
DeepL Traduttore e didattica della traduzione dall'italiano in tedesco

Alcune valutazioni preliminari

By Christine Heiss & Marcello Soffritti (Università di Bologna, Italy)

Abstract & Keywords

English:

What effects can the availability of a tool such as DeepL Translator have on the teaching of translation from Italian into German? This paper examines the most relevant factors to be taken into account, beyond simple standardised measurements. Starting from samples contained in two theses of Bachelor and Master's degree courses in translation, we explore quality levels of DeepL Translator's raw output (with some reflections on BLEU scores) and successive improvements achievable with a simple kind of post-editing (without formal check-lists). Our findings confirm that this application provides quite promising results even without domain-related adaptation. Although some residual failures to be addressed in post-editing are insidious and challenging, DeepL Translator is likely to become an invaluable tool for more advanced and efficient practices in teaching and training. Accordingly, post-editing should become an essential training activity.

Italian:

Quali effetti può avere la disponibilità di uno strumento come DeepL Translator sull'insegnamento della traduzione dall'italiano al tedesco? Questo articolo esamina i fattori più rilevanti da prendere in considerazione, al di là delle semplici misure standardizzate. Sulla base di due tesi (svolte in corsi di laurea e master in traduzione), esploriamo i livelli di qualità dell'output grezzo di DeepL Translator (con alcune riflessioni sui punteggi BLEU) e i successivi miglioramenti ottenibili con un primo intervento di post-editing (senza check-list formali). I riscontri ottenuti confermano che questa applicazione fornisce risultati piuttosto promettenti anche senza adattamenti relativi agli ambiti di specialità. Sebbene alcuni residui difetti richiedano attenti interventi di post-editing, è probabile che DeepL Translator diventi uno strumento insostituibile per pratiche più avanzate ed efficienti nell'insegnamento e nella formazione. Di conseguenza, il post-editing dovrebbe diventare un'attività di formazione essenziale.

Keywords: Machine Translation, translator training, formazione traduttori, traduzione automatica, post-editing

1. Introduzione – una risorsa efficiente e inaspettata

L'apparizione di DeepL Traduttore all'inizio di settembre 2017 ha colto di sorpresa gli operatori professionali e i docenti di traduzione. Sia gli uni che gli altri si sono trovati improvvisamente alle prese con uno strumento liberamente accessibile e molto più promettente di quanto si sarebbe ritenuto possibile fino a quel momento. Il prodotto mostrava infatti una chiara competitività rispetto ad altre applicazioni disponibili per la traduzione automatica, e in particolare rispetto a quelle finora rilasciate da Google, che almeno in teoria dispone di risorse finanziarie e quantità di dati ben superiori. Presentiamo di seguito alcune considerazioni sull'impatto di DeepL Traduttore nella didattica universitaria della traduzione, sia specializzata sia per il web. Cercheremo in particolare di chiarire con quali vantaggi DeepL Traduttore sia integrabile come componente stabile nella didattica, e in che modo si configurino corrispondenti attività di *post-editing*. Come per molti prodotti tecnologici, queste considerazioni sono provvisorie, e potrebbero presto rivelarsi inadeguate, ad esempio in seguito a reazioni della concorrenza, a ulteriori riscontri sistematici e dettagliati sulle prestazioni dell'applicazione in ambito professionale o con determinate combinazioni linguistiche, a imprevisti nel percorso commerciale del prodotto e molto altro. Tutti i riscontri qui discussi rispecchiano il comportamento dell'applicazione al 31 gennaio 2018.

2. Le caratteristiche principali di DeepL Traduttore

Non è solo il livello di competitività che può creare perplessità e indecisione. Fra i ricercatori impegnati nello sviluppo di applicazioni di traduzione automatica (MT) e nelle relative verifiche di qualità, la curiosità sulle caratteristiche della nuova applicazione rimane ancora insoddisfatta. I produttori hanno esplicitamente comunicato che per il momento non daranno alcuna informazione specifica sull'architettura della rete neurale utilizzata[1]. Allo stesso tempo riportano tabelle comparative per illustrare la superiorità di DeepL Traduttore rispetto a 5 applicazioni precedentemente disponibili. Vengono svolti due tipi di confronto. Il primo riguarda 6 combinazioni linguistiche (3 dall'inglese e 3 verso l'inglese), ed è presentato come segue:

sono state tradotte 100 frasi da DeepL Traduttore e da Google, Microsoft e Facebook. Alcuni traduttori professionisti hanno poi valutato i risultati, senza sapere da quali sistemi provenissero le traduzioni. Le traduzioni di DeepL sono state scelte come appropriate tre volte più spesso rispetto a quelle tradotte dai sistemi della concorrenza.

Il secondo confronto viene svolto sulla base di punteggi BLEU[2] ricavati da una serie di pubblicazioni. In queste si riferiscono e discutono i risultati raggiunti dai sistemi concorrenti (<https://www.deepl.com/press.html>) "su specifici set predefiniti di test[3]", riferiti alle combinazioni fra inglese e tedesco. Questi punteggi vengono incorporati in tabelle contenenti anche i valori raggiunti da DeepL per le stesse combinazioni linguistiche. Vale la

pena riprodurre qui di seguito le tabelle (fig. 1), perché i valori BLEU sull'asse verticale sono rappresentati con notevole disinvolture. In termini numerici il distacco massimo (riscontrato sulla scala BLEU da 0 a 100) non supera i 6 punti BLEU (a fronte di un totale di 31 punti raggiunti da DeepL nel primo caso e di 44,7 circa nel secondo caso). La lunghezza delle barre colorate non è quindi per nulla proporzionale ai valori numerici. È fino a 25 volte più lunga delle altre, e suggerisce differenze ben superiori a quelle tra i valori percentuali effettivamente raggiunti:

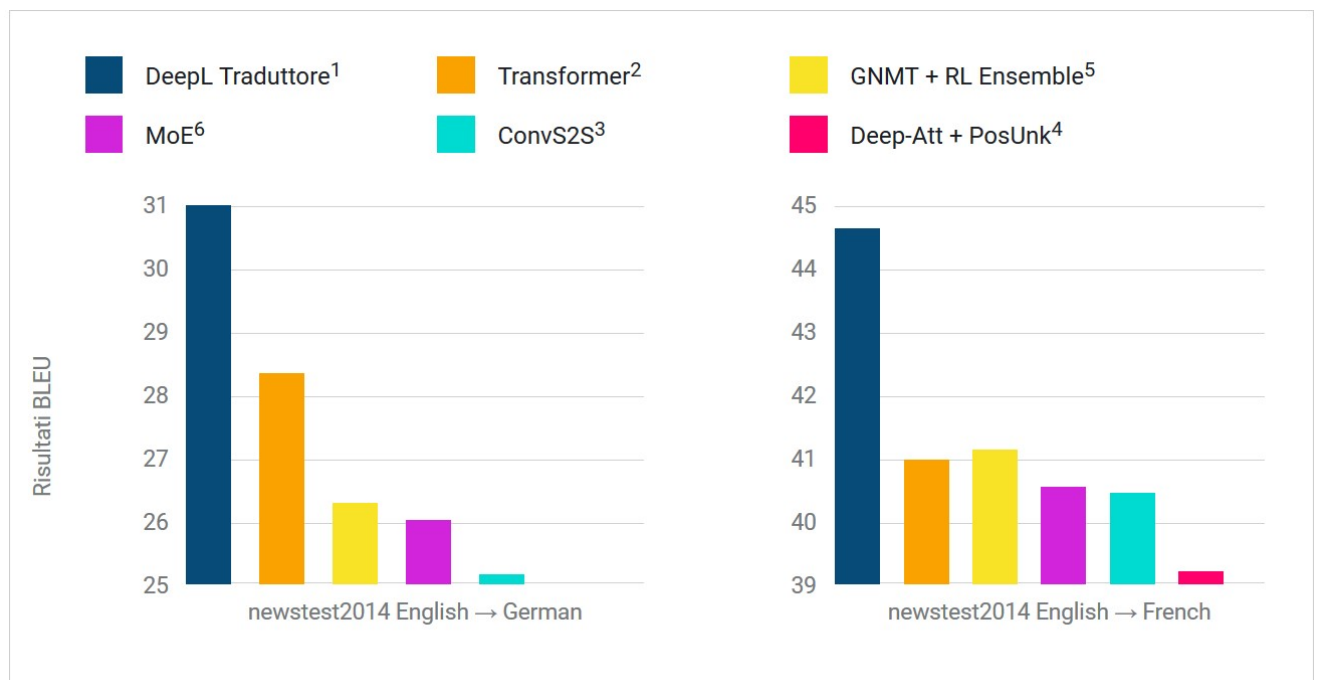


Fig. 1: Valori BLEU (fonte: <https://www.deepl.com/press.html>).

Questo tipo di visualizzazione corrisponde quindi alle finalità di una presentazione promozionale per il grande pubblico, ma non può certo soddisfare le esigenze di una comparazione scientificamente accettabile. Si può riconoscere, tuttavia, che scarti di 5 punti BLEU vengono considerati nella letteratura già molto significativi (cfr. ad es. Bentivogli *et al.*, 2016; Junczys-Dowmunt *et al.*, 2016). Dalle indicazioni riportate sulla pagina di accesso si desume poi che da un lato il sistema riesamina (in tempo reale) la qualità della traduzione fornita alla luce di eventuali modifiche apportate dall'utente al testo proposto, e dall'altro almeno una parte di queste modifiche viene utilizzata anche per riaggiustare il sistema stesso. Queste caratteristiche si accordano ovviamente con la natura neurale dell'applicazione, messa in rilievo come fattore principale dell'incremento di precisione ed efficienza. Per gli operatori commerciali e gli sviluppatori di applicazioni è disponibile un'*application programming interface* (API), che:

allows you to incorporate the DeepL Translator into your own products and services. Besides using it in your favourite CAT tools or integrating it into your intranet services, our API lends itself to many more of your in-house software applications.

Ai clienti che utilizzano l'API si assicura invece un diverso trattamento dei testi elaborati:

Your privacy and confidentiality are important to us - we don't save any texts sent to the DeepL Translator API, and we don't use them to improve the quality of the service. <https://www.deepl.com/api-contact.html>

Conseguentemente, il sistema potrà utilizzare per le dinamiche di riaggiustamento "dal basso" solo i dati provenienti da utilizzatori non paganti, vale a dire utenti non sempre qualificati a livello professionale e non sempre disposti a svolgere (e a condividere) un editing che effettivamente migliori la proposta ricevuta.

3. La misura della qualità dei sistemi NMT- criteri e metodi in uso

Facendo un piccolo passo indietro, bisogna ricordare brevemente che fino a settembre 2017 l'applicazione più usata era senz'altro Google Traduttore, reimpostata su un'architettura neurale (NMT) già dall'aprile dello stesso anno. Ad oggi, non tutte le comparazioni svolte hanno confermato pienamente la superiorità di questa architettura. Se da un lato, secondo gli esperti di MateCat (www.matecat.com), i sistemi NMT sembrano sopravanzare nella qualità tutti i sistemi di Statistical Machine Translation (SMT), tranne forse per quanto riguarda l'accuratezza nell'uso della terminologia, dall'altro pesano invece le considerazioni espresse da Castilho *et al.* (2017), che mettono in rilievo le discrepanze tra la valutazione automatica, basata sulle misure BLEU e simili, e quella ottenuta con la partecipazione di valutatori umani, basata sulla rilevazione di specifiche categorie di errori e sull'esito di diverse modalità di *post-editing*. Tenendo conto di queste considerazioni contrastanti, risulta quindi più difficile sostenere la superiorità dei sistemi NMT sugli altri, a meno che la comparsa sul mercato di DeepL non cambi significativamente la situazione. Nel complesso, le misurazioni finora pubblicate in questi ed altri studi sono tappe importanti ma ancora parziali di un procedimento collettivo di valutazione, estremamente denso di variabili e tutt'altro che concluso, su cui influiscono, insieme a interessi commerciali, numerosi condizionamenti tecnici e probabilmente anche pressioni settoriali di diversa natura. Rimangono in ogni caso da verificare, ed eventualmente da affrontare, alcune importanti inadeguatezze strutturali determinate dall'architettura stessa dei nuovi sistemi neurali:

[...] there are still problems and challenges of NMT need to be tackled: The training and decoding process is quite slow; the style of translation can be inconsistent for the same word; there exists an "out-of-vocabulary"

problem on the translation results; the “black-box” neural network mechanism leads to poor interpretability; thus the parameters for training are mostly picked based on experience. (Zhang, 2017)

Per contro, viene in genere riconosciuta una certa superiorità dei sistemi NMT nel generare segmenti d’arrivo più scorrevoli (cfr. ad es, Waswani *et al.*, 2017).

4. Alcune misure della qualità di DeepL

Tornando al tema della qualità di DeepL Traduttore, ed in particolare alle valutazioni automatiche in modalità BLEU, non risultano ancora disponibili indagini indipendenti atte a verificare se l’incremento qualitativo (fino a 5,5 punti nel confronto con la concorrenza) dichiarato dai produttori sia riproducibile in condizioni realistiche d’uso. Anche la valutazione umana indicata nella pagina web di descrizione del prodotto – citata più sopra – non è del tutto circostanziata, e non risulta che siano state svolte verifiche sperimentali indipendenti. Certamente una preferenza di tre volte superiore su un set di 100 frasi merita, se confermata, di essere presa sul serio, per quanto in questa modalità di rilevazione gli spunti impressionistici siano talvolta difficili da eliminare. Per questo discuteremo ora alcuni riscontri provvisori sui punteggi BLEU in contesti rilevanti per la traduzione dall’italiano in tedesco nella didattica universitaria, una prospettiva praticamente assente dagli esperimenti finora condotti. I tre settori tematici coinvolti (testi giuridici normativi, istruzione universitaria e istruzioni tecniche) sono conformi ai programmi e agli obiettivi formativi degli insegnamenti di traduzione specializzata dall’italiano in tedesco nel Corso di Laurea Magistrale in Specialized Translation presso il Dipartimento di Interpretazione e Traduzione dell’Università di Bologna. Trattandosi di settori tematici specializzati, la traduzione automatica trova qui le migliori prospettive di applicazione. In prospettiva didattica va tenuto presente anche il fatto che gli studenti di questo Corso di Laurea Magistrale usufruiscono già da alcuni anni di un modulo specifico di traduzione automatica e *post-editing*.

Ricordiamo che per calcolare il punteggio BLEU la misurazione della corrispondenza (nel nostro caso condotta con l’applicazione del sito Tilde Custom Machine Translation, disponibile all’indirizzo <https://www.letsmt.eu/Bleu.aspx>) avviene sulla base di segmenti allineati (in genere paragrafi o frasi), oltre che ovviamente in base alla percentuale di *n-gram* (di estensione 1-4) condivisi dai due testi nei segmenti allineati. Il punteggio è quindi direttamente proporzionale al numero delle corrispondenze lessicali/fraseologiche, e crolla se i segmenti non sono allineabili. Conseguentemente ci sono ben poche corrispondenze tra questi parametri di qualità e quelli in uso nella didattica universitaria, soprattutto se si pensa alla traduzione non tecnico-scientifica e alle situazioni in cui si devono operare rielaborazioni complesse della struttura sintattica.

Come testi ottimali di riferimento utilizziamo testi prodotti senza l’utilizzo di sistemi MT e ufficialmente approvati per l’uso[4].

4.1 Riscontro n. 1 – Traduzione giuridica

Si tratta di un campione di 9 segmenti (520 parole circa) contenuti nella Legge Provinciale 17 settembre 2013, n. 18 della Provincia Autonoma di Bolzano, tradotta dall’italiano in tedesco e pubblicata nel portale ufficiale della stessa provincia[5]. Il punteggio BLEU ottenuto da DeepL Traduttore (senza *post-editing*), calcolato con l’aiuto dell’applicazione <https://www.letsmt.eu/Bleu.aspx>, è 31,3, cioè quasi pari al valore raggiunto nel test (dall’inglese al tedesco) WMT 14 citato dai produttori in occasione del lancio del prodotto. Il punteggio raggiunto da Google Traduttore è 31,37. Consideriamo a mo’ di esempio la versione tedesca ufficiale e quella di DeepL Traduttore dell’art. 1 *quinquies*, comma 1:

ORIGINALE	TESTO PUBBLICATO	DEEPL TRADUTTORE
(1) Con frequenza bimestrale i comuni comunicano, per via telematica, alla Ripartizione provinciale Natura, paesaggio e sviluppo del territorio e all’Agenzia provinciale per l’ambiente le informazioni sui titoli abilitativi rilasciati nonché le comunicazioni di inizio lavori pervenute, con indicazione del tipo di impianto e della localizzazione.	(1) Alle zwei Monate teilen die Gemeinden der Landesabteilung Landschaft Raumentwicklung und der Landesagentur für Umwelt per Datenfernübertragung die Informationen über die erteilten Ermächtigungen mit sowie die eingegangenen Arbeitsbeginnmeldungen, mit Angabe der Art der Anlage und des Standorts.	(1) Alle zwei Monate kommunizieren die Gemeinden elektronisch an die provinzielle Verteilung der Natur-, Landschafts- und Landentwicklung und an das Landesumweltamt Informationen über die erteilten Genehmigungen und die Mitteilungen über den Beginn der Arbeiten, wobei die Art der Anlage und der Standort angegeben werden.

È evidente l’inadeguatezza della soluzione “Verteilung der Natur-, Landschafts- und Landentwicklung”, conseguente al fatto che non è stata riconosciuta la specificità semantica di *Ripartizione* come termine riferito a un ufficio pubblico dell’amministrazione provinciale (*Landesabteilung*). Per il resto le discrepanze riguardano solo la terminologia idiosincratca dell’amministrazione altoatesina, e non compromettono la comprensione dei dettagli da parte di qualsiasi lettore di lingua tedesca. La morfologia e la sintassi sono del tutto corrette, eppure la discrepanza tra il punteggio ottenuto automaticamente e una possibile valutazione didattica è rilevante. In un contesto didattico, infatti, la versione di DeepL Traduttore potrebbe essere considerata ancora sufficiente. Il punteggio BLEU riferito a questo singolo segmento, per contro, è un modesto 23,69, con una penalizzazione dovuta in gran parte al mancato utilizzo di una terminologia peculiare in uso esclusivamente in Alto Adige. Naturalmente anche questa discrepanza ha una sua rilevanza didattica, ma per il momento non è possibile tarare i sistemi di traduzione automatica verso il tedesco in base a parametri diatopici, e l’integrazione di corrispondenze terminologiche determinate dal singolo utente non è ancora consentita nella versione gratuita di DeepL Traduttore. In una progressione didattica, quindi, la traduzione così realizzata sarebbe sostanzialmente accettabile se si richiedesse di rendere genericamente comprensibile il testo a un lettore di lingua tedesca, ma senza precisare il vincolo dell’utilizzabilità per la popolazione dell’Alto Adige.

4.2 Riscontro n. 2 – Informazione istituzionale universitaria

Il secondo riscontro deriva da un campione di 58 segmenti (561 parole) tratti dalla pagina web di presentazione del Corso di Laurea in Servizio Sociale della Facoltà di scienze della Formazione della Libera Università di Bolzano (<https://www.unibz.it/it/faculties/education/bachelor-social-work/>). A norma di statuto, questa Università utilizza su base paritetica come lingue di lavoro l'italiano, l'inglese e il tedesco. Nel caso della presentazione del Corso di Studio, il testo di partenza è con ogni probabilità redatto in italiano, dovendosi basare sulla terminologia istituzionale delle università italiane. La pagina tedesca ufficiale è di impeccabile qualità linguistica e stilistica, malgrado le difficoltà che ancora permangono nella scelta (o nella creazione) di corrispondenze per i termini amministrativi italiani. Il punteggio raggiunto da DeepL Traduttore è in questo caso un modesto 22,83, molto lontano non solo da quello dichiarato in occasione del lancio del prodotto, ma anche da quello ottenuto nel riscontro precedente. Anche il punteggio ottenuto usando la versione di Google Traduttore è più alto: 25,63. Ecco un breve esempio:

ORIGINALE	VERSIONE UFFICIALE	DEEPL TRADUTTORE	GOOGLE TRADUTTORE
Abbiamo costruito un percorso triennale che affronta le principali tematiche della nostra società: migrazioni, povertà, disagio psichico, dipendenze, disabilità. Nel corso di laurea in servizio sociale, approfondirai lo studio delle loro cause e dei metodi più aggiornati per affrontarle. Ti approprierai delle conoscenze scientifiche e delle competenze per soddisfare efficacemente i bisogni dei cittadini e per poter esercitare un effetto concreto sulle questioni socio-politiche più impellenti del nostro tempo.	Wir haben einen Bachelor konzipiert, welcher diesen Anforderungen gerecht wird. Im Studium der Sozialarbeit lernen Sie die Ursachen und Zusammenhänge dieser sozialen Prozesse zu verstehen. Sie eignen sich wissenschaftlich fundierte Kompetenzen an, um sich den Bedürfnissen der Bürgerinnen und Bürger effizient widmen und auf sozialpolitische Entscheidungen einwirken zu können.	Wir haben einen dreijährigen Kurs aufgebaut, der sich mit den Hauptthemen unserer Gesellschaft beschäftigt: Migration, Armut, psychische Beschwerden, Sucht, Behinderung. Während Ihres Studiums im Sozialdienst studieren Sie die Ursachen und die modernsten Methoden, um mit ihnen umzugehen. Sie werden wissenschaftliche Kenntnisse und Fertigkeiten erwerben, um den Bedürfnissen der Bürgerinnen und Bürger gerecht zu werden und konkrete Auswirkungen auf die drängendsten gesellschaftspolitischen Fragen unserer Zeit haben zu können.	Wir haben einen dreijährigen Kurs aufgebaut, der die Hauptthemen unserer Gesellschaft behandelt: Migration, Armut, Geisteskrankheit, Süchte, Behinderung. Im Studiengang Soziales vertiefen Sie das Studium ihrer Ursachen und die aktuellsten Methoden, um mit ihnen umzugehen. Sie werden wissenschaftliche Kenntnisse und Fähigkeiten in die Hand nehmen, um die Bedürfnisse der Bürger effektiv zu erfüllen und konkrete Auswirkungen auf die drängendsten gesellschaftspolitischen Fragen unserer Zeit zu haben.

Questo quadro suggerisce alcune ulteriori osservazioni. La versione ufficiale è chiaramente 'condensata' a livello testuale, con alcuni allontanamenti rispetto alla microstruttura proposizionale del testo di partenza: un intervento di raffinamento stilistico, che può variare in base alla strategia e alla sensibilità di chi traduce. In ogni caso, è stata modificata in maniera sensibile la lunghezza e la struttura dei segmenti in cui le versioni tedesche vengono divise nella fase di calcolo del punteggio BLEU. Le due versioni prodotte automaticamente si presentano senza imperfezioni morfologiche o sintattiche, mentre lasciano a desiderare nell'ultima frase finale (Il termine *Auswirkungen* non è adeguato in riferimento ad un soggetto semantico "umano"). Malgrado il punteggio BLEU parli diversamente, due soluzioni di DeepL Traduttore sono decisamente migliori di quelle di Google Traduttore: "der Bürgerinnen und Bürger" (conforme a norme di correttezza di genere) e la collocazione *Fertigkeiten erwerben*. Si ripresenta qui una situazione analoga a quella discussa precedentemente: a un inferiore punteggio BLEU non corrisponde sempre un calo assoluto del livello di correttezza o di adeguatezza.

4.3 Riscontro n. 3 – Istruzioni tecniche

Per questo riscontro è stato utilizzato un campione di 56 segmenti (1022 parole) tratti da un manuale di istruzioni per il montaggio e l'uso di un apparecchio per la produzione di acqua calda industriale. Il manuale è stato tradotto dall'italiano in tedesco con l'aiuto di una *translation memory*, ripetutamente revisionato e verificato anche dal punto di vista terminologico, e presentato nell'ambito di una tesi di Laurea Magistrale (Pagin, 2018). Questa versione è stata utilizzata come riferimento per le misurazioni. Il punteggio BLEU ottenuto da una versione grezza prodotta da DeepL Traduttore è 30,29, molto vicino a quello dichiarato per la traduzione dall'inglese al tedesco nel test WMT 14 in occasione del lancio del prodotto. Il corrispondente punteggio ottenuto da Google Traduttore è 31,55. Nell'ambito di questa tesi di laurea l'output di DeepL Traduttore è stato anche sottoposto a *post-editing*. Il punteggio BLEU ottenuto dalla versione post-editata è 43,31. Riportiamo di seguito un breve campione:

ORIGINALE	DEEPL TRADUTTORE	DEEPL TRADUTTORE RIVISTO
9. DESCRIZIONE FUNZIONALE COSTRUTTIVA	9. FUNKTIONALE UND KONSTRUKTIVE BESCHREIBUNG	9. FUNKTIONALE UND KONSTRUKTIVE BESCHREIBUNG

<p>La funzione di questo apparecchio è di consentire lo scambio di calore, tra i prodotti della combustione del bruciatore premix e l'acqua presente nel serbatoio ad accumulo, attraverso uno scambiatore di calore a contatto con il bruciatore.</p> <p>La combustione avviene in maniera completamente stagna rispetto all'ambiente che contiene l'apparecchio, prelevando l'aria necessaria alla combustione dall'esterno e scaricando i prodotti della combustione stessa sempre all'esterno. La camera di combustione stagna, è posta nella parte frontale dell'apparecchio, all'interno dell'involucro. Nella parte inferiore invece è posto il sifone, il quale garantisce lo scarico condensa.</p>	<p>Die Funktion dieses Geräts besteht darin, den Wärmeaustausch zwischen den Verbrennungsprodukten des Vormischbrenners und dem Wasser im Speicher über einen Wärmetauscher in Kontakt mit dem Brenner zu ermöglichen.</p> <p>Die Verbrennung erfolgt vollkommen luftdicht in Bezug auf den Raum, in dem sich das Gerät befindet, wobei die für die Verbrennung erforderliche Luft von außen zugeführt wird und die Verbrennungsprodukte selbst nach außen abgegeben werden.</p> <p>Die Brennkammer ist wasserdicht und befindet sich an der Vorderseite des Gerätes, im Inneren des Gehäuses. Andererseits befindet sich der Siphon, Geruchsverschluss am Boden, der den Kondensatabfluss gewährleistet.</p>	<p>Die Funktion dieses Geräts besteht darin, den Wärmeaustausch zwischen den Verbrennungsprodukten des Vormischbrenners und dem Wasser im Speicher über einen Wärmetauscher zu ermöglichen, der in Kontakt mit dem Brenner steht.</p> <p>Die Verbrennung erfolgt vollkommen abgedichtet vom Raum, in dem sich das Gerät befindet, wobei die für die Verbrennung erforderliche Luft von außen zugeführt wird und die Verbrennungsprodukte selbst nach außen abgegeben werden.</p> <p>Die Brennkammer ist abgedichtet und befindet sich an der Vorderseite des Gerätes, im Inneren des Gehäuses. Im unteren Teil befindet sich der Siphon, der den Kondensatabfluss gewährleistet.</p>
---	---	--

Il *post-editing* sull'output di DeepL Traduttore, condotto peraltro senza rigorose istruzioni formali, determina quindi per questo campione un incremento di circa 13 punti BLEU. Lo sforzo richiesto dal *post-editing* è stato significativo, e si è concentrato in particolare su diverse lacune terminologiche (in questo brano la sostituzione di *Geruchsverschluss* con *Siphon*) da colmare con ricerche ad hoc. L'incremento di 13 punti ottenuto nelle condizioni specifiche di questa prova sembra indicare che i sistemi NMT richiedono ancora addestramenti relativi al dominio da affrontare, malgrado raggiungano già buone prestazioni di base. La qualità del testo emendato e regolarizzato si percepisce come discreta, ma il divario stilistico rispetto alla versione di riferimento rimane avvertibile. In complesso, la qualità dell'output di DeepL Traduttore nella traduzione dall'italiano in tedesco sembra aumentare solo moderatamente se si passa da un testo di legge a un manuale di istruzioni.

5. Post-editing e verifiche della qualità in ambito didattico

Malgrado la relativa frammentarietà dei riscontri ottenuti su campioni di testi specialistici, è abbastanza chiaro che i punteggi BLEU dichiarati dai produttori di DeepL Traduttore al momento del lancio del prodotto non trovano sempre conferma. È anche da notare che Google Traduttore sembra avere colmato (almeno nella combinazione linguistica dei nostri esempi) il distacco indicato dalle tabelle dell'autunno 2017. Va ribadito tuttavia che le verifiche comunemente in uso negli studi sulla traduzione automatica illuminano solo alcuni degli aspetti rilevanti. Queste verifiche impongono infatti di mantenere il confronto entro un quadro di corrispondenze fra singoli segmenti allineati di due versioni, di cui una è prodotta dal sistema MT e l'altra funge da standard di riferimento[6]. Verificare la qualità in termini più generali e inclusivi dell'attività didattica - soprattutto nei casi in cui non sia usata una memoria di traduzione - comporta invece strategie diverse: non sempre è sensato attenersi strettamente ad una corrispondenza tra frasi e segmenti, e inoltre non sempre è sensato o possibile individuare un'unica versione "perfetta" da usare come modello. Qualora lo si faccia, eventuali deviazioni rispetto al modello non significano automaticamente una perdita di qualità, pur determinando una diminuzione del punteggio BLEU.

Fatta salva questa prospettiva, si può certamente riconoscere che nel passaggio dall'italiano al tedesco DeepL Traduttore, Google Traduttore (e probabilmente anche le altre applicazioni NMT) hanno risolto in gran parte i problemi di gestione della morfologia e della sintassi del tedesco, e sembrano avere fatto notevoli progressi nel governare collocazioni, fraseologismi e stilemi di uso comune. Come già accennato, ne deriva in genere un effetto di sostanziale scorrevolezza e naturalezza, che può influenzare favorevolmente il giudizio di un utente che si concentri particolarmente sull'accettabilità del testo d'arrivo, distraendolo forse da altri difetti: "Even though the neural model demonstrates gains in fluency, it also shows a greater number of errors of omission, addition and mistranslation." (Castilho *et al.*, 2017: 118). Ma altrettanto importante è un'altra conseguenza: un testo grammaticalmente corretto e accattivante suscita molto probabilmente nell'operatore umano una maggiore propensione a percorrere la strada del *post-editing*, piuttosto che rigettarlo in toto - sebbene di fatto rimangano diversi interventi da compiere. Prima di passare all'esame di ulteriori casi specifici, è opportuno esaminare alcuni dei principali presupposti che riguardano questa fase del processo.

5.1 Criteri per il post-editing a confronto

Il *post-editing* è una componente essenziale sia nel processo di raffinamento, sia in quello di valutazione della qualità del prodotto. In questa seconda funzione esso serve soprattutto a generare dati per una valutazione più differenziata dell'*output* MT. La ricerca recente ha introdotto per questo tipo di *post-editing* numerosi pacchetti di criteri, ha proposto diversissime varianti operative, e ha individuato svariate priorità riferibili a diversi livelli della valutazione da compiere, con divergenze significative e incertezze non ancora risolte in maniera soddisfacente (per una discussione approfondita cfr. ad es. Flanagan & Paulsen Christensen, 2014). Tralasciamo,

in questa sede, le opzioni di *post-editing* incluse direttamente nell'interfaccia delle applicazioni, e limitiamo la discussione ad alcuni aspetti del *post-editing* condotto in una tappa successiva e separata. Sia in ambito professionale, sia nella didattica della traduzione specializzata la revisione del testo prodotto da un programma di traduzione automatica viene condotta sulla base di criteri predeterminati, a loro volta correlati o ad uno standard "good enough" o ad uno standard "publishable". Flanagan & Paulsen Christensen (2014: 261-270) discutono in una serie di sperimentazioni l'adeguatezza delle tabelle elaborate dall'organizzazione TAUS per i casi in cui si miri ad una qualità rispettivamente 'sufficiente' e 'professionale', e per quest'ultima elaborano un modello perfezionato[7]. Ecco una sinossi dei modelli TAUS:

Good enough	Publishable
Aim for semantically correct translation	Aim for grammatically, syntactically and semantically correct translation
Ensure that no information has been accidentally added or omitted	Ensure that no information has been accidentally added or omitted
Edit any inappropriate or unacceptable content	Edit any offensive, inappropriate or culturally unacceptable content
Use as much of the raw MT output as possible	Use as much of the raw MT output as possible
Basic rules regarding spelling apply	Basic rules regarding spelling, punctuation and hyphenation apply
No need to implement corrections that are of a stylistic nature only	Ensure that formatting is correct
No need to restructure sentences solely to improve the natural flow of the text	Ensure that key terminology is correctly translated and that untranslated terms belong to the client's list of 'Do Not Translate' terms

Ed ecco il modello professionale perfezionato:

G1	Use as much of the raw MT output as possible, but:
G2	Ensure the target reader perfectly understands the content of the target text
G3	Ensure the target text communicates the same meaning and message as the source text
G4	Ensure target-text language is appropriate, but do not restructure sentences solely to improve the flow of the target text
G5	Aim for a grammatically, syntactically and semantically correct target text
G6	Ensure key terminology is correctly translated (considering context and the client's list of preferred terminology, if available)
G7	Ensure any untranslated terms belong to the client's list of 'Do Not Translate' terms, if available
G8	Edit any offensive, inappropriate or culturally unacceptable content for the target reader
G9	Apply basic rules regarding spelling and punctuation
G10	Ensure the same ST tags are present and in the correct positions in the target text

Le priorità tra i criteri e il livello di dettaglio sono ancora difficili da gestire in maniera concorde fra i soggetti che propongono e valutano le liste. Per un confronto vale la pena di considerare anche la seguente tabella, in cui le categorie fondamentali sono più o meno le stesse, anche se combinate in maniera diversa:

Grammar	Any issues which affect language quality (e.g. morphology, word order, concordance etc.)
Mistranslation	The translation does not carry the same meaning as the source sentence.
Omission	Information from the original sentence is missing in the translation.
Spelling	Spelling or typographical errors.
Style	Linguistic issues which make the sentence sound awkward in the target language.
Terminology	Translation is correct, but the terminology is not adequate for the context.
None	No errors detected.

(da: Cattelan, 2017)

Queste tabelle globali rimangono tuttavia ancora generiche. Nella prassi professionale e nella didattica quotidiana, invece, si opera necessariamente con ulteriori differenziazioni, che dipendono anche dalla combinazione linguistica e dalle priorità connesse con specifiche forme testuali o comunicative. Per l'utilizzo concreto dei singoli criteri, ad esempio il G5 della seconda tabella, è più difficile elaborare formule condivise.

6. La dimensione didattica

Per la prassi dell'insegnamento la disponibilità di applicazioni MT sempre più raffinate impone di verificare regolarmente fino a che punto sia necessario o opportuno servirsene. Inoltre, anche nell'insegnamento della traduzione non specializzata si dovrà sempre più spesso decidere se a fronte di un determinato compito di traduzione convenga o no adottare una determinata applicazione MT da sottoporre successivamente a *post-editing*. Le valutazioni finora espresse sembrano indicare che la soglia per una giustificabile emendabilità di traduzioni grezze si aggira intorno a 30 punti BLEU (cfr. Parra Escartín & Arcedillo, 2015: 134 in riferimento a situazioni professionali). Nel caso di DeepL Traduttore, e se ci limitiamo a considerare la situazione di traduzione verso il tedesco, i punteggi raggiunti si collocherebbero ragionevolmente vicino a questa soglia, che, ricordiamo, è calcolata soprattutto in riferimento a criteri di produttività in ambito professionale. Ma questo parametro è da utilizzare con molta cautela, e comunque non è possibile calcolare un punteggio BLEU per un output grezzo in assenza di un testo di riferimento precedentemente elaborato. Conseguentemente, nella maggior parte dei casi affrontati insieme agli studenti, la decisione pro o contro il *post-editing* viene presa in base ad altre valutazioni. In aggiunta, e in particolare con traduzioni generalmente verso la prima o la seconda lingua straniera, già questa decisione risulta rischiosa se affidata esclusivamente allo studente. Le sue incertezze sulla correttezza e la qualità stilistica del testo grezzo possono infatti condurre ad una valutazione errata sulla sua emendabilità, oltre, ovviamente, a interventi impropri in fase di esecuzione. Qualora si proceda con il *post-editing*, è quindi in genere necessario ricorrere ad ulteriori risorse per la verifica delle proprie ipotesi di miglioramento (cfr. ad es. Zanettin 2009: 217-218). In linea più generale, vanno sempre messe in conto le differenze tra i risultati di una valutazione automatica e quelli di una valutazione data da revisori umani (Castilho *et al.*, 2017: 114):

Technical *post-editing* effort was reduced for NMT in all language pairs using measures of actual keystrokes [...] or the minimum number of edits required to go from pre- to post-edited text [...]. Feedback from the participants indicated that they found NMT errors more difficult to identify, whereas word order errors and disfluencies requiring revision were detected faster in SMT output (ibid. 117).

Di conseguenza il fattore della leggibilità diventa cruciale per il giudizio dato dai revisori umani. Le nozioni di scorrevolezza e adeguatezza da applicare alle traduzioni svolte in ambito didattico, cioè sia come riferimento per la valutazione, sia come orientamento che lo studente deve tenere presente quando traduce, vanno adeguatamente chiarite ed esemplificate. In particolare si concretizzano nei livelli di qualità riguardanti: problemi di coerenza transfrastica, risoluzione delle ambiguità, corrispondenza dei contenuti informativi, adeguatezza culturale e specificità dello *skopos*. Questi ambiti, essenziali per la maggior parte delle attuali teorie della traduzione, non sono affrontabili con procedimenti che prendano in considerazione segmenti isolati del testo da tradurre. Anche i nuovi sistemi MT sono soggetti a questa limitazione. È quindi lecito ipotizzare che le sacche residue di inadeguatezze (non meramente occasionali) da affrontare nel *post-editing* risiedano proprio in questi ambiti, per i quali si richiede competenza testuale, disciplinare, comunicativa e interculturale.

6.1 Strategie generali, traduzione non specializzata e utilizzo nella didattica

Malgrado queste note limitazioni strutturali, DeepL Traduttore ha destato, come si è detto, un'impressione sorprendentemente favorevole nel corso di alcune prove svolte fin dal momento della sua commercializzazione per la combinazione linguistica italiano-tedesco anche in ambito non specialistico. Di conseguenza si è ritenuto indispensabile verificare tempestivamente se e fino a che punto si dovesse tenere conto di questa applicazione, che appariva in grado di svolgere compiti di traduzione a un livello molto vicino alle soglie di accettabilità richieste dai corsi di traduzione nel corso di laurea triennale. Queste verifiche non sono ancora concluse, ma siamo tuttavia in grado di illustrare ora con alcuni esempi concreti il livello di utilizzabilità e adattabilità di DeepL Traduttore nell'insegnamento della traduzione dall'italiano al tedesco nell'ambito di una laurea triennale, con tipi di testo non specialistici.

I testi tradotti rientrano nella tematica generale del turismo (particolarmente rilevante per la preparazione dei futuri traduttori con questa combinazione linguistica). Dal punto di vista della qualità del prodotto entrano in gioco criteri di adeguatezza al pubblico dei fruitori e di efficacia funzionale rispetto alle finalità informative e promozionali del messaggio. Dal punto di vista della situazione didattica sono coinvolte traduttrici (o traduttori) in formazione, che traducono verso la lingua straniera. Si tratta di pagine web di siti turistici italiani, tradotte in elaborati finali (tesine), con la possibilità di riprendere e rielaborare anche le versioni prodotte da DeepL Traduttore. Nell'ambito di questo contributo ci limiteremo a discutere le soluzioni linguistiche, senza prendere in considerazione la dimensione multimodale dei testi affrontati. Fin dall'inizio, nei casi esaminati si sono delineati per il *post-editing* i seguenti principali punti di attenzione:

- il trattamento di vincoli e convenzioni testuali e culturali per il testo da realizzare;
- il trattamento della successione tema/rema;
- il trattamento delle implicazioni dell'interpunzione a livello sintattico;
- il trattamento dei nomi propri e dei *realia*;
- il trattamento dei termini specialistici nei brani non specialistici;
- il trattamento degli elementi di coesione e coerenza in porzioni di testo particolarmente lunghe
- la resa dei menu nelle pagine web;
- il cambio di prospettiva nella deissi e nei riferimenti da realizzare nel testo d'arrivo.

6.1.1 Riscontro n. 4

Si tratta di un brano contenuto nel sito "Cantina di Cesena" <http://www.cantinacesena.it/>, caratterizzato da funzioni informative e promozionali, con elementi di specializzazione terminologica nella denominazione dei vini.

ORIGINALE	OUTPUT TRADUTTORE	DEEPL POST-EDITING
Vini e profumi di Romagna "Romagna solatia, dolce paese, cui regnarono Guidi e Malatesta, cui tenne pure il Passator cortese, re della strada, re della foresta." (G. Pascoli) Sono le dolci colline romagnole ad accogliere i vigneti dei soci di Cantina di Cesena: la conformazione del terreno e il clima mite favoriscono le locali varietà di Sangiovese DOC, Trebbiano DOC e Albana	Weine und Düfte der Romagna Romagna solatia, süßes Dorf, in dem Guidi und Malatesta herrschten, die auch den höflichen Passator, König der Straße, König des Waldes, hielten. (G. Weiden) Es sind die sanften Hügel der Romagna, die die Weinberge der Mitglieder der Cantina di Cesena beherbergen: Die Beschaffenheit des Bodens und das milde Klima begünstigen die lokalen Sorten Sangiovese DOC, Trebbiano DOC und Albana.	Wein und Duft der Romagna "Romagna solatia, dolce paese, cui regnarono Guidi e Malatesta, cui tenne pure il Passator cortese, re della strada, re della foresta." (G. Pascoli) Die Weinberge der Mitglieder der Cantina di Cesena liegen auf den sanften Hügeln der Romagna: die Beschaffenheit des Bodens und das milde Klima begünstigen die lokalen Sorten Sangiovese DOC (Q.b.A.), Trebbiano DOC (Q.b.A) und Albana.

Nella versione fornita da DeepL Traduttore i difetti principali compaiono nel trattamento letterale del titolo, del cognome del poeta, e soprattutto della citazione (come del resto era facile prevedere). Quest'ultimo problema è stato eluso dalla laureanda, che nel *post-editing* ha ripristinato la versione italiana della strofa.

6.1.2 Riscontro n. 5

ORIGINALE	OUTPUT TRADUTTORE	DEEPL POST-EDITING
Una storia che guarda lontano Cantina Sociale di Cesena nasce nel 1933 dalla comune volontà di alcuni viticoltori di conferire l'intera produzione in una realtà nuova, in grado di esprimere al meglio i valori e la vocazione enologica del territorio. La filosofia della Cantina è subito chiara: portare innovazione nella vitivinicoltura romagnola nel rispetto del terroir e delle tipicità vinicole. Giorno dopo giorno, il sogno dei primi pionieri prende vita in numerosi progetti e in costanti investimenti nelle più moderne tecnologie di vinificazione.	Eine Geschichte, die weit weg schaut Die Kellerei Cantina Sociale di Cesena wurde 1933 durch den gemeinsamen Willen einiger Winzer gegründet, um die gesamte Produktion in eine neue Realität zu überführen, die in der Lage ist, die Werte und die Berufung des Weinbaugebiets am besten zum Ausdruck zu bringen. Tag für Tag wird der Traum der ersten Pioniere in zahlreichen Projekten und ständigen Investitionen in modernste Weinbereitungstechnologien verwirklicht.	Die Geschichte der Cantina di Cesena Die Cantina Sociale di Cesena wurde 1933 von einigen Winzern gegründet, um die gesamte Produktion in eine neue Realität zu überführen, die den traditionellen Werten und der besonderen Eignung des Weinbaugebiets Ausdruck verleihen kann. Die Philosophie der Cantina di Cesena ist sofort klar: Innovation im Weinbau der Romagna in Bezug auf Terroir und typische Weine. Tag für Tag wird der Traum der ersten Pioniere durch zahlreiche Projekte und ständige Investitionen in modernste Weinbereitungstechnologien verwirklicht.

Anche in questo raffronto risaltano nella resa di DeepL Traduttore una collocazione impropria ("Eine Geschichte, die weit weg schaut") e alcuni passaggi troppo letterali. Gli interventi della laureanda nel *post-editing* migliorano in parte la scorrevolezza del testo con l'eliminazione di ridondanze a livello informativo. È importante sottolineare che la collocazione errata costituisce un fallimento di DeepL Traduttore all'interno dei confini di frase e in un contesto non particolarmente metaforico, cioè in un ambito in cui dovrebbe incontrare minori difficoltà.

6.1.3 Riscontro n. 6

ORIGINALE	OUTPUT TRADUTTORE	DEEPL POST-EDITING
Facciamo parlare la terra Sono le dolci colline romagnole ad accogliere i	Sprechen wir über die Erde Es sind die sanften Hügel der Romagna, die die Weinberge	Lassen wir die Erde sprechen

vigneti dei soci di Cantina di Cesena: la conformazione del terreno e il clima mite consentono una ottima resa media (circa 80 quintali per ettaro) favorendo le locali varietà di Sangiovese DOC, Trebbiano DOC e Albana DOCG. I vini prodotti (venduti sfusi o confezionati) sono commercializzati con differenti marchi tra i quali: "Tenuta Amalia", regina della qualità dei nostri prodotti, "I Vigneti dei Tre Papi", nome coniato in omaggio ai tre cesenati saliti al soglio pontificio: Pio VI Braschi, Pio VII Chiaramonti e Pio VIII Castiglioni e il marchio "Le Tavolozze"

der Mitglieder der Cantina di Cesena willkommen heißen: Die Beschaffenheit des Bodens und das milde Klima ermöglichen einen ausgezeichneten Durchschnittsertrag (ca. 80 Doppelzentner pro Hektar), der die lokalen Sorten Sangiovese DOC, Trebbiano DOC und Albana DOCG begünstigt. Die produzierten Weine (lose oder verpackt) werden unter verschiedenen Markennamen vermarktet, darunter: "Tenuta Amalia", Königin der Qualität unserer Produkte, "I Vini dei Tre Papi", ein Name, der als Hommage an die drei Cesenati, die auf den päpstlichen Thron geklettert sind: Pius VI Braschi, Pius VII Chiaramonti und Pius VIII Castiglioni und die Marke "Castiglioni".

Die Weinberge der Mitglieder der Cantina di Cesena liegen auf den sanften Hügeln der Romagna: Die Beschaffenheit des Bodens und das milde Klima erlauben einen exzellenten Durchschnittsertrag (ca. 80 Doppelzentner pro Hektar), der die lokalen Sorten Sangiovese DOC (Q.b.A.), Trebbiano DOC (Q.b.A.) und Albana DOCG (Prädikatswein) begünstigt. Die Weine (ob offen oder in Flaschen) werden unter verschiedenen Marken angeboten: "Tenuta Amalia", Königin der Qualität unserer Produkte, "I Vigneti dei Tre Papi" (Die Weinberge der drei Päpste aus Cesena), Pius VI Braschi, Pius VII Chiaramonti und Pius VIII Castiglioni, und die Marke "Le Tavolozze" (wörtlich „Farbpalette“).

Le modifiche al testo introdotte in fase di post-editing sono in questo caso significative. Al di là della semplificazione di una metafora di sapore problematico per un lettore tedesco (le dolci colline che 'accolgono' i vigneti), è stato necessario emendare collocazioni errate (*lose* e *verpackt* in combinazione con *Weine*; *geklettert* in combinazione con *Thron*). Compare inoltre una fatale e inspiegabile sostituzione di parole (*Castiglioni* al posto di una traduzione per *Le Tavolozze*), che deriva probabilmente da fattori non attinenti al training del sistema. È infine dubbio se il termine italiano *Cesenati*, mantenuto inalterato in tedesco, rientri ancora nei limiti di comprensibilità per i lettori tedeschi non filologicamente preparati. Nel *post-editing* questo potenziale ostacolo è stato semplicemente rimosso, ma sarebbero ovviamente possibili anche soluzioni esplicative meno drastiche di questa. Si avverte poi il bisogno di una resa più informativa (per il lettore tedesco) delle sigle di qualità che accompagnano i nomi dei vini, che la laureanda ha cercato di realizzare nel *post-editing*.

7. Conclusioni

Le verifiche qui esposte poggiano su un numero ridotto di campioni, e si concentrano solo su aspetti della traduzione dall'italiano in tedesco. Entro questi confini, ci siamo limitati a considerare i prodotti grezzi di DeepL Traduttore, senza includere le modalità di editing in linea incorporate nell'interfaccia e senza verificare gli auto-aggiustamenti che ne potrebbero conseguire in tempo reale. È quindi scontato che stiamo presentando risultati parziali e indicazioni bisognose di ulteriori riscontri a diversi livelli.

Verificare la qualità di un'applicazione MT come DeepL Traduttore al di fuori delle metriche e delle procedure standardizzate, in una combinazione linguistica (dall'italiano in tedesco) non ancora sistematicamente esplorata dai valutatori, e per di più in contesti didattici reali (con le priorità specifiche della didattica) è un compito molto complesso. Si ripresentano infatti alcune delle difficoltà già riscontrate nei procedimenti automatizzati in uso nei settori dello sviluppo informatico e/o dell'impiego professionale, e in aggiunta compaiono difficoltà particolari riguardanti la didattica della traduzione generalistica e/o specialistica.

In primo luogo, i punteggi BLEU sono un punto di partenza utile, ma solo parzialmente idoneo per determinare la qualità di un'applicazione MT integrata nella didattica. Da un lato questi punteggi vengono calcolati di preferenza per alcuni campi tematico-funzionali fissati per convenzione e sulla base di campioni più o meno rappresentativi. Dall'altro derivano da livelli ancora molto semplificati di aderenza a versioni modello, con parametri probabilmente troppo livellanti. Nelle nostre prove, in particolare, i punteggi ottenuti hanno presentato oscillazioni difficili da mettere in relazione con specifiche differenze fra i testi (ad es. il passaggio da un dominio all'altro, il livello di specializzazione terminologica, la complicazione sintattica, ecc.). Anche le impressioni sulla leggibilità e la scorrevolezza dei testi grezzi proposti dall'applicazione sembrano distaccarsi dai valori rilevati e considerati caso per caso, e in generale le traduzioni di DeepL Traduttore si sono rivelate compatibili con una ragionevole soglia di emendabilità tramite *post-editing*, sia in ambito didattico, sia in un certo numero di situazioni professionali. È tuttavia inevitabile rimarcare che i punteggi BLEU calcolati per il nostro ciclo di verifiche di traduzioni specialistiche ottenute con DeepL Traduttore non hanno in nessun caso sopravanzato quelli ottenuti nelle stesse prove da Google Traduttore. Questo dato indica che la superiorità di DeepL Traduttore dichiarata al momento del lancio potrebbe essere nel frattempo sfumata. Appare quindi necessario allargare il quadro dell'indagine e condurre appena possibile un confronto sistematico tra le due applicazioni, in particolare dal punto di vista della rispettiva rilevanza e adattabilità per la formazione dei traduttori.

In secondo luogo, le priorità, i criteri e le modalità del *post-editing* necessitano ancora di raffinamenti significativi e contemporaneamente di una maggiore uniformità, in modo da ottenere risultati precisi, confrontabili e rilevanti per diversi ambienti (didattici, scientifici e professionali). Nelle nostre verifiche abbiamo constatato che studenti e studentesse con una buona formazione sulle priorità operative della traduzione funzionale riescono a ottenere buoni risultati ai fini della valutazione didattica anche senza applicare tabelle o *check-list*

parametrizzate. D'altro canto è emerso chiaramente che le traduzioni così elaborate possono essere ancora sensibilmente distanti da versioni ufficiali di riferimento.

Riconoscere e sanare fino in fondo le imperfezioni della versione di DeepL Traduttore ha messo seriamente alla prova la competenza delle persone coinvolte, che peraltro non avevano ancora concluso il percorso accademico di formazione. Alcune di queste imperfezioni, in particolare collocazioni errate o inadeguate, si annidano all'interno dei confini di frase, cioè per così dire sul terreno in cui i sistemi MT danno il meglio di sé stessi. Si tratta evidentemente di combinazioni non reperibili (o non correttamente elaborabili) sulla base dei repertori allineati di cui si serve l'applicazione, e si può supporre che la loro peculiarità le renda particolarmente difficili da trattare anche per esperti umani. Queste imperfezioni sono probabilmente destinate a diminuire progressivamente, con il progredire del feedback, il raffinamento del training e il perfezionamento degli algoritmi. Per eliminare altri tipi di imperfezione, invece, sono richieste potenzialità di per sé non (ancora) alla portata dell'architettura neurale, che si collocano al di là dei meccanismi di ottimizzazione attualmente in uso. Nel trattamento di casi attinenti alla sfera interculturale, ai riferimenti interdisciplinari e ad altri condizionamenti di diversa natura, l'applicazione è probabilmente destinata a fornire soluzioni solo occasionalmente adeguate. La didattica della traduzione universitaria deve quindi incorporare al più presto, come componente strategica e prioritaria, metodi e pratiche avanzate di *post-editing*: un traguardo problematico, se si considera che nella maggior parte dei corsi la competenza traduttiva viene perfezionata insieme a quella linguistica e comunicativa, oltre alla padronanza di tecniche avanzate di documentazione. Per arrivare a formule illuminate di *post-editing* è necessario colmare la distanza che ancora separa i criteri attualmente in uso (legati a protocolli applicabili solo a specifici ambiti della traduzione professionale) da criteri più complessi e articolati di revisione e valutazione derivanti dalle teorie di impostazione funzionale (in particolare Vermeer 1983; Nord, 1993), fino agli approcci integrati degli ultimi anni.

Riferimenti bibliografici

- Bentivogli L., Bisazza A., Cettolo M. & Federico M. (2016), Neural versus Phrase-Based Machine Translation Quality: a Case Study. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*: 257–267. www.aclweb.org/anthology/D/D16/D16-1025.pdf
- Castilho S., Moorkens J., Gaspari F., Calixto I., Tinsley J. & Way A. (2017), Is Neural Machine Translation the New State of the Art?, in *The Prague Bulletin of Mathematical Linguistics*, Number 108 June 2017: 109–120.
- Cattelan A. (2017), PBMT vs. NMT: Which helps translators the most?, <https://locworld-localizationworld.netdna-ssl.com/wp-content/uploads/2017/11/TS4---Alessandro-Cattelan.pdf>
- Cattelan A. & Di Lorenzo C. (2017), *Phrase-Based MT vs. Neural MT*. MateCat Webinar, <https://www.matecat.com/webinar/>
- Flanagan M. & Paulsen Christensen T. (2014), Testing post-editing guidelines: how translation trainees interpret them and how to tailor them for translator training purposes, in *The Interpreter and Translator Trainer*, 8:2: 257-275, DOI: 10.1080/1750399X.2014.93611
- Giavelli F. (2018). DeepL: la nuova frontiera della traduzione automatica neurale a confronto con il linguaggio enologico: uno studio basato sulla traduzione del sito della Cantina di Cesena, Tesi di Laurea Triennale, Università di Bologna, DIT.
- Junczys-Dowmunt M., Dwojak T. & Sennrich R. (2016), The AMU-UEDIN Submission to the WMT16 News Translation Task: Attention-based NMT Models as Feature Functions in Phrase-based SMT, in *Proceedings of the First Conference on Machine Translation*, Volume 2: *Shared Task Papers*: 319–325, <https://arxiv.org/abs/1605.04809v3>
- Madnani N., Resnik P., Dorr B. J. & Schwartz R. (2008), *Are Multiple Reference Translations Necessary? Investigating the Value of Paraphrased Reference Translations in Parameter Optimization*, <http://www.mt-archive.info/AMTA-2008-Madnani.pdf>
- Niño A. (2008), Evaluating the use of machine translation *post-editing* in the foreign language class, in *Computer Assisted Language Learning*, Vol. 21, No. 1: 29–49.
- Nord C. (1993), *Einführung in das funktionale Übersetzen*, Francke Verlag, Tübingen und Basel.
- Pagin E. (2018), La traduzione al servizio dell'internazionalizzazione d'impresa: l'output del programma di traduzione automatica DeepL a confronto con i risultati della traduzione assistita, Tesi di Laurea Magistrale, Università di Bologna, DIT.
- Parra Escartin C. & Arcedillo M. (2015), Machine translation evaluation made fuzzier: A study on *post-editing* productivity and evaluation metrics in commercial settings, in *Proceedings of MT Summit XV*, vol.1: 131-144.
- Papinemi K., Roukos S., Ward T., Henderson J., & Reeder F. (2002), BLEU: a method for automatic evaluation of machine translation, in *of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, Philadelphia: 311–318.
- Popović M., Arčan M. & Lommel A. (2016), Potential and Limits of Using Post-edits as Reference Translations for MT Evaluation, in: *Baltic J. Modern Computing*, Vol. 4 (2016), No. 2: 218–229.
- Qin Y., Specia L. (2015), *Truly Exploring Multiple References for Machine Translation Evaluation*, www.aclweb.org/anthology/W15-4915
- Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser Ł. & Polosukhin I. (2017), *Attention Is All You Need*, <https://arxiv.org/pdf/1706.03762>
- Vermeer H.-J. (1983), *Grundlagen einer allgemeinen Translationstheorie*, Niemeyer, Tübingen.
- Zhang M. (2017), *History and Frontier of the Neural Machine Translation*, <https://syncedreview.com/2017/08/17/history-and-frontier-of-the-neural-machine-translation/>
- Zanettin F. (2009), Corpus-based Translation Activities for Language Learner, in *The Interpreter and Translator Trainer*, 3:2, 209-224, DOI: 10.1080/1750399X.2009.10798789.

Note

[1] Per una panoramica delle caratteristiche di base delle reti neurali attualmente impiegate nella traduzione automatica, cfr. Cattelan & Di Lorenzo, 2017.

[2] Il punteggio BLEU si riferisce alla corrispondenza puramente formale fra un testo prodotto dall'applicazione di MT e un testo ottimale di riferimento, elaborato e verificato a regola d'arte. Questa corrispondenza viene calcolata sulla base di segmenti allineati, e all'interno di ogni coppia si verifica la percentuale di parole e di *n-gram* condivise. Per una descrizione completa cfr. Papinemi *et al.*, 2002).

[3] In base all'etichetta "newstest2014" riportata nei grafici si tratta con ogni probabilità del pacchetto di test distribuito nel febbraio 2014 per ACL 2014 Ninth Workshop on Statistical Machine Translation (<http://statmt.org/wmt14/>), che è servito per quell'anno da banco di prova per le verifiche di qualità. Nel frattempo sono in uso altre specifiche (WMT 2016), ma è probabile che si sia usata la versione 2014 perché è quella che permette i migliori raffronti con i valori della concorrenza.

[4] In alternativa, alcuni gruppi di ricerca ricorrono per la misura della qualità a testi prodotti con sistemi MT e successivamente sottoposti a *post-editing*. Questa metodologia, ovviamente non priva di effetti autoreferenziali, richiederebbe a sua volta ulteriori raffinamenti piuttosto dispendiosi, e ai fini di questa discussione non appare comunque vantaggiosa (pur tenendo conto di una serie di verifiche in Popović *et al.*, 2016).

[5] Il sito ufficiale della Provincia Autonoma di Bolzano contiene 3 versioni ufficiali della legge: in italiano, tedesco e ladino, ove la versione italiana deve ritenersi quella vincolante: http://lexbrowser.provinz.bz.it/doc/it/197684/legge_provinciale_17_settembre_2013_n_18.aspx?view=1

[6] Fanno eccezione i procedimenti di calcolo basati su "multiple reference translations", in genere piuttosto complessi e non ancora introdotti su vasta scala (ad es. Madnani *et al.*, 2008; Qin & Specia, 2015).

[7] Le autrici prospettano anche la necessità di un analogo set di regole per le situazioni in cui si miri ad una "qualità sufficiente" Flanagan & Paulsen Christensen (2014: 273), rimandandone però la proposta ad un momento successivo.

©inTRAlinea & Christine Heiss & Marcello Soffritti (2018).

"DeepL Traduttore e didattica della traduzione dall'italiano in tedesco Alcune valutazioni preliminari", *inTRAlinea* Special Issue: Translation And Interpreting for Language Learners (TAIL).

Stable URL: <http://www.intraline.org/archive/article/2294>