BET*WIXT*

Studies in Linguistics and Communication

23

MARIA PIA DI BUONO

GIORNALISMO ALGORITMICO E TRADUZIONE AUTOMATICA

Una valutazione della traduzione neurale

PAOLO LOFFREDO

On the cover:
Immagine realizzata da Bruna Troise

Finito di stampare nel mese di maggio 2023

ISBN 979-12-81068-15-5

ISSN 2611-1349 (collana)

Via Ugo Palermo, 6 - 80128 Napoli
www.loffredoeditore.com
paololoffredoeditore@gmail.com

INDICE

Capitolo Quattro
La Qualità della Traduzione Neurale

Elenco di Figure e Tabelle

Figure

Tabelle

A M.S.
A S.

Introduzione

Il giornalismo a livello globale sta vivendo una fase di transizione storica grazie ai rapidi progressi delle tecnologie digitali (Ali & Hassoun 2019) che determinano importanti trasformazioni nelle organizzazioni aziendali e nelle funzioni dei *media*.

Le tradizionali aziende dei *media* si stanno confrontando con molte sfide derivanti dalla radicale trasformazione digitale dell'industria editoriale (Leppänen *et al.* 2018), che porta alla ricerca di nuove soluzioni in seguito ai cambiamenti imposti al mercato delle notizie (Marconi & Siegman 2017). La digitalizzazione e la diffusione di sistemi intelligenti, infatti, rimodellano radicalmente le agenzie di stampa (Galily 2018), soprattutto per quanto riguarda gli aspetti della produzione e diffusione delle notizie (Ali & Hassoun 2019).

L'intelligenza artificiale (IA) si sta diffondendo in tutte le sfere multicreative, incluso l'ambito giornalistico (Ali & Hassoun 2019), che è stato già scenario di cambiamenti, principalmente alla luce della persistente perturbazione economica e della trasformazione digitale (Newman *et al.* 2018). L'IA si sta facendo strada trasversalmente sia nel processo di produzione delle notizie sia nella struttura e funzionamento del sistema dei *media* (Túñez-López *et al.* 2021).

Gli algoritmi di IA, considerati la rivoluzione più importante del giornalismo nell'era digitale, non solo sono parte integrante dell'ecosistema dei nuovi *media*, ma modificano il processo di selezione, produzione, diffusione e fruizione delle notizie, rimettendo nuovamente in discussione il quadro teorico dei valori delle notizie, originariamente proposto da Galtung & Ruge (1965), nonché il processo giornalistico.

D'altra parte, queste tecnologie presentano un grande potenziale per migliorare il giornalismo odierno e cambiarne le pratiche – ad esempio consentendo ai giornalisti di processare velocemente alti volumi di dati, creare notizie da dati strutturati e consegnarle automaticamente, garantire una copertura diversificata (Ali & Hassoun 2019).

Il tema è al centro dell'interesse di diverse entità collegate al settore privato e pubblico, come l'*European Broadcasting Union*[1] (EBU) che ha dato

[1] https://www.ebu.ch/home Ultimo accesso: 28/02/2023.

vita all'iniziativa *Artificial Intelligence and data*[2], con la finalità di aiutare i *media* del servizio pubblico a sfruttare le potenzialità dell'IA, considerando questi dei temi centrali, in particolar modo per rafforzare e personalizzare la relazione con i cittadini. Anche gli studi annuali della Reuters e dell'Università di Oxford negli ultimi anni hanno incluso l'IA e le tecnologie di trascrizione, traduzione automatica e conversione audio/testo e testo/audio tra gli elementi in grado di avere un impatto nel settore (Newman 2020). Nel 2022, i dati raccolti dalla Reuters confermano che i *media* continuano a scommettere sull'IA per aumentare l'efficienza della produzione e offrire servizi con una personalizzazione migliore, perché "*next generation technologies like artificial intelligence (AI), cryptocurrencies, and the metaverse (virtual or semi-virtual worlds) are already creating a new set of challenges for societies as well as new opportunities to connect, inform, and entertain*" (Newman 2022: 5).

Il rapporto tra IA e giornalismo comporta nuove sfide e delinea i confini di un nuovo settore accademico, che sta producendo diverse e interessanti linee di indagine (Parratt-Fernández *et al.* 2021). Le ricerche in ambito di *media* e comunicazione stanno mettendo in risalto l'impatto che l'impiego dell'IA ha sul settore, in termini di processo e prodotto, e su coloro che operano al suo interno, in particolar modo i giornalisti. Mentre le ricerche in IA, e più in particolare nel settore del trattamento automatico del linguaggio (TAL), conducono spesso esperimenti nel dominio delle notizie, che si presta, per le sue caratteristiche intrinseche ed estrinseche, allo sviluppo di sistemi specializzati e al raffinamento dei risultati ottenuti nei relativi compiti di TAL, come ad esempio nella generazione di linguaggio naturale, inclusa la traduzione automatica.

Tuttavia, l'approccio computazionale alle notizie manca di un *framework* teorico integrato per analizzare e comprendere il *giornalismo algoritmico* (Anderson 2013) o *computazionale* (Thurman 2019b), che risulta ancora un prodotto incompiuto, nato da un'adozione opportunistica, caso per caso, di tecnologie che hanno origine in altri campi, ma che sono in grado di fornire un ampio e interessante set di strumenti per il giornalismo (Caswell 2019). L'urgenza di includere queste tecnologie nel maggior numero possibile di contesti ha portato a una progettazione scadente che spesso non è ottimizzata e che non funziona come promesso (Broussard 2018). Secondo Túñez-López *et al.* (2021), il punto di convergenza di Caswell e Broussard ha a che

[2] https://www.ebu.ch/aidi Ultimo accesso: 28/02/2023.

fare con la coerenza e lo scopo specifico. La possibilità di perseguire con successo questi due aspetti è, a mio avviso, intrecciata alla comprensione del processo giornalistico nello scenario algoritmico e alla definizione del contributo che la traduzione automatica (TA) può dare.

Infatti, come la traduzione (umana) ha da sempre rivestito un ruolo importante in diverse fasi della produzione e disseminazione delle notizie straniere (Zanettin 2021), rappresentando la base per l'informazione o la disinformazione (Valdeón 1995), anche la traduzione automatica risulta fondamentale in tutte le fasi del processo giornalistico, inclusa la fruizione delle notizie. I giornalisti usano la traduzione automatica durante la fase di selezione e verifica delle fonti e dei contenuti, mentre gli utenti accedono alle informazioni globali soprattutto attraverso i sistemi di traduzione automatica delle piattaforme *social*. La caratteristica di *invisibilità della traduzione* nel giornalismo (Zanettin 2021), come vedremo, si amplifica con l'uso delle tecnologie e delle piattaforme *social*.

Fino a qualche anno fa, infatti, i sistemi di traduzione automatica producevano risultati ancora lontani dalla qualità della traduzione umana, riconoscibili come testo non nativo o non tradotto manualmente. Gli attuali sistemi hanno raggiunto, per alcune coppie di lingue e alcuni domini specifici[3], livelli qualitativi di fluidità tali da far pensare che il testo tradotto automaticamente possa essere nativo o frutto di una produzione umana. Inoltre, l'integrazione di sistemi di traduzione automatica nelle piattaforme *social* rende possibile agli utenti l'accesso immediato a traduzioni di notizie provenienti da tutto il mondo, spesso superando l'intermediazione dei *gatekeeper* tradizionali.

All'invisibilità della traduzione si affianca sia quella che potremmo definire l'*invisibilità del processo*, cioè la mancanza di *trasparenza algoritmica* (*algorithmic transparency*), di *spiegabilità* (*explainability*) (Leiser 2022) e *accessibilità* (*accessibility*) degli algoritmi e dei dati utilizzati per la produzione dei contenuti (Buhmann & Fieseler 2021), sia la non trasparenza delle agenzie di stampa che non sempre dichiarano che i contenuti sono stati generati automaticamente.

Mentre la tecnologia avanza e tende ad occuparsi di un numero crescente di processi nella creazione di notizie, è importante fornire strumenti efficaci che aiutino a creare un giornalismo di qualità, che consenta, inoltre, di

[3] Come vedremo, nel 2018 è stato dichiarato il raggiungimento della *parità umana* (*human parity*) per la coppia cinese→inglese proprio nel dominio della traduzione automatica delle notizie (Hassan *et al.* 2018).

ridurre i due fenomeni di invisibilità, in considerazione di un'altra grande sfida che entra in gioco nello scenario algoritmico: le implicazioni etiche dell'IA (Dörr & Hollnbucher 2017), in termini di deontologia professionale e responsabilità.

Il presente volume intende offrire, quindi, uno spunto di riflessione attraverso l'analisi dello scenario attuale nell'ambito del trattamento automatico delle notizie, con una particolare enfasi sulla traduzione automatica di questa tipologia di testi. L'obiettivo è quello di delineare un'area di convergenza tra traduttologia, linguistica, giornalismo e TAL nella quale discutere le possibilità e i limiti delle tecnologie del linguaggio e riflettere sull'apporto delle diverse discipline allo sviluppo di sistemi che siano supportati da studi e approcci tradizionalmente legati ai saperi umanistici.

Nel primo capitolo viene fornita un'introduzione al concetto di giornalismo algoritmico e discussi i cambiamenti che l'introduzione delle tecnologie di IA in tale ambito apporta al processo e al prodotto giornalistico. Il discorso giornalistico è qui inteso in maniera ampia come un insieme di dinamiche e processi propri delle agenzie di stampa e di professionisti del settore che hanno una diretta influenza sul testo giornalistico e su come questo viene prodotto e poi presentato. Queste componenti del discorso giornalistico sono assolutamente rilevanti ai fini di comprendere gli aspetti che i *Translation Studies* affrontano quando si confrontano con la traduzione giornalistica e di evidenziare le caratteristiche linguistiche, testuali e anche culturali, che i sistemi di IA basati su approcci di TAL, e più nello specifico di traduzione automatica, devono affrontare.

La relazione tra computazione e giornalismo viene analizzata nel secondo capitolo presentando le applicazioni computazionali del TAL che possono supportare le diverse fasi del processo giornalistico, cioè la selezione, la creazione e la diffusione delle notizie. La ricerca che si occupa TAL ha mostrato un profondo interesse per il dominio giornalistico, sviluppando e testando numerosi sistemi sui testi delle notizie: a partire dalle tecniche di estrazione di informazioni e dalle soluzioni per la verifica delle notizie e delle fonti, a cui i giornalisti possono fare ricorso nella fase di selezione delle notizie da pubblicare, fino ad arrivare alle analisi, come quella del sentimento, che permettono di ottimizzare la distribuzione delle notizie secondo gli interessi degli utenti.

Nel terzo capitolo, il *focus* si sposta sulla traduzione automatica delle notizie, una tecnologia fondamentale nel giornalismo globale. A partire dal contributo dei *Translation Studies*, la traduzione delle notizie assume caratteristiche peculiari tanto da rappresentare un'area specifica di studi teorici

ma anche di applicazione dei sistemi di traduzione automatica tradizionali e di quelli più attuali basati sulle reti neurali.

Le dichiarazioni del raggiungimento della parità umana (*human parity*) (Hassan *et al.* 2018) nella traduzione automatica delle notizie per la coppia di lingue cinese→inglese e della prestazione *super-human* (*super-human performance*) per la coppia inglese→ceco (Bojar *et al.* 2018) hanno fatto sì che si moltiplicassero le sperimentazioni per testare i sistemi più recenti e si ridiscutessero le metriche utilizzate per valutare la qualità dei risultati della TA.

Il cambio di paradigma avvenuto nell'ambito della TA, grazie allo sviluppo di sistemi e modelli del linguaggio neurali[4] in grado di ottenere dei risultati insperati fino a qualche anno fa, viene indagato attraverso una sperimentazione per la valutazione della TA per la coppia inglese→italiano nel quarto capitolo.

Infine, vengono presentate delle brevi riflessioni su come giornalisti ma anche linguisti e traduttologi possano contribuire allo sviluppo di tecnologie traduttive e non solo che stanno riconfigurando il giornalismo e il suo rapporto con la computazione.

Per concludere, vorrei ringraziare le persone che hanno reso possibile la stesura di questo volume, supportandomi e incoraggiandomi in diversi modi: in primo luogo la mia famiglia per il sostegno incondizionato; Johanna Monti, in qualità di amica e riferimento professionale, per la revisione della prima bozza del volume, ma, soprattutto, per le preziose indicazioni e le numerose opportunità di confronto e crescita; Felice Addeo, per gli indispensabili suggerimenti; Giulia Speranza e Raffaele Manna che mi hanno aiutato nella revisione finale di questo volume e Bruna Troise per averne realizzato la copertina.

Ringrazio anche il direttore della collana Giuseppe Balirano e l'editore Paolo Loffredo per avermi dato la possibilità di dare vita a questo progetto editoriale.

Infine, un pensiero di stima e affetto per Bojana Dalbelo Bašić e Jan Šnajder con cui ho avuto la fortuna di lavorare presso l'Università di Zagabria e grazie ai quali ho iniziato, qualche anno fa, a occuparmi di trattamento automatico dei testi giornalistici.

[4] Nel TAL un modello del linguaggio, o modello linguistico, è una probabilità di distribuzione di una sequenza di parole. Data una qualsiasi sequenza di parole di lunghezza *n*, un modello del linguaggio assegna una probabilità $P(w_1, \ldots, w_n)$ all'intera sequenza. I modelli linguistici vengono utilizzati per una serie di compiti, come la comprensione (*natural language understanding*) e la generazione di linguaggio naturale (*natural language generation*).