più frequenti per una parola o per un insieme di parole. I corpora monolingui servono per misurare quanto è probabile che una specifica parola o sequenza di parole, la loro struttura grammaticale e il loro ordine all'interno di una frase corrispondano realmente all'uso che se ne fa nella lingua di arrivo. Proprio perché si basa su tecniche di calcolo statistiche alla ricerca delle corrispondenze tra lingue, il sistema prende il nome di traduzione automatica statistica.

La SMT opera tramite tre diversi tipi di approcci statistici: il primo basato sulla singola parola e chiamato *Word-based SMT*, il secondo basato su una sequenza di più parole, denominato *Phrase-based SMT* e l'ultimo che si serve di modelli sintattici, detto *Syntax-based SMT*. Nel primo modello la traduzione si attua tramite una corrispondenza parola-per-parola tra le due lingue usando la traduzione a livello lessicale. Ogni parola della lingua di partenza ha un suo corrispettivo nella lingua di arrivo e per la scelta di queste, il calcolo si basa su un ampio *parallel corpus*. Nel secondo modello il sistema suddivide il testo in base a frasi intere, cercando di superare alcuni limiti del *Word-based SMT*. Infine nell'ultimo, si cerca di individuare la struttura sintattica delle frasi, per migliorare la traduzione tra due lingue con un diverso ordine di parole come le lingue europee ed extra-europee.

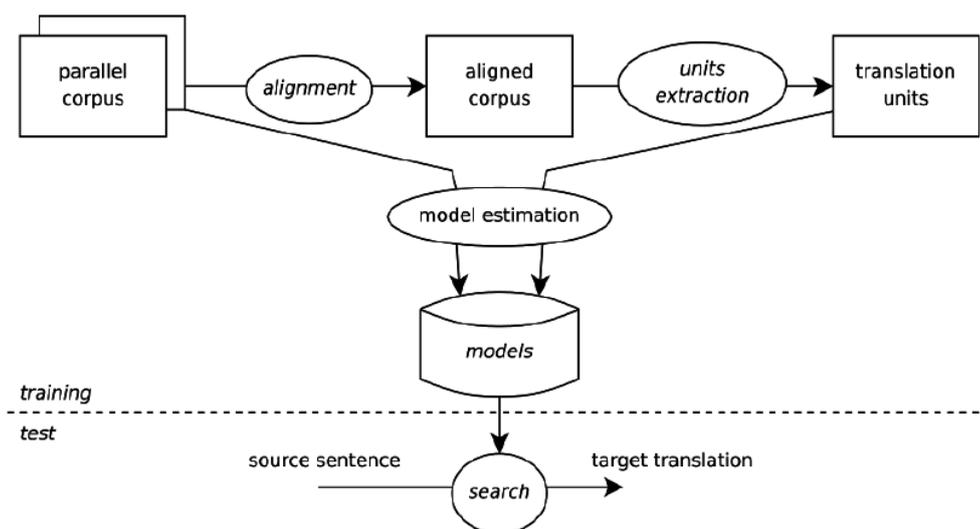


Figura 2. *Processo di funzionamento della Statistical Machine Translation*
 (Methods for Leveraging Lexical Information in SMT – K. Singla, 2015, p.4)

Nonostante questo modello statistico abbia ottenuto dei risultati migliori rispetto a quelli di prima generazione, presentava comunque delle limitazioni: per prima cosa, si rivelò efficace solo per le combinazioni linguistiche di cui erano presenti svariate traduzioni all'interno dei corpus paralleli. In secondo luogo, presentava spesso errori o omissioni nelle traduzioni, specialmente tra lingue che avevano strutture sintattiche

diverse. Il sistema, poi, falliva anche nella concordanza di genere e numero delle parole, essendo incapace di cogliere i rapporti di dipendenza tra le proposizioni. Anche questo secondo tipo di approccio sembrava, quindi, non soddisfare i requisiti richiesti per la creazione di una traduzione automatica ben fatta e fedele all'originale. A partire dal 2010, la *Statistical Machine Translation* venne rapidamente abbandonata a causa di nuovi sviluppi nel campo del *Deep Learning* che crearono un nuovo approccio alla traduzione artificiale, questa volta di tipo neurale: la *Neural Machine Translation*, introdotta per la prima volta nel 2014. Questa si basa sul principio di voler riprodurre l'attività neuronale biologica di un essere umano tramite neuroni artificiali. Nello stesso modo in cui i nostri neuroni trasmettono le informazioni per l'elaborazione di concetti complessi al nostro cervello, così i neuroni artificiali vengono stimolati a comunicare le informazioni che apprendono da un *input*, grazie a una rete, per produrre un *output*. Le reti neurali artificiali sono costituite da gruppi di neuroni chiamati con il termine inglese *layer* e le trasmettono ai neuroni a cui sono collegati sui livelli successivi per arrivare al risultato finale (*output*).

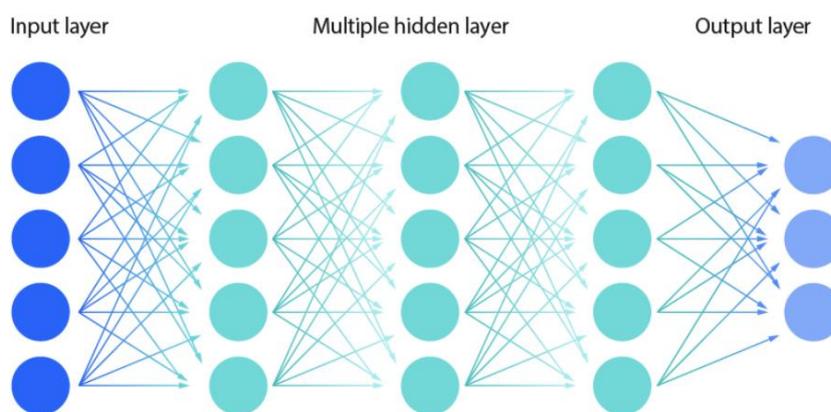


Figura 3. Sistema di rete neurale

(IBM - <https://www.ibm.com/topics/neural-networks>)

Ogni neurone artificiale rappresenta un nodo della rete e ogni connessione della rete (rappresentato dalle frecce nella figura 3.), equiparabile alle sinapsi umane, prende un input e lo moltiplica per un peso che un nodo esercita su quello successivo fino a creare l'output.

Il processo di funzionamento della *Neural Machine Translation* si articola in tre fasi:

- TRAINING. Questa fase, detta anche allenamento o *deep learning*, consente al sistema di assimilare un ampio corpus di traduzioni bilingui e di accumulare informazioni rilevanti sulle lingue di partenza e di arrivo. È come se la macchina svolgesse un lavoro di allenamento per le traduzioni che dovrà affrontare. La NMT analizza i testi di riferimento di entrambe le lingue e li scompone in unità come singole parole o elementi grammaticali. Questo processo, chiamato *word embedding*, determina le connessioni tra diverse unità, collocando concetti simili vicini nello spazio multidimensionale.
- ENCODER. In questa fase di codifica, si replica l'operazione di *word embedding* appresa nella fase precedente. Viene creata una rappresentazione vettoriale che permette di far apprendere alla macchina le combinazioni linguistiche e semantiche tra gli elementi del testo di partenza. Tramite la rappresentazione, la macchina trasmette gli stimoli da un neurone all'altro.
- DECODER. Dopo aver ottenuto l'immagine vettoriale, il sistema interpreta i valori delle singole unità nella lingua di partenza e li converte in un output nella lingua di arrivo. L'obiettivo è quello di produrre un risultato che sia quanto più vicino al "*gold standard*", ovvero una traduzione di riferimento creata da un traduttore professionista.

Una particolarità delle reti neurali è il modo in cui sono in grado di apprendere dagli errori, correggendo i parametri durante la fase di apprendimento. Questa tipologia di traduzione automatica consente di tradurre in poco tempo parti di testo molto ampie ed è, infatti, indicata per la gestione di traduzioni urgenti.

Per quanto innovativa, la traduzione automatica neurale presenta alcuni limiti come per esempio la traduzione non precisa di testi molto tecnici e specifici oppure di testi creativi, essendo il meccanismo addestrato a tradurre su modelli di traduzione lineari. Per arrivare a un risultato impeccabile è necessario l'intervento di un professionista umano; la NMT rimane, comunque, un sistema in continua evoluzione.

Questo approccio fu comunque adottato da tutti i principali attori nel campo della traduzione automatica, come ad esempio Google Translate, Microsoft Translator e DeepL, di cui si parlerà più nello specifico nel capitolo successivo.

CAPITOLO DUE

I LIMITI DELLA TRADUZIONE AUTOMATICA

2.1 I principali sistemi di traduzione automatica su Internet

Il continuo progresso della tecnologia e la rapida digitalizzazione delle informazioni hanno trasformato radicalmente i modi di comunicare. Oggi è possibile trovare strumenti come libri o giornali anche in forma digitale (senza che sia stata abbandonata la versione originale cartacea); la comunicazione avviene tramite l'utilizzo di telefoni di ultima generazione, Internet, televisione via satellite. Persino la comunicazione orale ha subito una grande trasformazione con l'introduzione di videochiamate e videoconferenze grazie ai nuovi sviluppi tecnologici, sempre più numerosi. L'elemento chiave che accomuna tutti questi cambiamenti è l'avvento di Internet. In una società sempre più globalizzata, è importante e necessario che vengano abbattute tutte le barriere linguistiche che ci impediscono di comunicare in tempo reale. Per questo la tecnologia, e in particolare Internet, possono darci una mano. Questi strumenti giocano un ruolo cruciale nel facilitare la comunicazione tra varie parti del mondo nella società contemporanea grazie ai nuovi sistemi di traduzione automatica sempre più diffusi, disponibili online e gratuitamente. In passato questi servizi venivano usati da un pubblico ristretto per scopi lavorativi, come aziende multinazionali o enti governativi che avevano bisogno di un gran numero di traduzioni. Tuttavia, l'accessibilità che oggi hanno questi sistemi ha contribuito a renderli conosciuti a un grande pubblico. Le conseguenze di questo fenomeno sono, in primis, che c'è stato un aumento significativo della domanda di traduzione di testi, come e-mail e pagine web. Ad esempio, quando un utente ha la necessità di comprendere in tempo reale il contenuto di un testo in lingua straniera, gli basterà inserire il testo sul portale di traduzione automatica e dopo alcuni secondi il sistema fornirà il significato del testo. I servizi di traduzioni online si sono diffusi su Internet a partire dagli anni '90, ma da allora i progressi sono stati tanti. Le attuali tecnologie rappresentano indubbiamente strumenti molto importanti nel mercato della traduzione, rendendo il lavoro più semplice e rapido. Oggi i principali sistemi di traduzione automatica che troviamo su Internet sono basati per lo più su sistemi neurali, che rappresentano la nuova frontiera tecnologica e di cui si è parlato nel capitolo precedente. Il loro

obiettivo è infatti quello di ricreare il lavoro della mente umana nel processo di traduzione utilizzando neuroni artificiali. I servizi di *Machine Translation* sono programmi sviluppati per computer che traducono senza l'intervento umano, alcuni lo fanno parola per parola, altri, più sofisticati, sono in grado di interpretare frasi o periodi più lunghi in maniera più accurata.

Un impulso significativo alla diffusione dei servizi di traduzione automatica online è stato dato dalla collaborazione tra Microsoft Internet e AltaVista, il famoso primo motore di ricerca che nel 1997 ha introdotto il primo servizio di traduzione automatica online in tempo reale, noto come Babel Fish. Il nome deriva da un pesce immaginario, il Pesce di Babele, di cui parla lo scrittore britannico Douglas Adams che lo descrive così:

«Il Babelfish è piccolo, giallo, ricorda una sanguisuga ed è forse la cosa più strana dell'universo. Si ciba dell'energia delle onde cerebrali, assorbendo frequenze inconscie ed espellendo una matrice di frequenze conscie verso i centri cerebrali del linguaggio. La conseguenza pratica è che se ve ne ficcate uno nell'orecchio, comprenderete istantaneamente qualunque cosa, in qualunque lingua.»⁴

Il suo nome deriva a sua volta dalla biblica Torre di Babele, durante la cui costruzione, secondo il racconto della Bibbia, gli uomini parlavano tutti la stessa lingua.

Il traduttore era capace di tradurre 12 lingue: cinese, olandese, francese, tedesco, greco, italiano, giapponese, coreano, portoghese, russo e spagnolo tutte da o verso l'inglese, anche se alcune coppie di lingue non erano traducibili tra di loro. Le traduzioni che erano possibili con la lingua italiana erano le seguenti:

- dall'italiano all'inglese
- dall'inglese all'italiano
- dall'italiano al francese
- dal francese all'italiano

Gli utenti potevano inserire un testo di massimo 150 parole e ricevere una traduzione istantanea, oppure potevano inserire l'indirizzo del sito web che volevano fosse

⁴ D. Adams, (1979). *Guida galattica per gli autostoppisti*. Mondadori.

tradotto ed essere indirizzati a una versione già tradotta del sito in questione. Babel Fish registrò un notevole successo, con un numero di richieste giornaliere di traduzione automatica che si aggirava intorno ai 500.000 nel maggio del 1998 fino ad arrivare a 1.3 milioni nel 2000. Nel 2008, il sistema venne acquisito da Yahoo e, dopo alti e bassi, fu definitivamente chiuso un anno dopo. Tuttavia, Babel Fish ha aperto la strada a una formula che si è rivelata il trampolino di lancio per la traduzione automatica, stimolando la diffusione di altri servizi gratuiti online a partire dal 1997. Nel 2003, erano già presenti su Internet circa 55 servizi di traduzione automatica, di cui 25 supportavano la lingua italiana.

2.1.1 Google Translate, Microsoft Translator e DeepL

Il traduttore online più diffuso e utilizzato al mondo è Google Translate che registra quasi 500 milioni di accessi al giorno. Google Traduttore, in italiano, è stato sviluppato da Google LLC, l'azienda statunitense che offre servizi online, e fu lanciato il 28 aprile del 2006. Originariamente, si basava sul sistema di traduzione automatica con approccio statistico (Capitolo 1) e utilizzava le trascrizioni delle Nazioni Unite e del Parlamento Europeo per raccogliere i suoi dati linguistici su cui basare la sua traduzione. Nel 2010, Google introdusse anche l'app per Android e la versione iOS a febbraio del 2011, per fornire una sorta di traduttore portatile per chi ne avesse bisogno in tempo reale. Il sistema arriva a tradurre 103 lingue diverse; l'ultimo lancio del traduttore avvenne nel dicembre del 2016 con l'introduzione del programma vocale in khmer e singalese. A novembre del 2016, Google annuncia l'abbandono dell'approccio statistico della traduzione in favore di quello neurale, introducendo così il sistema di "Neural Machine Translation" che, utilizzando tecniche di Deep Learning, prometteva di essere più preciso e affidabile di quello precedente. Google Traduttore può tradurre forme di testo, ma anche contenuti multimediali come immagini, video o testo parlato. Nello specifico è specializzato in:

- Traduzione di siti web. Google Translate è in grado di esaminare parole o frasi singolarmente, semplificando così la traduzione completa di siti web. La traduzione restituisce un'interpretazione accurata del contenuto del sito,

consentendo l'utilizzo dei menu come se ci si trovasse su una pagina italiana o della lingua richiesta. Una volta aperto il sito nella lingua straniera di origine, è sufficiente utilizzare il browser Google Chrome e cliccare sull'opzione "Traduci", scegliendo la lingua stabilita. È uno dei pochi sistemi di traduzione che presenta questa funzione anche su dispositivi mobili e quindi su app;

- Traduzione di documenti. Con Google Traduttore è possibile tradurre interi documenti che siano in pdf, word o generici semplicemente caricando dai documenti il file che si desidera tradurre;
- Creazione di un frasario personale. Capita spesso che si vogliano salvare frasi o parole precedentemente tradotte per poterle avere sempre a portata di mano e accedervi facilmente, senza doverle digitare di nuovo sul traduttore. Per questo Google Traduttore permette di salvare delle traduzioni già svolte precedentemente che vengono "conservate" nel proprio account, a cui si può accedere rapidamente;
- Ascolto della pronuncia delle parole. Questa funzione può essere utile per coloro che desiderano imparare la pronuncia di una lingua o semplicemente per togliersi dei dubbi riguardo una parola o una frase più complicata. Non bisogna far altro che inserire la parola nel campo del testo e premere sull'icona a forma di altoparlante;
- Traduzione di email. Google Traduttore è integrato anche nel servizio di posta elettronica Gmail. Quando ci si trova davanti ad una mail con un testo in lingua straniera, è possibile tradurlo tramite l'opzione che appare nel messaggio;
- Traduzione di chat e messaggi. È possibile tradurre messaggi o chat da app come WhatsApp, Telegram o Facebook ricevuti da una persona straniera che scrive in un'altra lingua semplicemente utilizzando il tasto "Traduci" dopo aver selezionato il testo. Questa operazione è disponibile solo con sistema Android;
- Utilizzo della funzione "Interprete". Il sistema permette di avere un interprete vocale e un traduttore istantaneo a portata di mano e riesce a mettere in contatto in tempo reale persone che parlano lingue diverse. Bisogna che si scelga la lingua di partenza e poi, dopo aver premuto il tasto con il microfono, pronunciare la frase che vogliamo venga tradotta. Google Traduttore fornirà

una traduzione parlata da far ascoltare alla persona con cui si vuole comunicare;

- Traduzione di immagini. Ci sono due modi per utilizzare questa funzione: scattare una foto sul momento all'immagine che ci interessa oppure importarla dalla nostra galleria. Dopo aver schiacciato sull'icona della fotocamera, il sistema tradurrà facilmente l'immagine lasciando il risultato corrispondente al colore e al carattere del testo originale;
- Traduzione della scrittura a mano. Oltre al testo scritto e alla voce, il sistema di traduzione automatica è anche capace di interpretare la scrittura a mano, premendo l'icona della penna e scrivendo con il polpastrello il testo che vogliamo tradurre.

In base alle varie statistiche, si è riuscito a stimare che le lingue più tradotte, e quindi più parlate, con Google Traduttore sono l'inglese, lo spagnolo, l'arabo, il russo, il portoghese e l'indonesiano, mentre il Paese in cui il sistema di traduzione automatica viene utilizzato di più è il Brasile. Seppur sia ancora impreciso, Google Translate è uno dei sistemi di traduzione automatica che si è evoluto di più nel corso degli anni e che sta lavorando ancora per migliorarsi ulteriormente.

Spostiamo adesso l'attenzione su un altro servizio di Machine Translation molto diffuso e utilizzato in tutto il mondo: Microsoft Translator, un servizio cloud multilingue che fa parte dei servizi di Microsoft Cognitive. La prima versione del sistema fu sviluppata tra il 1999 e il 2000 all'interno del Microsoft Research. Il traduttore si basa sull'approccio "data driven", ovvero invece che affidarsi a regole grammaticali per tradurre il testo di partenza, si basa su testi paralleli già tradotti e in questo modo viene addestrato a tradurre nuovi testi in un linguaggio naturale. Il sito, noto agli utenti anche con il nome di Bing Translator, è stato lanciato nel 2007 e fornisce traduzioni gratuite su Internet. Sono disponibili tre versioni del sistema: Microsoft Translator per computer, Microsoft Translator a cui si accede tramite il browser del telefono e l'app, sia per sistema operativo Android che per iOS. Tra il 2011 e il 2018, ha sviluppato numerose funzioni che andremo adesso a vedere nello specifico, molto simili a quelle offerte da Google Translate. Il sistema è in grado di:

- Tradurre testi tramite la digitazione. Basterà inserire il messaggio di testo nella lingua straniera e selezionare la lingua di arrivo per ottenere la traduzione desiderata;
- Tradurre immagini tramite la fotocamera. Cliccando sull'icona della fotocamera, il sistema permette di tradurre istantaneamente cartelli, insegne, menu, ecc.;
- Digitare il testo tramite voce. Questa opzione serve a sostituire la digitazione del testo con la pronuncia orale delle parole o frasi che vogliamo vengano tradotte;
- Far ascoltare la pronuncia delle parole;
- Far conversare persone che non parlano la stessa lingua. Microsoft Translator ha introdotto questa funzione innovativa che molti altri traduttori non possiedono ancora. Il sistema è in grado di replicare una reale conversazione tra due persone che parlano lingue diverse. Questa opzione si attiva tenendo premuta l'icona con due persone che parlano. Basterà scansionare con il proprio dispositivo il codice QR che appare sul telefono della persona con cui si vuole conversare e il dispositivo metterà in contatto i due interlocutori permettendogli di comunicare, nonostante non parlino la stessa lingua e a volte queste siano completamente diverse tra di loro, come le lingue europee e quelle extra-europee;
- Favorire l'apprendimento delle lingue tramite il frasario. Il sistema di traduzione automatica dispone, nella barra del menu, del cosiddetto frasario. Una volta aperta questa opzione, appariranno una serie di macro argomenti come viaggi, ristoranti, tecnologia, salute, emergenza, nei quali sono raggruppate espressioni o parole tipiche dell'argomento scelto, che non solo si possono rivelare utili in casi di emergenza, ma possono aiutare gli utenti nel caso vogliano imparare una nuova lingua partendo dalle basi e dalle espressioni più ricorrenti nella vita quotidiana;
- Personalizzare le impostazioni. In questa sezione si possono modificare parametri come: la pronuncia delle traduzioni scritte, che si può attivare o meno; il genere della sintesi vocale, ovvero se sarà la voce di un uomo o una donna a pronunciare le frasi; rallentare la velocità di riproduzione dell'audio,

utile per capire bene quello che il traduttore formula; aggiornare le traduzioni in modalità offline, nel caso non ci sia connessione a Internet.

Oggi Microsoft Translator è capace di tradurre più di 100 lingue e si propone l'obiettivo di ampliare ancora di più la selezione di lingue straniere. Per quanto riguarda la traduzione, come Google Translate anche Microsoft Translator è ancora impreciso su alcuni aspetti, soprattutto per quanto riguarda testi particolarmente complessi a livello grammaticale e semantico, ma il sistema sta cercando di rendere il livello di traduzioni sempre più alto, tramite il supporto dell'intelligenza artificiale.

L'ultimo servizio di traduzione automatica presente su Internet che si prenderà come oggetto di studio in questo capitolo è il sistema di traduzione automatica di DeepL. DeepL fa parte di un'azienda tedesca che ha sviluppato il sistema di traduzione nel 2016 e lo ha lanciato nel 2017. Il suo fondatore e amministratore delegato è il polacco Jaroslaw Kutylowski, residente oggi in Germania, che detiene un dottorato in informatica. Andiamo a vedere nel dettaglio come nasce e come si è evoluto negli anni questo sistema di Machine Translation.

Per la sua creazione nel 2016, furono utilizzati i dati accumulati dal sistema di Linguee⁵, basati su una serie di reti neurali già esistenti. Nel 2017 veniva lanciato al pubblico e offriva traduzioni tra le lingue inglese, tedesco, francese, spagnolo, italiano, polacco e olandese. Nel marzo del 2018 al servizio online gratuito si affianca anche una versione a pagamento, chiamata DeepL Pro, che utilizza il sistema API per la traduzione. API, acronimo di Application Programming Interface, è un software che permette a due applicazioni di comunicare tra di loro e permette dunque di integrare dati, applicazioni e dispositivi per facilitare la comunicazione e la collaborazione, ad esempio all'interno di un'azienda. Sempre nel 2018, a dicembre, al sistema di traduzione si aggiungono due nuove lingue: il russo e il portoghese. A settembre 2019 DeepL lancia un'applicazione per Windows e MacOS che permette a chi la usa di tradurre da qualsiasi applicazione sul proprio computer. Nel febbraio del 2020 il sistema fa ulteriori progressi nella ricerca delle reti neurali, andando a perfezionare ancora di più il livello di traduzioni. Ancora nel 2020, a marzo, alle lingue già presenti

⁵ Linguee è un servizio web che presenta un dizionario online bilingue con esempi nella lingua di origine e circa un miliardo di traduzioni redatte da esseri umani.

si aggiungono il cinese e il giapponese. Nell'aprile del 2021 DeepL ha proposto ai suoi utenti la creazione di un'app per iPhone e iPad che permette di consultare il sito direttamente dal cellulare. L'ultima innovazione introdotta dal sistema il 25 gennaio del 2024 è l'aggiunta della lingua araba, la quinta lingua più parlata al mondo e la prima lingua supportata da DeepL che si scrive e si legge da destra a sinistra. Il sistema di origine tedesca offre diversi piani agli utenti. Il primo è quello base, gratuito e accessibile a tutti da smartphone o computer. Quello a pagamento prende invece il nome di DeepL Pro e a sua volta si divide in tre piani di abbonamento: il livello Starter, quello Advanced e l'Ultimate, rispettivamente con un costo di 7,49 €, 24,99 € e 49,99 € al mese.

Vediamo adesso quali sono le principali funzioni che DeepL mette a disposizione, simili a quelle degli altri due sistemi di traduzione automatica, ma con qualche funzionalità in più. DeepL Translate permette agli utenti di:

- Tradurre testi o parole da una lingua ad un'altra digitandoli nell'apposito spazio;
- Ascoltare la pronuncia delle parole;
- Selezionare parole o espressioni alternative che permettono di risparmiare tempo. Cliccando sulla parola desiderata si aprirà un elenco di traduzioni alternative tra le quali è possibile scegliere se si vuole utilizzare un sinonimo o se non si è soddisfatti della prima traduzione fornita;
- Andare oltre il limite di traduzione di 5000 caratteri del traduttore gratuito con i piani offerti da DeepL Pro;
- Tradurre interi file. Si tratta di una novità rispetto agli altri traduttori automatici. Con questa funzione si possono tradurre tutti i documenti PowerPoint e Word mantenendo la loro formattazione originale.
- Scegliere il tono giusto per il proprio testo. Tramite i piani di abbonamento di DeepL Pro, si può scegliere di tradurre il testo in un tono formale o informale a seconda della situazione richiesta. Il sistema aggiornerà automaticamente i pronomi e le parole circostanti. Questa funzionalità è prevista per le lingue francese, italiano, olandese, polacco, portoghese, portoghese brasiliano, russo, spagnolo e tedesco.

- Creare un glossario. In questo modo è possibile risparmiare tempo nella traduzione impostando delle regole per tradurre termini specifici sempre allo stesso modo;
- Fare clic su una parola per scoprirne il significato. Gli utenti possono cercare il significato di una parola sia nel testo di partenza che in quello di arrivo. Questa funzione fornisce anche esempi, sinonimi e informazioni generali che altrimenti dovrebbero essere ricercate separatamente;
- Integrare DeepL nei CAT Tools. Questa è una soluzione completa, utile ai traduttori professionisti, accessibile con i piani Advanced e Ultimate della versione Pro. I traduttori possono integrare la tecnologia di DeepL nei loro software di traduzione assistita per pre-tradurre automaticamente i file, estendere le memorie di traduzione e mantenere la formattazione del testo originale. Alcuni dei CAT Tools supportati sono per esempio Trados Studio, memoQ e Across (di cui si parlerà più avanti).

Il sistema di traduzione automatica DeepL può tradurre in circa 31 lingue diverse. Il sito vanta più di 10 milioni di utenti attivi ogni mese, 500mila dei quali sono iscritti ai piani a pagamento della versione Pro. La maggior parte del business di DeepL è dovuto alle 20mila aziende clienti che utilizzano il software per le loro traduzioni; tra queste citiamo: Mercedes Benz, Fujitsu e la società ferroviaria tedesca Deutsche Bahn. Secondo la piattaforma Pitchbook⁶, nel gennaio 2023 la startup tedesca ha raccolto circa 100 milioni di dollari di finanziamenti da società di venture capital globali.

Completata la panoramica dei tre siti di traduzione automatica presi in considerazione, andiamo a vedere le principali differenze tra i tre e come si comportano nella traduzione.

Per quanto riguarda le funzionalità, Google Translate e Microsoft Translator condividono approssimativamente le stesse funzioni, mentre DeepL propone più soluzioni, seppur a pagamento. Si può dire che quest'ultimo risulta più utile per i

⁶ PitchBook è un fornitore di dati che offre informazioni sul mercato azionario, sulle acquisizioni delle fusioni, sulle società private, sul capitale privato e sul capitale di rischio e ha sede negli Stati Uniti d'America.

professionisti del settore, traduttori e interpreti, dal momento che mette a disposizione anche l'utilizzo dei CAT tools combinato con il sistema di traduzione automatica.

Un articolo pubblicato dal DDay⁷ nel maggio del 2018 mette alla prova i tre sistemi, facendo loro tradurre tre testi di diversi argomenti, con una difficoltà crescente. Il primo paragrafo preso come oggetto di prova riguarda una notizia giornalistica tradotta dall'inglese all'italiano. In questo caso il traduttore che rende meglio la notizia giornalistica è Google Translate, seguito da DeepL e per ultimo da Microsoft Translator, che fa un errore abbastanza importante: nel testo viene citato il famoso servizio di spesa online "Amazon Prime Now" che, essendo un nome proprio, va trascritto nella stessa maniera, lasciandolo quindi nella sua forma originale. Il sistema di traduzione automatica di Microsoft lo traduce come "*il primo servizio di Amazon ora*", una traduzione non accettabile che cambia il senso della frase e che fa classificare il sito all'ultimo posto in classifica.

Come seconda traduzione si prende in oggetto una notizia scientifica riguardante lo studio di un nuovo farmaco per l'emicrania. In questo caso tutte e tre le traduzioni fornite dai sistemi automatici sono ben fatte, ma anche questa volta il sito elegge come vincitore Google Traduttore. Le versioni di Microsoft e DeepL sono comunque accettabili, ma hanno bisogno di qualche piccolo ritocco.

Passando all'ultimo caso, in cui il paragrafo scelto fa riferimento ad una sentenza giuridica, le differenze sono notevoli. L'ambito giuridico è di per sé un campo abbastanza complicato da tradurre, che richiede più tempo e ricerca anche per i professionisti del settore e che, ancora oggi, è difficile da interpretare nonostante i miglioramenti dell'intelligenza artificiale. In tutti i tre casi di traduzione c'è, infatti, bisogno di un intervento correttivo umano. Per la terza volta, Google Translate è in testa a DeepL e Microsoft che si classificano rispettivamente secondo e terzo. Google e DeepL dimostrano comunque di conoscere il lessico giuridico, con traduzioni aderenti alla sentenza scelta. La versione di Microsoft Translator, seppur corretta, richiede un intervento maggiore a livello grammaticale e sintattico.

C'è da dire che questo test venne eseguito nel 2018, quando i sistemi erano un po' più arretrati e soprattutto quando DeepL, creato solo un anno prima, nel 2017, era ancora agli inizi della sua "carriera" e in fase di sperimentazione. Come abbiamo visto, infatti,

⁷ <https://www.dday.it/redazione/26760/google-translate-microsoft-deepl-quale-il-migliore-traduttore>

nei paragrafi precedenti riguardanti la sua storia, molte delle sue innovazioni e funzionalità vennero lanciate a partire dal 2018. In ogni caso, questo articolo fornisce un primo spunto di riflessione sulla valutazione in campo dei vari sistemi di traduzione automatica presenti ad oggi su Internet, messi a confronto e alla prova tra loro.

2.1.2 I CAT Tools

Altri strumenti di traduzione assistita sono i CAT Tools. Il termine deriva dall'acronimo inglese *Computer Assisted Translation* o *Computer Aided Translation* (Traduzione Assistita dal Computer) e sta ad indicare tutti quegli strumenti informatici utilizzati dai traduttori come ausilio al processo di traduzione. I programmi informatici CAT si basano sull'utilizzo di memorie di traduzione, che conservano elementi già utilizzati nelle traduzioni precedenti e che indirizzano verso un processo di traduzione più veloce e di qualità. Sul mercato se ne possono trovare molti e diversificati, alcuni dei quali sono usati prevalentemente in alcune aree geografiche per gestire contenuti particolari, come sottotitoli o software. In linea generale, questi sono software installati sui computer che consentono di suddividere un testo in singole frasi. Successivamente, una volta tradotte, le frasi vengono immagazzinate in una memoria di traduzione, la quale rappresenta una collezione di tutte le traduzioni già realizzate, corrette e revisionate dai professionisti del settore. Tali strumenti, inoltre, consentono di creare un glossario terminologico che include tutti i termini specifici utilizzati in uno o più testi, insieme alla loro traduzione relativa nel contesto specifico. In commercio esistono vari tipi di CAT Tools, ognuno con funzionalità diverse. Tra i più comuni e utili ai traduttori ci sono:

- Software di memoria di traduzione. In questo caso il testo viene diviso in parti chiamate "segmenti" che memorizzano le traduzioni apportate e le suggeriscono quando si presenta una frase simile;
- Software di allineamento: si carica il testo originale di partenza e quello tradotto e il programma provvede a suddividerli in paragrafi, registrare come vengono tradotti e creare un database in cui memorizzarli;

- Programma per la gestione della terminologia che consente di ricercare la traduzione di un termine specifico nel database e di aggiungere nuove coppie di parole tradotte, per poi scegliere la traduzione corretta in base al contesto tra varie alternative.

Nel 2018 l'Associazione Italiana Traduttori e Interpreti (AITI) ha svolto un sondaggio⁸ aperto a tutti i Traduttori e Interpreti che lavorano con l'italiano come lingua di partenza o di arrivo. Più precisamente, il sondaggio si è svolto tra il 20 gennaio 2018 e il 19 marzo 2018. Questo fa seguito ad altri due sondaggi analoghi svolti già nel 2013 e nel 2007. L'indagine è stata messa a punto dalla Commissione dei Traduttori Tecnico-Scientifici (CTTS) di AITI con la collaborazione della Commissione Interpreti (CI), della Commissione Formazione e Aggiornamento (CFA) e di tutto il Consiglio Direttivo Nazionale (CDN) dell'Associazione. Il tasso di adesione ha registrato un numero di partecipanti pari a 543, rispetto ai 445 traduttori e interpreti che avevano partecipato nel 2013. Le 543 interviste sono state raccolte con metodologia CAWI (Computer Assisted Web Interviewing).

In una parte dell'indagine sono state rivolte alcune domande ai professionisti del settore per quanto riguarda l'uso dei CAT Tools. La prima domanda posta è se gli intervistati lavorino o meno con i sistemi di traduzione assistita CAT.

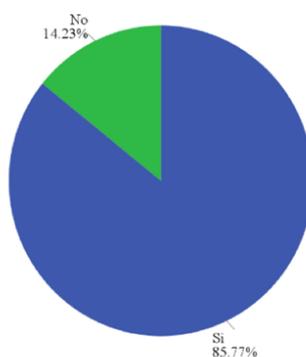


Figura 4. *Percentuale di Traduttori e Interpreti che lavorano con sistemi CAT*
(Indagine sul mercato dei Traduttori e degli Interpreti 2018 – AITI, p.60)

⁸ <https://aiti.org/it/indagine-sul-mercato-dei-traduttori-e-degli-interpreti-2018>

Il grafico sopra riportato indica le risposte alla domanda. La maggioranza dei rispondenti (85,77%) dichiara di lavorare con sistemi CAT, mentre la minoranza (14,23%) non li utilizza.

Sul mercato è disponibile una grande varietà di sistemi CAT Tools. Sono presenti servizi di traduzione assistita a pagamento e gratuiti. Quelli a pagamento più famosi e utilizzati sono i seguenti:

- SDL Trados Studio, forse il più usato. Si tratta di una soluzione completa per traduzione, gestione della terminologia, editing e assicurazione della qualità linguistica. È disponibile in diverse versioni, con piani annuali o permanenti, per traduttori freelance;
- memoQ. Anche se meno conosciuto del precedente, rappresenta un altro CAT Tool di rilievo. Offre molte funzionalità, come il pannello di anteprima di traduzione per visualizzare il contesto delle sezioni tradotte. memoQ è compatibile con Trados e altri software di traduzione e offre una versione gratuita per uso personale o progetti di piccole dimensioni, oltre a una versione professionale con licenza permanente;
- Wordfast è una versione integrata di Microsoft Word per la traduzione temporanea di documenti bilingue. Supporta vari formati di file utilizzabili da Word. Esiste una soluzione più completa con gestione terminologica e memoria di traduzione, chiamata Wordfast Pro;
- Déjà Vu, rinomato tra i vari CAT Tools. Offre funzionalità avanzate per la gestione della terminologia e della memoria di traduzione, con un'interfaccia altamente personalizzabile e supporto per una vasta gamma di formati di file, inclusi Trados e memoQ.
- Across. Include: memoria di traduzione, gestione della terminologia, assicurazione della qualità e gestione dei progetti. Offre una versione base gratuita per uso online e una versione premium con un piano di abbonamento mensile;
- XTM. Questo sistema è disponibile in due diverse versioni: una licenza di software tradizionale chiamata XTM Suite o una soluzione con abbonamento online chiamata XTM Cloud;

- Star Transit. Questo CAT Tool si differenzia dagli altri poiché trattiene nella sua memoria di traduzione anche la struttura complessiva del documento originale. Per questo motivo, non ci sono sezioni isolate dal contesto, consentendo di visualizzare e controllare il contesto originale di una determinata sezione in qualsiasi momento se ne abbia la necessità. Questo garantisce una soluzione comprensibile senza sezioni fuori contesto. Star Transit offre opzioni di abbonamento mensili o annuali;
- Poedit, traduce con le interfacce e può essere integrato per lavorare su Wordpress;
- MetaTaxis. Nonostante sia un sistema a pagamento, vanta un prezzo notevolmente più basso rispetto ai suoi competitori, anche per la sua versione avanzata;
- Pairapharase, sistema basato sul web che combina la traduzione automatica con quella umana. Presenta, inoltre, una funzionalità di *voice over*. È disponibile una versione gratuita, ma con funzioni limitate, e varie versioni con abbonamenti mensili.

I CAT Tools invece gratuiti presenti sul mercato sono:

- CafeTran Espresso, un programma utilizzabile sui sistemi operativi Windows, Mac o Linux. È inoltre compatibile con altri importanti programmi di traduzione. Questo sistema è gratuito fino a quando le memorie di traduzione o i glossari non arrivano alla loro capacità massima;
- OmegaT, forse il CAT gratuito più conosciuto. È un programma *open-source* ed è compatibile con sistemi operativi Windows, Mac e Linux;
- Similis, meno conosciuto, supporta file HTML, XML, PDF, MS Word, OpenOffice e Trados;
- Wordfast Anywhere. Si utilizza online e fornisce connessioni a database di traduzione pubblici;
- MateCat. Con l'accesso ai database di traduzione automatica che offre agli utenti, è in grado di migliorare la velocità di lavoro;
- SmartCAT, che fornisce una piattaforma d'incontro tra traduttori e clienti;

- Virtaal, utilizza il linguaggio di programmazione Python e comprende strumenti di gestione della traduzione e l'accesso a traduzioni automatiche;
- Pootle. A differenza della maggior parte dei CAT Tools, è specificamente disegnato per la traduzione di applicazioni piuttosto che di documenti;
- Lokalize, orientato alla produttività e alla garanzia di qualità;
- GlobalSight, che semplifica e gestisce il processo di localizzazione. Può essere usato da *project manager*, traduttori ed editori;
- gtranslator, un programma specializzato per la traduzione di software.

Seguendo i dati del sondaggio svolto dall'Associazione Italiana Traduttori e Interpreti (AITI) nel 2018, si vanno ad analizzare quali sono i CAT Tools maggiormente utilizzati, dopo averne presentato un elenco esaustivo di quelli presenti sul mercato, a pagamento e non. Nell'immagine riportata di seguito sono presenti quasi tutti i sistemi di traduzione assistita sopracitati.

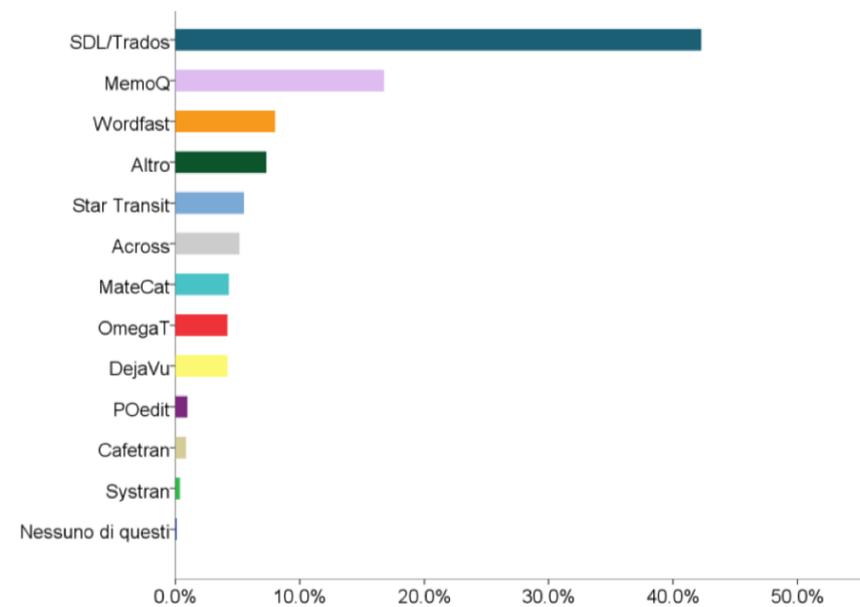


Figura 5. *CAT Tools maggiormente utilizzati da Traduttori e Interpreti*
(Indagine sul mercato dei Traduttori e degli Interpreti 2018 – AITI, p. 61)

Questa era una domanda a risposta multipla. Gli intervistati che hanno risposto erano 506. Tra i vari CAT indicati, al primo posto figura il sistema SDL Trados con il 42,3%

delle risposte, seguito al secondo posto da MemoQ con il 16,8%. Poi, nell'ordine, Wordfast, Star Transit, Across, MateCat, Déjà Vu, OmegaT, Poedit, Cafetran, Systran e altri software.

I vantaggi delle memorie di traduzione e degli strumenti CAT utilizzati da traduttori e interpreti professionisti sono molteplici. Ad esempio, si riduce il rischio di lasciare parti non tradotte poiché i CAT Tools non consentono la presenza di segmenti con il testo di arrivo assente. Si garantisce inoltre, l'uniformità terminologica e stilistica del testo, sia interna che esterna. Questo è un aspetto essenziale perché risulta di grande importanza che un testo venga tradotto con una precisione terminologica non indifferente per la comprensione e la fruizione di un prodotto. Un altro vantaggio è che si aumenta la produttività, ovvero il numero di parole tradotte in un determinato intervallo di tempo, poiché i segmenti identici, chiamati anche "100% match", vengono tradotti una sola volta, mentre quelli simili, chiamati "fuzzy match" vengono tradotti parzialmente dal programma. In questo modo il lavoro risulta più rapido, proprio per il fatto che si recupera ciò che era già stato tradotto in precedenza riducendo i tempi di traduzione. In aggiunta a questi aspetti, i CAT Tools permettono di valutare e individuare ripetizioni e concordanze, esportare la traduzione nello stesso formato dell'originale, calcolare rapidamente il numero di cartelle di traduzione, utilizzare strumenti di traduzione automatica integrati nei CAT, garantire la precisione nella traduzione di brevetti.

Il sondaggio del 2018 pone agli intervistati ulteriori domande riguardo ai CAT Tools: viene chiesto a traduttori e interpreti professionisti se l'utilizzo di questi ultimi abbia aumentato o meno la loro produttività.

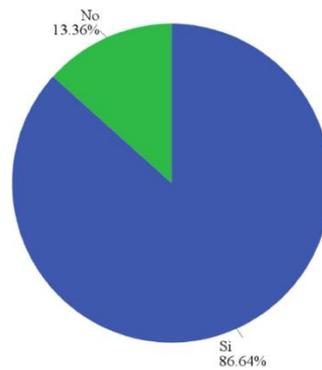


Figura 6. *Percentuale di Traduttori che hanno trovato vantaggi nell'uso dei CAT*
 (Indagine sul mercato dei Traduttori e degli Interpreti 2018 – AITI, p. 64)

Su 434 rispondenti, l'86,64% dichiara che l'utilizzo dei CAT ha effettivamente permesso di aumentare la produttività, mentre il 13,36% non ha trovato nessun miglioramento.

Viene loro domandato di quanto sia aumentata la produttività nel lavoro con l'utilizzo dei sistemi CAT. Su un totale di 339 rispondenti, il 42,18% segnala un aumento della produttività compreso tra il 21% e il 40%, il 25,96% un aumento della produttività compreso tra il 41% e il 60%, il 20,06% un aumento della produttività fino al 20%. Il rimanente 11,80% dichiara un aumento della produttività di oltre il 60%. Sembra, quindi, essere uno strumento utile alla produttività dei traduttori.

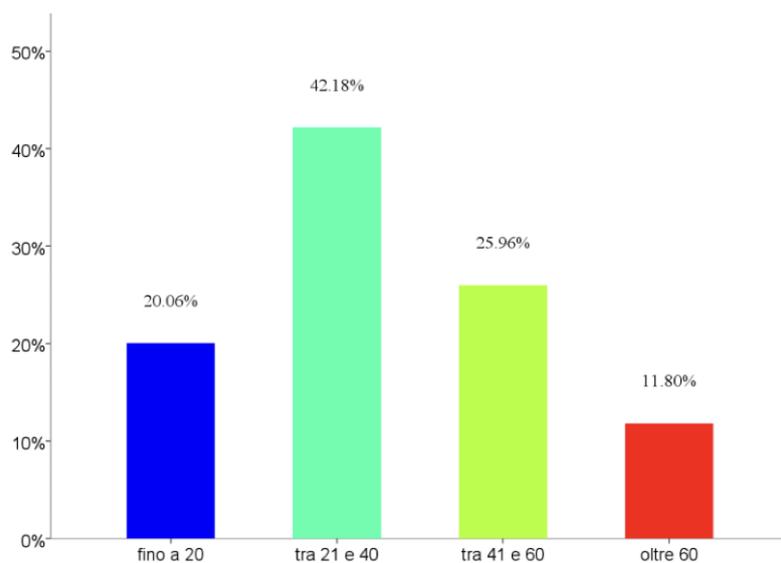


Figura 7. *Percentuale di aumento della produttività grazie all'uso di CAT Tools*
(Indagine sul mercato dei Traduttori e degli Interpreti 2018 – AITI, p.64)

È però importante tener presente che quando si fa riferimento agli strumenti CAT di traduzione assistita non si sta parlando di sistemi di traduzione automatica, di cui si è parlato in precedenza. Tra questi due vi è infatti una differenza, che spesso non viene riconosciuta: mentre gli strumenti di traduzioni assistita CAT automatizzano i processi di traduzione utilizzando dei database e suggerendo frammenti di traduzioni precedenti, fornendo, così, solo un aiuto ai professionisti del settore, la traduzione automatica affida completamente a una macchina o ai suoi algoritmi il controllo e il risultato finale della traduzione; i risultati, in questo caso, sono meno precisi poiché non vi è alcun intervento umano che possa supervisionare o correggere il testo tradotto. Difatti, se ci si volesse affidare completamente a questi ultimi in un processo traduttivo, sarebbe indispensabile un lungo processo di post-editing, di cui si parlerà più avanti, da parte di un traduttore professionista.

2.1.3 Dispositivi tecnologici per l'interpretariato

Fino ad ora si è parlato dei sistemi di traduzione automatica o di traduzione assistita che possono rivelarsi utili nel campo della traduzione scritta. È sicuramente più facile che le macchine e i sistemi dotati di intelligenza artificiale traducano testi o frasi scritte,

mentre sembra più difficile immaginare gli stessi sistemi che vengono adottati dai professionisti del settore nel campo della traduzione orale e quindi dell'interpretariato. L'interpretariato è l'atto di convertire il parlato da una lingua di partenza ad un'altra di arrivo in tempo reale. Nella pratica dell'interpretariato si distinguono due diversi tipi: quello simultaneo, in cui la trasposizione nella lingua di arrivo avviene contemporaneamente a quando l'interlocutore pronuncia le parole nella lingua di origine, e quello consecutivo, nel quale l'interprete parla dopo che l'interlocutore ha completato una frase, magari aiutandosi con la presa di appunti tramite simboli. Anche in questo caso, la tecnologia ha fatto del suo meglio per proporre strumenti di traduzione orale sempre a portata di mano. Vediamo quali sono i principali, in grado di catturare le frasi pronunciate in una determinata lingua e restituirle velocemente in un altro idioma:

- Auricolari con traduttore simultaneo. Si tratta di un sistema sempre pronto all'uso che consente di tradurre in tempo reale ciò che viene detto da altre persone. L'apparecchio comprende due cuffie wireless e può tradurre varie lingue con i relativi accenti. Tra le caratteristiche di questi strumenti vi sono l'estrema semplicità di utilizzo, poiché basta semplicemente installare un'app su smartphone per configurarli, una batteria di lunga durata e varie modalità di funzionamento tra cui scegliere, in base alle proprie esigenze e comodità;
- App di traduttori simultanei gratis. Tra questi vi è ad esempio quello di Google Traduttore che consente di dettare frasi al telefono o al tablet ascoltandone la traduzione in pochi istanti o l'app "Traduci" di Apple, in cui basta cliccare sulla scheda della Conversazione per tradurre una conversazione tra persone che parlano lingue diverse in tempo reale. L'ultima app citata è iTranslate Voice, specifica per le traduzioni simultanee vocali. Offre sia una versione gratuita sia una a pagamento che permette di accedere a traduzioni illimitate e di eliminare la pubblicità;
- Traduttori simultanei per computer. Se si preferisce usare il computer oppure se si lavora nel campo della traduzione con questo dispositivo, potrebbero essere utili alcuni siti come ad esempio QTranslate, gratis, che supportano tutte

le principali lingue del mondo e sfruttano i migliori servizi di traduzione automatica disponibili online;

- Traduttori simultanei online. Questi vengono utilizzati se si vuole avere un'alternativa alle app da scaricare. Tra questi ci sono Google Translate, Bing Translator, WordReference, ImTranslator, alcuni dei quali già citati e analizzati precedentemente nel paragrafo 2.1.1.
- Traduttori simultanei per conferenze. Esistono alcune piattaforme in grado di integrare un traduttore simultaneo da utilizzare all'occorrenza. Tra queste c'è, per esempio, Skype che permette di fare tutto ciò sfruttando la funzione "Conversazioni tradotte", da attivare nella scheda Chat premendo sulla voce "Invia richiesta di traduzione" (che dovrà poi essere accettata dal contatto a cui viene inviata).

In questo paragrafo sono stati presentati tutti i sistemi che sfruttano l'intelligenza artificiale nel campo della traduzione, scritta e orale. Tutti disponibili su Internet, alcuni gratuiti altri a pagamento e utilizzati sia in conversazioni quotidiane se si ha la necessità di comunicare, sia in ambienti lavorativi per svolgere traduzioni professionali.

2.2 Pre-editing, post-editing e linguaggio controllato

L'avanzamento della traduzione automatica e la sua integrazione con altre tecnologie, come le memorie di traduzione e i CAT Tools, hanno portato a una crescente adozione di questi processi. Secondo uno studio condotto dal TAUS Machine Translation Market Report nel 2017⁹, i profitti del mercato della traduzione automatica erano stimati intorno ai 130 milioni di dollari e sono destinati ad aumentare sempre di più. L'utilizzo della traduzione automatica può dare un importante contributo nell'accorciare i tempi lavorativi, ma spesso la qualità della traduzione viene sacrificata. Per evitare ciò vengono in soccorso tre diversi processi: il pre-editing, il post-editing e il linguaggio controllato. Difatti, di pari passo con la crescita della

⁹ <https://www.taus.net/think-tank/reports/translate-reports/taus-machine-translation-market-report-2017>

traduzione automatica e dei sistemi che si rifanno ad essa, è aumentata anche la domanda dei servizi di pre-editing e post-editing. Nonostante la traduzione automatica sia il più delle volte accettabile, è indispensabile l'intervento umano per un risultato di buona qualità. A testimonianza del suo uso sempre più frequente, nel 2017 è stato emesso lo standard internazionale *ISO 18587:2017 (en) Translation services – Post-editing of machine translation output – Requirement*¹⁰, che ufficializza il post-editing come un servizio di traduzione a sé stante, stabilendo i suoi requisiti e delineando le competenze necessarie per i post-editor. Il processo di traduzione basato sulla traduzione automatica richiede un'organizzazione diversa rispetto ai tradizionali approcci umani, insieme a specifiche competenze tecnologiche e procedurali da parte dei traduttori. L'intervento umano avviene prima della traduzione automatica con la fase del pre-editing, oppure successivamente alla traduzione nella lingua di arrivo, durante la fase di post-editing. Questo tipo di organizzazione del processo traduttivo basato sulla traduzione automatica è stato evidente sin dalle prime fasi della ricerca in questo settore. Già Bar-Hillel nel 1951 aveva sottolineato che per i destinatari che richiedevano un'alta precisione traduttiva, la traduzione automatica pura doveva essere sostituita con una sorta di approccio misto che coinvolgesse l'intervento umano. La questione principale era quali parti di questo processo dovessero essere affidate e gestite dalla parte umana. Il primo a introdurre i concetti di pre-editor e post-editor è stato probabilmente Erwin Reifler (1950). Secondo quest'ultimo, il pre-editor, che possiede una buona conoscenza degli aspetti linguistici e culturali del testo di partenza, si doveva occupare della preparazione del testo per il processo di traduzione automatica, rimuovendone qualsiasi ambiguità o segnalando le parti più difficili da tradurre automaticamente. Il post-editor invece, secondo Reifler non doveva necessariamente conoscere la lingua di partenza, ma si occupava solo di sistemare e di risolvere eventuali problemi presenti nel testo di arrivo.

Di seguito si vanno ad analizzare i due processi di revisione, facendo anche un breve excursus sull'uso dei linguaggi controllati, altro sistema utile per la successiva traduzione affidate alle macchine.

Il pre-editing è un processo di revisione che mira a ottenere un testo iniziale privo di ambiguità e complessità che possano rendere difficile l'elaborazione da parte dei

¹⁰ <https://www.iso.org/obp/ui/#iso:std:iso:18587:ed-1:v1:en>

motori di traduzione automatica. L'obiettivo principale è intervenire prima del processo automatico per ridurre i tempi di revisione del testo finale. Questa fase è monolingue e si concentra esclusivamente sul testo iniziale, senza la necessità di conoscere la lingua di destinazione. È essenziale che il pre-editor sia consapevole delle difficoltà previste che potrebbero incontrare i traduttori automatici e che influirebbero, quindi, negativamente sul risultato finale. In questo modo possono essere eliminate o rielaborate per ottenere una traduzione grezza facilmente rivisitabile.

Le modifiche generalmente includono la correzione di errori ortografici e grammaticali, la chiarificazione di parti ambigue o complesse attraverso esplicitazioni, semplificazioni o normalizzazioni. Un'altra forma di pre-editing coinvolge l'uso dei cosiddetti “linguaggi controllati” o “semplificati”, sviluppati principalmente per applicazioni industriali al fine di migliorare la comprensione dei documenti tecnici e specialistici. Questi linguaggi, derivati da lingue naturali come l'inglese o il tedesco, mirano a rendere la documentazione tecnica facilmente comprensibile e a ridurre le possibilità di interpretazioni errate da parte dei lettori.

Alcuni esempi di linguaggi controllati includono Caterpillar-Fundamental English, Siemens Technical German, Multinational customized English e Perkins Approved Clear English. Questi linguaggi sono stati sviluppati per migliorare la comprensione del personale tecnico, spesso con competenze linguistiche limitate.

Poche sono le lingue sviluppate specificamente per la traduzione automatica, come il Caterpillar Technical English o il Controlled Language Optimized for Uniform Translation (CLOUT). Quest'ultimo si basa su dieci regole fondamentali¹¹:

1. Scrivere frasi più brevi di 25 parole.
2. Esprimere solo un'idea per frase.
3. Ripetere la stessa frase per esprimere lo stesso concetto.
4. Scrivere frasi grammaticalmente complete.
5. Utilizzare una struttura grammaticale semplice.
6. Utilizzare la forma attiva delle frasi.
7. Ripetere il nome invece di usare un pronome.
8. Utilizzare articoli per identificare i nomi.

¹¹ https://works.bepress.com/uwe_muegge/88/download/

9. Utilizzare parole da un dizionario generale.
10. Utilizzare solo parole con corretta ortografia.

I benefici derivanti dall'utilizzo dei linguaggi controllati nella traduzione sono i seguenti: consentono la redazione di testi facilmente elaborabili dai sistemi automatici e migliorano la qualità del risultato finale in termini di comprensibilità, rendendo più efficiente la fase di post-editing. Tuttavia, uno svantaggio significativo è rappresentato dal fatto che gli autori sono molto limitati nella redazione dei testi e impiegano molto tempo per apprendere le regole necessarie.

In generale, la ricerca sul pre-editing e sull'uso dei linguaggi controllati ha ricevuto meno attenzione rispetto agli studi sul post-editing. Questo potrebbe essere dovuto al fatto che il pre-editing è spesso considerato una fase opzionale, talvolta persino anti-economica, e la sua efficacia nel ridurre i tempi e i costi complessivi per produrre traduzioni pubblicabili non è stata chiaramente dimostrata da esperimenti su larga scala. Solo di recente alcuni studi hanno evidenziato che il pre-editing accelera il post-editing e che combinare i due processi può portare vantaggi significativi. Inoltre, è interessante notare che i lavori dedicati alle fasi preliminari di preparazione dei testi per la traduzione automatica sono principalmente legati agli approcci linguistici basati su regole, nei quali è possibile individuare problematiche ricorrenti e intervenire sui testi di input. Con l'avvento dei sistemi statistici, questo aspetto ha ricevuto meno attenzione da parte degli studiosi. Tuttavia, sono stati proposti diversi approcci e strategie per il pre-editing al fine di migliorare la leggibilità, la correttezza e la coerenza dei testi. Alcuni contributi relativi all'uso dei linguaggi controllati sono stati sviluppati anche in relazione agli approcci statistici.

Alcuni studiosi si sono concentrati sull'implementazione di regole linguistiche efficaci per risolvere le difficoltà dei testi di partenza. Altri hanno affrontato la questione della traducibilità dei testi, focalizzandosi su aspetti linguistici e culturali che possono influenzare la traduzione automatica. Inoltre, il pre-editing è stato studiato in relazione alla produzione di documenti di qualità, come la documentazione tecnica, e recentemente sono emersi studi che esaminano gli effetti degli interventi di normalizzazione su contenuti generati dagli utenti. Il pre-editing sembra, quindi, avere meno rilevanza negli studi sulla traduzione.

Il post-editing, invece, ha trovato molto più spazio nel campo della revisione della traduzione: si tratta di un'operazione di revisione e adattamento dei testi tradotti automaticamente. Nel tempo, questa pratica è stata descritta come "la correzione dell'output della traduzione automatica da parte di linguisti/redattori umani" secondo Veale & Way nel 1997, "il controllo, la correzione e la revisione delle traduzioni effettuate da qualsiasi tipo di automazione traduttiva" secondo Gouadec (2007), e infine "la correzione dell'output della traduzione generata dalla macchina per garantire che raggiunga un livello di qualità concordato preventivamente tra cliente e post-editor"¹² (Rico & Torreòn 2012). Quest'ultima definizione è forse la più esaustiva, poiché non solo specifica la natura dell'attività, cioè la correzione della traduzione automatica, ma evidenzia anche un aspetto cruciale, ovvero che la qualità finale prodotta è il risultato di un accordo preventivo tra cliente post-editor. Si distinguono due livelli di post-editing: il *light post-editing*, detto anche *fast post-editing*, *gist post-editing*, *rapid post-editing* e il *full post-editing*, detto anche *conventional post-editing*. Il post-editing si compone di tre fasi: la lettura del testo originale, l'analisi del testo tradotto automaticamente per individuare le parti da correggere e la revisione per produrre il testo finale, apportando le modifiche necessarie per adattarlo agli scopi specifici.

Il *light post-editing* implica una revisione minimale del testo tradotto automaticamente, concentrandosi sulle correzioni essenziali: queste correzioni riguardano gli errori che compromettono il senso del testo e, di conseguenza, la sua corretta interpretazione da parte del destinatario. Il risultato finale deve quindi essere un testo comprensibile, ottenuto con il minimo sforzo di revisione, che corregga solo gli errori lessicali, grammaticali e sintattici che impediscono la trasmissione accurata delle informazioni contenute nel testo originale. Gli errori grammaticali e stilistici che non influenzano la comprensibilità del testo finale devono essere ignorati. Questo tipo di post-editing è spesso richiesto per ottenere documenti tradotti per uso interno (in inglese, "inbound translations") in tempi rapidi e con aspettative di qualità finale basse. Dall'altro lato, il *full post-editing* è un intervento di revisione più approfondito sul testo tradotto automaticamente, il cui risultato finale è un testo pubblicabile (in inglese "outbound translation"), teoricamente indistinguibile da quelli tradotti da traduttori

¹² (TAUS/CNGL 2011 cit. in Rico & Torreòn 2012: 168)

professionisti. Questo processo è più lento rispetto al *light post-editing* e richiede correzioni mirate per garantire il massimo della precisione nel trasferimento del messaggio dal testo originale al testo tradotto e la massima fruibilità del testo finale. Nonostante i tempi di revisione siano molto più lunghi rispetto al post-editing leggero, il cliente si aspetta comunque tempi rapidi di realizzazione, ma con elevate aspettative di qualità finale dei testi revisionati.

L'associazione Translation Automation User Society (TAUS), che mira a condividere conoscenze, risorse e metriche nell'industria linguistica, ha elaborato nel 2016 linee guida sul post-editing. Queste linee guida affrontano temi cruciali per le organizzazioni che intendono adottare processi automatici e necessitano di identificare, formare e valutare le prestazioni dei post-editor. Vengono forniti anche esempi di interventi da effettuare per il light e il full post-editing. Per il light post-editing, le linee guida raccomandano una traduzione semanticamente corretta, senza aggiungere o omettere informazioni e cercando di mantenere intatta la traduzione grezza il più possibile. Per il full post-editing, invece, il risultato finale è una traduzione semanticamente, sintatticamente e grammaticalmente corretta, che trasmette in modo completo e accurato le informazioni del testo originale, cercando di utilizzare quanto più possibile la traduzione grezza.

Sorge il dubbio se il post-editing abbia delle differenze rispetto alla revisione di una traduzione umana: si tratta di un'attività diversa, principalmente a causa dei tipi di errori da correggere (Loffer-Laurian 1985: 71). Ad esempio, McElhaney & Vasconcellos (1988) affermano che, data la natura automatica della traduzione, il post-editor può essere più sicuro che nessuna parte del testo sia stata trascurata e che non ci siano errori di ortografia, mentre le costruzioni errate sono presenti sia nelle traduzioni grezze che in quelle umane. Geoffrey Koby (2001: 7) sottolinea anche le differenze nella frequenza, ripetitività e tipologia degli errori. “Mentre un traduttore umano può commettere un errore di traduzione una sola volta [...], il sistema di traduzione automatica continuerà a commettere lo stesso errore ripetutamente”. Al di là di queste variazioni nel risultato finale, la valutazione degli errori e della qualità rappresenta una fase cruciale per determinare se sia conveniente utilizzare un sistema di traduzione automatica oppure se sia preferibile optare per una traduzione da zero.

Le attività e la tipologia di post-editing dipendono da una serie di fattori, alcuni dei quali non facilmente confrontabili con la traduzione tradizionale: la durata d'uso della traduzione, la tipologia, la lunghezza e la "visibilità del documento", i tempi di consegna, le necessità e le aspettative degli utenti finali, nonché la capacità dei lettori di utilizzare testi non perfetti. Il tempo d'uso della traduzione e la capacità di accettare traduzioni non precise sono fattori chiave nella scelta tra *light* e *full* post-editing. Ad esempio, se la traduzione è destinata solo a una comprensione generale con un breve tempo d'utilizzo e se la traduzione grezza non è sufficiente a questo scopo, il *light post-editing* può rappresentare una soluzione valida. Tuttavia, è importante che il lettore non si aspetti un'eccellenza qualitativa e sia consapevole dei limiti della revisione rapida del testo.

La ricerca sul post-editing è estremamente vasta e presenta un numero significativo di contributi rispetto al pre-editing. Uno dei primi studi condotti da Orr & Small (1967) ha confrontato il grado di comprensibilità di traduzioni umane, traduzioni automatiche e traduzioni automatiche sottoposte a post-editing. Il test di comprensione ha coinvolto studenti specializzati nei domini dei testi scientifici e ha evidenziato poche differenze nella comprensione tra traduzioni umane e quelle automatiche revisionate, mentre ha mostrato differenze significative tra traduzioni umane e quelle automatiche non revisionate. Negli anni '90, l'industria della traduzione ha iniziato a interessarsi al ruolo professionale del post-editor, con una serie di contributi che si sono intensificati dagli anni 2000 in poi. In questi anni, l'interesse dei Translation Studies¹³ per le implicazioni delle tecnologie nei processi di traduzione, in particolare per il post-editing, è cresciuto, poiché i traduttori tendono a partire da testi tradotti automaticamente. Questo ha portato alla focalizzazione della ricerca sugli aspetti umani del post-editing, con un'attenzione particolare alla pratica del post-editing nell'industria della traduzione specialistica.

L'indagine svolta nel 2018 da AITI (Associazione Italiana Traduttori e Interpreti) ha chiesto ai suoi intervistati se facessero o meno uso della pratica del post-editing.

¹³ I Translation Studies (Traduttologia e studi traduttivi) sono un'area accademica di studi interdisciplinari che indaga sulla teoria, descrizione e utilizzo della traduzione, interpretazione e localizzazione.

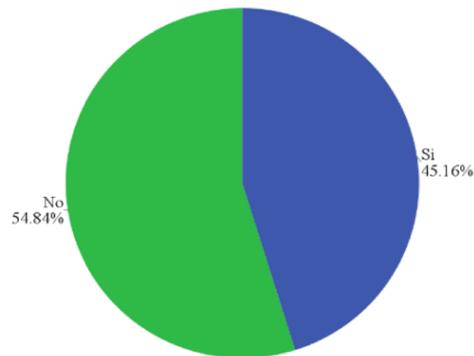


Figura 8. *Grafico della risposta alla domanda: "Hai svolto lavori di post-editing?"*
(Indagine sul mercato dei Traduttori e degli Interpreti 2018 – AITI, p.83)

Il 45,16% dei 496 rispondenti, ovvero 224, conferma di aver già utilizzato questo tipo di servizio di revisione. La maggioranza invece, con una percentuale del 54,84%, afferma di non aver ancora fatto uso del post-editing nel lavoro di traduzione.

2.3 Le metriche di valutazione qualitativa della traduzione automatica: BLEU, DQE, MQM

Nonostante i recenti progressi, la traduzione automatica continua a incontrare limitazioni nel convertire il testo di partenza a livello della "manifestazione lineare" (Eco 2003), poiché non riesce a superare gli ostacoli delle lingue e dalle asimmetrie culturali e storiche. Sempre più spesso la traduzione automatica e quella assistita vengono integrate nel processo di traduzione; tuttavia, la valutazione della sua qualità rimane una delle principali preoccupazioni per gli sviluppatori e per gli utenti.

La valutazione della qualità della traduzione rappresenta quindi uno dei temi centrali per l'industria delle lingue. La valutazione della traduzione automatica è ancora più complessa rispetto a quella umana e questo rappresenta una sfida che verrà analizzata nel corso di questo capitolo.

Uno dei problemi principali è la mancanza di accordo su quali parametri considerare nella valutazione delle traduzioni prodotte tramite tecnologie di traduzione, conosciute anche come traduzioni grezze: è considerata una delle tappe più cruciali per misurare l'efficacia delle tecnologie nel loro utilizzo. Prima di esaminare lo stato attuale della

ricerca in questo settore, è necessario fare un breve excursus storico, che illustra le metodologie adottate nel tempo per valutare la qualità della traduzione.

Il primo tentativo di valutare la traduzione automatica risale al famoso rapporto ALPAC del 1966, intitolato "Languages and machines: computers in translation and linguistics", commissionato dall'Accademia Nazionale delle Scienze americana per esaminare la fattibilità della traduzione automatica. All'epoca non esistevano metodi affidabili per valutare la qualità, quindi la commissione si basò su un esperimento condotto da John B. Carroll che confrontava traduzioni automatiche con traduzioni umane. Questo primo esperimento di valutazione era basato su aspettative molto alte, che mirava a ottenere una traduzione automatica di alta qualità che fosse paragonabile alla traduzione umana. Tuttavia, le conclusioni del rapporto ALPAC rivelarono che la qualità delle traduzioni prodotte dai sistemi di traduzione automatica era così scadente da non giustificare il livello di supporto finanziario erogato in tutti quegli anni, decretando di fatto il fallimento delle ricerche finanziate dal governo americano. Nonostante ciò, il rapporto ALPAC sottolineò la necessità di ulteriori progressi tecnologici in questo settore.

Un altro documento interessante in merito al dibattito sulle metodologie di valutazione è il Van Slype Report del 1979, uno studio commissionato dalla Commissione Europea a seguito dell'adozione di SYSTRAN. Tale rapporto sottolineava come la valutazione dovesse essere condotta su due livelli:

- *macrovalutazione* (o valutazione complessiva) in cui i criteri da analizzare erano a livello cognitivo: intelligibilità, fedeltà, coerenza, usabilità e accettabilità; a livello linguistico: ricostruzione della relazione semantica, sintattica e coerenza semantica, qualità assoluta, valutazione sintattica e analisi degli errori;
- *microvalutazione* (o valutazione dettagliata) in cui si ricorre all'analisi degli errori e allo sforzo necessario per la revisione, nonché allo studio delle cause degli errori e delle possibilità di miglioramento del sistema.

Dal rapporto ALPAC a quello di Van Slype si assiste dunque ad una evoluzione del processo di valutazione della traduzione automatica nella direzione di una maggiore complessità del modello, che non tiene conto solo di criteri linguistici, bensì anche di

criteri di tipo socio-situazionale, economico e operativo. Un altro momento rilevante fu il lancio delle campagne di valutazione della traduzione automatica e delle metodologie proposte dall'ARPA (Advanced Research Projects Agency Software), un'agenzia statunitense che finanzia e formula progetti tecnologicamente innovativi per uso principalmente militare. Agli inizi degli anni '90 l'ARPA si diede l'obiettivo di promuovere progressi rivoluzionari nel campo della traduzione automatica attraverso un programma dedicato ad una valutazione *black box*, ovvero in cui si attua una misurazione qualitativa di un sistema solamente in base al suo prodotto, senza considerare il processo che l'ha portato a quel risultato.

A partire dagli anni 2000, con l'avvento dei sistemi di tipo statistico, i ricercatori si sono dedicati allo sviluppo e alla sperimentazione di metriche di valutazione quantitativa che non prevedevano alcun intervento umano, con l'obiettivo di dare vita ad una misurazione obiettiva e più veloce dei risultati della traduzione automatica. Sono nati così alcuni programmi software, che si basano sul principio di *similarità* della traduzione automatica rispetto alla traduzione umana. I criteri utilizzati per valutare la similarità sono la *precisione* e il *recupero*. Per calcolare la precisione, si conta il numero di segmenti presenti nella traduzione automatica e correttamente resi rispetto alla traduzione di riferimento. Per il calcolo del recupero, invece, si conta il numero di segmenti corretti nella traduzione automatica rispetto al numero di segmenti nella traduzione di riferimento. L'algoritmo di valutazione più utilizzato è BLEU (BiLingual Evaluation Understudy), inizialmente sviluppato per l'inglese, che confronta i risultati della traduzione automatica con un corpus di riferimento, costituito da testi tradotti da traduttori umani. La qualità viene misurata segmentando il testo prodotto dalla traduzione automatica in segmenti di *n*-grammi. BLEU quindi non usa singole parole per la valutazione ma *n*-grammi di lunghezza 4, ovvero sequenze di 4 parole. Questi vengono poi confrontati con il corpus di riferimento per individuare i segmenti condivisi tra la traduzione automatica e il testo di riferimento. Un punteggio BLEU di 1 indica che i due testi condividono lo stesso numero di segmenti, mentre un punteggio di 0 indica che non ci sono segmenti condivisi. Quindi, quanto più il punteggio è vicino a 1, tanto migliore è considerata la traduzione automatica.

Negli ultimi anni la ricerca accademica ha esplorato nuove metodologie per la valutazione, basati ancora sui concetti di adeguatezza e fruibilità. Due approcci di

misurazione più recenti sono il TAUS Dynamic Quality Framework (DQF) e QTL Launchpad Multidimensional Quality Metrics (MQM).

Il DQF venne lanciato da TAUS nel febbraio del 2014. Il nome stesso del progetto, "dynamic", suggerisce l'intento di creare un sistema di valutazione flessibile, in grado di adattarsi alle diverse esigenze del mercato. Il sistema di TAUS nasce nel settore privato, dall'unione di aziende, prevalentemente nel settore della localizzazione, di fornitori di servizi linguistici e istituzioni accademiche. Queste entità hanno collaborato per creare il sistema alla base del DQF, il quale consiste in una piattaforma connessa a una base di conoscenze contenente pratiche, relazioni, modelli e una serie di strumenti di valutazione sviluppati attraverso contributi umani e automatici.

Il Dynamic Quality Framework (DQF)¹⁴, dove la qualità della traduzione è considerata dinamica poiché i requisiti di valutazione cambiano a seconda del tipo di contenuto, dello scopo del contenuto e del pubblico a cui è rivolto, si basa sui criteri di adeguatezza e fruibilità, così definiti:

- **Adeguatezza:** quanto del significato espresso nella traduzione standard o nel testo di origine è espresso anche nella traduzione di destinazione" (Consorzio dei Dati Linguistici). La scala a 4 punti prevede i seguenti valori: Tutto, La maggior parte, Poco, Niente (*Everything, Most, Little, None*)
- **Fruibilità:** in che misura la traduzione è ben formata grammaticalmente, contiene ortografie corrette, aderisce all'uso comune dei termini, dei titoli e dei nomi, è intuitivamente accettabile e può essere interpretata sensatamente da un madrelingua. La scala a 4 punti prevede i seguenti valori: Impeccabile, Buona, Non fluida, Incomprensibile (*Flawless, Good, Dis-fluent, Incomprehensible*).

Il Multidimensional Quality Metrics, MQM¹⁵, è stato sviluppato all'interno di un progetto finanziato dall'Unione Europea e quindi di natura pubblica, non privata come nel caso del DQF. Questo sistema sembra essere più preciso dal punto di vista terminologico, poiché mette a disposizione una metrica di qualità basata su caratteristiche testuali: offre, quindi, un quadro completo per l'individuazione di

¹⁴ <https://www.taus.net/resources/blog/category/dynamic-quality-framework>

¹⁵ <https://themqm.org>

problemi nei testi originali e in quelli tradotti. L'approccio teorico alla base del MQM non si basa su un unico modello, ma su una serie di modelli che coprono diversi tipi di testi. Questo approccio è composto da diversi elementi:

- Un vocabolario utilizzato per classificare le problematiche relative alla qualità;
- Un sistema di assegnazione di punteggi basato sul conteggio degli errori o sulla loro annotazione;
- Una serie di regole di markup⁴, che includono l'identificazione delle caratteristiche testuali.

La base teorica su cui si fonda questo modello è di stampo funzionalista, per cui il primo valore su cui si concentra è la valutazione della vicinanza del testo tradotto ai fini comunicativi proposti dall'originale.

Il MQM identifica oltre 100 categorie di problematiche, emerse da un'analisi di sistemi di valutazione della qualità, sia basati su rilevazione automatica dei problemi, sia affidati al lavoro di revisori umani. Queste categorie sono organizzate in una gerarchia che va dal livello più generale a quello più specifico. In linea generale, nel documento originale, agli utenti viene consigliato di controllare che ci sia il minor numero possibile di problemi per soddisfare i requisiti degli utenti. Le problematiche che il sistema MQM prende in considerazione nella misurazione della qualità traduttiva sono le seguenti:

- *Aggiunta (addition)*: il prodotto di arrivo include porzioni di testo non presenti nel prodotto di partenza;
- *Contenuto specifico del locale (locale-specific-content)*: problematiche relative al mantenimento di contenuto specifico al locale di partenza;
- *Convenzioni del locale (locale-convention)*: comprende problematiche relative al mancato rispetto di convenzioni relative al locale di arrivo (come ad esempio l'utilizzo di determinati simboli numerici o traslitterazioni);
- *Errori di traduzione (mistranslation)*: il contenuto del prodotto di arrivo non rappresenta accuratamente il contenuto del prodotto di partenza;

- Fluidità (*fluency*): comprende problematiche relative all'aspetto linguistico del prodotto (ovvero alla sua aderenza a un determinato codice linguistico) che, in quanto tali, possono essere applicate tanto a un testo tradotto quanto a un testo originale indipendentemente;
- Grammatica (*grammar*): problematiche relative alla grammatica o alla sintassi del testo;
- Incoerenza (*inconsistency*): il testo presenta incoerenze interne;
- Indefinito (*unintelligible*): problematica la cui esatta natura non è determinabile, ma che ciò nonostante causa gravi problemi nella fruibilità del testo;
- Mancata traduzione (*untranslated*): parte del contenuto che sarebbe dovuto essere tradotto è stato lasciato in originale;
- Omissione (*omission*): mancanza nel testo di arrivo di contenuto presente nel prodotto di partenza;
- Ortografia (*orthography*): tutti quegli errori che non rispettano le regole convenzionali specifiche di una determinata lingua;
- Registro grammaticale (*grammatical-register*): problematiche relative a un'incongruenza tra il registro richiesto e quello utilizzato;
- Style (*style*:) comprende problematiche relative sia agli stili definiti formalmente sia a quelli definiti informalmente;
- Terminologia (*terminology*): comprende problematiche relative all'utilizzo di una terminologia;
- Tipografia (*typography*): problematiche relative alla rappresentazione tecnica del prodotto;
- Altro (*other*:) comprende problematiche non definibili in nessuna delle dimensioni precedenti.

In un articolo¹⁶ scritto da Gennaro Nolano, Data Scientist del Mare Digital, e da Johanna Monti, professoressa di Lingue Moderne presso l'Università Orientale di Napoli, sono stati usati questi parametri di problematiche del sistema MQM per analizzare la qualità della traduzione automatica del testo “King, Warrior, Magician,

¹⁶ http://opar.unior.it/1962/1/mqm_articolo.pdf

Lover” di Robert Moore e Douglas Gillette. Il testo, dopo essere stato analizzato linguisticamente, è stato tradotto attraverso i due traduttori automatici Google Translate e DeepL per poi essere completato in fase di post-editing da traduttore umano. Fase essenziale del processo è stata la valutazione degli errori riscontrati nelle due traduzioni automatiche che hanno messo in evidenza i seguenti risultati:

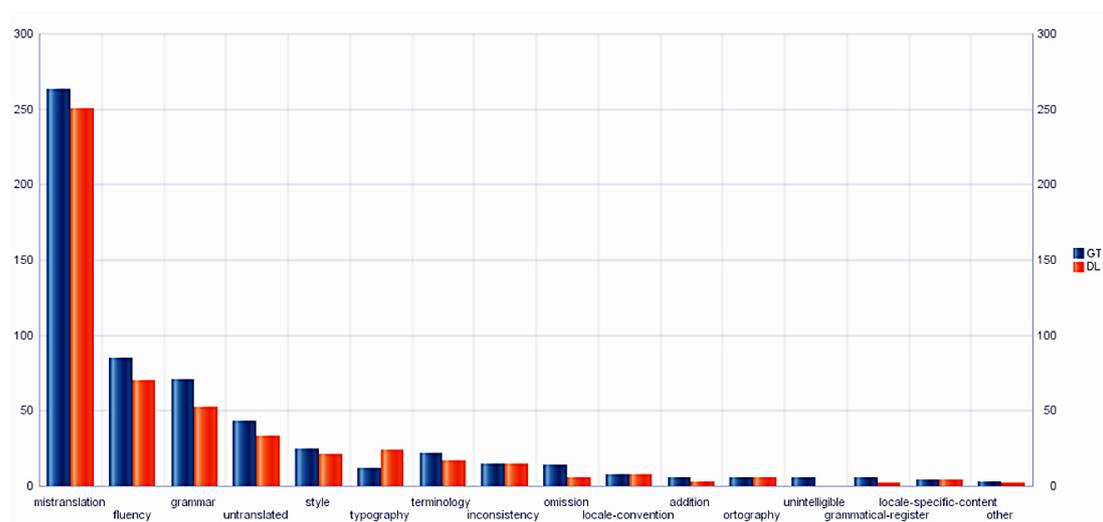


Figura 9. Tipologia di errori riscontrati tra i traduttori Google e DeepL

(Il Multidimensional Quality Metrics nell’analisi qualitativa della traduzione automatica – G. Nolano, J. Monti, 2019, p.9)

In generale, si può dire che si sono riscontrati più errori nella traduzione di Google Traduttore rispetto a quella di DeepL e che le tre problematiche con maggior percentuale di errore nella grafica sono quelle di *mistranslation*, *fluency* e *grammar*. Andando ad approfondire ulteriormente il concetto di errore di *mistranslation*, si considera l’aggiunta di problematiche più specifiche quali: una traduzione ambigua (*ambiguous-translation*) in cui è presente un’ambiguità non presente nel testo originale, problemi di data o orario (*date-time*) in cui c’è mancanza di coerenza tra le date o gli orari nel testo originale e in quello tradotto, entità (*entity*) con discrepanze nei nomi propri o nei luoghi, falsi amici (*false-friend*) dove la traduzione include erroneamente una parola che superficialmente assomiglia a quella nel testo originale, numeri (*number*) con la presenza di discrepanze nei numeri tra i due testi, traduzione troppo letterale (*overly-literal*) o parti che non dovrebbero essere tradotte (*no-*

translate) e che dovrebbero rimanere nella lingua originale. Per quanto riguarda la problematica “fluidità” i suoi errori rappresentano la seconda categoria più frequente. Questo tipo di problematica, come descritto nella documentazione ufficiale del MQM, riguarda questioni che possono essere individuate anche senza confrontare il testo originale con quello tradotto, in quanto sono indipendenti dallo stato di traduzione del testo analizzato: potevano risultare poco fluenti nella lettura, senza tuttavia compromettere completamente la comprensione del testo.

Infine la terza problematica più frequente è quella relativa alla grammatica. Gli errori riscontrati in questa categoria comprendono: uso incorretto, aggiunta o omissione di parole funzionali (*function-words*) nella traduzione, errori relativi alla forma grammaticale delle parole (*word-form*) o problemi riguardanti l'ordine delle parole (*word-order*). Indipendentemente dalla gravità di tali problematiche, queste si sono dimostrate, insieme alle questioni di fluidità più generiche, tra le più facili da individuare.

Essendo percentuali piuttosto alte di errori, sono quelle che hanno inciso di più sul risultato finale e quindi sulla qualità della traduzione automatica. Nella pratica, la discussione sulla valutazione e il controllo della qualità traduttiva resta aperta e viene alimentata da un lato dai contributi provenienti dall'accademia e dall'altro dai contributi provenienti dalla professione. I contributi accademici cercano di definire criteri precisi e forniscono modelli teorici, ma talvolta sono molto complessi. Ad ogni modo, per il controllo della qualità traduttiva, l'industria della traduzione si affida normalmente alle *best practice*, ossia una serie di pratiche consolidate, che comprendono controlli basati su metriche di valutazione, quali quelle appena descritte DQF e MQM, ancora in fase di aggiornamento, ma molto utili per valutare la qualità di un prodotto.

2.4 I limiti dell'intelligenza artificiale nel campo della traduzione

L'avanzamento delle tecnologie nel campo della traduzione automatica ha suscitato un crescente interesse e ha rivoluzionato il modo in cui le lingue vengono tradotte e interpretate. Spesso si fa affidamento a sistemi di *machine translation* quando si vuole

avere un riscontro veloce e immediato, quando si vuole sapere sommariamente il significato di una mail o di un sito web in una lingua diversa dalla propria oppure, nella maggior parte dei casi, la traduzione automatica che troviamo su Internet, come quella offerta da Google Traduttore o DeepL, viene utilizzata perché gratuita. Nonostante i grandi progressi fatti e i vari studi su questo campo per migliorare i sistemi, per esempio passando da un approccio statistico ad uno neurale, i limiti che questi hanno sono molti e importanti, e probabilmente non si riuscirà mai a trovare una soluzione a questo ostacolo. In questo paragrafo si andranno ad analizzare alcune tra le principali limitazioni dei sistemi traduttivi automatici di oggi e si andrà a dimostrare come la figura di un traduttore o interprete professionista sia insostituibile e necessaria per un risultato di qualità. Per prima cosa, un sistema di traduzione automatica è uno strumento troppo impreciso, che non è in grado di cogliere le reali esigenze di chi chiede una traduzione e del risultato che quest'ultimo desidera ottenere. Non riesce a cogliere le sfumature di significato: tradurre non significa solamente trasportare il significato di una parola da una lingua ad un'altra, ma vuol dire tener conto di tutto l'insieme, ovvero del contesto culturale e delle sfumature linguistiche. Solo un traduttore umano è in grado fare tutto ciò, quindi di interpretare e di trasferire nella lingua di arrivo quello che viene espresso nella lingua di partenza, magari utilizzando altre parole ma rimanendo sempre fedele al testo originale. Un altro problema che si riscontra nell'utilizzo dei traduttori automatici è che questi non sono capaci di disambiguare gli omofoni, ovvero parole che si scrivono e pronunciano nello stesso modo ma che hanno un significato diverso. Questo avviene perché non sono in grado di analizzare il contesto in cui vengono utilizzati questi termini e spesso confondono i significati, sbagliando e mal interpretando il testo. L'intelligenza artificiale non considera il contesto linguistico e culturale né i cambi di registro nel testo: ad esempio, se un testo viene scritto in un registro colloquiale con l'utilizzo della lingua parlata, andrà tradotto in maniera differente rispetto, ad esempio, ad un testo giuridico che richiede un registro più alto e un utilizzo di parole formali e adeguate al contesto. Di conseguenza, non restituisce traduzioni professionali, al contrario di come farebbe uno specialista del settore.

Il sistema automatico ragiona in questa direzione perché funziona su base statistica e, quindi, non prevede una reale lettura del testo di partenza, ma si limita a tradurre

seguendo esempi di testi o parole presenti nel suo database e che poi riutilizza nelle nuove traduzioni. Un essere umano, invece, può andare oltre anche davanti a ciò che è del tutto nuovo o assolutamente improbabile e lo può fare appoggiandosi ad una esperienza del linguaggio che va oltre le semplici parole già dette o ascoltate.

Prendendo come esempio Google Traduttore, si può vedere come questo sia relativamente affidabile nella lingua inglese, che è quella più diffusa e su cui il sistema è riuscito a crearsi una sorta di base di esperienza per poter tradurre efficacemente, soprattutto se si tratta di testi dalla sintassi semplice. Nella maggior parte delle lingue, invece, poiché sono meno diffuse e hanno una sintassi più complessa, come ad esempio il tedesco, l'algoritmo è ancora lontano dall'essere soddisfacente, non avendo avuto una base sufficiente da cui "imparare" e formarsi.

Un'altra problematica di fondamentale importanza è quella della sicurezza. Una delle ragioni principali per cui Google Traduttore è così popolare tra gli utenti di tutto il mondo è sicuramente il fatto che sia gratuito. In cambio, però, Google richiede agli utenti la possibilità di utilizzare i loro dati per alimentare e perfezionare costantemente il suo algoritmo. Ciò significa che anche informazioni sensibili, come risultati di test di laboratorio, dati finanziari altamente riservati o cartelle cliniche inseriti in Google Traduttore per ottenere una traduzione rapida, finiscono per essere inclusi nell'enorme contenitore di dati che Google immagazzina e utilizza. Le Norme sulla Privacy di Google, infatti, affermano che "Utilizziamo diverse tecnologie per elaborare le tue informazioni per gli scopi indicati. Utilizziamo sistemi automatizzati che analizzano i tuoi contenuti [...]. Usiamo anche algoritmi per individuare eventuali modelli nei dati. Ad esempio, Google Traduttore aiuta le persone a comunicare in più lingue rilevando i modelli del linguaggio comune nelle frasi che gli chiedi di tradurre".

Oltre a traduttori e interpreti, anche i doppiatori si trovano oggi a dover affrontare la "lotta" con l'intelligenza artificiale. Nel campo del doppiaggio questa può essere utilizzata per creare voci sintetiche che potrebbero potenzialmente competere con i doppiatori umani in alcuni contesti, specialmente per progetti a basso budget o in cui è richiesta una voce generica, riducendo, in questo modo, la necessità di lavoro umano in alcune fasi della produzione. Nonostante i progressi, anche per questo settore i problemi non mancano. Quello principale è sicuramente quello legato alla privacy e alla sicurezza: la voce di un individuo può essere sintetizzata dall'IA, falsificando

audio che potrebbero essere utilizzati in modo fraudolento. Un altro limite da citare è il fatto che le voci artificiali, seppur ben sviluppate, non possono paragonarsi alle voci umane, che riescono a restituire profondità emotiva e autenticità: infatti i doppiatori professionisti possono offrire interpretazioni ricche e una connessione più genuina con il pubblico.

La necessità di affrontare temi più vicini alla realtà professionale e coinvolgere maggiormente i traduttori nel dibattito sul miglioramento delle tecnologie è stata evidenziata anche durante il recente Machine Translation Summit XVII. Per la prima volta è stata organizzata una sessione dedicata esclusivamente al mondo della traduzione e ai traduttori, durante la quale sono stati trattati argomenti come la misurazione della produttività, il suo impatto sull'adozione della traduzione automatica, il ruolo nel lavoro del traduttore, questioni etiche e di riservatezza legate all'uso delle tecnologie e aspetti psico-sociali come la predisposizione psicologica all'utilizzo delle tecnologie e i pregiudizi.

È innegabile che la tecnologia stia e continuerà a modificare profondamente il ruolo del traduttore. Sicuramente l'utilizzo della tecnologia semplifica il compito di quest'ultimo, consentendogli, ad esempio, di evitare la traduzione completa partendo da zero, e facendo invece affidamento su testi pre-tradotti generati dalle macchine, sui quali poi apportare eventuali miglioramenti nella fase di post-editing. Soprattutto in un contesto professionale in cui la professione dei traduttori e degli interpreti ha conosciuto un'evoluzione tecnologica significativa negli ultimi anni grazie all'impiego delle tecnologie più avanzate, ci si chiede quali debbano essere le competenze e le abilità essenziali per queste figure in un'era sempre più digitalizzata. Quando si affronta un processo di traduzione, è essenziale considerare attentamente sia il contesto di partenza che quello di arrivo. Ad esempio, nel mondo professionale si parla sempre più frequentemente di "transcreation", un processo di traduzione che si colloca tra la traduzione e il *copywriting*, e che richiede un mix articolato di competenze, soprattutto la creatività del traduttore. Quindi, in situazioni in cui è necessario modificare completamente il nome di un prodotto o uno slogan per esempio, si impiega il processo di transcreazione, vista come una forma di "traduzione creativa", poiché prevede l'adattamento del messaggio da una lingua all'altra, senza seguire necessariamente una traduzione letterale parola per parola. Questo accento sull'aspetto creativo della

traduzione mantiene i traduttori al centro dell'azione, almeno per il momento, impedendo l'adozione delle macchine nel processo transcreativo. Sia la localizzazione che la transcreazione rappresentano forme di traduzione più sofisticate che non possono essere eseguite da sistemi di traduzione automatica, ma che richiedono l'intervento umano per essere eseguite in modo accurato e adeguato.

Anche nel post-editing è evidente come la figura di un traduttore specializzato sia essenziale, poiché un professionista del settore è l'unico capace di giudicare l'accuratezza di una traduzione individuando eventuali errori nella traduzione automatica. Quest'ultimo ha, inoltre, una conoscenza approfondita del trasferimento concettuale tra lingue e dispone di risorse tecniche che sa come utilizzare in caso di dubbi, valutandone l'adeguatezza in relazione alle esigenze tecniche e ai costi.

Anche le abilità linguistiche sono essenziali per qualsiasi traduttore professionista, così come le competenze di base. Queste ultime sono specificamente orientate verso competenze attitudinali o psico-fisiologiche che includono, per esempio, la capacità di elaborare e applicare le norme del post-editing alla traduzione, la gestione delle aspettative del cliente sulla qualità finale e la capacità di prendere decisioni informate riguardo alle strategie di post-editing da adottare. Certamente un altro aspetto fondamentale per il traduttore, stavolta più pratico, è la competenza strumentale, che riguarda le abilità tecniche necessarie per utilizzare consapevolmente e accuratamente le tecnologie. Queste competenze riguardano la conoscenza dei vari sistemi di traduzione automatica e delle loro caratteristiche, la gestione della terminologia, la manutenzione dei dizionari, la valutazione della qualità di un corpus, il pre-editing attraverso l'uso di linguaggi controllati e, infine, le competenze di programmazione.

Sebbene tutte queste competenze possano apparire generiche e in qualche modo convenzionali, sono in realtà elementi indispensabili per una figura professionale flessibile, chiamata a operare in un contesto in cui le tecnologie si evolvono rapidamente e costantemente. La capacità di apprendere costantemente, inoltre, è fondamentale per consentire al traduttore di adattarsi prontamente ai progressi tecnologici.

Dall'analisi appena fatta si vede come nonostante i progressi nell'IA, la mediazione linguistica richieda ancora competenze umane specifiche per garantire la comprensione culturale, l'adattamento contestuale e la sensibilità alle sfumature

linguistiche. La comprensione di una cultura diversa, la sensibilità e la capacità di adattarsi alle esigenze dei diversi utenti rimangono aspetti fondamentali che solo un mediatore umano può offrire. Un traduttore o un supporto automatico possono essere strumenti preziosi di supporto, grazie ai quali si possono migliorare la comunicazione e la mediazione linguistica, agevolando la traduzione, l'interpretazione e l'interazione multilingue e possono rivelarsi uno strumento utile per ammorbidire le barriere linguistiche. Diventa però essenziale mantenere un equilibrio tra l'uso dell'intelligenza artificiale e l'intervento umano per garantire risultati precisi e di alta qualità nella comunicazione interlinguistica: la presenza umana rimane, comunque, essenziale per una comunicazione interculturale efficace.

CAPITOLO TRE

L'ESEMPIO DI *STRAPPARE LUNGO I BORDI* DI ZEROCALCARE

3.1 L'adattamento del dialetto e dell'ironia

La traduzione automatica ha rivoluzionato il modo in cui interagiamo con testi scritti in diverse lingue, offrendo una soluzione immediata per comprendere il significato generale di un messaggio. Questa tecnologia si è rivelata particolarmente efficace quando si tratta di tradurre testi di natura più generale o standardizzata, che possono essere ad esempio mail o siti web. Tuttavia, quando ci si confronta con testi più specifici o intrisi di conoscenze specialistiche, in cui la comprensione richiede una conoscenza profonda dei contesti culturali o scientifici presi in considerazione, la traduzione automatica mostra i suoi limiti evidenti, in cui si sfidano le capacità delle macchine. In queste situazioni, emerge l'importanza e l'insostituibilità del ruolo del traduttore umano, che porta con sé non solo la competenza linguistica, ma anche la conoscenza approfondita del contesto e delle implicazioni sottili del testo originale: nell'affrontare questa sfida, i traduttori umani dimostrano la loro abilità di catturare la complessità e la ricchezza di significati. Uno dei tanti contesti in cui emerge la limitatezza delle macchine nella traduzione è rappresentato dal dialetto e dall'umorismo, due elementi che sfuggono alla comprensione dei traduttori automatici. Prima di andare ad analizzare alcuni esempi pratici per dimostrare quanto sia difficile, o addirittura impossibile, tradurre queste variazioni linguistiche per un traduttore artificiale, si andranno ad illustrare le loro principali caratteristiche cercando di capire perché sono così complesse da trasporre in altre lingue, anche per un traduttore umano. Partendo dal dialetto, il termine (dal greco "conversare") fa riferimento a quelle varietà linguistiche parlate affianco ad una lingua maggiore, ossia quella ufficiale, e usate in ambito informale. Il numero di parlanti è generalmente ristretto ad una determinata zona, nonostante ci siano casi in cui un dialetto è riuscito a diffondersi oltre i propri confini socio-linguistici. Nella linguistica, il concetto di "dialetto" è definito come una variante di una lingua, ma a volte può risultare difficile distinguere se un idioma è considerato una lingua o un dialetto. Per risolvere questa questione, si possono considerare tre parametri fondamentali, mettendo a confronto due parlanti di idiomi diversi:

1. Comprensione reciproca: se i due parlanti sono in grado di capirsi quando parlano, allora si tratta probabilmente di due dialetti della stessa lingua, poiché esiste una mutua intelligibilità tra di loro;
2. Lessico di base in comune: se più dell'80% delle parole di uso comune utilizzate dai due parlanti sono le stesse, anche se pronunciate in modo diverso, allora si possono considerare i due idiomi come dialetti della stessa lingua;
3. Morfologia e sintassi omogenee: se i due parlanti seguono le stesse regole grammaticali per esprimersi, allora è probabile che stiano parlando due dialetti della stessa lingua.

Questi parametri forniscono un quadro utile per distinguere tra lingue e dialetti, consentendo di comprendere meglio la natura delle varietà linguistiche presenti nelle diverse comunità linguistiche. Se consideriamo la lingua come un colore, si può dire che i dialetti si definiscono come gradazioni diverse dello stesso colore, e quindi della stessa lingua. Dal punto di vista sociolinguistico, invece, sono stati individuati diversi criteri per distinguere tra lingua e dialetto:

1. Estensione geografica: mentre una lingua è diffusa su un territorio vasto, il dialetto è confinato in un'area geografica più ristretta;
2. Assenza di standardizzazione: il dialetto manca di una forma "corretta" accettata universalmente dai parlanti, risultando in una varietà frammentata e locale, con ogni comunità che lo parla in modo unico;
3. Percezione sociale: il dialetto è spesso considerato dalla sua comunità parlante come un linguaggio informale e meno prestigioso, associato principalmente alle classi sociali meno abbienti e meno istruite;
4. Utilizzo informale: i dialetti vengono impiegati principalmente in contesti sociali informali, come in famiglia o tra amici, mentre non sono comunemente utilizzati in ambienti formali come le istituzioni pubbliche, le scuole o gli ambienti di lavoro;

5. Limitata produzione letteraria: i dialetti sono principalmente orali e mancano spesso di una tradizione letteraria significativa, quando presente, di valore limitato;
6. Influenza della lingua dominante: nei luoghi in cui è parlato, il dialetto è spesso influenzato dalla presenza di una lingua dominante, dalla quale riceve prestiti linguistici e costrutti grammaticali;
7. Limitatezza del lessico: i dialetti si concentrano sul vocabolario relativo alla vita quotidiana, spesso mancando di termini necessari per discutere argomenti scientifici, tecnologici, filosofici e altri ambiti di conoscenza.

Oltre al dialetto, anche l'ironia rappresenta uno strumento linguistico abbastanza complicato da tradurre. Innanzitutto, si tratta di una forma di comunicazione sottile, che non mira a nascondere, ma piuttosto a esporre contenuti sarcastici in modo indiretto. È rappresentata da un'interazione tra il linguaggio verbale e non verbale, ad esempio fatto da gesti o espressioni facciali. Secondo Muecke (1969; 1970), il messaggio ironico deve essere chiaro, ma non immediatamente evidente ed è comunemente utilizzato in contesti umoristici o per ottenere un effetto emotivo nella comunicazione. Durante il processo vocale ironico, le frasi possono, ad esempio, assumere caratteristiche specifiche come una pronuncia più nasale, una diminuzione della velocità del discorso, parole allungate e sillabe enfatizzate in modo esagerato, che danno vita a una forte enfasi emotiva nella comunicazione. La comprensione dell'ironia, che talvolta può mancare, è facilitata dalla condivisione di un background comune tra gli interlocutori e dalla conoscenza dell'argomento oggetto di ironia. Tuttavia, la mancata ricezione del messaggio ironico può derivare da un malinteso o da un rifiuto di comprendere l'ironia stessa.

Talvolta nell'ironia, nel sarcasmo o nello *humor* può esserci la presenza del turpiloquio, ovvero l'uso di termini o espressioni volgari o offensive. È spesso presente in molti dialetti o lingue regionali. I termini e le espressioni utilizzate spesso fanno riferimento a diversi tabù linguistici che possono riguardare il sesso, l'aggressività, la religione, ecc. I termini sono ricchi di valore emotivo, tant'è vero che molto spesso esprimono i sentimenti primari dell'uomo come la rabbia, la sorpresa, la paura, il divertimento. Le parole possono essere usate per imprecazioni, in cui si instaura un

vero e proprio dialogo con sé stessi e che servono a sfogare simbolicamente la propria aggressività magari contro oggetti inanimati; per insulti, di solito rivolti ad altre persone; per maledizioni, in cui le espressioni servono per augurare il male alla persona a cui ci si rivolge. Gli elementi del turpiloquio incontrano spesso una forte censura quando arriva il momento di tradurli in un'altra lingua di destinazione. La loro traduzione è influenzata da diversi fattori: in primo luogo, la ripetitività, soprattutto delle espressioni comuni nel doppiaggio, e poi l'originalità, in quanto durante la fase di "transcreazione" (Capitolo 2 paragrafo 2.4) si possono trovare dei sostituti che si discostano troppo dalla lingua originale. Inoltre, vi è la questione della corrispondenza, o mancanza di essa, tra la lingua di origine e quella di destinazione, sia dal punto di vista linguistico che concettuale, oltre alle possibili restrizioni imposte dal contesto audiovisivo. Talvolta, poi, i traduttori possono essere limitati nelle loro trasposizioni a causa di direttive o di esigenze delle case di produzione. Tuttavia, il turpiloquio è principalmente un mezzo di espressione emotiva e, proprio come gli elementi umoristici discussi in precedenza, non richiede una traduzione letterale, ma piuttosto un equivalente nella lingua di destinazione.

3.2 Analisi di parti selezionate dalla serie televisiva

La traduzione di tutti gli elementi comici e dialettali citati nel paragrafo precedente rappresenta un'ardua sfida per i traduttori di oggi, a cui viene richiesta non solo competenza linguistica, ma anche una profonda comprensione della cultura di origine e di quella di destinazione, insieme a una dose di creatività per preservare l'essenza e l'effetto emotivo degli elementi originali. Se già per i professionisti del settore questo è un compito ostico, che richiede un lungo studio e tanto impegno, si può solo immaginare che cosa sarebbe capace di fare un sistema di traduzione automatica se gli venisse affidato il compito di trasportare un testo ricco di sfumature linguistiche, culturali, sociali, ironia e comicità in un'altra lingua, lasciando inalterato il senso del testo originale e rispettando ciò che esso vuole comunicare.

Nel seguente paragrafo si andranno ad analizzare alcuni esempi pratici di traduzione verso l'inglese della serie Netflix in romanesco "Strappare lungo i bordi" del fumettista Zerocalcare. Per prima cosa verranno mostrate le sfide incontrate dai traduttori umani

nella realizzazione dell'adattamento inglese della serie, per poi dimostrare come un traduttore automatico propone la trasposizione di alcuni elementi comici, di modi dire e del turpiloquio.

Nel 2020 Michele Rech, in arte Zerocalcare, annuncia una collaborazione con Netflix per la sua prima serie animata, *Strappare lungo i bordi*, che verrà distribuita l'anno successivo, nel 2021, sulla piattaforma streaming. Prima di arrivare alla creazione di questa serie, l'autore era stato spesso al centro del panorama fumettistico e letterario italiano tramite una serie di partecipazioni a mostre ed esposizioni e grazie alla pubblicazione dei suoi tanti libri e fumetti, a cui si dedica dai primi anni 2000. Zerocalcare ha saputo trasformare le sue vignette in opere multiformi, trovando spazio su diverse piattaforme creative, tra cui riviste, giornali, audiolibri, cortometraggi animati, murali, copertine di album musicali, nonché produzioni cinematografiche e televisive. Inizia la sua carriera nel 2001 scrivendo un racconto sugli avvenimenti del G8 di Genova, per poi proseguire con la stampa dei suoi libri che in ordine cronologico sono: *La profezia dell'armadillo* (2011), *Un polpo alla gola* (2012), *Ogni maledetto lunedì su due* (2013), *Dodici* (2013), *Dimentica il mio nome* (2014), *L'elenco telefonico degli accolti* (2015), *Kobane Calling* (2016), *Macerie prime* (2017- seconda edizione 2018), *La scuola di pizze in faccia del professor Calcare* (2019), *Scheletri* (2020), *A Babbo morto. Una storia di Natale* (2020), *Niente di nuovo sul fronte Rebibbia* (2021), *No sleep till Shengal* (2022), *Dopo il botto* (2023), *Animation Art Book* (2023), *Enciclopedia calcarea. Guida ragionata all'universo di Zerocalcare* (2023). Oltre alla serie *Strappare lungo i bordi* del 2021, viene tratto un film intitolato "La profezia dell'armadillo", nel 2018, adattamento dell'omonimo libro a fumetti del 2011, e una seconda serie Netflix "Questo mondo non mi renderà cattivo" del 2023. Grazie a queste pubblicazioni, molto apprezzate dal pubblico e per le quali ha ottenuto una serie di premi e riconoscimenti, Michele Rech è diventato uno dei fumettisti più acclamati e influenti dell'attuale panorama italiano. L'artista è nato nel 1983 ad Arezzo e ha trascorso parte della sua giovinezza in Francia prima di stabilirsi a Roma, città che ha plasmato profondamente la sua identità artistica e alla quale è fortemente legato. Nei suoi lavori Zerocalcare si distingue per uno stile che unisce ironia e profondità, affrontando tematiche diverse e apparentemente inconciliabili. Da un lato, le sue narrazioni esplorano questioni personali, come il processo di crescita e maturazione;

dall'altro, manifestano un forte impegno verso tematiche di carattere sociale o politico, come testimoniano i suoi reportage sulle battaglie di Kobane o la sensibilità nei confronti dei problemi sociali della sua generazione. Inoltre, le opere di Zerocalcare sono permeate da una sottile autoironia e arricchite da riferimenti alla cultura popolare, elementi che contribuiscono a renderlo un artista trasversale, capace di conquistare un pubblico di diverse età e interessi. Le opere del fumettista spesso traggono ispirazione dalle sue esperienze di vita, le quali vengono reinterpretate e fuse con elementi inventati. L'autobiografismo si intreccia con elementi fantastici, dando vita a trame narrative in cui il confine tra realtà e finzione è sottile, permettendo al lettore di identificarsi nei sentimenti dell'autore pur mantenendo una certa distanza dalla sua vita.

Il successo di "Strappare Lungo i Bordi" è stato molto rapido, arrivando in cima alle classifiche italiane in soli quattro giorni dalla sua pubblicazione. La serie è ambientata principalmente a Roma, in particolare nel quartiere di Rebibbia, caro all'autore fin dall'infanzia e in cui ha vissuto la maggior parte della sua vita, e racconta di un viaggio verso Biella, una piccola città piemontese, che il protagonista Zero fa insieme ai suoi amici Secco e Sarah, per andare al funerale di una loro amica, Alice, che si è tolta la vita. L'obiettivo principale della serie, evidente già dal titolo, è quello di descrivere la pressione che spesso si sente nel dover seguire un sentiero tratteggiato, un percorso predeterminato verso la felicità e la realizzazione personale, fatto di obiettivi come il conseguimento della laurea, la ricerca del lavoro ideale e la formazione di una famiglia, e le difficoltà che ne derivano quando non si riesce a seguire questo percorso, come la depressione, la solitudine e persino il suicidio. Pur affrontando tematiche intense, la narrazione è resa più leggera dall'umorismo cinico tipico del fumettista, spesso incarnato dal personaggio dell'Armadillo, la coscienza del protagonista. La serie è stata resa disponibile in diverse lingue, incluso l'inglese, con doppiaggio e sottotitolazione. Per l'adattamento inglese sono state adottate alcune scelte traduttive che si andranno ad analizzare tenendo conto degli elementi descritti nel paragrafo precedente e che verranno confrontati con una traduzione affidata ai vari servizi di traduzione automatica descritti nel capitolo 2. Per l'analisi di queste sezioni si terrà conto principalmente del sottotitolaggio inglese e non del doppiaggio, dal momento che l'analisi si propone di analizzare i risultati di una traduzione automatica scritta.

La serie si compone di sei episodi, ognuno di 20 minuti circa, nei quali sono presenti gli elementi sopracitati come il dialetto romano, elemento principale della serie, elementi comici, modi di dire ed espressioni tipiche. Nelle tabelle proposte di seguito sono presentati tutti questi elementi spiegati tramite vari esempi presenti negli episodi, affiancati dal sottotitolaggio inglese e messi a confronto con una traduzione svolta dai traduttori automatici Google Traduttore, Microsoft Translator e DeepL che, a parte per alcune differenze, hanno dato tutti risultati molto simili tra loro.

Episodio 1

	Testo originale	Adattamento inglese	TA
1.	Sottone	Loser	Underling
2.	Monnezza	Sort of trash	Rubbish
3.	Sta' bona	Give me a break/be quiet	She's good

Dal primo episodio sono stati tratti tre esempi di riferimento utili all'analisi in questione. Nella prima riga della tabella, osserviamo la presenza della parola "sottone": il protagonista, Zero, la utilizza in una conversazione interna in cui esprime la speranza di non apparire così quando parla con la ragazza che gli piace, Alice, al fine di non rivelarle il suo interesse. Nell'adattamento inglese, questa parola viene resa con *loser* poiché il senso è quello di evitare di mostrare eccessivamente la sua indipendenza, facendolo sembrare quasi un "perdente" agli occhi della ragazza. Tuttavia, nella traduzione automatica, viene erroneamente utilizzato il termine *underling*, il quale indica invece una posizione inferiore o subordinata, alterando il significato del dialogo. Nel secondo esempio si utilizza la parola "monnezza" per riferirsi a qualcosa fatto male, qualcosa di grottesco o spiacevole. La sua traduzione inglese è infatti *trash* (*sort of trash* per il contesto del discorso) inteso come "robaccia"; tuttavia i traduttori automatici interpretano il significato letterale di "immondizia" e lo traducono come *rubbish*, che indica i rifiuti e la spazzatura in senso pratico. L'ultimo esempio preso in considerazione è l'espressione romanesca "sta' bona" che esprime il significato di "stare tranquilli", "dare tregua", tradotto efficacemente con il sottotitolaggio "give me a break", "be quiet", che cattura il senso dell'espressione

originale e la sua connotazione culturale romana. L'intelligenza artificiale non riesce, invece, a interpretare il contesto e a cogliere la sfumatura culturale romana, interpretando erroneamente l'espressione come un complimento verso una donna o una ragazza, e suggerendo la traduzione "she's good".

Episodio 3

	Testo originale	Adattamento inglese	TA
1.	[...] erano roba che lei ormai se la legava al cazzo.	[...] were just fucking child's play.	[...] was stuff that she tied to her cock.
2.	[...] per lei era come bere un bicchier d'acqua.	[...] it was a piece of cake for her.	[...] it was like drinking a glass of water for her.
3.	[...] che per comodità chiameremo pizza "stocazzo".	[...] which for brevity's sake we'll call it the "whatever".	[...] which for convenience we'll call it "fucking pizza".
4.	Boh.	Dunno.	I don't know.

Nel terzo episodio sono stati selezionati quattro esempi da analizzare, per lo più riguardanti frasi intere. La prima riga va così spiegata: il protagonista, Zero, parla di Alice, sua cara amica, raccontando della sua vita e dicendo che per potersi pagare l'università dava ripetizioni di matematica a dei bambini nel doposcuola. Era diventata talmente tanto esperta nei problemi di algebra e trigonometria che "erano roba che lei ormai se la legava al cazzo". Questa espressione, tutta romanesca, indica che per lei era diventata un'azione di estrema facilità. Per mantenere lo stesso significato del testo italiano originale, in inglese viene usata l'espressione "were just fucking child's play", ovvero "era un gioco da ragazzi". Il traduttore automatico traduce parola per parola la frase, dando vita ad un'espressione che non ha alcun senso e che sarebbe incomprensibile per un pubblico inglese. Molto simile è l'esempio proposto nella seconda riga, in cui troviamo il detto italiano: "facile come bere un bicchier d'acqua".

Il suo corrispettivo inglese è il famoso detto *like a piece of cake*, ampiamente conosciuto anche al di fuori dei paesi anglofoni. Se, però, si prova a digitare la stessa espressione su Google Traduttore o DeepL, il risultato sarà una traduzione letterale della frase, che per un madrelingua inglese avrà il significato letterale di “bere un bicchiere d’acqua”, perdendo il significato figurato che il detto originariamente intendeva trasmettere.

Nella penultima riga ci si trova nella scena in cui Zero, a cena in pizzeria con l’amico Secco, è indeciso su quale pizza ordinare e non sa se scegliere tra la solita margherita oppure una nuova pizza che legge sul menù e a cui dà il nome di “pizza stocazzo”. È evidente qui l’intento dell’autore di evitare di menzionare qualsiasi altro tipo di pizza per non risultare offensivo, adottando questo nome identificativo un po’ volgare, in linea con lo stile della serie. L’adattamento inglese utilizza il termine *whatever* per esprimere il significato di “una pizza qualsiasi”. La TA crede invece che si voglia associare un nome volgare alla pizza, quasi come se il protagonista avesse qualcosa contro quel particolare alimento, traducendolo quindi come “fucking pizza”.

L’ultimo esempio riguarda lo stile e il registro linguistico. È ovvio che in questa serie, in cui si parla principalmente il dialetto romano, il linguaggio non sia elevato, ma anzi colloquiale. Ad esempio, è spesso presente l’espressione “Boh”, molto colloquiale e comune tra i giovani e che corrisponde all’inglese *dunno*, usato anch’esso in contesti informali. Il traduttore non coglie questa sfumatura e lo traduce semplicemente con “I don’t know”, adottando quindi un inglese standard che farebbe perdere tutta l’atmosfera informale e leggera che il fumettista desidera trasmettere nella sua serie.

Episodio 4

	Testo originale	Adattamento inglese	TA
1.	Ammazza!	Wow!	Kill/slay!
2.	Scrauso	Crappy	Scrawny/scraggly

Nel quarto episodio sono stati presi come oggetto di analisi solamente due esempi. Il primo è l’espressione “ammazza!” che denota stupore, meraviglia, e che viene resa efficacemente in inglese con l’espressione *wow!* Tuttavia, il traduttore automatico

interpreta erroneamente l'espressione come un verbo imperativo alla seconda persona singolare, in cui viene dato l'ordine di uccidere qualcuno con le frasi: "kill!" o "slay!" Nel secondo caso, viene analizzato il termine dialettale "scauso" per definire qualcosa che vale poco, "da quattro soldi", stesso significato espresso dall'aggettivo inglese *crappy*. L'intelligenza artificiale dei vari traduttori automatici comprende il significato di *scrawny*, che significa magro, scheletrico, oppure quello di *scraggly* con il significato di incolto, trasandato.

Episodio 5

	Testo originale	Adattamento inglese	TA
1.	Comunque, se semo accannati.	We'd tell each other about breakups, though.	Anyway, we got stuck.
2.	[...] io te raggiungo col coso [...] te raggiungo col cazzo.	[...] I'll follow you with [...] I'll fucking stay in bed.	[...] I'll join you with the thing [...] I'll join you with my dick.
3.	[...] pure se non lo strappi quello se ciancica.	[...] it gets crumpled anyway.	/

Nell'episodio 5 si analizzano tre esempi di frasi. Nella prima riga si legge una tipica espressione della Capitale molto diffusa tra i ragazzi di oggi quando si parla di relazioni sentimentali: "Se semo accannati", che sta ad indicare quando una relazione non va bene e si decide quindi di interromperla. La frase viene tradotta in inglese come "we'd tell each other about breakups", con il significato di "lasciarsi", "mollarsi" come espresso nel testo originale italiano. Il traduttore automatico rende l'espressione con "we got stuck" che significa "ci siamo bloccati". Anche se facendo attenzione al contesto e alla scena in cui appare questa frase si potrebbe arrivare al significato espresso dalla traduzione automatica, resta il fatto che uno spettatore deve poter leggere e capire il sottotitolo senza necessariamente dover fare attenzione alla scena o al contesto della frase.

Nel secondo esempio è presente una frase iconica della serie, diventata quasi il simbolo di *Strappare lungo i bordi*. In questa scena Alice chiama Zeno in preda alla disperazione dopo aver interrotto la sua storia con il fidanzato e gli chiede di incontrarsi per poter parlare. Lo fa però alle 2 di notte, quando Zero sta dormendo. Nonostante ciò, decide comunque di aiutare la sua amica e di ascoltarla. Per trovare la voglia chiede all'Armadillo, personificazione della sua coscienza, di accompagnarlo, ma lui rifiuta. Per esprimere questo concetto, all'Armadillo viene affidata questa battuta volgare, in linea con lo stile della serie: “tu vai... io te raggiungo col coso [...] te raggiungo col cazzo”, per indicare che non ha nessuna voglia o intenzione di seguirlo. Nei sottotitoli inglesi si legge: “I’ll follow you with...[...] I’ll fucking stay in bed”, come a voler dire “non ci penso proprio a venire con te, rimango a letto”. È interessante notare anche la trasposizione del doppiaggio inglese, dove l’espressione romanesca “col coso” viene tradotta con “with the thingamajing” che in gergo vuol dire “aggeggio”, “coso”, “affare”, mentre la parte “te raggiungo col cazzo” viene resa con “I’ll follow you in the absolutely not”, sempre per esprimere assoluta negazione. Anche in questo caso Google Traduttore & company non riescono a svolgere un buon lavoro di traduzione e rendono la frase con “I’ll join you with the thing [...] I’ll join you with my dick” facendo un lavoro di trasposizione letterale che in inglese non ha alcun senso.

Nell’ultimo esempio del quinto episodio, la frase originale “[...] pure se non lo strappi quello se ciancica” fa riferimento ad un foglio, dove il verbo in forma dialettale “ciancicarsi” significa “stropicciarsi”, “sgualcirsi”. Questo viene correttamente reso in inglese con il verbo “to crumple” che ha appunto questo significato. Questa volta il traduttore automatico non ci prova nemmeno: il verbo viene riportato esattamente com’è scritto in italiano, dando vita ad una frase scritta per metà in inglese e per metà in italiano, per altro dialettale. Pensiamo a cosa potrebbe capirci un madrelingua inglese: assolutamente nulla.

Episodio 6

	Testo originale	Adattamento inglese	TA
1.	Te lo sapevi che Alice forse mezzo me batteva i pezzi?	Did you know that Alice kind of liked me?	Did you know that Alice beat my pieces?
2.	Lella	Gay	Lella

Nell'ultimo episodio della serie, troviamo la traduzione di due esempi. Il primo è un'espressione romana molto precisa, che potrebbe risultare di difficile comprensione anche per un pubblico italiano che non risiede a Roma o nei dintorni. L'espressione romanesca "battere i pezzi" ha il significato di "corteggiare", "provarci con qualcuno", la quale viene tradotta con il verbo "like" per rendere il significato. La traduzione automatica, non essendo a conoscenza del significato di questa espressione, la traduce letteralmente con "beat my pieces", dando vita ad un'espressione del tutto priva di senso e incomprensibile.

L'ultimo esempio proposto riguarda il linguaggio gergale relativo all'omosessualità, in cui si utilizza il termine "lella" per indicare una persona lesbica. In inglese, viene usata la parola "gay", mentre la traduzione automatica lascia il termine inalterato, presumendo che si tratti del nome proprio di una persona.

Conclusioni

In questo progetto di tesi, è stato analizzato l'impatto che la tecnologia e l'intelligenza artificiale hanno avuto sulla pratica della traduzione, rendendola una disciplina sempre più digitalizzata. Sono stati presi in esame diversi esempi pratici oltre che teorici che sono stati presentati nel dettaglio attraverso i tre diversi capitoli.

Nel primo capitolo si è voluta esplorare tutta la storia e l'evoluzione della traduzione automatica, partendo dagli anni Quaranta del Novecento con la pubblicazione del Memorandum dello scienziato e matematico Warren Weaver, passando per gli studi contrastanti dello studioso Bar-Hillel e il rapporto ALPAC, fino ad arrivare ai giorni nostri con i più recenti sistemi di traduzione online. Questi ultimi sono basati sul moderno approccio di tipo neurale, che ricrea il lavoro di una mente umana, sviluppato verso la fine del 2014. Prima di esso, i servizi di traduzione online erano basati su sistemi a regole o su approcci statistici.

Nel secondo capitolo si è poi passati ad un'analisi dettagliata dei traduttori automatici presenti su Internet, quali Google Translate, Microsoft Translator e DeepL, di cui sono state descritte tutte le potenzialità e funzionalità. Si è poi parlato di sistemi di valutazione qualitativa di traduzioni automatiche quali BLEU, MQM, DQF che, basandosi su diversi parametri come fluidità, stile e terminologia utilizzati, sono in grado di misurare l'efficacia e la qualità della resa traduttiva. L'ultimo paragrafo del secondo capitolo ha rappresentato il fulcro vero e proprio della tesi, poiché ha preso in analisi tutte le limitazioni e gli svantaggi a cui si va incontro se si fa uso della traduzione automatica. Si è poi concluso il capitolo con un breve excursus riguardante le competenze che, invece, ha un professionista del settore ben formato da un punto di vista lavorativo.

Il terzo capitolo è stato dedicato ad un'attenta descrizione di elementi quali l'ironia, il dialetto, il turpiloquio e lo humor, presenti in grandi quantità nella serie televisiva Netflix "Strappare lungo i bordi" del fumettista Zerocalcare. La serie, scritta in dialetto romano e ricca di elementi comici, ha fornito uno spunto interessante per la dimostrazione pratica della limitatezza della traduzione automatica. Tramite il confronto tra l'adattamento inglese proposto dalla serie e la traduzione affidata ai sistemi di traduzione online descritti nel secondo capitolo, è stata messa in luce la resa poco soddisfacente di questi elementi specifici, che quindi, deve essere affidata

esclusivamente a traduttori umani in grado di proporre un risultato di qualità che rimanga fedele al testo originale.

Nel corso di questa ricerca, ho scelto di esplorare il mondo della traduzione automatica non solo come un argomento di interesse accademico, ma anche come una riflessione personale sulla mia esperienza di traduttrice. Attraverso questo percorso, ho acquisito una consapevolezza ancora più profonda delle sfide e delle complessità che caratterizzano il mondo della traduzione, che non è solo una questione di trasferimento meccanico di parole da una lingua all'altra, ma richiede una profonda comprensione delle culture, delle sfumature linguistiche e dei contesti specifici.

Come traduttrice, mi rendo conto di quanto sia fondamentale e irrinunciabile il ruolo del traduttore umano: le macchine possono svolgere compiti ripetitivi e automatizzati, ma mancano dell'intelligenza, della sensibilità e della comprensione profonda che solo l'essere umano può offrire. Anche di fronte ai progressi straordinari nell'intelligenza artificiale, la capacità di interpretare, adattare e creare in modo flessibile rimane un tratto distintivo dell'umanità.

Pertanto, ritengo che nessun tipo di macchina o artificio potrà mai sostituire l'essenziale contributo del traduttore umano: il suo ruolo rimane imprescindibile nel garantire una comunicazione autentica, efficace e significativa tra culture e lingue diverse. La traduzione automatica può essere un valido ausilio, ma è solo attraverso l'intervento umano che la vera essenza della traduzione può emergere e dare il suo pieno valore alla comprensione e alla condivisione tra i popoli del mondo.

Introduction

The practice of translation has always played a key role in progress and communication and, with more than 6700 languages spoken in the world, it is therefore essential to break down language barriers through this discipline. Due to the rapid technological evolution of the last few years, also translation has undergone a significant transformation. Whereas in the past it was considered a 'primitive' practice, today it is transforming into an increasingly technological discipline that aims to provide a translation in record time with a simple *click*.

Machine translation presents itself as a potentially revolutionary resource for breaking down language barriers and facilitating the exchange of information between individuals and different cultures, but despite its significant progress, it continues to show obvious limitations that undermine its effectiveness.

The aim of this thesis project is to analyze the limitations of machine translation and underline the importance of the human translator in intercultural and interlingual communication. The birth of machine translation will be analyzed in detail, and its evolution up to the present day will be described. Its progress has followed different technological approaches: from the first rule-based systems, we have moved on to statistical models and, more recently, to neural systems, which exploit artificial neural networks to learn data and improve performance, recreating the work done by the human brain. Today, many online services make use of this latter approach to translation in data processing.

Since not all machine translators are the same, a section of this thesis will be devoted to a detailed comparison between them, in order to highlight their differences and offer an overview of their potential. Furthermore, various evaluation techniques have been developed to measure the quality and accuracy of translations, which will be discussed in detail in the second chapter.

Subsequently, various reasons why machine translation is still very limited today will be outlined. Then some practical examples will be provided to prove these claims: to do so, a few scenes from the Netflix TV series "Strappare lungo i bordi" by cartoonist Zerocalcare, written in Roman dialect and rich in comic elements, will be selected. The choice is not random: the aim is to demonstrate how translating texts of this genre is a complex challenge, requiring not only linguistic skills, but also a deep cultural

background and an in-depth knowledge of cultural nuances that is sometimes complex even for a human mind. Proposing a faithful translation that not only conveys the literal meaning, but also preserves the humor and cultural context of the original text is a daunting task, which a machine translation would not be able to do, as it lacks the necessary knowledge to offer a satisfactory rendering.

With this thesis project, I wish to demonstrate how machine translation is still very limited today, despite the great evolution it has undergone in the last period. At the same time, however, there are areas in which technology can be positively integrated into the translation process, for example through the use of CAT tools, which exploit their potential to improve the efficiency and productivity of the human translator. Through their use, in fact, the task of professionals is facilitated, transposition times are optimized and there is a basis from which to start.

However, the systems found online today are not yet able to capture the nuances of meaning that only a human being, with the necessary studies, is capable of understanding, and it is therefore good to emphasize the irreplaceable importance of the human translator. Through the analysis of personal experiences and challenges encountered in the field of translation, an attempt will be made to provide a significant contribution to the understanding of this phenomenon, which is increasingly relevant in the context of linguistic and cultural globalization.

CHAPTER ONE

MACHINE TRANSLATION

1.1 The birth of machine translation

The term "Machine Translation" (MT) refers to the process by which artificial intelligence is employed to automatically translate a text from one language to another, without any human intervention. It is an area of computational linguistics and translation science, which deals with the analysis and implementation of computer programs for machine translation of texts.

The idea that a machine could translate human language was considered unrealistic for a long time. However, in the 20th century, the first concrete research in this field began to take shape. The first attempts date back to the 1930s, specifically 1933, when Pyotr Petrovič Smirnov-Trojansky, a Russian engineer and inventor, theorized a system for creating an automatic bilingual dictionary based on the interaction between human and machine. Although he made the suggestion that human translation could be mechanized, Trojansky's ideas did not receive proper recognition and his project failed, partly due to his early death. However, his role was fundamental to subsequent studies in the field of machine translation and his idea laid the foundation for the development of further theories in this area. It was in the early 1950s, more precisely in 1949, that a real breakthrough in the field of machine translation occurred with the publication of the memorandum entitled "Translation," authored by the American scientist Warren Weaver, director of the Rockefeller Foundation's Division of Natural Sciences. He was considered the father of Machine Translation precisely because he was the first to consider the potential that computers could have in this field and to set translation goals before anyone else could have an idea. The 1949 memorandum was structured as follows: it opened with a brief introduction on the goals that were going to be addressed, followed by four parts explaining step by step how to actually translate a text with the help of technology.

First of all, he focused on the problem of the context of a text, specifying that a word, taken alone without considering the whole, could have an incorrect meaning comparing with the one obtained by looking at the whole sentence. His second proposal involved overcoming the problem of logical comprehension of language by

an electronic computer, and to do so he used the theorem of McCulloch and Pitts (1943), developed in the context of research on mathematical modeling of the neural structure of the human brain according to which a robot constructed with regenerative loops of a certain formal character is capable of deducing any legitimate conclusion from a set of premises. From this assumption, however, Weaver excluded literary translation because it was full of elements without formal logic such as personal styles, expressions, metaphors, sayings, puns that could not be summarized in formal rules. The third assumption concerned the possible applicability of cryptographic methods. Weaver was impressed by cryptography based on, as he put it, "letter frequencies, letter combinations, intervals between letters and letter combinations, letter patterns, etc. that are somewhat independent of the language used."

For his fourth proposal, Warren Weaver became more utopian: he thought that there could be logical elements common to all languages that, when thoroughly analyzed, could help identify absolute linguistic patterns and structures useful in developing a mechanization of translation. After the publication of the memorandum there was a dual reaction: on one hand, many people rejected the idea that mechanization of translation could ever be possible, because of its complexity in understanding the many nuances of a meaning; after all, this is the thinking that many professional translators still have today. On the other hand, there were people who were more positive towards this innovation and were less hostile towards the goals proposed by the American scientist. An early criticism came from the Israeli mathematician, philosopher and linguist Yehoshua Bar-Hillel, who immediately extinguished the dreams of Warren Weaver and all the scientists who pursued the same goal as him: according to him, a translation entrusted entirely to a machine, (FAHQ MT), could not exist. To prove this, he developed a paper entitled, "A Demonstration of the Non-Feasibility of Fully-Automatic High Quality Machine Translation." He believed, therefore, that machine translation could in no way be equated with that of a professional translator and that research should be directed towards more applicable goals such as, for example, mixed MT, a translation done by a computer, but with the help of a specialized figure in the field of translation who would intervene either before or immediately after the automatic process so that it could arrive at a well-rendered transposition.

To figure out whether or not to fund research for machine translation, in 1966 the U.S. government requested a report from the Automatic Language Processing Advisory Committee (ALPAC) to study its feasibility: the quality, cost and future prospects analyzed did not lead to a good result. The ALPAC report stated that there was no point in investing in machine translation systems because the quality was lacking, the costs were very high and the time of process was excessively long. Rather, it was proposed to invest in systems that could help specialized translators in shortening the transposition time. It was obvious that this report discredited the usefulness that machine translation was already losing and many funders cut research funding.

However, many scientific communities disagreed with the ALPAC's negative report, and many research groups were formed especially in Canada and Europe, given also the fact that the demand for English-to-French translations in Canada was growing more and more. The most interesting development of the 1980s was the spread of the first commercial machine translation products: SYSTRAN and LOGOS. Having enjoyed significant economic funding in the previous years, the two systems were now ready to introduce themselves to the market. The SYSTRAN program was adopted not only by the European Community, but also by large corporations such as NATO, General Motors, Dornier, Aérospatiale and Rank Xerox. At the same time, the LOGOS Development Corporation, founded in 1969, was expanding the number of clients and was working on more language pairs. While SYSTRAN is still in the market today, LOGOS, following some economic difficulties, closed down for good in 2000.

Always in the 1980s, most of the activities concerning TA took place mainly in Japan, where large IT companies developed translation systems from Japanese to English and other way round, mostly concerning the IT sector. In the 1990s, further researches began to circulate in the area of automatic speech translation through new technologies for speech recognition and speech generation. During this period, it was developed a more realistic approach aimed at listening the needs of: first, the possibility that it could be created a technological tool that could translate like a real human being was abandoned. The idea was to provide the public with practical help with the translation process and, for this reason, machine translation systems, assisted translation, translator work environments, translation memories and online systems provided over the Internet were born. Developments in the field were, however, mainly aimed at

professional translators who, in this way, could shorten working time and facilitate work that, in some cases, required a great deal of research and in-depth study of the topics covered. The first platform created for translators was the Translator Workstation, consisting of two main components: a bilingual lexical database system and a system that creates bilingual text archives. The translation process worked in this way: first, the source text was loaded onto the system and the language of the target text was selected; at this point the workstation would search for equivalences between the two languages and, when it found them, it would create a link with the corresponding words or phrases in the chosen language. Matching still proceeded word by word until the end of the text was reached. At the end of the process, the result was shown to the user in terms of the number of "matches" found between the two languages. The operation was considered to have failed if the number of equivalences was less than a specific percentage of the total source text. Once the translation was processed, it was recorded in the system, which could then reuse it for future transposition work.

1.2 Rule-based, Statistical and Neural Machine Translation

The previous systems of Machine Translation mainly used the Rule-based MT model, also known as the rule-based approach that operated through three different strategies: Direct Translation, Transfer-based Machine Translation and Interlingua-based Machine Translation. The Direct Translation method translates the source text into the target language through very simple techniques, so much so that it is called a "first generation" system. This type of system produces a word-for-word translation through the use of a bilingual dictionary and syntactic analysis of the source text. This was considered a very limited method, since it translated only through correspondents between the two languages without going into further semantic or syntactic analysis. Nevertheless, it survived in the MT field due to its ability to produce fairly comprehensible results. Unlike the latter, the Transfer-based MT and Interlingua-based MT approaches are systems based on indirect translation and were termed "second-generation." Most machine translation systems dating back to the 1980s used the transfer approach, which also consisted of three stages. In the first there was the

transformation of the source language into its natural form. This stage laid the foundation for the second step in which a corresponding representation of the target language was created. At this point there was the so-called generation, which again converted the abstract representation of the target language into its natural form.

Many scholars of the time began to evaluate the problems presented by the Rule-based Machine Translation model and realized that it could not be a good alternative to human translation, so it was abandoned in favor of a statistical approach.

The Statistical Machine Translation system was first presented by British mathematician Philip J. Brown at the 12th International Conference on Computational Linguistics in Budapest in 1988. Its name derives from its statistical computational techniques in the search for correspondences between languages. The statistical approach is based on data-driven and machine learning techniques in which no explicit linguistic knowledge is used. Instead of entering large amounts of grammatical, lexical and syntactic information into the system, which was a very complex and expensive process, the machine was provided with an archive of translations already performed in both languages under consideration so that the computer, through statistical calculation, can identify which word or locution has the greatest probability of being the correct one in the target language. The main source of linguistic information is provided by corpora, which are divided into parallel corpora and monolingual corpora. SMT operates through three different types of statistical approaches: the first based on a single word and called Word-based SMT, the second based on a sequence of multiple words, called Phrase-based SMT, and the last using syntactic models, called Syntax-based SMT. Although this statistical model achieved better results than the first-generation models, it still had limitations and seemed to fall in creating a well-done machine translation, faithful to the original.

Starting in 2010, Statistical Machine Translation was quickly abandoned due to new developments in the field of Deep Learning that created a new approach to machine translation, this time a neural one: Neural Machine Translation, first introduced in 2014. This is based on the principle of wanting to reproduce the biological neuronal activity of a human being through artificial neurons. In the same way that our neurons transmit information for processing complex concepts to our brain, so artificial neurons are stimulated to communicate the information they learn from an input,

thanks to a network, to produce an output. Artificial neural networks are made up of groups of neurons called layers, and they transmit them to the neurons they are connected to on subsequent layers to arrive at the final result (output).

The operation process of Neural Machine Translation consists of three stages:

- **TRAINING.** This phase, also deep learning, allows the system to assimilate a large corpus of bilingual translations and accumulate relevant information about the source and target languages. The Neural Machine Translation analyzes reference texts from both languages and forms some units, such as individual words or grammatical elements. This process, called word embedding, determines connections between these different units, placing similar concepts close together in multidimensional space.
- **ENCODER.** In this encoding stage, the word embedding of the previous stage is replicated. A vector representation for the machine to learn the linguistic and semantic combinations among the elements of the source text is created. Through the representation, the machine transmits the stimuli from one neuron to another.
- **DECODER.** After obtaining the vector image, the system interprets the values of individual units in the source language and converts them to an output in the target language. The goal is to produce an output that is as close to the "gold standard," which is a reference translation created by a professional translator.

A special feature of neural networks is the way they are able to learn from errors, correcting parameters during the learning phase. Although innovative, neural machine translation has some limitations such as inaccurate translation of technical or creative texts, as the mechanism is trained to translate on linear translation models.

However, this approach was adopted by all major players in the field of machine translation, such as Google Translate, Microsoft Translator and DeepL, which will be discussed more specifically in the next chapter.

CHAPTER TWO

THE LIMITS OF MACHINE TRANSLATION

2.1 The main systems of machine translation

Continuing advances in technology and rapid digitization of information have radically transformed ways of communicating. Today, tools such as books or newspapers can also be found in digital form (without the original paper version having been abandoned); communication takes place through the use of the latest generation of telephones, Internet and satellite television. Even oral communication has undergone a major transformation with the introduction of video calls and video conferencing thanks to new technological developments, which are becoming more and more numerous. The key element that unites all these changes is the advent of the Internet. In an increasingly globalized society, it is important that all language barriers that prevent us from communicating be broken down. For this purpose, the technology could be a solution, adopting new and increasingly popular machine translation systems, available online and free of charge. In the past, these services were used by a limited audience for business purposes, such as multinational companies or government agencies that needed a large number of translations. However, accessibility today is also opened to a large audience. Machine Translation services are programs developed for computers that translate without human intervention, some do so word for word, others, more sophisticated, are able to interpret sentences or longer periods more accurately.

The first significant impetus for the spread of online machine translation services came from the collaboration between Microsoft Internet and AltaVista, the famous search engine that introduced the first real-time online translation service, known as Babel Fish, in 1997. The translator was capable of translating 12 languages: Chinese, Dutch, French, German, Greek, Italian, Japanese, Korean, Portuguese, Russian, and Spanish- all from or into English, although some language pairs were not translatable between them. In 2008, the system was acquired by Yahoo and, after ups and downs, was finally shut down a year later.

Today, the most popular and widely used online translator in the world is Google Translate, which records nearly 500 million accesses per day. It was developed by

Google LLC, the U.S.-based online services company, and was launched on April 28, 2006. Originally, it was based on the statistical approach machine translation system, but in November 2016, Google announced that it was abandoning this approach in favor of the neural system, which, using Deep Learning techniques, promised to be more accurate and reliable than its predecessor. Google Translate can translate forms of text as well as multimedia content such as images, videos or spoken text. Specifically, it is specialized in: translating websites, documents, creating a personal phrasebook in which to save translations already made, listening to the pronunciation of words, translating emails, chats, messages or images. It is also specialized in using the "interpreter" function and in translating handwriting. Based on various statistics, it has been able to estimate that the most translated languages with Google Translator are English, Spanish, Arabic, Russian, Portuguese and Indonesian, while the country where the machine translation system is used the most is Brazil.

Another Machine Translation service that is very popular and used worldwide is Microsoft Translator, a multilingual cloud service that is part of the Microsoft Cognitive services. The first version of the system was developed between 1999 and 2000 within Microsoft Research; the translator is based on the "data driven" approach, so it relies on parallel texts that have already been translated. The site, also known to users as "Bing Translator", was launched in 2007 and provides free translations over the Internet. Three versions of the system are available: Microsoft Translator for computers, Microsoft Translator accessed through the browser and the app, for both Android and iOS operating systems. Between 2011 and 2018, it developed many functions: translating images using the camera, listening to the pronunciation of words, connecting two people who speak different languages, creating a phrasebook, and customizing settings to suit the needs of the users.

Today, Microsoft Translator is capable of translating more than 100 languages. However, Microsoft Translator is also still inaccurate in some aspects, especially in grammatically and semantically complex texts, but the system is trying to make the level of translations higher through the support of artificial intelligence.

The latest machine translation service on the Internet that will be taken as subject of study in this chapter is the system of DeepL. It is part of a German company that developed the translation system in 2016 and launched it in 2017. Its founder and CEO

is Polish Jaroslaw Kutylowski, now living in Germany, who holds a doctorate in computer science. The latest innovation introduced by the system in January 2024, is the addition of the Arabic language, the fifth most spoken language in the world and the first language supported by DeepL that is written and read from right to left. The German-originated system offers several plans to users: the basic one and others at a cost, called DeepL, divided into three subscription plans: the Starter, the Advanced and the Ultimate level, respectively costing \$8.74, \$28.74 and \$57.49 per month.

Let's see what are the main functions that DeepL provides, similar to those of the other two machine translation systems, but with a few more features. DeepL Translate allows users to: translate texts, listen to the pronunciation of words, select alternative words to the proposed translation, translate entire files while keeping their original formatting, choose a formal or colloquial tone (only with the pro subscription), create a glossary, integrate DeepL into CAT tool systems. The site boasts more than 10 million monthly active users.

In terms of functionality, Google Translate and Microsoft Translator share approximately the same functions, while DeepL offers more solutions, even if for a fee. It can be said that the latter is more useful for business professionals, translators and interpreters, since it also makes available the use of CAT tools combined with the machine translation system.

Other assisted translation tools are CAT Tools. The term comes from the acronym Computer Assisted Translation or Computer Aided Translation and stands for all those computer tools used by translators useful for the translation process. These programs are based on the use of translation memories, which retain elements that have already been used in previous translations and direct towards a faster and higher quality translation process. A wide variety of CAT Tools systems are available on the market. There are paid and free CAT services; the most popular and widely used paid ones are:

- SDL Trados Studio, perhaps the most widely used. It is a comprehensive solution for translation, terminology management, editing and language quality assurance. It is available in several versions, with annual or permanent plans, for freelance translators;

- memoQ. Although less well known than the previous one, it represents another influential CAT Tool. It offers many features and it is compatible with Trados and other translation software. memoQ offers a free version for personal use or small projects, as well as a professional version with a permanent license;
- Wordfast is an integrated version of Microsoft Word for temporary translation of bilingual documents. It supports various file formats that Word can use. There is a more comprehensive solution with terminology management and translation memory, called Wordfast Pro;
- Déjà Vu, which offers advanced terminology and translation memory management features, with a highly customizable interface and support for a wide range of file formats, including Trados and memoQ.
- Across. It includes: translation memory, terminology management, quality assurance and project management. The CAT offers a free basic version for online use and a premium version with a monthly subscription plan;
- XTM. This system is available in two different versions: a traditional software license called XTM Suite or an online subscription solution called XTM Cloud;
- Star Transit. This CAT Tool differs from the others because it also retains the overall structure of the original document in its translation memory. For this reason, there are no sections isolated from context, allowing the original context of a given section to be viewed and checked at any time. This ensures an understandable solution without out-of-context sections. Star Transit offers monthly or annual subscription options;
- Poedit, which translates with interfaces and can be integrated to work on Wordpress;
- MetaTaxis. Despite being a paid system, it boasts a significantly lower price than its competitors, even for its advanced version;
- Pairapharase, a web-based system that combines machine translation with human translation. It presents, in addition, a voice over feature. There is a free version available, but with limited functions, and several versions with monthly subscriptions.

The free CAT Tools available on the market are:

- CafeTran Espresso, a program that can be used on Windows, Mac or Linux operating systems. It is also compatible with other major translation programs. This system is free until the translation memories or glossaries reach their maximum capacity;
- OmegaT, perhaps the most famous free CAT. It is an open-source program and is compatible with Windows, Mac and Linux operating systems;
- Similis, less well known, supports HTML, XML, PDF, MS Word, OpenOffice and Trados files;
- Wordfast Anywhere. It is used online and provides connections to public translation databases;
- MateCat. With the access to machine translation databases it provides to users, it can improve the speed of work;
- SmartCAT, which provides a matching platform between translators and clients;
- Virtaal, uses the Python programming language and includes translation management tools and access to machine translation;
- Pootle. Unlike most CAT Tools, it is specifically designed for translation of applications rather than documents;
- Lokalize, geared toward productivity and quality assurance;
- GlobalSight, which simplifies and manages the localization process. It can be used by project managers, translators and publishers;
- gtranslator, a specialized program for software translation.

There are numerous advantages by using translation memories and CAT Tools: for example, the risk of leaving parts untranslated is reduced since CAT Tools do not allow segments with the target text absent. Also, terminological and stylistic uniformity of the text is ensured: this is an essential aspect because it is of great importance that a text is translated with a terminological accuracy for the understanding and enjoyment of a product. In addition to these aspects, CAT Tools are able to identify repetition and concordances, export the translation in the same format as the original, quickly calculate the number of translation folders and ensure accuracy in patent translation.

However, it is important to say that when we refer to CAT tools for computer-aided translation, we are not talking about machine translation systems, which were discussed earlier. In fact, there is a difference between these two, which often goes unrecognized: while CAT assisted translation tools only help the translation processes by using databases and suggesting fragments of previous translations, the machine translation completely entrusts a machine or its algorithms with the control and final result of the translation; the results, in this case, are less accurate since there is no human intervention to supervise or correct the translated text. In fact, if the translation were completely entrusted to automated systems, for the careful post-editing, discussed in the next paragraph, a translator would be essential.

2.2 Pre-editing, post-editing and controlled language

The advancement of machine translation and its integration with other technologies have led to increasing adoption of these processes. The use of machine translation can make an important contribution in shortening working time, but often the quality of the translation is sacrificed. Three different processes come to the rescue to avoid this: pre-editing, post-editing and controlled language. It is stated that human intervention is essential for a good quality result and it can occur before machine translation with the pre-editing stage, or after translation into the target language, during the post-editing stage. Pre-editing is a revision process that aims to obtain an initial text free of ambiguities and complexities that may make it difficult for machine translation engines to process. The main goal is to intervene before the automatic process to reduce the revision time of the final text. This phase is monolingual and focuses only on the initial text, without the need to know the target language. It is essential for the pre-editor to be aware of the anticipated difficulties that the automatic translators might encounter and which would, therefore, negatively affect the final result. In this way they can be eliminated or reworked to produce a rough translation that can be easily revisited. Editing generally includes correcting spelling and grammatical errors, clarifying ambiguous or complex parts through explications, simplifications or normalizations. Another form of pre-editing involves the use of so-called "controlled" or "simplified" languages, developed primarily for industrial applications to improve

the understanding of technical and specialized documents. These languages, derived from natural languages such as English or German, aim to make technical documentation easily understandable and reduce the chances of misinterpretation by readers. Examples of controlled languages include Caterpillar-Fundamental English, Siemens Technical German, Multinational customized English, and Perkins Approved Clear English. Few languages have been developed specifically for machine translation, such as Caterpillar Technical English or Controlled Language Optimized for Uniform Translation (CLOUT). The latter is based on ten basic rules:

1. Write sentences shorter than 25 words.
2. Express only one idea per sentence.
3. Repeat the same sentence to express the same concept.
4. Write grammatically complete sentences.
5. Use simple grammatical structure.
6. Use the active form of sentences.
7. Repeat the noun instead of using a pronoun.
8. Use articles to identify nouns.
9. Use words from a general dictionary.
10. Use only words with correct spelling.

The benefits of using controlled languages in translation are that they enable the editing of texts that are easily processed by automated systems and improve the quality of the final result in terms of comprehensibility, making the post-editing phase more efficient. However, a significant disadvantage is that authors are very limited in their text editing and take a long time to learn the necessary rules. In general, research on pre-editing and the use of controlled languages has received less attention than studies on post-editing, due to the fact that they are often considered an optional phase.

Post-editing, on the other hand, has found much more of a place in the field of translation revision: it is the revision and adaptation of automatically translated texts. Over time, this practice has been described as "the correction of machine translation output by human linguists/editors" according to Veale & Way in 1997, "the checking, correcting and revising of translations made by any kind of translation automation"

(Gouadec 2007), and finally "the correction of machine-generated translation output to ensure that it reaches a level of quality agreed upon in advance between client and post-editor." Two levels of post-editing are distinguished: light post-editing, also called fast post-editing, gist post-editing, rapid post-editing, and full post-editing, also known as conventional post-editing. Post-editing consists of three stages: reading the original text, analyzing the automatically translated text to identify the parts that need to be corrected and revising to produce the final text, making the necessary changes. Light post-editing involves minimal revision of the automatically translated text, focusing on essential corrections: these corrections concern errors that compromise the sense of the text and, consequently, its correct interpretation by the recipient. The end result should therefore be an understandable text, achieved with minimal effort, correcting only those lexical, grammatical and syntactical errors that prevent the accurate transmission of the information contained in the original text. Full post-editing, on the other hand, is a more thorough revision intervention on the automatically translated text, the end result of which is a publishable text ("outbound translation"), theoretically indistinguishable from those translated by professional translators. This process is slower than light post-editing and requires targeted corrections to ensure maximum accuracy in transferring the message from the original text to the translated text. Is a crucial step in determining whether it is cost-effective to use a machine translation system or whether it is preferable to opt for a translation from scratch.

The activities and type of post-editing to be chosen depend on a number of factors: the length of time of use of the translation, the type, length and "visibility of the document," the delivery time, the needs and expectations of end users, and the ability of readers to use less-than-perfect texts. The time of use of the translation and the ability to accept inaccurate translations are key factors in choosing between light and full post-editing. For example, if the translation is intended only for general understanding with a short time of use and if the rough translation is not sufficient for this purpose, light post-editing may be the right a solution. However, it is important that the reader does not expect qualitative excellence and is aware of the limitations of rapid text revision.

2.3 The limits of AI and the importance of human translators

The advancement of technologies in the field of machine translation has changed the way languages are translated and interpreted. Machine translation systems are often used when there is the necessity of a quick and immediate feedback, when one wants to know summarily the meaning of an email or website in another language. In most cases, the machine translation we find on the Internet, such as that offered by Google Translator or DeepL, is used because it is free. Despite the great progress to improve these systems, for example by moving from a statistical approach to a neural one, the limitations that they have are many and important, and a solution to its obstacle will probably never be found. In this section we are going to analyze some of the main limitations and demonstrate how the figure of a professional translator is necessary for a quality result.

First of all, a machine translation system is an inaccurate tool that fails to grasp the real needs of the person requesting a translation and the result the latter wishes to achieve. It fails to capture the nuances of meaning: translating does not mean merely transposing the meaning of a word from one language to another, but it means considering the whole, that is, the cultural context and linguistic nuances. Only a human translator is able to do this, thus to interpret and transfer into the target language what is expressed in the source language, perhaps using other words but always remaining faithful to the original text. Another problem with the use of these translators is that they are unable to distinguish homophones, that is, words that are spelled and pronounced the same way but have a different meaning. This occurs because they cannot analyze the context in which these terms are used and often confuses the meanings of the words, misreading and misinterpreting the text. Artificial intelligence does not consider changes of register in the text: for example, if a text is written in a colloquial register with the use of spoken language it will have to be translated differently than, for example, a legal text that requires a higher register and the use of formal, context-appropriate words. Consequently, it does not return professional translations as opposed to the way a specialist in the field would. The automated system works in this way because it uses a statistical basis and, therefore, does not provide for an actual reading of the source text, but merely translates by

following examples of texts or words in its database, which it then reuses in new translations. A human being, on the other hand, can work even on what is entirely new or improbable, and can do so by leaning on an experience of language that goes beyond mere words already spoken or heard.

Taking Google Translator as an example, it can be seen that this is relatively reliable in the English language, which is the most widely used and on which the system has managed to create a kind of experience base to be able to translate effectively, especially when it comes to texts with simple syntax. On the other hand, in those languages that are less widely used and have more complex syntax, such as German, the algorithm is still far from satisfactory, not having a sufficient base on which to "learn" and train.

Another issue of fundamental importance is that of security. One of the main reasons why Google Translator is so popular with users around the world is certainly the fact that it is free. In return, however, Google requires users to use their data to constantly feed and refine its algorithm. This means that even sensitive information, such as lab test results, highly confidential financial data, or medical records, which entered into Google Translator to get a quick translation, end up being included in the huge container of data that Google uses. Google's Privacy Policy, in fact, states that "We use a variety of technologies to process your information for the purposes stated. We use automated systems that analyze your content [...]. We also use algorithms to detect patterns in the data. For example, Google Translator helps people communicate in multiple languages by detecting common language patterns in the sentences you ask it to translate."

In addition to translators and interpreters, voice actors also face the "struggle" with artificial intelligence today. In the field of dubbing, this can be used to create synthetic voices that could potentially compete with human voice actors in some contexts, especially for low-budget projects or where a generic voice is required, thus reducing the need for human labor at some stages of production. Despite the progress, there is no shortage of problems for this sector as well. The main one is certainly one related to privacy and security: an individual's voice can be synthesized by AI, falsifying audio that could be used fraudulently. Another limitation to mention is the fact that artificial voices, while well-developed, cannot compare to human voices, which are able to

restore emotional depth and authenticity: in fact, professional voice actors can offer rich interpretations and a more genuine connection with the audience.

The need to involve translators more closely in the debate on improving technologies was also highlighted during the recent Machine Translation Summit XVII. For the first time, a session dedicated exclusively to the world of translation and translators was organized, during which topics such as the measurement of productivity, its impact on the adoption of machine translation, the role in the work of the translator, ethical and confidentiality issues related to the use of technologies, and psycho-social aspects such as psychological predisposition to the use of technologies and biases were discussed.

It is evident that technology is and will continue to profoundly change the role of the translator. Certainly, the use of technology simplifies the latter's task, enabling him, for example, to avoid complete translation from scratch, and instead rely on machine-generated pre-translated texts, on which he can then make any improvements in the post-editing phase. Especially in a professional context in which the profession of translators and interpreters has undergone significant technological evolution with the use of the advanced technologies, one wonders what the essential skills and abilities for these figures should be in an increasingly digitized age. When approaching a translation process, it is essential to carefully consider both the source and target contexts. For example, there is increasing talk in the professional world of "transcreation," a translation process that lies between translation and copywriting and that requires an articulated mix of skills, especially the creativity of the translator. Thus, in situations where it is necessary to completely change the name of a product or slogan for example, the process of transcreation is employed, which is seen as a form of "creative translation," since it involves adapting the message from one language to another, without necessarily following a literal word-for-word translation. This emphasis on the creative aspect of translation keeps translators at the center of the action, at least for the time being, preventing the adoption of machines in the "transcreative" process. Both localization and transcreation represent more sophisticated forms of translation that cannot be performed by machine translation systems, but require human intervention to perform accurately and adequately.

Even in post-editing, it is evident how the figure of a specialized translator is essential, since a professional is the only one capable of judging the accuracy of a translation by

detecting any errors in machine translation. In addition, he has in-depth knowledge of conceptual transfer between languages and has technical resources that he knows how to use when in doubt.

Language skills are also essential for any professional translator, as are basic skills. The latter are specifically oriented towards attitudinal or psycho-physiological skills that include, for example, the ability to process and apply post-editing rules to translation, the management of client expectations of final quality, and the ability to make informed decisions about which post-editing strategies to adopt. Certainly, another key aspect for the translator, more practical, is instrumental competence, which concerns the technical skills needed to consciously and accurately use technologies. These skills involve knowledge of the various machine translation systems and their features, terminology management, dictionary maintenance, assessing the quality of a corpus, pre-editing through the use of controlled languages, and, finally, programming skills.

Although all these skills may appear generic and somewhat conventional, they are indispensable elements for a flexible professional figure called upon to work in an environment where technologies are rapidly and constantly evolving. The ability to constantly learn, moreover, is essential to enable the translator to readily adapt to technological advances.

From the analysis just made, we can see that despite advances in AI, language mediation still requires specific human skills to ensure cultural understanding, contextual adaptation, and sensitivity to linguistic nuances. Understanding a different culture, sensitivity and the ability to adapt to the needs of different users remain key aspects that only a human mediator can provide. A translator or automated support can be valuable supporting tools through which communication and language mediation can be improved, facilitating translation, interpretation, and multilingual interaction, and can prove to be a useful tool for softening language barriers. However, it becomes essential to maintain a balance between the use of artificial intelligence and human intervention to ensure accurate and high-quality results in interlingual communication: human presence remains, however, essential for effective cross-cultural communication.

2.4 The adaptation of dialect and humor

Machine translation has revolutionized the way we interact with texts written in different languages, offering an immediate solution for understanding the general meaning of a message. This technology has proven particularly effective when it comes to translating texts of a more general nature, which may be emails or websites, for example. However, when confronted with texts that are more specific or that request a specialized knowledge of a cultural or scientific context, machine translation shows its limitations. In these situations, the importance and of the role of the human translator emerges, who brings with him not only linguistic expertise, but also in-depth knowledge of the context and subtle implications of the original text. In meeting this challenge, human translators demonstrate their ability to capture complexity and richness of meaning. One of the many contexts in which the limitation of machines in translation emerges is dialect and humor, two elements that escape the understanding of machine translators. Before going on to analyze some practical examples to show how difficult, or even impossible, it is to translate these two linguistic variations for an artificial translator, we are going to illustrate their main characteristics by trying to understand why they are so complex to transpose, even for a human translator.

Starting with dialect, the term refers to those language varieties spoken alongside a major language, the official language, and used in informal settings. The number of speakers is generally restricted to a particular area, although there are cases where a dialect has managed to spread beyond its socio-linguistic boundaries. In linguistics, the concept of "dialect" is defined as a variant of a language, but it can sometimes be difficult to distinguish whether an idiom is considered a language or a dialect. To resolve this issue, three basic parameters can be considered by comparing two speakers of different idioms:

1. Mutual understanding: if the two speakers are able to understand each other when they speak, then they are probably two dialects of the same language, since there is mutual intelligibility between them;
2. Common basic vocabulary: if more than 80 percent of the commonly used words used by the two speakers are the same, although pronounced differently, then the two idioms can be considered as dialects of the same language;

3. Homogeneous morphology and syntax: if the two speakers follow the same grammatical rules to express themselves, then it is likely that they are speaking two dialects of the same language.

These parameters provide a useful framework for distinguishing between languages and dialects, enabling a better understanding of the nature of the language varieties found in different language communities. If we consider language as a color, dialects can be said to be defined as different gradations of the same color, and thus of the same language. From the sociolinguistic point of view, instead, several criteria have been identified to distinguish between language and dialect:

1. Geographical extent: while a language is spread over a vast territory, dialect is confined to a narrower geographical area;
2. Absence of standardization: dialect lacks a "correct" form universally accepted by speakers, resulting in a fragmented and local variety, with each community speaking it uniquely;
3. Social perception: dialect is often regarded by its speaking community as an informal and less prestigious language, associated primarily with the less affluent and less educated social classes;
4. Informal use: dialects are mainly used in informal social settings, such as with family or friends, while they are not commonly used in formal settings such as public institutions, schools, or workplaces;
5. Limited literary production: dialects are mainly oral and often lack a significant literary tradition;
6. Influence of dominant language: in places where it is spoken, dialect is often influenced by the presence of a dominant language, from which it receives linguistic borrowings and grammatical constructs;
7. Vocabulary limitation: dialects focus on vocabulary related to everyday life, often lacking terms needed to discuss scientific, technological, philosophical and other areas of knowledge.

In addition to dialect, irony also represents a linguistic tool that is quite complicated to translate. First of all, it is a form of communication, which aims to expose sarcastic content in an indirect way. It is represented by an interaction between verbal and nonverbal language, made by gestures or facial expressions. According to Muecke (1969; 1970), the ironic message should be clear but not immediately obvious and it is commonly used in humorous contexts or to achieve an emotional effect in communication. During the ironic speech process, sentences may take on specific characteristics such as a more nasal pronunciation, a decrease in speech speed and very emphasized syllables, resulting in a strong emotional emphasis in communication. Understanding of irony, which can sometimes be lacking, is facilitated by sharing a common background between the interlocutors and knowledge of the subject matter of irony. However, failure to receive the ironic message may result from a misunderstanding or a refusal to understand the irony itself.

Sometimes in irony, sarcasm or humor there may be the presence of vulgar language, with the use of foul or offensive terms and expressions. It is often present in many dialects or regional languages. The terms and expressions used often refer to various linguistic taboos that may relate to sex, aggression, religion, etc. The terms are rich in emotional value, because they often express human feelings such as anger, surprise, fear, amusement. The words can be used for expletives, to symbolically express an aggressiveness perhaps against inanimate objects; for insults, usually addressed to other people; for curses, in which the expressions serve to wish evil on the person addressed. Elements of vulgar language often encounter strong censorship when it comes time to translate them into another target language. Their translation is affected by several factors: first, repetitiveness, especially of expressions common in dubbing, and then originality, since during the "transcreation" phase (Section 2.3) substitutes may be found that deviate too much from the original language. In addition, there is the issue of correspondence, or lack of this one, between the source and target languages, both linguistically and conceptually, as well as possible restrictions imposed by the audiovisual context. Sometimes, then, translators may be restricted in their transpositions due to directives or requirements of production companies. However, the vulgar language is primarily a means of emotional expression, and just

like the humorous elements discussed earlier, it does not require a literal translation, but rather an equivalent in the target language.

Translating all these comic and dialectal elements is a complex challenge for today's translators, who are required to have not only linguistic competence but also a deep understanding of the source and target cultures, along with a dose of creativity to preserve the essence and emotional effect of the original elements. If this is already a tough task for professionals in the field, requiring long study and a lot of effort, one can only imagine what a machine translation system would be capable of, if it were entrusted with the task of transposing a text rich in linguistic, cultural, social, irony and comic nuances into another language, while leaving the sense of the original text and respecting what it wants to communicate.

All these elements are present in the Italian Netflix TV series "Strappare lungo i bordi" by Italian cartoonist Zerocalcare, released for streaming in 2021. The author had often been at the center of the Italian comics and literary scene through a series of participations in exhibitions and thanks to the publication of his many books and comics, to which he has been dedicated since the early 2000s. Michele Rech, his real name, has become one of the most acclaimed and influential cartoonists on the current Italian scene. He spent part of his youth in France before settling in Rome, precisely in "Rebibbia" neighborhood, a city that has profoundly shaped his artistic identity.

In his works Zerocalcare combines irony and depth, describing different and apparently irreconcilable themes. On one hand, his narratives explore personal issues, such as the process of growing up and maturing; on the other, they manifest a strong interest in social or political issues, or his sensitivity to the social problems of his generation. Moreover, Zerocalcare's works are permeated by a subtle self-irony and enriched by references to popular culture, elements that contribute to making him a transversal artist, capable of winning over an audience of different ages and interests. The cartoonist's works often draw inspiration from his life experiences, which are reinterpreted and fused with invented elements.

The translation of comic elements, vulgar language, dialect terms, and expressions referring to the sexual sphere present in his TV series, required a long work of adaptation from Italian to English. If we try to enter all these elements on automatic translators such as Google Translate, Microsoft Translator or DeepL, the rendering is

not at all satisfactory: many times, Italian expressions, terms, and idioms have been literally translated or not even translated; incorrect terms have been used that do not correspond to the true meaning. Because of automatic translation's inability to capture the true meaning of words and the socio-cultural context of the text, the result is a translation that is totally incomprehensible to an English audience. This is a practical demonstration of how crucial a human mind is to be entrusted with a translation, especially if, as in this TV series, it requires a great deal of knowledge and in-depth study of the language and culture of other peoples.

Conclusion

In this thesis project, the impact that technology and artificial intelligence have had on the practice of translation was analyzed, making it an increasingly digitalized discipline. Several practical as well as theoretical examples were examined and presented in detail through the three different chapters.

The first chapter explored the entire history and evolution of machine translation, starting in the 1940s with the publication of the Memorandum of the scientist and mathematician Warren Weaver, passing through the contrasting studies of the scholar Bar-Hillel and the ALPAC report, and ending in the present day with the most recent online translation systems. The latter are based on the modern neural approach, developed in late 2014, which recreates the workings of a human mind. Prior to this, online translation services were based on rule-based systems or statistical approaches. The second chapter then went on to a detailed analysis of automatic translators on the Internet, such as Google Translate, Microsoft Translator and DeepL, whose skills were described. It then went on to discuss quality assessment systems for machine translations such as BLEU, MQM and DQF, which, based on various parameters such as fluency, style and terminology, are able to measure the effectiveness and quality of the translation output. The last paragraph of the second chapter was the real focus of the thesis, as it analyzed all the limitations and disadvantages of using machine translation. The chapter was then concluded with a brief excursus on the skills that a well-trained professional in the field has.

Also, a careful description of elements such as irony, dialect, vulgar language and humor, which are present in large quantities in the Italian Netflix TV series “Tear Along the Dotted Line” by cartoonist Zerocalcare. The series, written in Roman dialect and rich in comic elements, provided an interesting cue for the practical demonstration of the limitations of machine translation. Through the comparison between the English adaptation proposed by the series and the translation entrusted to the online translation systems described in chapter two, the unsatisfactory rendering of these specific elements was highlighted, which must therefore be entrusted exclusively to human translators capable of proposing a quality result that remains faithful to the original text.

In the course of this research, I chose to explore the world of machine translation not only as a topic of academic interest, but also as a personal reflection on my experience as a translator. Through this journey, I gained an even deeper awareness of the challenges and complexities that characterize the world of translation, which is not just a mechanical transfer words from one language to another, but requires a deep understanding of cultures, linguistic nuances and specific contexts.

As a translator, I realize how crucial and indispensable the role of the human translator is: machines may perform automated tasks, but they lack the intelligence, sensitivity and deep understanding that only the human being can offer. Even in the face of extraordinary advances in artificial intelligence, the ability to interpret, adapt and create remains a distinguishing feature of humanity.

Therefore, I believe that no machine or artifice can ever replace the essential contribution of the human translator: his role remains indispensable in ensuring authentic, effective and meaningful communication between different cultures and languages. Machine translation can be a valuable aid, but it is only through human intervention that the true essence of translation can emerge and give its full value to understanding and sharing between the peoples of the world.

Introducción

La práctica de la traducción siempre ha desempeñado un papel clave en la base del progreso y la comunicación y, con más de 6700 idiomas hablados hoy en día en el mundo, se vuelve esencial romper las barreras lingüísticas a través de esta disciplina. En la era digital en la que vivimos, la traducción también ha sufrido una importante transformación: si antiguamente se consideraba una práctica "primitiva", en la que el traductor dependía exclusivamente del papel, el bolígrafo y diccionario para traducir textos, hoy en día esto se está transformando en una disciplina cada vez más avanzada tecnológicamente y pretende ofrecer una traducción en un tiempo récord con un simple *click*. La traducción automática se presenta como un recurso revolucionario para romper barreras lingüísticas y facilitar el intercambio de información entre diferentes personas y culturas, pero, a pesar de sus importantes avances, sigue mostrando claras limitaciones.

Este proyecto de tesis tiene como objetivo analizar los límites de la traducción automática y subrayar la importancia insustituible del traductor humano en la comunicación intercultural e interlingüística. Se analizará en detalle el nacimiento de la traducción automática, cuya historia tiene sus raíces en los años 40, para luego describir toda su evolución hasta la actualidad y que ha seguido diferentes enfoques tecnológicos: de los primeros sistemas basados en reglas, se ha pasado a los modelos estadísticos y, más recientemente, a los sistemas neuronales que recrean el trabajo realizado por el cerebro humano. Actualmente, numerosos servicios en línea, ampliamente distribuidos, utilizan este último enfoque de traducción en el procesamiento de datos.

Dado que no todos los traductores automáticos son iguales, una sección de la tesis se dedicará a una comparación entre ellos, con el fin de resaltar sus diferencias, potencialidades y rendimiento. Además, para evaluar su eficacia se han desarrollado diversas técnicas que permiten medir la calidad de las traducciones y que se analizarán en detalle en el segundo capítulo.

A continuación, una vez expuestas las razones que hacen la traducción automática todavía muy limitada, se aportarán varios ejemplos prácticos desde la serie italiana Netflix "Cortar por la Linea de Puntos" por el dibujante Zerocalcare, distribuido en streaming en 2021, escrito en dialecto romano y lleno de elementos cómicos. La

traducción de textos de este género es un desafío complejo, que requiere no sólo habilidades lingüísticas, sino también un profundo bagaje cultural y un profundo conocimiento de las culturas que, a veces, resulta complejo incluso para la mente humana. Proponer una traducción fiel que no sólo transmita el significado literal, sino que también preserve el humor y el contexto cultural del texto original es una tarea ardua, que una traducción automática no sería capaz de implementar porque no tiene los conocimientos necesarios para ofrecer un rendimiento satisfactorio.

Con este proyecto de tesis se quiere, por tanto, demostrar cómo la traducción automática sigue siendo hoy en día muy limitada, a pesar de la gran evolución que ha experimentado en el último periodo. Al mismo tiempo, sin embargo, se identificarán áreas en las que la tecnología puede integrarse positivamente en el proceso de traducción, por ejemplo, mediante el uso de herramientas de traducción asistida (CAT Tools), que mejoran la eficiencia y la productividad del traductor. De hecho, mediante su uso se facilita la tarea de los profesionales del sector y se optimizan los tiempos de transposición.

En cualquier caso, los sistemas que se encuentran hoy en línea todavía no son capaces de captar esos matices de significado que sólo un ser humano, con los estudios necesarios, es capaz de comprender y, por tanto, es bueno subrayar la importancia insustituible del traductor humano. Gracias al análisis de experiencias personales y desafíos encontrados en el campo de la traducción, se intenta dar una contribución significativa a la comprensión de este fenómeno, cada vez más relevante en el contexto de la globalización lingüística y cultural.

CAPÍTULO UNO

LA TRADUCCIÓN AUTOMÁTICA

1.1 El nacimiento de la traducción automática

El término Traducción Automática (TA) se refiere al proceso mediante el cual se utiliza inteligencia artificial para traducir automáticamente el texto fuente de un idioma a otro, sin ningún tipo de intervención humana. Se trata de un área de la lingüística computacional que se ocupa del análisis e implementación de programas informáticos para la traducción automática de textos.

La idea de que una máquina pudiera traducir el lenguaje humano se consideró durante mucho tiempo una fantasía imposible. Sin embargo, en el siglo XX comenzaron a gestarse las primeras investigaciones concretas en este campo. Los primeros intentos se remontan a la década de 1930, concretamente a 1933, cuando el estudioso Pëtr Petrovič Smirnov-Trojanskij, ingeniero e inventor ruso, teorizó un sistema para la creación de un diccionario bilingüe automático basado en la interacción entre hombre y máquina. Su proyecto era encontrar dos personas monolingües para traducir de un idioma a otro. La teoría de Trojansky, innovadora en aquel momento, era que todas las lenguas tenían una gramática común, llamada "gramática universal" por Chomsky que por tanto sería más fácil para la máquina traductora si el texto fuente sufriera un proceso de transformación desde un período complejo a una forma básica. En concreto, el proyecto constó de tres pasos: un análisis lógico-sintáctico de las palabras para descomponerlas en forma básica, la traducción de estas formas básicas en la lengua de destino mediante una máquina automática; finalmente, la transformación de estas formas básicas en oraciones complejas mediante un proceso realizado por un experto en el idioma de destino. Aunque sugirió que la traducción humana podía mecanizarse, las ideas de Trojansky no obtuvieron el debido reconocimiento y su proyecto fracasó, en parte debido a su temprana muerte. Sin embargo, su papel fue fundamental para estudios posteriores en el campo de la traducción automática y su idea sentó las bases para el desarrollo de futuras teorías y herramientas en esta área.

Fue a principios de los años 50, más precisamente en 1949, cuando se produjo un verdadero punto de inflexión en el campo de la traducción automática con la publicación del *memorandum* titulado "Translation", escrito por el científico

estadounidense Warren Weaver, director de la División de Ciencias Naturales de la Fundación Rockefeller. Fue considerado el padre de la traducción automática porque fue el primero en considerar el potencial que podían tener los ordenadores en este campo y en fijar objetivos de traducción antes de que nadie tuviera idea. El *memorandum* de 1949 se estructuró de la siguiente manera: se abrió con una breve introducción sobre los objetivos que se pretendían abordar, seguida de cuatro partes en las que se explicaba cómo traducir eficazmente un texto con ayuda de la tecnología. En primer lugar, de hecho, se centró en el problema del contexto de un texto, precisando que una palabra, tomada sola sin considerar el conjunto, podría tener un significado incorrecto en comparación con el obtenido observando la frase completa. Su segunda propuesta, en cambio, preveía superar el problema de la comprensión lógica del lenguaje mediante una computadora electrónica. El tercer supuesto se refería a la posible aplicabilidad de los métodos criptográficos.

Para su cuarta propuesta, Warren Weaver se volvió más utópico. Se basaba en la creencia de que podía haber elementos lógicos comunes a todas las lenguas que, analizados en profundidad, podrían ayudar a identificar estructuras lingüísticas absolutas útiles para desarrollar una mecanización de la traducción.

Tras la publicación del memorándum hubo una doble reacción: por un lado, muchos rechazaron la idea de que la mecanización de la traducción fuera posible, debido a su complejidad: al fin y al cabo, esta es la idea que todavía tienen muchos traductores profesionales en la actualidad. Por otro lado, hubo personas que pensaron más positivamente y fueron menos hostiles hacia los objetivos propuestos por el científico estadounidense.

Una primera crítica vino del matemático, filósofo y lingüista israelí Yehoshua Bar-Hillel, que inmediatamente apagó los sueños de Warren Weaver y de todos los científicos que perseguían el mismo objetivo que él: según él, una traducción confiada enteramente a una máquina (FAHQ MT), no podría existir. Para demostrarlo, escribió un ensayo titulado: "A Demonstration of the Non-Feasibility of Fully-Automatic High Quality Machine Translation". Por tanto, considera que una traducción automática no puede ser en modo alguno comparable a la de un traductor profesional, que es algo inalcanzable y que la investigación debe orientarse hacia objetivos más aplicables como la traducción automática mixta o la traducción realizada por ordenador, pero con

la ayuda de una persona especializada en el campo de la traducción para garantizar una transposición bien terminada. Esta euforia inicial de los años 50 pronto dio paso al desánimo, debido a los malos resultados obtenidos de la investigación. Los proyectos financiados anteriormente fueron duramente criticados, especialmente por quienes creyeron en ellos primero y los apoyaron financieramente.

En 1966 el gobierno de Estados Unidos solicitó un informe al “Automatic Language Processing Advisory Committee” (ALPAC) para estudiar la viabilidad de la traducción automática. La calidad, los costes y las perspectivas de futuro analizadas no condujeron a un buen resultado. El informe ALPAC afirmaba que no tenía sentido invertir en sistemas de traducción automática porque faltaba calidad, los costes eran muy elevados (casi el doble que las traducciones humanas) y los tiempos excesivamente largos. Más bien se propuso invertir en sistemas que pudieran ayudar a los traductores especializados y acortar los tiempos de transposición. Era obvio que este informe desacreditaba aún más la utilidad que la traducción automática ya estaba perdiendo y, por eso, muchos financistas del sector recortaron la financiación de la investigación.

Sin embargo, muchas comunidades científicas no estuvieron de acuerdo con el informe negativo de ALPAC y por eso se formaron muchos grupos de investigación, especialmente en Canadá y Europa. Una de las pocas empresas que sobrevivió al informe ALPAC fue SYSTRAN, System Translation, creada por Peter Toma, investigador de la Universidad de Georgetown, en 1968. Fue una de las primeras empresas en abordar la traducción automática. Otros sistemas de traducción comenzaron a extenderse en varias partes de Europa como el grupo GETA de Grenoble en Francia, el grupo de Saarbrücken en Alemania y la empresa BSO en Holanda. Otros grupos de investigación en el campo de la MT también se extendieron fuera del continente europeo, en Japón, Corea, Taiwán, China, la Unión Soviética y Malasia.

El desarrollo más interesante de la década de 1980 fue la difusión de los primeros productos comerciales de traducción automática: SYSTRAN y LOGOS. Después de haber disfrutado de una considerable financiación económica de los Estados Unidos en años anteriores, los dos sistemas estaban ahora listos para aparecer en el mercado. Mientras que SYSTRAN sigue presente en el mercado, LOGOS, tras algunas dificultades económicas, cerró definitivamente en el año 2000.

En la década de 1980, la mayor parte de la actividad comercial relacionada con la traducción automática se desarrolló principalmente en Japón, donde algunas grandes empresas desarrollaron sistemas de traducción del japonés al inglés y viceversa y se referían principalmente al sector informático.

El paso más importante dado en los últimos años fue cuando la comercialización de sistemas de traducción asistida por ordenador se diseñó para su uso en ordenadores y no sólo a través de grandes calculadoras que, además, tenían un coste muy elevado. Sin embargo, la evolución del sector se dirigió principalmente a los traductores profesionales, que de esta manera podían acortar los tiempos de trabajo y facilitar un trabajo que, en algunos casos, requería mucha investigación y estudio. La primera plataforma creada para traductores fue la “Translator Workstation”. El sistema se describió por primera vez en 1990. Su diseño tuvo en cuenta el hecho de que los profesionales de la industria necesitaban herramientas rápidas, fáciles de usar y relativamente económicas. La estación de trabajo constaba de dos componentes principales: un sistema de *database* léxico bilingüe y un sistema que crea y gestiona archivos de texto bilingües. En cuanto al proceso de traducción real, se avanzaba de esta manera: primero se cargaba el texto de origen en el sistema y se seleccionaba el idioma del texto de destino; en este punto la estación de trabajo buscaba equivalencias entre los dos idiomas y, cuando las encontraba, creaba un enlace con las palabras o frases correspondientes en el idioma elegido. Al final del proceso, el resultado se mostraba al usuario en términos del número de "coincidencias" encontradas entre los dos idiomas. Se consideró que la operación había fracasado si el número de equivalencia era inferior a un porcentaje determinado del texto fuente total. Una vez procesada la traducción, quedaba registrada en el sistema que así podía reutilizarla para futuros trabajos de transposición.

1.2 Rule-based, Statistical y Neural Machine Translation

La tecnología en la que se basaba la traducción automática de los años 1990 era un poco diferente de la que está disponible hoy: utilizaba principalmente colecciones de traducción ya realizadas en archivos, diccionarios y gramáticas que permitían analizar el texto de partida y crear el equivalente en el idioma de destino. Casi todos los

sistemas de traducción automática de aquella época utilizaban el modelo de traducción automática “Rule-based”, también conocido como enfoque de reglas, que operaba a través de tres estrategias diferentes: “Direct Translation” (traducción directa), “Transfer-based Machine Translation” y “Interlingua-based Machine Translation”. El método de *Direct Translation* traduce el texto de origen al idioma de destino utilizando técnicas muy sencillas, hasta el punto de que se define como un sistema de "primera generación"; se denomina sistema directo ya que la traducción se realiza directamente desde la lengua de origen a la lengua de destino sin pasos intermedios, como ocurre con los otros dos enfoques. Este tipo de sistemas produce una traducción palabra por palabra gracias al uso de un diccionario bilingüe y un análisis sintáctico del texto inicial. Este se consideró un método muy aproximado y sobre todo limitado, dado que se traducía únicamente a través de correspondencia entre las dos lenguas sin realizar mayores análisis semánticos o sintácticos. Todavía, fue capaz de producir resultados bastante comprensibles. A diferencia de este último, los enfoques de TA *Transfer-based* y *Interlingua-based* son sistemas basados en traducción indirecta y fueron definidos como de "segunda generación". El primero enfoque se basaba sobre tres paajes: en el primero se realizaba un análisis de la lengua fuente en el que se llegaba a su forma natural. Esta fase sentaba las bases para el segundo paso, la transferencia, en el que se creaba una representación abstracta intermedia correspondiente del idioma de destino. En este momento se voleaba a la generación, que devolvía la representación abstracta de la lengua de destino a su forma natural.

El último enfoque de RBMT es aquello basado en *Interlingua*. Por definición, los sistemas interlenguas son sistemas multilingües, es decir, en los que se crea una representación abstracta común a todas las lenguas, la interlengua, tras haber realizado un obligatorio trabajo de análisis léxico, semántico y sintáctico. Aunque la representación es neutra respecto al texto original, contiene toda la información necesaria para generar el texto en el idioma de destino. La idea era crear una interlengua universal, válida para todas las combinaciones lingüísticas y común a todas las lenguas, siendo independiente de las naturales.

Muchos estudiosos de la época comenzaron a evaluar los problemas que presentaba el modelo de traducción automática basada en reglas y comprendieron que éste no podía constituir una alternativa válida a la traducción humana. Los puntos débiles que

llevaron a su abandono definitivo fueron, en primer lugar, el gran trabajo de *post-editing* que requirió, en segundo lugar, la evidente incapacidad para traducir metáforas, modismos, formas idiomáticas y matices de significado y, por último, la imposibilidad de traducir dos lenguas que tienen estructuras completamente diferentes entre sí, como por ejemplo un Lengua europea y una oriental. Por todas estas razones, se abandonó el enfoque basado en reglas en favor de un enfoque estadístico.

El sistema de traducción automática estadística fue presentado por primera vez por el matemático británico Philip J. Brown durante la XII Conferencia Internacional sobre Lingüística Computacional en Budapest en 1988. El enfoque estadístico se basaba en datos (*data-driven*) y técnicas de aprendizaje automático (*Machine Learning*) en las que no se utiliza ningún tipo de conocimiento lingüístico explícito. En lugar de introducir en el sistema grandes cantidades de informaciones gramaticales, lexicales y sintácticas, se dotaba a la máquina de un archivo de traducciones ya realizadas en ambos idiomas, de manera que el ordenador, mediante un cálculo estadístico, podía identificar qué palabra o frase tenía mayor probabilidad de ser la correcta en el idioma de destino. La principal fuente de información lingüística la proporcionaban los corpus, que se dividen en corpus paralelos y corpus monolingües. La traducción automática estadística operaba a través de tres tipos diferentes de enfoques estadísticos: el primero basado en palabras únicas y llamado “Word-based SMT”, el segundo basado en una secuencia de múltiples palabras, llamado “Phrase-based SMT” y el último que utiliza modelos sintácticos, llamado “Syntax-based SMT”.

En el primer modelo, la traducción se realiza mediante una correspondencia palabra por palabra entre los dos idiomas, utilizando la traducción a nivel léxico. Cada palabra en la lengua de origen tiene su contraparte en la lengua de destino y para la elección de estas el cálculo se basaba en un gran corpus paralelo. En el segundo modelo, el sistema dividió el texto en oraciones completas, intentando superar algunas limitaciones del SMT basado en palabras. Finalmente, en el último, se identificaba la estructura sintáctica de las oraciones, para mejorar la traducción entre dos idiomas con diferente orden de palabras como son los idiomas europeos y no europeos. Aunque este modelo estadístico obtuvo mejores resultados que los de primera generación, todavía tenía limitaciones: en primer lugar, demostró ser eficaz sólo para las combinaciones lingüísticas para las que había varias traducciones dentro de los corpus

paralelos. En segundo lugar, a menudo presentaba errores u omisiones en las traducciones, especialmente entre idiomas que tenían estructuras sintácticas diferentes. El sistema, entonces, también falló en la concordancia de género y número de palabras, siendo incapaz de captar las relaciones de dependencia entre las proposiciones.

Por lo tanto, este segundo tipo de enfoque tampoco parecía satisfacer los requisitos necesarios para la realización de una traducción automática bien realizada y fiel al original.

A partir de 2010, la traducción automática estadística se abandonó rápidamente debido a nuevos desarrollos en el campo del aprendizaje profundo que crearon un nuevo enfoque de traducción artificial, esta vez neuronal: la “Neural Machine Translation”, introducida por primera vez en 2014. Esta se basa en el principio de querer reproducir la actividad neuronal biológica de un ser humano a través de neuronas artificiales. De la misma manera que nuestras neuronas transmiten información para procesar conceptos complejos a nuestro cerebro, las neuronas artificiales son estimuladas para comunicar la información para producir un *output*.

El proceso operativo de Neural Machine Translation se divide en tres fases:

- TRAINING. Esta fase, también llamada formación o aprendizaje profundo, permite al sistema asimilar un gran corpus de traducciones bilingües y acumular informaciones sobre los idiomas de origen y de destino. Es como si la máquina estuviera realizando un trabajo de formación para las traducciones que tendrá que afrontar. NMT analiza textos de referencia de ambos idiomas y los divide en unidades, como palabras individuales o elementos gramaticales. Este proceso, llamado *word-embedding*, determina conexiones entre diferentes unidades, colocando conceptos similares juntos en un espacio multidimensional;
- ENCODER. En esta fase de codificación, se replica la operación de *word-embedding* de la fase anterior. Se crea una representación vectorial que permite a la máquina aprender las combinaciones lingüísticas y semánticas entre los elementos del texto fuente. A través de la representación, la máquina transmite estímulos de una neurona a otra;

- **DECODER.** Después de obtener la imagen vectorial, el sistema interpreta los valores de las unidades individuales en el idioma de origen y los convierte en un *output* en el idioma de destino. El objetivo es producir un resultado que se acerque lo más posible al "gold standard", es decir, una traducción de referencia creada por un traductor profesional.

Una peculiaridad de las redes neuronales es la forma en que son capaces de aprender de los errores, corrigiendo los parámetros durante la fase de aprendizaje. Aunque innovadora, la traducción automática neuronal tiene algunas limitaciones, como la traducción imprecisa de textos muy técnicos y específicos o textos creativos, ya que el mecanismo está entrenado para traducir en modelos de traducción lineal. Para lograr un resultado impecable es necesaria la intervención de un profesional humano; el NMT sigue siendo un sistema en continua evolución.

Sin embargo, este enfoque fue adoptado por todos los principales actores en el campo de la traducción automática, como Google Translate, Microsoft Translator y DeepL, que se analizarán más específicamente en el próximo capítulo.

CAPÍTULO DOS

LOS LÍMITES DE LA TRADUCCIÓN AUTOMÁTICA

2.1 Los principales servicios de traducción automática

El continuo avance de la tecnología y la rápida digitalización de la información han transformado radicalmente las formas de comunicarnos. Hoy es posible encontrar herramientas como libros o periódicos también en formato digital; la comunicación se realiza mediante el uso de teléfonos de última generación, Internet, televisión vía satélite. Incluso la comunicación oral ha sufrido una gran transformación con la introducción de las videollamadas y las videoconferencias gracias a los numerosos nuevos avances tecnológicos. El elemento clave que todos estos cambios tienen en común es la llegada de Internet. En una sociedad cada vez más globalizada, es importante y necesario que se derriben todas las barreras lingüísticas que nos impiden comunicarnos en tiempo real. Por eso la tecnología, y en particular Internet, puede ayudarnos. Estas herramientas desempeñan un papel crucial a la hora de facilitar la comunicación entre diversas partes del mundo en la sociedad contemporánea gracias a los nuevos sistemas de traducción automática, cada vez más extendidos, disponibles online y de forma gratuita. En el pasado, estos servicios eran utilizados por un público limitado con fines comerciales, como empresas multinacionales o agencias gubernamentales que necesitaban una gran cantidad de traducciones. Sin embargo, la accesibilidad que hoy tienen estos sistemas ha contribuido a darlos a conocer al público en general.

Los servicios de traducción *online* se han generalizado en Internet desde la década de 1990, pero desde entonces ha habido muchos avances. Sin duda, las tecnologías actuales representan herramientas muy importantes en el mercado de la traducción, haciendo el trabajo más sencillo y rápido. Hoy en día los principales sistemas de traducción automática que encontramos en Internet se basan en su mayoría en sistemas neuronales, que representan la nueva frontera tecnológica y de la que se habló en el capítulo anterior. La colaboración entre Microsoft Internet y AltaVista, el famoso primer motor de búsqueda que en 1997, ha introducido el primer servicio de traducción automática en línea en tiempo real, conocido como Babel Fish. Los usuarios pueden ingresar texto de hasta 150 palabras y recibir una traducción instantánea, o pueden ingresar la dirección del sitio web que desean traducir y ser dirigidos a una versión ya

traducida del sitio en cuestión. En 2008, el sistema fue adquirido por Yahoo y finalmente fue cerrado un año después. Sin embargo, Babel Fish fue pionero en una fórmula que resultó ser la plataforma de lanzamiento de la traducción automática, estimulando la difusión de otros servicios gratuitos en línea a partir de 1997.

El traductor en línea más difundido y utilizado en el mundo es Google Translate, que registra casi 500 millones de visitas por día. Google Translate, fue desarrollado por Google LLC, la empresa estadounidense que ofrece servicios en línea, y fue lanzado el 28 de abril de 2006. Originalmente se basaba en el sistema de traducción automática con enfoque estadístico (Capítulo 1). En noviembre de 2016, Google anunció el abandono del enfoque estadístico de la traducción en favor del neuronal. Específicamente, se utiliza para la traducción de sitios web, documentos, chats de correo electrónico y mensajes. También puede crear un libro de frases personal en el que guardar las traducciones ya realizadas. Finalmente, puede escuchar la pronunciación de las palabras, utilizar la función de intérprete y traducir textos escritos a mano.

A partir de diversas estadísticas se pudo estimar que los idiomas más traducidos y más hablados con Google Translate son el inglés, español, árabe, ruso, portugués e indonesio, mientras que el país donde más se utiliza el sistema de traducción automática es Brasil. Aunque todavía es impreciso, Google Translate es uno de los sistemas de traducción automática que más se ha evolucionado con el paso de los años y todavía trabaja para mejorar aún más.

Pasemos ahora a centrar nuestra atención en otro servicio de traducción automática muy extendido y utilizado en todo el mundo: Microsoft Translator. La primera versión del sistema se desarrolló entre 1999 y 2000 dentro de Microsoft Research. El traductor se basa en el enfoque "datadrive", es decir, en lugar de depender de reglas gramaticales para traducir el texto fuente, se basa en textos paralelos ya traducidos. El sitio, también conocido por los usuarios como Bing Translator, se lanzó en 2007 y ofrece traducciones gratuitas a través de Internet. Hay tres versiones del sistema disponibles: Microsoft Translator para computadoras, Microsoft Translator para el navegador y la aplicación, para sistemas operativos Android y iOS. Entre 2011 y 2018 desarrolló numerosas funciones entre las que destacan: traducción de textos, imágenes, escucha

de la pronunciación de palabras, conversación entre personas que no hablan el mismo idioma, personalización de información y creación de un libro de frases personal.

Hoy en día, Microsoft Translator es capaz de traducir más de 100 idiomas y tiene como objetivo ampliar aún más la selección de idiomas extranjeros. En cuanto a la traducción, al igual que Google Translate, Microsoft Translator sigue siendo impreciso en algunos aspectos, especialmente en los textos particularmente complejos a nivel gramatical y semántico.

El último servicio de traducción automática en Internet que será objeto de estudio en este párrafo es el sistema de traducción automática DeepL. Esto es parte de una empresa alemana que desarrolló el sistema de traducción en 2016 y lo lanzó en 2017. Su fundador y director ejecutivo es el polaco Jaroslaw Kutylowski, que ahora vive en Alemania y tiene un doctorado en informática. La última innovación introducida por el sistema el 25 de enero de 2024 es la incorporación del idioma árabe, el quinto idioma más hablado del mundo y el primer idioma soportado por DeepL que se escribe y lee de derecha a izquierda. El sistema de origen alemán ofrece diferentes planes a los usuarios. El primero es el básico, gratuito y accesible para todos desde un smartphone o un ordenador. El de pago se llama DeepL Pro y se divide a su vez en tres planes de suscripción: los niveles Starter, Advanced y Ultimate, respectivamente, con un coste de 7,49 €, 24,99 € y 49,99 € al mes.

Veamos ahora cuáles son las principales funciones que DeepL pone a disposición, más numerosas que las de Google Translate y Microsoft Translator. El sistema de traducción permite traducir archivos completos dejando los documentos en su formato original, imágenes, escuchar la pronunciación de las palabras, elegir un tono formal o informal según las solicitudes, crear un glosario e integrarse al uso de herramientas CAT. El sitio cuenta con más de 10 millones de usuarios activos cada mes, 500 mil de los cuales están suscritos a los planes pagos de la versión Pro. La mayor parte del negocio de DeepL se debe a las 20 mil empresas clientes que utilizan el software para sus traducciones; entre ellas mencionamos: Mercedes Benz, Fujitsu y la empresa ferroviaria alemana Deutsche Bahn. Según la plataforma Pitchbook, en enero de 2023 el *startup* alemán recaudó alrededor de 100 millones de dólares en financiación de empresas mundiales de capital riesgo. Veamos las principales diferencias entre los tres y cómo se comportan en la traducción.

En cuanto a las funciones, Google Translate y Microsoft Translator comparten aproximadamente las mismas funciones, mientras que DeepL ofrece más soluciones, aunque de pago. Se puede decir que este último es más útil para los profesionales del sector, traductores e intérpretes, ya que también pone a disposición el uso de herramientas CAT combinadas con el sistema de traducción automática.

Otras herramientas de traducción asistida son las CAT Tools. El término deriva del acrónimo inglés *Computer Assisted Translation* o *Computer Aided Translation* e indica todas aquellas herramientas informáticas que utilizan los traductores para ayudar en el proceso de traducción. Los programas informáticos CAT se basan en el uso de memorias de traducción, que preservan elementos ya utilizados en traducciones anteriores y que permiten un proceso de traducción más rápido y de mayor calidad. Además, estas herramientas permiten crear un glosario terminológico que incluye todos los términos específicos utilizados en uno o más textos, junto con su traducción relativa en el contexto específico.

Existen varios tipos de herramientas CAT en el mercado, cada una con una funcionalidad diferente. Entre los más comunes y útiles para los traductores se encuentran:

- Software de memoria de traducción. En este caso el texto se divide en partes llamadas "segmentos" que memorizan las traducciones realizadas y las sugieren cuando aparece una frase similar;
- Software de alineación: se carga el texto fuente original y el texto traducido y el programa los divide en párrafos, registra cómo se traducen y crea una base de datos para almacenarlos;
- Programa de gestión terminológica que permite buscar la traducción de un término específico en la base de datos y agregar nuevos pares de palabras traducidas, para luego elegir la traducción correcta según el contexto entre varias alternativas.

Existe una gran variedad de sistemas CAT Tools disponibles en el mercado, algunos gratuitos y otros de pago. Los pagos más famosos y utilizados son los siguientes:

- SDL Trados Studio, quizás el más utilizado. Es una solución completa para traducción, gestión terminológica, edición y control de calidad lingüística. Está disponible en diferentes versiones, con planes anuales o permanentes, para traductores autónomos;
- memoQ. Aunque menos conocida que la anterior, representa otra herramienta CAT importante. Ofrece muchas funciones, como el panel de vista previa de la traducción para ver el contexto de las secciones traducidas. memoQ es compatible con Trados y otros *software* de traducción y ofrece una versión gratuita para uso personal o pequeños proyectos, así como una versión profesional con licencia permanente;
- Wordfast es una versión integrada de Microsoft Word para traducción temporal de documentos bilingües. Admite varios formatos de archivo que Word puede utilizar. Existe una solución más completa con gestión terminológica y memoria de traducción, llamada Wordfast Pro;
- Déjà Vu, reconocido entre varias herramientas CAT. Ofrece funciones avanzadas de gestión de memorias de traducción y terminología, con una interfaz altamente personalizable y compatibilidad con una amplia gama de formatos de archivos, incluidos Trados y memoQ.
- Across. Incluye: memoria de traducción, gestión terminológica, control de calidad y gestión de proyectos. Ofrece una versión básica gratuita para uso online y una versión *premium* con un plan de suscripción mensual;
- XTM. Este sistema está disponible en dos versiones diferentes: una licencia de software tradicional llamada XTM Suite o una solución de suscripción en línea llamada XTM Cloud;
- Star Transit. Esta herramienta CAT se diferencia de otras porque también conserva la estructura general del documento original en su memoria de traducción. Por esta razón, no hay secciones aisladas del contexto, lo que le permite ver y controlar el contexto original de una sección determinada en cualquier momento si lo necesita. Esto garantiza una solución comprensible sin secciones fuera de contexto. Star Transit ofrece opciones de suscripción mensual o anual;
- Poedit, traduce con interfaces y puede integrarse para trabajar en Wordpress;

- MetaTaxis. A pesar de ser un sistema pago, presume de un precio significativamente más bajo que sus competidores, incluso para su versión avanzada;
- Paraphrase, un sistema basado en web que combina traducción automática y humana. También cuenta con una función de *voice over*. Hay una versión gratuita disponible, pero con funciones limitadas, y varias versiones con suscripciones mensuales.

Las Herramientas TAO gratuitas del mercado son:

- CafeTran Espresso, un programa que se puede utilizar en sistemas operativos Windows, Mac o Linux. También es compatible con otros programas de traducción importantes. Este sistema es gratuito hasta que las memorias de traducción o glosarios alcancen su capacidad máxima;
- OmegaT, quizás el CAT libre más conocido. Es un programa *open-source* y es compatible con los sistemas operativos Windows, Mac y Linux;
- Similis, menos conocido, soporta archivos HTML, XML, PDF, MS Word, OpenOffice y Trados;
- Wordfast Anywhere. Se utiliza *online* y proporciona conexiones a bases de datos de traducción públicas;
- MateCat. Con el acceso a *database* de traducción automática que ofrece a los usuarios, es capaz de mejorar la velocidad de trabajo;
- SmartCAT, que proporciona una plataforma de encuentro entre traductores y clientes;
- Virtaal, utiliza el lenguaje de programación Python e incluye herramientas de gestión de traducción y acceso a traducciones automáticas;
- Pootle. A diferencia de la mayoría de las herramientas CAT, está diseñada específicamente para traducir aplicaciones en lugar de documentos;
- Lokalise, orientado a la productividad y la calidad;
- GlobalSight, que simplifica y gestiona el proceso de localización. Puede ser utilizado por directores, traductores y editores;
- gtranslator, un programa especializado en traducción de software.

Las ventajas de las memorias de traducción y las herramientas TAO utilizadas por traductores e intérpretes profesionales son muchas. Por ejemplo, el riesgo de dejar partes sin traducir se reduce porque las herramientas CAT no permiten segmentos en los que falta texto de destino. Además, se garantiza la uniformidad terminológica y estilística del texto: este es un aspecto esencial porque es de gran importancia que un texto sea traducido con una precisión terminológica considerable para la comprensión y uso de un producto. Otra ventaja es que aumenta la productividad, es decir, el número de palabras traducidas en un intervalo de tiempo determinado, ya que los segmentos idénticos se traducen sólo una vez, mientras que los similares se traducen parcialmente. De esta forma el trabajo es más rápido, porque se recupera lo que ya se había traducido anteriormente, reduciendo los tiempos de traducción. Además de estos aspectos, las herramientas CAT permiten evaluar e identificar repeticiones y concordancias, exportar la traducción en el mismo formato que el original, calcular rápidamente el número de carpetas de traducción, utilizar herramientas de traducción automática integradas en CAT, garantizar la precisión en traducción de patentes.

Hasta ahora hemos hablado de traducción automática o sistemas de traducción asistida que pueden resultar útiles en el ámbito de la traducción escrita. Ciertamente es más fácil para las máquinas y sistemas equipados con inteligencia artificial traducir textos u oraciones escritas, mientras que parece más difícil imaginar los mismos sistemas que adoptan los profesionales del sector en el campo de la traducción oral y, por tanto, de la interpretación. Incluso en este caso, la tecnología se ha esforzado al máximo para ofrecer herramientas de traducción oral siempre a mano. Veamos cuáles son los principales, capaces de capturar frases habladas en un determinado idioma y devolverlas rápidamente en otro idioma:

- Auriculares con traductor simultáneo. Es un sistema que siempre está listo para usar y permite traducir lo que dicen otras personas en tiempo real. El dispositivo incluye dos auriculares *wireless* y puede traducir varios idiomas con sus acentos. Entre las características de estas herramientas se encuentran la extrema sencillez de uso, ya que simplemente necesitas instalar una aplicación en tu smartphone para configurarlas, una batería de larga duración

y varios modos de funcionamiento para elegir, según tus necesidades y comodidad;

- Aplicación gratuita de traducción simultánea. Entre ellas se encuentra, por ejemplo, la de Google Translate, que permite dictar frases al teléfono o a la tableta escuchando la traducción en unos instantes, o la aplicación "Translate" de Apple, en la que basta con hacer clic en la pestaña Conversación para traducir una conversación entre personas que hablan diferentes idiomas en tiempo real. La última aplicación mencionada es iTranslate Voice, específica para traducciones vocales simultáneas. Ofrece una versión gratuita y otra de pago que te permite acceder a traducciones ilimitadas y eliminar publicidad;
- Traductores simultáneos para ordenadores. Si prefieres usar la computadora o si trabajas en el campo de la traducción con este dispositivo, algunos sitios pueden ser útiles como QTranslate, gratuito, que soporta todos los principales idiomas del mundo y utiliza los mejores servicios de traducción automática disponibles *online*;
- Traductores simultáneos en línea. Se utilizan si desea tener una alternativa a las aplicaciones que descarga. Entre estos se encuentran Google Translate, Bing Translator, WordReference, ImTranslator, algunos de los cuales ya han sido mencionados y analizados anteriormente en este párrafo;
- Traductores simultáneos para conferencias. Existen algunas plataformas capaces de integrar un traductor simultáneo para utilizarlo cuando sea necesario. Entre ellos se encuentra, por ejemplo, Skype, que permite hacer todo esto mediante la función "Conversaciones traducidas", que se activa en la pestaña Chat.

En este párrafo se han presentado todos los sistemas que explotan la inteligencia artificial en el ámbito de la traducción escrita y oral. Todas disponibles en Internet, algunas gratuitas, otras de pago y utilizadas tanto en conversaciones diarias si necesitas comunicarte, como en entornos laborales para realizar traducciones profesionales.

2.2 Pre-editing, post-editing y lenguaje controlado

El avance de la traducción automática y su integración con otras tecnologías, como las memorias de traducción y las herramientas CAT, han propiciado una creciente adopción de estos procesos. El uso de la traducción automática puede contribuir de manera importante a acortar los tiempos de trabajo, pero a menudo se sacrifica la calidad de la traducción. Para evitarlo, acuden al rescate tres procesos diferentes: *pre-editing*, *post-editing* y lenguaje controlado. La intervención humana ocurre antes de la traducción automática con la fase de *pre-editing*, o después de la traducción al idioma de destino, durante la fase de *post-editing*. Este tipo de organización del proceso de traducción ha sido evidente desde las primeras etapas de la investigación en esta área. Bar-Hillel ya había subrayado en 1951 que, para los destinatarios que necesitaban una alta precisión en la traducción, la traducción automática debía ser sustituida por una especie de enfoque mixto que implicaba la intervención humana. El primero en introducir los conceptos de *pre-editor* y *post-editor* fue probablemente Erwin Reifler (1950). Según este último, el *pre-editor*, que conoce bien los aspectos lingüísticos y culturales del texto fuente, debía preparar el texto para el proceso de traducción automática, eliminando cualquier ambigüedad o señalando las partes más difíciles de traducir automáticamente. Por otra parte, el *post-editor*, según Reifler, no tenía por qué conocer necesariamente el idioma de origen, sino que sólo era responsable de solucionar y resolver los posibles problemas presentes en el texto de destino.

El *pre-editing* es un proceso de revisión que tiene como objetivo obtener un texto inicial libre de ambigüedades y complejidades que puedan dificultar su procesamiento por parte de los motores de traducción automática. El objetivo principal es intervenir antes del proceso automático para reducir el tiempo necesario para revisar el texto final. Esta fase es monolingüe y se centra exclusivamente en el texto inicial, sin necesidad de conocer el idioma de destino. Es fundamental que el *pre-editor* sea consciente de las dificultades previstas que podrían encontrar los traductores automáticos y que, por tanto, afectarían negativamente al resultado final. De esta manera se pueden eliminar o reelaborar para obtener una traducción aproximada que pueda revisarse fácilmente.

Las ediciones generalmente incluyen corregir errores ortográficos y gramaticales, aclarar partes ambiguas o complejas mediante aclaraciones, simplificaciones o

normalizaciones. Otra forma de *pre-editing* implica el uso de los lenguajes “controlados” o “simplificados”, desarrollados principalmente para aplicaciones industriales para mejorar la comprensión de documentos técnicos y especializados. Estos idiomas, derivados de lenguajes naturales como el inglés o el alemán, tienen como objetivo hacer que la documentación técnica sea fácilmente comprensible y reducir la posibilidad de malas interpretaciones por parte de los lectores.

Algunos ejemplos de idiomas controlados incluyen el inglés fundamental de Caterpillar, el alemán técnico de Siemens, el inglés personalizado multinacional y el inglés claro aprobado por Perkins. Estos lenguajes se desarrollaron para mejorar la comprensión del personal técnico, a menudo con habilidades lingüísticas limitadas.

Existen pocos lenguajes desarrollados específicamente para la traducción automática, como el inglés técnico de Caterpillar o el lenguaje controlado optimizado para traducción uniforme (CLOUT). Este último se basa en diez reglas fundamentales:

1. Escribir oraciones de menos de 25 palabras.
2. Expresar sólo una idea por frase.
3. Repetir la misma frase para expresar el mismo concepto.
4. Escribir oraciones gramaticalmente completas.
5. Utilizar una estructura gramatical sencilla.
6. Utilizar la forma activa de oraciones.
7. Repetir el nombre en lugar de utilizar un pronombre.
8. Utilizar artículos para identificar sustantivos.
9. Utilizar palabras de un diccionario general.
10. Utilizar únicamente palabras escritas correctamente.

Los beneficios que se derivan del uso de lenguajes controlados en la traducción son los siguientes: permiten redactar textos fácilmente procesados por sistemas automáticos y mejoran la calidad del resultado final en términos de comprensibilidad, haciendo más fácil la fase de *post-editing*. Sin embargo, una desventaja importante es que los autores tienen muchas limitaciones en la redacción de textos y tardan mucho en aprender las reglas necesarias.

En general, las investigaciones sobre el *pre-editing* y el uso de lenguajes controlados han recibido menos atención que los estudios sobre el *post-editing*. Esto puede deberse al hecho de que el proceso de *pre-editing* se considera un paso opcional, a veces incluso antieconómico, y no se ha demostrado claramente su eficacia para reducir el tiempo y los costos generales de producción de traducciones publicables.

El *post-editing*, por el contrario, ha encontrado mucho más espacio en el campo de la revisión de traducciones: es una operación de revisión y adaptación de textos traducidos automáticamente. Con el tiempo, esta práctica se ha descrito como "la corrección de la producción de traducción automática por parte de lingüistas/editores humanos" según Veale & Way en 1997, "la verificación, corrección y revisión de las traducciones realizadas mediante cualquier tipo de traducción automática" (Gouadec 2007), y finalmente "la corrección del resultado de la traducción generada por la máquina para asegurar que alcanza un nivel de calidad acordado previamente entre el cliente y el post-editor" (TAUS/CNGL 2011 citado en Rico & Torreòn 2012: 168). Hay dos niveles de *post-editing*: *light post-editing*, también llamado *fast post-editing*, *gist post-editing*, *rapid post-editing* y *full post-editing*, también llamado *conventional post-editing*. El *post-editing* consta de tres fases: leer el texto original, analizar el texto traducido automáticamente para identificar las partes a corregir y revisarlo para producir el texto final, realizando los cambios necesarios para adaptarlo a propósitos específicos.

El *light post-editing* implica una revisión mínima del texto traducido automáticamente, centrándose en las correcciones esenciales: estas correcciones se refieren a los errores que comprometen el significado del texto y, en consecuencia, su correcta interpretación por parte del destinatario. El resultado final debe, por tanto, ser un texto comprensible, obtenido con un mínimo esfuerzo de revisión, que corrija únicamente los errores léxicos, gramaticales y sintácticos que impiden la transmisión fiel de la información contenida en el texto original. Por otra parte, el *full post-editing* es una intervención de revisión más profunda sobre el texto traducido automáticamente, cuyo resultado final es un texto publicable, "outbound translation", teóricamente indistinguible de los traducidos por traductores profesionales. Este proceso es más lento que el *light post-editing* y requiere correcciones específicas para garantizar la

máxima precisión en la transferencia del mensaje del texto original al texto traducido y la máxima usabilidad del texto final.

Aunque los tiempos de revisión son mucho más largos que con el *light post-editing*, el cliente sigue esperando tiempos de finalización rápidos, pero con altas expectativas en la calidad final de los textos revisados.

Para el *light post-editing*, las pautas recomiendan una traducción semánticamente correcta, sin agregar ni omitir información y tratando de mantener la traducción original intacta tanto como sea posible. Sin embargo, en el caso del *full post-editing*, el resultado final es una traducción semántica, sintáctica y gramaticalmente correcta, que transmite la información del texto original de forma completa y precisa, intentando utilizar la traducción en bruto en la medida de lo posible.

La investigación sobre el *post-editing* es sumamente extensa y cuenta con un número importante de aportaciones respecto al *pre-editing*.

2.3 Los límites de la traducción automática

El avance de las tecnologías en el campo de la traducción automática ha despertado un interés creciente y ha revolucionado la forma en que se traducen e interpretan los idiomas. A menudo se confía en los sistemas de traducción automática cuando queremos tener una respuesta rápida e inmediata, cuando queremos saber brevemente el significado de un correo electrónico o de una página web en un idioma distinto al nuestro o, en la mayoría de los casos, la traducción automática que encontramos en Internet, como la que ofrece Google Translate o DeepL, se utiliza porque es gratuita. A pesar de los grandes avances realizados y de los diversos estudios en este campo para mejorar los sistemas, por ejemplo, pasando de un enfoque estadístico a uno neuronal, las limitaciones que estos tienen son muchas e importantes, y probablemente nunca será posible encontrar una solución a este problema. En este párrafo analizaremos algunas de las principales limitaciones de los sistemas de traducción automática actuales y demostraremos cómo la figura de un traductor o intérprete profesional es insustituible y necesaria para un resultado de calidad.

En primer lugar, un sistema de traducción automática es una herramienta demasiado imprecisa, que no es capaz de captar las necesidades reales de quien solicita una

traducción y el resultado que desea obtener. No capta los matices del significado: traducir no significa simplemente transponer el significado de una palabra de un idioma a otro, sino que significa tenerlo todo en cuenta el contexto cultural y los matices lingüísticos. Sólo un traductor humano es capaz de hacer todo esto, es decir, interpretar y trasladar al idioma de destino lo que se expresa en el idioma fuente, quizás utilizando otras palabras, pero siempre siendo fiel al texto original.

Otro problema que surge al utilizar traductores es que no son capaces de desambiguar las palabras homófonas, es decir, palabras que se escriben y pronuncian de la misma manera pero que tienen un significado diferente. Esto sucede porque no es capaz de analizar el contexto en el que se utilizan estos términos. La inteligencia artificial no considera el contexto lingüístico y cultural ni los cambios de registro en el texto: por ejemplo, si un texto está escrito en un registro coloquial con el uso del lenguaje hablado se traducirá de forma diferente a, por ejemplo, un texto legal que requiere un registro superior y el uso de palabras formales adecuadas al contexto. En consecuencia, no devuelve traducciones profesionales como lo haría un especialista del sector. El sistema automático razona en esta dirección porque funciona sobre una base estadística. Un ser humano, por el contrario, puede traducir también textos completamente nuevos o absolutamente improbables y puede hacerlo apoyándose en una experiencia del lenguaje que va más allá de las simples palabras ya dichas o escuchadas.

Otro tema de fundamental importancia es el de la seguridad. Una de las principales razones por las que Google Translate es tan popular entre los usuarios de todo el mundo es el hecho de que es gratuito. Sin embargo, a cambio, Google exige que los usuarios puedan utilizar sus datos para alimentar y perfeccionar constantemente su algoritmo. Esto significa que incluso la información confidencial, como resultados de pruebas de laboratorio, datos financieros altamente confidenciales o registros médicos, ingresados en Google Translate para una traducción rápida, terminan incluidos en el enorme contenedor de datos que Google almacena y utiliza.

Además de los traductores e intérpretes, los actores de doblaje también se enfrentan hoy en día a la "lucha" con la inteligencia artificial. En el campo del doblaje, esto se puede utilizar para crear voces sintéticas que potencialmente podrían competir con los actores de voz humanos en algunos contextos, especialmente para proyectos de bajo

presupuesto o donde se requiere una voz genérica, reduciendo así la necesidad de mano de obra humana en algunas fases de la producción. A pesar de los avances, también en este sector no faltan los problemas. El principal está ciertamente relacionado con la privacidad y la seguridad: la voz de un individuo puede ser sintetizada por IA, falsificando audio que podría usarse de manera fraudulenta. Otra limitación que vale la pena mencionar es el hecho de que las voces artificiales, aunque bien desarrolladas, no se pueden comparar con las voces humanas, que logran transmitir profundidad emocional y autenticidad: de hecho, los locutores profesionales pueden ofrecer interpretaciones ricas y una conexión más genuina con el público.

Es innegable que la tecnología está y seguirá cambiando profundamente el papel del traductor. Especialmente en un contexto profesional en el que esta profesión está experimentando una importante evolución tecnológica en los últimos años gracias al uso de las tecnologías más avanzadas, surge la pregunta de cuáles deberían ser las competencias y habilidades imprescindibles para estas figuras en un contexto cada vez más digitalizado. Al abordar un proceso de traducción, es esencial considerar cuidadosamente tanto el contexto de origen como el de destino. Por ejemplo, en el mundo profesional hablamos cada vez más de "transcreation", un proceso que requiere una compleja combinación de habilidades, especialmente la creatividad del traductor. Por ello, en situaciones en las que es necesario cambiar completamente el nombre de un producto o un eslogan, por ejemplo, se recurre al proceso de "transcreation", entendido como una forma de "traducción creativa", ya que implica adaptar el mensaje de un idioma a otro, sin seguir necesariamente una traducción literal palabra por palabra. Este énfasis en el aspecto creativo de la traducción mantiene a los traductores en el centro de la acción, al menos por el momento, impidiendo la adopción de máquinas en el proceso transcreativo.

Las habilidades lingüísticas también son esenciales para cualquier traductor profesional, al igual que las habilidades básicas. Estos últimos están específicamente orientados hacia habilidades actitudinales o psicofisiológicas que incluyen, por ejemplo, la capacidad de desarrollar y aplicar estándares de *post-editing* a la traducción, la gestión de las expectativas del cliente sobre la calidad final y la capacidad de tomar decisiones informadas sobre el *post-editing*.

Otro aspecto fundamental para el traductor, esta vez más práctico, es la competencia instrumental, que se refiere a las habilidades técnicas necesarias para utilizar las tecnologías de forma consciente y precisa.

Si bien todas estas habilidades pueden parecer genéricas y un tanto convencionales, en realidad son elementos indispensables para una figura profesional flexible, llamada a operar en un contexto en el que las tecnologías evolucionan rápida y constantemente. La comprensión de una cultura diferente, la sensibilidad y la capacidad de adaptarse a las necesidades de diferentes usuarios siguen siendo aspectos fundamentales que sólo un mediador humano puede ofrecer. Un traductor o un soporte automático pueden ser valiosas herramientas de apoyo, gracias a las cuales se puede mejorar la comunicación y la mediación lingüística, facilitando la traducción, la interpretación y la interacción multilingüe, y pueden resultar una herramienta útil para suavizar las barreras lingüísticas. Sin embargo, resulta esencial mantener un equilibrio entre el uso de la inteligencia artificial y la intervención humana para garantizar resultados precisos y de alta calidad en la comunicación interlingüística: por eso, la presencia humana, sigue siendo esencial para una comunicación intercultural eficaz.

2.4 La adaptación del dialecto y de la ironía

La traducción automática ha revolucionado la forma en que interactuamos con textos escritos en diferentes idiomas, ofreciendo una solución inmediata para comprender el significado general de un mensaje. Esta tecnología ha demostrado ser especialmente eficaz a la hora de traducir textos de carácter más general, pero, ante textos más específicos o que requieren un conocimiento especializado de un contexto cultural o científico, la traducción automática muestra sus limitaciones. En estas situaciones, emerge la importancia y el papel del traductor humano; al afrontar este desafío, los traductores humanos demuestran su capacidad para captar la complejidad y la riqueza de significado. Uno de los muchos contextos en los que emerge la limitación de las máquinas en la traducción es el dialecto y la ironía, dos elementos que escapan a la comprensión de los traductores automáticos.

Comenzando con dialecto, el término se refiere a aquellas variedades lingüísticas que se hablan junto con una lengua principal, la lengua oficial, y se utilizan en entornos

informales. El número de hablantes suele estar restringido a una zona determinada, aunque hay casos en los que un dialecto ha logrado extenderse más allá de sus fronteras sociolingüísticas. Si consideramos la lengua como un color, se puede decir que los dialectos se definen como diferentes gradaciones del mismo color y, por tanto, de la misma lengua. Desde el punto de vista sociolingüístico se han identificado varios criterios para distinguir entre lengua y dialecto:

1. Extensión geográfica: mientras que una lengua se extiende por un vasto territorio, el dialecto se limita a un área geográfica más estrecha;
2. Ausencia de estandarización: el dialecto carece de una forma "correcta" universalmente aceptada por los hablantes, lo que resulta en una variedad fragmentada y local, donde cada comunidad lo habla de manera única;
3. Percepción social: la comunidad hablante suele considerar el dialecto como una lengua informal y menos prestigiosa, asociada principalmente con las clases sociales menos ricas y menos educadas;
4. Uso informal: los dialectos se utilizan principalmente en entornos sociales informales, como con familiares o amigos, mientras que no se utilizan comúnmente en entornos formales como instituciones públicas, escuelas o lugares de trabajo;
5. Producción literaria limitada: los dialectos son principalmente orales y a menudo carecen de una tradición literaria significativa;
6. Influencia de la lengua dominante: en los lugares donde se habla, el dialecto suele verse influenciado por la presencia de una lengua dominante, de la que recibe préstamos lingüísticos y construcciones gramaticales;
7. Limitación del vocabulario: los dialectos se centran en vocabulario relacionado con la vida cotidiana, y a menudo carecen de los términos necesarios para discutir áreas científicas, tecnológicas, filosóficas y otras áreas del conocimiento.

La ironía también representa una herramienta lingüística bastante complicada de traducir. Esta es una forma de comunicación cuyo objetivo es exponer el contenido sarcástico de forma indirecta. Está representado por una interacción entre el lenguaje verbal y no verbal, realizada mediante gestos o expresiones faciales. Durante el proceso del habla irónica, las oraciones pueden adquirir características específicas

como una pronunciación más nasal, una disminución de la velocidad del habla y sílabas muy acentuadas. resultando en un fuerte énfasis emocional en la comunicación. La comprensión de la ironía se facilita al compartir un trasfondo común entre los interlocutores y el conocimiento del tema de la ironía.

A veces en la ironía, el sarcasmo o el humor puede haber presencia de lenguaje vulgar, con el uso de términos y expresiones groseros u ofensivos. A menudo está presente en muchos dialectos o lenguas regionales. Los términos y expresiones utilizados se refieren a diversos tabúes lingüísticos que pueden estar relacionados con el sexo, la agresión, la religión, etc. Los términos son ricos en valor emocional, porque expresan sentimientos humanos como la ira, la sorpresa, el miedo y la diversión. Su traducción se ve afectada por varios factores: en primer lugar, la repetitividad, especialmente de las expresiones habituales en el doblaje, y luego la originalidad, ya que durante la fase de "transcreation" (apartado 2.3) se pueden encontrar sustitutos que se desvíen demasiado de la lengua original. En ocasiones, los traductores pueden verse restringidos en sus transposiciones debido a directivas o exigencias de las productoras. Sin embargo, el lenguaje vulgar es principalmente un medio de expresión emocional y, al igual que los elementos humorísticos comentados anteriormente, no requiere una traducción literal, sino un equivalente en el idioma de destino.

Traducir todos estos elementos cómicos y dialectales es un desafío complejo para los traductores de hoy en día, a quienes se les exige no sólo tener competencia lingüística sino también un profundo conocimiento de las culturas de origen y de destino, junto con una dosis de creatividad para preservar la esencia y el efecto emocional de los elementos originales. Si ésta ya es una tarea difícil para los profesionales del sector, que exige largos estudios y mucho esfuerzo, sólo podemos imaginar lo que sería capaz de hacer un sistema de traducción automática si se le confiara la tarea de transponer un texto rico en contenidos lingüísticos, matices culturales, sociales, irónicos y cómicos a otro idioma, al mismo tiempo dejando el sentido del texto original y respetando lo que se quiere comunicar. Si intentamos introducir todos estos elementos en traductores automáticos como Google Translate, Microsoft Translator o DeepL, la interpretación no será satisfactoria: muchas veces, estos traducen expresiones, términos y dichos literalmente o ni siquiera traducidos; utilizan términos incorrectos que no corresponden al verdadero significado. Debido a la incapacidad de la

traducción automática para captar el verdadero significado de las palabras y el contexto sociocultural del texto, el resultado sería una traducción totalmente incomprensible para un público extranjero.

Al final se puede decir que el proceso de traducción requiere un gran conocimiento y estudio en profundidad de la lengua y la cultura de otros pueblos y solo se puede confiar a la mente humana.

Conclusión

En este proyecto de tesis se ha analizado el impacto que la tecnología y la inteligencia artificial han tenido en la práctica de la traducción, convirtiéndola en una disciplina cada vez más digitalizada. Se han examinado y presentado en detalle varios ejemplos prácticos y teóricos a lo largo de los dos capítulos diferentes.

En el primer capítulo se ha explorado toda la historia y evolución de la traducción automática, desde los años 40 con la publicación del Memorandum del científico y matemático Warren Weaver, pasando por los estudios contrastantes del estudioso Bar-Hillel y el informe ALPAC, hasta la actualidad con los más recientes sistemas de traducción online. Estos últimos se basan en el moderno enfoque neuronal, que recrea el trabajo de la mente humana, desarrollado a finales de 2014. Antes, los servicios de traducción en línea se basaban en sistemas basados en reglas o enfoques estadísticos.

A continuación, en el segundo capítulo hemos pasado a un análisis detallado de los traductores automáticos presentes en Internet, como Google Translate, Microsoft Translator y DeepL, de los que se han descrito todas sus potencialidades y funcionalidades. El penúltimo párrafo del segundo capítulo ha representado el verdadero punto de apoyo de la tesis, ya que ha analizado todas las limitaciones y desventajas que se encuentran si se utiliza la traducción automática. Luego el capítulo se ha concluido con un breve discurso sobre las competencias que tiene un profesional bien formado cuando tiene que traducir textos llenos de elementos complejos como la ironía, el dialecto y los lenguajes obscenos. Se puso de relieve la interpretación insatisfactoria de estos elementos específicos, que por tanto deben confiarse exclusivamente a traductores humanos capaces de proponer una traducción de calidad que se mantiene fiel al texto original.

Durante esta investigación, elegí explorar el mundo de la traducción automática no sólo como un tema de interés académico, sino también como una reflexión personal sobre mi experiencia como joven traductora. A lo largo de este camino, he adquirido una conciencia aún más profunda de los desafíos y complejidades que caracterizan el mundo de la traducción, que no es sólo una cuestión de transferir palabras de un idioma a otro, sino que requiere una comprensión profunda de las culturas, los matices lingüísticos y contextos específicos.

Como traductora, me doy cuenta de lo fundamental e indispensable que es el papel del traductor humano: las máquinas pueden realizar tareas repetitivas y automatizadas, pero carecen de la inteligencia, la sensibilidad y la comprensión profunda que sólo los seres humanos pueden ofrecer. Incluso frente a avances extraordinarios en inteligencia artificial, la capacidad de interpretar, adaptar y crear sigue siendo un sello distintivo de la humanidad.

Por tanto, creo que ningún tipo de máquina o artificio podrá sustituir jamás la contribución esencial del traductor humano: su papel sigue siendo esencial para garantizar una comunicación auténtica, eficaz y significativa entre diferentes culturas e idiomas. La traducción automática puede ser una ayuda válida, pero sólo a través de la intervención humana puede emerger la verdadera esencia de la traducción y dar todo su valor a la comprensión y el intercambio entre los pueblos del mundo.

BIBLIOGRAFIA

Angelone, E., Ehrensberger-Dow, M., & Massey, G. (A cura di). (2019). *The Bloomsbury Companion to Language Industry Studies*. Bloomsbury Academic.

Bersani Berselli, G. (Ed.). (2011). *Usare la traduzione automatica*. CLUEB.

Eberharter, C. (2020). *Le problematiche e le sfide della traduzione automatica. Un confronto tra Google Translate, Bing Translator e DeepL*. GRIN Verlag.

Eco, U. (2003). *Dire quasi la stessa cosa. Esperienze di traduzione*. Bompiani.

Forcada, M., L., (2017). Making sense of neural machine translation. *Translation Spaces*, 6:2, 291-309.

Giacomarra, M., G., (2017). *Translation Studies. Tradurre: manipolare e costruire realtà*. Libreriauniversitaria.it Edizioni.

Koehn, P. (2020). *Neural Machine Translation*. Cambridge University Press.

Koponen, M., Mossop, B., Robert, I., S., & Scocchera, G. (A cura di). (2021). *Translation Revision and Post-editing: Industry Practices and Cognitive Process*. Routledge.

Monti, J. (2019). *Dalla Zairja alla traduzione automatica. Riflessioni sulla traduzione nell'era digitale*. Paolo Loffredo.

Morini, M. (2007). *La traduzione: teorie, strumenti e pratiche*. Sironi Editore.

Naldi, M. (2014). *Traduzione automatica e traduzione assistita*. Società Editrice Esculapio.

O'Brien, S., Winther Balleng, L., Carl, M., Simard, M., & Specia, L., (A cura di). (2014). *Post-editing of Machine Translation: Processes and Applications*. Cambridge Scholars Publishing.

Sammut, C., & Webb, G. (A cura di). (2011). *Encyclopedia of Machine Learning*. Springer.

Zuccato, E. (2022). *Trasferimenti in loco: saggi sulla traduzione, il dialetto e la poesia*. Mucchi Editore.

SITOGRAFIA

- http://opar.unior.it/1962/1/mqm_articolo.pdf (consultato il 6 febbraio 2024)
- <http://www.popmag.it/google-translate-traduce-ma-lo-fa-come-un-traduttore-umano/> (consultato il 14 febbraio 2024)
- <https://aclanthology.org/W19-6725.pdf> (consultato il 18 gennaio 2024)
- <https://aws.amazon.com/it/what-is/machine-translation/> (consultato l'8 gennaio 2024)
- <https://caledonian.it/blog/traduttore-professionista-vs-google-translate-tutti-i-vantaggi/> (consultato il 13 febbraio 2024)
- <https://creative-words.com/cose-e-come-funziona-la-traduzione-automatica-neurale/> (consultato il 20 gennaio 2024)
- https://elearning.humnet.unipi.it/pluginfile.php/236303/mod_resource/content/0/16.%20Valutazione%20delle%20traduzioni%20automatiche%20-%205%20dicembre%202019.pdf (consultato il 10 febbraio 2024)
- https://en.wikipedia.org/wiki/Microsoft_Translator (consultato il 24 gennaio 2024)
- <https://forbes.it/2023/08/14/deepl-unicorno-tedesco-sfida-google-translate-chatgpt/> (consultato il 5 febbraio 2024)
- <https://faturanetwork.eu/interviste-al-futuro/749-3742/chatgpt-a-traduttori-e-traduttrici-lintelligenza-artificiale-non-vi-sostituira> (consultato il 23 gennaio 2024)
- https://it.wikipedia.org/wiki/Google_Traduttore (consultato il 23 gennaio 2024)
- https://it.wikipedia.org/wiki/Pesce_di_Babele (consultato il 23 gennaio 2024)
- <https://it.wikipedia.org/wiki/SYSTRAN> (consultato il 16 gennaio 2024)
- https://it.wikipedia.org/wiki/Turpiloquio_nella_lingua_italiana (consultato il 16 febbraio 2024)
- <https://open.unive.it/hitrade/books/HutchinsWarren.pdf> (consultato il 18 dicembre 2023)
- <https://patrimonilinguistici.it/dialetto-definizione/> (consultato il 16 febbraio 2024)
- https://sites.cs.ucsb.edu/~lilei/course/dl4mt21fa/Weaver_1949_Translation.pdf (consultato il 18 dicembre 2023)
- <https://spiegato.com/cose-babel-fish> (consultato il 23 gennaio 2024)

<https://ssmlsandomenico.wordpress.com/author/astrambientegiovani/> (consultato il 13 febbraio 2024)

https://static.deepl.com/files/press/companyProfile_IT.pdf (consultato il 5 febbraio 2024)

<https://translationtheories.wordpress.com/2013/03/20/storia-della-traduzione-1-le-origini/> (consultato il 23 novembre 2023)

<https://translationtheories.wordpress.com/2013/03/20/storia-della-traduzione-2-800-e-900/> (consultato il 23 novembre 2023)

<https://travelholicsouls.com/microsoft-translator-come-funziona/> (consultato il 24 gennaio 2024)

<https://web.archive.org/web/20101008031230/http://www.hutchinsweb.me.uk/IntroMT-1.pdf> (consultato il 27 novembre 2023)

<https://www.aranzulla.it/traduttore-simultaneo-63546.html> (consultato il 9 febbraio 2024)

<https://www.dday.it/redazione/26760/google-translate-microsoft-deepl-quale-il-migliore-traduttore> (consultato il 6 febbraio 2024)

<https://www.deepl.com/it/blog> (consultato il 5 febbraio 2024)

<https://www.deepl.com/it/features> (consultato il 5 febbraio 2024)

https://www.federlingue.it/export/sites/unione/federlingue/doc/5-SYSTRAN_Dal-pre-editing-al-post-editing.pdf (consultato il 29 novembre 2023)

<https://www.focus.it/cultura/curiosita/quante-lingue-si-parlano-al-mondo> (consultato il 22 novembre 2023)

<https://www.gregoriosettimo.eu/notizie/264-le-origini-della-traduzione-e-la-nascita-del-traduttore.html> (consultato il 23 novembre 2023)

<https://www.illibraio.it/news/fumetti/zerocalcare-libri-fumetti-1411699/> (consultato il 17 febbraio 2024)

<https://www.intuition.it/blog/machine-translation-e-traduzione-automatica-pro-e-contro> (consultato il 27 novembre 2023)

https://www.lingue.unich.it/sites/st07/files/lm37-lm38_-_04_-_cat_tools.pdf (consultato l'8 febbraio 2024)

<https://www.linguise.com/it/blog/guida/qual-e-la-traduzione-automatica-neurale/> (consultato il 20 gennaio 2024)

<https://www.linguise.com/it/blog/guida/quale-servizio-di-traduzione-automatica-e-il-piu-accurato/> (consultato il 23 gennaio 2024)

<https://www.link-lab.it/articoli/i-rischi-di-google-traduttore/> (consultato il 14 febbraio 2024)

<https://www.lionbridge.com/it/blog/translation-localization/machine-translation-in-translation/> (consultato il 19 gennaio 2024)

<https://www.lionbridge.com/it/blog/translation-localization/whats-the-difference-between-translation-interpretation-services/> (consultato il 9 febbraio 2024)

<https://www.lionspeech.com/2020/08/06/storia-della-traduzione-eta-contemporanea/#:~:text=Il%20XIX%20secolo%20è%20pertanto,tratta%20di%20un%20testo%20straniero.> (consultato il 24 novembre 2023)

<https://www.navigaweb.net/2017/06/7-modi-di-usare-google-traduttore-con.html> (consultato il 23 gennaio 2024)

https://www.polilingua.it/blog/post/cat_tools.htm (consultato il 7 febbraio 2024)

<https://www.salesforce.com/it/blog/definizione-application-programming-interface-api/> (consultato il 24 gennaio 2024)

<https://www.ssmicarlobo.it/non-categorizzato/cat-tool/> (consultato il 7 febbraio 2024)

<https://www.traduzioni-legalizzate.it/i-dieci-anni-di-google-traduttore/> (consultato il 23 gennaio 2024)

<https://www.translatorcorner.com/cat-tool-cosa-sono-perche-importanti/> (consultato il 7 febbraio 2024)