



Angewandte Linguistik

Giulia Gobbo

La traduzione automatica nei CAT-Tool

Tre modalità di visualizzazione a confronto: interattiva, nel segmento o come suggerimento?

Graduate Papers
in Applied
Linguistics 22



Die vorliegende Arbeit wurde am Departement Angewandte Linguistik der ZHAW Zürcher Hochschule für Angewandte Wissenschaften im Frühlingssemester 2023 als Masterarbeit im Masterstudiengang Angewandte Linguistik, Vertiefung Fachübersetzen, verfasst (Referenten: Dr. Hellmut Riediger und Dr. Martin Kappus) und mit dem Supertext-Preis für die beste Masterarbeit im Bereich Fachübersetzen ausgezeichnet.

Das Departement Angewandte Linguistik der ZHAW betreibt Angewandte Linguistik als transdisziplinär orientierte Sprachwissenschaft. Diese befasst sich mit den Problemen der realen Welt, in denen Sprache eine zentrale Rolle spielt. Sie identifiziert, analysiert und löst diese Probleme einerseits durch die Anwendung linguistischer Theorien, Methoden und Resultate, andererseits durch die Entwicklung neuer theoretischer und methodischer Ansätze.

In den *Graduate Papers in Applied Linguistics* veröffentlicht das Departement Angewandte Linguistik der ZHAW preisgekrönte Abschlussarbeiten von Studierenden des Bachelorstudiengangs Angewandte Sprachen, des Bachelorstudiengangs Kommunikation, des Masterstudiengangs Angewandte Linguistik und des MAS Communication Management and Leadership

Kontakt

ZHAW Angewandte Linguistik
Theaterstrasse 15c
Postfach
8401 Winterthur

info.linguistik@zhaw.ch

+41 (0) 58 934 60 60

Giulia Gobbo 2023: La traduzione automatica nei CAT-Tool. Tre modalità di visualizzazione a confronto: interattiva, nel segmento o come suggerimento? Winterthur: ZHAW Zürcher Hochschule für Angewandte Wissenschaften. (*Graduate Papers in Applied Linguistics* 22).

DOI 10.21256/zhaw-2814 (<https://doi.org/10.21256/zhaw-2814>)

Indice

	Abstract	3
1	Introduzione	4
2	La traduzione automatica	7
2.1	Il post-editing	8
2.2	La traduzione automatica interattiva.....	9
2.3	La traduzione automatica adattiva	12
3	I CAT-Tool.....	14
3.1	Cenni storici.....	14
3.2	Componenti e funzionalità.....	15
3.3	Integrazione della traduzione automatica nei CAT-Tool	16
3.4	Lilt.....	17
3.5	Ergonomia nei CAT-Tool.....	18
4	Metodi di ricerca per il processo di traduzione	20
4.1	Misurazione dello sforzo necessario per il post-editing	20
4.2	Screen recording e keystroke logging	22
4.3	Feedback dei partecipanti	23
4.4	Qualità del prodotto finale	23
4.5	Confronto tra professionisti e studenti	24
5	Descrizione dell'esperimento.....	25
5.1	Partecipanti.....	25
5.2	Incarichi di traduzione	25
5.3	Materiale	28
5.4	Raccolta dei dati	29
5.5	Valutazione della qualità	29
6	Risultati e analisi.....	31
6.1	Sforzo temporale.....	31
6.2	Sforzo tecnico	33
6.3	Sforzo cognitivo	34
6.4	Qualità dei testi di arrivo.....	36
6.5	Opinioni dal questionario.....	39
6.6	Ulteriori considerazioni.....	42

7	Discussione.....	47
7.1	Relazione tra sforzo e qualità.....	47
7.2	Confronto tra opinioni e dati misurati.....	48
7.3	Confronto con altre ricerche.....	48
7.4	Confronto tra studenti e professionisti.....	50
7.5	Qualità dell'output di Lilt.....	52
7.6	Limitazioni.....	53
8	Conclusione.....	55
	Indice delle abbreviazioni.....	57
	Indice delle figure e delle tabelle.....	58
	Bibliografia.....	59
	Allegati.....	62
	Allegato A: Testi tradotti durante l'esperimento.....	62
	Allegato B: Risultati del processo di traduzione.....	65
	Allegato C: Qualità dei testi finali.....	68
	Allegato D: Dati demografici.....	70
	Allegato E: Risposte al questionario.....	70
	Allegato F: Istruzioni date ai partecipanti (esempio P1).....	73

Quando in questo documento, unicamente a scopo di semplificazione, è usato il maschile, la forma è da intendersi riferita in maniera inclusiva a tutte le persone, sia di genere maschile che femminile.

Abstract

Nel settore della traduzione l'importanza di tradurre contenuti in modo rapido e qualitativo è sempre più marcata. La tesi ha l'obiettivo di identificare la modalità migliore di visualizzare i risultati della traduzione automatica (TA) nei CAT-Tool per aumentare la produttività, senza porre limiti alla qualità. A tal fine è stato condotto un esperimento con sette partecipanti (traduttori professionisti e studenti) e tre modalità a confronto: TA direttamente nel segmento (post-editing convenzionale), TA all'esterno del segmento come suggerimento e TA interattiva. Ogni partecipante ha tradotto tre brevi testi, uno per ogni modalità. Il processo di traduzione è stato registrato attraverso il keystroke logging e la registrazione dello schermo. Al termine dell'esperimento i partecipanti hanno completato un questionario online sulle loro preferenze e opinioni riguardo alle tre modalità. Infine, tre esperti hanno valutato la qualità dei testi finali. I risultati indicano che il post-editing convenzionale ha richiesto lo sforzo temporale, tecnico e cognitivo minore e ha permesso di ottenere una qualità finale leggermente maggiore rispetto alle altre due modalità. La maggior parte dei partecipanti ha preferito tradurre con il post-editing convenzionale. Si può concludere che il post-editing convenzionale permette di aumentare la produttività senza compromettere la qualità dei testi di arrivo.

Finding productive ways to translate while maintaining high quality levels is becoming increasingly important. This thesis aims to identify the best way to display machine translation (MT) results in the CAT-Tool to boost productivity without compromising quality: directly in the segment (conventional post-editing), as a suggestion outside of the segment or interactively. An experiment with seven participants (professional translators and students) has been conducted for this purpose. Each participant translated three short texts, one for each mode. The translation process was registered with screen recording and keystroke logging. After the experiment, the participants completed an online questionnaire with questions about their preferences and opinions regarding the three modes. The quality of the final texts was then assessed by three experts. The results show that conventional post-editing required the least temporal, technical, and cognitive effort and produced target texts with a slightly higher quality. Most participants preferred to translate with conventional post-editing. The main conclusion that is drawn is that conventional post-editing increases productivity without compromising the quality of the target texts.

1 Introduzione

Nel settore della traduzione l'importanza di tradurre contenuti in modo rapido, qualitativo e soprattutto economico è sempre più marcata. Per soddisfare questa domanda è utile scoprire quali metodi di lavoro sono i più efficienti e permettono di aumentare la produttività risparmiando tempo e denaro, senza trascurare la qualità.

I CAT-Tool e la traduzione automatica (TA) fanno ormai parte della quotidianità nella professione. La TA può essere integrata e visualizzata nei CAT-Tool in diversi modi: come suggerimento, direttamente nel segmento o in modo interattivo. Scegliere il giusto approccio all'uso della traduzione automatica nei CAT-Tool è utile da diversi punti di vista, oltre che da quello puramente economico.

In primo luogo, per i traduttori stessi, che possono diventare più consapevoli del funzionamento della TA e trovare il modo migliore per integrarla nel proprio CAT-Tool. Inoltre, aumentando la produttività in testi in cui è possibile usare molto la TA, rimane più tempo per le altre traduzioni che richiedono più creatività o per altre attività. L'aumento della produttività permette a sua volta di ridurre lo stress e aiutare i traduttori in situazioni in cui i tempi di consegna sono molto ristretti.

Il risultato della ricerca potrebbe anche aiutare studenti di traduzione che non hanno ancora esperienza con i CAT-Tool, in modo che sappiano come interagirci per aumentare al massimo la loro produttività. Infatti, una migliore comprensione e consapevolezza della TA e del post-editing può preparare gli studenti alle sfide professionali future legate all'evoluzione della tecnologia (Jia et al., 2019).

Infine, anche gli sviluppatori di sistemi di traduzione automatica e di CAT-Tool possono beneficiare di queste conoscenze per innovare i loro prodotti in base alle modalità rivelatesi più efficienti o alle funzioni mancanti desiderate dai traduttori. Infatti, i sistemi di TA interattiva e adattiva non sono ancora molto presenti all'interno dei CAT-Tool commerciali.

Sono già stati condotti diversi studi sull'uso di sistemi di TA interattivi. La storia della TA interattiva parte da TransType, il primo sistema di questo tipo, per passare a piattaforme come Cairta e Casmacat. I risultati di queste ricerche hanno mostrato che con la TA interattiva solo raramente è stato possibile aumentare la produttività rispetto al post-editing convenzionale (Läubli & Green, 2019).

Tuttavia, bisogna precisare che la maggior parte di questi studi è stata condotta diversi anni fa con sistemi di TA statistici, motivo per cui è interessante svolgere una nuova ricerca usando la TA neurale, che offre un salto di qualità. Inoltre, la TA è in costante evoluzione e i risultati trovati qualche anno fa potrebbero già essere cambiati.

Uno studio più recente condotto da Knowles et al. (2019), invece, ha mostrato che oltre la metà dei partecipanti ha tradotto in modo più rapido con la TA neurale interattiva e l'ha preferita rispetto al post-editing.

Le ricerche menzionate, però, si sono concentrate sulla TA interattiva, ma non sulla differenza tra la presentazione della TA direttamente nel segmento o all'esterno del segmento come suggerimento. La presente tesi intende quindi confrontare le tre diverse modalità di visualizzazione della TA citate, tenendo conto dei progressi più recenti della TA neurale.

L'obiettivo della tesi è di trovare un metodo adatto per aumentare la produttività traducendo nei CAT-Tool con l'aiuto della TA, senza porre limiti alla qualità.

La domanda di ricerca a cui la tesi cerca di rispondere è:

Qual è il modo migliore per aumentare l'efficienza e la produttività con l'aiuto della traduzione automatica nei CAT-Tool?

A questa si legano le seguenti domande di approfondimento:

- Qual è la modalità migliore per presentare la TA ai traduttori umani? Interattiva, nel segmento o come suggerimento? Con quale di queste si lavora in modo più efficiente ed esercitando meno sforzo?
- In che modo si può aumentare la produttività con la TA nei CAT-Tool senza porre limiti alla qualità?
- Quali sono le impressioni e opinioni di chi ha partecipato allo studio riguardo alle tre modalità?

Nella tesi sarà dapprima presentata una parte teorica sulla traduzione automatica, sui CAT-Tool e sui metodi di ricerca per la traduzione per poi passare alla parte empirica, che comprende l'osservazione del lavoro di tre traduttori professionisti e quattro studenti di master.

Ogni partecipante ha tradotto tre brevi testi usando le tre modalità: TA nel segmento del testo di arrivo, TA come suggerimento all'esterno del segmento, TA interattiva. Sono stati proposti dei testi sia con parti da tradurre in modo creativo, generalmente meno adatte alla TA, sia con parti piuttosto tecniche. L'analisi si è concentrata sulla combinazione linguistica tedesco-italiano, poco ricercata finora in questo ambito (Läubli et al., 2019).

Durante le traduzioni, i dati sono stati raccolti tramite registrazione dello schermo e key-stroke logging. Questi dati quantitativi ottenuti hanno permesso di fare un confronto diretto tra le tre modalità e hanno mostrato lo sforzo temporale, lo sforzo tecnico e lo sforzo cognitivo.

Per completare i dati raccolti, dopo aver concluso le traduzioni, i partecipanti hanno ricevuto un questionario con domande chiuse e aperte che ha permesso di capire le loro opinioni in merito alle varie modalità. Con questi dati prevalentemente qualitativi è stato possibile capire soprattutto cosa hanno o non hanno apprezzato e perché.

La qualità dei testi tradotti è poi stata valutata da tre esperti esterni allo studio secondo uno schema definito. Questo passaggio ha permesso di stabilire se un eventuale aumento di velocità nella traduzione ha un impatto negativo sulla qualità.

Il confronto tra professionisti e studenti ha permesso di mostrare il ruolo dell'esperienza di chi lavora nel settore da diversi anni e quello delle conoscenze degli studenti sugli ultimi sviluppi tecnologici.

I dati raccolti sono stati anonimizzati e all'interno del questionario i partecipanti hanno acconsentito al trattamento dei dati.

2 La traduzione automatica

I sistemi di traduzione automatica (TA) sono largamente diffusi nella pratica professionale dei traduttori e vengono integrati nei sistemi di traduzione assistita (CAT-Tool) (cfr. cap. 3). Oggi si usa principalmente la TA neurale, una tecnologia introdotta nel 2016 che ha preso il posto dei sistemi basati su regole e statistici usati in precedenza. “L’avvento della TA basata su reti neurali [...] ha provocato un’impennata nelle aspettative, anche per quanto riguarda generi testuali prima non ritenuti adatti, quali per esempio la saggistica o addirittura la narrativa” (Riediger, 2018, p. 8).

Come indica il nome, la TA neurale lavora con delle reti neurali artificiali, ovvero dei sistemi matematici che trasformano una frase in una serie di numeri ed eseguono milioni di operazioni in una frazione di secondo (Pérez-Ortiz et al., 2022). Queste reti sono composte da migliaia o milioni di unità artificiali che assomigliano ai neuroni e vengono attivate in base ai segnali che ricevono da altri neuroni e alla forza delle connessioni che trasportano questi segnali (Pérez-Ortiz et al., 2022). I sistemi di TA neurale vengono allenati attraverso esempi presi da corpora di testi paralleli che contengono milioni di frasi di partenza con le relative traduzioni (Pérez-Ortiz et al., 2022). Una grande quantità di dati a disposizione per una determinata combinazione linguistica e un determinato genere testuale permette al sistema di lavorare meglio e di ottenere risultati più precisi. Col tempo, i sistemi di TA sono in grado di “perfezionarsi attraverso un’intelligenza basata su capacità di autoapprendimento (machine learning)” (Riediger, 2018, p. 20).

Un sistema di TA funziona particolarmente bene quando incontra “testi tecnici o informativi con una struttura lineare e culturalmente neutri” (Riediger, 2018, p. 8), mentre ha difficoltà quando deve tradurre espressioni idiomatiche, termini connotati o ambigui e concetti che richiedono conoscenze extralinguistiche. O’Brien (2022) definisce la TA una tecnologia imperfetta. Infatti, per una frase può produrre una traduzione accurata e accettabile, ma per un’altra può portare a errori nel significato, omissioni, aggiunte o problemi stilistici. Il tipo e il numero di errori presenti in un testo prodotto attraverso la traduzione automatica dipendono da vari fattori, tra cui la combinazione linguistica, il contenuto nonché i dati o le tecniche usate per allenare il sistema di TA (O’Brien, 2022).

Per i sistemi di TA statistica e neurale vale quanto indicato da Riediger (2018, p. 10):

- “più si usano (bene), meglio funzionano;
- più stanno online («nella cloud» o in altre basi dati condivise), più diventano accessibili a un largo numero di persone, e più saranno usate;
- più passa il tempo, più imparano e migliorano le loro performance”.

Molti sistemi di TA sono disponibili gratuitamente sul web. Tuttavia, raccolgono i dati che vengono inseriti e di conseguenza non sono adatti a testi confidenziali. I sistemi a pagamento, invece, garantiscono una maggiore sicurezza dei dati e permettono l'integrazione nei CAT-Tool (cfr. cap. 3.3).

2.1 Il post-editing

Con post-editing si intende l'identificazione e la correzione degli errori nell'output della traduzione automatica da parte di traduttori professionisti (O'Brien, 2022). Questo procedimento è necessario soprattutto quando un testo è destinato alla pubblicazione e richiede una buona qualità.

Già nel 1952, durante la prima conferenza sui traduttori automatici al MIT, si è capito che insieme alla traduzione automatica era necessario un intervento umano (Riediger, 2018). Infatti, la pratica del post-editing è stata svolta fin dagli inizi della tecnologia di traduzione automatica, ma si è sviluppata molto negli ultimi anni (Vieira, 2019). Inizialmente, il post-editing era considerato un'attività passiva, in cui i traduttori non dovevano sempre conoscere la lingua di partenza e aiutavano la macchina. Questa visione incentrata sulla macchina ha compromesso la reputazione di tale attività e di conseguenza molti traduttori ne avevano un'opinione piuttosto negativa. Gli sviluppi più recenti hanno permesso di migliorare l'uso della TA, ad esempio rendendola disponibile nei sistemi di traduzione assistita (CAT-Tool) per combinarla con memorie di traduzione (TM) e banche dati terminologiche. I ruoli si sono quindi rovesciati e la macchina è ora vista come uno strumento che aiuta i traduttori umani ad aumentare la loro produttività (Vieira, 2019).

In scenari di traduzione reali, il post-editing solitamente non è un processo isolato perché, come anticipato, i suggerimenti della TA vengono integrati nei CAT-Tool e combinati con i suggerimenti della TM (cfr. cap. 3.3) (Daems & Macken, 2019). Il post-editing è quindi visto come un'attività con molte sfaccettature, che va oltre la semplice correzione dei risultati della TA e può comprendere diversi compiti (Vieira, 2019).

L'avvento dei sistemi di TA neurali ha portato delle difficoltà per il post-editing. Infatti, la maggiore fluidità dell'output di tali sistemi può rendere più difficile l'identificazione e la correzione degli errori di traduzione (Vieira, 2019).

A seconda del tempo a disposizione e delle esigenze di qualità per il testo di arrivo, si può intervenire nell'output della TA in diversi modi (Riediger, 2018, p. 4):

- *post-editing leggero*: traduzione per usi interni; “il traduttore umano corregge solo gli errori che pregiudicano la comprensione del testo”;
- *post-editing completo*: “traduzione e revisione umana di alta qualità”, destinata alla pubblicazione; “il traduttore umano corregge anche tutti gli errori terminologici e grammaticali e interviene con ritocchi stilistici”.

L'idea di riutilizzare il più possibile l'output della traduzione automatica è un aspetto fondamentale del post-editing (O'Brien, 2022).

Due criteri importanti per la valutazione del post-editing considerati nella maggior parte delle ricerche sono la qualità e lo sforzo (Vieira, 2019). Le ricerche hanno generalmente dimostrato che il post-editing richiede meno sforzo rispetto alla traduzione umana non assistita e non ha alcun effetto negativo sulla qualità del testo di arrivo (Vieira, 2019). Tuttavia, i risultati possono variare a seconda della combinazione linguistica e a causa del fatto che alcuni testi possono essere più o meno adatti al post-editing (Vieira, 2019). Si è dimostrato che il post-editing è spesso (ma non sempre) più veloce rispetto alla traduzione umana da zero e porta a risultati di qualità comparabile o addirittura migliore, anche se l'impressione dei traduttori non corrisponde sempre ai risultati (Gaspari et al., 2014). Uno studio molto attuale (Cui et al., 2023) ha confermato i risultati delle ricerche precedenti. In particolare, è stato mostrato che per la combinazione linguistica inglese-cinese il post-editing richiede meno sforzo rispetto alla traduzione umana e permette di aumentare la produttività e la qualità della traduzione. Questi risultati valgono soprattutto per i testi pubblicitari, informativi e letterari, ma meno per quelli legali (Cui et al., 2023).

Nonostante i risultati incoraggianti delle ricerche, i traduttori professionisti non hanno sempre un'opinione positiva sulla traduzione automatica. Come riportato da Läubli e Green (2019, p. 370), talvolta il post-editing è visto come un processo "noioso e disumanizzante".

Il post-editing può avvenire sia in modo statico (in questa tesi definito post-editing convenzionale) che interattivo. Nel primo caso, l'output della TA viene generato prima e poi corretto in una fase separata (Vieira, 2019). Nell'approccio interattivo, gli errori vengono corretti nello stesso momento in cui viene generato l'output della TA e talvolta si mantiene il nome post-editing anche se l'attività non viene svolta alla fine (O'Brien, 2022). Questo approccio viene illustrato nel prossimo capitolo.

2.2 La traduzione automatica interattiva

In questo approccio alla traduzione i traduttori umani collaborano in modo interattivo con la macchina. Il sistema cerca di auto-completare il testo che il traduttore sta scrivendo e aggiorna costantemente i suggerimenti di traduzione automatica in base a cosa viene digitato (Daems & Macken, 2019). Chi traduce può accettare un suggerimento premendo un tasto oppure rifiutarlo digitando una traduzione alternativa (Knowles et al., 2019) e il sistema reagisce alla decisione aggiustando l'output in tempo reale (O'Brien, 2022).

Si tratta quindi di un processo di previsione e correzione, in cui ogni volta che l'utente rifiuta una parola, il sistema offre una nuova ipotesi di traduzione migliore di quella precedente (Peris et al., 2017). In altre parole, a ogni interazione il sistema genera una nuova ipotesi

di traduzione producendo un suffisso che completa il prefisso già digitato. Chi traduce conferma un prefisso sempre più grande, finché l'ipotesi corrisponde alla traduzione desiderata (Domingo et al., 2017).

Il sistema lavora in modo simile alle tastiere degli smartphone che propongono dei suggerimenti mentre si scrive.

2.2.1 Stato della ricerca

Il primo sistema di TA interattiva, TransType, presentato da Langlais et al. (2000), era in grado di auto-completare parole e brevi espressioni, proponendo suggerimenti in un menu drop-down e aggiornandoli in base a ciò che veniva digitato (Läubli & Green, 2019). Il tool non ha permesso di aumentare la produttività e gli utenti spesso continuavano a digitare le parole anche se venivano già presentate nel menu drop-down (Läubli & Green, 2019). Il primo prototipo era talmente rudimentario che i traduttori l'hanno definito "inutilizzabile" (Langlais et al., 2000, p. 50). Il successore, TransType2 (Esteban et al., 2004), includeva più lingue ed era in grado di auto-completare anche intere frasi (Läubli & Green, 2019).

Caitra, un sistema di TA interattiva sviluppato all'Università di Edimburgo da Koehn e Haddow (2009), offriva delle opzioni di traduzioni alternative oltre all'auto-completamento delle frasi (Läubli & Green, 2019). Per testare la produttività con il programma è stato condotto un esperimento. I risultati hanno mostrato che rispetto al post-editing convenzionale e alla traduzione non assistita, la TA interattiva è sembrata la modalità più utile ai partecipanti, che secondo i risultati, però, sono stati più produttivi con il post-editing convenzionale (Läubli & Green, 2019). Sia con TransType sia con Caitra, gli utenti hanno espresso la loro frustrazione riguardo all'inabilità del sistema di imparare dalle loro correzioni (Läubli & Green, 2019).

Il progetto Casmacat (Sanchis-Trilles et al., 2014) ha seguito Caitra e usava la sua stessa struttura di TA, ma con una nuova interfaccia utente (Läubli & Green, 2019). Il tool offriva anche diversi tipi di visualizzazioni: ad esempio, una stima di quanto l'output della TA fosse affidabile, indicata con diversi colori (Sanchis-Trilles et al., 2014). Questa è stata la caratteristica meno apprezzata dai traduttori che hanno partecipato alle sperimentazioni legate al progetto (Vieira, 2019). Casmacat non era in grado di generare nuovi suggerimenti alla velocità con cui i traduttori digitavano e le proposte arrivate in ritardo spesso venivano ignorate anche se erano corrette (Läubli & Green, 2019). I partecipanti hanno espresso la loro difficoltà ad abituarsi al post-editing interattivo perché il testo di arrivo continuava a cambiare (Läubli & Green, 2019). Come indicato da Vieira (2019), i risultati di uno studio longitudinale con Casmacat (Alabau et al., 2016) hanno anche permesso di confermare che i traduttori diventavano gradualmente più veloci abituandosi alla modalità interattiva, una tendenza che non si è osservata con il post-editing convenzionale.

L'effetto della TA interattiva sulla produttività è stato valutato attraverso studi empirici, nei quali si sono osservati soprattutto il tempo necessario per la traduzione e lo sforzo tecnico (Knowles et al., 2019). Secondo i primi studi di valutazione del tool TransType, la TA interattiva ha portato a riduzioni di produttività fino al 35% rispetto al post-editing convenzionale (Langlais & Lapalme, 2002). Macklovitch (2006) ha scoperto che TransType 2 permetteva di aumentare la produttività del 15-20% rispetto alla traduzione non assistita, producendo testi di qualità comparabile. Koehn e Haddow (2009) hanno mostrato che la TA interattiva permetteva di tradurre in modo più rapido e qualitativo rispetto alla traduzione non assistita, ma non rispetto al post-editing convenzionale (Knowles et al., 2019). Vengono condivisi risultati simili anche da Underwood et al. (2014), Sanchis-Trilles et al. (2014), Alabau et al. (2016) e Alves et al. (2016). Queste ricerche sono state svolte con Casmacat e le combinazioni linguistiche analizzate sono state inglese-spagnolo, inglese-portoghese, inglese-tedesco e francese-inglese (Knowles et al., 2019). In base a queste conclusioni, la TA interattiva statistica non si è dimostrata più veloce del post-editing convenzionale (Knowles et al., 2019).

Altri studi, più recenti, si sono invece concentrati sulla TA interattiva neurale. Knowles et al. (2019) e Daems e Macken (2019) sono stati i primi a condurre studi empirici che hanno osservato la produttività della TA interattiva neurale e in generale il feedback dei traduttori e gli indicatori temporali, tecnici e di qualità hanno mostrato che è una valida alternativa al post-editing convenzionale (Knowles et al., 2019). La ricerca di Knowles et al. (2019), per esempio, ha rivelato che oltre la metà dei partecipanti ha tradotto in modo più veloce con la TA neurale interattiva e l'ha preferita rispetto al post-editing convenzionale.

Di recente, una tesi scritta presso l'Università di Ginevra (Bays, 2022) ha comparato il post-editing convenzionale alla TA interattiva/adattiva di Lilt per la combinazione linguistica inglese-francese. I risultati hanno mostrato che la TA interattiva ha richiesto uno sforzo temporale e tecnico maggiore rispetto al post-editing convenzionale, ma ha permesso di raggiungere una qualità finale migliore. Inoltre, tutti i partecipanti hanno preferito lavorare con la TA interattiva.

L'impatto della TA interattiva sulla qualità dei testi post-editati è stato analizzato meno frequentemente rispetto all'influenza sulla produttività (Vieira, 2019). Come indicato da Vieira (2019), alcune ricerche hanno mostrato che il post-editing interattivo richiede più tempo rispetto al post-editing convenzionale, ma potrebbe portare a testi di qualità maggiore. Underwood et al. (2014), al contrario, hanno constatato che la TA interattiva non ha un effetto significativo sulla qualità del testo di arrivo. Inoltre, la ricerca di Underwood et al. è una delle poche basate su valutazioni di qualità umane per esaminare l'efficacia della TA interattiva (Vieira, 2019). Secondo Vieira (2019) sono necessarie ulteriori ricerche che prendono in considerazione metodi di valutazione umana in questo ambito.

2.3 La traduzione automatica adattiva

Un altro modo con il quale i traduttori lavorano in costante interazione con la macchina è la TA adattiva. Si tratta di un sistema che impara dalle correzioni effettuate e viene continuamente allenato in tempo reale (Daems & Macken, 2019). Quando un traduttore conferma un segmento, la coppia testo di partenza e testo di arrivo viene considerata corretta e usata come esempio per allenare la macchina e ottenere migliori suggerimenti di traduzione futuri. Chi traduce insegna al sistema di traduzione automatica ad adattarsi alle proprie scelte di stile, grammatica e lessico (Lilt Inc., 2023). In questo modo si può evitare che la macchina commetta gli stessi errori più volte all'interno del medesimo documento e ridurre il tempo e lo sforzo necessario per il post-editing. Si affronta così uno dei punti deboli della TA che prima i traduttori trovavano frustrante: continuare a correggere sempre lo stesso errore (O'Brien, 2022).

Il tempo necessario per l'adattamento può variare. La terminologia corretta potrebbe già apparire nel segmento successivo oppure richiedere più tempo (Lilt Inc., 2023). Anche il contesto di un segmento gioca un ruolo importante: il suggerimento adattato viene mostrato con più probabilità se il traduttore ha corretto una parola in un segmento vicino, mentre se la correzione è più lontana nel documento, la traduzione automatica potrebbe tornare a suggerimenti diversi (Lilt Inc., 2023).

Lilt Translate (cfr. cap. 3.4) è l'unico CAT-Tool commerciale che combina la traduzione automatica interattiva con quella adattiva (Martikainen, 2022). Il sistema di TA di Lilt si adatta su due livelli. In primo luogo, Lilt impara dal traduttore umano e con il tempo produce soluzioni sempre migliori, come la maggior parte degli attuali sistemi di TA. In secondo luogo, il sistema adattivo migliora l'output della TA anche durante il processo di traduzione. Questo significa che ogni parola tradotta viene subito presa in considerazione per adattare l'output della TA e cercare prevedere cosa verrà digitato dal traduttore umano (Martikainen, 2022).

Il post-editing convenzionale sembra un compito meno creativo e meno piacevole rispetto alla traduzione. La TA adattiva è stata quindi proposta come un modo per riportare la persona al centro del processo e favorire un'interazione più autentica con la macchina grazie a un'usabilità migliore (Martikainen, 2022). Il processo di traduzione diventa così più dinamico e si differenzia molto dal post-editing convenzionale, più ripetitivo (Martikainen, 2022). La TA adattiva è stata sviluppata appositamente per i processi di traduzione professionale. Tuttavia, al contrario di quanto immaginato inizialmente, questa tecnologia non sembra essersi affermata e ha ricevuto un'attenzione limitata dai ricercatori (Martikainen, 2022).

Martikainen (2022) ha condotto uno studio per analizzare le opinioni di studenti sulla TA adattiva di Lilt. Secondo i risultati, la maggior parte degli studenti che hanno partecipato all'esperimento ha ritenuto la tecnologia di TA interattiva/adattiva promettente, con un adattamento rapido e adeguato. Inoltre, gli studenti hanno apprezzato la riduzione dello sforzo

di digitazione e si sono sentiti meno influenzati dai suggerimenti della TA, perché venivano presentati all'esterno del segmento del testo di arrivo. Si sono però osservati anche degli aspetti negativi: l'adattamento in tempo reale ha diminuito la qualità e introdotto degli errori. Ogni modifica all'output della TA suggerito ha provocato un suggerimento grammaticalmente scorretto o un cambiamento del significato. Nell'esperimento, la qualità dell'output iniziale della TA per la combinazione linguistica inglese-francese era molto scarsa, soprattutto a confronto della concorrenza (ad esempio DeepL o Google Translate). Lo studio ha preso in considerazione diverse combinazioni linguistiche e ha mostrato che la qualità di Lilt dipende proprio da queste. Infatti, per la combinazione inglese-portoghese, la qualità di Lilt è sembrata maggiore a quella della concorrenza (Martikainen, 2022).

In questo capitolo, così come in tutta la tesi, si intende l'adattamento che viene effettuato continuamente, dopo il post-editing di ogni segmento. Esiste anche un altro tipo di adattamento della TA, in cui si personalizza un sistema in base al settore, alla terminologia, allo stile e ad altre esigenze specifiche di un'azienda. In questo caso l'adattamento non avviene in tempo reale dopo ogni segmento confermato, ma in una fase precedente al post-editing. Il sistema viene prima allenato con testi generali e in seguito con una grande quantità di testi paralleli specifici di un settore, ad esempio memorie di traduzione di un'azienda.

3 I CAT-Tool

I sistemi di traduzione assistita (CAT-Tool) sono il tipo di software più usato nel settore della traduzione professionale e della localizzazione (Reinke, 2018). Sono diventati fondamentali per le “tipologie testuali caratterizzate da qualche forma di ripetività (testi tecnici, manuali, istruzioni d’uso)” (Riediger, 2018, p. 6), perché permettono di salvare i testi di partenza con le rispettive traduzioni umane per riutilizzarne delle parti in futuro (Reinke, 2018). In questi programmi il testo viene diviso in coppie di segmenti (lingua di partenza e lingua di arrivo) che vengono registrate in una memoria di traduzione (TM). Quando appare la stessa frase o una simile, il sistema ripropone i segmenti tradotti in precedenza, sotto forma di una corrispondenza esatta (exact match) o parziale (fuzzy match) (Reinke, 2018). Il grado di somiglianza viene indicato in percentuale e corrisponde a una stima di quanto il segmento è simile a una traduzione precedente. Inoltre, “l’applicazione segnala al traduttore con precisione le differenze riscontrate” (Riediger, 2018, p. 6). Chi traduce può accettare la traduzione proposta o modificarla e ciò che conferma viene a sua volta inserito nella memoria di traduzione. Più si lavora con le TM, più “crescono e diventano un patrimonio di lavoro” (Riediger, 2018, p. 6).

I CAT-Tool permettono di aumentare la produttività di chi traduce e assicurano che la terminologia e le espressioni usate siano consistenti, aumentando la qualità della traduzione (Reinke, 2018). Inoltre, hanno il vantaggio di restituire il testo di arrivo con lo stesso layout del testo di partenza (Riediger, 2018).

Possono però anche influenzare in modo negativo la qualità della traduzione: il sistema lavora a livello di frasi e chi traduce potrebbe rischiare di concentrarsi troppo su frasi isolate, senza considerare il contesto circostante (Reinke, 2018). Infatti, in alcuni casi una corrispondenza esatta dalla TM potrebbe suggerire una traduzione sbagliata per un determinato contesto (Reinke, 2018).

I recenti sviluppi e i vari modi di integrare proposte nei CAT-Tool hanno cambiato la qualità e quantità dei suggerimenti e il modo in cui vengono presentati a chi traduce. Il lavoro del traduttore si è quindi spostato dal creare il proprio contenuto a selezionare la migliore opzione tra vari suggerimenti provenienti da diverse fonti (Kappus & Ehrensberger-Dow, 2020). Tuttavia, il traduttore umano rimane al centro del processo di traduzione, perché decide come e quando usare macchine o dati (Mitkov, 2022). Infatti, al contrario della TA, i CAT-Tool lasciano ai traduttori umani il controllo sul processo di traduzione, svolgendo i lavori ripetitivi per loro e permettendo alla traduzione di rimanere un’attività creativa quando è richiesta l’abilità linguistica di una persona (Reinke, 2018).

3.1 Cenni storici

Gli inizi dei CAT-Tool risalgono agli anni 1960, quando la Comunità europea del carbone e dell’acciaio (CECA) ha sviluppato e usato un sistema informatico per recuperare i termini e

il loro contesto da traduzioni umane memorizzate (Reinke, 2018). Era ben diverso dagli attuali sistemi di traduzione assistita e si trattava di un sistema bilingue che serviva principalmente a mostrare i termini della lingua di partenza e i loro equivalenti nella lingua di arrivo nel loro rispettivo contesto (Reinke, 2018).

Nei primi anni 1970, Krollmann, a capo del servizio di traduzione dell'esercito tedesco, ha sviluppato un modello in grado di riutilizzare frammenti di testo identici (oggi gli *exact match*) identificando i passaggi rimasti invariati (Reinke, 2018).

L'idea di un sistema di traduzione assistita di portata più vasta è arrivata alla fine degli anni 1970 dal traduttore della Commissione europea Peter Arthern. Il suo sistema non era solo in grado di identificare frasi identiche come quello di Krollmann, ma anche di trovare frasi simili con le relative traduzioni (oggi i *fuzzy match*) (Reinke, 2018). Così è nato il concetto di memoria di traduzione (TM). Arthern aveva persino già previsto di integrare la traduzione automatica (Reinke, 2018).

Ci è però voluto un altro decennio prima che le idee di Krollmann e Arthern venissero trasformate in veri sistemi commerciali (Reinke, 2018). La registrazione e il recupero automatico degli *exact match* sono stati applicati agli inizi degli anni 1980 da ALPS Inc. con una semplice componente chiamata *Repetitions Processing*, come parte del sistema aziendale *Translation Support System (TSS)* (Reinke, 2018).

Solo a partire dai primi anni 1990 è stato possibile riutilizzare i *fuzzy match*, con i primi sistemi di traduzione assistita per il vasto pubblico come *Translation Manager* di IBM e *Translator's Workbench* di Trados (Reinke, 2018). Tuttavia, a causa dei prezzi alti e degli elevati requisiti hardware, questi sistemi non si sono diffusi subito tra i traduttori autonomi, fino alla comparsa di software economici o gratuiti alla fine degli anni '90 e nel primo decennio del 2000 (Riediger, 2018).

3.2 Componenti e funzionalità

Oltre alla memoria di traduzione, i CAT-Tool dispongono di svariate componenti e funzionalità. Reinke (2018, p. 58–59) ne elenca alcune:

- *Editor multilingue*: permette di visualizzare i testi di partenza e di scrivere le traduzioni in tutti i formati di file rilevanti provenienti da diversi programmi, proteggendo i tag di layout per evitare che vengano cancellati o modificati accidentalmente.
- *Programma di gestione della terminologia*: permette di creare, consultare e aggiornare banche dati terminologiche (TB) in base alla terminologia specifica di un argomento, un cliente o un progetto.
- *Funzione di riconoscimento automatico dei termini*: permette di cercare automaticamente tutti i termini che appaiono nel testo di partenza nella banca dati terminologica.

- *Tool di concordanza*: permette agli utenti di consultare tutti gli esempi di una specifica stringa di ricerca (singole parole, gruppi di parole, frasi) da una TM e di visualizzare queste occorrenze nel loro contesto.
- *Funzione di statistica e analisi*: mostra un conteggio delle parole e dei caratteri nel testo di partenza e dà una panoramica sulla quantità di testo che può essere riutilizzato da una TM per tradurre un nuovo documento. Durante la traduzione appare inoltre una percentuale che indica il progresso.
- *Programma per l'allineamento dei file*: permette di creare delle TM da testi tradotti precedentemente che sono disponibili solo come documenti di testo separati.

Inoltre, i CAT-Tool includono delle funzionalità di controllo della qualità, che comprendono il controllo ortografico e l'identificazione di termini ripetuti, spazi vuoti, formati diversi, tag mancanti. Queste permettono di automatizzare compiti ripetitivi e lasciare più tempo a chi traduce per l'attività di traduzione. Si può anche attivare una modalità di revisione che tiene traccia delle modifiche effettuate e permette al traduttore iniziale di accettarle o rifiutarle con un solo clic. In alcuni CAT-Tool esistono anche dei programmi per la gestione del progetto (Riediger, 2018).

3.3 Integrazione della traduzione automatica nei CAT-Tool

L'integrazione della traduzione automatica nei CAT-Tool sta cambiando il modo tradizionale di svolgere il post-editing. Infatti, i risultati dalla TA vengono presentati al traduttore sotto forma di segmenti, in modo simile alle corrispondenze della memoria di traduzione e, a causa dei recenti miglioramenti nella fluidità della TA neurale, non sono sempre facilmente distinguibili da essi (Kappus & Ehrensberger-Dow, 2020). L'integrazione di diverse tecnologie nei CAT-Tool ha reso meno netti i confini tra le tecnologie e i diversi tipi di assistenza che possono essere usati nel processo di traduzione (Vieira, 2019).

A seconda del CAT-Tool, il sistema di traduzione automatica può essere già integrato oppure può essere collegato tramite plug-in (Kappus & Ehrensberger-Dow, 2020). Spesso per integrare i sistemi di TA nei CAT-Tool è necessaria una licenza a pagamento.

Integrando la traduzione automatica per svolgere il post-editing nel CAT-Tool si ha il vantaggio di poter salvare i segmenti corretti nella memoria di traduzione per poterli usare in futuro. Ci sono però anche degli svantaggi. Infatti, il traduttore riceve molte informazioni da elaborare: il match della TM con la percentuale, il suggerimento della TA, la terminologia, metadati come chi ha creato il match ecc. (O'Brien, 2022). Contemporaneamente deve decidere quale match o suggerimento scegliere nonché cosa modificare e come. Questa interfaccia richiede quindi uno sforzo cognitivo elevato (O'Brien, 2022).

Nel seguente elenco vengono presentate le varie modalità di visualizzare i risultati della TA nei CAT-Tool:

- *TA nel segmento*: quando si crea un nuovo progetto si può attivare la pre-traduzione, che completa automaticamente in background l'intero documento, inserendo in ogni segmento risultati dalla TA o corrispondenze dalla TM (quando la percentuale della somiglianza supera la soglia selezionata) (Kappus & Ehrensberger-Dow, 2020). Chi traduce trova nell'editor un documento già completato da modificare. In questa tesi la modalità viene definita "post-editing convenzionale".
- *TA nella finestra dei risultati di traduzione*: la traduzione automatica può anche essere visualizzata nella finestra dei risultati di traduzione, insieme ai match della TM e ai termini della TB. I suggerimenti della TA vengono contrassegnati con un altro colore, in modo da distinguerli dagli altri risultati, e possono essere inseriti nel testo di arrivo in modo analogo ai match della TM (Kappus & Ehrensberger-Dow, 2020). Prima di iniziare, i segmenti nell'editor rimangono vuoti e la TA viene presentata come suggerimento segmento per segmento. Il traduttore può poi scegliere se inserire il suggerimento nell'editor o digitare un'altra soluzione. In questa tesi la modalità viene definita "suggerimento".
- *TA interattiva*: come visto nel capitolo 2.2, in questa modalità i suggerimenti di TA vengono calcolati e mostrati in tempo reale in base a ciò che si digita.

3.4 Lilt

Lilt è un'azienda che offre servizi e strumenti di traduzione tra cui Lilt Translate, un sistema di traduzione assistita uscito sul mercato nel 2016 (Kappus & Ehrensberger-Dow, 2020) e differente dai CAT-Tool convenzionali usati frequentemente, come ad esempio SDL Trados Studio e memoQ (Daems & Macken, 2019). La differenza principale è l'approccio di traduzione automatica. Infatti, Lilt usa la TA interattiva e adattiva, che suggerisce traduzioni in base alle azioni di chi traduce e impara dalle correzioni effettuate. Il sistema di traduzione automatica viene anche allenato con la memoria di traduzione collegata (Daems & Macken, 2019).

Un'altra differenza è la presentazione dei segmenti sullo schermo. Lilt mostra il testo di partenza e quello di arrivo uno sopra l'altro, in modo verticale invece che orizzontale (Daems & Macken, 2019). Tuttavia, questa modalità può essere cambiata e impostata come nei CAT-Tool convenzionali. I suggerimenti della TA vengono presentati in modo interattivo sotto alla linea in cui appare il testo di arrivo e possono essere accettati direttamente premendo enter (Kappus & Ehrensberger-Dow, 2020).

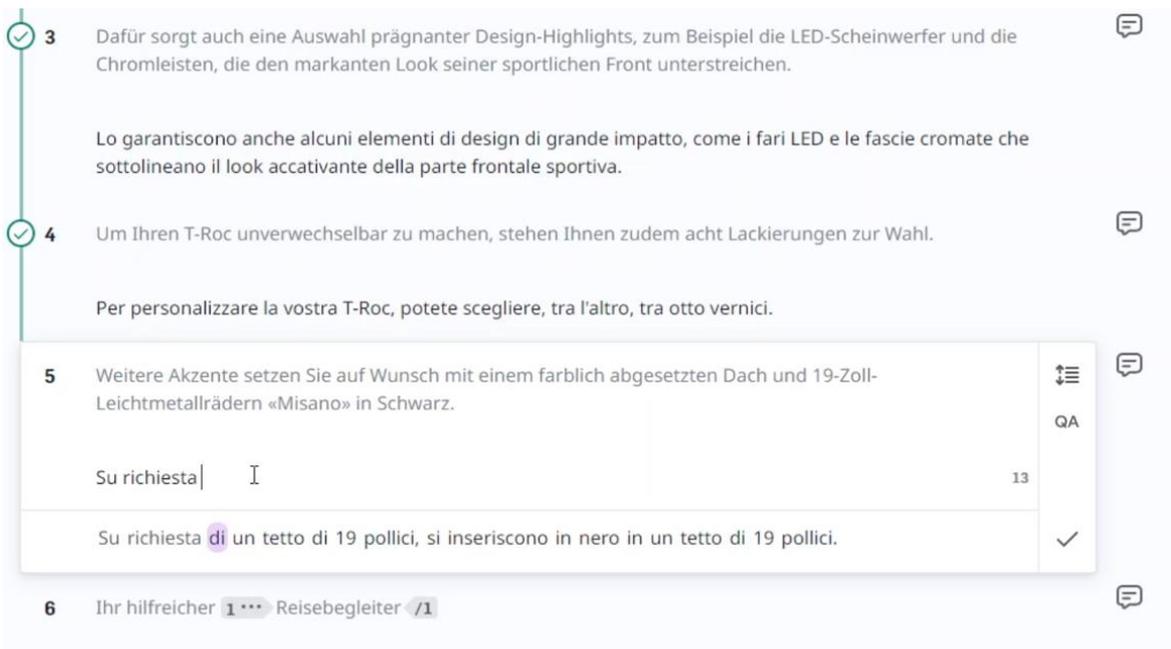


Figura 1: Screenshot interfaccia di Lilt

Il tool si distingue anche per l'interfaccia utente minimalista (Daems & Macken, 2019).

Lilt ha due modelli di traduzione automatica: un modello base, che viene condiviso da tutti i clienti di una combinazione linguistica, e un modello adattivo, che viene usato esclusivamente per un determinato progetto e viene allenato in tempo reale (Lilt Inc., 2023).

3.5 Ergonomia nei CAT-Tool

Svolgere attività complesse guardando per molto tempo lo schermo richiede concentrazione. “Gli aspetti cognitivi ergonomici legati all’uso di tecnologie per la traduzione includono il design, l’organizzazione e il funzionamento dell’interfaccia utente nonché la quantità e la complessità delle funzioni” (Kappus & Ehrensberger-Dow, 2020, p. 387). Più i CAT-Tool sono facili da usare, più chi traduce ha tempo e capacità cognitive per le attività di traduzione. Infatti, un CAT-Tool con funzioni complicate e poco intuitive compromette l’interazione tra uomo e macchina e porta a insoddisfazione (Kappus & Ehrensberger-Dow, 2020). Un CAT-Tool con funzioni che non rispecchiano ciò di cui un traduttore ha bisogno in termini di usabilità può quindi portare a una diminuzione dell’efficienza.

Anche la giusta integrazione della traduzione automatica nel CAT-Tool può contribuire alla soddisfazione e a un aumento di efficienza grazie a una buona usabilità.

Lo studio di Martikainen (2022) ha indicato che oltre all’ergonomia dei CAT-Tool anche la qualità della TA sembra avere un’importante influenza sull’usabilità e sulla soddisfazione dei traduttori.

La ricerca condotta da Moorkens & O'Brien (2017) ha mostrato i desideri dei traduttori in termini di usabilità della TA nei CAT-Tool: vorrebbero vedere l'origine della traduzione suggerita, una pre-elaborazione migliorata nonché dei cambiamenti e miglioramenti dinamici del sistema di TA in caso di errori ricorrenti che devono essere corretti più volte durante il post-editing. Una soluzione a questo desiderio potrebbe essere la traduzione automatica adattiva. Inoltre, i partecipanti vorrebbero vedere in ogni segmento delle percentuali per sapere quanto possono fidarsi dei suggerimenti della TA, vorrebbero che il suggerimento di TA migliore apparisse direttamente nel segmento del testo di arrivo quando non c'è nessun match dalla TM e apprezzerebbero un suggerimento che combina TA e TM (Moorkens & O'Brien, 2017).

Kappus e Ehrensberger-Dow (2020) si sono chiesti se fosse più ergonomico avere il maggior numero possibile di risorse linguistiche nell'area del testo di arrivo (come in Lilt) o in un'area separata (ad esempio in SDL Trados). I risultati della ricerca hanno mostrato che il tool con l'interfaccia utente più semplice (Lilt) è sembrato più ergonomico rispetto a quello con un'interfaccia più complessa (Trados) e la qualità dei testi di arrivo era maggiore, probabilmente perché ci si può concentrare maggiormente sul processo di traduzione (Kappus & Ehrensberger-Dow, 2020). Non è chiaro se i due fattori siano indipendenti uno dall'altro o se l'aumento della produttività traducendo con Lilt lascia più tempo per raggiungere una qualità maggiore (Kappus & Ehrensberger-Dow, 2020).

4 Metodi di ricerca per il processo di traduzione

4.1 Misurazione dello sforzo necessario per il post-editing

Diversi studi precedenti hanno paragonato il post-editing alla traduzione non assistita in termini di sforzo e produttività. Krings (2001) ha stabilito un quadro di riferimento ricorrente per misurare il livello di sforzo richiesto dal post-editing, dividendolo in tre categorie: sforzo temporale, sforzo tecnico e sforzo cognitivo.

I tre tipi di sforzo sono correlati tra loro: lo sforzo temporale viene determinato dalla combinazione di sforzo cognitivo e sforzo tecnico (Daems et al., 2017a). Tuttavia, le relazioni tra i vari tipi di sforzo sono influenzate dalla situazione traduttiva e dalle differenze individuali tra traduttori con diverse competenze o stili (Lacruz, 2017). I seguenti sotto-capitoli presentano in dettaglio i tre tipi di sforzo delineati da Krings (2001, p. 178–179).

4.1.1 Sforzo temporale

Lo sforzo temporale rappresenta il tempo necessario per trasformare l'output della TA in una traduzione di alta qualità. È il tipo di sforzo più importante e il più facile da misurare. Attraverso questo indicatore si può osservare quanto tempo viene risparmiato con il post-editing rispetto alla traduzione non assistita. Si tratta quindi di un indicatore fondamentale perché permette di calcolare la redditività economica della traduzione automatica (Krings, 2001).

4.1.2 Sforzo tecnico

Con sforzo tecnico si intendono tutte le azioni fisiche necessarie per post-editare un testo, come cancellare, inserire, riordinare parole. Questo tipo di sforzo è misurabile in modo diretto (Krings, 2001).

Lo sforzo tecnico viene generalmente indicato con il numero di keystroke (battute) e clic del mouse necessari per produrre il testo di arrivo (Läubli & Green, 2019). Il post-editing è una revisione testuale che richiede un uso intensivo del mouse e della tastiera e può portare a un affaticamento delle mani e dei polsi (O'Brien, 2022). Se l'output è buono, la TA può ridurre la quantità di battute necessarie (O'Brien, 2022).

Inoltre, nel contesto della TA interattiva, lo sforzo di digitazione è un indicatore interessante per misurare quanto spesso i traduttori usano i suggerimenti di traduzione, ad esempio accettando il completamento di una parola invece di digitarla (Läubli & Green, 2019). Questo tipo di sforzo è solitamente rappresentato dal numero di keystroke in relazione al numero di caratteri nel testo di arrivo (Läubli & Green, 2019). Barrachina et al. (2009) si sono riferiti a questo sforzo con il termine *keystroke ratio* e hanno considerato anche il numero

di movimenti del mouse in relazione al numero dei caratteri, definito *mouse-action ratio*, e il totale dei due indicatori (*keystroke and mouse-action ratio*). Le prime ricerche sulla TA interattiva hanno mostrato che lo sforzo di digitazione misurato e quello percepito possono differire (Läubli & Green, 2019).

Lo sforzo tecnico può anche essere misurato attraverso le cosiddette *edit distance metrics*, che misurano il numero minimo di operazioni necessarie per trasformare una parte di testo in un'altra (O'Brien, 2022).

4.1.3 Sforzo cognitivo

Questo tipo di sforzo comprende i processi mentali e il carico cognitivo nella mente del traduttore mentre esegue il post-editing. È difficile da misurare perché riguarda processi nascosti che avvengono nel cervello come leggere, comprendere, comparare il significato del testo di partenza con quello dell'output della TA, prendere decisioni, tenere conto di linee guida e aspettative (O'Brien, 2022). Tuttavia, è un aspetto importante da considerare, perché anche se i traduttori sono più rapidi con la TA, forse si stancano più facilmente rispetto a quando traducono da zero (O'Brien, 2022) e questo a lungo termine può portare a una diminuzione della produttività (Daems et al., 2017b).

Lo sforzo cognitivo può essere misurato in diversi modi. In primo luogo, si può chiedere ai partecipanti di commentare le difficoltà cognitive incontrate; durante l'esperimento oppure in modo retrospettivo (O'Brien, 2022). Un'altra possibilità per analizzare questo tipo di sforzo sono i dati ricavati dall'eye tracking (durata e numero delle fissazioni, dilatazione delle pupille) (O'Brien, 2022).

Come indicato da Jia et al. (2019), anche le pause tra un keystroke e l'altro sono generalmente considerate un indicatore efficace per misurare lo sforzo cognitivo nel processo di traduzione. O'Brien (2006) ha introdotto il concetto di *pause ratio*, che esprime il tempo totale trascorso in pause in relazione al tempo totale necessario per il post-editing. Lacruz et al. (2012) hanno poi scoperto che le correzioni più impegnative a livello cognitivo erano spesso accompagnate da gruppi di pause brevi, che difficilmente avevano un grande impatto sul *pause ratio*. È quindi stato introdotto il concetto di *average pause ratio*, che considera anche queste brevi pause apparentemente importanti. Il valore viene calcolato dividendo la durata media di ogni pausa per la durata media del post-editing per parola (Lacruz, 2017). Più avanti, Lacruz et al. (2014) hanno introdotto una nuova misura efficace quanto l'*average pause ratio* per misurare lo sforzo cognitivo nel post-editing: il *pause to word ratio*, cioè il numero di pause nel post-editing di un segmento diviso il numero di parole nel segmento del testo di partenza. Più alto è il valore, più sforzo cognitivo esercitano i traduttori (Lacruz, 2017).

Jia et al. (2019) hanno fornito una spiegazione sul collegamento tra sforzo cognitivo e post-editing. Secondo loro, quando la TA si rivela sufficientemente buona, probabilmente i partecipanti smettono di dedicare altro sforzo cognitivo alla ricerca di possibili traduzioni alternative. Nella traduzione da zero, invece, i traduttori generano diverse varianti di traduzione e poi ne selezionano una. Con il post-editing si potrebbe quindi risparmiare lo sforzo cognitivo necessario per prendere decisioni quando ci sono varie alternative nonché il tempo necessario per consultare risorse esterne (Jia et al., 2019).

4.1.4 Lo sforzo come indicatore di qualità della TA

Secondo O'Brien (2022), la quantità di sforzo necessario per il post-editing dovrebbe indicare indirettamente la qualità dell'output prodotto da uno specifico sistema di TA, per una combinazione linguistica e un determinato argomento. Pertanto, lo sforzo necessario per il post-editing può essere un modo per valutare la qualità del sistema di TA. Più bassa è la qualità dell'output, più modifiche e tempo saranno necessari. La qualità della TA può essere misurata anche in altri modi, ad esempio attraverso l'identificazione, la classificazione o il conteggio del numero degli errori. Tuttavia, considerare lo sforzo necessario per il post-editing può fornire ancora più informazioni, perché rivela quanto sia facile o difficile lavorare con l'output della TA per produrre un determinato livello di qualità (O'Brien, 2022).

4.2 Screen recording e keystroke logging

Lo screen recording è una tecnica che permette di registrare tutto ciò che accade sullo schermo per raccogliere informazioni qualitative sul processo di traduzione. Si può ad esempio osservare come reagiscono i partecipanti di fronte a diversi incarichi di traduzione.

Un programma di keystroke logging, invece, è un software specifico progettato per osservare i processi di scrittura su un computer. Il programma registra e cronometra i singoli movimenti dei tasti e del mouse per ricostruire e descrivere in modo preciso i processi di produzione testuale (Leijten & Van Waes, 2013). Il software permette anche di identificare le pause tra le azioni consecutive della tastiera o del mouse (Lacruz, 2017).

Entrambi i programmi vengono eseguiti in background e hanno quindi il grande vantaggio di non disturbare i normali processi di traduzione (Ehrensberger-Dow & Massey, 2019). È però importante precisare che viene registrato l'intero schermo di chi traduce, quindi non solo le proprie attività, ma anche eventuali e-mail o comunicazioni con altre persone. Anche se l'anonimità dei dati viene garantita e i partecipanti firmano una dichiarazione di consenso, non è possibile ottenere il consenso delle altre persone con cui interagiscono (Tang et al., 2006).

4.3 Feedback dei partecipanti

L'esperienza utente con le tecnologie di traduzione interattiva è stata spesso misurata attraverso osservazioni e interrogazioni dei partecipanti. Molti studi includono dei sondaggi, perché le osservazioni non sono sufficienti per comprendere appieno il comportamento degli utenti (Läubli & Green, 2019). Nelle ricerche sulla traduzione automatica interattiva sono più comuni i sondaggi al termine della traduzione, perché le interrogazioni durante il processo di traduzione possono influenzare altre variabili, come ad esempio il tempo utilizzato per terminare l'incarico (Läubli & Green, 2019). In diversi studi, i risultati empirici sono in contrasto con le percezioni di chi partecipa allo studio (Läubli & Green, 2019).

Anche se il post-editing si rivela più veloce e di qualità comparabile, è comunque importante che i traduttori siano soddisfatti delle loro prestazioni (Daems et al., 2017b).

4.4 Qualità del prodotto finale

Oltre a osservare il processo di traduzione, è importante considerare la qualità del testo di arrivo, perché il tempo e lo sforzo cognitivo risparmiati attraverso il post-editing valgono la pena solo se non compromettono la qualità del prodotto finale (Jia et al., 2019). Infatti, il tempo risparmiato con il post-editing è rilevante solo se la qualità delle traduzioni prodotte rimane costante e il pubblico di arrivo non riesce a distinguere i testi post-editati da quelli tradotti da zero (Läubli et al., 2013). Se, ad esempio, una funzione di auto-completamento delle frasi permette di aumentare la velocità di traduzione del 30% ma spinge gli utenti a produrre traduzioni peggiori, i risultati ottenuti sono meno significativi (Läubli & Green, 2019).

La qualità dei testi finali può essere valutata sia da persone che da macchine. La valutazione automatica è un'alternativa più veloce e meno costosa a quella umana e consiste nella comparazione della traduzione prodotta a una o più traduzioni di riferimento (Läubli & Green, 2019). BLEU, ad esempio, è la metrica più usata per valutare i sistemi di traduzione automatica e misura la corrispondenza tra la traduzione generata dal sistema di TA e quella prodotta da un traduttore umano (Peris et al., 2017). La valutazione automatica ha lo svantaggio di non prendere in considerazione alcuni fenomeni linguistici come la sinonimia e la parafrasi (Callison-Burch et al., 2006).

Vista la scarsa quantità di ricerche basate su valutazioni umane in questo ambito (cfr. cap. 2.2.1) (Vieira, 2019), in questa tesi si è scelta la valutazione umana per i testi finali. Per mostrare le differenze tra i due sistemi di TA in alcuni esempi, invece, si è fatto ricorso alla valutazione automatica con BLEU (cfr. cap. 6.6.1).

4.5 Confronto tra professionisti e studenti

Come precisato da Kappus e Ehrensberger-Dow (2020), con la costante evoluzione nel campo della traduzione e i grandi progressi della TA neurale, i professionisti delle generazioni meno giovani potrebbero trovarsi in difficoltà di fronte a schermi con un'alta densità di informazioni. Non sarebbe quindi ideale trarre conclusioni solo dalle opinioni e dall'esperienza dei professionisti, senza considerare la flessibilità e le conoscenze tecnologiche dei colleghi più giovani, ad esempio studenti (Kappus & Ehrensberger-Dow, 2020). Il confronto tra i due gruppi permette di vedere queste differenze.

Inoltre, è stato dimostrato che gli studenti di traduzione lavorano in modo diverso rispetto ai traduttori professionisti e anche il loro atteggiamento nei confronti della traduzione automatica è differente (Daems et al., 2017). Infatti, come indicato da Volkart et al. (2022), diversi studi hanno mostrato che quando gli studenti svolgono il post-editing, i testi finali contengono più costruzioni non idiomatiche e grammaticalmente scorrette. Gli studenti tendono a essere più tolleranti nei confronti dell'output della TA e ad accettare anche traduzioni subottimali (Volkart et al., 2022).

5 Descrizione dell'esperimento

5.1 Partecipanti

La partecipazione all'esperimento è stata volontaria. Hanno partecipato sette persone: tre traduttrici professioniste che lavorano rispettivamente nel team interno di un'azienda di e-commerce a Zurigo e in una LSP svizzera a Berlino e quattro studenti all'ultimo semestre del Master in Linguistica applicata alla ZHAW di Winterthur. I partecipanti avevano un'età compresa tra 24 e 35 anni e un'esperienza professionale nella traduzione e nel post-editing che spazia tra 0,5 e 7 anni. Anche gli studenti avevano già una minima esperienza professionale nella traduzione. I partecipanti sono di madrelingua italiana e hanno ottime conoscenze del tedesco.

Complessivamente, la visione dei partecipanti sulla traduzione automatica è positiva. Alla domanda quanto apprezzano usare la TA su una scala da 1 a 10 (1= poco, 10= tanto), quattro persone hanno indicato 7, mentre due 9 e una 6. Nessuno ha dichiarato di vedere la TA come grande minaccia per la professione in futuro. Su una scala da 1 a 5, tre persone la trovano molto utile (5), tre utile (4) e una né utile né inutile (3).

Nella loro quotidianità lavorativa, cinque partecipanti usano la TA come suggerimento quando hanno dei dubbi su una struttura/parola, mentre due basano la loro traduzione su di essa, correggendo le parti errate (post-editing). Inoltre, tre partecipanti hanno indicato di aver già usato un sistema di TA interattiva/adattiva prima.

A causa di limitazioni tecniche e di privacy sui computer aziendali, non era sempre possibile scaricare il programma di keystroke logging Inputlog. Di conseguenza, dei cinque professionisti che inizialmente avevano accettato di prendere parte allo studio, quattro hanno dovuto rinunciare perché avevano un Mac o non avevano un computer privato. Sono poi state trovate altre due persone come sostitute. Inoltre, da due partecipanti la registrazione di Inputlog si è bloccata ed è quindi stato possibile raccogliere dati solo dallo screen recording, dal questionario e dai testi finali. Per questo motivo, l'analisi legata allo sforzo tecnico e allo sforzo cognitivo (cfr. cap. 6.2 e 6.3) considera solo i risultati di cinque partecipanti.

5.2 Incarichi di traduzione

In modo simile allo studio di Jia et al. (2019), si è scelto di lasciare lavorare i traduttori in condizioni il più simile possibile al loro ambiente di lavoro naturale, con accesso alle loro preferenze nel browser e ai loro dizionari. La ricerca in un ambiente di lavoro realistico offre un'alta validità ecologica e permette di studiare fenomeni che riflettono le pratiche reali dei traduttori (Ehrensberger-Dow & Massey, 2019).

Tuttavia, a causa di limitazioni del software Inputlog e di memoQ, disponibili solo per Windows, una partecipante con Mac ha lavorato sul computer di un'altra persona che ha partecipato allo studio. Inoltre, a causa di limitazioni aziendali per il download di programmi, le traduttrici professioniste hanno dovuto lavorare sui loro computer privati.

Tutti i partecipanti hanno integrato il sistema di traduzione automatica DeepL Pro nel CAT-Tool memoQ.

I traduttori hanno ricevuto delle istruzioni scritte su come procedere, che illustravano in dettaglio ogni passo da seguire. Le spiegazioni erano personalizzate, in base all'ordine degli incarichi e alla combinazione di modalità e testo.

I partecipanti hanno ricevuto una breve introduzione scritta a Lilt e hanno avuto la possibilità di familiarizzare con l'interfaccia e il funzionamento del tool con un testo di prova (cfr. allegato).

Dopodiché, a ogni traduttore è stato chiesto di tradurre tre brevi testi dal tedesco all'italiano con le seguenti modalità:

1. TA come suggerimento, all'esterno del segmento (CAT-Tool memoQ): con questa modalità i traduttori potevano inserire attivamente il suggerimento nell'editor se lo consideravano utile oppure potevano direttamente digitare la loro traduzione.
2. TA nel segmento, post-editing convenzionale (CAT-Tool memoQ): attivando la funzione di pre-traduzione, la TA appare direttamente in tutti i segmenti.
3. TA interattiva (CAT-Tool Lilt)

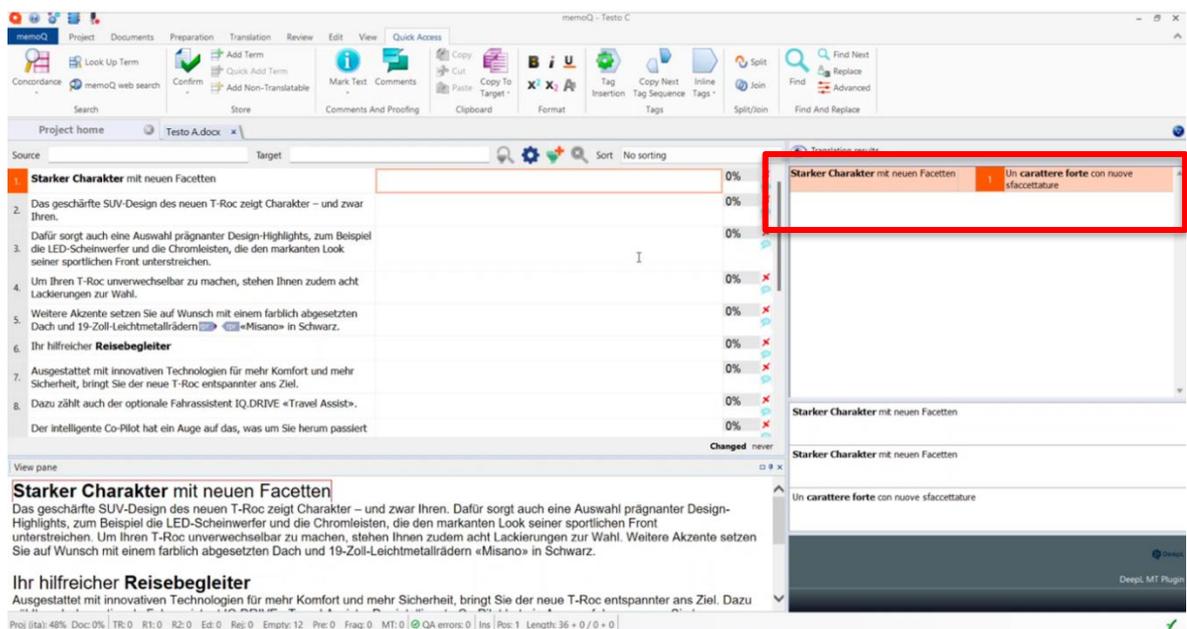


Figura 2: Screenshot modalità 1, TA come suggerimento all'esterno del segmento

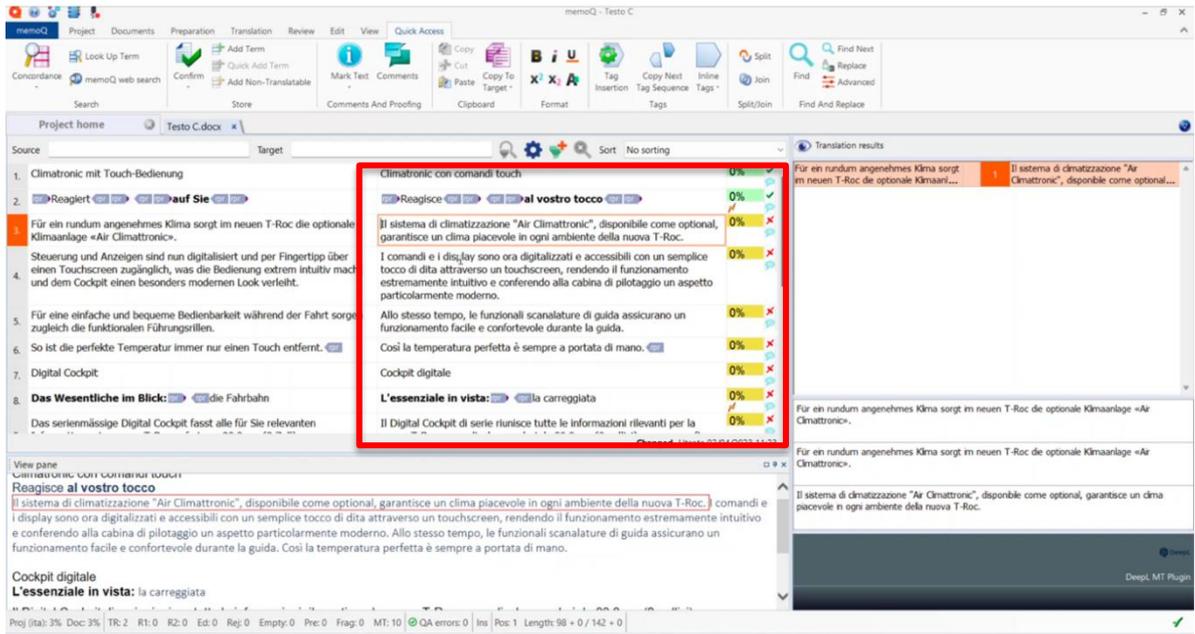


Figura 3: Screenshot modalità 2, TA nel segmento (pre-traduzione)

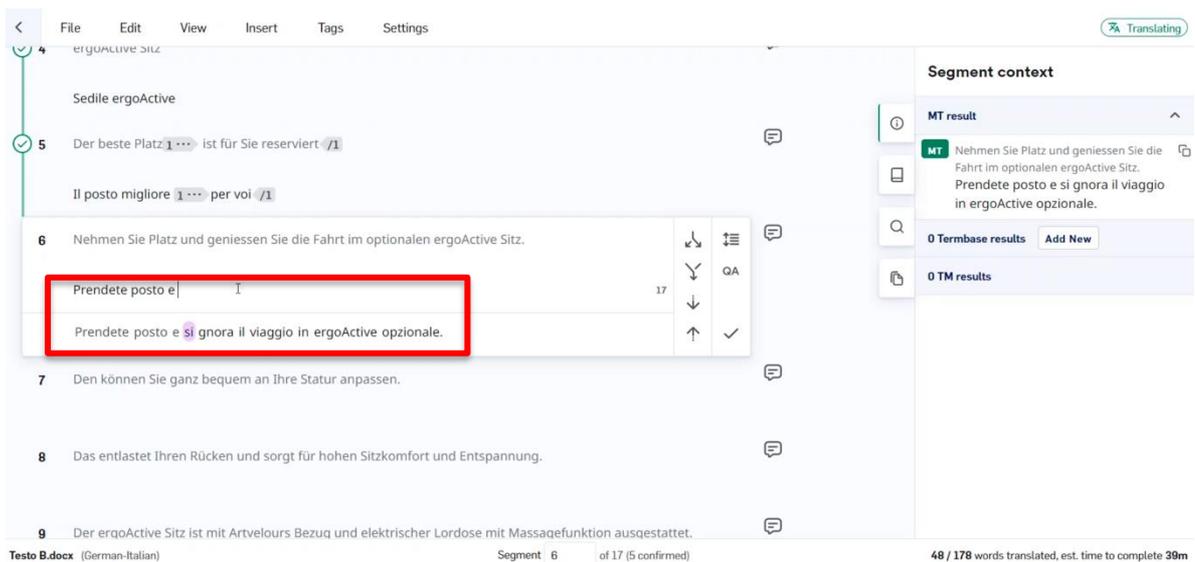


Figura 4: Screenshot modalità 3, TA interattiva

Si è richiesto ai partecipanti di fornire un testo di qualità umana pubblicabile, che corrisponde a un post-editing completo (cfr. cap. 2.1). Siccome i testi automobilistici presenti sul web esistono anche in italiano, nelle istruzioni è stato espressamente chiesto ai partecipanti di non consultare il testo già tradotto. È stato inoltre indicato di svolgere gli incarichi in circa 20-30 minuti, senza però doversi fermare in caso non si abbia terminato. Le istruzioni chiedevano di tradurre i singoli testi senza pause o interruzioni, mentre tra un testo e l'altro era possibile fare delle pause.

Per controllare i potenziali effetti della familiarità e della stanchezza nonché le eventuali difficoltà per un determinato testo, l'ordine degli incarichi e l'assegnazione alla modalità sono stati casuali, secondo la tabella 1:

Partecipante	Categoria	Incarico 1		Incarico 2		Incarico 3	
		Modalità	Testo	Modalità	Testo	Modalità	Testo
P1	Professionista	1	A	2	B	3	C
P2	Studente	2	A	3	B	1	C
P3	Studente	3	A	1	B	2	C
P4	Studente	1	B	3	C	2	A
P5	Studente	2	B	1	C	3	A
P6	Professionista	3	B	2	C	1	A
P7	Professionista	1	C	2	A	3	B

Tabella 1: Assegnazione testi e modalità

5.3 Materiale

Per le traduzioni si sono scelti tre estratti di una pagina pubblicitaria e informativa di un'automobile. I testi contengono sia frasi connotate da tradurre in modo creativo, generalmente più difficili per la TA, sia frasi piuttosto tecniche, relative alle caratteristiche dell'automobile. Nessuno dei partecipanti aveva una particolare esperienza di traduzione nel settore automobilistico.

Si sono scelti testi con una lunghezza e un livello di complessità comparabili:

	Testo A	Testo B	Testo C
Numero di caratteri	1300	1347	1239
Numero di parole	175	180	168
Numero di parole uniche ¹	129	126	122
Densità lessicale (parole uniche/totale)	0.737	0.700	0.726
Lunghezza media per frase (in parole)	12.64	9	13.08
Indice di leggibilità Gunning Fog ²	16.27	18.59	19.10
Indice di leggibilità Flesch ³	46	39	32

Tabella 2: Comparazione dei testi

Nonostante si volesse simulare un ambiente di lavoro il più realistico possibile, è stato deciso di non fornire una memoria di traduzione né una banca dati terminologica per l'esperimento. In primo luogo, perché l'analisi dell'interazione con questi strumenti non rientra negli obiettivi di questa tesi. In secondo luogo, perché si tratta di testi relativamente corti, che anche in un ambiente realistico potrebbero non contenere suggerimenti dalla TM o TB.

¹ Ricavato da: <https://planetcalc.com/3205/>

² Ricavato da: <http://gunning-fog-index.com/index.html>

³ Ricavato da: <https://fleschindex.de/berechnen/>

5.4 Raccolta dei dati

I dati sono stati raccolti tra il 16 marzo e il 2 aprile 2023. Il processo di traduzione è stato registrato con il programma di keystroke logging Inputlog (Leijten & Van Waes, 2013) e il software per lo screen recording OBS Studio. Prima di iniziare, si è chiesto ai partecipanti di scaricare i software sui loro computer.

Inputlog ha permesso di raccogliere dati precisi relativi al tempo impiegato per la traduzione, al numero di keystroke e clic del mouse nonché al numero e alla durata delle pause. I risultati in formato XML sono stati esportati in una tabella Excel nella quale sono stati filtrati secondo i criteri desiderati. I video ricavati con lo screen recording hanno permesso di osservare alcune particolarità del processo di traduzione.

Le opinioni e le impressioni dei partecipanti sono state raccolte tramite un questionario Google, che è stato completato immediatamente dopo gli incarichi di traduzione e conteneva sia domande aperte che domande chiuse. Il questionario era composto da due parti. Nella prima, le domande erano piuttosto generali sulla propria esperienza di traduzione e sull'opinione della traduzione automatica. La seconda parte, invece, conteneva domande specifiche sull'esperimento (cinque domande aperte e quattro domande chiuse), in particolare sulle preferenze, sullo sforzo percepito e sulla qualità raggiunta con le varie modalità (cfr. allegato).

Il questionario indicava che tutte le informazioni sarebbero state trattate con la massima attenzione in termini di protezione dei dati e sarebbero state usate solo in relazione alla ricerca. Si è anche precisato che i riferimenti personali sarebbero stati codificati in modo da non poter risalire a singole persone. Procedendo con il questionario, i partecipanti hanno confermato di avere ricevuto informazioni sul contenuto e sullo scopo del progetto e di acconsentire alla raccolta e all'uso dei dati.

Alla fine dell'esperimento è stato formato un corpus di 21 testi tradotti (sette partecipanti e tre modalità), che sono poi stati inoltrati in forma anonima agli esperti per la valutazione della qualità.

5.5 Valutazione della qualità

La qualità di ogni traduzione è stata valutata da tre professori di traduzione della ZHAW di lingua madre italiana, con un'esperienza approfondita nella correzione di traduzioni dal tedesco all'italiano. I valutatori non hanno ricevuto informazioni sull'origine delle traduzioni, quindi non erano a conoscenza di quali testi erano stati tradotti da chi e con quale modalità.

Per la valutazione delle traduzioni si è fatto riferimento allo schema usato da Läubli et al. (2019), che prevede l'assegnazione di punti su una scala da uno a sei (1 = peggiore, 6 =

migliore) per le seguenti categorie: coerenza, coesione, grammatica, stile, adeguatezza culturale e impressione generale.

I professori hanno ricevuto i testi in formato Word. In ogni documento era già presente una tabella con lo spazio per inserire i punteggi per ogni categoria. Inoltre, è stata aggiunta una casella per eventuali commenti, per ottenere anche dei dati qualitativi oltre a quelli quantitativi ricavati dai punteggi.

Le valutazioni sono poi state inserite in una tabella Excel, nella quale si è calcolata la media e la mediana dei punteggi.

6 Risultati e analisi

6.1 Sforzo temporale

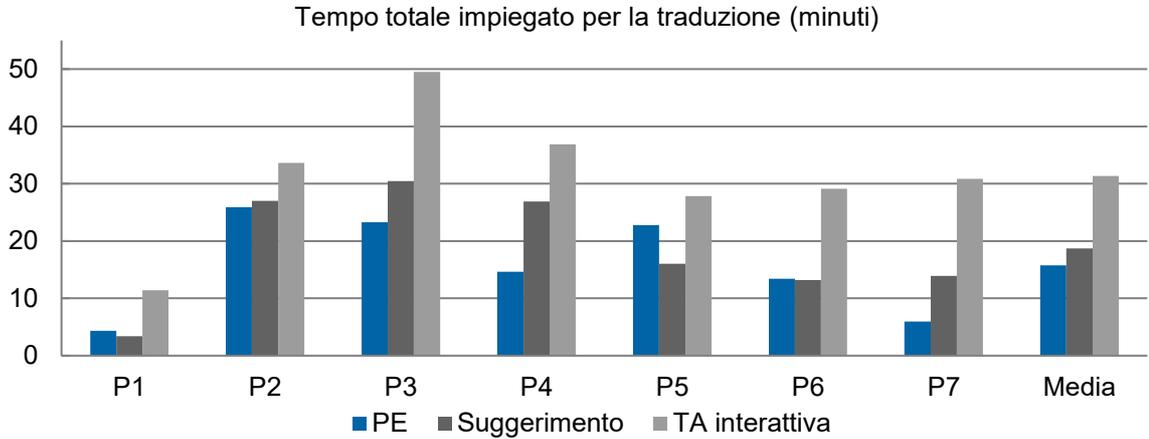


Figura 5: Tempo totale impiegato per la traduzione (minuti)

Come mostra la figura 5, la TA interattiva è stata la modalità con cui tutti i partecipanti hanno impiegato più tempo per produrre il testo di arrivo. In media sono stati necessari 31,3 minuti con la TA interattiva, 18,7 minuti con la TA come suggerimento all'esterno del segmento e 15,7 minuti con il post-editing convenzionale (PE). Con la TA interattiva si è quindi impiegato in media il doppio del tempo rispetto al post-editing convenzionale. Nella modalità con la TA all'esterno del segmento P1 e P6 hanno scelto di inserire il suggerimento ogni volta direttamente nel segmento. Per questo motivo, i valori tra post-editing convenzionale e TA come suggerimento non si differenziano molto.

I testi avevano lunghezze leggermente differenti, quindi è stata calcolata anche la velocità in secondi per parola del testo di partenza.

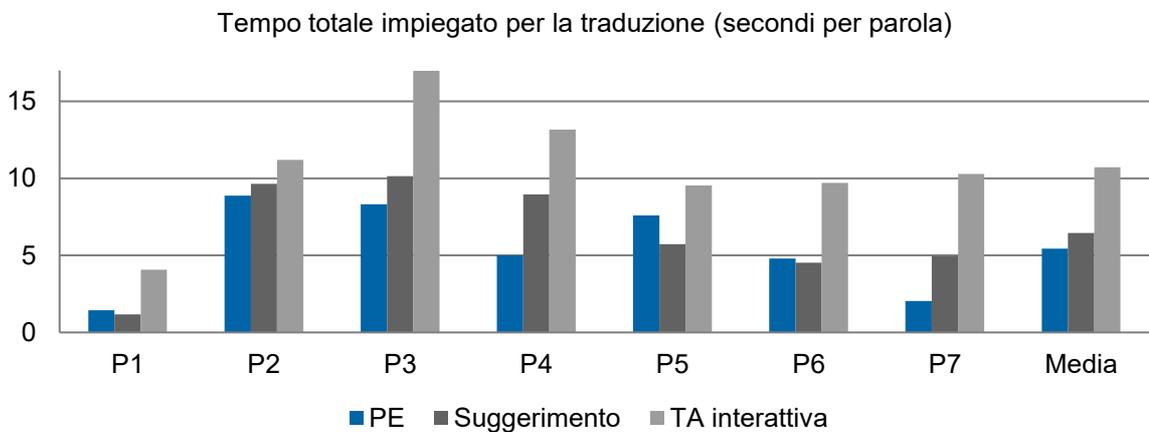


Figura 6: Tempo totale impiegato per la traduzione (secondi per parola)

Come mostra la figura 6, la lunghezza dei testi non ha avuto un particolare impatto sul tempo totale impiegato per la traduzione e mostra dei risultati molto simili alla figura 5.

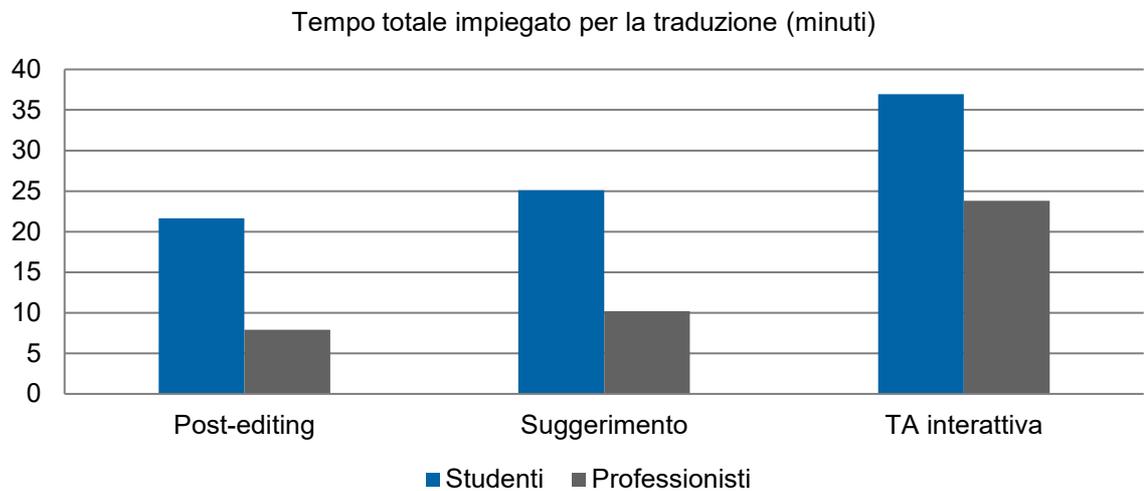


Figura 7: Tempo totale impiegato per la traduzione (confronto tra studenti e professionisti)

Il confronto tra studenti e professionisti (figura 7) mostra che per entrambi i gruppi la TA interattiva è la modalità che ha richiesto più tempo, mentre il post-editing convenzionale è quella che ne ha richiesto meno. Si notano però delle differenze tra una modalità e l'altra: nel post-editing convenzionale e nella TA come suggerimento gli studenti hanno impiegato in media più del doppio di tempo (21,6 e 25,1 minuti) rispetto ai professionisti (7,9 e 10,1 minuti). Nella TA interattiva la differenza è meno marcata, ma esiste comunque (studenti: 36,9 minuti; professionisti: 23,8 minuti).

6.2 Sforzo tecnico

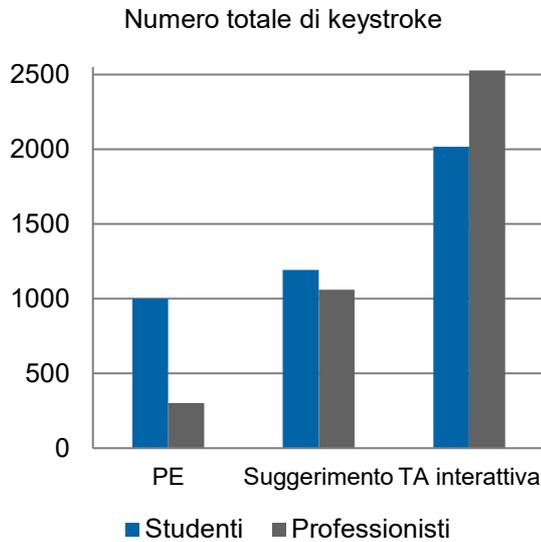


Figura 8: Numero totale di keystroke

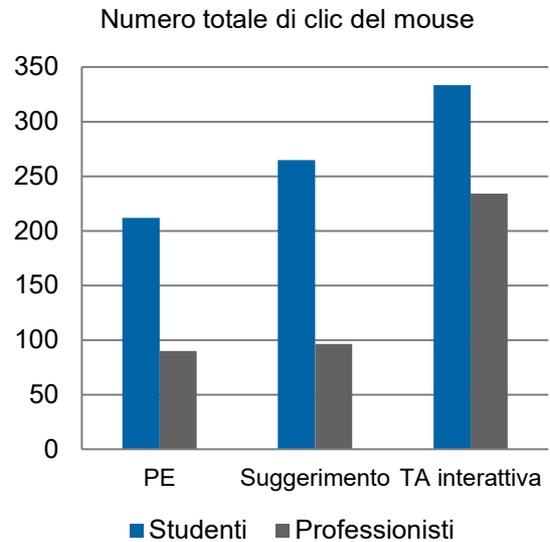


Figura 9: Numero totale di clic del mouse

Le figure 8 e 9 mostrano che in media la TA interattiva ha richiesto più *keystroke* (2221) e clic del mouse (294) per produrre il testo di arrivo, seguita dalla modalità con il suggerimento e dal post-editing convenzionale, con il quale è stata necessaria una quantità tre volte minore di *keystroke* (722) e quasi due volte minore di clic del mouse (163).

I professionisti hanno generalmente esercitato meno sforzo tecnico rispetto agli studenti, tranne per quanto riguarda il numero di *keystroke* con la TA interattiva. La differenza tra i due gruppi è particolarmente evidente nei clic del mouse. Infatti, nelle modalità di post-editing convenzionale e TA come suggerimento gli studenti hanno effettuato in media più del doppio di clic del mouse (212 e 265) rispetto ai professionisti (90 e 97). Anche il numero di *keystroke* effettuati dagli studenti con il post-editing convenzionale (1001) è nettamente superiore rispetto ai professionisti (303).

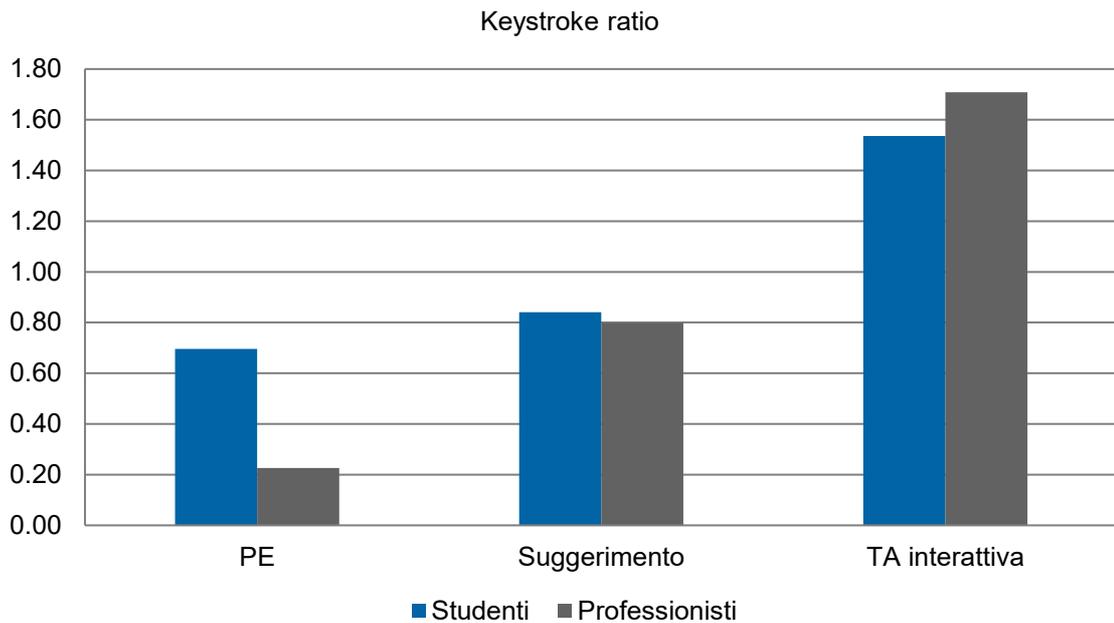


Figura 10: Keystroke ratio

Il *keystroke ratio* (sforzo di digitazione) rappresenta il numero totale di keystroke in relazione al numero di caratteri nel testo di arrivo. I risultati (figura 10) mostrano un andamento simile alla figura 8. Con la TA interattiva, entrambi i gruppi hanno raggiunto un valore superiore a uno: questo significa che il numero di keystroke è superiore al numero di caratteri nel testo di arrivo. Non sembra quindi che con la TA interattiva si sia accettato il completamento di parole o frasi premendo un solo tasto invece di digitarle.

6.3 Sforzo cognitivo

Lo sforzo cognitivo esercitato durante le traduzioni è stato misurato attraverso le pause tra un'azione della tastiera o del mouse e l'altra. I seguenti due grafici mostrano il numero di pause effettuate durante il processo di traduzione.

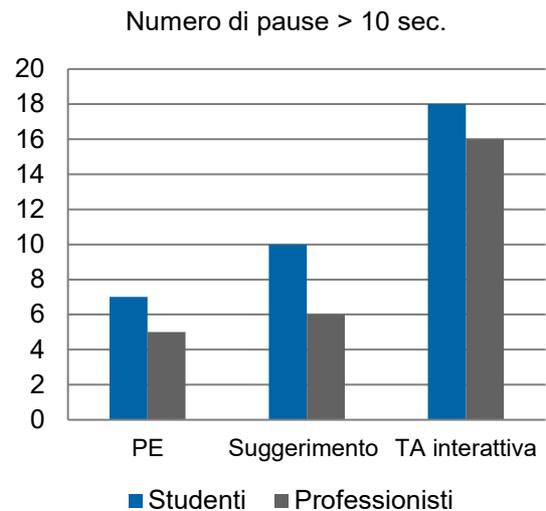
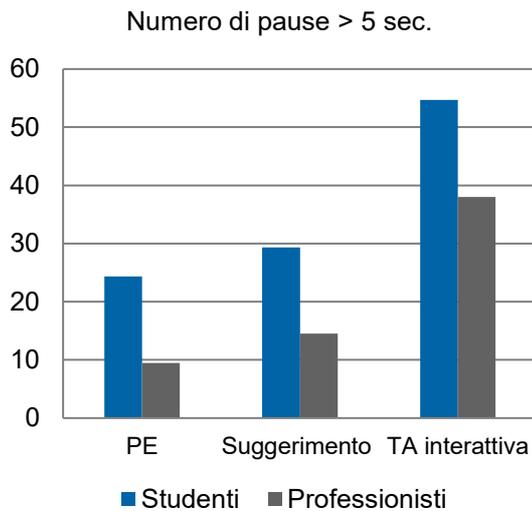


Figura 11: Numero di pause di durata > 5 sec

Figura 12: Numero di pause di durata > 10 sec.

Se si osserva il numero totale di pause della durata superiore a 5 rispettivamente 10 secondi (figure 11 e 12), si nota una chiara tendenza: i partecipanti hanno fatto il maggior numero di pause mentre traducevano con la TA interattiva (media: 48 e 17) e il minor numero di pause durante il post-editing convenzionale (media: 18 e 6). La differenza tra post-editing convenzionale e suggerimento è meno marcata rispetto a quella con la TA interattiva.

In tutte le modalità, gli studenti hanno fatto più pause e hanno quindi esercitato uno sforzo cognitivo maggiore rispetto ai professionisti. Nel post-editing convenzionale, ad esempio, gli studenti hanno registrato più del doppio delle pause superiori a cinque secondi rispetto ai professionisti. Anche in questo caso con il post-editing convenzionale si nota una differenza più marcata tra i due gruppi mentre con la TA interattiva il divario è meno evidente.

Per gli indicatori seguenti, la soglia delle pause è stata posta a cinque secondi. Una pausa è quindi considerata come tale se supera la durata di cinque secondi.

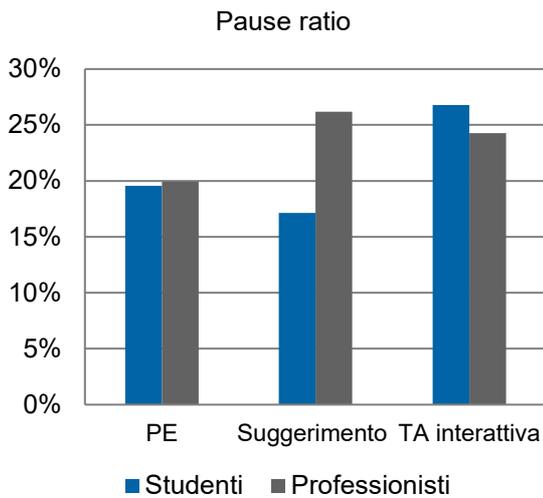


Figura 13: Pause ratio

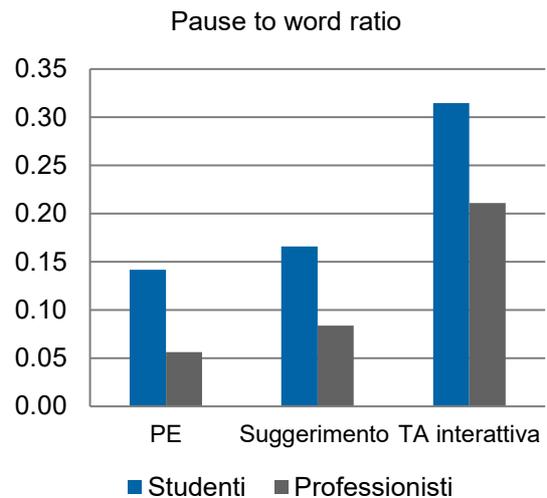


Figura 14: Pause to word ratio

Il *pause ratio* (figura 13) rappresenta il tempo totale trascorso in pause rispetto al tempo totale impiegato per la traduzione. In media, nella modalità di post-editing convenzionale, i due gruppi hanno trascorso il 19,7% del tempo totale in pause. Nella modalità con la TA come suggerimento la percentuale si attesta a 20,8%, mentre con la TA interattiva a 25,8%. Anche secondo questo indicatore, quindi, la modalità che ha richiesto lo sforzo cognitivo maggiore è la TA interattiva. Se si osservano le differenze tra studenti e professionisti, si nota che nel post-editing convenzionale viene raggiunto un valore simile tra i due gruppi, nella modalità con il suggerimento i professionisti hanno registrato una percentuale più alta rispetto agli studenti, mentre nella TA interattiva al contrario.

Il *pause to word ratio* (figura 14), invece, rappresenta il numero totale di pause in relazione al numero totale di parole nel testo di partenza. Anche in questo caso, per entrambi i gruppi la TA interattiva ha richiesto lo sforzo cognitivo maggiore (media: 0,27) e il post-editing convenzionale quello minore (media: 0,11).

6.4 Qualità dei testi di arrivo

La figura 15 indica il livello di qualità raggiunto dai partecipanti con le tre modalità. I risultati di ogni partecipante sono una media di 18 punteggi: tre valutatori e sei categorie. Il punteggio minimo da assegnare era 1 e quello massimo 6. Più alto è il punteggio, migliore è la qualità. A differenza degli indicatori di sforzo, che mostrano un andamento simile tra i partecipanti, la valutazione della qualità ha portato a risultati diversi per ogni persona. Tre partecipanti hanno ottenuto il punteggio migliore con il post-editing convenzionale, mentre tre con il suggerimento. Tuttavia, nessuno ha raggiunto la qualità migliore con la TA interattiva. In complesso, non ci sono differenze marcate tra le tre modalità e la media dei partecipanti indica una qualità leggermente migliore con il post-editing convenzionale.

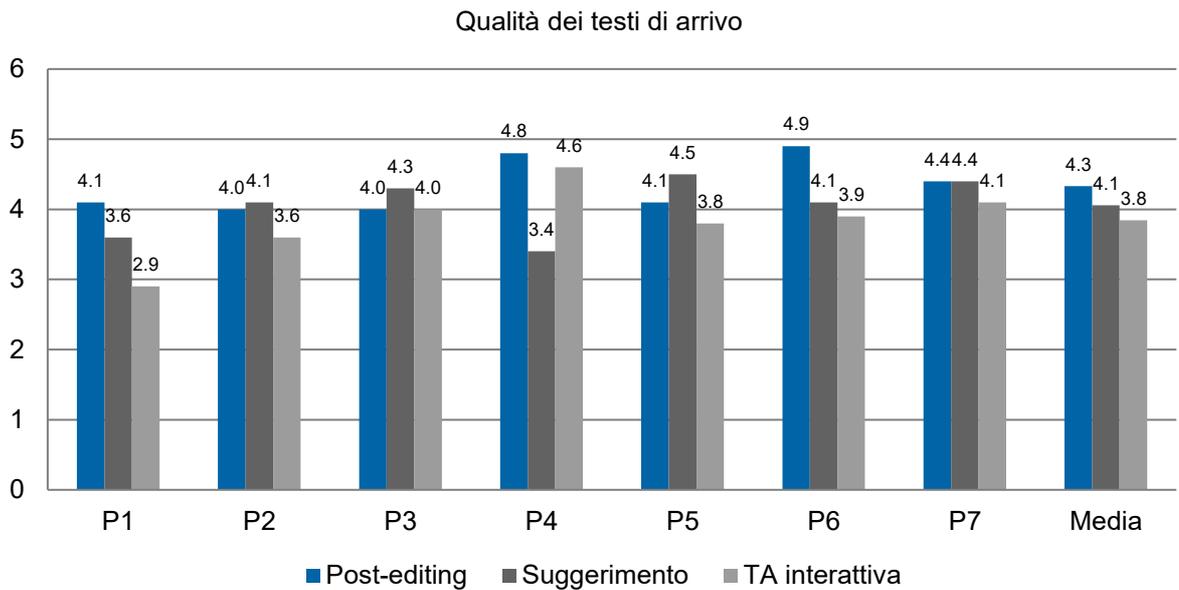


Figura 15: Qualità dei testi di arrivo (media)

La tabella 3 mostra i risultati della valutazione della qualità divisi nelle categorie. Ogni valore è una media di 21 punteggi: sette testi valutati da tre persone.

Modalità	Post-editing	Suggerimento	TA interattiva
Coerenza	4,4	4,3	4,1
Coesione	4,1	3,9	3,9
Grammatica	4,4	4,1	3,7
Stile	4,5	4,0	3,9
Adeguatezza culturale	4,2	4,2	3,9
Impressione generale	4,3	4,0	3,5
Media	4,3	4,1	3,8
Media studenti	4,2	4,1	4,0
Media professionisti	4,5	4,0	3,6

Tabella 3: Valutazione della qualità (media)

Osservando la media dei punteggi di tutte le categorie, si nota che il post-editing convenzionale ha ottenuto una valutazione leggermente superiore alla modalità con il suggerimento (+0,2) e alla TA interattiva (+0,5).

Le differenze più marcate tra le tre modalità appaiono nelle categorie “stile”, “grammatica” e “impressione generale”. Il post-editing convenzionale ha permesso di raggiungere un testo grammaticalmente più corretto (+0,3; +0,7) e uno stile migliore (+0,5; +0,6) rispetto alla modalità con il suggerimento e alla TA interattiva. Anche nella categoria “impressione generale” il post-editing convenzionale ha ottenuto decisamente più punti rispetto alle altre modalità (+0,3; +0,8).

Il confronto tra i due gruppi indica che con il post-editing convenzionale i professionisti hanno ottenuto un punteggio migliore rispetto agli studenti (+0,3), mentre con la modalità con il suggerimento si trovano alla pari (nella tabella il valore è leggermente diverso per arrotondamento). Con la TA interattiva, invece, gli studenti hanno ottenuto un punteggio migliore rispetto ai professionisti (+0,4).

È stata calcolata anche la mediana (tabella 4). Ogni valore è una mediana di 21 punteggi: sette testi valutati da tre persone.

Modalità	Post-editing	Suggerimento	TA interattiva
Coerenza	4,5	4,0	4,5
Coesione	4,0	4,0	4,0
Grammatica	4,5	4,5	4,0
Stile	4,5	4,0	4,0
Adeguatezza culturale	4,0	4,0	4,0
Impressione generale	4,5	4,0	4,0
Mediana	4,5	4,0	4,0
Mediana studenti	4,1	4,3	4,0
Mediana professionisti	4,5	4,0	4,0

Tabella 4: Valutazione della qualità (mediana)

Anche secondo la mediana di tutte le categorie, i testi tradotti con il post-editing convenzionale hanno ottenuto il punteggio migliore (+0,5), mentre quelli tradotti con le altre due modalità sono alla pari. Nelle categorie “coesione” e “adeguatezza culturale” tutte le modalità hanno ottenuto lo stesso punteggio. Il post-editing si è distinto positivamente soprattutto nelle categorie “stile” (+0,5) e “impressione generale” (+0,5). La modalità con il suggerimento e la TA interattiva hanno ottenuto spesso lo stesso punteggio.

Nel post-editing gli studenti hanno raggiunto una qualità peggiore rispetto ai professionisti (-0,4), mentre nel suggerimento migliore (+0,3) e nella TA interattiva uguale.

6.4.1 Altri commenti sulla qualità

Nonostante l’aggiunta di commenti fosse opzionale, nelle valutazioni i professori ne hanno inseriti alcuni, che vengono brevemente riassunti in questo capitolo.

Per i testi tradotti con il post-editing convenzionale si è fatto riferimento principalmente a scelte lessicali non appropriate (es. “aggiungere accenti”, “il davanti sportivo”), a termini non tradotti (es. “co-pilot”, “stop and go”) ed espressioni poco chiare (es. “Il nuovo T-Roc con il suo design SUV particolarmente accentuato”). Altri commenti si legano al tipo di testo: titoli non convincenti e poco incisivi (es. “Le nuove sfaccettature di un carattere forte”, “Fascino sicuro di sé”) ed espressioni non adeguate (tra cui talvolta anche termini troppo tecnici).

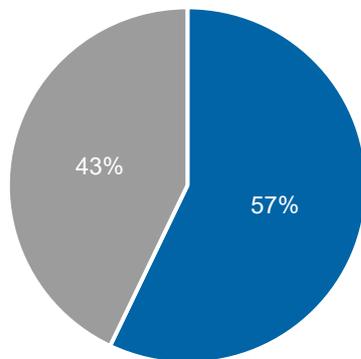
Nei testi tradotti con la TA come suggerimento si è notata più volte la presenza di refusi. Sono inoltre stati indicati errori di accordo, termini non tradotti, scelte lessicali non idiomatiche (es. “elementi di design concisi”, “un fascino incisivo”) e scelte traduttive discutibili (es. “luogo di benessere” per “Wellness-Ausflug” e “riscaldato” per “beheizbar”). In un testo è stata omessa un’informazione. Infine, come nei testi tradotti con il post-editing convenzionale, i titoli sono sembrati poco incisivi (es. “Fascino consapevole”, “Illuminazione seducente”).

Anche nei testi tradotti con la TA interattiva sono stati segnalati diversi refusi e scelte lessicali non idiomatiche (es. “design affinato”, “le moderate luci posteriori”). Inoltre, è stato indicato un errore di traduzione (“cambio” per “Führungsrillen”). Una particolarità notata quasi solo con la TA interattiva sono le omissioni di informazioni (parole o frasi), che sono state rilevate più volte, in diversi testi.

6.5 Opinioni dal questionario

La prima parte del questionario era composta da domande chiuse legate all’esperimento che si riferivano alla modalità preferita, alla velocità, alla qualità raggiunta e allo sforzo cognitivo.

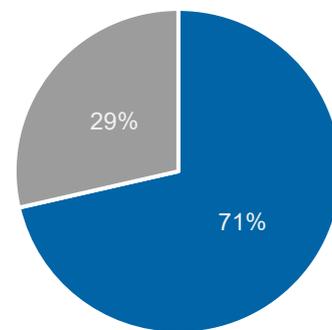
Con quale modalità hai preferito lavorare?



■ PE convenzionale ■ TA come suggerimento

Figura 16: Modalità preferita

Con quale modalità ti è sembrato di lavorare in modo più veloce ed efficiente?



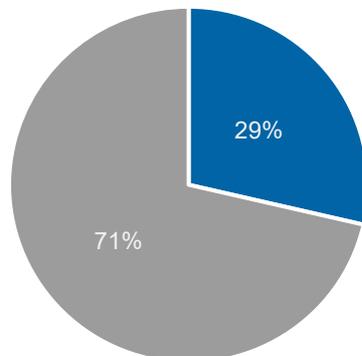
■ PE convenzionale ■ TA come suggerimento

Figura 17: Opinioni sullo sforzo temporale

Come mostra la figura 16, poco più della metà dei partecipanti ha preferito lavorare con la modalità di post-editing convenzionale, quindi con l’output della TA già presente nel segmento. Si nota però ancora una preferenza abbastanza grande anche per la modalità con

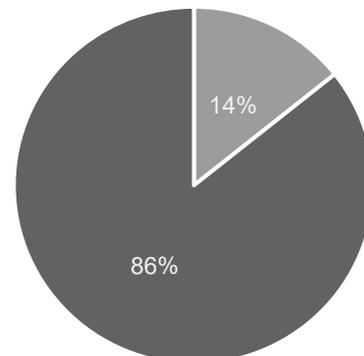
il segmento vuoto. Inoltre, a quasi tre quarti dei partecipanti è sembrato di lavorare in modo più veloce ed efficiente con il post-editing convenzionale (figura 17).

Con quale modalità ti è sembrato di raggiungere una qualità maggiore?



■ PE convenzionale ■ TA come suggerimento

Quale modalità ha richiesto lo sforzo cognitivo maggiore?



■ TA come suggerimento ■ TA interattiva/adattiva

Figura 18: Opinioni sulla qualità raggiunta

Figura 19: Opinioni sullo sforzo cognitivo

Per raggiungere una qualità maggiore, invece, a quasi tre quarti dei partecipanti la modalità più adatta è sembrata la TA come suggerimento all'esterno del segmento (figura 18). Alla domanda quale fosse la modalità che richiede più sforzo cognitivo tutti tranne una persona hanno indicato la TA interattiva (figura 19).

Inoltre, si è chiesto ai partecipanti quale dei tre testi abbiano trovato più difficile da tradurre e perché. Nonostante la combinazione tra testi e modalità fosse diversa tra i partecipanti, tutti hanno indicato che il testo più difficile è stato quello tradotto con Lilt. Il motivo citato maggiormente è il CAT-Tool; in due casi ci si è riferiti anche alla terminologia più difficile e una persona ha indicato anche il fatto di non vedere il layout del testo.

La seconda parte del questionario presentava delle domande aperte sulle opinioni dei partecipanti riguardo ai tre metodi.

Secondo chi ha partecipato allo studio, i vantaggi del post-editing convenzionale con la TA già nel segmento sono stati la velocità, la qualità già abbastanza alta e il fatto che vengono inseriti i tag e la formattazione.

Sono solita fare PE e trovo sia comodo, quindi la funzione in generale mi è piaciuta e la qualità era anche buona. Il risparmio di tempo è sicuramente un vantaggio non indifferente. (P6)

Per quanto riguarda la ricerca terminologica, ci sono state opinioni contrastanti. Una persona ha dichiarato che avere già la TA nel segmento permette di risparmiare tempo nella ricerca di termini tecnici, perché bisogna "solo guardare che siano effettivamente utilizzati" (P3). Un'altra persona, invece, ha indicato che "dove ci sono termini tecnici spesso [la TA]

sbaglia e perdo quasi più tempo a mettere a posto l'output della TA [...] che a tradurre da zero" (P5).

Gli altri svantaggi indicati per il post-editing convenzionale si legano alla limitazione della creatività, a preferenze personali e al tipo di testo:

Spesso sono fuorviata dalla traduzione errata proposta dalla MT oppure mi sento influenzata dalla traduzione già presente e non riesco a trovare soluzioni alternative/più creative. Mi sento spesso limitata. (P7)

[...] non mi piace particolarmente il fatto di ritrovare già una proposta di traduzione davanti a me. (P3)

La TA rende un testo che generalmente è corretto ma che non ha ancora la forma diciamo del testo "pubblicitario", serve quindi un adattamento in quel senso ma non molto di più. (P2)

Le opinioni sui vantaggi della TA come suggerimento all'esterno del segmento erano piuttosto univoche: lascia libertà di pensare a una propria variante, di essere più creativi e di scegliere se consultare la TA quando necessario, ad esempio se ci si sente bloccati.

Per alcuni partecipanti, la TA come suggerimento si è rivelata scomoda:

Ho notato che mi sono ritrovato a importare l'opzione fornita a lato da DeepL nel segmento per fare post editing "tradizionale", perché trovo vada più veloce e poi non sia molto ergonomica come opzione [...]. Per come lavoro io, non avendo una TM e dovendo guardare a destra "solo" per la TA non ha gran senso, meglio importare direttamente la TA nel segmento e partire da lì. (P2)

Essendo fuori dal segmento dovevo sforzarmi a guardare la proposta della MT. (P7)

Un po' scomodo non avere già l'output dentro i segmenti, anche se è come se fossero dei fuzzy. (P6)

Per quanto riguarda la TA interattiva, in questo caso Lilt, tutti hanno espresso dei giudizi negativi. Vari partecipanti hanno condiviso le stesse opinioni e l'accento è stato posto principalmente sulla scarsa qualità della traduzione automatica e sulla lentezza del sistema. Una persona ha inoltre indicato che vedere le frasi e le parole muoversi continuamente l'ha distratta. Quindi, per molti Lilt è stato un ostacolo fuorviante piuttosto che un aiuto. Una persona ha anche dichiarato di aver avuto un "attimo di nervosismo" (P2).

L'output della TA era di pessima qualità e, pur essendo interattivo, non migliorava scrivendo parti di frasi. (P6)

La traduzione proposta era spesso errata, mostrava spesso parole non presenti nel source. Non si adattava a quello che scrivevo in modo coerente. Mi è stato quasi impossibile utilizzarlo. (P7)

La TA in LILT era di pessima qualità, ciò che più che aiutare fuorviava. Trattandosi di un testo commerciale/pubblicitario, con un linguaggio quindi piuttosto figurato e evocativo, se la traduzione automatica non è più che buona rende la vita difficile. Infatti sono ricorso a DeepL in esterna per avere almeno una traduzione accettabile di termini tecnici e quindi dover abbellire solo il linguaggio per rimanere fedele alla funzione evocativa. (P2)

Mi ha disturbato il fatto di vedere le frasi/parole muoversi a seconda di cosa inserivo, perché volevo concentrarmi su quello che scrivevo ma automaticamente guardavo dove la TA continuava a cambiare (mi ha distratto un po'). (P5)

Tuttavia, è stato notato che con il testo di prova assegnato per la fase di familiarizzazione con il tool, un breve testo informativo e pubblicitario sulle funzioni di DeepL, Lilt ha funzionato meglio.

Alcuni partecipanti hanno definito la TA interattiva uno strumento interessante con grandi potenzialità, “che potrebbe rivelarsi utile in ambito traduttivo, se l’output della TA fosse più corretto/riutilizzabile” (P7). Tuttavia, nessuno ha dichiarato di volerla usare in futuro, ma quattro persone hanno indicato che se la qualità delle traduzioni dovesse migliorare la userebbero volentieri.

Ci si è inoltre riferiti alla familiarizzazione con il tool:

Probabilmente è anche una questione di abitudine, ma non penso che mi ci abituerei o comunque con fatica. (P5)

Infine, dal questionario è emersa una nuova idea per integrare la TA nei CAT-Tool:

Se potessi scegliere preferirei avere la TA di DeepL integrata nel segmento come nella App di DeepL, ovvero che cliccando su una parola è possibile vedere alternative e modificarla a piacimento, il tutto integrato in un CAT. (P2)

Si tratta di un desiderio e di una proposta che contiene una parte interattiva, quella di poter cliccare su una parola proposta da DeepL per vedere alternative, ma che rimane comunque piuttosto statica all’interno del CAT-Tool. È sicuramente una funzione che i produttori di CAT-Tool e di TA potrebbero prendere in considerazione e integrare nei loro prodotti in futuro.

6.6 Ulteriori considerazioni

Osservando le registrazioni dello schermo e i risultati del keystroke logging si sono notate altre particolarità legate al processo di traduzione, che vengono riassunte in questo capitolo.

I partecipanti hanno interagito con Lilt in modo diverso. Due persone hanno approfittato della possibilità di accettare direttamente le parole suggerite con un solo tasto o clic invece di continuare a digitarle. Infatti, le istruzioni indicavano che è possibile farlo tramite i tasti “enter” o “shift+enter” per inserire intere frasi. Tuttavia, cinque persone hanno quasi sempre continuato a digitare le parole anche se corrispondevano alla proposta di Lilt. Questo anche per frasi intere, in cui sarebbero bastati alcuni keystroke o clic per accettare i suggerimenti e tradurre in modo più veloce.

Inoltre, come indicato da alcuni partecipanti nel questionario, nello screen recording si vede che Lilt propone nuove soluzioni lentamente. In alcuni casi, chi traduceva ha impiegato quasi meno tempo a digitare che ad aspettare le nuove proposte interattive.

Si è anche constatato che due persone hanno guardato il documento word con il testo di partenza, perché su Lilt non si può vedere la formattazione e l’impaginazione.

Lo screen recording e il keystroke logging hanno anche mostrato che un partecipante ha fatto ricorso a DeepL mentre lavorava con Lilt. Il motivo di questa scelta viene indicato nelle risposte al questionario ed è la scarsa qualità dell’output di Lilt, che viene brevemente approfondita nel prossimo sotto-capitolo.

Nelle traduzioni con Lilt sono state segnalate delle omissioni (cfr. cap. 6.4.1). La causa potrebbe essere il fatto che la parola sia apparsa nei suggerimenti e la persona abbia pensato di averla scritta perché l’ha letta. Nell’esempio riportato (figura 20 e 21) è stato dimenticato “è sempre”.



Figura 20: Screenshot Lilt (prima)

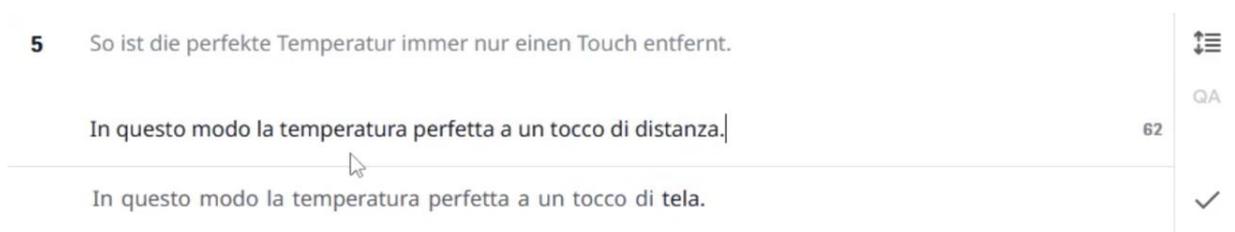


Figura 21: Screenshot Lilt (dopo)

Nella modalità con il suggerimento all’esterno del segmento i partecipanti erano liberi di scegliere come interagire con la TA. Negli screen recording si è visto che due persone

hanno scelto di accettare direttamente tutti i suggerimenti della traduzione automatica inserendoli nel segmento e svolgendo quindi una sorta di post-editing convenzionale. Gli altri cinque partecipanti hanno invece scelto di accettare il suggerimento e inserirlo nell'editor in base al tipo di segmento.

Generalmente, nelle frasi brevi e idiomatiche hanno scelto di digitare una loro soluzione (es. "So ist die perfekte Temperatur nur einen Touch entfernt"; "So wird sogar die Fahrt zur Arbeit zum Wellness-Ausflug"). Bisogna però notare che lo screen recording non fornisce informazioni su dove si trovava lo sguardo dei partecipanti, quindi non si può sapere se hanno guardato il suggerimento della TA mentre digitavano o se hanno scelto una propria variante da zero. Per i segmenti lunghi e con un'alta densità di informazioni, ad esempio con liste e specifiche tecniche, hanno invece inserito il suggerimento della TA direttamente nell'editor svolgendo poi il post-editing.

C'è anche chi ha scelto di copiare solo alcune parti di una frase suggerita da DeepL nell'editor invece di riprendere l'intero segmento.

Tra i vari partecipanti si è notata anche una differenza nel controllo finale: alcuni hanno attivato la funzione di controllo della qualità nel CAT-Tool per individuare eventuali errori ortografici, spazi vuoti, tag mancanti ecc., mentre altri no.

6.6.1 Qualità dell'output della TA

Gli screen recording confermano la scarsa qualità di Lilt indicata nelle risposte al questionario. Alcune frasi non avevano un senso compiuto ed erano praticamente inutilizzabili. C'erano però anche delle frasi tradotte bene, in cui bisognava cambiare solo alcune parti.

In questo capitolo vengono mostrati degli esempi tratti dagli screen recording in cui la differenza tra l'output di Lilt e quello di DeepL Pro è particolarmente evidente. In alcuni casi Lilt proponeva diverse soluzioni a seconda dei partecipanti.

La qualità degli esempi è stata valutata automaticamente con la metrica BLEU⁴ (cfr. cap. 4.4). Più alto è il punteggio, migliore è la qualità della traduzione automatica. Per la valutazione automatica è necessaria una traduzione umana di riferimento; in questo caso è stata usata la traduzione ufficiale in italiano.

Bisogna però ricordare che l'output della TA di Lilt cambia continuamente e quindi una valutazione statica è difficile. È stata scelta la prima frase presentata, ma non si è potuto tenere conto dei cambiamenti apparsi dopo che i traduttori avevano già iniziato a digitare

⁴ Punteggio calcolato attraverso il seguente sito: <https://www.letsmt.eu/Bleu.aspx>.

la frase. Questo è anche il motivo per cui si è calcolato il punteggio solo per alcune frasi, a titolo esemplificativo.

Testo di partenza (B)	Der ergoActive Sitz ist mit Artvelours Bezug und elektrischer Lordose mit Massagefunktion ausgestattet.	Punteggio BLEU
Traduzione ufficiale	Il sedile ergoActive è completo di rivestimento in Artvelours e funzione di massaggio antilordosi elettrico.	
TA Lilt	Il sedile ergoActive è dotato di un servizio di eccellenza di Artvelours e di un potere di controllo elettrolitico.	15,77
TA DeepL Pro	Il sedile ergoActive è dotato di rivestimento Artvelours e di supporto lombare elettrico con funzione di massaggio.	24,69

Tabella 5: Confronto dell'output di Lilt e DeepL Pro (esempio 1)

Testo di partenza (B)	So wird sogar die Fahrt zur Arbeit zum Wellness-Ausflug.	Punteggio BLEU
Traduzione ufficiale	Così anche il viaggio per recarsi al lavoro si trasforma in un momento di benessere.	
TA Lilt	I viaggi per il lavoro per il viaggio di pendolarità sono in realtà i viaggi per la sicurezza.	5,54
TA DeepL Pro	Anche il tragitto in auto per andare al lavoro diventa un'escursione all'insegna del benessere.	7,76

Tabella 6: Confronto dell'output di Lilt e DeepL Pro (esempio 2)

Testo di partenza (A)	Weitere Akzente setzen Sie auf Wunsch mit einem farblich abgesetzten Dach und 19-Zoll-Leichtmetallrädern «Misano» in Schwarz	Punteggio BLEU
Traduzione ufficiale	Su richiesta, potete personalizzare ulteriormente il veicolo con un tettuccio in colore a contrasto e i cerchi in lega leggera da 19 pollici «Misano» neri.	
TA Lilt	Su richiesta di un tetto di 19 pollici, si inseriscono in nero in un tetto di 19 pollici.	4,23
TA DeepL Pro	È possibile creare ulteriori accenti con il tetto in tinta e i cerchi in lega da 19 pollici "Misano" in nero.	18,06

Tabella 7: Confronto dell'output di Lilt e DeepL Pro (esempio 3)

Testo di partenza (B)	Von vorn bis hinten begeistert der neue T-Roc mit seinem neu gestalteten Lichtdesign: Schon von Weitem fallen die charakteristischen LED-Tageslichtscheinwerfer und moderne, abgedunkelte Rückleuchten ins Auge.	Punteggio BLEU
Traduzione ufficiale	La T-Roc conquista da davanti fino alla parte posteriore con il design rivisitato delle luci: già da lontano non passano inosservati i caratteristici fari LED di marcia diurna e le moderne luci posteriori oscurate.	
TA Lilt	Da un lato e dietro è entusiasta, il nuovo T-Roc con il suo design di luce: già da molto tempo si vede i segnali luminosi della LED e i segnali moderati e oscurati di riflessioni.	7,65
TA DeepL Pro	Dalla parte anteriore a quella posteriore, la nuova T-Roc ispira con il suo design luminoso ridisegnato: Anche da lontano, i caratteristici fari diurni a LED e le moderne luci posteriori oscurate catturano l'attenzione.	20,81

Tabella 8: Confronto dell'output di Lilt e DeepL Pro (esempio 4)

Il punteggio complessivo degli esempi presentati è 19,49 per DeepL Pro e 7,13 per Lilt. Si osserva quindi una differenza abbastanza evidente tra i due sistemi.

Inoltre, in alcuni casi Lilt ha proposto dei suggerimenti con nessun collegamento alla lingua di partenza (figure 7, 8 e 9). Dopo aver digitato la parola seguente, però, il suggerimento era di nuovo riferito al resto della frase.

10 Das macht ihn vor allem auf Langstrecken zum idealen Reisebegleiter.

Motivo per il c|

Motivo per il quale la Commissione ha presentato una proposta di direttiva che modifica la direttiva 77/389/CEE del Consiglio.

15

QA

✓

Figura 22: Screenshot Lilt (contesto sbagliato, esempio 1)

8 Das entlastet Ihren Rücken und sorgt für hohen Sitzkomfort und Entspannung.

Il Parlamento europeo ha approvato la relazione che ha presentato il suo progetto di bilancio per il 1992 e che ha approvato la relazione dell'onorevole Woltjer (doc.

0

QA

✓

Figura 23: Screenshot Lilt (contesto sbagliato, esempio 2)

8 Das entlastet Ihren Rücken und sorgt für hohen Sitzkomfort und Entspannung.

È un problema che si pone in modo che i nostri cittadini si sentano sicuri e che si possano trovare i posti di lavoro e che si tratti di un problema di sicurezza sociale.

0

QA

✓

Figura 24: Screenshot Lilt (contesto sbagliato, esempio 3)

Negli ultimi due esempi (figura 23 e figura 24) si può osservare che a seconda del partecipante Lilt ha proposto diverse soluzioni per la stessa frase, entrambe legate a un contesto completamente diverso dal testo di partenza.

7 Discussione

I risultati dello sforzo temporale mostrano che il post-editing convenzionale è stata la modalità più rapida per completare la traduzione (in media 15,7 minuti), mentre la TA interattiva ha richiesto il doppio del tempo (in media 31,3 minuti) e si è collocata in fondo alla classifica. Il post-editing convenzionale ha quindi permesso di raddoppiare la produttività rispetto alla TA interattiva e di aumentarla del 19% rispetto alla modalità con la TA come suggerimento. Bisogna però ricordare che con quest'ultima i partecipanti erano liberi di scegliere come interagire con la TA e alcuni hanno inserito tutti i suggerimenti direttamente nell'editor.

I dati sullo sforzo tecnico indicano un andamento simile. Sia in termini di keystroke, sia di clic del mouse, sia di keystroke ratio, il post-editing convenzionale ha richiesto meno sforzo tecnico rispetto alle altre due modalità, mentre la TA interattiva ha richiesto lo sforzo tecnico maggiore. La TA interattiva ha quindi avuto un effetto contrario di quanto ci si potrebbe aspettare. Con la possibilità di accettare singoli suggerimenti tramite un solo keystroke, lo sforzo di digitazione dovrebbe diminuire significativamente. Tuttavia, come descritto nel capitolo 6.6, la maggior parte delle persone non ha usufruito di questa funzione e ha continuato a digitare le parole anche se corrispondevano ai suggerimenti di Lilt. Il motivo potrebbe essere il periodo di familiarizzazione con il tool troppo breve.

Secondo l'analisi delle pause, anche lo sforzo cognitivo mostra una classifica delle tre modalità in linea con gli altri indicatori di sforzo citati sopra: la TA interattiva ha richiesto lo sforzo cognitivo maggiore, la modalità con il suggerimento si è collocata al centro, mentre con il post-editing convenzionale i partecipanti hanno esercitato lo sforzo cognitivo minore.

Per riassumere, in tutti i tipi di sforzo il post-editing convenzionale ha ottenuto i risultati migliori, mentre la TA interattiva quelli peggiori. La TA come suggerimento si è collocata al centro.

7.1 Relazione tra sforzo e qualità

Anche le valutazioni della qualità dei testi finali mostrano un andamento simile a quello delineato finora. Sia osservando la media che la mediana delle valutazioni, il post-editing ha ricevuto un punteggio leggermente maggiore rispetto alle altre modalità. Considerando le singole categorie, si è osservato che il post-editing convenzionale ha permesso di raggiungere uno stile migliore e ha dato un'impressione generale del testo migliore. Si tratta di due categorie fondamentali per i testi pubblicitari, che richiedono uno stile bello per convincere ancora di più i lettori.

Rispetto alle altre due modalità, il post-editing convenzionale ha quindi permesso di aumentare non solo la produttività ma anche la qualità dei testi tradotti. Infatti, il tempo guadagnato non ha compromesso la qualità dei testi. Pertanto, si può affermare che i risultati sono significativi.

Tuttavia, bisogna notare che le valutazioni spaziano tra 4,5 e 3,5. Avrebbero potuto essere anche molto più distanti tra loro. Inoltre, nessun punteggio medio o mediano ha superato il 4,5. Questo indica che in generale i testi non hanno soddisfatto molto i valutatori. Infatti, nei commenti aggiuntivi si è fatto riferimento a scelte lessicali non idiomatiche e titoli poco incisivi per tutte le modalità. Tale risultato potrebbe essere dovuto alla pressione legata all'esperimento, visto che i partecipanti sapevano che stavano registrando lo schermo.

7.2 Confronto tra opinioni e dati misurati

In diversi studi, i risultati empirici sono stati in contrasto con le impressioni dei partecipanti (Läubli & Green, 2019). In questa tesi, le opinioni dei partecipanti riguardo all'esperimento hanno quasi sempre coinciso con i dati misurati. Il 71% dei partecipanti ha avuto l'impressione di lavorare in modo più rapido ed efficiente con il post-editing convenzionale. Questa opinione rispecchia i risultati ottenuti durante l'esperimento.

Le opinioni sulla qualità raggiunta nel testo finale, invece, hanno coinciso meno con i dati misurati. Infatti, al 71% dei partecipanti è sembrato di ottenere una qualità migliore con la TA come suggerimento all'esterno del segmento, mentre solo il 29% dei partecipanti ha pensato di aver prodotto un testo di arrivo migliore con il post-editing convenzionale. Tale risultato potrebbe mostrare ancora una certa sfiducia nei confronti della traduzione automatica e del post-editing.

Le percezioni sullo sforzo cognitivo, invece, sono in linea con i dati misurati. L'86% dei partecipanti ha avuto l'impressione di esercitare più sforzo cognitivo con la TA interattiva e tutti hanno trovato più difficile il testo tradotto con Lilt. I dati sulle pause confermano queste opinioni.

Inoltre, nelle domande aperte del questionario molti hanno indicato che, a causa della scarsa qualità dell'output della TA, Lilt è stato un ostacolo fuorviante piuttosto che un aiuto. I dati quantitativi sullo sforzo temporale, tecnico e cognitivo con Lilt lo dimostrano.

Si può quindi affermare che i dati oggettivi misurati corrispondono in gran parte ai dati soggettivi raccolti con il questionario.

7.3 Confronto con altre ricerche

È difficile comparare questa tesi alle ricerche passate a causa delle diverse condizioni: sistemi di TA, combinazioni linguistiche, interfaccia dei CAT-Tool, setting e obiettivo

dell'esperimento, metodi di misurazione, livello di esperienza e numero di partecipanti, durata del periodo di raccolta dei dati, tipi di post-editing o traduzione a confronto, tipo e lunghezza dei testi. Tuttavia, in questo capitolo vengono mostrate correlazioni o differenze con ricerche citate nella parte teorica della tesi.

A differenza dello studio condotto da Knowles et al. (2019), in questa ricerca la TA interattiva non si è dimostrata migliore del post-editing convenzionale, né in termini di produttività, né di qualità finale né di soddisfazione dei partecipanti. La causa potrebbe essere la combinazione linguistica. Infatti, Knowles et al. hanno lavorato con la combinazione inglese-spagnolo, mentre questa tesi si è concentrata su tedesco-italiano. Anche la qualità dell'output della TA interattiva potrebbe aver contribuito alla differenza.

I risultati sono invece piuttosto in linea con quelli delle prime ricerche sulla TA interattiva statistica svolte da Langlais et al. (2000), Macklovitch (2006), Koehn e Haddow (2009), Sanchis-Trilles et al. (2014), Underwood et al. (2014), Alabau et al. (2016) e Alves et al. (2016), in cui la TA interattiva non ha permesso di aumentare la produttività rispetto al post-editing convenzionale.

La tesi di Bays (2022) ha ottenuto dei risultati in parte simili e in parte contrastanti a questa ricerca. Con la TA interattiva i partecipanti hanno esercitato uno sforzo temporale e tecnico maggiore rispetto al post-editing convenzionale. Tuttavia, la qualità dei testi finali era migliore con la TA interattiva e tutti i partecipanti hanno preferito lavorare con quest'ultima.

Al contrario di questa tesi, lo studio di Martikainen (2022) ha mostrato che i partecipanti avevano una visione piuttosto positiva della TA interattiva. Tuttavia, come in questa tesi, si è osservato che la qualità dell'output di Lilt era scarsa e i suggerimenti interattivi erano in parte scorretti.

Anche nel processo di traduzione si sono notate delle somiglianze con studi già condotti. Ad esempio, i partecipanti hanno continuato a digitare le parole anche se venivano già presentate nei suggerimenti della TA interattiva, in modo simile a quanto osservato nel progetto TransType (Langlais & Lapalme, 2002) e Casmacat (Alabau et al., 2016). Bisogna però notare che i partecipanti non hanno avuto molto tempo per familiarizzare con il tool e questo potrebbe essere il motivo per cui non hanno usufruito della funzione di accettare una parola suggerita premendo un solo tasto.

In modo simile a quanto riportato da Alabau et al. (2016) in relazione a Casmacat, una persona ha espresso la sua difficoltà a concentrarsi e abituarsi alla TA interattiva perché i suggerimenti continuavano a cambiare.

In alcuni casi, gli screen recording e le risposte al questionario hanno mostrato che il sistema era lento e che i traduttori impiegavano meno tempo a digitare una loro soluzione invece che aspettare i suggerimenti di Lilt. Si era già osservato un andamento simile con

Casmacat, in cui a causa della lentezza del sistema i suggerimenti interattivi venivano ignorati anche se erano corretti (Läubli & Green, 2019).

In base alle ricerche che hanno confrontato il post-editing alla traduzione umana, si può affermare che il post-editing richiede meno sforzo e non ha alcun effetto negativo sulla qualità del testo finale (Vieira, 2019). Tuttavia, questa tesi non ha analizzato la traduzione umana non assistita; infatti, anche nella modalità con il segmento vuoto era presente un suggerimento della TA che i traduttori potevano seguire.

A livello di ergonomia, la ricerca di Kappus e Ehrensberger-Dow (2020) ha mostrato che Lilt, con l'interfaccia utente più semplice, è sembrato più ergonomico rispetto a SDL Trados e ha permesso di raggiungere una qualità maggiore nei testi di arrivo. In questa tesi, secondo le risposte al questionario, il layout e l'interfaccia utente di Lilt, diversi dagli altri CAT-Tool abituali, non sembrano aver avuto una particolare influenza sul processo di traduzione. L'insoddisfazione dei partecipanti è stata piuttosto causata dalla lentezza del sistema e dalla scarsa qualità dei suggerimenti. Tuttavia, una persona ha indicato che il fatto di non vedere il layout del testo di partenza nell'interfaccia di Lilt l'ha disturbata.

7.4 Confronto tra studenti e professionisti

Per quanto riguarda lo sforzo temporale, la differenza tra studenti e professionisti è piuttosto marcata: nella modalità con il suggerimento e nel post-editing convenzionale gli studenti hanno in media impiegato più del doppio di tempo per produrre il testo di arrivo rispetto ai professionisti. Con la TA interattiva la differenza è meno evidente, ma i professionisti sono comunque stati più rapidi degli studenti.

Se si osserva il numero totale di keystroke, che si ripercuote anche sul keystroke ratio, si nota una differenza tra i due gruppi a seconda della modalità. Nella TA interattiva, i professionisti hanno registrato più keystroke degli studenti. Nella modalità con il suggerimento, i due gruppi sono stati quasi alla pari, mentre nel post-editing convenzionale i professionisti hanno effettuato un numero nettamente minore di keystroke (tre volte minore) rispetto agli studenti. Ciò potrebbe suggerire che i professionisti tendono a fare meno modifiche all'output della TA.

Gli studenti hanno inoltre effettuato in media più del doppio di clic del mouse rispetto ai professionisti nelle modalità di post-editing e di TA come suggerimento. Questo indica che i professionisti sono più abituati a usare i shortcut della tastiera al posto del mouse. Come per lo sforzo temporale, con la TA interattiva la differenza di clic del mouse è stata meno evidente tra i due gruppi. Secondo quanto indicato da Volkart et al. (2022), gli studenti tendono a essere più tolleranti nei confronti dell'output della TA e ad accettare anche soluzioni subottimali. I dati relativi allo sforzo tecnico sembrerebbero mostrare il contrario: gli studenti hanno esercitato uno sforzo tecnico maggiore, che dovrebbe corrispondere a più modifiche all'output della TA.

I professionisti hanno registrato un numero minore di pause rispetto agli studenti e hanno quindi esercitato meno sforzo cognitivo. Anche in questo caso il divario è meno marcato con la TA interattiva. Il pause to word ratio ha indicato un andamento simile. Secondo il rapporto tra tempo totale trascorso in pause e il tempo totale impiegato per la traduzione (pause ratio), invece, i risultati sono piuttosto simili tra i due gruppi: con il post-editing convenzionale sono alla pari, con il suggerimento i professionisti hanno raggiunto un pause ratio maggiore, mentre con la TA interattiva hanno registrato uno sforzo cognitivo lievemente inferiore rispetto agli studenti.

Come indicato nei paragrafi precedenti, in tutti i tipi di sforzo la differenza più marcata tra i due gruppi appare nella modalità di post-editing convenzionale. Con la TA interattiva, gli studenti recuperano: il divario è minore e secondo alcuni indicatori gli studenti esercitano meno sforzo rispetto ai professionisti. Questi risultati potrebbero suggerire che i professionisti sono più abituati a svolgere il post-editing in modo rapido e sono quindi in vantaggio rispetto agli studenti, mentre con la TA interattiva entrambi i gruppi hanno avuto un periodo di familiarizzazione piuttosto breve.

I risultati hanno dimostrato che l'esperienza gioca un ruolo importante. Per quanto riguarda l'adattabilità alle nuove tecnologie, i due partecipanti che hanno usufruito della funzione di completamento delle parole con un solo tasto sono studenti. Oltre a questa osservazione non si possono trarre particolari conclusioni sulle conoscenze tecnologiche più avanzate delle generazioni più giovani, visto che i partecipanti avevano un'età massima di 35 anni. Infatti, come indicato da Kappus e Ehrensberger-Dow (2020) (cfr. cap. 4.5) i professionisti delle generazioni meno giovani potrebbero trovarsi in difficoltà con la costante evoluzione delle tecnologie di traduzione. Inizialmente erano stati selezionati anche professionisti di un'età superiore a 35 anni, che potrebbero essere abituati a lavorare in modo più tradizionale, ma hanno dovuto rinunciare a causa dei problemi tecnici legati a Inputlog. Dal questionario è emerso che i partecipanti avevano un'attitudine generalmente positiva nei confronti della TA. Sarebbe stato interessante vedere se i risultati di partecipanti delle generazioni meno giovani e meno favorevoli alla TA fossero stati diversi.

Per quanto riguarda la qualità finale dei testi, i due gruppi hanno ottenuto risultati abbastanza simili. I valori medi indicano una qualità leggermente maggiore per i professionisti nel post-editing, ma minore nella TA interattiva. Con il suggerimento all'esterno del segmento entrambi i gruppi hanno ottenuto quasi lo stesso risultato. Secondo la mediana dei valori, invece, la TA interattiva ha portato a risultati uguali, il post-editing a una qualità migliore per i professionisti, mentre il suggerimento una qualità migliore per gli studenti. Anche le valutazioni della qualità sembrano quindi indicare una tendenza di divario maggiore tra professionisti e studenti per la modalità di post-editing.

Inoltre, nella modalità con il suggerimento, due professionisti hanno consapevolmente deciso di inserire direttamente tutte le proposte di DeepL nell'editor. Questo mostra che secondo la loro esperienza sapevano che avrebbero impiegato meno tempo svolgendo il post-editing convenzionale.

Anche nelle opinioni si sono notate delle differenze tra i due gruppi. Il 75% degli studenti ha preferito la modalità con il suggerimento, mentre il 100% dei professionisti ha preferito il post-editing convenzionale. Inoltre, il 50% degli studenti ha avuto l'impressione di tradurre in modo più rapido con il suggerimento all'esterno del segmento, mentre il 100% dei professionisti ha indicato il post-editing convenzionale.

7.5 Qualità dell'output di Lilt

Come precisato nel capitolo 6.6.1, per la combinazione linguistica tedesco-italiano, spesso la qualità dell'output di Lilt era scarsa. Questo fattore potrebbe aver contribuito a rendere l'incarico di traduzione più difficile e di conseguenza richiedere più sforzo temporale, tecnico e cognitivo. Non si può però trarre una conclusione generale per tutti i tipi di testo. Infatti, nel questionario un partecipante ha dichiarato che con il testo di prova i suggerimenti di Lilt erano migliori. Si trattava di un testo informativo su DeepL, con frasi semplici e piuttosto corte. Il genere e la complessità del testo di partenza possono avere un impatto sulla qualità della TA e, di conseguenza, influenzare il livello di sforzo necessario per il post-editing (Vieira, 2019).

Se la qualità di Lilt fosse stata allo stesso livello di DeepL Pro, probabilmente i risultati sarebbero stati diversi. In una ricerca futura sarebbe opportuno confrontare l'output dei due sistemi di TA prima dell'esperimento per individuare già dall'inizio le differenze di qualità ed eventualmente scegliere un altro tipo di testo, ponderare i risultati in base alla qualità dell'output o usare un sistema di TA interattiva con una qualità simile a quella di DeepL. Bisogna però ricordare che l'output di Lilt cambia costantemente in base a ciò che viene digitato e quindi è difficile basare la valutazione su una frase statica.

La tesi di Bays (2022) ha confrontato la TA interattiva al post-editing convenzionale usando per entrambe le modalità l'output di Lilt. Ha usato lo stesso sistema di TA e quindi la qualità dell'output dovrebbe essere alla pari. Tuttavia, i risultati hanno mostrato che la TA interattiva ha comunque richiesto uno sforzo temporale e tecnico maggiore.

Come indicato da O'Brien (2022) (cfr. cap. 4.1.4), la tesi ha anche mostrato che una qualità peggiore dell'output della TA si ripercuote sugli indicatori di sforzo e talvolta anche sulla qualità del testo finale. Quindi non solo ha permesso di misurare lo sforzo richiesto per le varie modalità, ma anche di mostrare che la qualità dell'output di Lilt è minore per il tipo di testo e la combinazione linguistica analizzati. Sono stati mostrati solo alcuni esempi, nei quali il punteggio BLEU ha confermato che l'output di DeepL Pro aveva una qualità migliore.

Tuttavia, questi dati servono solo a scopo esemplificativo e non possono essere generalizzati per il resto del testo. In futuro, sarebbe opportuno svolgere un'analisi più approfondita. Si potrebbero selezionare alcuni segmenti simili in termini di lunghezza e contenuti e farne valutare la qualità dell'output della TA da uno o più esperti per metterla in relazione allo sforzo temporale, tecnico e cognitivo necessari per quella determinata frase.

7.6 Limitazioni

La ricerca ha diverse limitazioni da considerare in uno studio futuro per ottenere risultati più approfonditi e più rappresentativi.

In primo luogo, l'esperimento è stato condotto con un numero relativamente piccolo di partecipanti. È quindi difficile generalizzare i risultati. Anche il numero di esperti che si sono occupati della valutazione della qualità era basso; i risultati non sono quindi completamente oggettivi. Sarebbe stato utile coinvolgere più valutatori.

Bisogna tenere presente che gli indicatori di sforzo potrebbero essere influenzati non solo dalla modalità, ma anche da altri fattori come la quantità di ricerche effettuate, la velocità individuale, la concentrazione e le competenze di chi traduce. Inoltre, anche se si sono scelti dei testi con lunghezze simili assegnati in modo casuale alle modalità, determinate caratteristiche e difficoltà dei testi di partenza potrebbero aver influito sui risultati.

Sarebbe stato interessante analizzare diversi generi testuali, per sapere se i risultati possono essere generalizzati o se valgono solo per i testi pubblicitari. Siccome i testi scelti contenevano sia parti più tecniche che parti più creative, si sarebbero potute comparare singole frasi di lunghezza e complessità simili per vedere la differenza tra i due tipi di testo in termini di sforzo, qualità raggiunta e opinioni dei partecipanti.

Inoltre, i risultati valgono per la combinazione linguistica tedesco-italiano; in studi futuri sarebbe quindi opportuno includere altre coppie linguistiche per sapere se i risultati sono generalizzabili.

Anche se i partecipanti hanno svolto l'esperimento a casa, in un ambiente di lavoro il più realistico possibile, e attraverso software eseguiti in background che non disturbano il processo di traduzione, potrebbero essersi sentiti sotto pressione perché sapevano che stavano svolgendo un esperimento.

Il programma di keystroke logging Inputlog è stato un grande ostacolo, siccome su alcuni computer non funzionava correttamente o non era permesso scaricarlo per questioni di privacy (nei computer aziendali). In studi futuri sarebbe opportuno controllarne prima il funzionamento o scegliere un software alternativo per non dover cercare dei partecipanti sostituiti a causa di limitazioni tecniche.

Lilt è un CAT-Tool con un'interfaccia e con funzioni diverse da quelle che i partecipanti sono soliti a usare. La scarsa familiarità con il sistema potrebbe avere influenzato i risultati. Per questo motivo, con maggiori risorse temporali e finanziarie sarebbe stato utile includere un periodo di familiarizzazione più lungo oppure svolgere uno studio longitudinale. I risultati della ricerca di Alabau et al. (2016) hanno mostrato che sono necessarie in media sei settimane per abituarsi alla TA interattiva. Dopo questo periodo, i traduttori dovrebbero essere in grado di interagirci in modo efficace. Infatti, si ottiene un aumento della produttività con la TA interattiva solo se si smette di digitare le parole già presenti nei suggerimenti accettandole con una singola battuta (Alabau et al., 2016).

Inoltre, i testi tradotti durante l'esperimento erano troppo brevi per permettere al sistema di TA di adattarsi, quindi i partecipanti non hanno potuto trarre vantaggio dalla funzione adattiva di Lilt.

Infine, come indicato nelle risposte al questionario (cfr. cap. 6.5) e nell'analisi dell'output della TA (cfr. cap. 6.6.1), la qualità dell'output dei due sistemi di TA non era comparabile e potrebbe avere influenzato i risultati.

8 Conclusione

I risultati hanno permesso di rispondere alle domande di ricerca. In primo luogo, si è potuto constatare che il modo migliore per visualizzare la TA nei CAT-Tool è direttamente nel segmento del testo di arrivo (post-editing convenzionale). Secondo i dati raccolti, questa modalità ha richiesto lo sforzo temporale, tecnico e cognitivo minore e ha permesso di aumentare l'efficienza e la produttività, senza porre limiti alla qualità. Infatti, le valutazioni hanno mostrato che con il post-editing convenzionale si è ottenuta una qualità leggermente migliore rispetto alle altre due modalità.

Con la ricerca si sono anche potute individuare le preferenze e le opinioni dei partecipanti riguardo alle tre modalità. Le impressioni soggettive sullo sforzo esercitato hanno in gran parte coinciso con i dati oggettivi misurati. Inoltre, nessuno ha scelto la TA interattiva come modalità preferita, mentre il 57% ha indicato il post-editing convenzionale. È interessante notare che ancora una grande percentuale dei partecipanti (43%) ha preferito lavorare con il suggerimento all'esterno del segmento. Si potrebbe concludere che questa modalità è stata apprezzata per via del genere testuale, che in alcuni punti richiedeva più creatività e in altri presentava contenuti piuttosto tecnici. Infatti, le osservazioni degli screen recording hanno mostrato che per frasi brevi e idiomatiche i partecipanti hanno digitato la loro soluzione, mentre per segmenti più lunghi e con un'alta densità di informazioni (specifiche tecniche) hanno inserito il suggerimento della TA nel segmento (cfr. cap. 6.6).

I dati raccolti suggeriscono che la TA interattiva è stata la modalità peggiore per la scarsa qualità dell'output della TA, la lentezza del sistema e il breve periodo di familiarizzazione con Lilt.

Inoltre, i risultati indicano che complessivamente i professionisti hanno esercitato meno sforzo temporale, tecnico e cognitivo rispetto agli studenti. Tuttavia, la qualità dei testi tradotti dai professionisti è stata solo leggermente migliore di quella degli studenti per la modalità di post-editing, alla pari per quella con il suggerimento e lievemente minore con la TA interattiva.

La ricerca ha mostrato quale modalità permette di lavorare in modo più efficiente e di raggiungere una qualità maggiore. Tuttavia, non si è fatto riferimento alla creatività, un criterio importante soprattutto nei testi pubblicitari. La TA neurale offre risultati sempre più buoni, ma la traduzione umana rimane fondamentale quando è richiesta molta creatività. Nei commenti della valutazione della qualità (cfr. cap. 6.4.1) si è fatto più volte riferimento a titoli non convincenti e poco incisivi, espressioni non adeguate al tipo di testo e scelte lessicali non idiomatiche. Sarebbe interessante scoprire in che modo le tre modalità influenzano e/o limitano la creatività dei traduttori. In uno studio futuro sarebbe quindi opportuno includere anche un'analisi della creatività. Infatti, nelle risposte al questionario proprio la libertà di pensare a una propria variante e di essere più creativi, consultando la TA solo quando

necessario, è stata indicata come vantaggio della modalità con il suggerimento all'esterno del segmento. Vedere la traduzione automatica direttamente nel segmento ha limitato e influenzato chi traduceva, ostacolando la ricerca di soluzioni più creative.

I risultati dovrebbero incoraggiare i (futuri) traduttori a visualizzare la traduzione automatica direttamente nel segmento, quindi pre-traducendo il documento, per aumentare la produttività senza porre limiti alla qualità. Allo stesso tempo, dovrebbero spingere i produttori di sistemi di TA interattiva a lavorare sulla qualità dei suggerimenti, soprattutto per la combinazione linguistica tedesco-italiano. Alcuni partecipanti hanno indicato che se la qualità della TA interattiva fosse migliore, la userebbero volentieri in futuro.

Viste le differenze marcate tra studenti e professionisti, soprattutto con la modalità di post-editing convenzionale, gli studenti dovrebbero svolgere più esercizi di post-editing per esercitarsi anche in questo tipo di approccio e prepararsi alle sfide professionali future.

In seguito a questa tesi dovrebbero emergere altre ricerche che prendono in considerazione le limitazioni di questo studio (cfr. cap. 7.6) per verificare se con altre combinazioni linguistiche e generi testuali si possono ottenere i risultati promettenti per la TA interattiva ottenuti da Knowles et al. (2019). Se in base a ricerche future la TA interattiva dovesse rivelarsi utile, sarebbe opportuno integrarla in più CAT-Tool commerciali per renderla più accessibile e favorire la familiarizzazione con questo tipo di approccio e la ricerca in un ambiente realistico.

Dal questionario è emersa una nuova idea interessante per integrare e visualizzare la traduzione automatica nei CAT-Tool che i produttori potrebbero prendere in considerazione in futuro. Un partecipante vorrebbe visualizzare la TA nel segmento, ma con la possibilità di cliccare sulle parole per vedere delle proposte alternative, come nell'app o sul sito di DeepL.

Infine, in futuro potrebbero esserci possibilità di interazione nei CAT-Tool ancora diverse. Si pensi ad esempio all'intelligenza artificiale che è in costante e rapido sviluppo e potrebbe permettere a chi traduce di riformulare una frase, cercare sinonimi e farsi proporre diverse varianti di traduzione direttamente nel CAT-Tool. Come mostra un post su LinkedIn (Sabani, 2023), nel settore della traduzione emergono già i primi progetti che integrano chatbot basati sull'intelligenza artificiale in CAT-Tool come SDL Trados Studio.

Indice delle abbreviazioni

CAT	Computer Assisted Translation (traduzione assistita dal computer)
LSP	Language service provider (fornitore di servizi linguistici)
PE	Post-editing
TA	Traduzione automatica
TB	Banca dati terminologica
TM	Memoria di traduzione
ZHAW	Università di Scienze applicate Zurigo

Indice delle figure e delle tabelle

Figura 1: Screenshot interfaccia di Lilt	18
Figura 2: Screenshot modalità 1, TA come suggerimento all'esterno del segmento	26
Figura 3: Screenshot modalità 2, TA nel segmento (pre-traduzione)	27
Figura 4: Screenshot modalità 3, TA interattiva.....	27
Figura 5: Tempo totale impiegato per la traduzione (minuti)	31
Figura 6: Tempo totale impiegato per la traduzione (secondi per parola)	31
Figura 7: Tempo totale impiegato per la traduzione (confronto tra studenti e professionisti)	32
Figura 8: Numero totale di keystroke.....	33
Figura 9: Numero totale di clic del mouse	33
Figura 10: Keystroke ratio	34
Figura 11: Numero di pause di durata > 5 sec	35
Figura 12: Numero di pause di durata > 10 sec.	35
Figura 13: Pause ratio	36
Figura 14: Pause to word ratio	36
Figura 15: Qualità dei testi di arrivo (media).....	37
Figura 16: Modalità preferita	39
Figura 17: Opinioni sullo sforzo temporale	39
Figura 18: Opinioni sulla qualità raggiunta	40
Figura 19: Opinioni sullo sforzo cognitivo.....	40
Figura 20: Screenshot Lilt (prima)	43
Figura 21: Screenshot Lilt (dopo).....	43
Figura 22: Screenshot Lilt (contesto sbagliato, esempio 1).....	46
Figura 23: Screenshot Lilt (contesto sbagliato, esempio 2).....	46
Figura 24: Screenshot Lilt (contesto sbagliato, esempio 3).....	46
Tabella 1: Assegnazione testi e modalità	28
Tabella 2: Comparazione dei testi	28
Tabella 3: Valutazione della qualità (media).....	37
Tabella 4: Valutazione della qualità (mediana).....	38
Tabella 5: Confronto dell'output di Lilt e DeepL Pro (esempio 1)	45
Tabella 6: Confronto dell'output di Lilt e DeepL Pro (esempio 2)	45
Tabella 7: Confronto dell'output di Lilt e DeepL Pro (esempio 3)	45
Tabella 8: Confronto dell'output di Lilt e DeepL Pro (esempio 4)	45
Tabella 9: Tempo impiegato per la traduzione (minuti)	65
Tabella 10: Tempo impiegato per la traduzione (secondi per parola).....	65
Tabella 11: Numero di keystroke necessari per la traduzione	65
Tabella 12: Numero di clic del mouse necessari per la traduzione.....	66
Tabella 13: Keystroke ratio (keystroke/caratteri testo di arrivo).....	66
Tabella 14: Numero di pause > 5 secondi	66
Tabella 15: Numero di pause > 10 secondi	67
Tabella 16: Pause ratio	67
Tabella 17: Pause to word ratio.....	67
Tabella 18: Valutazione della qualità dei testi finali (media PE)	68
Tabella 19: Valutazione della qualità dei testi finali (media suggerimento)	68
Tabella 20: Valutazione della qualità dei testi finali (media TA interattiva)	68
Tabella 21: Valutazione della qualità dei testi finali (mediana PE)	69
Tabella 22: Valutazione della qualità dei testi finali (mediana suggerimento)	69
Tabella 23: Valutazione della qualità dei testi finali (mediana TA interattiva)	69

Bibliografia

- Alabau, V., Carl, M., Casacuberta, F., Martínez, M. G., González-Rubio, J., Mesa-Lao, B., Ortiz-Martínez, D., Schaeffer, M., & Sanchis-Trilles, G. (2016). Learning Advanced Post-editing. In M. Carl, S. Bangalore, & M. Schaeffer (A cura di), *New Directions in Empirical Translation Process Research* (pp. 95–110). Springer International Publishing. https://doi.org/10.1007/978-3-319-20358-4_5
- Alves, F., Koglin, A., Mesa-Lao, B., García-Martínez, M., de Lima Fonseca, N. B., de Melo Sá, A., Gonçalves, J. L., Szpak, K. S., Sekino, K., & Aquino, M. (2016). Analysing the Impact of Interactive Machine Translation on Post-editing Effort. In M. Carl, S. Bangalore, & M. Schaeffer (A cura di), *New Directions in Empirical Translation Process Research* (pp. 77–94). Springer International Publishing. <https://doi.org/10.1007/978-3-319-20358-4>
- Barrachina, S., Casacuberta, F., Cubel, E., Lagarda, A., Tomás, J., Vilar, J.-M., Bender, O., Civera, J., Khadivi, S., Ney, H., & Vidal, E. (2009). Statistical Approaches to Computer-Assisted Translation. *Computational Linguistics*, 35(1), 3–28.
- Bays, C. (2022). *Post-édition interactive vs statique: Étude comparative de Liit et COPECO* [Tesi di Master, Université de Genève]. <https://archive-ouverte.unige.ch/unige:163114>
- Callison-Burch, C., Osborne, M., & Koehn, P. (2006). Re-evaluating the Role of Bleu in Machine Translation Research. *11th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 249–256). Association for Computational Linguistics. <https://aclanthology.org/E06-1032>
- Cui, Y., Liu, X., & Cheng, Y. (2023). A Comparative Study on the Effort of Human Translation and Post-Editing in Relation to Text Types: An Eye-Tracking and Key-Logging Experiment. *SAGE Open*, 13(1), 1–15. <https://doi.org/10.1177/21582440231155849>
- Daems, J., & Macken, L. (2019). Interactive adaptive SMT versus interactive adaptive NMT: A user experience evaluation. *Machine Translation*, 33(1–2), 117–134. <https://doi.org/10.1007/s10590-019-09230-z>
- Daems, J., Vandepitte, S., Hartsuiker, R. J., & Macken, L. (2017a). Identifying the Machine Translation Error Types with the Greatest Impact on Post-editing Effort. *Frontiers in Psychology*, 8, 1–15. <https://doi.org/10.3389/fpsyg.2017.01282>
- Daems, J., Vandepitte, S., Hartsuiker, R. J., & Macken, L. (2017b). Translation Methods and Experience: A Comparative Analysis of Human Translation and Post-editing with Students and Professional Translators. *Meta*, 62(2), 245–270. <https://doi.org/10.7202/1041023ar>
- Domingo, M., Peris, Á., & Casacuberta, F. (2017). Segment-based interactive-predictive machine translation. *Machine Translation*, 31(4), 163–185. <https://doi.org/10.1007/s10590-017-9213-3>
- Ehrensberger-Dow, M., & Massey, G. (2019). Translation workplace-based research. In M. O'Hagan (A cura di), *The Routledge Handbook of Translation and Technology* (1^a ed., pp. 354–369). Routledge. <https://doi.org/10.4324/9781315311258>
- Esteban, J., Lorenzo, J., Valderrábanos, A. S., & Lapalme, G. (2004). TransType2 – An Innovative Computer-Assisted Translation System. *Proceedings of the ACL Interactive Poster and Demonstration Sessions* (pp. 94–97). Association for Computational Linguistics. <https://aclanthology.org/P04-3001>
- Gaspari, F., Toral, A., Naskar, S. K., Groves, D., & Way, A. (2014). Perception vs. reality: Measuring machine translation post-editing productivity. *Proceedings of the 11th Conference of the Association for Machine Translation in the Americas* (pp. 60–72). Association for Machine Translation in the Americas. <https://aclanthology.org/2014.amta-wptp.5>
- Jia, Y., Carl, M., & Wang, X. (2019). How does the post-editing of neural machine translation compare with from-scratch translation? A product and process study. *The Journal of Specialised Translation*, 31, 60–86.
- Kappus, M., & Ehrensberger-Dow, M. (2020). The ergonomics of translation tools: Understanding when less is actually more. *The Interpreter and Translator Trainer*, 14(4), 386–404. <https://doi.org/10.1080/1750399X.2020.1839998>
- Knowles, R., Sanchez-Torron, M., & Koehn, P. (2019). A user study of neural interactive translation prediction. *Machine Translation*, 33(1–2), 135–154. <https://doi.org/10.1007/s10590-019-09235-8>
- Koehn, P., & Haddow, B. (2009). Interactive Assistance to Human Translators using Statistical Machine Translation Methods. *Proceedings of Machine Translation Summit XII: Papers*. <https://aclanthology.org/2009.mtsummit-papers.8>

- Krings, H. P. (2001).** Repairing texts: Empirical investigations of machine translation post-editing processes. The Kent State Univ. Press.
- Lacruz, I. (2017).** Cognitive Effort in Translation, Editing, and Post-editing. In J. W. Schwieter & A. Ferreira (A cura di), *The Handbook of Translation and Cognition* (1^a ed., pp. 386–401). Wiley. <https://doi.org/10.1002/9781119241485.ch21>
- Lacruz, I., Denkowski, M., & Lavie, A. (2014).** Cognitive demand and cognitive effort in post-editing. *Proceedings of the 11th Conference of the Association for Machine Translation in the Americas* (pp. 73–84). Association for Machine Translation in the Americas. <https://aclanthology.org/2014.amta-wptp.6>
- Lacruz, I., Shreve, G. M., & Angelone, E. (2012).** Average Pause Ratio as an Indicator of Cognitive Effort in Post-Editing: A Case Study. *Workshop on Post-Editing Technology and Practice*. Association for Machine Translation in the Americas. <https://aclanthology.org/2012.amta-wptp.3>
- Langlais, P., Foster, G., & Lapalme, G. (2000).** TransType: A Computer-Aided Translation Typing System. *ANLP-NAACL 2000 Workshop: Embedded Machine Translation Systems* (pp. 46–51). <https://aclanthology.org/W00-0507>
- Langlais, P., & Lapalme, G. (2002).** Trans Type: Development-Evaluation Cycles to Boost Translator's Productivity. *Machine Translation*, 17(2), 77–98. <https://doi.org/10.1023/B:COAT.0000010117.98933.a0>
- Läubli, S., Amrhein, C., Düggelein, P., Gonzalez, B., Zwahlen, A., & Volk, M. (2019).** Post-editing Productivity with Neural Machine Translation: An Empirical Assessment of Speed and Quality in the Banking and Finance Domain. *Machine Translation Summit XVII* (pp. 267–272). European Association for Machine Translation. <https://doi.org/10.5167/UZH-179869>
- Läubli, S., Fishel, M., Massey, G., Ehrensberger-Dow, M., & Volk, M. (2013).** Assessing post-editing efficiency in a realistic translation environment. *MT Summit XIV Workshop on Post-editing Technology and Practice* (pp. 83–91). European Association for Machine Translation. <https://doi.org/10.5167/UZH-80891>
- Läubli, S., & Green, S. (2019).** Translation technology research and human–computer interaction (HCI). In M. O'Hagan (A cura di), *The Routledge Handbook of Translation and Technology* (1^a ed., pp. 370–383). Routledge. <https://doi.org/10.4324/9781315311258>
- Leijten, M., & Van Waes, L. (2013).** Keystroke Logging in Writing Research: Using Inputlog to Analyze and Visualize Writing Processes. *Written Communication*, 30(3), 1–35. <https://doi.org/10.1177/0741088313491692>
- Lilt Inc. (2023).** *What is Contextual AI?* <https://support.lilt.com/kb/what-is-interactive-adaptive-mt>
- Macklovitch, E. (2006).** TransType2: The Last Word. *Proceedings of the Fifth International Conference on Languages Resources and Evaluation (LREC 06)* (pp. 167–172). European Language Resources Association. <https://aclanthology.org/L06-1008/>
- Martikainen, H. (2022).** *Ghosts in the machine: Can adaptive MT help reclaim a place for the human in the loop?* hal-03548696v3. <https://hal.science/hal-03548696v3>
- Mitkov, R. (2022).** Translation memory systems. In S. Deane-Cox & A. Spiessens (A cura di), *The Routledge Handbook of Translation and Memory* (1^a ed., pp. 364–380). Routledge. <https://doi.org/10.4324/9781003273417>
- Moorkens, J., & O'Brien, S. (2017).** Assessing user interface needs of post-editors of machine translation. In D. Kenny & J. Williams (A cura di), *Human Issues in Translation Technology* (pp. 109–130). Routledge.
- O'Brien, S. (2006).** Pauses as Indicators of Cognitive Effort in Post-editing Machine Translation Output. *Across Languages and Cultures*, 7(1), 1–21. <https://doi.org/10.1556/Acr.7.2006.1.1>
- O'Brien, S. (2022).** How to deal with errors in machine translation: Post-editing. In D. Kenny (A cura di), *Machine translation for everyone: Empowering users in the age of artificial intelligence* (pp. 105–120). Language Science Press. <https://doi.org/10.5281/zenodo.6653406>
- Pérez-Ortiz, J. A., Forcada, M. L., & Sánchez-Martínez, F. (2022).** How neural machine translation works. In D. Kenny (A cura di), *Machine translation for everyone: Empowering users in the age of artificial intelligence* (pp. 141–164). Language Science Press. <https://doi.org/10.5281/zenodo.6760020>
- Peris, Á., Domingo, M., & Casacuberta, F. (2017).** Interactive neural machine translation. *Computer Speech & Language*, 45, 1–21. <https://doi.org/10.1016/j.csl.2016.12.003>
- Reinke, U. (2018).** State Of The Art In Translation Memory Technology. In G. Rehm, F. Sasaki, D. Stein, & A. Witt (A cura di), *Language Technologies For A Multilingual Europe* (pp. 55–84). Language Science Press. <https://doi.org/10.5281/ZENODO.1291930>
- Riediger, H. (2018).** *Tradurre col computer: Dalla traduzione automatica al CAT e ritorno*. <http://www.fondazionemilano.eu/blogpress/weaver/?wpdmact=process&did=MTEuaG90bGluaw==>, PDF, (consultato il 1.6.23).

- Sabani, A. (marzo 2023).** *I know, I know, you've probably heard enough about ChatGPT. But, This could be of interest to all you translators.* [Post]. LinkedIn. https://www.linkedin.com/posts/arben-sabani-339074128_translation-project-translators-ugcPost-7039991834646892544-s1id?utm_source=share&utm_medium=member_desktop
- Sanchis-Trilles, G., Buck, C., Carl, M., Casacuberta, F., García-Martínez, M., Germann, U., González-Rubio, J., Hill, R. L., Koehn, P., Leiva, L. A., Mesa-Lao, B., Ortiz-Martínez, D., Saint-Amand, H., Tsoukala, C., Vidal, E., & Alabau, V. (2014).** Interactive translation prediction versus conventional post-editing in practice: A study with the CasMaCat workbench. *Machine Translation*, 28(3–4), 217–235. <https://doi.org/10.1007/s10590-014-9157-9>
- Tang, J. C., Liu, S. B., Muller, M., Lin, J., & Drews, C. (2006).** Unobtrusive but invasive: Using screen recording to collect field data on computer-mediated interaction. *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work - CSCW '06* (pp. 479–482). <https://doi.org/10.1145/1180875.1180948>
- Underwood, N., Mesa-Lao, B., Martínez, M. G., Carl, M., Alabau, V., González-Rubio, J., Leiva, L. A., Sanchis-Trilles, G., Ortíz, D., & Casacuberta, F. (2014).** Evaluating the effects of interactivity in a post-editing workbench. *Proceedings of the Ninth International Conference on Language Resources and Evaluation* (pp. 553–559). European Language Resources Association. <https://aclanthology.org/L14-1261/>
- Vieira, L. N. (2019).** Post-editing of machine translation. In M. O'Hagan (A cura di), *The Routledge Handbook of Translation and Technology* (1ª ed., pp. 319–335). Routledge. <https://doi.org/10.4324/9781315311258>
- Volkart, L., Girletti, S., Gerlach, J., Mutal, J. D., & Bouillon, P. (2022).** Source or target first? Comparison of two post-editing strategies with translation students. *Journal of Data Mining and Digital Humanities, Towards robotic translation?*, 1–7. <https://doi.org/10.46298/jdmdh.9067>

Allegati

Allegato A: Testi tradotti durante l'esperimento

Testo A:

Starker Charakter mit neuen Facetten

Das geschärfte SUV-Design des neuen T-Roc zeigt Charakter – und zwar Ihren. Dafür sorgt auch eine Auswahl prägnanter Design-Highlights, zum Beispiel die LED-Scheinwerfer und die Chromleisten, die den markanten Look seiner sportlichen Front unterstreichen. Um Ihren T-Roc unverwechselbar zu machen, stehen Ihnen zudem acht Lackierungen zur Wahl. Weitere Akzente setzen Sie auf Wunsch mit einem farblich abgesetzten Dach und 19-Zoll-Leichtmetallrädern «Misano» in Schwarz.

Ihr hilfreicher **Reisebegleiter**

Ausgestattet mit innovativen Technologien für mehr Komfort und mehr Sicherheit, bringt Sie der neue T-Roc entspannter ans Ziel. Dazu zählt auch der optionale Fahrassistent IQ.DRIVE «Travel Assist». Der intelligente Co-Pilot hat ein Auge auf das, was um Sie herum passiert und erhöht den Fahrkomfort, indem er Sie in eintönigen und ermüdenden Fahrsituationen unterstützt. Das macht ihn vor allem auf Langstrecken zum idealen Reisebegleiter.

- Besonders auf der Autobahn und gut ausgebauten Landstrassen kann der IQ.DRIVE «Travel Assist» für Sie die Spur und die Geschwindigkeit halten und den Abstand zum vorausfahrenden Verkehr berücksichtigen.
- In Verbindung mit dem optionalen Doppelkupplungsgetriebe DSG fahren Sie entspannt durch Stau und Stop-and-go-Verkehr.

(Fonte: <https://www.volkswagen.ch/de/modelle/t-roc.html>)

Testo B:

Lichtdesign

Selbstbewusste Ausstrahlung

Von vorn bis hinten begeistert der neue T-Roc mit seinem neu gestalteten Lichtdesign: Schon von Weitem fallen die charakteristischen LED-Tageslichtscheinwerfer und moderne, abgedunkelte Rückleuchten ins Auge. Optional verleiht das illuminierte Kühlergitter der markanten Front eine besondere Ausstrahlung.

ergoActive Sitz

Der beste Platz **ist für Sie reserviert**

Nehmen Sie Platz und geniessen Sie die Fahrt im optionalen ergoActive Sitz. Den können Sie ganz bequem an Ihre Statur anpassen. Das entlastet Ihren Rücken und sorgt für hohen Sitzkomfort und Entspannung. Der ergoActive Sitz ist mit Artvelours Bezug und elektrischer Lordose mit Massagefunktion ausgestattet. So wird sogar die Fahrt zur Arbeit zum Wellness-Ausflug. Ausserdem bietet Ihnen der ergoActive Sitz:

- integrierte Sitzheizung
- verstellbare Beinauflage

Multifunktionslenkrad

Hier dreht sich alles um Sie

Mit dem Multifunktionslenkrad steuern Sie nicht nur den neuen T-Roc, sondern auch Radio, Mediaplayer, Sprachsteuerung, Telefon und die Ansicht des Kombiinstrument Digital Cockpit. Im T-Roc Life und Style ist es auf Wunsch beheizbar und sorgt mit edlem Lederbezug für ein exklusives Ambiente. Und wenn Sie sich für den T-Roc R-Line entscheiden, erwartet Sie das optional beheizbare Sport Multifunktionslenkrad inklusive Touchfunktion.

(Fonte: <https://www.volkswagen.ch/de/modelle/t-roc.html>)

Testo C:

Climatronic mit Touch-Bedienung

Reagiert **auf Sie**

Für ein rundum angenehmes Klima sorgt im neuen T-Roc die optionale Klimaanlage «Air Climatronic». Steuerung und Anzeigen sind nun digitalisiert und per Fingertipp über einen Touchscreen zugänglich, was die Bedienung extrem intuitiv macht und dem Cockpit einen besonders modernen Look verleiht. Für eine einfache und bequeme Bedienbarkeit während der Fahrt sorgen zugleich die funktionalen Führungsrillen. So ist die perfekte Temperatur immer nur einen Touch entfernt.

Digital Cockpit

Das Wesentliche im Blick: die Fahrbahn

Das serienmässige Digital Cockpit fasst alle für Sie relevanten Informationen im neuen T-Roc auf einem 20,3 cm (8 Zoll) grossen Farbdisplay mit ansprechenden Grafiken zusammen. So können sie während der Fahrt über das per Multifunktionslenkrad bedienbare Kombiinstrument bequem auf Funktionen wie Audio, Telefonie oder Assistenzsysteme zugreifen. Optional können Sie Ihren T-Roc auch mit dem Digital Cockpit Pro ausstatten. Mit einem 26,04 cm (10,25 Zoll) grossen, hochauflösenden Display bietet es ihnen vielfältige Möglichkeiten zur individuellen Konfiguration und zusätzliche Anzeigen, wie zum Beispiel die Darstellung der Navigationskarte direkt hinter dem Lenkrad.

(Fonte: <https://www.volkswagen.ch/de/modelle/t-roc.html>)

Testo di prova per familiarizzare con Lilt:

Ein Tool für all Ihre Übersetzungsanforderungen

Erstklassige Qualität

Integrieren Sie die unvergleichliche Qualität und Präzision von DeepL-Übersetzungen in Ihre Produkte und Anwendungen.

Dokumentübersetzung

Übersetzen Sie ganze Word- (.docx), PowerPoint- (.pptx), PDF- (.pdf), Text- (.txt) und HTML-Dateien (.html).

Höchste Datensicherheit

Profitieren Sie von weltweit führenden Datenschutzstandards und der sofortigen Löschung Ihrer Texte nach der Übersetzung.

Anredeform

Bestimmen Sie, wie formell oder informell Ihre Übersetzungen für Zielsprachen mit mehreren Anredeformen klingen sollen.

Spracherkennung

Nutzen Sie die Spracherkennung der DeepL API, um die Ausgangssprache eines Textes automatisch erkennen zu lassen.

Monatliche Kostenkontrolle

Begrenzen Sie Ihre monatlichen nutzungsabhängigen Kosten mit der Kostenkontrolle.

Verarbeitung von XML

Übersetzen Sie Texte, die XML-Markup enthalten. Weitere Informationen finden Sie in unserer API-Dokumentation.

API-Dokumentation

Greifen Sie jederzeit auf unsere API-Dokumentation zu, um mehr darüber zu erfahren, wie die API funktioniert.

Glossar

Legen Sie fest, wie bestimmte Wörter und Ausdrücke übersetzt werden sollen, indem Sie Ihr eigenes Glossar erstellen.

(Fonte: <https://www.deepl.com/de/pro-api>)

Allegato B: Risultati del processo di traduzione

Sforzo temporale

Partecipante	Durata in minuti		
	Post-editing	Suggerimento	TA interattiva
P1	4.33	3.40	11.42
P2	25.89	27.02	33.62
P3	23.29	30.45	49.53
P4	14.62	26.91	36.86
P5	22.76	16.06	27.85
P6	13.42	13.22	29.15
P7	5.94	13.91	30.87
Media	15.75	18.71	31.33
∅ Studenti	21.64	25.11	36.97
∅ Professionisti	7.90	10.17	23.81

Tabella 9: Tempo impiegato per la traduzione (minuti)

Partecipante	Durata in secondi per parola		
	Post-editing	Suggerimento	TA interattiva
P1	1.4	1.2	4.1
P2	8.9	9.7	11.2
P3	8.3	10.1	17.0
P4	5.0	9.0	13.2
P5	7.6	5.7	9.5
P6	4.8	4.5	9.7
P7	2.0	5.0	10.3
Media	5.4	6.5	10.7
∅ Studenti	7.45	8.63	12.73
∅ Professionisti	2.76	3.55	8.03

Tabella 10: Tempo impiegato per la traduzione (secondi per parola)

Sforzo tecnico

Partecipante	Numero totale di keystroke		
	Post-editing	Suggerimento	TA interattiva
P2	1525	927	2149
P3	747	1377	2323
P4	732	1276	1579
P6	346	582	2372
P7	259	1537	2683
Media	722	1140	2221
∅ Studenti	1001	1193	2017
∅ Professionisti	303	1060	2528

Tabella 11: Numero di keystroke necessari per la traduzione

Partecipante	Numero totale di clic del mouse		
	Post-editing	Suggerimento	TA interattiva
P2	297	242	332
P3	189	258	374
P4	150	294	294
P6	115	98	218
P7	65	95	250
Media	163.2	197	294
∅ Studenti	212	265	333
∅ Professionisti	90	97	234

Tabella 12: Numero di clic del mouse necessari per la traduzione

Partecipante	Keystroke ratio (keystroke/caratteri testo di arrivo)		
	Post-editing	Suggerimento	TA interattiva
P2	0.95	0.69	1.50
P3	0.58	0.95	1.71
P4	0.55	0.88	1.40
P6	0.26	0.43	1.67
P7	0.19	1.17	1.75
Media	0.51	0.82	1.60
∅ Studenti	0.70	0.84	1.54
∅ Professionisti	0.23	0.80	1.71

Tabella 13: Keystroke ratio (keystroke/caratteri testo di arrivo)

Sforzo cognitivo

Partecipante	Numero di pause > 5 secondi		
	Post-editing	Suggerimento	TA interattiva
P2	13	23	42
P3	35	26	64
P4	25	39	58
P6	18	20	49
P7	1	9	27
Media	18.40	23.40	48.00
∅ Studenti	24.3	29.3	54.7
∅ Professionisti	9.5	14.5	38.0

Tabella 14: Numero di pause > 5 secondi

Partecipante	Numero di pause > 10 secondi		
	Post-editing	Suggerimento	TA interattiva
P2	2	9	12
P3	10	6	16
P4	9	15	26
P6	10	10	21
P7	0	2	11
Media	6.20	8.40	17.20
∅ Studenti	7.0	10.0	18.0
∅ Professionisti	5.0	6.0	16.0

Tabella 15: Numero di pause > 10 secondi

Partecipante	Pause ratio		
	Post-editing	Suggerimento	TA interattiva
P2	6.9%	15.4%	18.7%
P3	23.6%	11.7%	19.0%
P4	28.2%	24.3%	42.7%
P6	37.9%	41.7%	33.3%
P7	2.0%	10.7%	15.2%
Media	19.7%	20.8%	25.8%
∅ Studenti	19.57%	17.14%	26.79%
∅ Professionisti	19.94%	26.19%	24.26%

Tabella 16: Pause ratio

Partecipante	Pause to word ratio		
	Post-editing	Suggerimento	TA interattiva
P2	0.07	0.14	0.23
P3	0.21	0.14	0.37
P4	0.14	0.22	0.35
P6	0.11	0.11	0.27
P7	0.01	0.05	0.15
Media	0.11	0.13	0.27
∅ Studenti	0.14	0.17	0.31
∅ Professionisti	0.06	0.08	0.21

Tabella 17: Pause to word ratio

Allegato C: Qualità dei test finali

Media valutatori (testi tradotti con il post-editing convenzionale)							
Testi PE	Coerenza	Coesione	Grammatica	Stile	Adequat. culturale	Impr. generale	Media
P1- B	4.0	3.7	4.7	4.5	3.8	3.8	4.1
P2- A	4.2	4.0	3.7	4.2	4.2	4.0	4.0
P3- C	4.5	3.7	4.2	3.8	4.0	3.8	4.0
P4- A	4.8	4.5	4.5	5.2	4.5	5.0	4.8
P5- B	4.0	4.3	4.3	4.3	3.8	4.0	4.1
P6- C	4.7	4.3	4.7	5.5	5.2	5.2	4.9
P7- A	4.5	4.0	4.8	4.3	4.2	4.3	4.4
ø PE	4.4	4.1	4.4	4.5	4.2	4.3	4.3
ø stud.	4.4	4.1	4.2	4.4	4.1	4.2	4.2
ø prof.	4.4	4.0	4.7	4.8	4.4	4.4	4.5

Tabella 18: Valutazione della qualità dei test finali (media PE)

Media valutatori (testi tradotti con il suggerimento all'esterno del segmento)							
Testi SUGG	Coerenza	Coesione	Grammatica	Stile	Adequat. culturale	Impr. generale	Media
P1- A	4.0	3.8	3.5	3.7	3.5	3.3	3.6
P2- C	4.2	3.7	4.7	4.0	4.0	4.0	4.1
P3- B	4.7	4.2	4.7	3.5	4.5	4.2	4.3
P4- B	4.0	3.7	2.8	3.7	3.5	3.0	3.4
P5- C	4.3	4.0	5.0	4.3	4.8	4.5	4.5
P6- A	4.3	3.7	3.5	4.2	4.7	4.2	4.1
P7- C	4.7	4.2	4.5	4.3	4.2	4.7	4.4
ø SUGG	4.3	3.9	4.1	4.0	4.2	4.0	4.1
ø stud.	4.3	3.9	4.3	3.9	4.2	3.9	4.1
ø prof.	4.3	3.9	3.8	4.1	4.1	4.1	4.0

Tabella 19: Valutazione della qualità dei test finali (media suggerimento)

Media valutatori (testi tradotti con la TA interattiva)							
Testi LILT	Coerenza	Coesione	Grammatica	Stile	Adequat. culturale	Impr. generale	Media
P1- C	3.3	3.2	2.8	3.5	2.7	2.2	2.9
P2- B	4.2	4.0	3.3	3.5	3.3	3.0	3.6
P3- A	4.3	4.0	3.5	4.2	3.7	4.2	4.0
P4- C	4.8	4.2	4.7	4.7	5.0	4.5	4.6
P5- A	4.3	4.0	3.8	2.8	4.0	3.5	3.8
P6- B	3.5	3.7	3.8	4.5	4.3	3.5	3.9
P7- B	4.0	4.3	4.2	3.8	4.3	3.8	4.1
ø LILT	4.1	3.9	3.7	3.9	3.9	3.5	3.8
ø stud.	4.4	4.0	3.8	3.8	4.0	3.8	4.0
ø prof.	3.6	3.7	3.6	3.9	3.8	3.2	3.6

Tabella 20: Valutazione della qualità dei test finali (media TA interattiva)

Mediana valutatori (testi tradotti con il post-editing convenzionale)							
Testi PE	Coerenza	Coesione	Grammatica	Stile	Adequat. culturale	Impr. generale	Mediana
P1- B	4.0	4.0	5.0	5.0	4.5	4.5	4.5
P2- A	4.5	4.0	4.0	4.0	4.0	4.5	4.0
P3- C	4.5	4.0	4.0	4.0	4.0	4.0	4.0
P4- A	5.0	4.0	4.5	5.0	4.5	5.0	4.8
P5- B	4.0	4.0	4.0	4.5	4.0	4.0	4.0
P6- C	5.0	4.5	5.0	5.5	5.0	5.0	5.0
P7- A	4.0	4.0	5.0	4.0	4.0	4.5	4.0
Mediana	4.5	4.0	4.5	4.5	4.0	4.5	4.5
stud.	4.5	4.0	4.0	4.3	4.0	4.3	4.1
prof.	4.0	4.0	5.0	5.0	4.5	4.5	4.5

Tabella 21: Valutazione della qualità dei testi finali (mediana PE)

Mediana valutatori (testi tradotti con il suggerimento all'esterno del segmento)							
Testi SUGG	Coerenza	Coesione	Grammatica	Stile	Adequat. culturale	Impr. generale	Mediana
P1- A	4.0	4.0	3.5	3.5	3.5	3.5	3.5
P2- C	4.0	3.0	5.0	4.0	4.0	4.0	4.0
P3- B	4.5	4.0	4.5	4.0	5.0	4.5	4.5
P4- B	4.0	4.0	2.5	3.0	3.0	2.5	3.0
P5- C	5.0	4.5	5.0	4.5	5.0	5.0	5.0
P6- A	4.0	3.0	3.0	4.0	4.0	4.0	4.0
P7- C	5.0	4.0	4.5	4.5	4.0	4.5	4.5
Mediana	4.0	4.0	4.5	4.0	4.0	4.0	4.0
stud.	4.3	4.0	4.8	4.0	4.5	4.3	4.3
prof.	4.0	4.0	3.5	4.0	4.0	4.0	4.0

Tabella 22: Valutazione della qualità dei testi finali (mediana suggerimento)

Mediana valutatori (testi tradotti con la TA interattiva)							
Testi LILT	Coerenza	Coesione	Grammatica	Stile	Adequat. culturale	Impr. generale	Mediana
P1- C	3.0	3.0	3.0	3.5	2.5	2.5	3.0
P2- B	4.5	4.0	3.0	3.0	3.0	3.0	3.0
P3- A	5.0	4.0	4.0	4.0	4.0	4.0	4.0
P4- C	5.0	4.0	4.5	4.5	5.0	4.5	4.5
P5- A	5.0	4.0	4.0	3.0	4.0	3.5	4.0
P6- B	3.0	3.5	4.0	5.0	5.0	4.0	4.0
P7- B	4.0	5.0	4.5	4.5	5.0	4.5	4.5
Mediana	4.5	4.0	4.0	4.0	4.0	4.0	4.0
stud.	5.0	4.0	4.0	3.5	4.0	3.8	4.0
prof.	3.0	3.5	4.0	4.5	5.0	4.0	4.0

Tabella 23: Valutazione della qualità dei testi finali (mediana TA interattiva)

Allegato D: Dati demografici

Partecipante	P1	P2	P3	P4	P5	P6	P7
Categoria	profess.	stud.	stud.	stud.	stud.	profess.	profess.
Età	30	35	24	27	24	27	29
Sesso	F	M	M	F	F	F	F
Anni di esperienza professionale nella traduzione	7	2	0.5	4	2	4	7
Anni di esperienza nel post-editing	7	2	0.5	0	2	4	2
Quanto ti piace usare la TA? 1= poco; 10 = tanto	9	9	7	7	7	7	6
Come vedi la TA? 1= nessuna minaccia per la professione in futuro; 5= grande minaccia per la professione in futuro	2	1	3	2	3	1	3
Come vedi la TA? 1= per niente utile; 5= molto utile	5	4	4	5	4	5	3
Di solito usa	PE	PE	sugg.	sugg.	sugg.	sugg.	sugg.
Già usato TA interattiva prima	Sì	Sì	Sì	No	No	No	No

Allegato E: Risposte al questionario

Partecipante	P1	P2	P3	P4	P5	P6	P7
Con quale modalità hai preferito lavorare?	PE	PE	sugg.	sugg.	sugg.	PE	PE
Con quale modalità ti è sembrato di lavorare in modo più veloce ed efficiente?	PE	PE	sugg.	PE	sugg.	PE	PE
Con quale modalità ti è sembrato di raggiungere una qualità maggiore?	PE	PE	sugg.	sugg.	sugg.	sugg.	sugg.
Secondo te, quale modalità ha richiesto lo sforzo cognitivo maggiore?	TA interattiva	TA interattiva	sugg.	TA interattiva	TA interattiva	TA interattiva	TA interattiva

Quale dei tre testi ti è sembrato più difficile da tradurre? A, B, o C? Perché?

P1	C. Non mi piace Lilt
P2	B. La TA in LILT era di pessima qualità, ciò che più che aiutare fuorviava. Trattandosi di un testo commerciale/pubblicitario, con un linguaggio quindi piuttosto figurato e evocativo, se la traduzione automatica non è più che buona rende la vita difficile. Infatti sono ricorso a DeepL in esterna per avere almeno una traduzione accettabile di termini tecnici e quindi dover abbellire solo il linguaggio per rimanere fedele alla funzione evocativa.
P3	A: Credo che la mia inefficienza e le mie difficoltà (p. es. nelle formulazioni) siano in parte dovute al fatto che ho dovuto prima familiarizzarmi con questo tipo di testo. Inoltre, l'«aiuto» da parte di Lilt non è stato un granché.
P4	C - terminologia tecnica, CAT-Tool utilizzato (Lilt)

P5	Testo A perché è il testo per il quale ho dovuto cercare di più i termini automobilistici (p. es. Leichtmetallrädern, Co-Pilot, ecc.)
P6	B- Perché la traduzione automatica era molto fuorviante e non si vedeva il layout del testo. Era un po' confusionario.
P7	B perché è stato tradotto con un tool che dava suggerimenti errati.

Cosa ti è particolarmente piaciuto/non piaciuto della traduzione automatica interattiva?

Part.	TA interattiva (+)	TA interattiva (-)
P1		Inserire i suggerimenti nella traduzione è lento.
P2	Ha delle grandi potenzialità, serve sicuramente esercizio,	ma la qualità di molti passaggi lasciava a desiderare. Per inciso, il TEST in LILT ha funzionato molto meglio che l'incarico vero e proprio. In quest'ultimo ho avuto un attimo di nervosismo perché la qualità era molto scarsa, e quindi era un ostacolo.
P3	Mi piace molto l'idea che la traduzione automatica sia interattiva e adattiva e che dunque proponga più varianti a seconda di quello che viene scritto.	Però la qualità, in confronto a DeepL, è molto più scarsa. Inoltre, ci sono occasioni in cui la parte interattiva/adattiva «si inceppa» e non segue il traduttore/la traduttrice.
P4	È interessante la possibilità di inserire una parola alla volta, così come i suggerimenti adattivi.	Purtroppo la qualità dei suggerimenti è piuttosto bassa.
P5		Mi ha disturbato il fatto di vedere le frasi/parole muoversi a seconda di cosa inserivo, perché volevo concentrarmi su quello che scrivevo ma automaticamente guardavo dove la TA continuava a cambiare (mi ha distratto un po').
P6	Utile la funzione di caratteri ben visibile.	L'output della TA era di pessima qualità e, pur essendo interattivo, non migliorava scrivendo parti di frasi.
P7	È sicuramente uno strumento che potrebbe rivelarsi molto utile in ambito traduttivo, se l'output della MT fosse più corretto/riutilizzabile.	La traduzione proposta era spesso errata, mostrava spesso parole non presenti nel source. Non si adattava a quello che scrivevo in modo coerente. Mi è stato quasi impossibile utilizzarlo.

Cosa ti è particolarmente piaciuto / non piaciuto del post-editing convenzionale?

Part.	Post-editing (+)	Post-editing (-)
P1	Il post-editing è molto più rapido.	
P2	La qualità della traduzione automatica è già abbastanza alta, mi toglie diverso lavoro di ricerca per quelli che sono i termini tecnici. Devo solo guardare che siano effettivamente utilizzati ma mi risparmia tutto una parte di ricerca.	La TA rende un testo che generalmente è corretto ma che non ha ancora la forma diciamo del testo "pubblicitario", serve quindi un adattamento in quel senso ma non molto di più.
P3	Apprezzo il fatto che aggiunge già i tag e le formattazioni nei segmenti,	però non mi piace particolarmente il fatto di ritrovare già una proposta di traduzione davanti a me.
P4	Il plug-in di DeepL fornisce traduzioni quasi sempre di qualità, quindi il post editing si rivela molto più semplice e meno invasivo.	
P5	Mi è piaciuto che comunque velocizza il lavoro	ma non tanto che con testi come questi dove ci sono termini tecnici spesso sbaglia e perdo quasi più tempo a mettere a posto l'output della TA (che sembrano giusti ma invece no, false fluency) che a tradurre da zero.
P6	Sono solita fare PE e trovo sia comodo, quindi la funzione in generale mi è piaciuta e la qualità	

	era anche buona. Il risparmio di tempo è sicuramente un vantaggio non indifferente.	
P7	Il post-editing mi permette di lavorare in modo più rapido e mi dà suggerimenti spesso utili che posso sfruttare nella mia traduzione.	Spesso sono fuorviata dalla traduzione errata proposta dalla MT oppure mi sento influenzata dalla traduzione già presente e non riesco a trovare soluzioni alternative/più creative. Mi sento spesso limitata.

Cosa ti è particolarmente piaciuto / non piaciuto della modalità con i suggerimenti all'esterno del segmento?

Part.	Suggerimento (+)	Suggerimento (-)
P1		Non mi piace il fatto che devo cliccare sopra per inserirli nel segmento.
P2		Ho notato che mi sono ritrovato a importare l'opzione fornita a lato da DeepL nel segmento per fare post editing "tradizionale", perché trovo vada più veloce e poi non sia molto ergonomica come opzione (guardare a destra tutte le volte, avendo due schermi bello grandi è un po' scomodo), Per come lavoro io, non avendo una TM e dovendo guardare a destra "solo" per la TA non ha gran senso, meglio importare direttamente la TA nel segmento e partire da lì.
P3	Questo metodo lascia la libertà di pensare ad una propria variante e di consultare, se necessario, la traduzione automatica.	
P4	Questa modalità permette di lavorare con più libertà ed essere più creativi.	
P5	Mi è piaciuto che ho avuto la libertà di scelta: se tradurre io direttamente o prendere la TA come suggerimento.	
P6		Un po' scomodo non avere già l'output dentro i segmenti, anche se è come se fossero dei fuzzy.
P7	Mi è piaciuto avere un suggerimento quando mi sentivo bloccata.	Essendo fuori dal segmento dovevo sforzarmi a guardare la proposta della MT.

Se potessi scegliere, useresti la traduzione automatica interattiva in futuro? Motiva brevemente la tua risposta.

P1	No. Preferisco il post-editing con segmento già pre-tradotto.
P2	Se potessi scegliere preferirei avere la TA di DeepL integrata nel segmento come nella App di DeepL, ovvero che cliccando su una parola è possibile vedere alternative e modificarla a piacimento, il tutto integrato in un CAT. La mia migliore soluzione al momento è ctrl+C+C, importare il segmento in DeepL pro, modificare a piacimento e reimportarlo nel segmento in versione definitiva. Se si potesse evitare l'esporto e il reimporto, sarebbe perfetto.
P3	Se la qualità della traduzione automatica migliora, sì. Altrimenti posso utilizzare un qualsiasi CAT tool e il risultato sarebbe lo stesso.
P4	Se la qualità dei segmenti migliorasse, sarei felice di utilizzarla.
P5	No, tuttavia la trovo interessante perché probabilmente sarà il futuro, ma personalmente mi turba appunto il fatto che si muovono le frasi/parole mentre scrivo distogliendo l'attenzione. Probabilmente è anche una questione di abitudine, ma non penso che mi ci abituerei o comunque con fatica.
P6	No, a meno che la qualità non sia decisamente migliore, altrimenti dà solo suggerimenti inutili che fuorviano e fanno perdere tempo.
P7	Certo, se fosse ben fatta. Velocizzerebbe il mio lavoro, lo trovo uno strumento affascinante.

Allegato F: Istruzioni date ai partecipanti (esempio P1)

Fase 1: familiarizzazione con Lilt

Se conosci e hai già usato Lilt puoi saltare questo passaggio.

Lilt è un CAT-Tool che offre un sistema di traduzione automatica interattivo e adattivo. I suggerimenti della traduzione automatica vengono adattati ogni volta che si digita qualcosa. Premendo “enter” puoi inserire direttamente le parole evidenziate proposte dalla traduzione automatica.

Di default Lilt mostra il testo di partenza e quello di arrivo uno sopra l'altro e non uno accanto all'altro. Se vuoi, puoi cambiare la visualizzazione su “view” → “split view”.

Per familiarizzare con Lilt, traduci il breve testo “TEST Lilt” o una parte di esso.

Ecco i shortcut più importanti:

Confirm/unconfirm segment	Ctrl + Enter
Insert highlighted phrase suggestion	Enter
Insert full suggestion	Shift + Enter

Se ti servono altre informazioni su Lilt:

<https://support.lilt.com/kb/introduction-to-lilt-translate>

Altri shortcut:

<https://support.lilt.com/kb/hotkeys>

Fase 2: traduzioni

Informazioni generali valide per tutti gli incarichi	
Incarico di traduzione	<ul style="list-style-type: none"> • È richiesta una traduzione di qualità umana pubblicabile. • Puoi usare tutti gli strumenti di ricerca di cui hai bisogno. • Per favore, non consultare la traduzione in italiano già presente sul web. • Traduci il testo senza pause o interruzioni (tra un testo e l'altro, invece, puoi fare delle pause). • Cerca di non metterci più di 20-30 minuti per testo (ma non devi fermarti se non hai terminato). • I testi sono tre estratti di una pagina informativa e pubblicitaria di un'automobile.
Procedimento	<ul style="list-style-type: none"> • Per avere più privacy, durante lo screen recording e il keystroke logging chiudi le applicazioni del tuo computer che potrebbero contenere dati sensibili (es. e-mail, notifiche,...) e non digitare password. • Fai una prova per vedere se la registrazione dello schermo e il keystroke logging funzionano.

- | | |
|--|--|
| | <ul style="list-style-type: none"> • Se lavori con due schermi, lo screen recording deve registrare lo schermo con il CAT-Tool. |
|--|--|

- **Incarico 1: traduzione del testo A con la modalità 1 – segmento vuoto**
 - Importa il testo A in memoQ, senza TM né TB
 - I suggerimenti della traduzione automatica appaiono al di fuori del segmento
 - Avvia la registrazione dello schermo e il keystroke logging
 - Traduci il testo
 - Ferma la registrazione dello schermo e il keystroke logging
 - Esporta il testo
- **Incarico 2: traduzione del testo B con la modalità 2 – pre-translate**
 - Importa il testo B in memoQ, senza TM né TB
 - Abilita la funzione pre-translate, in modo da avere la traduzione automatica già nei segmenti
 - Avvia la registrazione dello schermo e il keystroke logging
 - Traduci il testo
 - Ferma la registrazione dello schermo e il keystroke logging
 - Esporta il testo
- **Incarico 3: traduzione del testo C con la modalità 3 – TA interattiva/adattiva (Lilt)**
 - Attiva il tuo account Lilt tramite l'e-mail che hai ricevuto
 - Avvia la registrazione dello schermo e il keystroke logging
 - Apri il testo C che ti è stato assegnato in Lilt
 - Traduci il testo
 - Ferma la registrazione dello schermo e il keystroke logging
 - Esporta il testo

Conclusione esperimento

- | |
|---|
| <ul style="list-style-type: none"> • Mandami i vari testi tradotti esportati dai CAT-Tool • Mandami le registrazioni dello schermo (se il file è troppo grande, eventualmente via https://wettransfer.com oppure chiavetta USB o altro) • Mandami l'analisi del keystroke logging (file xml) per ogni incarico |
|---|

Fase 3: questionario

Per completare i dati raccolti durante il processo di traduzione, per favore completa il questionario finale al seguente link: <https://forms.gle/sPv7p6xx2rsHmUtL8>.

Grazie per aver partecipato!

Allegato: Guida Inputlog.pdf