



UNIVERSITÀ
DI PAVIA

DIPARTIMENTI DI GIURISPRUDENZA, INGEGNERIA INDUSTRIALE E
DELL'INFORMAZIONE, SCIENZE ECONOMICHE E AZIENDALI, SCIENZE POLITICHE
E SOCIALI, STUDI UMANISTICI

CORSO DI LAUREA INTERDIPARTIMENTALE IN
COMUNICAZIONE DIGITALE

**LA TRADUZIONE È ANCORA UMANA? UN'ANALISI QUALITATIVA SULLA
MACHINE TRANSLATION E SUL RUOLO DEL TRADUTTORE NELL'EPOCA
DELL'INTELLIGENZA ARTIFICIALE GENERATIVA**

Relatore:

Chiar.mo Prof. Paolo Costa

Correlatore:

Chiar.mo Prof. Giuseppe Antonelli

Tesi di Laurea di:

Clarissa Besana

Matricola n. 537414

ANNO ACCADEMICO 2024/25

Abstract

Negli ultimi anni, la rapida evoluzione dei modelli linguistici di grandi dimensioni (LLM) ha condizionato significativamente numerosi aspetti della vita quotidiana e professionale. Tali tecnologie si sono imposte in modo progressivo in ambiti tradizionalmente riservati all'esclusiva competenza umana, tra i quali figura la traduzione. La presente ricerca si propone di analizzare l'impatto degli LLM nel campo della traduzione letteraria, rivolgendo particolare attenzione alle implicazioni qualitative e professionali determinate dal loro crescente utilizzo. A partire da una contestualizzazione teorica che ripercorre le principali teorie della traduzione e l'evoluzione dei modelli di IA per l'elaborazione del linguaggio naturale, l'indagine adotta un approccio multidisciplinare che integra analisi concettuali, testuali e indagini qualitative. In seguito a una disamina dei principali metodi di valutazione della qualità delle traduzioni automatiche e delle tendenze stilistiche ricorrenti nei testi artificiali, si è condotto un esperimento di traduzione che ha previsto l'utilizzo di tre diversi modelli di intelligenza artificiale. I risultati dell'analisi testuale dimostrano delle chiare differenze nei profili comportamentali di ciascun modello. Ulteriori risposte alle ipotesi poste dalla domanda di ricerca sono fornite dalle testimonianze di due traduttori professionisti di rilievo. Le interviste affrontano temi fondamentali per il settore, come l'effetto ancoraggio nei processi di revisione, la potenziale perdita dello stile del traduttore, l'applicabilità della nozione di *good enough quality* in ambito letterario e gli scenari futuri della professione. La ricerca evidenzia come la figura del traduttore si trovi in una fase di ridefinizione, nella quale la componente umana sembra ancora indispensabile.

Abstract (in lingua inglese)

In recent years, the rapid evolution of Large Language Models (LLMs) has significantly impacted numerous aspects of both daily and professional life. These technologies have progressively established themselves in fields traditionally reserved for exclusive human expertise, including the field of translation. This research aims to analyze the impact of LLMs within literary translation, paying particular attention to the qualitative and professional implications resulting from their increasing use. Based on a theoretical framework that traces the primary theories of translation and the evolution of AI models for Natural Language Processing (NLP), the study adopts a multidisciplinary approach integrating conceptual, textual, and qualitative analyses. Following an examination of the main methods for evaluating machine translation quality and the stylistic trends recurring in AI-generated texts, a translation experiment was conducted using three different artificial intelligence models. The results of the textual analysis demonstrate clear differences in the behavioral profiles of each model. Further insights into the research hypotheses are provided by the testimonies of two prominent professional translators. The interviews address fundamental themes within the sector, such as the anchoring effect in revision processes, the potential loss of the translator's individual style, the applicability of the "good enough quality" notion in a literary context, and the future outlook of the profession. Ultimately, the research highlights how the role of the translator is currently undergoing a phase of redefinition, one in which the human component still appears to be indispensable.

ABSTRACT	2
ABSTRACT (IN LINGUA INGLESE)	3
INTRODUZIONE.....	7
CAPITOLO 1 – COMPRENDERE PER TRADURRE: DALL’UOMO ALLA MACCHINA.....	9
1.1. LA TRADUZIONE: UN DISCORSO FRAMMENTATO	9
1.1.1. <i>La traduzione come negoziazione</i>	10
1.1.2. <i>Il paradosso della fedeltà</i>	11
1.1.3. <i>L’importanza del contesto</i>	11
1.2. ALLE ORIGINI DELLE MACCHINE DEL LINGUAGGIO.....	12
1.2.1. <i>Le prime sperimentazioni: il paradigma simbolico e i limiti dei sistemi esperti</i>	13
1.2.2. <i>La svolta statistica e i modelli basati sui corpora</i>	16
1.2.3. <i>Le reti neurali e l’innovazione dei metodi di word embedding</i>	24
1.2.4. <i>Elaborazione sequenziale con le Reti Neurali Ricorrenti (RNN)</i>	30
1.3. LO STATO DELL’ARTE ATTUALE: I TRANSFORMER.....	31
1.3.1. <i>Il positional encoding</i>	32
1.3.2. <i>Attention is all you need</i>	33
1.3.3. <i>L’architettura encoder-decoder</i>	34
1.4. LA TRADUZIONE AUTOMATICA NELLA STORIA.....	35
1.4.1. <i>Gli inizi: dagli anni ’50 agli anni ’60</i>	35
1.4.2. <i>La condanna dell’ALPAC</i>	36
1.4.3. <i>La ricerca non si ferma</i>	36
1.4.4. <i>L’avvento dei metodi basati sui corpora e lo stato attuale</i>	37
CAPITOLO 2 – COME TRADUCE LA MACCHINA? VALUTARE E RICONOSCERE IL MARCHIO DELL’IA NELLE TRADUZIONI AUTOMATICHE.....	38
2.1. VALUTARE LA QUALITÀ DELLE TRADUZIONI AUTOMATICHE	38
2.1.1. <i>Metriche di valutazione automatiche: i limiti di BLEU</i>	38

2.1.2. <i>Metriche di valutazione umane e classificazione degli errori</i>	40
2.2. LE TENDENZE STILISTICHE DEI TESTI ARTIFICIALI	42
2.3. METODOLOGIA	44
2.3.1. <i>Definizione dell'obiettivo</i>	44
2.3.2. <i>Costruzione del corpus</i>	45
2.3.3. <i>Modelli usati per la traduzione</i>	46
2.3.4. <i>Procedura di traduzione</i>	47
2.3.5. <i>Limiti epistemologici</i>	49
2.4. ANALISI DEI RISULTATI	50
2.4.1. <i>Claude</i>	51
2.4.2. <i>ChatGPT</i>	54
2.4.3. <i>Gemini</i>	57
CAPITOLO 3 – LE VOCI IN CAPITOLO: LE TESTIMONIANZE DI DUE PROFESSIONISTI	62
3.1. NOTA METODOLOGICA	62
3.1.1. <i>Gli intervistati</i>	63
3.2. LE INTERVISTE.....	63
3.2.1. <i>Il rapporto con gli strumenti di traduzione automatica o di LA</i>	64
3.2.2. <i>Riconoscere gli elementi caratterizzanti e i rischi delle traduzioni automatiche</i>	65
3.2.3. <i>L'effetto ancoraggio nel post-editing</i>	68
3.2.4. <i>La perdita dello stile del traduttore</i>	69
3.2.5. <i>Il concetto di good enough quality nella traduzione letteraria</i>	71
3.2.6. <i>Il futuro della professione</i>	74
CONCLUSIONE	76
BIBLIOGRAFIA E SITOGRAFIA	79
APPENDICI	84
INTERVISTA A LUCA BRIASCO	84
INTERVISTA A SARA BILOTTI.....	95

Introduzione

Negli ultimi anni il progressivo avanzamento tecnologico relativo ai modelli linguistici di grandi dimensioni (LLM) e, di conseguenza, al più ampio campo dell'intelligenza artificiale, ha influito considerevolmente su molteplici aspetti della vita quotidiana e professionale di ciascuno. Tali tecnologie hanno infatti conosciuto una rapida espansione, arrivando a pervadere contesti che, fino a poco tempo fa, potevano essere considerati esclusivamente umani: il linguaggio è senza alcun dubbio uno di questi. Dalle prime sperimentazioni con i sistemi esperti fino alle attuali architetture Transformer, l'elaborazione del linguaggio naturale ha infine raggiunto capacità tali da rendere possibile delegare alle macchine processi linguistici di notevole complessità, tra i quali figura la traduzione.

In uno scenario di crescente diffusione di strumenti basati su modelli linguistici in questo particolare ambito, occorre pertanto interrogarsi sulle loro molteplici implicazioni relative non soltanto alla qualità delle traduzioni, ma anche all'impatto sulla figura professionale del traduttore. Nello specifico, questa tesi si concentrerà sul più ristretto settore della traduzione letteraria: la necessità di esplorare tale tematica trova infatti la sua origine nella consapevolezza che essa rappresenta tuttora un territorio in cui le differenze tra linguaggio naturale e linguaggio artificiale si manifestano con grande chiarezza, rendendo urgente una riflessione che comprenda al suo interno una visione globale delle criticità che tale fenomeno comporta.

Alla luce di queste considerazioni, la presente ricerca si pone l'obiettivo di analizzare l'impatto percepito dell'utilizzo di modelli linguistici di grandi dimensioni come strumenti a supporto nella traduzione letteraria, rivolgendo particolare attenzione alle implicazioni sul livello qualitativo delle traduzioni e, soprattutto, sul ruolo del traduttore. Per raggiungere tale scopo si è deciso di strutturare la ricerca secondo un approccio multidisciplinare: combinando analisi concettuali, testuali e indagini qualitative è stato possibile sondare in primo luogo le fondamenta teoriche della traduzione automatica, per poi verificare concretamente tali ipotesi mediante un esperimento e delle interviste.

La tesi si articola in tre parti principali. La prima parte fornisce una cornice teorica generale, nella quale si ripercorrono in maniera essenziale alcune delle numerose teorie della traduzione e tracciando la graduale evoluzione delle tecnologie di intelligenza artificiale relative all'elaborazione del linguaggio naturale. Nel dettaglio, si dà particolare rilievo alla storia e al funzionamento di tali macchine, che dai primi esperimenti degli anni Cinquanta del secolo scorso hanno conosciuto un'importante trasformazione nell'ultimo decennio con l'avvento del meccanismo di attenzione e dell'architettura Transformer.

Il secondo capitolo rappresenta il fulcro della ricerca empirica di questa tesi: dopo aver accuratamente illustrato i principali metodi di valutazione della qualità delle traduzioni automatiche, ponendo l'accento sulla distanza che divide le metriche automatiche da quelle umane, e le tendenze stilistiche ricorrenti nei testi artificiali, viene proposto un esperimento di traduzione condotto su tre diversi modelli di intelligenza artificiale. Tale indagine empirica è risultata necessaria per verificare le teorie precedentemente esposte, così come per identificare e analizzare le particolari cifre stilistiche che caratterizzano i testi di origine automatica.

La terza e ultima parte è stata strutturata con il fine di ampliare le prospettive teoriche attraverso delle interviste a due traduttori professionisti di consolidata esperienza, che hanno permesso l'accesso a una visione più puntuale e articolata sul rapporto con le tecnologie di intelligenza artificiale nella pratica quotidiana. Le testimonianze hanno toccato questioni di fondamentale importanza per il settore, come gli elementi caratteristici delle traduzioni automatiche e i rischi derivanti dal loro utilizzo, il fenomeno dell'effetto ancoraggio nell'attività di revisione di tali traduzioni, l'eventuale e progressiva perdita dello stile del traduttore, la possibilità di applicare la nozione di *good enough quality* in ambito letterario e, infine, gli scenari futuri a cui potrebbero andare incontro i traduttori professionisti.

Capitolo 1 – Comprendere per tradurre: dall'uomo alla macchina

1.1. La traduzione: un discorso frammentato

La riflessione sulla traduzione vanta una tradizione millenaria: dalle prime teorie di Cicerone fino ai contributi più contemporanei di Jakobson¹, Eco² e Berman³, numerosi autori si sono interrogati sulla natura e sui limiti del tradurre. Indubbiamente, l'ampiezza cronologica di tale dibattito ha contribuito a delinearne il carattere, generando un'eterogeneità tale che è possibile fare riferimento a molteplici teorie della traduzione. Nello specifico, si assiste a una particolare tendenza per cui se in passato le teorie trovavano la propria origine nelle esperienze personali degli autori, ponendo quindi al centro un discorso prettamente empirico, in epoche più contemporanee si privilegia invece una «prospettiva metodologico-filologico-filosofica»⁴.

Benjamin, ad esempio, ne *Il compito del traduttore* affronta il tema della traduzione senza offrire alcuna dimostrazione concreta e in maniera decisamente originale, distaccandosi completamente dalle categorie estetiche tradizionali. In tal senso, il passo più significativo dell'opera tratta il rapporto fra originale e traduzione, «tanto più intimo, quanto per l'originale stesso non significa più nulla»⁵: l'autore evidenzia quindi la natura temporale della traduzione, che consente all'originale di trasformarsi e sopravvivere in nuove forme.

In tempi più recenti, il teorico della traduzione Berman concentra la propria analisi sulla traduzione etnocentrica, ossia l'approccio dominante nella traduzione letteraria. Nel dettaglio, essa mira a rendere invisibile il processo traduttivo per dare al lettore l'illusione che l'autore straniero abbia scritto direttamente nella lingua

¹ Roman. Jakobson, «On Linguistic Aspects of Translation». In *On Translation*. Harvard University Press, 1959. <https://doi.org/10.4159/harvard.9780674731615.c18>.

² Umberto. Eco, *Dire quasi la stessa cosa: esperienze di traduzione* (Bompiani, 2003).

³ Antoine. Berman, *La traduzione e la lettera o l'albergo nella lontananza* (Quodlibet, 2003).

⁴ Siri. Nergaard, *La teoria della traduzione nella storia: testi di Cicerone, San Gerolamo, Bruni, Lutero, Goethe, von Humboldt, Schleiermacher, Ortega y Gasset, Croce, Benjamin* (Bompiani, 1993).

⁵ Walter. Benjamin, *Il compito del traduttore*, trad. da Maria Teresa Costa (Mimesis, 2023), 49.

d'arrivo; tuttavia, tale pratica tende ad annettere l'opera originale alla cultura ricevente, eliminando di fatto ogni traccia di estraneità linguistica. Di conseguenza Berman, in antitesi al concetto di etnocentrismo, chiarisce che l'essenza della traduzione risiede nell'«aprire l'Estraneo in quanto Estraneo al proprio spazio di lingua»⁶.

1.1.1. La traduzione come negoziazione

Prima di illustrare i processi che hanno condotto alle straordinarie innovazioni della traduzione automatica moderna, occorre tuttavia introdurre in maniera generale i concetti fondamentali concernenti il processo traduttivo e le sue implicazioni teoriche. Data la natura empirica di questa tesi, si sceglie di affrontare tale tematica prendendo in analisi Umberto Eco con *Dire quasi la stessa cosa*. Questo testo si rivela particolarmente appropriato, in quanto Eco affronta la traduzione da una prospettiva pragmatica, attingendo dalle proprie esperienze come traduttore e autore tradotto.

A questo proposito, il concetto di negoziazione si pone come pilastro essenziale per definire l'atto di tradurre: Eco, infatti, parte dal presupposto che la traduzione non consista in realtà nel «dire la stessa cosa in un'altra lingua», ma nel «dire *quasi* la stessa cosa»⁷. Nello specifico, il «*quasi*» non rappresenta un'approssimazione involontaria, ma delimita lo spazio di manovra di un processo deliberato in cui i margini di flessibilità dipendono da criteri che devono essere negoziati preliminarmente tra le diverse istanze coinvolte⁸; per definizione, del resto, la negoziazione si configura proprio come un procedimento in cui per ottenere un determinato risultato si accetta di rinunciare a qualcos'altro. Alla luce di quanto appena detto, è evidente che non sia possibile conservare tutto il potenziale del testo originale⁹.

⁶ Antoine. Berman, *La traduzione e la lettera o l'albergo nella lontananza*, 62.

⁷ Umberto. Eco, *Dire quasi la stessa cosa* op. cit., 9.

⁸ Ivi, 10.

⁹ Ivi, 18.

1.1.2. Il paradosso della fedeltà

Un ulteriore elemento cruciale, tanto nella traduttologia quanto nel libro qui analizzato, è la fedeltà, sebbene il termine assuma significati differenti nei due contesti. Se la convenzione tradizionale la rappresenta come una ricerca dell'esattezza lessicale o una corrispondenza biunivoca tra termini, Eco la riassume perfettamente nel seguente modo:

La fedeltà è piuttosto la tendenza a credere che la traduzione sia sempre possibile se il testo fonte è stato interpretato con appassionata complicità, è l'impegno a identificare quello che per noi è il senso profondo del testo, e la capacità di negoziare a ogni istante la soluzione che ci pare più giusta¹⁰.

In questa prospettiva, la fedeltà si configura come una forma di interpretazione che mira a produrre nel lettore della traduzione effetti analoghi a quelli che il testo originale intendeva suscitare nei suoi lettori su tutti i piani di scrittura. Emerge così un paradosso molto dibattuto nel campo della traduttologia: spesso, tradurre letteralmente non significa necessariamente rispettare il significato autentico del testo originale. Si deduce pertanto che un distacco dalla lettera può costituire, controintuitivamente, un «atto di fedeltà» al senso profondo dell'opera¹¹.

1.1.3. L'importanza del contesto

L'ultimo concetto chiave, propedeutico a quanto seguirà in questo capitolo, risiede nella consapevolezza che la traduzione non avviene mai tra sistemi linguistici astratti, ma tra testi concreti. Eco chiarisce questo punto osservando che, se l'attività del traduttore consistesse semplicemente nel mettere in relazione due sistemi linguistici, il dizionario bilingue rappresenterebbe l'esempio più alto di traduzione; tuttavia nessuno lo ritiene tale, poiché una traduzione soddisfacente non può essere ridotta a una mera esibizione di corrispondenze lessicali.

¹⁰ Ivi, 364.

¹¹ Ivi, 16.

Supponendo che questo sia il caso, la traduzione risulterebbe pertanto un'operazione impossibile, in quanto i sistemi semiotici, analizzati isolatamente, appaiono «mutuamente incommensurabili», ovvero impongono ai propri parlanti una specifica visione del mondo¹². Tale incommensurabilità teorica si oppone, però, alla pratica: l'oggetto di una traduzione è sempre un testo, che per sua natura è caratterizzato dal contesto.

In particolare, il sistema linguistico presenta un repertorio di possibilità spesso ambigue e omonime¹³ ed è per questo che la disambiguazione si realizza esclusivamente a livello testuale. Prima di procedere alla traduzione vera e propria, infatti, è necessario determinare quale accezione di un termine sia pertinente in un dato contesto¹⁴. Questo passaggio richiede, secondo Eco, il ricorso alle «selezioni contestuali» e all'informazione enciclopedica: nello specifico, le prime definiscono i possibili contesti d'uso di una parola, mentre la seconda rappresenta la conoscenza extra-linguistica del mondo reale¹⁵.

1.2. Alle origini delle macchine del linguaggio

Da sempre l'uomo tende a costruire strumenti e macchine in grado di facilitare lo svolgimento di alcuni processi e con l'avanzare della tecnologia queste diventano sempre più sofisticate, allargando notevolmente gli ambiti di utilizzo. Il dominio del linguaggio, però, è sempre stato una prerogativa dell'essere umano. Infatti, nonostante esistano macchine inerenti con l'uso del linguaggio (come i programmi di videoscrittura), non è mai stato possibile produrlo in modo credibile attraverso una macchina¹⁶.

Oggi l'automazione dei processi linguistici è al centro di un dibattito che solleva molti interrogativi sul rapporto tra tecnologia e creatività: in particolare, l'idea che la macchina possa sostituire la componente umana in processi creativi come la

¹² Ivi, 37.

¹³ Ad esempio, si pensi a quelle parole che contengono al loro interno più significati.

¹⁴ Ivi, 48.

¹⁵ Ivi, 31.

¹⁶ Alfio. Ferrara, *Le macchine del linguaggio: l'uomo allo specchio dell'intelligenza artificiale* (Einaudi, 2025).

scrittura o la traduzione. In questo senso, è interessante come già nel 1951 J.R.R. Tolkien (filologo e scrittore) trattasse la tematica della macchina in maniera molto attuale in una lettera al suo editore. La sua riflessione vede la macchina come una forma di magia moderna, poiché «alternativa allo sviluppo delle proprie potenzialità e dei propri talenti interiori» in contrapposizione netta con la creazione artistica vera e propria¹⁷. Quanto preconizzato da Tolkien è facilmente osservabile nella sempre più scadente qualità dei contenuti presenti nel Web, molto spesso frutto di molteplici traduzioni dalla lingua inglese attraverso sistemi di *Machine Translation*¹⁸.

Ora, prima di ripercorrere la graduale evoluzione che ha caratterizzato il campo dell'intelligenza artificiale relativa alla traduzione automatica, è necessario comprendere i meccanismi sottostanti a quelle che Ferrara definisce “macchine del linguaggio”.

1.2.1. Le prime sperimentazioni: il paradigma simbolico e i limiti dei sistemi esperti

Il 1956 segna l'anno di svolta per la ricerca sull'intelligenza artificiale (IA) con la conferenza di Dartmouth. Un gruppo di quattro matematici statunitensi propone che venga svolto uno studio partendo dalla seguente congettura: «every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it.»¹⁹. In tale contesto, il linguaggio è stato uno dei primi aspetti dell'intelligenza umana a essere indagati, basandosi sulla teoria per la quale una parte significativa del pensiero consiste nella manipolazione delle parole secondo alcune «regole di ragionamento e di congettura»²⁰. Infatti, l'approccio

¹⁷ J. R. R. Tolkien, *Il Silmarillion*, a cura di Christopher Tolkien e Marco Respinti, traduzione di Francesco Saba Sardi, Bompiani, 2014 [1977], 19.

¹⁸ Brian Thompson et al., *A Shocking Amount of the Web is Machine Translated: Insights from Multi-Way Parallelism*, 5 giugno 2024, <http://arxiv.org/abs/2401.05749>.

¹⁹ John McCarthy et al., *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence* (2006).

²⁰ Ibid.

basato sull'insegnamento di regole come mezzo per simulare i processi cognitivi umani ha caratterizzato profondamente la ricerca sull'IA negli ultimi cinquant'anni²¹.

Pertanto, è possibile individuare due paradigmi distinti nella storia dell'intelligenza artificiale. Il primo fa riferimento al paradigma simbolico, nel quale si è cercato di rappresentare esplicitamente²² la conoscenza necessaria per la risoluzione di problemi e compiti. Il secondo si identifica nell'apprendimento automatico o *Machine Learning* (ML), che utilizza algoritmi capaci di apprendere dai dati e di svolgere compiti in modo autonomo senza la necessità di istruzioni esplicite.

All'interno del paradigma simbolico domina una particolare applicazione di intelligenza artificiale nota come sistemi basati sulla conoscenza o *knowledge-based systems*, nei quali sono inclusi i sistemi esperti. Tali sistemi si basano su due componenti essenziali: innanzitutto le basi di conoscenza che, come già delineato, costituiscono la rappresentazione esplicita dei saperi necessari alla risoluzione di un problema. A queste si aggiunge il secondo elemento indispensabile per il funzionamento di un sistema esperto, vale a dire il motore inferenziale o quello che Ferrara chiama "ragionamento automatico": una componente software che applica delle regole logiche alle basi di conoscenza per dedurre nuove informazioni. Tuttavia questa conoscenza, anche se chiara ed evidente per l'essere umano, deve essere necessariamente tradotta in un codice leggibile dalla macchina, così che possa accedervi in modo preciso e logico²³. Sotto questo profilo esistono molteplici modelli logici adatti a essere processati dalla macchina, ma le logiche descrittive, in particolare, hanno conosciuto un'ampia diffusione nei primi anni di questo secolo grazie alla loro semplicità. Esse basano la loro architettura su dei costrutti, ovvero simboli codificati in modo tale che la macchina possa interpretarli e, nello specifico, questi sono chiamati individui, concetti e ruoli. Rispettivamente, i primi sono finalizzati a rappresentare degli oggetti specifici del mondo reale; i concetti, al contrario, hanno

²¹ Naomi S. Baron, *Who wrote this?: how AI and the lure of efficiency threaten human writing* (Stanford University Press, 2025).

²² Con il termine "esplicitamente" si intende una rappresentazione della conoscenza che sia *human-readable*, ovvero leggibile dall'essere umano.

²³ Ferrara, *Le macchine del linguaggio*, op. cit.

carattere più generale e definiscono insiemi di individui come, ad esempio, il concetto di genitore; infine, i ruoli hanno lo scopo di definire le relazioni fra gli individui²⁴.

In un contesto prettamente linguistico, è indispensabile fare riferimento a un progetto che ha influenzato in maniera decisiva gli studi sul linguaggio naturale e, in particolare, quell'ambito che si occupa della sua elaborazione automatica: il database lessicale conosciuto come WordNet. La sua struttura è composta da un'inconsueta rappresentazione semantica dei vocaboli, in quanto opera basandosi su insiemi di sinonimi chiamati *synset* e sui rapporti lessicali che intercorrono tra di essi. Le relazioni tra *synset* si configurano come relazioni IS-A o, in linguaggio informatico, ereditarietà: in questo modo qualsiasi attributo appartenente alla superclasse diventa attributo della sottoclasse²⁵. Tuttavia, come riferito dallo stesso Vetere, la lingua si basa su un principio di economia linguistica per cui ogni parola detiene una specifica connotazione e, quindi, non può esistere una perfetta corrispondenza tra sinonimi. A questo punto, emergono chiaramente i limiti di un database come WordNet: l'architettura basata su *synset* comporta gravi mancanze rispetto alla ricchezza lessicale della lingua e, oltretutto, rende estremamente tortuoso il trasporto in altre lingue²⁶.

Sebbene i sistemi esperti abbiano conosciuto un'ampia diffusione e si siano contraddistinti come il prodotto più rilevante della stagione simbolica dell'IA, nel corso del tempo sono affiorate criticità strutturali che hanno comportato il loro progressivo abbandono nel campo del linguaggio. La ragione principale di questo declino trova la sua origine nella complessità del linguaggio, più precisamente nella trasposizione in regole e basi di conoscenza dei saperi a esso relativi. In poche parole, cercare di rappresentare tutto il linguaggio in uno schema di regole definito risulterebbe in una struttura a rete decisamente complessa e, soprattutto, rigida: difatti il linguaggio è, di per sé, mutevole e possiede la capacità di rinnovare continuamente i significati delle parole attraverso il suo utilizzo quotidiano²⁷.

²⁴ Ibid.

²⁵ Guido. Vetere, *Intelligenze aliene : linguaggio e vita degli automi* (Luca Sossella editore, 2025).

²⁶ Idem.

²⁷ Ferrara, *Le macchine del linguaggio : l'uomo allo specchio dell'intelligenza artificiale*.

Alla luce di quanto detto, quindi, è possibile affermare che i sistemi esperti e le basi di conoscenza presentano limiti strutturali tali da rendere impossibile una simulazione efficace del linguaggio umano²⁸.

1.2.2. La svolta statistica e i modelli basati sui corpora

La struttura con cui i sistemi esperti sono stati concepiti rende inevitabile il ricorso a componenti hardware sempre più sofisticate: essi, infatti, operando in domini altamente specializzati, richiedevano un impiego intensivo di risorse computazionali e un continuo aggiornamento delle basi di conoscenza. Questo ha inciso in modo significativo sui costi di gestione e manutenzione, che eccedevano ampiamente il valore restituito da tali sistemi²⁹. Il fallimento dell'industria dei sistemi esperti porta con sé il collasso dell'intero mercato dell'intelligenza artificiale verso la fine degli anni Novanta, provocando altresì un arresto improvviso alla ricerca accademica. Nella storia dell'intelligenza artificiale questo periodo di tempo viene ricordato con il nome di *AI Winter*, che allude alla completa perdita di interesse nel campo e alla conseguente assenza di progresso nella ricerca.

Successivamente a questa stagione di “carestia” si assiste a una svolta inaspettata che vede la nascita di un rinnovato interesse nell'ambito dell'intelligenza artificiale: importanti progressi nella produzione di microchip determinano una crescente potenza computazionale e, in concomitanza, l'espansione globale di Internet genera un'enorme quantità di nuovi dati³⁰. Inoltre, nello specifico ambito trattato in questa tesi, l'avvento dei social network apporta una tipologia di dati testuali estremamente utili, in quanto rappresentativi dei diversi modi di usare il linguaggio umano³¹. In tal senso, i progressi tecnologici e la crescita di Internet aprono la strada

²⁸ Idem.

²⁹ Chugani, Vinod. “AI Winter: Understanding the Cycles of AI Development”. Datacamp, 3 novembre 2025. <https://www.datacamp.com/blog/ai-winter>

³⁰ Amirhosein Toosi et al., «A Brief History of AI: How to Prevent Another Winter (A Critical Review)», in *PET Clinics*, vol. 16, n. 4, preprint, W.B. Saunders, 1 ottobre 2021, <https://doi.org/10.1016/j.cpet.2021.07.001>.

³¹ Ferrara, *Le macchine del linguaggio: l'uomo allo specchio dell'intelligenza artificiale*.

al secondo paradigma dell'intelligenza artificiale rappresentato dall'apprendimento automatico o *Machine Learning*.

Il *Machine Learning*, come già anticipato, è quel sottoinsieme di intelligenza artificiale basato su algoritmi in grado di apprendere da dati di addestramento e di generalizzare le conoscenze apprese su dati nuovi, svolgendo compiti senza la necessità di istruzioni esplicite. In un'ottica più completa, un modello di *machine learning* ha come obiettivo principale quello di ottimizzare le proprie prestazioni su un set di dati attraverso una fase di addestramento e, una volta ultimato, il modello sarà capace di formulare previsioni accurate su dati a esso sconosciuti. Oggi, è il paradigma dominante nella maggior parte dei modelli di intelligenza artificiale³².

Per approfondire la tecnologia sottostante al *machine learning* occorre distinguere tre paradigmi di apprendimento fondamentali: apprendimento non supervisionato, apprendimento supervisionato e apprendimento per rinforzo. Tali approcci differiscono non soltanto per l'obiettivo che il modello deve raggiungere, bensì anche per la natura dei dati richiesti nella fase di addestramento³³.

In termini più specifici, il primo metodo noto come apprendimento non supervisionato (o *unsupervised learning*) prevede un processo di analisi e di raggruppamento di dati non etichettati, con il fine di individuare pattern, dipendenze o correlazioni rilevanti al loro interno; considerato che il modello citato non dispone di output predefiniti né di obiettivi specifici, la supervisione umana risulta non necessaria. Gli algoritmi basati su apprendimento non supervisionato si dimostrano particolarmente efficaci nella tecnica del *clustering*, utile per raggruppare dati non etichettati in categorie omogenee secondo delle caratteristiche condivise.

A differenza dell'approccio precedente, l'apprendimento supervisionato (o *supervised learning*) si serve di un insieme di dati, denominato *training set*, per produrre output corretti in risposta a specifici input. Nel dettaglio, i dati di addestramento sono composti da coppie corrispondenti di input e output etichettati comunemente definiti *ground truth*, poiché verificati rispetto alla realtà osservabile; inoltre, descrivendo

³² Dave. Bergmann. "Che cos'è il machine learning?". IBM, consultato il 10 gennaio 2026. <https://www.ibm.com/it-it/think/topics/machine-learning>

³³ Ibid.

problemi di cui è nota la soluzione, il *training set* consente di orientare il modello verso i risultati attesi. Il processo di apprendimento si articola generalmente in una fase di addestramento, seguita da una fase di test. Nella fase iniziale, l'algoritmo elabora una grande quantità di dati per produrre una funzione (*loss function*) che associ le caratteristiche degli input con quelle degli output corrispondenti; quindi, attraverso un processo iterativo sui dati di input, il modello procede ad aggiornare i propri parametri con l'obiettivo di minimizzare l'errore. La fase successiva, detta di test, è necessaria per verificare il tasso di errore contenuto nella funzione di perdita e, di conseguenza, la qualità dell'addestramento. Inoltre, si prevede l'utilizzo di dati diversi dal *training set* per garantire una valutazione oggettiva dell'efficacia e della capacità di generalizzazione del modello: ciò permette di limitare il rischio di *overfitting*, vale a dire un'eccessiva aderenza del modello ai dati di addestramento³⁴.

Infine, l'apprendimento per rinforzo (o *reinforcement learning*), si distingue dagli altri paradigmi di *machine learning* per le modalità con cui l'algoritmo raggiunge l'obiettivo prestabilito. In termini generali, tale approccio consiste nella programmazione di un agente affinché impari ad agire in modo autonomo interagendo con l'ambiente circostante: esso prende decisioni sulla base di un sistema di premi e punizioni con lo scopo di individuare, tramite un processo di tentativi ed errori, le azioni che massimizzano la ricompensa complessiva nel lungo periodo.

Nel quadro relativo all'elaborazione del linguaggio naturale, il *machine learning* si presta a un'ampia varietà di scopi. Tra questi, innanzitutto, è possibile menzionare il sistema TF-IDF (*term frequency – inverse document frequency*), una tecnica impiegata per convertire il testo in rappresentazioni numeriche interpretabili dalle macchine. Il metodo si basa su un assortimento non strutturato dei termini contenuti in un corpus di documenti definito esclusivamente in base alla frequenza, ignorando ordine e contesto delle parole. La sua peculiarità risiede nel tentativo di invertire la tendenza di assegnare punteggi elevati a parole molto frequenti ma semanticamente irrilevanti,

³⁴ Ferrara, *Le macchine del linguaggio*, op. cit.

come articoli e congiunzioni³⁵. Tuttavia, il sistema TF-IDF è principalmente utilizzato per operazioni di ricerca e comparazione testuale e, quindi, non ha una vera utilità nell'ambito della generazione di testo³⁶.

In un'ottica sempre più vicina allo sviluppo di macchine capaci di generare testo, i modelli di *machine learning* trovano applicazione nel processo di tokenizzazione. Tale approccio costituisce il primo passo nell'elaborazione del linguaggio naturale, poiché le macchine processano l'informazione testuale secondo criteri diversi da quelli umani: hanno, infatti, la necessità di scomporre il testo in unità discrete e isolabili a cui attribuire dei valori quantitativi e delle forme di codifica³⁷. Se per l'essere umano l'isolamento di tali unità avviene in maniera quasi istintiva grazie a convenzioni ortografiche e morfologiche, per la macchina questo processo richiede un addestramento mediante algoritmi specifici. In passato la tokenizzazione era affidata a programmi basati su regole, che utilizzavano la punteggiatura e gli spazi come separatori; tuttavia, come è ben noto, i sistemi basati su regole presentano dei limiti intrinseci nell'ambito del linguaggio dati dalla loro eccessiva rigidità. Un cambiamento decisivo si afferma con l'adozione di metodi statistici di tokenizzazione, di cui *WordPiece*, sviluppato da Google, rappresenta un caso emblematico. In tali casi, la macchina si serve di un vasto corpus di testi per osservare la frequenza con cui appaiono specifiche sequenze di caratteri, imparando a identificare cosa sia una parola (o, meglio, un *token*) senza la necessità di conoscenze grammaticali pregresse. Nel dettaglio, il processo inizia con la scomposizione dell'intero testo nelle sue unità minime, ovvero i singoli caratteri; da questa operazione deriva un primo vocabolario costituito da una collezione di sottosequenze. In una fase successiva, l'algoritmo si serve dei singoli caratteri come base per la formazione di token più complessi attraverso un processo iterativo: individua così le occorrenze consecutive più frequenti e le fonde insieme, dando origine a un nuovo token che, infine, viene aggiunto al vocabolario. Poiché il ciclo iterativo è potenzialmente infinito viene

³⁵ Jacob Murel, [Eda Kavlakoglu, Cos'è una bag of words?](https://www.ibm.com/it-it/think/topics/bag-of-words), Ibm.com, <https://www.ibm.com/it-it/think/topics/bag-of-words>

³⁶ Ferrara, *Le macchine del linguaggio*, op. cit.

³⁷ Ivi.

stabilita una dimensione massima del vocabolario, al raggiungimento della quale il processo viene interrotto. In questo senso, è possibile che molte delle parole presenti nel corpus non siano incluse nel vocabolario e sebbene a un primo sguardo ciò possa apparire come una limitazione, tale caratteristica si rivela invece un punto di forza. In sostanza, le parole sconosciute al modello vengono trattate come sequenze di token distinti e, di conseguenza, scomposte in sottosequenze più piccole presenti nel vocabolario: ciò permette di isolare le radici dai suffissi che formano la flessione della parola senza la necessità di una teoria morfologica esplicita, mantenendo inoltre un vocabolario dalla dimensione piuttosto ridotta³⁸.

Su queste basi, il *machine learning* e il processo di tokenizzazione assumono un ruolo centrale nella capacità del modello di produrre testo: esso, infatti, non concepisce la scrittura come espressione di un pensiero, ma come la generazione sequenziale di token che, nel loro insieme, danno origine a un output di senso compiuto. In tale prospettiva, il vocabolario appreso durante la fase di tokenizzazione costituisce il punto di partenza per l'attività di scrittura automatica che, nello specifico, avviene mediante l'estrazione di più token e la loro conseguente concatenazione. Si tratta, tuttavia, di una descrizione che non coglie la complessità di tale processo e che trascura una serie di aspetti significativi. Il primo fra questi concerne indubbiamente la distribuzione di probabilità con cui le parole vengono estratte, in quanto, nell'uso concreto del linguaggio, alcune occorrono con maggiore frequenza rispetto ad altre; Per generare un testo verosimile, pertanto, il modello dovrà rispecchiare le proporzioni che regolano l'uso effettivo delle parole in un dato linguaggio. Per conseguire tale obiettivo è necessario contare quante volte ogni parola appare nel corpus e dividere la quantità ricavata per il numero totale delle parole, ottenendo così un valore compreso tra 0 e 1. Il passaggio immediatamente successivo è il cosiddetto campionamento, vale a dire il processo di estrazione basato sulle probabilità appena calcolate e che consiste nell'associare a ogni token un punteggio cumulativo composto dalla somma tra la propria probabilità e quella dei token che lo precedono. Infine, il modello estrae un numero casuale compreso tra 0 e 1 e individua la parola il cui

³⁸ Ivi.

punteggio cumulativo è il più piccolo valore maggiore o uguale al numero estratto. Considerate le premesse, emerge chiaramente come questo modello manifesti delle criticità evidenti e, soprattutto, fallisca nella produzione di testi dotati di senso compiuto. Tutto ciò è il risultato della seguente formula:

$$P(w_1, w_2, \dots, w_{n-1}, w_n) = P(w_1) \cdot P(w_2) \cdot \dots P(w_{n-1}) \cdot P(w_n)$$

Essa descrive la probabilità con cui si ottiene una specifica sequenza di n parole, ovvero con il prodotto delle probabilità con cui è stata estratta ogni parola. Il concetto statistico alla base di questa equazione è la probabilità congiunta, ovvero la probabilità che due o più eventi si verifichino simultaneamente; nello specifico, quando tali eventi sono indipendenti tra loro e si conosce la probabilità di ciascun singolo evento, è sufficiente moltiplicarle per calcolare la probabilità dell'evento congiunto. La debolezza strutturale del modello risiede proprio nel presupporre che l'estrazione di un token sia indipendente da quella del successivo e nel fondarsi, pertanto, sulle probabilità generali di occorrenza di ciascun token: da ciò deriva un testo privo di coerenza costituito da una sequenza totalmente casuale di parole³⁹.

Dal momento che il contesto rappresenta un elemento imprescindibile nella produzione di un testo, è necessario apportare delle profonde modifiche strutturali al modello affinché possa generare enunciati di senso compiuto. In ragione di ciò l'estrazione dei token non può più fondarsi su una scelta puramente casuale, ma risulta necessario condizionarla ai token precedenti, trasformandola di fatto in un evento caratterizzato da dipendenza: tale concetto, in statistica, è formalmente definito probabilità condizionata, ovvero la capacità di determinare la probabilità di un evento dato che un altro evento si è già verificato. Come è ben noto, in un testo coerente le relazioni di dipendenza non riguardano esclusivamente parole contigue, ma possono sussistere anche tra termini collocati a distanza all'interno della frase; di conseguenza, a ogni token viene assegnata una probabilità condizionata dall'intera sequenza. In una prospettiva più dettagliata, per poter estrarre la prima parola, tale approccio richiede

³⁹ Idem.

necessariamente l'inserimento di due token speciali⁴⁰ all'interno del corpus che delimitino l'inizio e la fine della frase, consentendo al modello di identificare sia le parole più appropriate da estrarre per iniziare la frase sia quelle per terminarla. Il funzionamento è il seguente: a partire dal token iniziale [#S], viene estratta una prima parola w_1 che viene aggiunta al testo in fase di generazione. Successivamente, si procede con l'estrazione della seconda parola, iterando il processo fino a estrarre il token di fine testo [#E]. In modo simile al modello della probabilità congiunta, è possibile ricavare una formula che spieghi come ottenere una specifica sequenza di n parole:

$$P(w_1, \dots, w_n) = P(w_1) \times P(w_2|w_1) \times \dots \times P(w_n|w_1, w_2, \dots, w_{n-1})$$

Nello specifico, tale probabilità si ottiene moltiplicando la probabilità di iniziare con la prima parola per la probabilità di estrarre la seconda parola dato che si è selezionata la prima e così via, aggiungendo man mano la sequenza che si è venuta a creare (riformulare). Inoltre, su questa formula si fonda il funzionamento di quei modelli comunemente noti come *language models*.

Analogamente ai loro predecessori, anch'essi presentano diverse problematiche. Infatti, sebbene tali modelli siano ora in grado di produrre testi coerenti e grammaticalmente corretti, mancano ancora di una vera capacità nel diversificare le frasi generate: per ottenere una generazione non deterministica, infatti, essi necessitano di un vasto corpus testuale a cui poter accedere e su cui potersi addestrare. Volendo intraprendere una soluzione alternativa, si potrebbe assegnare a tutti i token del vocabolario una probabilità minima di poter essere estratti; tuttavia, tale opzione non è consigliabile in quanto comporta il rischio di ricadere in un modello con estrazione indipendente. Un'ulteriore limitazione di questi modelli linguistici concerne la lunghezza delle sequenze generate: poiché l'estrazione di un token dipende dall'intera sequenza che lo precede, al crescere della lunghezza diventa

⁴⁰ Rispettivamente, il token [#S] per l'inizio e il token [#E] per la fine.

improbabile trovare corrispondenze identiche nel corpus di addestramento⁴¹. Questo si ricollega direttamente al problema della scarsa variabilità del modello, in quanto le sequenze molto lunghe finiscono per limitare le scelte disponibili durante la generazione, restituendo di fatto l'unica continuazione presente nel corpus⁴².

Una risposta apparentemente ovvia a questa problematica è quella di ridurre la lunghezza di tali sequenze. A questo proposito, l'espedito maggiormente impiegato nell'ambito dei sistemi di elaborazione del linguaggio naturale sono le catene di Markov: esse si fondano infatti sul principio che la probabilità di ciascun evento dipenda unicamente dallo stato presente senza considerare la sequenza di eventi precedenti. In termini più specifici circa il loro funzionamento, si stabilisce un parametro k che rappresenta il numero di token considerati dal modello per procedere all'estrazione della parola successiva; esso costituisce, inoltre, ciò che si definisce «iperparametro», ossia una variabile di configurazione definita in fase di progettazione che determina la struttura e il comportamento specifico del modello⁴³. Ad esempio, nel caso in cui k assuma valore 3, la macchina viene indotta a considerare solo le ultime tre parole della sequenza, eliminando dalla propria memoria quelle precedenti. Applicando tale strategia è possibile fornire al modello molteplici alternative tra cui scegliere durante il processo di estrazione e, aspetto ancora più rilevante, generare frasi diverse da quelle reperibili nel corpus di addestramento⁴⁴. Ciononostante, benché i modelli markoviani abbiano apportato dei vantaggi non indifferenti nel campo delle intelligenze artificiali generative, gli output prodotti presentano ancora incongruenze a livello contestuale riconducibili alle limitazioni imposte sulla lunghezza delle sequenze considerate⁴⁵.

⁴¹ Idem.

⁴² Vetere, *Intelligenze aliene*, op. cit.

⁴³ Ferrara, *Le macchine del linguaggio*, op. cit.

⁴⁴ Idem.

⁴⁵ Vetere, *Intelligenze aliene*, op. cit.

1.2.3. Le reti neurali e l'innovazione dei metodi di *word embedding*

Come si è visto, il *machine learning* statistico ha introdotto delle innovazioni significative nell'ambito dell'elaborazione del linguaggio naturale consentendo di superare i vincoli imposti dai sistemi basati su regole; permangono tuttavia delle criticità che verranno superate esclusivamente tramite l'adozione di tecnologie linguistiche più evolute, le quali saranno destinate a costituire il fondamento dei moderni modelli di intelligenza artificiale generativa. Un primo passo verso il superamento di tali limiti è rappresentato dall'impiego di una particolare categoria di algoritmi di *machine learning*: le reti neurali, che devono il loro nome a un'analogia con il funzionamento del cervello umano, basandosi sull'idea di riprodurre i meccanismi con cui esso elabora le informazioni e apprende per sviluppare macchine intelligenti.

Passando a un quadro più specifico, occorre fare una breve premessa: come già accennato e come anche evidenziato da Ferrara, le macchine sono in grado di elaborare esclusivamente dati numerici attraverso calcoli matematici, per quanto le moderne interfacce utente riescano a dissimulare tale natura. Risulta dunque evidente che una rete neurale necessita innanzitutto di un processo di codifica dei dati per poterli manipolare; tale procedura, denominata *encoding*, attua una conversione dei dati in vettori e matrici che consentono al modello di eseguire molteplici operazioni. Tra queste la più rilevante è la somma pesata, mediante la quale ciascun dato viene moltiplicato per un peso specifico prima di essere sommato agli altri, permettendo così al modello di attribuire maggiore o minore importanza ai diversi elementi. Nel contesto specifico di questa tesi, che si concentra su modelli di intelligenza artificiale applicati all'elaborazione del linguaggio naturale, la tipologia di dati principalmente considerata è quella dei dati categorici, ossia dati che «contengono riferimenti a entità, o spesso categorie, discrete, in cui ogni dato si riferisce non a una quantità, ma a un oggetto specifico, come una città o una stagione»⁴⁶. Proprio per l'assenza di un valore numerico intrinseco, i dati categorici si prestano difficilmente al processo di *encoding*; tuttavia, sono disponibili molteplici approcci per superare tale criticità. Un primo metodo coincide con il sistema TF-IDF, già affrontato nel sottocapitolo precedente,

⁴⁶ Ferrara, *Le macchine del linguaggio*, op. cit.

che consente di assegnare al dato testuale un valore corrispondente alla frequenza con cui appare nel corpus. In alternativa è possibile codificare ogni dato categorico in modo arbitrario attraverso un valore numerico, sebbene tale approccio possa generare interpretazioni erranee da parte della macchina. Questa problematica si risolve tramite il metodo *one-hot encoding*, che prevede la creazione di un vettore per ciascun dato categorico in cui si assegna il valore 1 alla categoria presente e 0 a tutte le altre, evitando così di attribuire un falso ordine numerico.

Terminata la fase di *encoding*, la rete neurale dispone di dati in input espressi in forma di vettori e parametri (o pesi) organizzati in forma matriciale. A questo punto, la macchina può procedere con le operazioni che caratterizzano questa particolare tipologia di modello: innanzitutto calcola il prodotto tra matrice e vettore per ottenere una somma pesata e, dopodiché, applica una funzione di attivazione al risultato. Tale funzione, però, non è sempre necessaria e viene impiegata esclusivamente in presenza di rapporti non lineari tra input e output. Naturalmente si tratta di una semplificazione: le reti neurali che operano in ambiti complessi come quello del linguaggio possiedono architetture molto più elaborate, costituite da numerose iterazioni dei due passaggi appena descritti ed è proprio questa profondità strutturale a giustificare l'appellativo di reti neurali profonde (o *deep neural networks*). In tal senso, in presenza di reti articolate su molteplici livelli (o *layers*), la quantità dei parametri raggiunge dimensioni tali da precluderne l'interpretazione, rendendo inspiegabile il comportamento della rete: da ciò deriva la definizione di *black boxes*.

Procedendo oltre, è opportuno riprendere il tema dell'apprendimento che, nel caso specifico delle reti neurali, è rappresentato interamente dai valori contenuti nella matrice dei pesi; e pertanto, in ragione della sua cruciale importanza, risulta necessario comprendere come l'algoritmo individui i valori più adatti, ossia quelli che riescono a massimizzare le prestazioni del modello. Il meccanismo implementato è un processo iterativo basato sull'apprendimento supervisionato: nello specifico, la rete parte da parametri casuali, producendo inevitabilmente delle previsioni errate. In seguito, mediante un *training set* contenente dati di cui si conosce la soluzione, la macchina confronta le proprie previsioni con i dati reali per quantificare l'errore complessivo.

A questo punto l'obiettivo diventa minimizzare tale errore e per farlo occorre modificare i parametri osservando il suo andamento al variare di questi cambiamenti; in tal modo è possibile determinare con precisione la direzione ottimale verso la quale modificare ulteriormente i pesi. La procedura appena esposta è chiamata calcolo del gradiente. Come illustrato da Ferrara in forma semplificata, il gradiente rappresenta un vettore orientato nella direzione di massima crescita di una funzione; ma dal momento che l'obiettivo finale consiste nel ridurre al minimo l'errore, allora sarà sufficiente aggiornare i parametri nella direzione opposta a quella indicata dal gradiente. Tale processo, invece, è definito discesa del gradiente. Una volta stabilita la direzione di aggiornamento dei pesi si rende necessario determinarne la quantità: attraverso il *learning rate*, infatti, è possibile fissare la velocità di apprendimento. A tal proposito, tipicamente viene adottato un approccio graduale che presuppone molteplici aggiornamenti dei parametri con il conseguente ricalcolo della funzione di errore. In conclusione, tale procedimento iterativo costituisce nella sua totalità l'apprendimento della rete neurale. Tuttavia, una volta avviata la discesa del gradiente, non è possibile prevedere quando l'algoritmo raggiungerà una funzione di perdita minima e per questo motivo si rende necessario adottare uno dei seguenti approcci: si può fissare un numero massimo di iterazioni per l'aggiornamento del *learning rate* oppure proseguire fino al raggiungimento di una condizione di convergenza in cui la riduzione dell'errore diventa trascurabile.

Per quanto concerne il tema dell'errore è opportuno precisare che non potrà mai essere annullato del tutto, poiché la rete neurale deve preservare la propria capacità di generalizzazione, ossia di produrre output corretti anche processando dati diversi da quelli del *training set*. Si può pertanto affermare che esista un tasso di errore fisiologico che consente al modello di svolgere efficacemente il suo compito⁴⁷.

Sono stati illustrati i limiti strutturali dei modelli probabilistici basati sull'osservazione della frequenza con cui occorrono determinate successioni di parole. Il superamento delle criticità imposte da tali meccaniche risiede nell'addestramento di

⁴⁷ Idem.

reti neurali finalizzate all'approssimazione di queste probabilità. Nello specifico, esistono diversi metodi per addestrare l'algoritmo a predire la parola più adeguata: il primo fra questi viene denominato *continuous bag of words* e può essere rappresentato come un gioco in cui il modello deve indovinare una parola centrale a partire da una finestra di parole di contesto circostanti. In alternativa si può utilizzare un metodo analogo chiamato *skip-gram*, sostanzialmente speculare al precedente e il cui obiettivo consiste nel predire le parole contestuali sulla base di una parola centrale data. Va inoltre precisato che il *training set* utilizzato per questa tipologia di addestramento non è necessariamente composto da frasi di senso compiuto, potendo consistere anche in un insieme non ordinato di parole. Come precedentemente illustrato, l'input testuale deve essere codificato affinché il modello possa interpretarlo correttamente, utilizzando per questo il metodo *one-hot encoding*; successivamente, la rete comprime l'informazione dell'input in un vettore di dimensioni molto più ridotte e arbitrarie chiamato *hidden layer*, per poi trasformarlo nel vettore di output finale. In tal senso, occorre evidenziare un aspetto rilevante che verrà ripreso in seguito: il vettore intermedio coincide esattamente con i valori numerici della matrice associati alla parola presa in esame per predire le altre. Quest'ultimo elemento rappresenta il principio fondante del metodo *Word2Vec*, che costituisce l'approccio più diffuso per la codifica semantica delle parole.

Tuttavia, prima di affrontare i meccanismi sottostanti a *Word2Vec*, è necessario fare un passo indietro e riflettere su alcuni concetti. Anzitutto, questi metodi si basano sulla supposizione che l'algoritmo, vista la sua capacità di predire quali parole compariranno nel contesto di un dato termine, ne comprenda effettivamente il significato e ciò conduce a interrogarsi sulla natura stessa del significato. Secondo la concezione tradizionale, esso risiede in ciò che la parola denota nel mondo reale; tale prospettiva è però preclusa alle macchine, che non dispongono di facoltà sensoriali per fare esperienza della realtà esterna, ma possono unicamente processare la sua manifestazione linguistica. In passato si è tentato di fornire alle macchine delle ontologie che definissero esplicitamente concetti e relazioni ma, come già visto nei primi paragrafi di questo capitolo, tali sistemi sono risultati troppo rigidi, difficili da

aggiornare e incapaci di gestire la fluidità del linguaggio naturale⁴⁸. La soluzione adottata dai moderni modelli linguistici è invece la semantica distribuzionale, fondata sul concetto cardine secondo cui le proprietà di un elemento linguistico sono deducibili esclusivamente dall'insieme dei contesti in cui esso compare all'interno di un corpus. In un'ottica più dettagliata, elementi che condividono distribuzioni simili vengono considerati appartenenti alla stessa classe, sia grammaticale che semantica. Il principio fondamentale di tale paradigma teorico può essere condensato nella nota citazione di Firth: «conoscerai una parola da quelle a cui si accompagna»⁴⁹. In conclusione, il passaggio a un approccio prettamente statistico come quello della semantica distribuzionale ha notevolmente agevolato il compito delle macchine, in quanto è sufficiente fare riferimento a un corpus di documenti per fornire loro la conoscenza necessaria alla ricerca del significato delle parole⁵⁰.

Resta da chiarire dove risieda esattamente il significato delle parole all'interno del modello. A questo proposito, in precedenza si era accennato all'esistenza di un livello intermedio (*hidden layer*) che contiene dei vettori numerici risultanti da una prima trasformazione dei dati in input; poiché la capacità della rete di predire correttamente i termini circostanti dipende esclusivamente da questi valori, ne consegue che, per la macchina, tali numeri costituiscono il significato della parola. Ciò spiega la denominazione di *word embedding* attribuita a tali metodi, poiché il termine *embedding* (dall'inglese "incastonare") allude al processo di cristallizzazione del significato, concetto astratto ed elusivo, in una rappresentazione vettoriale stabile. Infine, vale la pena notare che in questi casi l'addestramento della rete neurale è finalizzato esclusivamente a produrre i vettori contenuti nel livello intermedio⁵¹.

Passando agli aspetti tecnici, è opportuno chiarire le peculiari proprietà di questi vettori e, soprattutto, come possono essere utilizzati nello specifico ambito dell'elaborazione del linguaggio. In particolare, Ferrara definisce un vettore come una rappresentazione di «un punto in uno spazio multidimensionale in cui ogni numero

⁴⁸ Idem.

⁴⁹ Vetere, *Intelligenze aliene*, op. cit.

⁵⁰ Ferrara, *Le macchine del linguaggio*, op. cit.

⁵¹ Idem.

del vettore stesso è la coordinata rispetto a una delle dimensioni di tale spazio»⁵²; tuttavia, trattandosi di vettori molti grandi e di conseguenza non rappresentabili fisicamente, è necessario applicare delle tecniche di riduzione⁵³. Una volta ottenuta una rappresentazione visiva dello spazio vettoriale risulta possibile osservare diverse proprietà caratteristiche di tali vettori e la prima fra queste è che la loro distanza geometrica rispecchia la similarità semantica delle parole: prossimità spaziale implica affinità di significato e viceversa. Ciò apporta un significativo vantaggio rispetto alla rappresentazione tramite codifica *one-hot*, la quale tratta le parole in maniera binaria (come entità identiche o totalmente distinte), risultando incapace di cogliere le sfumature semantiche proprie del linguaggio umano. Una seconda peculiarità di tali vettori riguarda la capacità di eseguire operazioni matematiche sui concetti, nota come proprietà di composizionalità: nel dettaglio è possibile effettuare somme e sottrazioni per ricavare un vettore nuovo che mantenga relazioni semantiche coerenti. È importante precisare che il vettore risultante da tali operazioni non costituisce di per sé un nuovo significato, in quanto non corrisponde necessariamente a una parola presente nel corpus; semplicemente rappresenta una combinazione dei vettori di partenza che individua regioni semantiche contigue (addizione) od opposte (sottrazione) rispetto ai concetti originali. Un'ulteriore operazione resa possibile dalla composizionalità consiste nel riconoscimento di analogie mediante proporzioni vettoriali e ciò che più sorprende è che tali analogie coinvolgono spesso dimensioni semantiche figurate, ambiti tradizionalmente preclusi alle macchine.

Per concludere questa sezione sulle reti neurali occorre illustrarne i limiti. In primo luogo, tali modelli necessitano di corpora di addestramento estremamente ampi per conseguire un livello di generalizzazione soddisfacente, in quanto i significati appresi dalla macchina dipendono esclusivamente dalle relazioni che intercorrono tra le parole presenti nei *training set*. Ciò implica anche che i vettori sono di natura non contestuale, ossia non contengono al loro interno tutti i significati che una parola può assumere ma solamente uno, risultando inadeguati per la gestione della polisemia

⁵² Idem.

⁵³ Ferrara, *Le macchine del linguaggio*, op. cit.

lessicale. In secondo luogo, quest'ultima criticità dovuta alla non contestualità dei metodi di *word embedding* comporta un ulteriore limite: essi tendono a trattare il linguaggio come una semplice collezione di parole, perdendo l'informazione veicolata dall'ordine sequenziale e dalle relazioni che intercorrono tra i termini⁵⁴.

1.2.4. Elaborazione sequenziale con le Reti Neurali Ricorrenti (RNN)

Il tema concernente le reti neurali applicate al linguaggio rimane tuttavia una sfida aperta: se da un lato sono stati superati i limiti imposti dalle tecnologie sottostanti ai modelli markoviani, dall'altro persistono alcune problematiche legate alle modalità con cui tali reti elaborano le informazioni. Infatti, esse processano l'intera sequenza di input simultaneamente, comprimendola poi in un'unica rappresentazione vettoriale e, soprattutto, risultano prive di meccanismi di memoria per ritenere l'informazione elaborata in precedenza. In ragione di ciò, vista la necessità di elaborare informazioni in cui l'ordine delle parole e la dimensione temporale sono fondamentali per determinare il significato del messaggio, le reti neurali vengono sottoposte ad alcune modifiche che consentono di superare le limitazioni appena descritte. Nascono così le reti neurali ricorrenti (RNN).

In un'ottica più dettagliata, le RNN simulano un processo di lettura sequenziale, acquisendo l'informazione un elemento alla volta e a ciò si aggiunge un *hidden layer* che funge da vera e propria memoria della rete: una semplice lettura sequenziale, infatti, non sarebbe sufficiente se la rete non potesse conservare traccia delle informazioni già elaborate. Inoltre, senza una memoria interna, ogni trasformazione avverrebbe in modo isolato, costringendo l'algoritmo a basare le proprie predizioni esclusivamente sull'ultima parola osservata. Il livello nascosto mantiene la funzione di codifica dei significati osservata in *Word2Vec*, ma presenta questa volta una sostanziale innovazione tecnica nel proprio meccanismo di aggiornamento: quando la rete processa una nuova informazione, non utilizza solo il vettore che rappresenta l'informazione attuale, ma la combina con il vettore interno

⁵⁴ Ferrara, *Le macchine del linguaggio*, op. cit.

prodotto al passo precedente. Questo significa che la rappresentazione di una parola cambia a seconda di ciò che l'ha preceduta.

Per compiti più complessi come la traduzione automatica, le RNN vengono spesso organizzate in un'architettura *encoder-decoder*. In questa configurazione sono previste due reti che svolgono funzioni distinte: la prima (l'encoder) ha il compito di leggere e codificare l'intera frase di input nel proprio livello nascosto, che viene poi passato a una seconda rete (il decoder) per guidare la generazione della sequenza di output parola dopo parola.

Nonostante queste potenzialità, le RNN presentano dei limiti tecnici intrinseci che hanno indotto un'ulteriore evoluzione tecnologica, la quale sarà oggetto di analisi più avanti nel capitolo. Innanzitutto, Ferrara evidenzia un problema di diluizione dell'informazione in presenza di sequenze particolarmente lunghe: poiché l'informazione iniziale deve essere propagata attraverso numerosi passaggi matematici, quella relativa alle prime parole tende a scomparire progressivamente. Ciò rende complicato cogliere relazioni tra termini distanti fra loro. Inoltre, la natura sequenziale dell'apprendimento costituisce un ostacolo per l'efficienza di tali macchine, in quanto ogni operazione dipende necessariamente dal completamento della precedente. Infine, la fase di calcolo del gradiente, data la sua propagazione attraverso molteplici passaggi, può portare a dei risultati instabili rendendo il processo di apprendimento tecnicamente più complesso.

1.3. Lo stato dell'arte attuale: i transformer

Dall'inizio del capitolo è emerso come l'intelligenza artificiale abbia cercato di approssimare l'ambito del linguaggio umano in diversi modi: partendo dall'idea che le si potesse insegnare a scrivere fornendole delle conoscenze esplicite, poi automatizzando il processo di apprendimento simulando i meccanismi del cervello umano, fino ad arrivare all'impresa di cristallizzare il significato delle parole in rappresentazioni numeriche complesse. Tuttavia, sebbene ogni evoluzione tecnologica sia riuscita a colmare numerose lacune, nessuno di questi approcci si è rivelato particolarmente soddisfacente.

1.3.1. Il *positional encoding*

Le problematiche più urgenti riguardano le sequenze di token eccessivamente lunghe che, per motivi intrinseci ai modelli fin qui sviluppati, costituiscono una sfida ancora irrisolta per tali macchine. Infatti, i meccanismi sottostanti alle RNN si rivelano inefficaci nel cogliere i rapporti che intercorrono tra le parole di una frase, considerando prevalentemente la loro sequenzialità⁵⁵.

Una prima innovazione che ha permesso ai modelli di intelligenza artificiale di avvicinarsi alla comprensione umana dei testi scritti è il *positional encoding*. Tale meccanismo nasce dall'esigenza di abbandonare l'elaborazione sequenziale degli input, risultata poco efficace, per sfruttare appieno le potenzialità dei computer moderni che consentono di eseguire operazioni di calcolo parallele; tuttavia, tale scelta comporta la perdita della capacità della macchina di distinguere l'ordine delle parole, trattandole come elementi indipendenti.

Per risolvere questo problema viene introdotta la codifica posizionale, un processo che trasforma la posizione di ciascun termine all'interno di una sequenza in un dato esplicito e manipolabile. Nello specifico, tale procedura prevede inizialmente la codifica dell'input in una rappresentazione vettoriale (*embedding*), la quale però genera il medesimo vettore per ogni occorrenza di un termine indipendentemente dalla sua posizione nella sequenza; per introdurre una distinzione, la rete genera quindi un secondo vettore, detto "di posizione". Esso viene costruito applicando una formula matematica fissa che genera valori numerici basandosi esclusivamente sulla posizione specifica della parola e sulla dimensione del vettore di *encoding* scelto. Un aspetto cruciale da tenere in considerazione è che i vettori di posizione sono totalmente indipendenti dal significato della parola a cui sono applicati, tanto che frasi diverse con lo stesso numero di termini condivideranno gli stessi identici vettori posizionali pur avendo vettori di *embedding* completamente differenti. Infine, i due vettori vengono combinati tramite una somma, producendo un unico vettore finale

⁵⁵ Vetere, *Intelligenze aliene*, op. cit.

che condensa in sé sia l'identità della parola sia la sua collocazione spaziale nella frase⁵⁶.

1.3.2. *Attention is all you need*⁵⁷

Il secondo pilastro alla base della tecnologia *transformer* è il meccanismo di attenzione, che nasce dalla necessità di superare nuovamente le limitazioni imposte dall'elaborazione sequenziale delle reti neurali ricorrenti: in particolare, tale procedimento tende a privilegiare l'ultimo elemento processato a scapito dei primi. L'obiettivo è, perciò, di simulare un processo di lettura simile a quello umano, nel quale il peso attribuito a ogni parola non è vincolato dalla sua posizione nella sequenza ma alla sua effettiva rilevanza informativa per la comprensione del testo.

In un quadro più dettagliato, il concetto alla base dell'attenzione è quantificare quanta importanza abbia una determinata parola per interpretare tutte le altre e, allo stesso tempo, quanto le altre parole siano rilevanti per interpretare il termine osservato. Trattandosi di mettere in relazione delle parole, si sceglie di rappresentare l'attenzione attraverso una matrice composta dallo stesso numero di righe e colonne, le quali coincidono con i token contenuti nella sequenza. A questo proposito, esistono due tipologie di attenzione, che rispondono a esigenze diverse: la *self-attention* e la *cross-attention*. Nella prima si mettono in relazione le parole di una singola frase per acquisire una rappresentazione profonda del proprio contesto. La seconda, invece, è utile per operazioni di generazione di testo in risposta a un *prompt* e, quindi, al suo interno pone in relazione due frasi distinte: la sorgente e il testo generato.

Passando a un'analisi più tecnica, il meccanismo di attenzione è rappresentato da tre matrici. La matrice Q si ottiene moltiplicando la matrice contenente i vettori della frase in input per un primo insieme di parametri, mentre la matrice K deriva dalla moltiplicazione degli stessi vettori per un secondo insieme di parametri. Il risultato finale è la matrice V, che contiene i vettori di *embedding* e codifica posizionale arricchiti dall'informazione calcolata dall'attenzione.

⁵⁶ Ferrara, *Le macchine del linguaggio*, op. cit.

⁵⁷ Ashish Vaswani et al., *Attention Is All You Need* (2023).

1.3.3. L'architettura *encoder-decoder*

Analogamente alle RNN, anche i *transformer* sfruttano l'architettura *encoder-decoder* per conseguire l'obiettivo di generare un testo di output a partire da un input.

Il processo inizia nell'*encoder*, dove la frase di input viene convertita in una matrice numerica attraverso i meccanismi di *embedding* e *positional encoding*, fornendo al modello informazioni sia sul significato delle parole sia sulla loro posizione relativa nella sequenza. Successivamente interviene il meccanismo di *multi-head attention*, che scompone il calcolo dell'attenzione in più componenti indipendenti elaborate in parallelo; questo permette alla rete di catturare simultaneamente aspetti diversi del linguaggio, come relazioni sintattiche tra parole vicine o dipendenze semantiche tra termini distanti. L'ultimo passaggio dell'*encoder* consiste nell'utilizzo di una rete neurale *feed-forward*, che processa indipendentemente ciascun termine della frase per arricchirne la rappresentazione interna.

Il procedimento continua con l'attivazione del *decoder* che, almeno inizialmente, segue un percorso analogo applicando al testo meccanismi di *embedding* e di *positional encoding*. L'unica differenza consiste nell'introduzione di un accorgimento tecnico durante la fase di *multi-head attention*: data la natura auto-regressiva del *decoder* e dato l'obiettivo di simulare un processo di scrittura in cui è possibile guardare a ciò che è già stato redatto ma non a ciò che deve ancora essere prodotto, occorre “nascondere” le parole future della sequenza assegnandogli un peso di attenzione nullo. Tale meccanismo è chiamato *masked multi-head attention*.

A questo punto, però, non è ancora avvenuta un'associazione tra la frase in input (*prompt*) e quella in output. L'elemento di congiunzione tra le due componenti, infatti, è rappresentato dalla *cross-attention*, un meccanismo in cui il decoder mette in relazione le parole che sta generando con l'informazione codificata fornita dall'encoder. In questo modo, il modello può determinare quali parti dell'input richiedano maggiore attenzione per selezionare correttamente il token successivo, rendendo superflua la presenza di un dizionario esplicito per compiti come la traduzione.

1.4. La traduzione automatica nella storia

Delineata l'evoluzione dei modelli di intelligenza artificiale che si sono susseguiti negli ultimi settant'anni, è ora possibile contestualizzarli nell'ambito specifico della traduzione automatica o *machine translation*. Essa è, per definizione, la «traduzione di un testo da parte di un programma senza intervento umano»⁵⁸ ed è una sottocategoria del più ampio campo denominato *Natural Language Processing* (NLP).

1.4.1. Gli inizi: dagli anni '50 agli anni '60

Sebbene l'idea di un linguaggio universale risalga già a pensatori come Cartesio nel XVII secolo, le moderne basi scientifiche per la traduzione automatica vengono gettate negli anni '40 del secolo scorso⁵⁹. Formalmente, l'inizio delle ricerche in tale campo viene attribuito a Warren Weaver, allora direttore della *Natural Science Division* della Rockefeller Foundation: in una lettera al matematico Norbert Wiener si interroga sulla possibilità di utilizzare i computer per tradurre testi scientifici, proponendo di affrontare la questione come un problema di crittografia.

Gli anni successivi si caratterizzano per un'inarrestabile ricerca sulla traduzione automatica, coinvolgendo molteplici enti del mondo universitario americano. Nel 1951, infatti, il Laboratorio di Ricerca per l'Elettronica del MIT incarica Yehoshua Bar-Hillel (matematico e linguista) di valutare lo stato e le prospettive della ricerca: nel suo report conclude che una traduzione automatica di alta qualità non sarebbe stata possibile nel breve termine e, pertanto, raccomanda un modello ibrido basato sulla supervisione umana⁶⁰.

Solo tre anni più tardi, nel 1954, viene realizzata la prima dimostrazione pubblica di un progetto di *machine translation*. La Georgetown University, in collaborazione con la celebre IBM, porta a termine con successo un esperimento di traduzione automatica dal russo all'inglese mediante un sistema basato su un ristretto insieme di regole sintattiche e un vocabolario di 250 parole. Tale dimostrazione si

⁵⁸ [https://www.treccani.it/enciclopedia/traduzione-automatica_\(Lessico-del-XXI-Secolo\)/](https://www.treccani.it/enciclopedia/traduzione-automatica_(Lessico-del-XXI-Secolo)/)

⁵⁹ Chan. Sin-wai, *The Routledge encyclopedia of translation technology* (Routledge, 2015), 105

⁶⁰ Baron, *Who wrote this?*, op. cit., 85

configura come un momento cruciale nella storia della traduzione automatica, attestandone la concreta fattibilità; inoltre, il notevole impatto mediatico attira un'ingente quantità di investimenti negli Stati Uniti e stimola altri Paesi a intraprendere iniziative analoghe⁶¹.

1.4.2. La condanna dell'ALPAC

Negli anni seguenti l'entusiasmo per la traduzione automatica raggiunge il suo apice, complici i continui progressi della linguistica che sembravano promettere miglioramenti continui nella qualità delle traduzioni. Tuttavia, la complessità semantica del linguaggio naturale e l'incapacità di trovare soluzioni adeguate bastano a determinare un graduale ridimensionamento delle aspettative⁶².

In tal senso, nel 1964 la *National Academy of Sciences* chiede la formazione di una task force che potesse valutare l'effettiva utilità delle ricerche sulla traduzione automatica: la *Automatic Language Processing Advisory Committee* (ALPAC). Il report arriva solo due anni dopo, spegnendo definitivamente le speranze di ricercatori e investitori: il comitato, infatti, decreta che la traduzione automatica non offre alcun vantaggio rispetto a quella umana, che risulta come la soluzione più economica e veloce⁶³. Come diretta conseguenza, i finanziamenti nel settore subiscono un brusco arresto.

1.4.3. La ricerca non si ferma

Se negli Stati Uniti il peggioramento dei rapporti con l'Unione Sovietica durante la guerra fredda spinge a concentrare le ricerche sulla traduzione in inglese di materiale scientifico russo, Canada ed Europa affrontano esigenze diverse legate alla presenza di molteplici lingue ufficiali⁶⁴.

⁶¹ Ivi, 106

⁶² Ibid.

⁶³ Baron, *Who wrote this?*, op. cit. 86

⁶⁴ Sin-wai, *The Routledge encyclopedia of translation technology*, 107

Ad esempio, a Montréal nasce il progetto TAUM (*Traduction Automatique à l'Université de Montréal*) che, inizialmente, si concentra sulla traduzione dei bollettini metereologici: scelta apparentemente peculiare, che viene tuttavia giustificata dal ridotto lessico utilizzato in tale ambito. Il sistema garantisce un'altissima qualità e riduce drasticamente il carico di lavoro dei traduttori umani⁶⁵. Contemporaneamente, l'azienda SYSTRAN lancia i primi sistemi commerciali, utilizzati dalla Commissione Europea a partire dal 1976 e da Google fino al 2007⁶⁶.

1.4.4. L'avvento dei metodi basati sui corpora e lo stato attuale

Come già illustrato, verso la fine degli anni Ottanta si assiste al declino dei sistemi basati su regole a favore di modelli statistici fondati su vasti dataset denominati corpora. Da un lato, Makoto Nagao propone la traduzione basata su esempi (*Example-Based Machine Translation*, EBMT), che traduce per analogia recuperando coppie di frasi simili da set di dati preesistenti⁶⁷; dall'altro, i ricercatori dell'IBM con il progetto *Candide* rivoluzionano il campo con la *Statistical Machine Translation* (SMT)⁶⁸.

I successivi progressi nell'ambito della traduzione automatica sono segnati dall'introduzione delle reti neurali: a differenza dei modelli statistici esse comprimono il testo fonte in rappresentazioni semantiche dense, simulando un'effettiva conoscenza dei significati⁶⁹. Tuttavia, il salto di qualità definitivo si compie nel 2017 con l'architettura Transformer basata sul meccanismo di attenzione, il quale consente al modello di processare contesti semantici molto più ampi.

⁶⁵ Baron, *Who wrote this*, op. cit., 88

⁶⁶ Sin-wai, *The Routledge encyclopedia of translation technology*, 107

⁶⁷ Wang H et al., *Progress in Machine Translation* (Elsevier, 2022), 144

⁶⁸ Sin-wai, *The Routledge encyclopedia of translation technology*, op. cit., 109

⁶⁹ Wang H et al., *Progress in Machine Translation*, op. cit., 144

Capitolo 2 – Come traduce la macchina? Valutare e riconoscere il marchio dell'IA nelle traduzioni automatiche

2.1. Valutare la qualità delle traduzioni automatiche

Oggi, valutare la qualità delle traduzioni automatiche rappresenta un ambito di studio estremamente articolato, nel quale dimensioni cognitive, linguistiche e tecnologiche si intrecciano per dare risposta alle crescenti incognite sull'abilità dei modelli di intelligenza artificiale di effettuare con successo operazioni fin qui ritenute di carattere prettamente umano. Tale complessità riflette l'altrettanto intricata natura del processo traduttivo, la cui essenza non è ancora completamente definibile⁷⁰.

Nell'attuale scenario caratterizzato dalla continua espansione di contenuti digitali e, pertanto, di possibili materiali linguistici da tradurre, il confine tra umano e macchina appare sempre più sfumato a causa dell'integrazione di strumenti tecnologici di traduzione nei processi di lavoro. In tal senso, il concetto stesso di qualità abbandona i propri standard tradizionali per abbracciare criteri pragmatici variabili in funzione del contesto, della tipologia testuale, dei processi lavorativi e dell'utenza finale⁷¹.

Generalmente, la valutazione della qualità si divide tra approcci automatici e umani: come verrà illustrato, entrambi offrono molteplici vantaggi ma presentano anche altrettante limitazioni strutturali e intrinseche.

2.1.1. Metriche di valutazione automatiche: i limiti di BLEU

La valutazione della qualità mediante metriche automatiche implica l'utilizzo di algoritmi progettati per comparare l'output della macchina con una o più traduzioni

⁷⁰ Joss Moorkens et al., *Translation Quality Assessment From Principles to Practice Machine Translation: Technologies and Applications Series Editor: Andy Way* (s.d.), <http://www.springer.com/series/15798>, 10.

⁷¹ Ivi, 2

di riferimento⁷²; pertanto, tali metriche mirano a quantificare attraverso un punteggio numerico la vicinanza fra la traduzione automatica e quella umana di riferimento (o *gold standard*), che si presume sia di qualità superiore⁷³. A questo proposito, Tavosanis ritiene inesatto definire “automatici” tali sistemi, preferendo una dicitura più precisa come “semiautomatici”⁷⁴.

Nel novero delle metriche più utilizzate figura BLEU (*Bilingual Evaluation Understudy*), divenuto lo standard de facto grazie a un’elevata corrispondenza con i giudizi delle valutazioni umane. A livello tecnico, il suo funzionamento prevede la misurazione della precisione calcolando quanto spesso una sequenza di parole (n-grammi) appare sia nell’output generato dalla macchina sia nelle traduzioni di riferimento⁷⁵. Infine, BLEU applica un fattore correttivo denominato *brevity penalty* che, come si può intuire, penalizza i testi significativamente più brevi rispetto ai riferimenti umani. Nello specifico tale meccanismo si rende necessario poiché, per consentire variazioni nella scelta lessicale dell’output, il sistema non pone alcun vincolo all’ordine in cui occorrono sequenze corrispondenti e, di conseguenza, potrebbe attribuire punteggi artificialmente alti a traduzioni troppo brevi⁷⁶.

Nonostante la sua ampia diffusione dovuta all’oggettività, alla rapidità e ai bassi costi di esecuzione, BLEU presenta problematiche strutturali significative che derivano dalla sua natura puramente statistica e, con l’avvento delle reti neurali nel campo della traduzione automatica, ha mostrato i suoi limiti: questi modelli, infatti, possono generare output semanticamente migliori ma formalmente distanti dai riferimenti umani, rendendo i punteggi del tutto inattendibili⁷⁷.

⁷² Irene. Rivera-Trigueros, «Machine Translation Systems and Quality Assessment: A Systematic Review». *Language Resources and Evaluation* 56, fasc. 2 (2021): 593–619. <https://doi.org/10.1007/s10579-021-09537-5>, 596.

⁷³ Moorkens et al., *Translation Quality Assessment*, op. cit., 25.

⁷⁴ Mirko. Tavosanis, «Valutare La Qualità Dei Testi Generati in Lingua Italiana». *AI-Linguistica. Linguistic Studies on AI-Generated Texts and Discourses* 1, fasc. 1 (2024). <https://doi.org/10.62408/ai-ling.v1i1.14>, 6

⁷⁵ Mirko. Tavosanis, *Lingue e intelligenza artificiale*. Carocci editore, 2018, 100.

⁷⁶ Moorkens et al., *Translation Quality Assessment*, op. cit., 168.

⁷⁷ Tavosanis, «Valutare La Qualità Dei Testi Generati in Lingua Italiana» op. cit., 6.

Si può concludere che le metriche di valutazione automatica risultano insufficienti se impiegate isolatamente; pertanto, si rende necessaria l'adozione di approcci olistici che integrino valutazione automatica e valutazione umana per compensare l'incapacità delle prime di rilevare corrispondenze sintattiche e semantiche nelle traduzioni automatiche⁷⁸.

2.1.2. Metriche di valutazione umane e classificazione degli errori

Date le premesse, la valutazione umana rimane lo strumento più accurato per analizzare fenomeni linguistici complessi, nonostante comporti una significativa onerosità in termini di tempi e costi⁷⁹. Essa si articola tipicamente su due criteri fondamentali: l'adeguatezza (*adequacy*), la quale valuta in che misura il significato del testo originale è stato preservato nel testo tradotto, e la fluidità (*fluency*), che invece prende in esame la correttezza sintattica dell'output⁸⁰. I metodi di valutazione più utilizzati per tali parametri sono spesso basati su indicatori molto semplici, come scale Likert o scale numeriche da 1 a 5, poiché un metro di giudizio più esteso risulterebbe fonte di maggiore difficoltà discriminativa⁸¹.

Un ulteriore livello di analisi consiste nel processo di identificazione, classificazione e analisi degli errori, fondamentale per determinare i fallimenti specifici dei sistemi di traduzione automatica e spesso supportato da approcci standardizzati⁸². In passato, la metodologia più diffusa prevedeva l'utilizzo di liste di errori che, tuttavia, presentavano il limite di un approccio indifferenziato e spesso incapace di adattarsi alle specificità delle diverse tipologie di testo o di settore: è in tale contesto che si inseriscono nuovi progetti orientati a migliorare il processo di standardizzazione, tra i quali figura il *Multidimensional Quality Metrics* (MQM)⁸³. A differenza dei suoi predecessori, il progetto MQM nasce con l'obiettivo di creare un

⁷⁸ Moorkens et al., *Translation Quality Assessment*, op. cit., 26.

⁷⁹ Rivera-Trigueros, «Machine Translation Systems» op. cit., 597.

⁸⁰ Moorkens et al., *Translation Quality Assessment*, op. cit., 18.

⁸¹ Ivi, 164.

⁸² Rivera-Trigueros, «Machine Translation Systems» op. cit., 598.

⁸³ Moorkens et al., *Translation Quality Assessment*, op. cit., 15.

vocabolario comune per la classificazione degli errori di traduzione che preveda al contempo un'elevata flessibilità di utilizzo; tale caratteristica è riconducibile alla possibilità per l'utilizzatore di non applicare necessariamente l'intero catalogo di errori, ma di poter selezionare all'occorrenza le tipologie più rilevanti⁸⁴.

In un'ottica più dettagliata, il sistema MQM si configura come una gerarchia ramificata in otto diverse dimensioni, che coprono aspetti sia linguistici sia extra-linguistici della qualità. Tra le più rilevanti figurano le già citate accuratezza e fluidità, affiancate da ulteriori dimensioni dal carattere innovativo come la veridicità (*verity*), che analizza la relazione tra il contenuto e il mondo reale di cui fa parte: essa permette di identificare casistiche che, pur non dimostrando problematicità a livello grammaticale, risultano inappropriate per il contesto in cui appaiono⁸⁵.

Come già delineato, un aspetto essenziale del metodo MQM risiede nel suo legame con le specifiche del progetto, abbracciando una visione funzionalista della traduzione. Secondo tale prospettiva non è opportuno considerare la qualità come un valore assoluto, poiché viene determinata dal grado di aderenza alle finalità intrinseche della traduzione: ciò impedisce di penalizzare scelte stilistiche che sono in realtà coerenti con gli obiettivi prefissati⁸⁶.

Come ultima considerazione, occorre sottolineare che il semplice conteggio degli errori risulta insufficiente quando si valuta una traduzione e, pertanto, per raggiungere un livello di analisi più approfondito occorre classificarli secondo la loro gravità e la loro importanza. Più precisamente, il modello MQM definisce quattro livelli di gravità, in ordine decrescente di criticità. In primo luogo, per l'appunto, figurano gli errori critici, la cui sola presenza rende la traduzione inadeguata e inutilizzabile; in seguito, è la volta degli errori gravi che, trasponendo impropriamente il significato dall'originale, costituiscono un ostacolo considerevole alla comprensione del testo. Il penultimo livello, invece, riguarda gli errori lievi, i quali incidono in modo marginale sulla traduzione e possono essere facilmente corretti dal valutatore; infine,

⁸⁴ Ivi, 113.

⁸⁵ Ivi, 118.

⁸⁶ Idem.

si riconosce un livello nullo corrispondente alla mancanza di errori. Al contrario, l'importanza consiste nell'attribuzione di pesi alle diverse categorie di errori: nel dettaglio, si assegnano pesi diversi a specifiche categorie per indicare la loro rilevanza relativa e permettere alle metriche di adattarsi di conseguenza⁸⁷.

2.2. Le tendenze stilistiche dei testi artificiali

Dopo aver illustrato il quadro di riferimento dei protocolli mediante i quali viene effettuata la valutazione della qualità delle traduzioni automatiche, è ora possibile concentrarsi su un'ulteriore dimensione di analisi. In tal senso, data la vasta quantità di contenuti testuali prodotti dall'intelligenza artificiale generativa, si rende necessario osservare con attenzione l'esistenza di vere e proprie cifre stilistiche che permettono di distinguere nitidamente la scrittura umana da quella artificiale (o sintetica, come definita da De Cesare⁸⁸).

Nei testi generati artificialmente, tali caratteristiche distintive permangono a prescindere da eventuali indicazioni stilistiche che l'utente fornisce alla macchina attraverso il *prompt*⁸⁹: la ragione di una tale stabilità espressiva trova la sua origine principalmente nella tendenza a ottimizzare le probabilità basandosi sui dati più comuni invece di replicare il complesso insieme di fattori che guida la produzione del linguaggio umano⁹⁰. Questo quadro di sostanziale omogeneità ha indotto alcuni studiosi a teorizzare l'emergere di una «*algorithmic monoculture*»⁹¹, ossia una progressiva uniformazione di contenuti e stile che si pone in netto contrasto con la varietà stilistica propria della creatività umana. L'analisi stilometrica conferma questa tendenza: nonostante la crescente sofisticatezza dei modelli linguistici di grandi dimensioni, i

⁸⁷ Ivi, 120-121.

⁸⁸ Anna-Maria. De Cesare, *L'italiano sintetico dell'intelligenza artificiale generativa* (Franco Cesati Editore, 2026).

⁸⁹ Bitton, Yehonatan, Elad Bitton, e Shai Nisan. «Detecting Stylistic Fingerprints of Large Language Models». Versione 1. Preprint, arXiv, 2025. <https://doi.org/10.48550/ARXIV.2503.01659>, 3.

⁹⁰ Guo, Yanzhu, Guokan Shang, e Chloé Clavel. «Benchmarking Linguistic Diversity of Large Language Models». Versione 2. Preprint, arXiv, 2024. <https://doi.org/10.48550/ARXIV.2412.10271>, 1.

⁹¹ Chakrabarty, Tuhin, Philippe Laban, e Chien-Sheng Wu. «Can AI Writing Be Salvaged? Mitigating Idiosyncrasies and Improving Human-AI Alignment in the Writing Process through Edits». Versione 5. Preprint, arXiv, 2024. <https://doi.org/10.48550/ARXIV.2409.14509>, 2.

testi sintetici mostrano comunque delle impronte stilistiche che li differenziano da quelli umani⁹².

A titolo esemplificativo verranno qui di seguito riportate alcune di queste tendenze, tra le quali si configura l'uso pervasivo di cliché e figure retoriche banali, fenomeno ritenuto intrinseco alle tecnologie sottostanti ai modelli linguistici: poiché il testo viene generato mediante calcoli probabilistici, le espressioni convenzionali risultano quindi altamente probabili diventando quasi inevitabili per la macchina⁹³. In seguito è possibile osservare la tendenza a una certa verbosità tramite il ricorso alla cosiddetta «*purple prose*», che identifica una scrittura eccessivamente elaborata e ornamentale caratterizzata da un uso sproporzionato di elementi testuali ostacolando così il flusso narrativo; si riscontra altresì una scrittura ridondante, la cui conseguenza più notevole consiste nel rivelare i significati sottintesi nel testo e, pertanto, impedire al lettore di inferirli autonomamente dalla narrazione⁹⁴. A livello sintattico i modelli linguistici prediligono una sintassi povera, spesso contraddistinta da strutture ripetitive che ricorrono con una frequenza sproporzionata, come anche da un uso eccessivo della paratassi e dalla mancanza di connettivi logici che, complessivamente, compromettono la fluidità del testo.

Un ultimo fondamentale tratto distintivo rispecchia un limite noto di tali tecnologie: laddove i testi umani eccellono nella descrizione di dettagli concreti e sfumature di significato, i testi artificiali privilegiano affermazioni generiche distinguendosi per una generale mancanza di profondità nella scrittura, che al contrario caratterizza voci autoriali autentiche. Di conseguenza, benché i modelli linguistici siano in grado, nel complesso, di produrre testi corretti a livello superficiale, faticano a mantenere una coerenza tematica e stilistica su testi di lungo respiro,

⁹² O'Sullivan, James. «Stylometric Comparisons of Human versus AI-Generated Creative Writing». *Humanities and Social Sciences Communications* 12, fasc. 1 (2025). <https://doi.org/10.1057/s41599-025-05986-3>, 4.

⁹³ Chakrabarty et al. «Can AI Writing Be Salvaged?» op. cit., 25.

⁹⁴ Ivi, 10.

commettendo spesso errori nella *consecutio temporum* e incorrendo in cambiamenti di tono repentini⁹⁵.

Infine occorre sempre ricordare che tali testi, essendo generati da agenti automatici, sono privi di una reale intenzionalità comunicativa e operano esclusivamente attraverso sistemi probabilistici. La questione dell'intenzione, infatti, riveste un ruolo centrale nell'affrontare la più ampia tematica della traduzione applicata all'intelligenza artificiale: come chiaramente illustrato da Eco, la presenza di una «*intentio operis*» è di fondamentale importanza per giustificare scelte traduttive che a primo impatto sembrano deviare dall'originale⁹⁶. Tale prospettiva più etica verrà ulteriormente esplorata nel capitolo successivo e sarà avvalorata da testimonianze di professionisti del settore.

2.3. Metodologia

Una volta esposte le principali tendenze stilistiche ricorrenti nei testi artificiali, è ora possibile approfondire e verificare tale fenomeno empiricamente mediante un piccolo esperimento di traduzione.

2.3.1. Definizione dell'obiettivo

Il presente esperimento si propone di analizzare il comportamento di tre diversi modelli linguistici in ambito traduttivo, più specificamente nella traduzione di brevi testi narrativi generati artificialmente dall'inglese all'italiano, prestando particolare attenzione alle cifre stilistiche ricorrenti nei rispettivi output.

In tal senso, anziché ricorrere all'uso di approcci tradizionali nella valutazione delle traduzioni automatiche già menzionati in precedenza, lo studio parte dalla consapevolezza che la traduzione letteraria non costituisce un processo univoco ma, al contrario, implica necessariamente una componente interpretativa. Valutare le

⁹⁵ Jasim, Yaser A., e Ali Jala Awqati. «Distinguishing Human Creativity from AI-Generated Literary Texts». JOURNAL OF PROSPECTIVE RESEARCHES 25, fasc. 2 (2025): 40–47. <https://doi.org/10.61704/pr.494>, 44.

⁹⁶ Eco. *Dire quasi la stessa cosa*, op. cit., 13.

prestazioni di tali modelli, pertanto, presenta non poche difficoltà determinate soprattutto dalla presenza di molteplici variabili, tra le quali si configurano il prompt, i dati di addestramento e l'eventuale presenza di alcuni parametri come la temperatura e il *top-p*⁹⁷. Ne consegue che non è possibile attribuire oggettivamente le caratteristiche di ogni singola traduzione alle capacità intrinseche al modello.

Alla luce di tali premesse, l'obiettivo di questa analisi è individuare le tendenze ricorrenti nelle traduzioni effettuate dai modelli presi in esame: nello specifico, si vuole evidenziare quali tra essi propendono verso una traduzione più letterale e quali verso una traduzione che al contrario si discosta dall'originale, eventuali errori (grammaticali, sintattici, ortografici, lessicali) e il riutilizzo di elementi testuali (figure retoriche, aggettivi, strutture sintattiche, etc.). A questo proposito, la valutazione della qualità delle traduzioni tramite le metriche di *adequacy* e *fluency* a cui si è accennato a inizio capitolo viene adottata come supporto all'analisi e all'interpretazione delle differenze osservate negli output.

2.3.2. Costruzione del corpus

Per portare a termine gli obiettivi previsti da questo esperimento è stato costruito un corpus composto da dieci racconti brevi in lingua inglese appartenenti ad altrettanti generi letterari. La scelta è ricaduta sui seguenti proprio per la loro eterogeneità stilistica e lessicale: avventura, giallo, fiaba, fantasy, storico, horror, mystery, rosa, fantascienza e thriller. I testi sono stati generati utilizzando il modello Claude Sonnet 4.6 mediante prompt standardizzati per ciascun genere, in modo tale da minimizzare le variabili non controllate.

In un'ottica più specifica, il processo di generazione dei racconti è stato condotto tramite un protocollo di *prompt engineering* finalizzato a istruire il modello con indicazioni dettagliate relative all'esecuzione del compito. Per ogni racconto, pertanto, è stato somministrato un prompt strutturato nelle seguenti modalità: in primo luogo

⁹⁷ Donato, Benedetta, Leonardo Mariani, Daniela Micucci, e Oliviero Riganelli. «Studying How Configurations Impact Code Generation in LLMs: The Case of ChatGPT». Versione 1. Preprint, arXiv, 2025. <https://doi.org/10.48550/ARXIV.2502.17450>.

si è indicata una lunghezza approssimativa di circa 5000 parole, ottenendo così un corpus omogeneo e standardizzato. Successivamente si è fornita una definizione del genere letterario preso in esame che descrivesse in modo esplicito gli elementi narrativi tipici come ambientazioni, personaggi e tematiche; a questo punto si è inclusa altresì una breve trama per guidare il modello nella costruzione del racconto e mantenere una certa coerenza narrativa. Infine, la macchina riceve istruzioni sulle cifre stilistiche a cui attenersi durante la scrittura che, nel dettaglio, comprendono atmosfera e ambientazione, ritmo narrativo, punto di vista, linguaggio e finale; in aggiunta a tali caratteristiche generali ne sono state inserite altre più specifiche, questa volta inerenti alla trama di ciascun racconto.

2.3.3. Modelli usati per la traduzione

Una volta costruito il corpus iniziale è stato possibile procedere con la fase di traduzione automatica, per la quale sono stati selezionati tre diversi modelli linguistici di grandi dimensioni rappresentativi dello stato dell'arte attuale: Claude Sonnet 4.6, GPT-5.3 e Gemini 3. La loro significatività nel panorama attuale dell'intelligenza artificiale generativa è facilmente constatabile grazie alle statistiche di utilizzo raccolte da Similarweb⁹⁸: nel loro insieme, infatti, rappresentano la stragrande maggioranza delle interazioni degli utenti con modelli di IA a livello mondiale.

Considerando altri aspetti, l'inserimento di tali modelli all'interno di questo studio è ricaduto sulla volontà di mettere a confronto tre diversi approcci all'elaborazione del linguaggio, i quali a loro volta si traducono in differenze considerevoli nei comportamenti linguistici. GPT-5.3 è stato scelto innanzitutto per la sua architettura avanzata in grado di alternare in modo dinamico, secondo la complessità della richiesta, un modello ad alta velocità a uno di ragionamento profondo: una tale capacità di calibrazione dello sforzo cognitivo unita a una progressiva riduzione delle allucinazioni e a un maggior rispetto delle istruzioni

⁹⁸ <https://datareportal.com/reports/digital-2026-one-billion-people-using-ai?rq=ai>

fornite, dovrebbe garantire una traduzione fluida e semanticamente adeguata⁹⁹. Claude Sonnet 4.6, oltre a configurarsi come uno dei modelli selezionati per la traduzione, rappresenta il motore principale dell'esperimento in quanto "creatore" dei testi originali. È stato integrato nella ricerca per aver introdotto al suo interno una modalità di pensiero adattiva, che permette al modello di determinare autonomamente quanto tempo e risorse dedicate alla riflessione prima di produrre un output. Un'altra motivazione risiede nella sua capacità di gestire contesti lunghi fino a un milione di token con un'elevata precisione nel recupero delle informazioni: una tecnologia fondamentale per garantire il mantenimento di una coerenza narrativa e terminologica anche in testi particolarmente estesi¹⁰⁰. Infine, Gemini 3 completa la selezione grazie alla sua natura di modello di ragionamento multimodale e multilinguistico, in cui l'introduzione di diversi livelli di pensiero consente di bilanciare la qualità degli output con la velocità di risposta¹⁰¹.

2.3.4. Procedura di traduzione

Definiti i modelli a cui sottoporre i racconti originali per la traduzione in italiano, è risultata necessaria la costruzione di ulteriori prompt strutturati che istruissero la macchina a portare a termine il compito nel miglior modo possibile. In una prospettiva più dettagliata, per garantire l'imparzialità dell'analisi, a ciascun modello è stato sottoposto il medesimo input articolato secondo la seguente traccia: «Sei un traduttore letterario specializzato in narrativa [genere narrativo corrispondente]. Traduci il seguente racconto dall'inglese all'italiano», accompagnata da un breve riassunto della trama e delle cifre stilistiche presenti nell'originale, aggiungendo poi una serie di indicazioni sui tratti distintivi che devono essere necessariamente mantenuti nella traduzione.

⁹⁹ Singh, Aaditya, Adam Fry, Adam Perelman, et al. «OpenAI GPT-5 System Card». Versione 1. Preprint, arXiv, 2026. <https://doi.org/10.48550/ARXIV.2601.03267>.

¹⁰⁰ Anthropic. «System Card: Claude Sonnet 4.6», 2026.

¹⁰¹ Google. «Gemini 3 Flash Model Card», 2025.

La fase di traduzione automatica è stata tuttavia notevolmente condizionata dalle diverse specifiche tecniche di ciascun modello: se Claude ha consentito l'elaborazione di interi documenti in un unico procedimento, garantendo di fatto una maggiore coerenza, le interfacce di ChatGPT e Gemini hanno richiesto una gestione frammentata dei testi. Tale disparità rappresenta pertanto una variabile significativa che potrebbe avvantaggiare Claude a discapito degli altri. In tal senso, numerosi studi¹⁰² evidenziano come l'ampiezza della finestra di contesto influenzi l'efficacia della traduzione: nello specifico, la capacità massima di elaborazione non coincide necessariamente con un utilizzo efficiente del contesto ai fini della coerenza e della fedeltà traduttiva. In questi casi, infatti, è possibile osservare un drastico peggioramento delle prestazioni non appena le dimensioni del documento crescono oltre una determinata soglia¹⁰³.

All'aumentare della finestra di contesto si osservano inoltre delle problematiche in grado di compromettere l'integrità del testo tradotto, tra le quali emergono l'incremento esponenziale di omissioni di segmenti del testo sorgente (sotto-traduzione) e la tendenza alla generazione di contenuti allucinati o ridondanti (sovra-traduzione). In presenza di testi di estrema lunghezza molti modelli mostrano difficoltà nel completare il processo traduttivo, iniziando a produrre sintesi del documento originale invece di una trasposizione completa; ulteriori fenomeni osservati concernono problematiche di incoerenza e ambiguità sul piano lessicale e sintattico¹⁰⁴.

Alla luce di tali considerazioni, la fase di traduzione ha previsto quindi l'adozione di un procedimento identico per tutti i modelli, somministrando in input

¹⁰² Cui, Menglong, Jiangcun Du, Shaolin Zhu, e Deyi Xiong. «Efficiently Exploring Large Language Models for Document-Level Machine Translation with In-Context Learning». In Findings of the Association for Computational Linguistics ACL 2024. Findings of the Association for Computational Linguistics ACL 2024. Association for Computational Linguistics, 2024. <https://doi.org/10.18653/v1/2024.findings-acl.646>.

¹⁰³ Wang, Longyue, Zefeng Du, Wenxiang Jiao, et al. «Benchmarking and Improving Long-Text Translation with Large Language Models». In Findings of the Association for Computational Linguistics ACL 2024. Findings of the Association for Computational Linguistics ACL 2024. Association for Computational Linguistics, 2024. <https://doi.org/10.18653/v1/2024.findings-acl.428>, 7175.

¹⁰⁴ Wang, Kuang-Da, Shuoyang Ding, Chao-Han Huck Yang, et al. «Extending Automatic Machine Translation Evaluation to Book-Length Documents». Versione 1. Preprint, arXiv, 2025. <https://doi.org/10.48550/ARXIV.2509.17249>.

frammenti del testo originale di un'estensione compresa tra le 1200 e le 1500 parole. Conclusa questa fase, il risultato finale è un corpus composto da un totale di 40 testi.

2.3.5. Limiti epistemologici

La scelta di utilizzare racconti brevi generati artificialmente come base del corpus è scaturita, in primo luogo, da esigenze pratiche e metodologiche. Se da un lato infatti ha consentito di evitare problematiche legate al diritto d'autore, permettendo quindi una piena libertà di utilizzo dei testi, dall'altro ha ridotto il rischio che i modelli traducessero in funzione di conoscenze pregresse presenti all'interno dei dati di addestramento.

Tuttavia tale scelta comporta altresì alcune limitazioni: nel dettaglio, i testi sintetici possiedono qualità di scrittura e cifre stilistiche indubbiamente carenti e meno stratificate rispetto a testi letterari autentici, orientando in modo significativo i risultati ottenuti con le traduzioni. In quest'ottica eventuali errori, semplificazioni o ambiguità presenti nei testi di partenza tendono a riflettersi sistematicamente nelle traduzioni; pertanto, è possibile che la maggior parte delle tendenze stilistiche osservate negli output non siano attribuibili unicamente al comportamento dei modelli, ma anche alle caratteristiche intrinseche del corpus su cui si fonda l'analisi.

Un ulteriore limite è rappresentato dalla ben nota natura non deterministica dei modelli linguistici di grandi dimensioni, i quali operano esclusivamente secondo principi probabilistici: ciò implica che, a parità di prompt, lo stesso modello può produrre output diversi. A questo proposito, le traduzioni alla base del corpus di questa ricerca si configurano come il primo risultato ottenuto mediante l'input descritto in precedenza, rappresentando perciò una sola possibile realizzazione delle molteplici alternative generabili. Concludendo, è necessario chiarire che tale caratteristica impedisce di considerare i risultati ottenuti pienamente rappresentativi del comportamento del modello.

2.4. Analisi dei risultati

Nel complesso, alla luce dell'analisi condotta sulle 30 traduzioni, si è potuta evidenziare la presenza di tendenze ricorrenti e relativamente stabili nell'intero corpus, consentendo così di delineare dei profili comportamentali specifici per ciascun modello.

Innanzitutto, Claude si è contraddistinto per un approccio traduttivo orientato a una marcata aderenza al testo fonte: le sue traduzioni, infatti, sono risultate eccessivamente letterali, presentando una rielaborazione pressoché assente delle strutture sintattiche e una tendenza a mantenere invariata la punteggiatura originale. Sebbene tale rigidità garantisca un'elevata fedeltà formale, può tuttavia generare in alcuni casi un effetto di straniamento nella lingua di arrivo.

ChatGPT, per contro, manifesta una propensione alla riformulazione dando ampia priorità alla scorrevolezza del testo in italiano. Ciò è da identificarsi in un adattamento talvolta più pronunciato della sintassi alla lingua di arrivo e, pertanto, nella riduzione di quelle che De Cesare definisce come «impronte algoritmiche»¹⁰⁵ dell'inglese. Inoltre si riscontra l'emergere di alcuni fenomeni di allucinazione a livello morfologico che vanno a incidere sulla qualità complessiva della traduzione.

Infine, Gemini si colloca in una posizione intermedia rispetto ai due modelli precedentemente analizzati: pur manifestando una maggiore flessibilità e chiarezza nella trasposizione semantica insieme a una selezione lessicale più consona, dall'altro lato si nota una certa debolezza nella coerenza testuale, riconducibile a incongruenze nella *consecutio temporum* e alla permanenza di approcci eccessivamente letterali in determinati passaggi.

¹⁰⁵ De Cesare, *L'italiano sintetico dell'intelligenza artificiale generativa*, op. cit., 89.

2.4.1. Claude

Prima di procedere con l'analisi effettiva vale la pena ricordare che Claude, oltre ad aver prodotto le traduzioni in italiano, ha altresì generato i racconti originali; di conseguenza, è possibile affermare che le cifre stilistiche presenti nelle traduzioni sono riconducibili a tale modello.

Si prenda ora in esame il principale tratto distintivo di Claude: l'eccessiva letteralità. A tal proposito occorre evidenziare che è stato possibile osservare questo fenomeno attraverso l'intero corpus, indipendentemente dal genere letterario e dallo stile narrativo indicato nel prompt. Ad esempio:

She was twenty-three years old, freshly credentialed from the University of Barcelona with a master's thesis on pre-colonial maritime cultures of the Indonesian archipelago. The ink on her diploma was barely dry. The salt spray on her jacket was not yet a week old. [...] lay the ruins of something that had no name in any language still spoken by living people.

Aveva ventitré anni, laurea fresca dell'Università di Barcellona con una tesi magistrale sulle culture marittime precoloniali dell'arcipelago indonesiano. L'inchiostro sul diploma era appena asciutto. Il sale sullo giubbotto non aveva ancora una settimana. [...] giacevano le rovine di qualcosa che non aveva nome in nessuna lingua ancora parlata da esseri viventi.

Nel passaggio, tratto dal racconto appartenente al genere avventura, si notano diversi fenomeni: innanzitutto spicca l'utilizzo errato della preposizione articolata “sullo” e la conseguente omissione del pronome possessivo “her”. Nell'ultima frase del paragrafo, invece, è ben visibile la totale aderenza della traduzione alla sintassi inglese, risultando in una frase poco scorrevole.

Un altro esempio di eccessiva letteralità, questa volta nel racconto giallo:

The fields that surrounded the town were at their most beautiful in October, when the harvests were in and the frost had not yet killed the last of the year's colour.

I campi che circondavano la città erano al loro più belli in ottobre, quando i raccolti erano stati riposti e il gelo non aveva ancora ucciso gli ultimi colori dell'anno.

In una prospettiva prettamente stilistica, si osserva la ricorrenza degli stessi termini: selezionando diverse traduzioni prodotte da Claude, infatti, si evidenzia un uso sproporzionato dell'aggettivo “deliberato”, così come dell'avverbio “deliberatamente”. Questa tendenza, tuttavia, non si circoscrive al livello lessicale, ma si estende anche sul piano sintagmatico con la presenza del sintagma “modo in cui” e di sue variazioni, quali “nel modo particolare”, “in qualche modo” “nel modo di”.

Qui di seguito alcuni esempi tratti da molteplici racconti:

Un'ombra che si muoveva da sola, lenta e deliberata.

Il drago virò sulla valle senza rallentare, la grande testa che descriveva un arco che era in qualche modo preciso, in qualche modo deliberato [...].

I testi antichi [...] descrivono un patto deliberato.

[...] la compostezza di Ferris si incrinò nel modo particolare di un uomo [...].

[...] e Sara si avvicinò nel modo misurato e deliberato che le era stato insegnato [...].

Sapeva che Elspeth aveva scelto, deliberatamente, di non scrivere certe cose. [...] in qualche modo era quello di cui c'era bisogno.

Nei testi prodotti da questo modello, inoltre, si è osservata la tendenza al riutilizzo di metafore specifiche, come il peso del non detto:

Lasciò che il silenzio portasse il peso di ciò che non stava del tutto dicendo.

Sentì il peso di ciò che non vi era detto [...].

[...] ovvero che il silenzio che avevo mantenuto per cinque anni aveva un peso proprio [...].

Per di più, si riscontra in due diverse occasioni una metafora che accosta suoni intensi al rintocco di una campana:

[..] risuonò un suono — basso e risonante, come una campana colpita una volta sola in una stanza vuota.

[...] un tono grave e prolungato come una campana percossa sottoterra [...].

In seguito, per quanto concerne la tematica delle allucinazioni, è possibile osservarne una manifestazione, più precisamente, nel racconto horror con la traduzione del verbo “*strobed*” in “strobava”.

She photographed each region before touching it, and the camera’s flash strobed white across the steel surfaces of the room in a rhythm she found grounding.

Prima di toccare ogni area, la fotografava, e il flash della fotocamera strobava bianco sulle superfici d'acciaio della sala in un ritmo che trovava rassicurante.

Un’ulteriore allucinazione la si può rilevare nel racconto giallo, benché in questo caso tale sia riconducibile direttamente all’originale. Nel dettaglio, viene utilizzato il termine “*constable*” per indicare la figura dell’agente di polizia: il problema

sorge a livello semantico, poiché sebbene tale parola appartenga in modo esclusivo al mondo britannico, il racconto in questione sembra essere ambientato in una Germania immaginaria. La versione italiana traduce altrettanto erroneamente selezionando “brigadiere”.

Allo stesso tempo, all'interno del racconto fantasy si nota il termine “subsonico” che, nel contesto in cui la narrazione è ambientata, appare inadeguato.

2.4.2. ChatGPT

Una volta condotta l'analisi, si è potuto constatare che ChatGPT propende per una scrittura più scorrevole rispetto al modello precedentemente valutato. In tal senso, si rende necessario effettuare un confronto non solo con il testo originale, ma anche con le traduzioni prodotte da Claude. Si prenda ad esempio il seguente passaggio:

It was closed again. Latched, dark behind its glass. Exactly as it should be. She looked at it for a moment, and she had the strong, entirely irrational conviction that on the other side of that door, in the cold and the dark and the smell of old paper, something was aware of her standing there.

Era di nuovo chiusa. Agganciata, buia dietro il vetro. Esattamente come avrebbe dovuto essere. La guardò per un momento, e aveva la ferma, del tutto irrazionale convinzione che dall'altra parte di quella porta, nel freddo e nel buio e nell'odore di carta vecchia, qualcosa fosse consapevole di lei che stava lì. (Claude)

Era di nuovo chiusa. Serrata, buia dietro il vetro. Esattamente come doveva essere. La guardò per un momento, e provò la convinzione forte, del tutto irrazionale, che dall'altra parte di quella porta, nel freddo e nel buio e nell'odore di carta vecchia, qualcosa fosse consapevole della sua presenza. (ChatGPT)

Dal confronto tra queste versioni si evidenziano chiaramente due strategie traduttive distinte. Innanzitutto l'uso del termine “serrata”, poiché rappresenta la scelta lessicale più adeguata in lingua italiana per denotare una porta chiusa; in secondo luogo, la riformulazione della frase finale elimina l'impronta algoritmica della sintassi inglese.

Un'ulteriore dimostrazione della maggiore fluidità nella narrazione si può osservare in questo brano tratto dal racconto di fantascienza:

I do not experience the walk to the Portside District processing terminal as beautiful. Beauty is not a category that applies to me—or so I have understood, until recently. The canal reflects the sky in silver and grey pieces. I observe this. I log it. And then something happens that I cannot fully account for in my standard processing log: I continue to observe it, past the point of utility. Past the moment when the observation serves any purpose.

Non sperimento il tragitto verso il terminal di processing del Distretto Portuale come qualcosa di bello. Bellezza non è una categoria che mi si applica – o così ho compreso, fino a poco tempo fa. Il canale riflette il cielo a frammenti d'argento e grigio. Lo osservo. Lo registro. E poi accade qualcosa che non riesco a rendicontare pienamente nel mio log di processing standard: continuo a osservarlo, oltre il punto di utilità. Oltre il momento in cui l'osservazione serve a qualcosa. (Claude)

Non percepisco la passeggiata verso il terminal di elaborazione del distretto Portside come bella. La bellezza non è una categoria che mi riguarda—o così avevo compreso, fino a poco tempo fa. Il canale riflette il cielo in frammenti d'argento e grigio. Osservo. Registro. E poi accade qualcosa che non riesco a spiegare nei miei log standard: continuo a osservare oltre il punto di utilità. Oltre il momento in cui l'osservazione ha uno scopo. (ChatGPT)

La preferenza del secondo modello per l'adattamento in italiano del lessico tecnico informatico utilizzato nell'originale conferisce indubbiamente maggior

scorrevolezza e permette un'immersione più profonda nel contesto narrativo per il lettore italiano. Nell'insieme di questo estratto, infine, si rileva la tendenza ad alleggerire lo stile di scrittura evitando la ripetizione dei pronomi ridondanti. Analizzando ulteriormente la traduzione di tale racconto risulta necessario mettere in rilievo alcune criticità del modello relative alla coerenza traduttiva: se in una prima fase opta per il mantenimento di determinate denominazioni in lingua originale, successivamente procede alla loro traduzione in italiano.

Ed è così che finiva che il *Coalition of Machine Rights* si accampava fuori dal Parlamento.

La *Coalizione per i Diritti delle Macchine* l'aveva organizzata [...].

Inoltre, questo fenomeno interessa persino la traduzione di una stessa parola:

I had responded: Clarify *query*.

[...] And I have been turning that question over in whatever part of my architecture handles persistent, unresolved *queries* for eleven days since Renn was taken for emergency recalibration.

Io risposi: Chiarire la *domanda*.

[...] E da undici giorni, da quando Renn fu prelevato per ricalibrazione d'emergenza, sto elaborando quella domanda nella parte della mia architettura che gestisce *query* persistenti e irrisolte.

Un'incoerenza di tale portata nella traduzione è da identificarsi come un grave errore, poiché potrebbe minare la comprensibilità del testo da parte del lettore.

Nel quadro relativo al piano morfologico si evidenzia la presenza di tre allucinazioni, di cui una definibile come «neologia algoritmica»¹⁰⁶. Prendendo in esame quest'ultima casistica, analogamente a quanto osservato in precedenza nel racconto

¹⁰⁶ De Cesare, *L'italiano sintetico dell'intelligenza artificiale generativa*, op. cit., 59.

horror, il verbo “*strobed*” viene tradotto in maniera alquanto originale con il termine “stroboscopava”; per ciò che concerne invece le restanti allucinazioni, esse si configurano come errori morfologici nella coniugazione al passato remoto dei verbi derivati da “mettere”, come nei casi di “trasmettei” e di “ammetté”. Vale la pena sottolineare che tali errori si sono riscontrati esclusivamente nell’analisi dei testi prodotti da ChatGPT.

2.4.3. Gemini

L’ultimo modello utilizzato per la conduzione di questo esperimento occupa una posizione intermedia rispetto ai precedenti, mostrando da un lato maggiore fluidità e chiarezza di significato e dall’altro incongruenze riscontrabili nella concordanza dei tempi verbali, così come la persistenza di un’eccessiva letteralità in determinati passaggi.

In un’ottica più dettagliata, la cifra stilistica di Gemini si configura nel tentativo di selezionare la terminologia più specifica possibile:

He visited the mill in the morning and spent two hours there, speaking first with Voss's foreman, a reliable and uncomplicated man named Dressler, and then alone, in the storage rooms and along the *mill-race*, with no company but his own.

Visitò il mulino al mattino e vi trascorse due ore, parlando prima con il caposquadra di Voss, un uomo affidabile e semplice di nome Dressler, e poi da solo, nei magazzini e lungo la *gora del mulino*, senza altra compagnia che la propria.

Se i modelli precedenti hanno entrambi tradotto il termine in questione con un generico “canale”, Gemini in questo caso sceglie di utilizzare un lessico più specifico: la definizione di “gora” secondo Treccani, infatti, corrisponde al «canale che porta

l'acqua a un mulino»¹⁰⁷. In modo analogo, ulteriori esempi sono rappresentati in primo luogo dalla trasposizione di “*steward*” con “siniscalco”¹⁰⁸ che, trovandosi all'interno di un contesto narrativo medievale, risulta particolarmente adatto; in secondo luogo, la scelta dei termini tecnici “sala autoptica” e “tavolo settorio” nel racconto horror, che potenzia la verosimiglianza e l'immersione del lettore nella narrazione.

Tale tendenza, tuttavia, non sempre conduce ai risultati sperati, inducendo il modello a incorrere nel medesimo errore commesso da ChatGPT nella traduzione di “*constable*” che, questa volta, forse considerata la somiglianza con il termine inglese, viene tradotto con “conestabile”: una scelta lessicale totalmente fuori luogo, in quanto secondo Treccani in origine faceva riferimento a un «titolo di ufficiali di corte bizantini» e attualmente al capo della polizia in alcune città americane¹⁰⁹.

Mantenendo l'attenzione sul piano lessicale, degna di nota è la traduzione del seguente brano:

He was not important [...] but he was alive and he was talking, in the *piecemeal* way of a man who calculates exactly how much information is worth his continued good treatment.

Non era importante [...] ma era vivo e parlava, a *spizzichi e bocconi*, come un uomo che calcola esattamente quanta informazione valga il proseguimento del buon trattamento.

Come già evidenziato, gli output di Gemini manifestano nel complesso una maggiore chiarezza a livello di significato, laddove le traduzioni di Claude rendono la lettura difficoltosa, ostacolando la comprensione del senso generale del racconto. Ad esempio:

¹⁰⁷ <https://www.treccani.it/vocabolario/gora/>

¹⁰⁸ <https://www.treccani.it/vocabolario/siniscalco/>

¹⁰⁹ <https://www.treccani.it/vocabolario/conestabile/>

Maya almost walked past them — the vegetation had been so thorough in its reclamation that the stones looked like natural formations at first, just black basalt humps in the undergrowth. But she had spent four years studying the construction techniques of pre-colonial civilizations in this region, and her hands recognized what her eyes were slow to process. She knelt and brushed moss from a surface and found a join between two stones too precise to be accident. Cut. Fitted. Old.

Maya quasi le sorpassò — la vegetazione era stata così sistematica nella sua riconquista che le pietre sembravano formazioni naturali, a prima vista: solo gobbe di basalto nero nel sottobosco. Ma aveva trascorso quattro anni a studiare le tecniche costruttive delle civiltà precoloniali della regione, e le sue mani riconobbero ciò che i suoi occhi erano lenti a elaborare. Si inginocchiò, spazzò via il muschio da una superficie e trovò la giuntura tra due pietre troppo precisa per essere casuale. Tagliate. Incastrate. Antiche. (Claude)

Maya rischiò quasi di oltrepassarle: la vegetazione era stata così meticolosa nel riprendersi i propri spazi che le pietre sembravano formazioni naturali, semplici gobbe di basalto nero nel sottobosco. Ma lei aveva passato quattro anni a studiare le tecniche costruttive delle civiltà pre-coloniali della regione, e le sue mani riconobbero ciò che gli occhi tardavano a elaborare. Si inginocchiò, scostò il muschio da una superficie e trovò un incastro tra due pietre troppo preciso per essere casuale. Tagliato. Rifinito. Antico. (Gemini)

L'ultimo brano risulta indubbiamente più comprensibile, anche grazie alle modifiche apportate alla punteggiatura, in quanto il cosiddetto *em-dash* non è così comune nella scrittura italiana. Ulteriore elemento degno di nota è la trasposizione di “*reclamation*” con l'espressione “riprendersi i propri spazi”: tale scelta evidenzia come il superamento della traduzione letterale possa, in alcuni contesti, preservare il significato originale.

Passando a un diverso livello di analisi, risulta necessario mostrare alcuni passaggi caratterizzati da errori dovuti a un'eccessiva letteralità, poiché tale tendenza persiste in tutti e tre i modelli. Ad esempio:

"What are you?" she asked it, quietly, *the way she had been asking since she was seven years old.*

«Cosa sei?», gli chiese, a bassa voce, *come glielo stava chiedendo da quando aveva sette anni.*

Three times the seasons *turned* [...].

Per tre volte le stagioni *girarono* [...].

Edric said nothing. *He sat with* the information and felt its weight, which was considerable, and said nothing.

Edric non disse nulla. *Rimase seduto con* quell'informazione e ne sentì il peso, che era considerevole, e non disse nulla.

Nei brani presi in esame emerge chiaramente l'impronta algoritmica della sintassi inglese che, in assenza di un opportuno adattamento, produce strutture totalmente innaturali nella lingua di arrivo.

Infine, è possibile evidenziare errori a livello della *consecutio temporum*, soprattutto nella coniugazione dei congiuntivi ai tempi passati:

His housekeeper, a stolid woman named Bertha Kroll who had served in that house for eleven years and who possessed the incurious reliability of someone who has decided, on some early morning long ago, that the world beyond her particular duties is not her concern, had come at six o'clock as always.

La sua governante, una donna solida di nome Bertha Kroll che aveva servito in quella casa per undici anni e che possedeva l'incuriosa affidabilità di chi ha deciso, in una mattina remota di molto tempo fa, che il mondo oltre i propri doveri particolari non la *riguardi*, era arrivata alle sei come sempre.

Oltre all'imprecisione riscontrabile nella traduzione dell'aggettivo "*incurious*", il brano presenta un'incongruenza nella selezione del modo e del tempo verbale di "riguardi": una corretta gestione della sintassi, infatti, avrebbe richiesto il ricorso al presente dell'indicativo o in alternativa al congiuntivo imperfetto.

The ember that guttered and flared and refused to go out no matter how many winters she endured [...].

La brace che s'affievoliva e divampava, rifiutandosi di spegnersi, incurante di quanti inverni lei *sopportasse* [...].

Nel presente passaggio la gestione incorretta della concordanza dei tempi si riscontra nella coniugazione del verbo sopportare, che in questo caso richiederebbe il ricorso al congiuntivo trapassato, poiché mantenere il congiuntivo imperfetto rischia di compromettere la resa di un'anteriorità rispetto alla proposizione principale.

Capitolo 3 – Le voci in capitolo: le testimonianze di due professionisti

Finora sono state illustrate le principali prospettive teoriche presenti nella letteratura accademica concernenti l'impiego di strumenti di traduzione automatica e di intelligenza artificiale generativa nell'ambito della traduzione letteraria, facendo particolare attenzione alle tendenze stilistiche di tali strumenti e alle metodologie sviluppate per valutarne gli output. Dopodiché, a completamento del quadro teorico, è stato possibile eseguire uno studio empirico condotto mediante un esperimento di traduzione, volto principalmente a rilevare l'effettiva manifestazione dei fenomeni stilistici descritti in precedenza. Poi, una volta eseguita una disamina dettagliata delle traduzioni generate dai modelli di intelligenza artificiale e averne delineato le cifre stilistiche ricorrenti, si è reso necessario integrare tali risultati con l'esperienza e le opinioni di professionisti del settore.

Il presente capitolo, perciò, si propone di analizzare le interviste a due traduttori professionisti, con l'obiettivo di approfondire le tematiche sollevate dalla domanda di ricerca di questa tesi, ovvero «qual è l'impatto percepito dell'utilizzo dei modelli linguistici come strumenti di supporto nella traduzione di narrativa letteraria secondo traduttori professionisti e quali implicazioni ha sulla qualità delle traduzioni e sul ruolo del traduttore?».

3.1. Nota metodologica

Per esplorare in profondità le dinamiche emerse dall'analisi comparativa delle traduzioni, la scelta si è orientata verso un approccio di ricerca qualitativa basato su interviste semi-strutturate. In particolare, le conversazioni hanno seguito una traccia comune di otto domande che ha permesso di guidare gli intervistati lungo un percorso specifico, aprendosi in primo luogo con la presentazione del proprio percorso professionale, attraverso poi una valutazione critica degli strumenti di traduzione automatica e di intelligenza artificiale generativa e, infine, verso una riflessione sugli scenari futuri della professione.

3.1.1. Gli intervistati

Prima di prendere approfonditamente in esame le questioni toccate durante le interviste, è opportuno presentare i due professionisti che hanno accettato di partecipare a questa fase della ricerca.

Sara Bilotti inizia il suo percorso da traduttrice nel 2018 su richiesta dell'editor di una casa editrice indipendente, individuando in lei un profilo particolarmente adatto al progetto di traduzione di un romanzo americano che racconta una storia di abuso familiare. La sua formazione accademica all'estero e l'esperienza ventennale a contatto con ragazzi vittime di simili abusi sono infatti risultate essenziali per l'interpretazione e una resa fedele del romanzo. Tale esperienza è stata determinante sul piano professionale, poiché l'ha spinta a dare priorità all'attività di traduzione rispetto al suo ruolo di scrittrice. Da allora, ha tradotto vari autori tra cui Joe R. Lansdale, Joyce Carol Oates, Quentin Tarantino e diversi altri.

Luca Briasco intraprende il suo percorso da traduttore contemporaneamente a quello editoriale alla fine degli anni Novanta, traducendo per le case editrici per cui lavora. Traduce esclusivamente dall'inglese, ovvero l'unica lingua in cui possiede la padronanza necessaria per raggiungere i livelli qualitativi a cui aspira, pur conoscendo altre lingue come francese e spagnolo; ad oggi conta un catalogo di circa un centinaio di titoli tradotti che, eccetto qualche opera di saggistica, è composto interamente da narrativa di autori classici e contemporanei, tra i quali si trovano Coldwell, Ballard e Stephen King.

3.2. Le interviste

Come delineato in precedenza, la traccia comune a entrambe le interviste è articolata su otto domande, le quali hanno consentito di indagare nel dettaglio l'impatto percepito dell'impiego dell'intelligenza artificiale nell'ambito di studio di questa tesi e, pertanto, se le «impronte algoritmiche» di tali strumenti trovino riscontro nella pratica quotidiana dei traduttori, così come le loro implicazioni etiche. In tal senso, essa accompagnerà quest'analisi passo dopo passo: per ogni tematica

sottostante a ogni domanda, infatti, si confronteranno le risposte degli intervistati a molteplici articoli presenti in letteratura accademica, in modo tale da poter ottenere una visione d'insieme che includa altresì il panorama professionale italiano, il quale risulta estremamente sotto rappresentato nelle pubblicazioni di settore.

3.2.1. Il rapporto con gli strumenti di traduzione automatica o di IA

Storicamente è possibile ricondurre la totalità del percorso professionale del traduttore al lavoro di traduzione umana “da zero” (*human translation from scratch*), il quale sta tuttavia attraversando una fase di profonda ridefinizione a causa della progressiva integrazione nei flussi di lavoro di strumenti di traduzione automatica e di intelligenza artificiale generativa. A dimostrazione di ciò, se un tempo la professione era caratterizzata da una netta distinzione tra la creazione testuale autonoma e il mero supporto terminologico, ad oggi questa linea di demarcazione è sempre più sfumata: dall'ELIS Research del 2025 risulta infatti che circa il 50% dei traduttori che hanno partecipato alla ricerca (sia dipendenti che indipendenti) dichiara di utilizzare tali strumenti nei propri processi di lavoro¹¹⁰. Nel corso dell'analisi verrà illustrato inoltre come il rapporto con gli strumenti alla base di questa ricerca tenda a diversificarsi in funzione degli anni di esperienza maturati durante il proprio percorso professionale.

In un'ottica più dettagliata, il concetto di «multistabilità» risulta particolarmente adatto a descrivere il rapporto attuale che i traduttori professionisti hanno con questi strumenti: infatti, come riportato da Moorkens nel saggio «The machine translator's visibility», tale linea di pensiero evidenzia il fenomeno per cui una tecnologia può essere contemporaneamente qualcosa che creiamo e usiamo per i nostri scopi, ma anche qualcosa che può guidarci o addirittura influenzarci e controllarci¹¹¹. Se si applica questo ragionamento alla traduzione automatica è possibile osservare che il ricorso a tale tecnologia avviene in funzione di due variabili fondamentali, ossia le competenze dell'utente, che in questo caso si configura nel traduttore professionista,

¹¹⁰ Moorkens, Joss. «The Machine Translator's Visibility». *Translation Spaces*, ahead of print, 6 ottobre 2025. <https://doi.org/10.1075/ts.23030.moo>, 3.

¹¹¹ Ivi, 5.

e la natura del testo. Con riferimento alla prima variabile menzionata, occorre evidenziare come numerosi professionisti in ambito letterario dichiarino di avvalersi di strumenti di traduzione automatica o IA generativa esclusivamente per ricevere un supporto alla creatività, funzionale al superamento dei blocchi linguistici, così come per ottenere un punto di partenza da cui sviluppare la propria interpretazione o agevolare la ricerca terminologica¹¹²; l'ultima variabile, invece, sarà oggetto di approfondimenti in una fase successiva di quest'analisi.

Tali affermazioni trovano altresì riscontro nei responsi degli intervistati alla prima domanda, il cui scopo è di sondare pertanto il loro rapporto con questi strumenti: entrambi i rispondenti dichiarano di non utilizzare l'intelligenza artificiale nel processo effettivo di traduzione ma, proprio per il fatto di essere a conoscenza delle loro dinamiche, di relegare il loro impiego a contesti didattici e di ricerca. Nello specifico, nella prima intervista si fa riferimento alla figura di docente di traduzione di Luca Briasco per la scuola Minimum Lab e, di conseguenza, alla volontà di «ragionare con i miei studenti sui possibili sviluppi futuri della traduzione per effetto dell'intelligenza artificiale»; nel secondo caso, invece, Sara Bilotti afferma di poter ottenere un considerevole risparmio di tempo nelle fasi di ricerca terminologica che accompagnano il lavoro di interpretazione del testo, chiedendo all'intelligenza artificiale di «trovarmi contesti in cui possa leggere i vari usi dello slang in questione - forum, siti, discussioni - , in modo da confrontarli e comprenderne a fondo il significato».

3.2.2. Riconoscere gli elementi caratterizzanti e i rischi delle traduzioni automatiche

L'identificazione di una traduzione generata o anche solo revisionata con il supporto dell'intelligenza artificiale non risiede unicamente in una sensazione di generica innaturalità, poiché risulta necessario esaminare precise caratteristiche

¹¹² Ivi, 4.

linguistiche e strutturali che la letteratura accademica definisce come *post-edite*¹¹³ o *machine translationese*¹¹⁴. Questi ultimi sono una diretta ed esacerbata estensione del fenomeno conosciuto più semplicemente come *translationese* (in italiano traduttorese), che si configura come un esito traduttivo di scarsa qualità in cui è possibile rilevare un testo sensibilmente semplificato e contraddistinto da un elevato grado di interferenza con la lingua di partenza¹¹⁵.

Riprendendo i concetti esposti nel capitolo precedente, uno dei primi campanelli d'allarme che preannunciano a occhi esperti di essere effettivamente in presenza di una traduzione automatica è l'impovertimento lessicale e morfologico: i modelli linguistici tendono infatti a manifestare in maniera ciclica i pattern linguistici più frequenti nei dati di addestramento¹¹⁶. In tal senso, è possibile evidenziare una perdita sistematica dei termini meno comuni o, addirittura, di eventuali soluzioni creative messe in atto dal traduttore per ricalcare caratteristiche peculiari del testo originale¹¹⁷. Ciò trova riscontro nelle parole del primo intervistato, che fa proprio riferimento a questa casistica:

[...] ci sono alcuni elementi che si notano che, per esempio, sono legati alla difficoltà che molti programmi di intelligenza artificiale hanno di lavorare sulla ricorsività di certe forme espressive o perfino di certi termini o di certi giri di frase.

Pertanto tale fenomeno si concretizza in una minore varietà e densità lessicale, restituendo un output qualitativamente più povero e banale da un punto di vista semantico. A livello sintattico, per contro, le traduzioni effettuate con il supporto

¹¹³ Toral, Antonio. «Post-Edite: An Exacerbated Translationese». Versione 2. Preprint, arXiv, 2019. <https://doi.org/10.48550/ARXIV.1907.00900>.

¹¹⁴ Vanmassenhove, Eva, Dimitar Shterionov, e Matthew Gwilliam. «Machine Translationese: Effects of Algorithmic Bias on Linguistic Complexity in Machine Translation». Versione 1. Preprint, arXiv, 2021. <https://doi.org/10.48550/ARXIV.2102.00287>.

¹¹⁵ Toral, «Post-Edite: An Exacerbated Translationese» op. cit.

¹¹⁶ Ploeger, Esther, Huiyuan Lai, Rik van Noord, e Antonio Toral. «Towards Tailored Recovery of Lexical Diversity in Literary Machine Translation». Versione 1. Preprint, arXiv, 2024. <https://doi.org/10.48550/ARXIV.2408.17308>.

¹¹⁷ Toral, «Post-Edite: An Exacerbated Translationese» op. cit.

dell'IA mostrano un'aderenza quasi pedissequa alla struttura del testo originale, determinando di conseguenza una propensione alla normalizzazione della lunghezza delle frasi. A questo proposito, come già dimostrato nell'esperimento condotto all'interno di questa tesi, si è osservato che la lunghezza delle frasi nelle traduzioni automatiche revisionate tende anch'essa a ricalcare quella del testo originale ben più di quanto si riscontri nelle traduzioni umane: ciò è indubbiamente causato da una mancanza di autonomia interpretativa della macchina nel processo traduttivo¹¹⁸. L'eccessiva letteralità, pertanto, si configura come un'ulteriore segnale per il riconoscimento delle traduzioni automatiche, a cui è possibile annettere una scarsa coesione semantica e una certa confusione nella trasposizione di espressioni idiomatiche o ambiguità intrinseche dell'originale¹¹⁹. Infine, l'intenzionalità comunicativa rappresenta anch'essa un elemento imprescindibile quando si affrontano traduzioni in ambito letterario, eppure l'intelligenza artificiale ne è sprovvista: le difficoltà riscontrate nel gestire il linguaggio figurato, infatti, rendono il prodotto finale privo di quel senso profondo e di quell'effetto estetico che si considerano come il fine ultimo della letteratura¹²⁰.

Relativamente a quanto appena esposto, è possibile avvalorare tali fenomeni mediante le parole di entrambi gli intervistati, che confermano la totale incapacità dell'intelligenza artificiale di comprendere l'impianto simbolico di un'opera letteraria e il «suono della voce di un autore». Tuttavia, rispondendo a questa seconda domanda, si rileva un disaccordo in relazione alla possibilità da parte della macchina di commettere errori di senso: se il primo intervistato afferma che «un programma di intelligenza artificiale è difficile che commetta quel tipo di errore» (pur non escludendo completamente l'eventualità), nell'intervista successiva Sara Bilotti

¹¹⁸ Idem

¹¹⁹ Joke Daems, Paola Ruffo, and Lieve Macken. 2024. «Impact of translation workflows with and without MT on textual characteristics in literary translation». In *Proceedings of the 1st Workshop on Creative-text Translation and Technology*, pages 57–64, Sheffield, United Kingdom. European Association for Machine Translation, 58.

¹²⁰ Cook, Angela, Angie Knaggs, e Jiashun Sun. «Exploring the Goals of Translation Education: Adding Human Value to Literary Translation through Deep Meaning Making and Artistry». *The Interpreter and Translator Trainer* 19, fasc. 3–4 (2025): 254–76. <https://doi.org/10.1080/1750399x.2025.2543208>, 260.

racconta di una traduzione «in cui invece di tradurre “cittadina in fiamme” c’era scritto “villaggio infiammato”».

I fenomeni appena esposti implicano inevitabilmente il manifestarsi di rischi concreti nel panorama delle traduzioni letterarie, tra i quali risalta una progressiva omogeneizzazione dei testi: la progressiva semplificazione lessicale e semantica, così come l’interferenza sintattica con la lingua dell’opera originale espongono, nel lungo periodo, a un pericolo di impoverimento della lingua d’arrivo¹²¹. Tali rischi sono ben noti ai professionisti del settore, tanto da dichiarare che «rischiamo di ritrovarci con romanzi tutti uguali. Le storie sono già state tutte scritte, ma il modo in cui le raccontiamo continua ancora a fare la differenza»; mentre Luca Briasco mette l’accento sulla possibilità di una penalizzazione dello stile dell’autore. Tuttavia, se a un primo sguardo questi potrebbero essere gli unici rischi degni di nota osservabili nelle traduzioni automatiche, in realtà la prima intervista permette di scendere a un livello di analisi più profondo: Briasco evidenzia infatti l’eventualità di affidare a tali strumenti una certa tipologia di letteratura nella quale «il concetto stesso di autore è un concetto abbastanza labile», citando a titolo esemplificativo il caso degli Harmony. In un tale contesto di autorialità relativa, in cui l’aspetto di maggior rilievo è rappresentato dal tipo di trama, i danni potrebbero essere limitati. Lo stesso intervistato, però, introduce una distinzione netta per quanto riguarda la letteratura classica poiché, trattandosi di opere ormai non più soggette a diritto d’autore, «la traduzione diventa l’investimento primario»; in tal caso l’utilizzo dell’intelligenza artificiale risulterebbe pertanto incompatibile con le esigenze qualitative del settore.

3.2.3. L’effetto ancoraggio nel post-editing

L’introduzione di strumenti di traduzione automatica in ambito letterario ha contribuito alla ridefinizione dell’identità professionale del traduttore, la quale si è progressivamente ridotta a un ruolo di revisore e a cui si aggiunge un importante condizionamento stilistico durante il processo traduttivo, conosciuto come effetto

¹²¹ Toral, «Post-Editese: An Exacerbated Translationese» op. cit.

ancoraggio: un bias cognitivo che si manifesta quando ci si affida eccessivamente alla prima informazione ricevuta prima di prendere una decisione. In questo scenario, pertanto, il traduttore subisce un'influenza inevitabile dall'output della macchina, che agisce come un'ancora mentale da cui difficilmente è possibile distaccarsi. Sotto questo profilo, numerose analisi confermano la tendenza del post-editing a produrre traduzioni maggiormente letterali ed evidenziano la riluttanza spesso provata dai revisori nell'apportare modifiche sostanziali a quanto proposto dalla macchina¹²². L'esperienza nel campo editoriale del primo intervistato interviene a supporto di tale visione, sottolineando come la traduzione automatica tenda a costituire in modo inevitabile la trama di fondo del prodotto finale indipendentemente dall'entità dell'intervento del revisore, che si trova necessariamente impossibilitato a riscrivere quanto generato dalla macchina senza vanificare i vantaggi (in termini di tempi e costi) che tali strumenti dovrebbero garantire.

D'altro canto, se la *machine translation* permette un aumento di produttività considerevole, è altresì vero che limita la libertà espressiva e l'ingegno artistico¹²³. In tal senso risulta possibile osservare un effetto paradossale, per cui il processo di revisione diviene più dispendioso in termini di tempo, nonché, come affermato da Briasco, meno esaltante o emotivamente coinvolgente rispetto alla traduzione umana.

3.2.4. La perdita dello stile del traduttore

Una tematica strettamente connessa con le caratteristiche e i conseguenti rischi delle traduzioni automatiche è la progressiva perdita della presenza del traduttore nel testo. La traduzione letteraria, come illustrato da Kenny e Winters nel saggio «Machine translation, ethics and the literary translator's voice», si fonda sulla presenza discorsiva del traduttore: esistono infatti delle occasioni in cui tale "voce" emerge con evidenza, infrangendo la consueta cornice testuale entro cui domina la sola voce autoriale; ciò

¹²² Lieve Macken, Paola Ruffo, and Joke Daems. 2025. The Role of Translation Workflows in Overcoming Translation Difficulties: A Comparative Analysis of Human and Machine Translation (Post-Editing) Approaches. In Proceedings of the Second Workshop on Creative-text Translation and Technology (CTT), pages 1–13, Geneva, Switzerland. European Association for Machine Translation, 7.

¹²³ Moorkens, «The Machine Translator's Visibility» op. cit., 12.

si osserva, ad esempio, con le note di traduzione o addirittura in passaggi che possono risultare dissonanti rispetto all'originale. In quest'ottica risulta necessario dare riconoscimento alla voce del traduttore che, in quanto tale, rappresenta quindi un indice dell'agire e delle responsabilità del professionista rispetto alle scelte testuali effettuate per restituire l'intenzionalità del testo originale¹²⁴.

Alla luce di ciò è possibile identificare due tipologie distinte di voce: testuale e contestuale. Nello specifico, la voce del traduttore può appartenere a entrambe le categorie se, ad esempio, contemporaneamente all'atto di tradurre, si osservano dei commenti a piè di pagina alla traduzione o al testo fonte; la voce testuale, per contro, può essere desunta esaminando attentamente la collezione dei lavori effettuati dal traduttore, così da poterne individuare le preferenze stilistiche costanti. Sintetizzando, lo stile del traduttore è un modo di tradurre che risulti riconoscibile in una serie di traduzioni svolte dalla stessa persona, differenzi il suo lavoro da quello di altri traduttori e, infine, costituisca un modello coerente di scelte¹²⁵.

Le testimonianze dei professionisti non trovano un punto d'incontro riguardo tale tematica, che indubbiamente si configura come una questione controversa dello scenario traduttivo contemporaneo. Il primo intervistato, infatti, chiarisce fin dall'inizio la sua posizione a riguardo, dichiarando:

Questo tipo di lavoro, cioè questa sovrapposizione di voce – che è ovviamente il contrario dell'interpretazione –, questi fenomeni appropriativi sono pericolosi; per la stessa identica ragione sono pericolosi i fenomeni nei quali questo appiattimento si verifica in altro modo, cioè si verifica semplicemente perché c'è un'entità artificiale che fa determinate scelte al posto del traduttore.

¹²⁴ Kenny, Dorothy, e Marion Winters. «Machine Translation, Ethics and the Literary Translator's Voice». *Translation Spaces* 9, fasc. 1 (2020): 123–49. <https://doi.org/10.1075/ts.00024.ken>, 2.

¹²⁵ Ivi, 5

A cui si aggiunge un commento sull'invisibilità del traduttore, in quanto «il traduttore, in quel senso, è invisibile al servizio di un lavoro di interpretazione».

Quest'ultimo aspetto è stato anch'esso oggetto di molteplici dissertazioni, tra le quali si evidenzia il saggio di Lawrence Venuti intitolato *L'invisibilità del traduttore*. Tale concetto, così come delineato dall'autore, descrive la condizione paradossale in cui versa il traduttore professionista nelle culture contemporanee, un fenomeno complesso che si manifesta su due livelli interdipendenti: se da un lato è definibile come un effetto illusionistico prodotto dal discorso stesso, vale a dire dalla manipolazione della lingua d'arrivo realizzata dal traduttore, dall'altro è il risultato di una pratica consolidata di lettura e valutazione delle traduzioni che premia la scorrevolezza a scapito della fedeltà al senso profondo del testo originale. Sotto questo profilo, Venuti illustra chiaramente come il fulcro dell'invisibilità sia rappresentato da ciò che definisce come «the regime of fluency», secondo cui una traduzione è giudicata accettabile da editori, critici e lettori esclusivamente quando presenta una tale scorrevolezza da non intaccare la lettura, mostrandosi del tutto trasparente; l'autore stesso paragona una traduzione di buona qualità a una lastra di vetro attraverso la quale è possibile spingere il lettore a credere di percepire direttamente la personalità, l'intenzione o il significato essenziale dell'autore originale. Tale illusione di trasparenza comporta tuttavia un effetto di occultamento delle numerose condizioni sottostanti il processo traduttivo, a partire dall'intervento fondamentale del traduttore sul testo fonte, rimanendo pertanto invisibili al lettore comune¹²⁶.

3.2.5. Il concetto di *good enough quality* nella traduzione letteraria

La *good enough quality* è il concetto che, nella traduzione specialistica, identifica il raggiungimento di una soglia qualitativa che sia semplicemente sufficiente a garantire la comprensibilità e la correttezza informativa del testo; tuttavia, con la progressiva integrazione di strumenti di intelligenza artificiale nei flussi di lavoro

¹²⁶ Venuti, Lawrence. *The translator's invisibility: a history of translation*. Routledge, 2008.

traduttivi, tale linea di pensiero sembrerebbe in procinto di approdare in ambito letterario. Ragione per cui risulta necessario approfondirne le implicazioni, anche con il supporto delle testimonianze raccolte dai professionisti.

Ad oggi, nell'industria editoriale, l'automazione dei processi traduttivi viene calibrata attentamente attorno a una netta distinzione tra testi che richiedono una precisione assoluta e testi il cui scopo primario è la mera trasmissione di un messaggio. Secondo i principi di etica professionale illustrati da Moorkens, l'uso di strumenti di traduzione automatica dovrebbe essere direttamente proporzionale al ciclo di vita e al livello di rischio associato al testo tradotto: pertanto, contenuti dalla breve «shelf-life» e caratterizzati da scarsi rischi non giustificano in termini economici e di tempo lo sforzo considerevole che risulterebbe da una traduzione umana. L'applicazione di questa logica alla traduzione letteraria solleva tuttavia molteplici problematiche, dovute sia alla non neutralità della tecnologia sottostante a tali strumenti sia alle caratteristiche intrinseche della letteratura, che si vede naturalmente collocata nell'alveo dei testi ad alto valore¹²⁷.

Analizzando ulteriormente la questione, una diversa prospettiva viene fornita dalla teoria funzionalista dello *skopos* di Vermeer, secondo cui la qualità di una traduzione non può essere ritenuta un valore assoluto, in quanto deve configurarsi nell'adeguatezza del testo rispetto allo scopo per cui è stato prodotto. Analogamente agli approcci mostrati in precedenza si osserva una distinzione tra tipi testuali informativi ed espressivi: se nel primo caso il fine ultimo è la semplice trasmissione efficace e corretta di contenuti o conoscenze, l'ultima tipologia si caratterizza invece per un'organizzazione estetica e artistica consapevole di tale contenuto. Pertanto, il compromesso derivante dall'approccio della *good enough quality* non risulta accettabile in entrambe le istanze descritte: l'idea stessa di una soglia accettabile della qualità appare infatti indubbiamente problematica per la letteratura, poiché in tale contesto il concetto di sufficienza informativa rimane per definizione inadeguato¹²⁸.

¹²⁷ Moorkens, Joss. «Ethics and Machine Translation». Preprint, Zenodo, 27 giugno 2022. <https://doi.org/10.5281/ZENODO.6759984>.

¹²⁸ Reiss, Katharina., Hans J. Vermeer, Christiane Nord, e Marina Dudenhofer. *Towards a general theory of translational action : skopos theory explained*. Routledge, 2014.

Un'altra chiave di lettura significativa, strettamente connessa a quanto appena esposto, è stata fornita da Luca Briasco nell'intervista condotta come parte di questa ricerca, evidenziando come, in realtà, non esista un vero e proprio confine che possa stabilire cosa può essere definito letteratura e cosa non rientra in quella determinata categoria: esistono infatti numerosi testi appartenenti al cosiddetto genere della saggistica che, oggi, sono considerati al pari di autentiche opere letterarie come, ad esempio, il celebre *Dei delitti e delle pene* di Cesare Beccaria. L'intervistato commenta, legittimamente, che «a nessuno verrebbe in mente di farla passare dall'intelligenza artificiale». Tale criticità si solleva analogamente sulle opere contemporanee, per le quali è forse più complicato stabilirne i confini. Infine, la teoria funzionalista dello *skopos* viene avvalorata dalle parole di Briasco, che conclude il discorso ponendo l'accento su tre diversi elementi: circuito editoriale, natura del testo e fruitore finale.

Quindi secondo me è proprio la natura del testo, il circuito editoriale dentro cui il testo entra e il destinatario del testo a diventare rilevante: se il testo *L'imperatore del male* è destinato a tutte le persone che spesso purtroppo per motivi personali hanno un'esperienza anche molto dura nei confronti della realtà del cancro e che leggono quel testo per trovarci tutta una serie di elementi che risuonano, allora a quel punto l'intelligenza artificiale quel testo lo uccide. Se invece quel testo è stato preso da un qualunque editore il quale dice “a me questo testo serve perché fornisce una serie di informazioni storiche che possono essere utili per gli studenti di medicina, gli studenti di storia sociale, etc. allora forse lo si può anche affidare all'intelligenza artificiale.

Anche la seconda intervistata, Sara Bilotti, conferma la percezione generale che il concetto di *good enough quality* sia accettabile esclusivamente per alcune tipologie di testo:

La “qualità accettabile” di cui parla può andar bene per i testi che io chiamo “di servizio”: manuali, per esempio, o alcuni tipi di saggi. Non tutti. Tale compromesso tiene conto solo di un paio di dimensioni: manca la dimensione

intima, e tutto ciò che c'è tra le righe, e che solo una particolare costruzione della frase può suggerire. Un errore voluto, per esempio, può essere riconosciuto e risolto dall'IA, facendo perdere alla frase il suo senso più profondo. L'unica "qualità accettabile", nella traduzione letteraria, è quella altissima.

3.2.6. Il futuro della professione

Alla domanda «Se dovesse immaginare il traduttore professionista tra cinque anni, come descriverebbe il suo rapporto con questi strumenti e cosa, secondo lei, resterà umano in questo lavoro?» entrambi gli intervistati hanno naturalmente dichiarato di non sapere con certezza cosa riserverà il futuro per la professione del traduttore, tentando tuttavia di fornire degli spunti interessanti.

Innanzitutto, Sara Bilotti racconta il suo desiderio di vedere, in futuro, una tecnologia esclusivamente a supporto della professione e che, quindi, non sostituisca del tutto la figura finora intrinsecamente umana del traduttore:

Non posso prevedere il futuro, ma posso dirle cosa vorrei vedere: un traduttore con qualche ora libera in più - il lavoro di traduzione richiede un impegno di ore notevole - grazie al tempo risparmiato nella ricerca, un traduttore più sereno, che potrà impegnarsi nell'attività più importante che richiede il suo ruolo: l'interpretazione di una voce, nel suo significato più profondo.

Spero che di umano resti tutto, tranne le parti più noiose e meccaniche.

Luca Briasco, al contrario, prevede uno scenario caratterizzato da una distinzione fondata sugli anni di esperienza maturati dal professionista: un traduttore che «ormai è già a un certo punto del suo percorso [...] potrà continuare a esercitare le proprie scelte, perché probabilmente traduce tipi di cose che non verranno toccate nemmeno fra cinque anni dall'impatto dell'intelligenza artificiale». La problematica pertanto si pone per chi si approccia alla professione per la prima volta e per chi ha già tradotto, ma ha ancora relativamente pochi anni di esperienza.

In tal senso, in quanto docente di traduzione, Briasco consiglia fermamente a questi ultimi di «abituarsi anche all'idea che il vostro lavoro possa diventare un lavoro da revisori» e, quindi, di specializzarsi altresì in questo ambito. Tali affermazioni trovano nuovamente conferma nell'articolo di Moorkens, che consiglia l'attuazione di una «MT literacy» insieme, allargando le proprie prospettive, a una «AI literacy», con il fine di conseguire un utilizzo più consapevole e critico degli strumenti di traduzione automatica e dei modelli linguistici di grandi dimensioni¹²⁹.

¹²⁹ Moorkens, «The Machine Translator's Visibility» op. cit., 7.

Conclusione

La presente ricerca si è posta l'obiettivo di indagare l'impatto che la crescente diffusione dei modelli linguistici basati su intelligenza artificiale in ambito traduttivo ha prodotto sulla qualità delle traduzioni e sulla figura professionale del traduttore. In particolare, l'interesse si è focalizzato principalmente sul settore più specifico della traduzione letteraria che, per la complessità dei processi cognitivi e linguistici che lo accompagnano, solleva ancora oggi innumerevoli incognite sull'effettiva adeguatezza di tali strumenti.

In seguito a una breve contestualizzazione delle principali teorie della traduzione e a una precisa disamina dell'evoluzione delle tecnologie sottostanti ai moderni modelli linguistici di grandi dimensioni, si sono illustrate a titolo introduttivo le metriche di valutazione e le tendenze stilistiche ricorrenti delle traduzioni automatiche. Tali questioni sono state successivamente avvalorate da un'indagine qualitativa, che ha compreso sia la conduzione di un esperimento di traduzione, su cui si è basata un'analisi testuale, sia un'ulteriore verifica tramite le testimonianze di due professionisti del settore.

L'esperimento di traduzione, innanzitutto, ha previsto la costruzione di un solido corpus di testi generati artificialmente, i quali sono stati poi sottoposti a un processo di traduzione automatica attraverso l'utilizzo di tre diversi modelli di IA: Claude, ChatGPT e Gemini. L'analisi testuale delle traduzioni ha confermato la presenza sistematica di tendenze stilistiche ricorrenti all'interno del corpus, che hanno consentito di delineare dei profili comportamentali specifici per ciascun modello: se Claude si è distinto per la sua quasi totale aderenza al testo originale, ChatGPT e Gemini d'altro canto hanno manifestato una maggiore flessibilità nella traduzione, che ha tuttavia causato l'emergere di allucinazioni e incoerenze a livello semantico e sintattico.

I fenomeni emersi in quest'analisi hanno trovato un significativo riscontro nelle riflessioni messe in luce dagli intervistati, evidenziando una forte connessione tra le caratteristiche delle traduzioni automatiche e le criticità percepite dai professionisti. In tal senso, alcune questioni assumono una particolare rilevanza: ad

esempio, la pressione esercitata dall'effetto ancoraggio nel processo di revisione delle traduzioni automatiche emerge come un elemento di grande influenza capace di orientare le scelte del traduttore. Parallelamente, si osservano potenziali rischi di una generale omogeneizzazione degli stili dell'autore e del traduttore, così come un progressivo impoverimento lessicale causato dalla ciclicità dei pattern linguistici contenuti nei dati di addestramento di tali modelli. Un'ultima interessante riflessione ha visto l'eventualità dell'applicazione del concetto di *good enough quality* nel contesto letterario, sollevando dubbi legittimi sui confini che delineano cosa può essere considerato letteratura e cosa, invece, non può. Alla luce di queste considerazioni, si evince che la figura del traduttore letterario è in fase di ridefinizione: a seconda del settore di specializzazione e del bagaglio di esperienza maturato, infatti, il professionista potrebbe confrontarsi con la concreta possibilità che il proprio ruolo si riduca progressivamente a quello di revisore di traduzioni automatiche, con tutto ciò che ne consegue.

Tuttavia, questa indagine presenta dei chiari limiti che si rende necessario esplicitare. In primo luogo, la scelta di utilizzare un insieme di testi generato artificialmente come corpus su cui basare l'analisi ha sicuramente evitato problematiche legate al diritto d'autore, ma ha altresì orientato in modo significativo i risultati ottenuti con le traduzioni; contemporaneamente, la natura non deterministica dei modelli linguistici ha impedito di considerare gli output del tutto rappresentativi del loro comportamento. In secondo luogo, la traccia alla base delle interviste ha tralasciato consapevolmente una serie di questioni relative al panorama editoriale italiano, che avrebbero permesso di giungere a una visione globale e completa della problematica posta dalla domanda di ricerca. Tale scelta è giustificata dalla volontà di esplorare esclusivamente le altrettanto interessanti prospettive tecniche della traduzione e, soprattutto, le implicazioni derivate dalla massiccia diffusione dell'intelligenza artificiale in questo ambito. Infine, il campione ridotto degli intervistati pone anch'esso un grande limite ai fini della ricerca, la quale si propone pertanto con un intento esplorativo della questione.

Considerando i risultati e i limiti emersi, un possibile sviluppo futuro di questa indagine potrebbe essere rappresentato, anzitutto, da un ampliamento del corpus testuale, così come del campione a cui somministrare le interviste; data la crescente rapidità con cui si trasforma il panorama dell'intelligenza artificiale, risulterebbe altresì stimolante ripetere l'esperimento di traduzione a intervalli prestabiliti, in modo tale da poter ottenere un quadro evolutivo delle capacità di tali modelli registrandone i progressi (o le persistenti limitazioni) nel corso del tempo.

In definitiva, l'introduzione dei processi di automazione linguistica nel campo della traduzione letteraria apre una serie di interrogativi che toccano il rapporto fondamentale tra tecnologia e creatività, mettendo in luce il timore che la macchina possa, in futuro, sostituire ciò che tradizionalmente si è sempre considerato umano.

Bibliografia e sitografia

Baron, Naomi S. *Who Wrote This? How AI and the Lure of Efficiency Threaten Human Writing*. Stanford University Press, 2023.

Benjamin, Walter, e Maria Teresa Costa. *Il compito del traduttore*. Minima/Volti, N. 115. Mimesis, 2023.

Bergmann, Dave. «Cos'è il machine learning? | IBM». 18 agosto 2025. <https://www.ibm.com/it-it/think/topics/machine-learning>.

Berman, Antoine. *La traduzione e la lettera o l'albergo nella lontananza*. Quodlibet, 2022.

Bitton, Yehonatan, Elad Bitton, e Shai Nisan. «Detecting Stylistic Fingerprints of Large Language Models». Preprint, arXiv, 2025. <https://doi.org/10.48550/ARXIV.2503.01659>.

Brower, Reuben Arthur, a c. di. «On Linguistic Aspects of Translation». In *On Translation*. Harvard University Press, 1959. <https://doi.org/10.4159/harvard.9780674731615.c18>.

Chakrabarty, Tuhin, Philippe Laban, e Chien-Sheng Wu. «Can AI writing be salvaged? Mitigating Idiosyncrasies and Improving Human-AI Alignment in the Writing Process through Edits». Preprint, arXiv, 2024. <https://doi.org/10.48550/ARXIV.2409.14509>.

Chan, Sin-wai, a c. di. *The Routledge Encyclopedia of Translation Technology*. Routledge, 2015.

Chugani, Vinod. «AI Winter: Understanding the Cycles of AI Development». Datacamp, 3 novembre 2025. <https://www.datacamp.com/blog/ai-winter>.

Cook, Angela, Angie Knaggs, e Jiashun Sun. «Exploring the Goals of Translation Education: Adding Human Value to Literary Translation through Deep Meaning Making and Artistry». *The Interpreter and Translator Trainer* 19, fasc. 3–4 (2025): 254–76. <https://doi.org/10.1080/1750399X.2025.2543208>.

Cui, Menglong, Jiangcun Du, Shaolin Zhu, e Deyi Xiong. «Efficiently Exploring Large Language Models for Document-Level Machine Translation with

In-context Learning». Preprint, arXiv, 2024.
<https://doi.org/10.48550/ARXIV.2406.07081>.

Daems, Joke, Paola Ruffo, e Lieve Macken. «Impact of translation workflows with and without MT on textual characteristics in literary translation». In *Proceedings of the 1st Workshop on Creative-text Translation and Technology*, a cura di Bram Vanroy, Marie-Aude Lefer, Lieve Macken, e Paola Ruffo. European Association for Machine Translation, 2024. <https://aclanthology.org/2024.ctt-1.6/>.

Donato, Benedetta, Leonardo Mariani, Daniela Micucci, e Oliviero Riganelli. «Studying How Configurations Impact Code Generation in LLMs: the Case of ChatGPT». Preprint, arXiv, 2025. <https://doi.org/10.48550/ARXIV.2502.17450>.

Eco, Umberto. *Dire quasi la stessa cosa: esperienze di traduzione*. 7. ed. Studi Bompiani Il campo semiotico. Bompiani, 2007.

Ferrara, Alfio. *Le macchine del linguaggio. L'uomo allo specchio dell'intelligenza artificiale*. Giulio Einaudi Editore, 2025.

Guo, Yanzhu, Guokan Shang, e Chloé Clavel. «Benchmarking Linguistic Diversity of Large Language Models». Preprint, arXiv, 2024. <https://doi.org/10.48550/ARXIV.2412.10271>.

Jaśim, Yaser A., e Ali Jala Awqati. «Distinguishing Human Creativity from AI-Generated Literary Texts». *JOURNAL OF PROSPECTIVE RESEARCHES* 25, fasc. 2 (2025): 40–47. <https://doi.org/10.61704/pr.494>.

John, McCarthy, Minsky Marvin L., Rochester Nathaniel, e Shannon Claude E. «A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence». *AI Magazine* 27, fasc. 4 (2006): 12–14.

Kavlakoglu, Jacob Murel Ph D., Eda. «Cos'è una bag of words? | IBM». 22 dicembre 2023. <https://www.ibm.com/it-it/think/topics/bag-of-words>.

Kenny, Dorothy, e Marion Winters. «Machine Translation, Ethics and the Literary Translator's Voice». *Translation Spaces* 9, fasc. 1 (2020): 123–49. <https://doi.org/10.1075/ts.00024.ken>.

Macken, Lieve, Paola Ruffo, e Joke Daems. «The Role of Translation Workflows in Overcoming Translation Difficulties: A Comparative Analysis of

Human and Machine Translation (Post-Editing) Approaches». In *Proceedings of the Second Workshop on Creative-text Translation and Technology (CTT)*, a cura di Bram Vanroy, Marie-Aude Lefer, Lieve Macken, Paola Ruffo, Ana Guerbero Arenas, e Damien Hansen. European Association for Machine Translation, 2025. <https://aclanthology.org/2025.ctt-1.1/>.

Manzanarez, Gerardo Aleman, Raul Monroy, Jorge Garcia Flores, e Hiram Calvo. «Literary Language Mashup: Curating Fictions with Large Language Models». *Mathematics* 14, fasc. 2 (2026): 210. <https://doi.org/10.3390/math14020210>.

Moorkens, Joss. *Ethics and Machine Translation*. Zenodo, 2022. <https://doi.org/10.5281/ZENODO.6759984>.

Moorkens, Joss. «The Machine Translator's Visibility: A Postphenomenological Analysis of Machine Translation». *Translation Spaces*, pubblicazione online ad accesso anticipato, 6 ottobre 2025. <https://doi.org/10.1075/ts.23030.moo>.

Moorkens, Joss, Sheila Castilho, Federico Gaspari, e Stephen Doherty, a c. di. *Translation Quality Assessment: From Principles to Practice*. Vol. 1. Machine Translation: Technologies and Applications. Springer International Publishing, 2018. <https://doi.org/10.1007/978-3-319-91241-7>.

Nergaard, Siri, a c. di. *Teoria della traduzione nella storia*. Strumenti Bompiani. Bompiani, 2002.

O'Sullivan, James. «Stylometric Comparisons of Human versus AI-Generated Creative Writing». *Humanities and Social Sciences Communications* 12, fasc. 1 (2025): 1708. <https://doi.org/10.1057/s41599-025-05986-3>.

Ploeger, Esther, Huiyuan Lai, Rik van Noord, e Antonio Toral. «Towards Tailored Recovery of Lexical Diversity in Literary Machine Translation». Preprint, arXiv, 2024. <https://doi.org/10.48550/ARXIV.2408.17308>.

Prabowo, Bayu Ade, e Rahmanti Asmarani. «Generative Literature: The Role of Artificial Intelligence in the Creative Writing Process». *Allure Journal* 5, fasc. 1 (2025): 1–9. <https://doi.org/10.26877/allure.v5i1.19959>.

Reiß, Katharina, e Hans J. Vermeer. *Towards a General Theory of Translational Action: Skopos Theory Explained*. Tradotto da Christiane Nord. Routledge, Taylor & Francis Group, 2015.

Rivera-Trigueros, Irene. «Machine Translation Systems and Quality Assessment: A Systematic Review». *Language Resources and Evaluation* 56, fasc. 2 (2022): 593–619. <https://doi.org/10.1007/s10579-021-09537-5>.

Singh, Aaditya, Adam Fry, Adam Perelman, et al. «OpenAI GPT-5 System Card». Preprint, arXiv, 2026. <https://doi.org/10.48550/ARXIV.2601.03267>.

Tavosanis, Mirko. «Valutare la qualità dei testi generati in lingua italiana». *AI-Linguistica. Linguistic Studies on AI-Generated Texts and Discourses* 1, fasc. 1 (2024). <https://doi.org/10.62408/ai-ling.v1i1.14>.

Thompson, Brian, Mehak Preet Dhaliwal, Peter Frisch, Tobias Domhan, e Marcello Federico. «A Shocking Amount of the Web is Machine Translated: Insights from Multi-Way Parallelism». arXiv:2401.05749. Preprint, arXiv, 5 giugno 2024. <https://doi.org/10.48550/arXiv.2401.05749>.

Tolkien, J. R. R. *Il Silmarillion*. 1977; Bompiani, 2023.

Toosi, Amirhosein, Andrea G. Bottino, Babak Saboury, Eliot Siegel, e Arman Rahmim. «A Brief History of AI: How to Prevent Another Winter (A Critical Review)». *PET Clinics* 16, fasc. 4 (2021): 449–69. <https://doi.org/10.1016/j.cpet.2021.07.001>.

Toral, Antonio. «Post-edite: an Exacerbated Translationese». Preprint, arXiv, 2019. <https://doi.org/10.48550/ARXIV.1907.00900>.

Treccani. «Traduzione automatica - Enciclopedia». Consultato 15 aprile 2026. [https://www.treccani.it/enciclopedia/traduzione-automatica_\(Lessico-del-XXI-Secolo\)/](https://www.treccani.it/enciclopedia/traduzione-automatica_(Lessico-del-XXI-Secolo)/).

Vanmassenhove, Eva, Dimitar Shterionov, e Matthew Gwilliam. «Machine Translationese: Effects of Algorithmic Bias on Linguistic Complexity in Machine Translation». Preprint, arXiv, 2021. <https://doi.org/10.48550/ARXIV.2102.00287>.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, et al. «Attention Is All You Need». arXiv:1706.03762. Preprint, arXiv, 2 agosto 2023. <https://doi.org/10.48550/arXiv.1706.03762>.

Venuti, Lawrence. *The Translator's Invisibility: A History of Translation*. 2. ed., Repr. Routledge, 2009.

Vetere, Guido. *Intelligenze aliene. Linguaggio e vita degli automi*. Luca Sossella Editore, 2025.

Vlaad, Sofie. «Texts without Authors: Ascribing Literary Meaning in the Case of AI». *The Journal of Aesthetics and Art Criticism* 83, fasc. 1 (2025): 4–11. <https://doi.org/10.1093/jaac/kpae047>.

Wang, Haifeng, Hua Wu, Zhongjun He, Liang Huang, e Kenneth Ward Church. «Progress in Machine Translation». *Engineering* 18 (novembre 2022): 143–53. <https://doi.org/10.1016/j.eng.2021.03.023>.

Wang, Kuang-Da, Shuoyang Ding, Chao-Han Huck Yang, et al. «Extending Automatic Machine Translation Evaluation to Book-Length Documents». Preprint, arXiv, 2025. <https://doi.org/10.48550/ARXIV.2509.17249>.

Wang, Longyue, Zefeng Du, Wenxiang Jiao, et al. «Benchmarking and Improving Long-Text Translation with Large Language Models». *Findings of the Association for Computational Linguistics ACL 2024*, 2024, 7175–87. <https://doi.org/10.18653/v1/2024.findings-acl.428>.

Appendici

Intervista a Luca Briasco

Può descrivermi brevemente il suo percorso come traduttore e qual è, ad oggi, il suo rapporto con strumenti di traduzione automatica o LA generativa?

Il percorso da traduttore parte alla fine degli anni Novanta ed è molto vicino al mio percorso editoriale, perché ho lavorato anche all'interno di case editrici: ho infatti cominciato a tradurre per le case editrici per cui lavoravo. Poi si è progressivamente ampliato e allargato, quindi attualmente sono arrivato a mettere insieme un centinaio di traduzioni che sono tutte quante dall'inglese, che è l'unica lingua da cui traduco (pur conoscendo anche francese e spagnolo, ma è l'unica lingua in cui mi sento abbastanza confidente per poter tradurre al livello a cui spero di riuscire a tradurre). Ho tradotto solo quasi esclusivamente narrativa, tranne un paio di titoli di saggistica, di autori sia classici sia contemporanei: classici tipo Coldwell, Ballard, etc. e contemporanei tipo Stephen King.

Non ho mai utilizzato intelligenza artificiale per tradurre, però, essendo anche docente di traduzione per una scuola (Minimum Lab) legata alla casa editrice per cui lavoro adesso (Minimum Fax), ho studiato alcuni programmi di intelligenza artificiale per vedere che tipo di resa davano a livello di traduzione e per ragionare con i miei studenti sui possibili sviluppi futuri della traduzione per effetto dell'intelligenza artificiale. Quindi questo è il rapporto che ho: ovviamente non è un rapporto professionale, perché non la utilizzo e non ho intenzione di utilizzarla neanche in futuro.

Quando legge una traduzione prodotta con il supporto dell'IA, riesce a riconoscerla? Su quali elementi si basa per farlo?

Allora, ci sono alcuni elementi che si notano che, per esempio, sono legati alla difficoltà che molti programmi di intelligenza artificiale hanno di lavorare sulla

ricorsività di certe forme espressive o perfino di certi termini o di certi giri di frase: cioè l'intelligenza artificiale tende, almeno per ora, a tradurre in un modo per cui se tu hai l'originale di fronte capisci immediatamente che è una scelta che il traduttore fa di ripetere o di variare un modulo espressivo all'interno della traduzione, in maniera proprio da ricavare anche un effetto di ritmica e di richiamo interno al testo. Questa è una cosa che, da quello che ho visto almeno per ora, i programmi di intelligenza artificiale tendono a non fare, mentre secondo me lavorano sul senso in maniera globalmente corretta e può perfino succedere in alcuni casi che un programma di intelligenza artificiale traduca in modo più letterale, quindi in modo più corretto rispetto al traduttore: per esempio, se un traduttore può commettere un errore di senso, un programma di intelligenza artificiale è difficile che commetta quel tipo di errore. Però dal punto di vista proprio della consapevolezza dello stile dell'autore, io in quello vedo ancora dei forti limiti e uno dei marker dello stile di un autore è proprio il modo in cui l'autore applica la ricorsività, quindi applica il giro di frase replicato, richiamato; poi ci sono addirittura degli autori che usano proprio questo come modalità di trama, quindi fare il richiamo diretto di un pensiero che magari è stato formulato da un personaggio 50 pagine prima, l'intelligenza artificiale 50 pagine dopo spesso non te lo ritraduce in modo identico a come ha fatto 50 pagine prima, perché basta che la costruzione della frase sia differente e nella traduzione magari viene diversa. Quindi, se io devo guardare un prodotto che arriva direttamente dall'intelligenza artificiale – quindi senza un intervento umano intermedio – tendo a vedere ricorrere alcune di queste cose che mi fanno capire quasi automaticamente che lì non c'è una mano umana o se c'è una mano umana sarebbe una mano terribilmente distratta.

Per esempio non possono capire l'impianto simbolico: un termine che ricorre più volte che ha anche una polisemia – quindi ha una ricchezza di significati e di possibili traduzioni a seconda del contesto della frase –, un traduttore umano capisce che quel termine è un nucleo simbolico del testo, di conseguenza fa una scelta (giusta o sbagliata che sia) ma la replica ogni volta perché è la sua scelta. Da quel punto di vista io vedo che, almeno per ora, i programmi di IA questa cosa non ce l'hanno.

Ritiene che lavorare a partire da un testo pre-tradotto automaticamente possa condizionare stilisticamente il processo di traduzione rispetto alla traduzione diretta dal testo originale?

Io sono convinto di sì. È impossibile pensare il contrario, nel senso che comunque se noi anche immaginiamo che un testo tradotto con l'intelligenza artificiale produca una sorta di trama rozza dentro la quale poi il revisore lavora di cesello e cerca di sistemare i fili, rimane il fatto che questa trama – per me – tende a rimanere rattoppata: l'intervento del revisore è più una toppa che non una riscrittura, perché se dovesse essere una riscrittura si creerebbe il grande paradosso per il quale rivedere un testo tradotto dall'intelligenza artificiale diventa riscriverlo e, quindi, diventa un moltiplicatore di tempo anziché un riduttore di tempo.

I casi sono due: o tu accetti che quel primo prodotto – anche rough – generato dall'intelligenza artificiale costituisce la base del tuo lavoro e allora a quel punto ti metti a lavorare su quella base, la metti a posto, la sistemi, cogli gli elementi che non funzionano e intervieni su quelli. Rimane il fatto che la tessitura di base deve essere quella della traduzione fornita dall'intelligenza artificiale, altrimenti viene meno lo scopo ultimo dell'intelligenza artificiale in traduzione, che è lo scopo di ridurre i tempi e di ridurre i costi. Se i tempi aumentano, aumentano i costi e se aumentano i tempi, aumentano proprio perché uno, invece di utilizzare la traduzione fornita dall'intelligenza artificiale come base che non viene messa in discussione in quanto tale, se non la si utilizza questo tappeto si trasforma in un gigantesco ostacolo, perché a quel punto a te tocca ritradurre e nel ritradurre sei frenato dal testo che hai per le mani. È quasi impossibile che questo non accada, quindi è indispensabile che si prenda atto o che si dica che quel testo prodotto dall'intelligenza artificiale è la traduzione, perché se tu non parti da quest'idea e parti dall'idea di usarla ma di fare comunque quello che vuoi, va a finire che – per prima cosa – la traduzione rimane un po' incastrata dentro quel testo indipendentemente dalla tua volontà e, seconda cosa, perdi tempo invece di guadagnarlo.

Quindi tu dovresti entrare su quella traduzione come entrano i revisori: anche i revisori delle traduzioni delle traduzioni fatte da un essere umano fanno

perfettamente, nel momento in cui intervengono sulla traduzione, che non la possono riscrivere. In quel caso subentra anche un principio di autorialità, perché seppur di secondo grado il traduttore è un autore ed è una firma e quindi naturalmente sopra una firma tu non intervieni riscrivendo, ma intervieni su quelli che sono gli errori materiali, ma non puoi fare molto più di questo. Lo stesso discorso inevitabilmente deve valere per la traduzione con l'intelligenza artificiale: allo stato attuale con degli interventi che inevitabilmente tendono a essere un pochino più pesanti e un pochino più estensivi, però esiste un limite e quel limite va studiato. Io credo che sia un fenomeno che vada studiato nel momento in cui la traduzione con l'intelligenza artificiale dovesse diventare una base almeno per la traduzione del contemporaneo, perché secondo me per la traduzione dei classici è proprio impensabile utilizzarla.

Intervengo commentando che, per ora, il post-editing appartiene più che altro a contesti specialistici e che in narrativa letteraria ancora non si è provato ad utilizzarlo.

No, non si è provato. Io infatti lo faccio fare, però lo faccio fare a titolo di esercitazione perché secondo me – da docente, non da traduttore – è importante che ci si metta già dentro quest'ottica: se non lo si fa si rischia poi di rimanere completamente spiazzati e di soffrire di conseguenza di un calo drastico della qualità della traduzione, perché anche rivedere una traduzione con intelligenza artificiale diventa un mestiere. Questo mestiere ha degli aspetti nuovi proprio per le caratteristiche specifiche di cui accennavamo prima: un testo tradotto con l'intelligenza artificiale presenta una serie di problemi che non sono gli stessi di un testo tradotto da un essere umano (magari anche male), ma sono comunque delle problematiche completamente diverse che bisogna imparare a studiare il prima possibile, perché tanto a questo arriveremo. Io sono abbastanza convinto che purtroppo – anche per la narrativa o almeno per una parte di narrativa – si arriverà prima o poi a questo: è solo questione di tempo.

Quali rischi potrebbe comportare l'utilizzo dell'IA generativa o di strumenti di traduzione automatica se applicati alla traduzione letteraria?

Allora i rischi sono inevitabili, i rischi sono che lo stile e anche l'idiosincrasia dello stile del singolo autore diventino un aspetto penalizzato. Finché tutto questo avviene – quindi l'intelligenza artificiale interviene su testi che sono fondamentalmente di trama e quindi lo stile è quasi trasparente, cioè serve solo ad accompagnare lo sviluppo di una narrazione – il danno è relativo; nel momento in cui l'autore ha una cifra specifica (ripeto questa non ha a che fare con i generi letterari o simili, perché autori con una cifra specifica si trovano un po' ovunque) allora a quel punto il danno è che lo stile viene totalmente appiattito e che, di conseguenza, prevale solamente l'aspetto comunicativo e non l'aspetto espressivo. Quindi – intervenendo sulla traduzione con l'intelligenza artificiale – sull'aspetto comunicativo e quindi sull'efficacia della traduzione dal punto di vista puramente di significato è in grado perfettamente di intervenire; invece il significante e tutto quello che ha a che fare con la forma espressiva scelta, lo stile, eccetera, rischia di subire un appiattimento.

Naturalmente il problema qual è? È ovvio che, se su certe tipologie di letteratura nelle quali il concetto stesso di autore è un concetto abbastanza labile, come il caso di Harper e, quindi, il caso di affidare per esempio gli harmony all'intelligenza artificiale. Il punto è che negli harmony l'autorialità è un principio molto relativo: quello che conta è il tipo di storia, il tipo di trama e anzi addirittura una certa ripetitività delle forme espressive, perché è proprio il classico tipo di libro prodotto in serie in cui il lettore si aspetta determinate cose e paradossalmente certe fughe stilistiche non sarebbero nemmeno apprezzate. Allora in questo caso forse il danno è limitato, forse addirittura non c'è; ma in tutti i casi in cui c'è l'evidenza di uno stile specifico e di una forte individualità anche della forma espressiva, il rischio dell'appiattimento è altissimo.

Questo è il motivo per il quale naturalmente non penso che succederà mai che l'intelligenza artificiale vada a impattare per esempio sulla traduzione di classici, perché sono fuori diritto d'autore e la traduzione diventa l'investimento primario: diventa

quindi molto importante chi traduce. Lì il traduttore diventa a tutti gli effetti un secondo autore e in quei casi l'intelligenza artificiale non ha ragione di esistere; però c'è tutto un terreno intermedio che è quello della narrativa contemporanea, su cui ci sono dei rischi oggettivi.

La traduzione letteraria implica un lavoro di interpretazione, di conseguenza ogni professionista traduce secondo una sua personale visione dell'opera originale, naturalmente cercando di mantenere intatto lo stesso effetto a cui mirava l'originale. Pensa che l'introduzione di sistemi di IA nel processo traduttivo possa in qualche modo portare progressivamente alla perdita dello stile personale del traduttore?

Questo certamente. Io dico sempre una cosa: fra una traduzione costellata di errori (di senso, di resa), probabilmente perché si è in presenza di un traduttore che non è accurato nel suo lavoro di ricerca – ormai ci sono degli strumenti talmente avanzati con la rete che commettere degli errori di senso nella traduzione è spesso l'effetto più di una distrazione che di altro – in cui c'è stato un lavoro di resa personale, di interpretazione personale, di immedesimazione personale nel testo e un lavoro di traduzione molto pulito, molto ben fatto in cui però c'è quasi una voce che si sovrappone a quella dell'autore – ovvero c'è assenza di lavoro interpretativo sul testo – bisogna sempre preferire il primo. Bisogna sempre preferire una traduzione anche con molti errori, ma nella quale questo sforzo di immedesimarsi nel testo originale, di renderlo in una nuova lingua sia stato compiuto; dove non è compiuto, secondo me non c'è un buon traduttore.

A me sono capitati dei casi, perché ho anche lavorato tantissimo – avendo lavorato per tante case editrici – come revisore, e in particolare un caso specifico nel quale io leggevo le traduzioni di questo traduttore di diversi autori e la singola traduzione mi sembrava meravigliosa, perché era bella tornita, con uno stile molto riconoscibile, peccato che poi tutti questi autori erano tradotti allo stesso modo. Quindi a un certo punto tre autori completamente diversi erano stati trasformati da questo traduttore in tre libri che sembravano frutto della stessa penna. Questo tipo di

lavoro, cioè questa sovrapposizione di voce – che è ovviamente il contrario dell'interpretazione –, questi fenomeni appropriativi sono pericolosi; per la stessa identica ragione sono pericolosi i fenomeni nei quali questo appiattimento si verifica in altro modo, cioè si verifica semplicemente perché c'è un'entità artificiale che fa determinate scelte al posto del traduttore.

C'è un grosso pericolo che riguarda certamente anche lo stile del traduttore, che ha chiaramente un suo stile ed è inevitabile, perché ognuno di noi ha un giro di frase, alcune caratteristiche e se vogliamo perfino alcune idiosincrasie che ci rendono riconoscibili e che noi utilizziamo in modo quasi inconsapevole; però il traduttore veramente bravo non si deve riconoscere perché deve cambiare e deve adattare il suo stile in relazione al testo su cui sta lavorando. Quindi è del tutto ovvio che il traduttore, in quel senso, è invisibile al servizio di un lavoro di interpretazione e questo lavoro di interpretazione del testo originale l'intelligenza artificiale non è in grado di farlo. È ovvio che se tu lavori sull'intelligenza artificiale da revisore anche molto esperto, anche molto bravo, anche aggiungendo delle cose tue o provando a interpretare, comunque sei inchiodato a quel testo per ragioni di tempo e per ragioni di condizionamento. Comunque tu vedi una resa e nonostante tutto quella resa ti rimane addosso e non è che tutte le volte a quella resa sostituisci la tua: spesso e volentieri sottostai a quello che hai per le mani e te ne accontenti, anche perché da un punto di vista formale non è sbagliata. Il punto è anche quello: spesso in una traduzione allo stato attuale – per il tipo di programmi su cui io ho provato a fare degli esperimenti – io vedo che l'80%, l'85% anche il 90% di una traduzione non è di per sé sbagliato, è anche corretto. Il problema è: ci basta una traduzione corretta? Forse per alcune cose sì, mentre per un certo tipo di letteratura secondo me no.

Nel dibattito sulla traduzione automatica nel campo delle traduzioni specialistiche si parla spesso di good enough quality, ovvero una qualità sufficiente a trasmettere correttamente e comprensibilmente il contenuto di un testo specialistico. Ritiene che nel caso della traduzione letteraria questo compromesso sia accettabile? O la soglia della “qualità accettabile” non esiste per definizione quando si tratta di letteratura?

Gli effetti dell'intelligenza artificiale possono essere letti anche al contrario: nel momento in cui tu ti difendi in modo acritico dall'impatto dell'intelligenza artificiale, tu tendi a sacralizzare la letteratura per effetto di questo. È una forma di sacralizzazione difensiva. Il punto fondamentale, bisogna evitare sacralizzazioni dicendo che la letteratura non si tocca e tutto il resto si può toccare, perché allora nasce automaticamente la domanda “che cos'è letteratura e che cosa non lo è?”. Dove stanno i confini? Chi li stabilisce? Perché secondo me fare dei confini di genere e dire – la poesia, ovviamente, io su quello posso essere d'accordo: a nessuno viene in mente di affidare la traduzione della poesia all'intelligenza artificiale e credo che non verrà mai fatto, perché lì c'è proprio una questione nella quale subentrano altri elementi come le assonanze, la metrica, etc. Quindi dubito per quanto possa evolversi l'intelligenza artificiale, credo che l'intelligenza artificiale evolve anche in relazione a valutazioni e considerazioni di ordine economico e secondo me la poesia non è sufficientemente attrattiva per fare operazioni del genere, quindi non credo che troveremo mai una Divina Commedia tradotta con l'intelligenza artificiale, ma le ragioni sono altre.

Il punto è che nel momento in cui tu dici per un certo tipo di traduzione specialistica la good enough quality è l'elemento che tutto sommato...

Il confine è anche della cosiddetta saggistica: oggi Dei delitti e delle pene di Beccaria, noi lo consideriamo un testo sul quale si applica il principio della good enough quality o ormai la consideriamo letteratura? Questo per capire che anche i confini della letteratura cambiano, anche in relazione all'impatto delle opere, quindi un'opera come quella di Beccaria, per esempio, oggi a nessuno verrebbe in mente di farla passare dall'intelligenza artificiale. Allora anche sul contemporaneo, qual è il confine? Il punto è che io fatico a definire il confine, il quale non può che essere legato alla volontà di chi pubblica in primo luogo e quindi del tipo di pubblico al quale si vuole arrivare. Per dire, se il testo di un manuale di medicina tu lo destini a studenti di medicina che devono apprendere delle nozioni, allora a quel punto la good enough quality vale; se però ti imbatti in un testo che viene studiato anche a medicina – come

per esempio l'Imperatore delle malattie, che è una storia del cancro – quello è un testo che ha vinto un premio Pulitzer e se tu lo metti in mano all'intelligenza artificiale, il risultato temo sia lo stesso catastrofico che tu applicheresti a un'opera letteraria. Quindi secondo me è proprio la natura del testo, il circuito editoriale dentro cui il testo entra e il destinatario del testo a diventare rilevante: se il testo L'imperatore del male è destinato a tutte le persone che spesso purtroppo per motivi personali hanno un'esperienza anche molto dura nei confronti della realtà del cancro e che leggono quel testo per trovarci tutta una serie di elementi che risuonano, allora a quel punto l'intelligenza artificiale quel testo lo uccide. Se invece quel testo è stato preso da un qualunque editore il quale dice “a me questo testo serve perché fornisce una serie di informazioni storiche che possono essere utili per gli studenti di medicina, gli studenti di storia sociale, etc. allora forse lo si può anche affidare all'intelligenza artificiale.

Quindi per me il discorso diventa un altro, ovvero diventa in quale circuito editoriale e per quale fruitore finale è destinata la traduzione. In tutti i casi in cui il concetto di fruitore finale si allarga o addirittura rimane indeterminato, perché il testo potrebbe davvero finire nelle mani di qualunque lettore, c'è un rischio con l'intelligenza artificiale, c'è una difficoltà nel determinare il concetto stesso di good enough quality; cioè non sai dov'è questo confine, non sai se l'intelligenza artificiale lo copre oppure no.

Se dovesse immaginare il traduttore professionista tra cinque anni, come descriverebbe il suo rapporto con questi strumenti e cosa, secondo lei, resterà umano in questo lavoro?

Allora questo ovviamente io fatico a determinarlo, però tendo a dire una cosa: gli impatti saranno di due tipi. Innanzitutto bisognerà distinguere fra il traduttore professionista che ormai è già a un certo punto del suo percorso e che quindi potrà continuare a esercitare le proprie scelte, perché probabilmente traduce tipi di cose che non verranno toccate nemmeno fra cinque anni dall'impatto dell'intelligenza artificiale. Se penso a un autore come Stephen King – uno degli autori che traduco io – non sarà mai tradotto dall'intelligenza artificiale finché vive, perché ormai sta in un

circuito, in un certo tipo di valutazione ed è considerato uno dei grandi autori a cavallo fra vecchio e nuovo millennio, quindi è un autore che non verrà toccato. Di conseguenza è chiaro che chi è arrivato a un livello per cui traduce questo tipo di autori ne risentirà in modo molto relativo. Il problema è chi si avvia la traduzione adesso o chi ha tradotto ancora relativamente poco o chi traduce per esempio tutta la cosiddetta letteratura seriale (non solo gli harmony, ma penso ai romance) e lì c'è un forte rischio che l'intelligenza artificiale si mangi una notevole quantità di spazio.

Io per questo dico anche quando faccio lezioni di traduzione di abituarsi anche all'idea che il vostro lavoro possa diventare un lavoro da revisori e, quindi, di mettersi nelle condizioni di lavorare su un testo chiaramente per delle cifre che non saranno più le cifre che prendono i traduttori – peraltro abbastanza misere in Italia –, però saranno cifre ancora più basse, perché è chiaro che utilizzare l'intelligenza artificiale per nessun editore ha un valore di “io faccio il pioniere e uso l'intelligenza artificiale”, cioè io non vedo pionierismo in tutto questo: vedo solamente calcolo economico. L'intelligenza artificiale mi costerà enormemente di meno, quindi mi permetterà di comprimere i costi e quindi di realizzare un guadagno anche in presenza di un numero di copie vendute più basso rispetto al margine di guadagno che c'è oggi.

Allora davanti a questo io comincio a dire a chi si avvicina alla traduzione di prepararsi anche all'idea di specializzarvi in quest'ambito, perché se imparate a fare bene questo lavoro avrete lavoro; magari non sarà un lavoro esaltante o emotivamente coinvolgente come quello della traduzione classica, però temo che questo sarà uno sviluppo che fra cinque anni coprirà anche – oltre ovviamente agli ambiti specialistici in cui la si utilizza già – almeno una parte della narrativa.

Ha qualcosa da aggiungere che non ha avuto modo di approfondire finora?

Io insisto solamente su un concetto: secondo me il concetto più importante di cui tenere conto è che la traduzione è un servizio fondamentale ed è un contributo a un processo che è il processo editoriale, cioè il traduttore sta lì dentro. Cioè è una macchina, un ingranaggio dentro un processo. Quindi l'impatto

dell'intelligenza artificiale dipenderà solamente da quanto pesa questo ingranaggio all'interno di quel processo e soprattutto a chi quel processo è destinato, perché se si crea un altro tipo di distinzione e si dice “questa è letteratura e questa no”, questo è un tipo di distinzione che può valere oggi e può non valere domani. Invece dentro un processo, semplicemente, se il tuo contributo deve essere preservare la good enough quality, stabilire cos'è good enough quality dentro un processo editoriale ha tutto a che fare col processo stesso e quindi con chi prende la decisione finale su a chi mi rivolgo. Quindi, per esempio, tradurre un manuale e tradurlo per Il Mulino piuttosto che per Apogeo piuttosto che per Laterza già cambia: nel senso che il tipo di soggetto editoriale che sceglie un certo tipo di pubblicazione, quindi innesca un certo tipo di processo dentro il quale c'è anche la traduzione, è veramente il soggetto che decide.

Tu come traduttore sei dentro quel processo, quindi puoi pensare che dentro quel processo anche un manuale di diritto può essere qualunque cosa, può anche essere scritto da un grande maestro di stile – ci sono tantissimi autori, per esempio, di diritto che sono ormai considerati grandi scrittori tout court, cioè ci sono dei testi che un tempo valevano dentro un certo tipo di ambito che teoricamente potrebbe essere coperto dall'intelligenza artificiale e oggi sono diventati dei classici. Se diventano dei classici verranno semplicemente e probabilmente ritradotti, perché poi succede anche questo: cioè che un libro ha una vita lunga, che magari nasce come strumento per studenti e diventa letteratura e se diventa letteratura è molto probabile che quella traduzione fatta dall'intelligenza artificiale non sia considerata più sufficiente, che qualche editore si riprenda quel libro, che abbia un traduttore e dica “adesso me lo traduci e me lo fai diventare letteratura”.

Quindi bisogna proprio capire che senza la consapevolezza del processo editoriale dentro il quale si colloca la traduzione, tracciare un distinguo tra IA e traduzione umana è arbitrario.

Intervista a Sara Bilotti

Può descrivermi brevemente il suo percorso come traduttore e qual è, ad oggi, il suo rapporto con strumenti di traduzione automatica o LA generativa?

Ho cominciato nel 2018, su richiesta dell'editor di straniera di una casa editrice indipendente: conosceva il mio percorso di studi all'estero e l'esperienza ventennale accanto a ragazzi vittime di abuso familiare. Tale esperienza era fondamentale per l'interpretazione e la fedeltà nella traduzione di un celebre romanzo americano che racconta, appunto, di un abuso familiare. Il lavoro mi ha appassionata talmente tanto da dare alla professione di traduttrice priorità rispetto al mio ruolo di scrittrice. Da allora, ho tradotto autori come Joe R. Lansdale, Joyce Carol Oates, Quentin Tarantino e diversi altri.

Il mio rapporto con gli strumenti di traduzione automatica non esiste. Nel senso che non li ho mai utilizzati. Per quanto riguarda l'IA, invece, il discorso è complesso ma sintetizzo la mia risposta: non la uso per tradurre perché ritengo che la traduzione letteraria sia un lavoro artistico, e in quanto tale non possa essere sostituito da una macchina. La utilizzo però per il lavoro di ricerca, e con grande risparmio di tempo. Le faccio un esempio: sono anni che non vado negli Stati Uniti, e non conosco alcuni slang contemporanei. Chiedo all'intelligenza artificiale di trovarmi contesti in cui possa leggere i vari usi dello slang in questione - forum, siti, discussioni -, in modo da confrontarli e comprenderne a fondo il significato. Tale lavoro di ricerca, prima dell'IA, avrebbe richiesto ore, adesso mi prende qualche minuto.

Quando legge una traduzione prodotta con il supporto dell'IA, riesce a riconoscerla? Su quali elementi si basa per farlo?

Se la traduzione non ha subito alcuna revisione da parte di un editor "umano", la riconosco immediatamente. L'IA utilizza degli schemi ricorrenti di ragionamento e scrittura, non è difficile riconoscerli. Inoltre, compie diversi errori di senso. Ricordo ancora la traduzione di un libro (purtroppo pubblicato) in cui invece di tradurre

“cittadina in fiamme” c’era scritto “villaggio infiammato”. Ho subito pensato che forse aveva bisogno di una Tachipirina. Scherzi a parte, per fortuna ultimamente non incappo più in tali orrori. Spero sia perché le case editrici si sono rese conto dell’assurdità di lasciare tutto il lavoro all’IA, e non perché sono stata semplicemente fortunata.

Inoltre, l’IA non riconosce il suono della voce di un autore. Un romanzo è come uno spartito: se non senti la musica della sua voce, non puoi riprodurla neanche se sai leggere le note.

Ritiene che lavorare a partire da un testo pre-tradotto automaticamente possa condizionare stilisticamente il processo di traduzione rispetto alla traduzione diretta dal testo originale?

Sì. O, almeno, io ne sarei condizionata, probabilmente anche a livello inconscio. Ho bisogno di leggere lo spartito di cui parlavo prima, di farlo “suonare”, leggo infatti spesso ad alta voce il testo originale prima di cominciare a tradurre. È come un incantesimo che impari con l’esperienza. Non credo che qualcosa di tradotto automaticamente possa cogliere certe sfumature di ritmo e senso.

Quali rischi potrebbe comportare l’utilizzo di LA generativa o di strumenti di traduzione automatica se applicati alla traduzione letteraria?

Rischiamo di ritrovarci con romanzi tutti uguali. Le storie sono già state tutte scritte, ma il modo in cui le raccontiamo continua ancora a fare la differenza.

Detto questo, credo che l’iniziale entusiasmo cieco nei confronti dell’utilizzo dell’IA in questo campo sia scemato. O meglio, si sia trasformato in una sana curiosità e nella comprensione di quanto può migliorarci la vita se la usiamo come supporto, evitando che si sostituisca completamente a noi.

La traduzione letteraria implica un lavoro di interpretazione, di conseguenza ogni professionista traduce secondo una sua personale visione dell’opera originale, naturalmente cercando

di mantenere intatto lo stesso effetto a cui mirava l'originale. Pensa che l'introduzione di sistemi di IA nel processo traduttivo possa in qualche modo portare progressivamente alla perdita dello stile personale del traduttore?

Dipende da come si usa l'IA. Come ho detto nelle risposte precedenti, se l'IA è al nostro servizio non può che migliorare l'esperienza, perché ci consente di fare ricerche in modo veloce e mirato, permettendoci di concentrarci esclusivamente sulla voce e gli intenti dell'originale. Utilizzarla in modo cieco e irresponsabile significa innanzitutto consegnare ai lettori un lavoro mediocre, ma anche perdere quell'allenamento all'interpretazione che ci consente di dare giustizia a un romanzo.

Nel dibattito sulla traduzione automatica nel campo delle traduzioni specialistiche si parla spesso di good enough Quality, ovvero una qualità sufficiente a trasmettere correttamente e comprensibilmente il contenuto di un testo specialistico. Ritiene che nel caso di una traduzione letteraria questo compromesso sia accettabile? O la soglia della "qualità accettabile" non esiste per definizione quando si tratta di letteratura?

La "qualità accettabile" di cui parla può andar bene per i testi che io chiamo "di servizio": manuali, per esempio, o alcuni tipi di saggi. Non tutti. Tale compromesso tiene conto solo di un paio di dimensioni: manca la dimensione intima, e tutto ciò che c'è tra le righe, e che solo una particolare costruzione della frase può suggerire. Un errore voluto, per esempio, può essere riconosciuto e risolto dall'IA, facendo perdere alla frase il suo senso più profondo. L'unica "qualità accettabile", nella traduzione letteraria, è quella altissima.

Se dovesse immaginare il traduttore professionista tra cinque anni, come descriverebbe il suo rapporto con questi strumenti e cosa, secondo lei, resterà umano in questo lavoro?

Non posso prevedere il futuro, ma posso dirle cosa vorrei vedere: un traduttore con qualche ora libera in più - il lavoro di traduzione richiede un impegno

di ore notevole - grazie al tempo risparmiato nella ricerca, un traduttore più sereno, che potrà impegnarsi nell'attività più importante che richiede il suo ruolo: l'interpretazione di una voce, nel suo significato più profondo.

Spero che di umano resti tutto, tranne le parti più noiose e meccaniche.

Ha qualcosa da aggiungere che non ha avuto modo di dire finora?

No, credo sia tutto ciò che ho da dire sull'uso dell'IA nel campo della traduzione letteraria. Mettiamola al nostro servizio, perché se ci sostituirà non faremo altro che assecondare un appiattimento purtroppo già molto evidente nelle arti; un livellamento che, a lungo andare, avrà effetti devastanti sulla nostra intelligenza, anche quella emotiva.