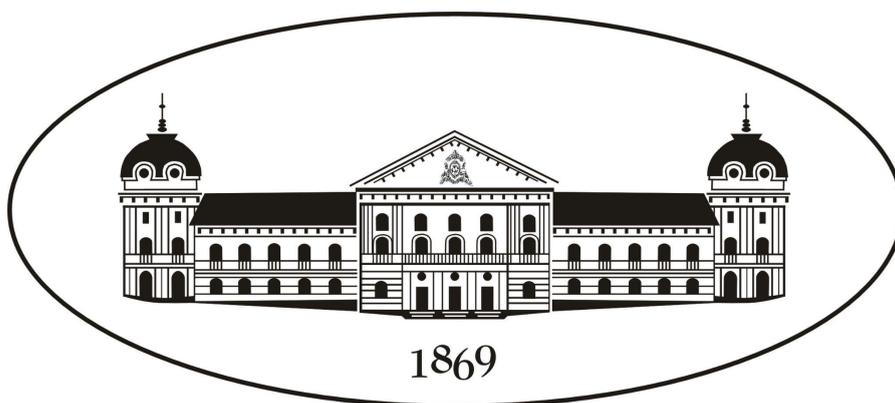


Sezione di Linguistica Computazionale
Istituto per la lingua bulgara Professor Lubomir Andreychin
Accademia bulgara delle scienze



Viktoriya Krasimirova Petrova-Lyubenova
SVILUPPO SEMIAUTOMATICO DI RISORSE
TERMINOLOGICHE MULTILINGUI

Astratto

della tesi per il conseguimento del titolo di Dottore di Ricerca

Settore dell'istruzione superiore: 2. Umanistica;

Direzione professionale: 2.1. Filologia;

Specializzazione: Linguistica generale e comparata

Supervisore:

Prof.ssa dott. ric. Svetla Koeva

Sofia

2022

Sommario

Caratteristiche generali della tesi	3
CAPITOLO UNO	8
I programmi di traduzione assistita e le loro componenti	8
I. I programmi di traduzione assistita e le loro componenti	8
II. L'atteggiamento dei traduttori nei confronti della tecnologia per traduzione assistita	14
CAPITOLO DUE	19
Le databasi terminologiche nei programmi di traduzione assistita	19
I. Termini, dizionari terminologici e databasi terminologiche	19
II. Confronto di databasi terminologiche con dizionari terminologici	22
III. Collegamento delle databasi terminologiche con la memoria di traduzione	24
IV. Creazione e gestione di databasi terminologiche nei programmi di traduzione assistita	25
V. Capacità di lavorare con la terminologia e le databasi terminologiche	26
CAPITOLO TRE	28
Creazione semiautomatica di databasi terminologiche multilingui	28
I. Creazione di risorse terminologiche nel campo della terminologia informatica	29
II. Metodologia per la creazione semiautomatica di una database terminologica nel campo della terminologia informatica	35
III. Descrizione delle caratteristiche della database terminologica creata nel campo della terminologia informatica	39
CAPITOLO QUATTRO	41
Misurare la qualità nei programmi di traduzione assistita e nelle loro parti	41
I. Misurazione dell'accuratezza della traduzione con programmi di traduzione assistita	42
II. Confronto dell'efficacia dei programmi di valutazione della traduzione	45
III. Classificazione degli errori	47
IV. Un confronto dei risultati delle misure automatiche di qualità	50
V. Standard internazionali per la qualità della traduzione	53
Conclusione	55
Elenco delle pubblicazioni relative all'argomento della tesi	57
BIBLIOGRAFIA	58

Caratteristiche generali della tesi

Nel periodo tra gli anni '60 e gli anni '80, le speranze che la traduzione automatica si sviluppasse a sufficienza per sostituire la traduzione umana si sono rivelate infondate¹. Pertanto, gli sforzi scientifici vengono reindirizzati allo sviluppo di programmi per assistere i traduttori nel loro lavoro. L'uso crescente di strumenti e risorse di traduzione computerizzata cambia il modo in cui lavorano i traduttori e la teoria della traduzione. Le nuove tecnologie sono così indissolubilmente legate alla professione del traduttore che oggi la traduzione sembra impensabile senza l'utilizzo di qualche strumento informatico (anche il più semplice editor di testo).

La necessità di avere traduzioni specialistiche e tecniche è sempre maggiore, poiché la quantità dei testi del genere sta crescendo rapidamente. Tali traduzioni non possono essere effettuate senza conoscere la terminologia corretta per l'area in questione. La situazione si complica ulteriormente dalla mancanza di risorse linguistiche per le lingue "piccole", come lo è il bulgaro. Pertanto, la presente tesi si propone di analizzare la situazione al momento e di proporre un modo per superare (almeno per quanto possibile) la mancanza di risorse terminologiche.

L'oggetto della ricerca comprende la ricerca, l'analisi e la selezione di metodi accessibili ed efficaci per la creazione di risorse terminologiche destinate ai programmi di traduzione assistita. Riguarda in particolare i metodi esistenti, per creare databasi terminologiche e per esplorare le possibilità di integrare altre tecnologie allo scopo di facilitare e accelerare (ove possibile) il processo di creazione di nuove risorse terminologiche.

L'oggetto d'indagine sono le databasi terminologiche e il loro utilizzo nei programmi di traduzione assistita, nonché i modi in cui le moderne tecnologie ne consentono e facilitano la creazione e l'utilizzo. Inoltre vengono analizzate le attitudini dei traduttori, i mezzi che utilizzano per creare databasi terminologiche e quali sono gli strumenti più diffusi (sia a pagamento che gratuiti).

Le lingue oggetto di attenzione sono il bulgaro e l'inglese e la direzione della traduzione nella tesi è dall'inglese al bulgaro. La creazione di risorse terminologiche è

¹ Soprattutto dal Comitato consultivo per l'elaborazione automatica del linguaggio o rapporto ALPAC del 1966.

illustrata da corpora paralleli, dizionari terminologici e una database terminologica nel campo della terminologia informatica. Il costante sviluppo delle tecnologie richiede anche il corrispondente rinnovamento della terminologia informatica.

L'obiettivo principale della ricerca è quello di proporre una metodologia per la creazione di risorse terminologiche per un settore specializzato. Per questo motivo sono state analizzate le possibilità dei programmi di traduzione assistita per la creazione di databasi terminologiche, programmi per l'estrazione di termini, nonché le risorse terminologiche attualmente disponibili per la lingua bulgara.

Il raggiungimento **dell'obiettivo principale** è pianificato attraverso i seguenti **obiettivi specifici:**

- Descrivere in dettaglio i moderni programmi di traduzione assistita, sottolineandone le componenti principali (memoria di traduzione, traduzione automatica, databasi terminologiche, comprese quelle multilingui).
- Descrivere le nuove conoscenze e competenze richieste ai traduttori in relazione all'uso delle tecnologie di traduzione assistita.
- Effettuare un sondaggio tra i traduttori in Bulgaria sull'uso dei programmi di traduzione assistita e di sintetizzarne e analizzarne i risultati.
- Presentare i tipi di dizionari terminologici e la loro struttura rispetto alla struttura delle databasi terminologiche.
- Descrivere la struttura e le caratteristiche delle databasi terminologiche.
- Presentare la creazione e la gestione delle databasi terminologiche nei programmi di traduzione assistita. In aggiunta saranno presi in considerazione gadget che possono essere utilizzati insieme ai programmi di traduzione assistita e quelli autonomi di gestione della terminologia.
- Dimostrare l'uso delle databasi terminologiche a seconda dei diversi tipi di testo per la traduzione e l'applicazione di lingue controllate.
- Presentare brevemente le caratteristiche della terminologia informatica, mirando a una generalizzazione dal punto di vista dell'inclusione dei termini informatici nelle risorse terminologiche.
- Descrivere i modi per creare risorse terminologiche, partendo dai dizionari terminologici tradizionali e arrivando all'estrazione automatica dei termini e presentando gli strumenti esistenti a tale scopo.

- Creare una metodologia per la creazione semiautomatica di una database terminologica nel campo della terminologia informatica e descrivere le fasi specifiche del lavoro sulla creazione di una database terminologica bilingue (per inglese e bulgaro).
- Presentare la misurazione dell'accuratezza della traduzione con programmi di traduzione assistita, rispettivamente senza e con databasi terminologiche.
- Presentare la misurazione dell'accuratezza della traduzione con funzionalità integrate o esterne ai programmi di traduzione assistita.
- Proporre una classificazione degli errori di traduzione con programmi di traduzione assistita in considerazione della loro rilevanza rispetto al contenuto del testo.
- Confrontare l'efficacia dei programmi di valutazione della traduzione secondo una metodologia predeterminata.

Nel lavoro sulla tesi sono stati utilizzati diversi **metodi moderni (statistici, elaborazioni informatiche, corpus)** oltre ad alcuni metodi **tradizionali (descrittivi, analitici, sperimentali)**. Gli approcci principali per raggiungere gli obiettivi della tesi sono l'analisi e la descrizione. Il metodo descrittivo comprende l'esaminazione della letteratura scientifica esistente in questo campo, l'esaminazione delle databasi terminologiche già note e l'esaminazione degli strumenti tecnici utilizzati nel lavoro quotidiano dei traduttori. L'analisi viene applicata nella selezione delle risorse linguistiche e degli strumenti appropriati per la loro elaborazione.

Sono state formulate delle **ipotesi** sull'atteggiamento dei traduttori bulgari nei confronti dei nuovi mezzi di lavoro per la traduzione e sulla quantità di termini che possono essere estratti a seconda del corpus parallelo di origine. Le ipotesi vengono testate sperimentalmente in diversi modi: attraverso un sondaggio condotto online; lavorando con programmi per l'elaborazione automatica di testi: per l'estrazione automatica di termini, per l'allineamento automatico di documenti paralleli per frasi, per la creazione automatica di databasi terminologiche bilingue.

Il lavoro sulla tesi ha richiesto un lavoro professionale con una serie di programmi di traduzione assistita e le loro componenti (memoria di traduzione e database terminologica), programmi autonomi per l'estrazione automatica delle informazioni, programmi per

l'estrazione automatica dei termini e l'allineamento delle frasi di testi paralleli e, ultimo ma non meno importante – programmi autonomi per il controllo della qualità della traduzione.

Lo sviluppo scientifico si compone di **quattro capitoli indipendenti**.

Capitolo uno - *I programmi per la traduzione assistita e le loro componenti*, in cui vengono esaminate le nuove caratteristiche dei programmi di traduzione assistita, il loro uso in funzione delle caratteristiche linguistiche dei diversi tipi di testi e le esigenze di questi programmi di traduzione assistita riguardo alle competenze dei traduttori. È stato condotto un sondaggio tra i traduttori in Bulgaria per riassumere e analizzare il grado di utilizzo delle nuove tecnologie e il loro atteggiamento nei confronti della traduzione automatica e dei programmi di traduzione assistita da computer.

Capitolo due - *Le databasi terminologiche nei programmi di traduzione assistita* presenta i termini, i glossari e le databasi terminologiche nel contesto della traduzione. È stato effettuato un confronto delle databasi terminologiche con i dizionari terminologici e della memoria di traduzione e delle databasi terminologiche. Sono stati analizzati i metodi per la creazione e la gestione delle databasi terminologiche nei programmi di traduzione assistita, i componenti aggiuntivi del programma principale, i sistemi di gestione della terminologia e le competenze per lavorare con la terminologia e le databasi terminologiche.

Capitolo tre - *Creazione semiautomatica di database terminologiche multilingui* presenta la creazione di risorse terminologiche nel campo della terminologia informatica e le sue caratteristiche. Sono stati analizzati i mezzi per l'estrazione automatica dei termini ed è stato effettuato un confronto tra di essi. È stata sviluppata una metodologia per la creazione semiautomatica di una datadatabase terminologica nel campo della terminologia informatica attraverso un corpus parallelo bilingue. Dopo l'estrazione dei termini, è stata effettuata una verifica manuale e un'impaginazione tecnica della datadatabase terminologica. Vengono descritte le caratteristiche della database terminologica creata nel campo della terminologia informatica.

Capitolo quattro - *Misurare la qualità nei programmi di traduzione assistita da computer e nei loro componenti* presenta la misurazione dell'accuratezza della traduzione con programmi di traduzione assistita senza base, con base e con programmi esterni agli strumenti di traduzione assistita. Viene proposta una classificazione degli errori e viene effettuato un confronto dei risultati degli strumenti automatici di misurazione della qualità. Vengono

esaminati gli standard internazionali, le metriche e i modelli relativi alla qualità della traduzione.

La dissertazione ha tre appendici, che mostrano integralmente i risultati del sondaggio tra i traduttori in Bulgaria, i termini inglesi estratti automaticamente nel campo della terminologia informatica e la database terminologica inglese-bulgara nel campo della terminologia informatica.

CAPITOLO UNO

I programmi di traduzione assistita e le loro componenti

I. I programmi di traduzione assistita e le loro componenti

Dopo il fallimento dei primi prototipi di traduzione automatica, viene proposto un nuovo approccio: un programma che non traduce automaticamente, ma facilita il lavoro del traduttore. Stiamo parlando di programmi di traduzione assistita (Computer Assisted / Aided Translation Tools) - software che consente al traduttore di tradurre le traduzioni assegnategli nel modo più rapido e semplice possibile. In questi programmi, i traduttori sono in grado di modificare, creare, archiviare e gestire i contenuti, avendo allo stesso tempo a disposizione database terminologica, traduzioni precedenti, traduzione automatica, controllo ortografico e altre funzioni utili (come ad es. l'anteprima del documento originale, l'utilizzo di combinazioni macro, l'opzione di dettatura, ecc.). Una caratteristica fondamentale di questi tipi di programmi è la loro capacità di suddividere i documenti in segmenti (singole frasi o paragrafi, a seconda delle impostazioni) che vengono memorizzati in una banca dati e che possono essere riutilizzati.

I creatori di programmi di traduzione assistita danno le seguenti definizioni per questo tipo di programma: «Gli strumenti di traduzione assistita sono applicazioni software che aiutano a tradurre testi da una lingua all'altra. In particolare, gli strumenti di traduzione assistita vengono utilizzati per commissionare, modificare, gestire e archiviare le traduzioni», secondo Memsource²; «Gli strumenti di traduzione assistita dividono i documenti multilingui di grandi dimensioni in segmenti (frasi e paragrafi), che vengono archiviati in una banca dati chiamata memoria di traduzione: i documenti già tradotti possono essere riutilizzati in qualsiasi momento», secondo memoQ³; «Questo è un software utilizzato da traduttori e linguisti. Ha vari scopi, ma soprattutto supporta il processo di traduzione. Permette l'editing, la creazione, la gestione delle traduzioni» secondo Trados⁴. Nel 2015 sono presenti sul mercato 67 programmi di traduzione assistita (Garcia 2015: 22), e nel 2020 il numero è salito a 164⁵, con cui si intendono "Sistemi di gestione delle traduzioni" (Translation Management

² <https://www.memsource.com/what-are-cat-tools/>

³ <https://www.memoq.com/tools/what-is-a-cat-tool>

⁴ <https://www.trados.com/solutions/cat-tools/translation-101-what-is-a-cat-tool.html>

⁵ Според годишния доклад на Нимзи за 2020 г.

<https://www.nimdzi.com/nimdzi-language-technology-atlas-2020/?hilite=trados>

Systems). I sistemi di gestione delle traduzioni sono progettati per organizzare e gestire l'intero processo di traduzione e di solito includono funzioni che non sono per niente correlate ad esse.

I programmi di traduzione assistita hanno due componenti principali: la memoria di traduzione e la database terminologica. Anche la traduzione automatica fa parte di alcuni dei programmi di traduzione assistita.

La memoria di traduzione (Translation Memory o TM) è una banca dati che memorizza espressioni da documenti già tradotti da un essere umano. Quando viene inserito nel programma di traduzione assistita, il testo viene segmentato (per impostazione predefinita a ogni segno di punteggiatura di fine frase) e allineato in coppie corrispondenti di unità o segmenti della lingua di origine e di destinazione. Tutti i segmenti della lingua di origine e di destinazione sono memorizzati. Quando si incontra nuovamente un'espressione nello stesso o in un nuovo documento, la sua traduzione viene suggerita dalla memoria. Il motivo per cui la memoria di traduzione è estremamente preziosa è che fa risparmiare tempo e denaro ai propri proprietari e aiuta ad aumentare l'efficienza e la produttività dei traduttori perché questi impiegano meno tempo a cercare la traduzione corretta e hanno solo bisogno di confrontare le ripetizioni identificate dalla memoria (Mitchell-Schuitevoerder 2020: 5). Sebbene la qualità della traduzione automatica sia notevolmente migliorata negli ultimi anni (principalmente con i progressi nel campo della traduzione automatica neurale), l'uso della memoria di traduzione è ancora preferito. In questo modo, non solo è possibile tradurre "perfettamente" frasi già tradotte, ma viene offerta anche una qualità di traduzione "quasi perfetta" quando si recuperano frasi simili dalla memoria (Bulté e Tezcan 2019).

L'unità di base della memoria, il segmento, è solitamente punteggiato ed è normalmente una frase, ma può anche essere un titolo, un motto, una formula, ecc. Un segmento della lingua di origine (segmento di origine) associato alla sua traduzione (segmento di destinazione), più i relativi metadati pertinenti (ad esempio, ora, data e nome del traduttore o del redattore che ha eseguito la correzione, nome del cliente, area tematica, ecc.) formano la cosiddetta "unità di traduzione" (Translation Unit o TU). La memoria contiene anche un algoritmo per riconoscere le corrispondenze: quando si incontra un segmento identico o simile in un nuovo testo. Se nella memoria viene trovato un segmento di origine nella banca dati che corrisponde esattamente al segmento del nuovo testo, la traduzione corrispondente viene offerta come corrispondenza esatta o corrispondenza al

100%. In una situazione del genere, ciò che il traduttore deve fare è verificare se il segmento può essere riutilizzato così com'è o se sono necessari aggiustamenti a causa delle differenze di contesto. Se nella memoria viene trovato un segmento di origine simile al segmento tradotto, questo viene offerto come Fuzzy Match insieme al relativo grado di somiglianza come percentuale. La somiglianza è calcolata dalla distanza di Levenshtein⁶, cioè il numero minimo di inserimenti, cancellazioni o sostituzioni di caratteri; il traduttore valuta se la traduzione proposta può essere adattata o se sarebbe meno faticoso fare una traduzione completamente nuova.

Per impostazione predefinita, vengono solitamente visualizzati solo i segmenti con una corrispondenza superiore al 70% (per Trados è 75%), in quanto le corrispondenze con una percentuale inferiore sono considerate più un ostacolo che un aiuto per il traduttore⁷. Questi vengono chiamati Fuzzy Match Threshold. Quando non ci sono frasi nella memoria, si verifica il "No Match" e il traduttore deve tradurre il segmento in modo tradizionale (Garcia 2015: 71-72). Le versioni più recenti degli strumenti di traduzione assistita hanno la funzione di matches di contesto (Context Matches). Un match di contesto è noto anche come "corrispondenza del 101%" e si verifica quando c'è una corrispondenza esatta con un segmento nella memoria di traduzione e una corrispondenza con il suo contesto: ad es. anche i segmenti precedenti e successivi sono una corrispondenza esatta, ad es. 100%.

La memoria di traduzione necessita di tre diversi tipi di software per funzionare correttamente:

- sistema di allineamento dei segmenti (aligner) mediante il quale vengono creati i file delle memorie di traduzione. Può essere sia un programma autonomo che una funzionalità integrata nei programmi di traduzione assistita. Abbina i segmenti di origine e di destinazione del testo già tradotto per creare la memoria di traduzione;
- un sistema di gestione del formato che è un'opzione integrata nei programmi di traduzione assistita. Assicura che la formattazione del testo di origine corrisponda alla formattazione del testo tradotto;

⁶ La distanza di Levenshtein tra due stringhe è il numero di cancellazioni, inserimenti o sostituzioni di un singolo carattere necessarie per trasformare una stringa in un'altra. È anche noto come distanza tra due stringhe.

⁷ Va notato che sono possibili differenze nel calcolo delle corrispondenze parziali. In misura minore o maggiore può essere influenzato da: ordine delle parole, punteggiatura (a volte l'influenza non è importante, altre volte può cambiare completamente il significato del segmento), formattazione e tag (elementi che contengono informazioni codificate sulla forma e sulla struttura di una parola, frase o segmento).

- un sistema di gestione della memoria di traduzione che consente di spostare il contenuto tra diverse memorie di traduzione, cancellarlo, esportarlo in parti del contenuto, nonché la cronologia di ciascun segmento situato nella memoria (Gouadec 2007: 271 - 272).

Parlando in generale, **la traduzione automatica** è un software speciale per tradurre il testo da una lingua naturale all'altra senza l'intervento umano. Le tendenze nella traduzione assistita stanno ponendo sempre più enfasi sulla traduzione automatica.

I due modelli più popolari e diffusi sono la traduzione automatica statistica e quella neurale. Nella traduzione automatica statistica (SMT), il decodificatore che "traduce" è essenzialmente un algoritmo di ricerca. Per ogni parola e gruppo di parole nella frase di traduzione, l'algoritmo controlla le risorse disponibili contenenti parole e gruppi di parole ed estrae il miglior equivalente. Questo metodo può anche essere definito paradigmatico: ogni parola crea uno spazio che può essere riempito da una data parola tra le parole e le frasi di traduzione disponibili (Bowker e Ciro 2019: 37). La traduzione automatica neurale (NMT) funziona in modo diverso. Il decoder non cerca un elemento specifico dalle risorse disponibili. Utilizza prima le reti neurali per apprendere e poi identificare le migliori sequenze per tradurre intere frasi. Il decodificatore viene addestrato da dati paralleli in grandi volumi. La traduzione automatica neurale cerca di costruire una sequenza di parole durante la traduzione lineare. Ogni sequenza di parole precedenti determina la parola successiva.

Per molti anni, la traduzione automatica e le tecnologie di traduzione assistita sono state messe l'una contro l'altra. Tuttavia, negli ultimi anni, la traduzione automatica (soprattutto dall'avvento della traduzione automatica neurale) ha iniziato ad assumere un ruolo sempre più centrale nei flussi di lavoro di traduzione assistita. Ad esempio, nell'ultima versione di Trados, l'uso della traduzione automatica diventa parte integrante del processo di ogni nuovo progetto: dopo che i documenti di traduzione, la memoria e la database terminologica sono caricati, il programma crea una traduzione automatica specifica per questo progetto⁸.

I primi tentativi di integrazione tra la memoria di traduzione e la traduzione automatica fanno riferimento alla possibilità di combinarle in due modi. Il primo modo è che l'utente riceva, per ogni segmento, suggerimenti dalla traduzione automatica, dalla memoria

⁸ La connessione viene eseguita in modo diverso per le versioni più recenti e precedenti del programma. <https://www.trados.com/products/machine-translation/>

di traduzione e dalla database terminologica e li visualizzi contemporaneamente. Il secondo modo è utilizzare le due tecnologie insieme per migliorare i risultati nella lingua di destinazione e quindi aumentare la produttività e ridurre gli sforzi di post-editing.

Allo stesso tempo, sebbene l'uso delle tecnologie di traduzione automatica sia inevitabile, in pratica, il risultato di qualsiasi traduzione automatica può richiedere una qualche forma di intervento umano. Il motivo più importante per cui la traduzione automatica non può essere utilizzata da sola⁹ rimane la qualità del testo originale. Un modo per migliorare la qualità della traduzione automatica è pre-elaborare i testi utilizzati per la sua formazione: vengono filtrati per rimuovere termini sconosciuti e ambiguità di diversa natura; i testi vengono convertiti in un linguaggio "semplice" con frasi brevi. Un altro modo per migliorare la qualità della traduzione automatica è attraverso l'"insegnamento" utilizzando una memoria di traduzione. Da un punto di vista qualitativo, questo metodo è considerato migliore, poiché di solito i segmenti confermati della memoria sono approvati da traduttori professionisti.

La crescente integrazione della traduzione automatica nel flusso di lavoro dei progetti di traduzione porta alla cosiddetta "traduzione aumentata". La traduzione aumentata unisce le competenze dei traduttori umani e la traduzione automatica. A. Arenas definisce questa attività come "revisionare un testo pre-tradotto generato dalla traduzione automatica del testo d'origine e correggere eventuali errori al fine di raggiungere i criteri di qualità prefissati"¹⁰ (Arenas 2020: 333). Si può fare una distinzione tra due tipi di editing della traduzione automatica: editing parziale e completo. Nel primo caso, il testo deve essere comprensibile, pur ammettendo alcuni errori grammaticali e ortografici. Nel secondo caso, il testo, lo stile, la grammatica, l'ortografia e la terminologia devono essere paragonabili a quelli di un testo tradotto da un essere umano. Lavorare con la traduzione automatica richiede anche conoscenze tecniche specialistiche, che potrebbero non appartenere alla competenza dei traduttori.

Le databasi terminologiche sono banche dati centralizzate che contengono termini specifici per soggetto o progetto di traduzione. Di solito, questi termini sono pre-approvati (in molti casi, i clienti li redigono da soli). Il vantaggio principale di una database terminologica è che rende il processo di traduzione più veloce (soprattutto se è anche automatizzato) perché

⁹ L'uso della traduzione automatica per scopi personali non è preso in considerazione.

¹⁰ È importante sottolineare che i criteri di qualità vengono definiti insieme al committente del progetto prima dell'inizio dei lavori.

risparmia ai traduttori la necessità di cercare il termine richiesto. Allo stesso modo della memoria di traduzione, il programma di traduzione assistita esegue la scansione di ogni nuovo segmento per verificare la corrispondenza con la database terminologica. Questo processo di lavoro garantisce anche un alto livello di coerenza, perché lo stesso termine viene tradotto in modo identico.

A prima vista, le databasi terminologiche sono come i dizionari: contengono un elenco di termini. Al contrario dei dizionari però, le databasi terminologiche possono essere configurate dall'utente della base, che ha piena libertà in qualsiasi momento di aggiungere, eliminare, modificare e classificare i termini in diverse categorie (ad esempio, un termine vietato o un termine da utilizzare solo in un determinato contesto). **Le databasi terminologiche multilingui** seguono la stessa struttura e funzionano allo stesso modo delle basi bilingui.

Il meccanismo con cui funzionano i programmi di traduzione assistita li rende estremamente adatti per testi monotipi, poiché le unità linguistiche ripetute vengono tradotte molto più velocemente. Migliorare la velocità di traduzione attraverso la tecnologia di mappatura sopra descritta evita la necessità di digitare la stessa traduzione più e più volte. La standardizzazione dei testi aiuta anche con l'accuratezza della traduzione. Con la memoria di traduzione, i traduttori non devono più ricordare ciò che hanno tradotto per tradurre in modo equivalente frammenti dati in testi di grandi dimensioni.

P. Newman nel suo libro *Manuale di traduzione* (A Textbook in Translation) dal 1988 afferma che: «Una traduzione soddisfacente è sempre possibile, ma un buon traduttore non ne è mai soddisfatto. Di solito può migliorarla. Non esiste una traduzione perfetta, ideale o "corretta". Il traduttore cerca sempre di ampliare le proprie conoscenze e migliorare i propri mezzi espressivi; insegue sempre fatti e parole. Il traduttore lavora su quattro livelli: la traduzione è innanzitutto una scienza, che comporta la conoscenza e la verifica dei fatti e del modo in cui sono descritti - a questo livello si possono identificare gli errori fattuali; in secondo luogo, è un'abilità che richiede un linguaggio appropriato e un uso corretto; terzo, è un'arte che distingue la traduzione buona, creativa, intuitiva, a volte ispirata, dalla trasmissione impersonale di contenuti in un'altra lingua; infine, è una questione di gusti, e la varietà delle traduzioni di qualità è un riflesso delle differenze individuali dei traduttori» (Newman 1988: 6).

II. L'atteggiamento dei traduttori nei confronti della tecnologia per traduzione assistita

Il profilo del traduttore assume nuove dimensioni con lo sviluppo della tecnologia. È necessario distinguere i traduttori di testi letterari dai traduttori di testi specialistici. I requisiti per questi ultimi sono cambiati in modo significativo negli ultimi anni e continuano a cambiare, rendendo necessario lavorare con i più recenti strumenti di traduzione assistita come Trados o altri strumenti e piattaforme. I traduttori eseguono un vasto gamma di compiti che includono, ma non si limitano a: traduzione, editing, correzione di bozze, utilizzo della lingua, gestione della terminologia e molto altro.

L'uso di programmi di traduzione assistita richiede un cambiamento nei requisiti per l'istruzione e le competenze dei traduttori. Il profilo di un traduttore non può più limitarsi alle sole competenze linguistiche. I traduttori diventano esperti tecnici altamente qualificati grazie al contenuto che traducono e agli strumenti e software che devono utilizzare nel loro lavoro quotidiano (legato alla conoscenza dei diversi programmi di traduzione assistita, alla gestione della memoria di traduzione, alla creazione e modifica di banche dati terminologiche, all'editing di traduzioni automatiche, la necessità di programmazione, il lavoro con sottotitoli, localizzazione, elaborazione di immagini, gestione di file in diversi formati, ecc.).

Uno dei mezzi utilizzati negli anni per verificare gli atteggiamenti e le opinioni dei traduttori sono i sondaggi. Secondo uno studio della società americana CSA¹¹ dal 2020 per il lavoro di 7.363 traduttori poco più di un terzo (37%) di coloro che utilizzano la traduzione automatica ritiene che la qualità complessiva sia buona. L'81% nota gravi variazioni nella qualità dei testi non pre-elaborati da un essere umano (ad es. eliminazione dell'ambiguità, sostituzione della fraseologia, ecc.). I traduttori che utilizzano la traduzione automatica generalmente preferiscono lavorare con sistemi adattivi come Lilt¹² (per la complessiva miglior qualità garantita dalla personalizzazione dei contenuti) rispetto a quelli con traduzione automatica grezza (71%), come le piattaforme aperte come Google Translate. I traduttori apprezzano il contributo dei dizionari ai risultati di qualità (il 91% degli intervistati fornisce una migliore qualità quando utilizza i dizionari e il 76% quando utilizza la memoria

¹¹ CSA Research - una società americana indipendente di ricerche di mercato. <https://csa-research.com/>

¹² Lilt <https://lilt.com/>

di traduzione). È interessante notare che quasi un quarto dei traduttori (23%) afferma di fornire una migliore qualità utilizzando la traduzione automatica, smentendo la convinzione comune che il prodotto finale della traduzione automatica sia inferiore alla traduzione umana. I traduttori continuano a valutare ancora l'uso della traduzione automatica come un rallentamento del processo di traduzione rispetto al lavoro con altri strumenti, il che è molto probabilmente dovuto al fatto che il risultato della traduzione automatica non è abbastanza buono e deve essere modificato.

Nella relazione annuale del *Sondaggio sull'industria linguistica europea* (European Language Industry Survey o ELIS¹³) per il 2022, secondo i risultati: «una percentuale significativamente più alta di professionisti indipendenti riferisce che la formazione tecnologica che ricevono dai creatori di strumenti che supportano la traduzione in vari modi è sufficiente per avere successo (48%, rispetto al 40% nel 2021), ma c'è ancora il 21% degli intervistati che non è d'accordo con questa conclusione, che è maggiore rispetto al risultato del 2018 - 18%». La traduzione automatica mostra il tasso di crescita più elevato in termini di utilizzo, ma è ancora molto indietro rispetto alla memoria di traduzione. Meno del 10% degli intervistati prevede di investire nell'acquisto di tecnologie che supportano la traduzione, esclusa la traduzione automatica (11%).

E' stato fatto un **sondaggio tra i traduttori in Bulgaria**. Il sondaggio è composto da 26 domande sull'uso degli strumenti di traduzione assistita, la memoria di traduzione, la database terminologica e la traduzione automatica. E' stato condotto nel periodo tra il 01.02 e il 31.03.2022 e vi hanno partecipato un totale di 73 traduttori professionisti.

La percentuale maggiore di intervistati ha risposto di tradurre testi specializzati (90,4%), seguita da coloro che traducono documenti amministrativi (69,9%), narrativa (23,3%) e altri tipi di documenti (31,5%). In altre parole, una percentuale maggiore del gruppo degli intervistati si occupa di documenti adatti alla traduzione assistita. Il 46,5% dei traduttori utilizza costantemente programmi di traduzione assistita e l'8,2% - relativamente spesso. Quasi la metà degli intervistati utilizza programmi di traduzione assistita raramente o mai (Figura 1).

¹³ <https://elis-survey.org/>

Използвате ли инструменти за компютърнопомогнат превод?

73 отговора

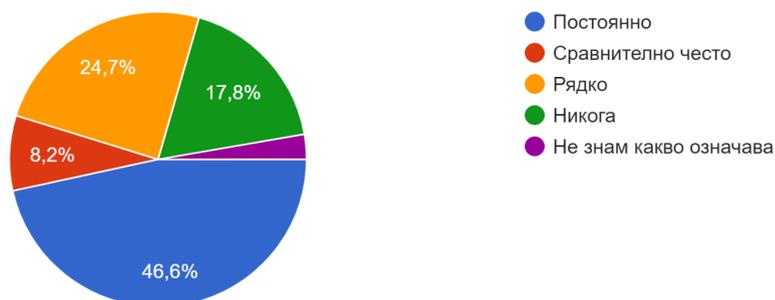


Figura 1: frequenza di utilizzo degli strumenti di traduzione assistita

I programmi più diffusi per la traduzione assistita sono Trados, Memsource, memoQ, Wordfast, SmartCAT. Quasi la metà degli intervistati li ha giudicati molto utili (49,3%) e più di un quarto – piuttosto utili (26%). Circa un quarto dei traduttori non ha opinioni o pensa che i programmi di traduzione assistita non siano utili. Da questi numeri si può concludere che i programmi di traduzione assistita sono utilizzati dai traduttori bulgari.

Alla domanda «Come pensi che gli strumenti di traduzione assistita dovrebbero essere sviluppati e migliorati?» hanno risposto in totale 55 partecipanti su 73. Qui sono elencate alcune delle risposte ricevute: «Avere adeguate istruzioni anche in bulgaro. Non tutti i traduttori usano l'inglese»; «Arricchire il vocabolario e includere espressioni idiomatiche, lavorando per una migliore grammatica e precisione lessicale»; «È necessario lavorare meglio con le memorie di traduzione e le databasi terminologiche; possibilità di ricerca diretta dei termini» ecc. Le risposte mostrano che (intuitivamente o no) i traduttori hanno aspettative realistiche per lo sviluppo di programmi di traduzione assistita e anche requisiti abbastanza ben definiti per le funzionalità di questi programmi, il che è (senza dubbio) dovuto alla loro esperienza di lavoro con questi.

Alla domanda se usano la memoria di traduzione, il 47,9% degli intervistati ha risposto che la usa costantemente e il 12,3% - relativamente spesso, questo dà più del 60% in totale a favore dell'uso della memoria di traduzione.

Le risposte alla domanda se i traduttori utilizzano databasi terminologiche sono paragonabili alle risposte alla domanda se utilizzano la memoria di traduzione: il 31,5% li utilizza costantemente e il 32,9% relativamente spesso, sempre oltre il 60%. Da ciò si può concludere che vengono utilizzate le basi, che si riferiscono alle risposte aperte sull'uso di programmi di traduzione assistita, secondo le quali alcuni traduttori vogliono avere un collegamento per cercare i termini nei repository terminologici. La mancanza di adeguate risorse terminologiche è un altro motivo per trovare un modo facile, universale e accessibile per creare databasi terminologiche.

Il 67,8% dei traduttori che utilizza le basi li crea autonomamente, il 39,4% utilizza le basi dei clienti, il 26,8% utilizza le basi delle agenzie e il 49,3% utilizza quelle disponibili pubblicamente.

Il 41% dei partecipanti al sondaggio utilizza basi aperte al pubblico. Qui va tenuto conto del fatto che in molti casi i traduttori necessitano di databasi terminologiche specializzate che non si riflettono nelle fonti pubbliche indicate. Quest'ultimo è a sostegno della possibilità di crearli da soli.

Alla domanda «Secondo lei, qual è l'importanza delle databasi terminologiche per la qualità della traduzione?» Il 43% ha risposto relativamente grande e il 37% - grande. Da queste risposte si può concludere che lo sviluppo di una metodologia unificante per la creazione di databasi terminologiche può essere vantaggioso per il lavoro quotidiano dei traduttori.

I risultati delle risposte alla domanda se i traduttori usano la traduzione automatica sono interessanti: il 42,5% la usa raramente e il 34,2% mai. Solo il 5,5% dei traduttori ha dichiarato di utilizzare regolarmente la traduzione automatica. Queste risposte possono essere dovute alla qualità ancora insoddisfacente della traduzione automatica, ma anche a una possibile riluttanza ad ammettere che vengono utilizzate tecnologie che dovrebbero (se non sostituire completamente) facilitare notevolmente il lavoro dei traduttori.

Esiste un'ampia varietà di risposte alla domanda «Come controlla la qualità di una traduzione?». Solo il 17,8% ha risposto di utilizzare le funzionalità di controllo della qualità degli strumenti di traduzione assistita; 9,6% utilizza programmi esterni, il 61,6% (il gruppo più numeroso) controlla personalmente il testo.

Le risposte dei traduttori bulgari non differiscono significativamente dagli altri sondaggi presentati. I traduttori cercano di utilizzare nuove tecnologie come la memoria di traduzione e le basi termiche, in quanto possono contribuire a migliorare la velocità e, in una certa misura, la coerenza della traduzione. D'altra parte, ci sono ancora traduttori che non solo non usano le nuove tecnologie, ma esprimono anche la loro preoccupazione nei loro confronti. Va inoltre segnalata la mancanza di databasi terminologiche disponibili al pubblico in molte aree della conoscenza umana, nonché l'insufficiente qualità della traduzione automatica, soprattutto da lingue diverse dall'inglese. Il sondaggio tra i traduttori bulgari, il cui lavoro quotidiano è legato alle tecnologie di traduzione assistita, conferma queste conclusioni.

CAPITOLO DUE

Le databasi terminologiche nei programmi di traduzione assistita

I. Termini, dizionari terminologici e databasi terminologiche

Secondo M. Popova, «i **termini** sono lo strato più informativo nel vocabolario del linguaggio, della parola e del testo scientifici (Popova 2019: 149). Le definizioni di cosa sia un termine saranno esaminate in vista dell'uso dei termini nelle databasi terminologiche.

M. Baker offre la seguente definizione di termine: «I termini si differenziano dalle parole in quanto hanno una forma speciale di riferimento, vale a dire, si riferiscono a singole unità concettuali, proprietà, attività o relazioni che si riferiscono alla conoscenza di una determinata area disciplinare. Per distinguere tra riferimento speciale e generale, viene fatta una distinzione tra termini che hanno un riferimento speciale all'interno di una particolare disciplina e parole che funzionano con riferimento generale in diverse aree tematiche» (Baker 2001: 216). J. Sager offre la seguente definizione di termine: «I termini differiscono dalle parole in quanto hanno una forma speciale di riferimento, vale a dire, si riferiscono a unità concettuali separate, proprietà, attività o relazioni che si riferiscono alla conoscenza di una particolare area disciplinare. Per distinguere tra riferimento speciale e generale, si fa una distinzione tra termini che hanno un riferimento speciale all'interno di una particolare disciplina scientifica e parole che funzionano con riferimento generale in diverse aree tematiche» (Sager 2001: 261).

A. Kis dà una definizione decisamente più breve: «il termine (l'autore lo chiama *terminus technicus*) è un'espressione appartenente al linguaggio tecnico» (Kis 2005: 106). A. Imre osserva che definizioni diverse enfatizzano aspetti diversi: alcune tengono conto della "forma" della parola o dell'espressione, mentre altre si concentrano sul significato. Un termine deve avere un solo significato, escludendo i sinonimi, ma soprattutto, il significato dei termini è ben stabilito (nessuna sovrapposizione con il significato di altri termini) in modo che non possa essere ampliato o ristretto (Imre 2013: 127).

Esiste una distinzione tra comprensione ampia e ristretta della terminologia: nella comprensione ampia della terminologia, un termine è considerato qualsiasi parola o frase che esprime un concetto utilizzato in un particolare contesto (Trumble e Stevenson 2002); la comprensione ristretta presuppone che un termine sia una parola o una frase usata con un significato ben definito in una data area tematica o disciplina.

I Dizionari terminologici sono caratterizzati da definizioni, parole e significati esperti che appartengono a una specifica area disciplinare tecnica o scientifica (come la medicina o la matematica) (De Caluwe e Van Santen 2003: 82). La scelta dei termini per il dizionario ne determina la portata tematica, e il suo ruolo conoscitivo si realizza o attraverso la rappresentazione delle relazioni semantiche tra termini e concetti, o attraverso informazioni semantiche implicite veicolate attraverso le definizioni dei termini inclusi. Da questo punto di vista, il dizionario terminologico diventa un mezzo di trasferimento di conoscenze professionali (Łukasik 2012: 100).

I dizionari terminologici possono includere uno o più dei seguenti componenti (Popova 2016: 71): il campo scientifico del termine; unità di vocabolario - intestazione; informazioni lessicali-grammaticali – come parte del discorso; informazioni linguistiche formali - caratteristiche grammaticali; funzione; dati etimologici; informazioni sulla compatibilità; parametro pragmatico – caratteristiche di utilizzo ed estensione della distribuzione; parametro illustrativo; parametro di registrazione (la data di registrazione del termine); parametro interpretativo; parametro di sistematicità concettuale-semantic; parametro di sistematicità lessicale-semantic; parametro temporale; parametro metadialeale (i termini di una data scuola scientifica); parametro idioletto (i termini di un dato studioso); parametro areale; parametro categoriale; informazioni sulla fonte.

Secondo ISO 30042:2008 **database terminologica** è «un banca dati che contiene informazioni su concetti linguistici specializzati e sui termini che denotano tali concetti, insieme a informazioni aggiuntive». In contesti commerciali, tuttavia, va notato che la database terminologica non si limita a "concetti linguistici specializzati" ma può contenere qualsiasi elemento lessicale che deve essere "gestito" per ottenere contenuti coerenti o di qualità. Un termbase è un base di termini relativi a un campo (o progetto) specifico integrato in uno strumento di traduzione assistita.

Di per sé, le basi sono come dei dizionari, ma il modo in cui vengono create, gestite e funzionano li rende fundamentalmente diverse. La loro caratteristica principale è che sono personali e personalizzabili per ogni istituzione, azienda privata e singolo traduttore. I loro creatori scelgono come configurarli, per quali combinazioni linguistiche, quali termini aggiungere, rimuovere, vietare e come classificare questi termini (Faber e Araúz 2021: 588).

In generale, le databasi terminologiche possono includere una o più delle seguenti componenti (Wright 2001: 573): definizione (una descrizione del significato del termine); fonte; contesto d'uso del termine; area a cui appartiene il termine¹⁴; informazioni grammaticali (verbo, sostantivo, ecc.); etichetta d'uso (ad esempio: figurativo, inglese americano, formale, ecc.); autore ("creato da" - l'utente che ha aggiunto il nuovo termine; in caso di utenti diversi, il traduttore sa quale termine da chi è stato creato/aggiunto); data di creazione/modifica ("creato/modificato il..."); stato ("verificato", "approvato", "preferito" o "proibito, da non utilizzare"); collegamenti/collegamenti tra singoli termini e definizioni, nonché collegamenti a fonti esterne; appunti.

Una precisazione necessaria è che i criteri di selezione dei termini dipendono dai requisiti del progetto specifico. Per questo motivo, si suggeriscono i seguenti passaggi per identificare se un termine è appropriato per un particolare progetto:

- Il termine appartiene o meno alla lingua ufficiale;
- Verificare se il termine è usato o meno;
- Se il termine è usato con lo stesso significato;
- Se c'è il rischio di un uso improprio del termine;
- Se l'uso del termine non richiede ulteriori riferimenti;
- Se ci sono sinonimi e quali differenze di significato porta ciascuno di essi.

È possibile trovare una grande quantità di risorse terminologiche liberamente disponibili. Di seguito sono elencati alcuni di loro. Alcuni sono di proprietà di istituzioni europee, altri di organizzazioni non governative.

- **IATE** (Terminologia interattiva per l'Europa¹⁵) è la risorsa terminologica più famosa per i traduttori in Bulgaria, come confermato dal sondaggio condotto tra di loro. Sulla

¹⁴ Mentre i dizionari elencano tutti o almeno i significati più popolari di un determinato termine, le basi terminologiche preferiscono l'univocità. Ogni base è creata per un'area, un cliente o un progetto specifico. Non si tiene conto di ciò che il termine dato significa in altri campi.

¹⁵ <http://iate.europa.eu>

piattaforma è possibile trovare un totale di 8.047.139 termini a luglio 2022. Sono supportate tutte le lingue ufficiali dell'Unione Europea.

- **Eurovoc**¹⁶ è un thesaurus multilingue e multidisciplinare che copre la terminologia dei settori di attività dell'UE. Come IATE, questa risorsa è disponibile nelle 24 lingue ufficiali dell'UE, nonché nelle lingue dei tre paesi candidati (Albania, Macedonia del Nord e Serbia).
- **lettropedia**¹⁷ è una banca dati della Commissione elettrotecnica internazionale (IEC) e fornisce il maggior numero al mondo di termini e definizioni nel campo dell'energia elettrica e dell'elettronica, contenente più di 22.000 voci terminologiche in inglese e francese, organizzate per area tematica e con termini equivalenti in varie altre lingue.
- **Il grande dizionario terminologico** (dal francese¹⁸ Le grand dictionnaire terminologique o GDT) è una banca terminologica creata dall'Ufficio della lingua francese del Quebec (da francese Office québécois de la langue française). Ogni termine si riferisce a un concetto relativo a un campo di utilizzo specializzato, con termini presentati in francese e inglese, e talvolta in altre lingue.
- **La database terminologica multilingue delle Nazioni Unite** (United Nations Multilingual Terminology Database¹⁹ o UNTERM) contiene termini relativi al lavoro delle Nazioni Unite. Le informazioni sono fornite nelle sei lingue ufficiali delle Nazioni Unite (arabo, cinese, inglese, francese, russo e spagnolo), e sono presenti anche voci in tedesco e portoghese.

II. Confronto di databasi terminologiche con dizionari terminologici

Quando si confrontano databasi terminologiche e dizionari terminologici online, è possibile determinare tra loro le seguenti caratteristiche comuni: a) Possibilità di riferirsi ad altri termini attraverso collegamenti tra loro; b) Possibilità di inserire immagini e altri aiuti nella voce per ogni termine; c) Di fronte ad un problema terminologico, i traduttori possono accedere rapidamente a databasi terminologiche ed a più dizionari se le informazioni ricercate sono disponibili; d) Possibilità di aggiornamento delle informazioni.

¹⁶ <https://op.europa.eu/en/web/eu-vocabularies>

¹⁷ <https://www.electropedia.org/iev/iev.nsf/Welcome?OpenForm>

¹⁸ <http://www.oqlf.gouv.qc.ca/>

¹⁹ <http://unterm.un.org>

Tuttavia, le somiglianze finiscono qui. Durante la traduzione, le databasi terminologiche facilitano, e molto spesso eliminano, la scelta di un determinato termine da parte del traduttore. È abituale inviare ai traduttori una base già preparata con termini approvati dal cliente da utilizzare nel testo di destinazione. Se si traccia un'analogia con i dizionari terminologici, il cliente assume il ruolo di compilatore perché è lui che determina cosa includere in ciascuna base. Un'altra differenza è l'accessibilità. Le databasi terminologiche sono disponibili solo negli strumenti di traduzione assistita (tranne quando una database o parte di essa è stata esportata per essere utilizzata per uno scopo specifico). Nella maggior parte dei casi, i dizionari online sono liberamente accessibili. Ultimo ma non meno importante, il modo di lavorare è diverso. Durante la traduzione nell'editor di testo del programma di traduzione assistita, il traduttore non deve cercare i termini. Il programma stesso li visualizza.

I dizionari possono essere suddivisi in online e cartacei. Per quanto riguarda i dizionari stampati, anche se affrontano molta concorrenza, sono ben lungi dall'essere respinti. Mentre le databasi terminologiche sono individuali per ogni azienda, progetto e traduttore, i dizionari terminologici tradizionali e le loro versioni online portano unificazione e conferma del significato dei termini. I loro creatori sono solitamente specialisti autorevoli con reputazione consolidata. Quest'ultimo non può dirsi per le databasi terminologiche, laddove, nonostante chiare indicazioni sull'uso o il divieto dell'uso di un dato termine, non sono escluse controversie tra committente e traduttore circa l'aggiunta o la rimozione di un determinato termine, il suo uso o il suo significato.

Un'alternativa alle databasi terminologiche sono le edizioni cartacee dei dizionari terminologici. Un esempio di questi, che include la lingua bulgara, è (senza pretese di elenco esaustivo): dizionario legale inglese-bulgaro, Hristo Danov, casa editrice Trud, 1991; Dizionario terminologico inglese-bulgaro degli studi a distanza, Romyana Kancheva, casa editrice accademica Prof. Marin Drinov, 2020; Dizionario inglese-bulgaro di cardiologia, Valentina Mincheva, casa editrice accademica Prof. Marin Drinov, 1991; Dizionario agrario, commerciale e forestale inglese-bulgaro, Tsvetelina Tsakova, casa editrice accademica Prof. Marin Drinov, 2012; Dizionario marittimo inglese-bulgaro con collocazioni, Vyara Petkova, casa editrice Steno, 2006; Dizionario tematico multilingue dell'integrazione europea, casa editrice Kolibri, 2007.

La particolarità di queste risorse è che ogni dizionario rappresenta un'area tematica diversa e ha un ambito diverso. Anche nei casi in cui più glossari rappresentano un'unica disciplina, il loro contenuto e la loro credibilità sono difficili da confrontare. Anche le ragioni di ciò sono molto diverse. L'intervallo di tempo negli anni di pubblicazione può essere limitante per il loro utilizzo. Anche le limitazioni alla circolazione possono costituire un ostacolo alla disponibilità di queste risorse. Proprio per quest'ultimo motivo, anche le risorse bilingue su Internet (compreso il bulgaro) sono un'alternativa, ad esempio: dizionario online inglese-bulgaro²⁰; Dizionario online PONS inglese-bulgaro²¹; Dizionario SA²².

I problemi con le risorse terminologiche su Internet derivano (con poche eccezioni) principalmente dalla mancanza di informazioni sulla loro affidabilità e credibilità. Nella maggior parte dei casi, non è chiaro se siano stati creati da esperti o meno. Inoltre, non è chiaro se gli elementi lessicali inclusi soddisfino i criteri del termine e i criteri di selezione. La differenza nell'ambito e nella selezione dei termini, così come gli specialisti che li hanno selezionati, è un altro fattore molto importante.

III. Collegamento delle databasi terminologiche con la memoria di traduzione

La memoria di traduzione è solitamente integrata nei programmi di traduzione assistita. Ogni frase tradotta è memorizzata nella memoria. Quando il traduttore incontra un nuovo segmento che corrisponde a un altro dalla base, la traduzione viene recuperata e offerta per essere riutilizzata (Tayebeh 2008: 97).

La memoria di traduzione può essere utile per arricchire la database terminologica. Le corrispondenze parziali possono aiutare a trovare dei termini simili a quelli utilizzati nel termbase. Il processo di creazione di databasi terminologiche utilizzando la memoria di traduzione non è automatizzato. Mentre l'aggiunta di tutti i segmenti alla memoria è un passaggio obbligatorio per il completamento di una traduzione, l'aggiunta di termini è un processo completamente manuale che si basa sul giudizio o sulle istruzioni del traduttore. Alcuni termbase hanno l'ulteriore vantaggio di offrire suggerimenti automatici sui termini.

²⁰ <https://www.rechnik-bg.com/>

²¹ <https://bg.pons.com/%D0%BF%D1%80%D0%B5%D0%B2%D0%BE%D0%B4>

²² <http://notrial.bg/software/windows/sa-dictionary/>

Con loro, il programma identifica possibili nuovi termini e li offre al traduttore per aggiungerli.

IV. Creazione e gestione di databasi terminologiche nei programmi di traduzione assistita

Qualsiasi programma moderno di traduzione assistita include la disponibilità di funzioni di gestione della terminologia. I modi per creare una database terminologica variano a seconda del programma utilizzato, ma ciò che hanno tutti in comune è che ogni programma di traduzione assistita ha impostazioni che consentono agli utenti di creare tutte le basi che desiderano. In Trados, ad esempio, la creazione di una nuova base è un passaggio che è parte dalla creazione di un nuovo progetto. All'interno del programma il traduttore ha la possibilità di tradurre un singolo file o creare un progetto. Prima di iniziare il lavoro di traduzione vero e proprio, il traduttore seleziona una delle basi già esistenti o ne crea una nuova. Questo viene fatto in una nuova finestra, dove passo dopo passo il traduttore sceglie se aggiungere file, basi online, ecc. Impostazioni aggiuntive abilitano l'importazione rapida, in cui il file è completamente compatibile con il formato XML MultiTerm. Ciò consente di risparmiare tempo controllando le voci nel file di importazione. La funzione "Esegui riorganizzazione completa dopo l'importazione" (Perform full reorganisation after import) viene utilizzata quando si aggiornano basi esistenti con termini nuovi o modificati. Con questo metodo di funzionamento, è possibile ottenere errori dopo l'importazione come duplicati indesiderati o ricerca parziale che non funziona correttamente.

I terminologi sono assistiti da strumenti "intelligenti" come software di estrazione di termini, sistemi di indicizzazione automatica e programmi di generazione automatica di testo. Nell'ultimo decennio, vari autori hanno evidenziato caratteristiche importanti di come le tecnologie moderne come la realtà virtuale e la terminologia basata su cloud (Varga 2012) si applicano alla tecnologia terminologica.

Oltre alle funzionalità standard relative alle databasi terminologiche nei programmi di traduzione assistita, i traduttori hanno la possibilità di installare ausili aggiuntivi (**gadget**). Non fanno parte del software di base perché non tutti i programmi di traduzione assistita li supportano e molti sono a pagamento. La maggior parte si connette al programma principale

tramite un plug-in, ma alcuni possono funzionare in modo indipendente. Un esempio di tali gadget sono²³: LogiTerm²⁴; Juremy²⁵; RWS AppStore²⁶.

Sistemi di gestione della terminologia (Terminology Management Systems o TMS) sono progettati per raccogliere, conservare e accedere ai dati terminologici. Possono essere sia indipendenti che parte dei programmi di traduzione assistita.

Esistono anche programmi dedicati alla localizzazione. La loro struttura differisce dagli strumenti di traduzione assistita. Attraverso questi la traduzione di contenuti web, applicazioni mobili e contenuti di file viene automatizzata.

Le lingue controllate sono state utilizzate con successo in molti campi negli ultimi 20 anni come metodo per migliorare la leggibilità dei documenti tecnici e per facilitarne la traduzione nella lingua dei loro clienti. «Un vocabolario controllato è la disposizione organizzata di parole e frasi utilizzate per l'indicizzazione dei contenuti e/o il recupero dei contenuti. I termini preferiti e le loro varianti possono essere aggiunti al vocabolario controllato e possono avere un ambito definito o descrivere un'area specifica»²⁷. Il vantaggio delle lingue controllate è che consentono agli autori di creare documenti di facile lettura e più coerenti in termini di vocabolario e stile utilizzato.

V. Capacità di lavorare con la terminologia e le databasi terminologiche

Il profilo del terminologo professionista²⁸ copre le conoscenze e le abilità relative alla conoscenza dei principi della terminologia (teorica e pratica), alla creazione e all'individuazione di risorse terminologiche per scopi e gruppi specifici, eccellenti capacità linguistiche, capacità di ricerca e capacità di identificare informazioni importanti, abilità nell'uso di databasi terminologiche, nonché come eccellente padronanza dei sistemi di gestione della terminologia, strumenti di traduzione assistita, software di traduzione automatica e strumenti di estrazione di termini²⁹. Devono inoltre avere conoscenze di

²³ Tutti gli esempi di tali programmi sono presi dall'elenco Nimzi è anche noto come:

<https://www.nimdzi.com/tbs/>

²⁴ <https://terminotix.com/index.asp?content=category&cat=4&lang=en>

²⁵ <https://juremy.com/about>

²⁶ <https://appstore.rws.com/>

²⁷ https://www.getty.edu/research/publications/electronic_publications/intro_controlled_vocab/what.pdf

²⁸ <https://termcoord.eu/why-terminology/31318-2/>

²⁹ <http://clearwordstranslations.com/language/en/7-terminology-management-best-practices/>

ontologie e reti semantiche. Competenze specialistiche come il trasferimento di conoscenze e insegnamento, conoscenza di aree tematiche speciali, capacità di pianificare e gestire processi e progetti, gestione del cambiamento, capacità di prendere e attuare decisioni strategiche in coordinamento con altre unità dell'organizzazione³⁰ sono necessari.

Un numero crescente di università offre opportunità di specializzazione e qualificazioni. Esempi di questi sono: aspetti della traduzione e dell'interpretazione relativi alla terminologia³¹, terminologia ed esigenze professionali³², lo sviluppo delle competenze traduttive di base e il conseguimento di competenze nel riconoscimento di registri in diverse lingue³³, lo sviluppo della ricerca in settori quali la traduzione tecnica, letteraria e dei media, l'interpretazione di conferenza e la gestione della terminologia³⁴.

³⁰ https://terminorgs.net/downloads/Terminology_skills.pdf

³¹ Master in terminologia presso l'Università del Lussemburgo
https://www.uni.lu/studies/fhse/master_in_learning_and_communication_in_multilingual_and_multicultural_contexts/programme

³² Programma di terminologia online presso l'Istituto di Linguistica Applicata, un centro di ricerca dell'Università Pompeo Fabra <https://www.upf.edu/en/web/terminologiaonline>

³³ Master in traduzione e studi terminologici presso l'Università di Malta
<https://www.um.edu.mt/courses/overview/PMTTFTT-2021-2-O>

³⁴ Centro per la ricerca sui cavi dell'Università di Vienna <https://transvienna.univie.ac.at/en/about-us/>

CAPITOLO TRE

Creazione semiautomatica di databasi terminologiche multilingui

Un'area che ha attirato notevole attenzione è l'estrazione automatica dei termini (Kilgarriff 2014 et al.; Costa et al. 2016). I primi metodi utilizzavano solo informazioni linguistiche come i modelli morfosintattici. Successivamente sono stati sviluppati metodi statistici sempre più sofisticati, applicando vari tipi di algoritmi di frequenza e collocazione. Ad un certo stadio dello sviluppo dell'estrazione automatica dei termini, compaiono sistemi ibridi che combinano entrambi i tipi di informazioni: linguistica e statistica.

Decidere quali termini includere nella base è una sfida importante e K. Warburton consiglia di basarsi sull'evidenza di corpus. Il processo di identificazione dei termini, inserendoli in una database terminologica, e l'aggiunta di informazioni varia a seconda che l'approccio scelto sia tematico (questo approccio presuppone che i termini e i loro concetti sottostanti siano parte di un sistema, e quindi la loro esistenza e il loro significato possono essere solo confermati dal confronto con altri termini nello stesso sistema), o il cosiddetto approccio ad-hoc, in cui viene risolto un problema specifico, di solito facendo riferimento a un solo concetto e alla sua denominazione (Warburton 2015: 652).

Le tecnologie di estrazione automatica non solo consentono ai terminologi di identificare più termini di quanto sarebbe possibile manualmente, ma l'uso di queste tecnologie aumenta la corrispondenza tra i termini nella base e nel corpus di cui intendono riflettere il contenuto. Riduce anche parzialmente la soggettività perché include termini che sono effettivamente usati nei corpora.

Le risorse linguistiche per le lingue piccole sono difficili da trovare. In gran parte del trasferimento interlinguistico, l'inglese è utilizzato come lingua di partenza (Søgaard et al. 2015: 1713). «L'inserimento di anglicismi si osserva anche nel vocabolario comune, ad esempio *toster* (*мочеп*), *stiker* (*цмукер*), *bidiguard* (*бодузард*), così come nel gergo giovanile» (Blagoeva, Koeva e Murdarov 2012: 14-15).

I. Creazione di risorse terminologiche nel campo della terminologia informatica

La definizione della terminologia "corretta" è legata allo sviluppo delle risorse terminologiche e alla loro gestione. I traduttori utilizzano la terminologia per rendere le loro traduzioni più accurate e coerenti.

Una caratteristica distintiva di **terminologia informatica** è il costante progresso e sviluppo delle tecnologie, che richiede anche l'aggiornamento e l'integrazione dei termini utilizzati (Hoisl et al., 2015). L'introduzione, la diffusione e l'utilizzo della terminologia informatica non è un fenomeno nuovo. Già all'inizio del nuovo millennio, L. Kirova affermava che: «Il vocabolario informatico si è sviluppato in modo incredibile in un periodo di soli vent'anni. Quanto più aumenta il numero di utenti della tecnologia informatica, tanto più decisiva diventa la loro partecipazione alla formazione del linguaggio per lavorare con esso» (Kirova 2002).

Per i termini che sono parole composte, la struttura sintattica più comune è la struttura di un aggettivo subordinato e di un nome principale (Koeva 2007: 61) (es. *риболовен сезон* - *stagione di pesca*), anche le strutture di un nome, una preposizione e un nome che formano un gruppo preposizionale subordinato sono relativamente comuni (ad esempio *подобряване на почвата* - *miglioramento del suolo*), ciascuno dei nomi - principale o subordinato può essere spiegato con aggettivi (ad esempio *електронен трансфер на фондове* - *trasferimento elettronico di fondi*), in alcuni casi possono esserci più aggettivi (ad esempio *европейска парична система* - *sistema monetario europeo*), ci sono anche combinazioni di parole di due sostantivi in cui il secondo è la parte principale (ad esempio *тенис корт* - *campo da tennis*).

I termini possono anche essere classificati come neutri nell'uso stilistico (ad esempio *търсеща система* - *motore di ricerca*), termini condivisi da un certo gruppo di specialisti in un dato campo (ad esempio *метатърсачка* - *metamotore di ricerca*), e gergo professionale (ad esempio *паяк* - *ragno*) (Kirova 2004).

Le abbreviazioni sono comuni nella (terminologia informatica). Come affermato: «le abbreviazioni prese in prestito dall'inglese iniziano a funzionare come parole complete e sempre più spesso devono entrare nei dizionari» (Kirova 2004). Ci sono vari casi di trasmissione di un'abbreviazione straniera in lingua bulgara - in alcuni casi viene adottata l'abbreviazione (*джиесем* o *gsm*, *есемес* - *sms*), e in altri possono essere usati sia

l'abbreviazione che il nome espanso (*циди - cd, компактдиск - compact disc*) (Kolkovska 2010: 2 – 3).

I dizionari cartacei in inglese e bulgaro nel campo della terminologia informatica hanno una portata e una struttura diverse e per gli stessi termini sono date definizioni diverse, ecc. I dizionari stampati con autori chiaramente indicati sono più autorevoli per quanto riguarda la qualità delle informazioni incluse. D'altra parte, anche se vengono ripubblicati, i dizionari cartacei diventano obsoleti molto rapidamente. Il problema con i dizionari su Internet è l'incertezza nella loro qualità, la mancanza di informazioni su come vengono selezionate le unità, se appartengono davvero alla terminologia informatica. Per questi motivi può essere utile anche la raccolta di dizionari monolingui in inglese e bulgaro nel campo della terminologia informatica.

I dizionari di terminologia informatica in lingua bulgara si distinguono per un numero limitato di termini, mancanza di rappresentatività e autori non noti, in molti casi presentati in modo errato dal punto di vista dell'ortografia bulgara del termine o mancanza di presentazione in lingua bulgara e definizioni.

Sul web e nelle librerie è possibile trovare numerosi dizionari in inglese e bulgaro nel campo della terminologia informatica. Esempi di questi sono:

- *Компютърен речник А-З - Dizionario informatico AZ*³⁵ (edizione cartacea del 2003;
- *Английско-български компютърен речник: А-З - Dizionario informatico inglese-bulgaro: А-З*³⁶ (edizione cartacea 2017;
- *English-Bulgarian Computer Terms*³⁷ (una base contenente 729 termini in inglese e le loro traduzioni in bulgaro;
- *Преводно-тълковен речник на някои компютърни термини - Un dizionario di traduzione-interpretazione di alcuni termini informatici*³⁸ (un sito in cui non sono indicate le fonti di informazione e le citazioni necessarie;
- Nikolava D. *Словарь компьютерных терминов / Речник на компютърните термини - Dizionario di termini informatici / Dizionario di termini informatici.*

³⁵ <https://m.helikon.bg/43115-Компютърен-речник-А-З.html> [20.07.2022]

³⁶ <https://www.book.store.bg/p5397/anglijsko-bylgarski-kompiutyren-rechnik-a-z.html> [20.07.2022]

³⁷

<https://elrc-share.eu/repository/browse/english-bulgarian-computer-terms-processed/f9d2a4e44afb11e9a7e100155d02670615329808b1384ab388941fcd4fba5029/> [20.07.2022]

³⁸ <http://kafence.com/Преводно-тълковен-речник-на-някои-компютърни-термини-201.html> [20.07.2022]

Shumen: Chimera, 2016. 258 pp. ISBN 978-619-7218-27-5. Il dizionario presenta termini popolari nel campo della tecnologia dell'informazione: dispositivo informatico, attrezzatura; informatica; prodotti software; lavoro col computer; Internet;

Sebbene questa presentazione probabilmente non copre tutti i dizionari disponibili che presentano la terminologia informatica in inglese, emergono le seguenti tendenze: i dizionari creati da grandi editori o organizzazioni contengono informazioni esaustive e affidabili; inoltre, ci sono molte fonti su Internet che offrono informazioni, ma solo gli specialisti possono certificare in alcuni casi se i termini e le definizioni sono appropriati.

Sono disponibili anche dizionari terminologici informatici per il bulgaro, ma sono elettronici:

- *Компютърни термини, които трябва да знаем (кратък списък с термини) - Termini informatici che dobbiamo conoscere*³⁹ (breve elenco di termini);
- *Речник на компютърните термини и акроними - Glossario dei termini informatici e degli acronimi*⁴⁰ (contiene oltre 250 termini e le loro definizioni);
- *Речник на компютърни термини по азбучен ред - Dizionario alfabetico dei termini informatici. Основни компютърни термини - Termini informatici di base. Il programma e la sua installazione su un computer*⁴¹ (il dizionario si trova su un sito di notizie di tecnologia dell'informazione, ancora una volta, alcuni termini e definizioni non sono tradotti correttamente in bulgaro e ci sono anche ripetizioni di errori da altre fonti.

I dizionari con terminologia informatica in lingua bulgara si distinguono per un numero limitato di termini, mancanza di rappresentatività e autore non indicato, in molti casi presentati in modo errato dal punto di vista dell'ortografia bulgara del termine o mancanza di presentazione in lingua bulgara e definizioni, che può anche non rappresentare adeguatamente il concetto descritto, e in alcuni casi ci sono chiare indicazioni di traduzione automatica.

³⁹ <https://urocibg.eu/компютърни-термини-които-трябва-да-зн/> [20.07.2022]

⁴⁰ <https://bg.chalized.com/речник-на-компютърните-термини-и/> [20.07.2022]

⁴¹

<https://roidmod.ru/bg/lighting/slovar-kompyuternyh-terminov-po-alfavitu-osnovnye-kompyuternye-terminy-programma-i-e-ustanovka-n/> [20.07.2022]

Sebbene alcuni dei dizionari elencati siano utili, possono rendere difficile il lavoro dei traduttori a causa del diverso scopo, della diversa struttura e, soprattutto, del fatto che agli stessi termini vengono date definizioni diverse (anche se i traduttori sono consapevoli della qualità del dizionario che usano). Un altro fattore che rende i dizionari esistenti non sempre adatti alla creazione di databasi terminologiche è il focus della traduzione specifica.

La questione dell' **estrazione automatica dei termini** non è nuova. Già nel 1996, K. Kageura e B. Umino hanno rivisto i principi e i metodi del riconoscimento automatico dei termini (Kageura e Umino 1996). I due mostrano due tendenze principali nel recupero dei termini: uno è il recupero delle informazioni e l'altro è il riconoscimento automatico dei termini.

L'estrazione automatica dei termini viene effettuata dai corpora linguistici che dovrebbero rappresentare una determinata area tematica. La ricerca e la raccolta di testi rilevanti può non essere un compito facile perché la qualità dei termini recuperati dipende anche da essi. Ai fini dell'estrazione automatica dei termini, vengono raccolte anche grandi quantità di documenti in modo automatico o semiautomatico. Come indicato, per alcune lingue le risorse disponibili in formato elettronico potrebbero non essere tante. Alcuni metodi di estrazione dei termini richiedono anche un ampio corpus di riferimento per illustrare il vocabolario di uso comune e il confronto con cui consentire un'adeguata identificazione dei termini. L'estrazione di termini in due o più lingue riguarda l'allineamento di parole e frasi parallele in corpora paralleli.

La questione dell'estrazione automatica dei termini non è nuova. Già nel 1996, K. Kageura e B. Umino hanno rivisto i principi e i metodi del riconoscimento automatico dei termini (Kageura e Umino 1996). I due mostrano due tendenze principali nell'estrazione dei termini: uno è il recupero delle informazioni e l'altro è il riconoscimento automatico dei termini. A. Šajatović ed al. suddividono i metodi di estrazione dei termini in: metodi di frequenza (basati sull'assunto che una maggiore frequenza di un termine candidato implichi una maggiore probabilità che si tratti effettivamente di tale termine); metodi contestuali (basati sul contesto in cui compare il termine candidato); modellazione tematica (le informazioni tematiche possono essere utilizzate al posto della frequenza) e riclassificazione (i metodi di questo gruppo utilizzano altri metodi per estrarre termini come caratteristiche e cercano di valutare l'importanza di ciascuna caratteristica) (A. Šajatović et al. 2019: 150).

Alcuni strumenti adottano un approccio statistico, quindi cercano semplicemente le ripetizioni del testo. In questo modo, tuttavia, esiste il rischio di estrarre potenziali termini che non sono validi (noti anche come "rumore"). Un altro approccio all'estrazione di termini è quello linguistico, per cui il programma cerca strutture come "nome + nome" o "aggettivo + nome" ecc. Un altro esempio di estrazione di termini è lo strumento di TM-Town⁴², una piattaforma di collaborazione tra traduttori e clienti. Il programma estrae automaticamente i termini chiave dai testi di origine e di destinazione. I termini sono valutati in base a un algoritmo che calcola la rilevanza di ciascun termine nel documento. Ciò avviene attraverso algoritmi statistici di elaborazione del linguaggio naturale che analizzano il contenuto dei documenti e identificano le parole chiave più rilevanti. Un punteggio di "Term Rank" più elevato indica una maggiore importanza del termine nei documenti. Dopo aver estratto i termini principali dal testo di destinazione, l'utente abbina il termine di origine con la traduzione corrispondente.

Nella maggior parte degli **strumenti di estrazione dei termini** da due lingue, i termini candidati nella lingua di partenza sono identificate sulla base di modelli predefiniti. Dopo di che vengono selezionati i termini candidati equivalenti nella traduzione nella lingua di destinazione. Un approccio alternativo consiste nel generare termini candidati direttamente dalle parole e frasi allineate nel corpus. Le informazioni sulla frequenza e gli n-grammi determinano quali parole o parole composte sono termini. In questo approccio, i modelli predefiniti dipendenti dalla lingua non vengono presi in considerazione, ma le frasi sono collegate in base a corrispondenze lessicali e somiglianza sintattica (Lefever 2009: 496).

Il confronto degli strumenti di estrazione automatica dei termini mostra che non ci sono molti strumenti disponibili gratuitamente, e quelli che offrono l'estrazione di termini in lingua bulgara sono ancora meno. Fivefilters (fivefilters.org⁴³) offre funzioni di recupero della terminologia posizionando il testo da elaborare o il collegamento Internet ad esso nella posizione designata. È possibile recuperare fino a 100 termini in cinque diversi formati. Questo strumento supporta anche la lingua bulgara.

In uno studio del 2016, A. Zaretskaya et al. (2016: 18) suggeriscono **una comparazione di strumenti di estrazione automatica dei termini**. Vengono analizzati i seguenti programmi indipendenti, che sono definiti come il "tipo di software più popolare" al

⁴² <https://www.tm-town.com/> [20.07.2022]

⁴³ <https://www.fivefilters.org/term-extraction/> [20.07.2022]

momento. In particolare, vengono confrontate le seguenti funzionalità: *Estrazione bilingue* (Bilingual extraction) per l'estrazione simultanea di termini dalla lingua d'origine e di destinazione; *Confrontare i contesti della lingua d'origine e di destinazione* (Source and target context comparison); *Controllo del termine* (Terms validation); *Completamento del dizionario bilingue* (Bilingual dictionaries compilation) per creare un dizionario bilingue; *Estrazione di termini contestuali* (Context extraction); *Supporto per vari formati di file* (Support various file formats); *Ordinazione di frequenza dei termini* (Rank terms by frequency); *Supporto multilingue*; *Impostazione di un numero minimo di occorrenze per termine* (Specify the minimum number of occurrences) per impostare il numero minimo di occorrenze di un determinato termine; *Visualizzazione delle informazioni linguistiche* (Show linguistic information); *Impostazione del numero minimo e massimo di traduzioni* (Specify the maximum number of translations) per il numero di equivalenti di traduzione di un dato termine; *Opzione da disattivare* (Stopword list option) per escludere parole che non possono essere termini (es. preposizioni); *Scelta del numero minimo e massimo di parole nella composizione di un termine* (Choose the minimum and maximum number of words per term); *Statistiche* (Term statistics).

Dal confronto tra i diversi programmi risulta chiaro che tutti includono informazioni statistiche. Solo tre dei nove offrono l'estrazione bilingue. SDL Multiterm offre la possibilità di confrontare i contesti della lingua di origine e di destinazione. Funzionalità come l'estrazione di termini contestuali, il supporto per diversi formati dei file, l'ordine di frequenza dei termini, l'impostazione di un numero minimo di occorrenze di un termine e il supporto per lingue diverse sono possedute da tutti i programmi senza Translated. La visualizzazione delle informazioni linguistiche è supportata da Terminus, SDL e Termsuite. La definizione di un numero minimo e massimo di traduzioni è una funzionalità di Termsuite. Un totale di 5 programmi su 9 (SDL Multiterm, Simple Extractor, Kia, Rainbow e JATE) forniscono la possibilità di escludere termini e di scegliere il numero minimo e massimo di parole nella composizione di un termine.

Seguendo il modello dello studio svolto da A. Zaretskaya et al., nel 2022, ai fini di sviluppare questa tesi, sono stati analizzati gli stessi strumenti per verificarne lo sviluppo. Le nuove funzionalità sono: *Estrazione di termini multilingui* (Multilingual term extraction); *Verifica della corrispondenza tra il testo e la database terminologica* (Consistency check between the text and the termbase) per verificare se viene utilizzata la terminologia della

base; *Possibilità di estrarre manualmente i termini* (Manual terms extraction); *Impostazioni del progetto* (Project settings); *Recuperare un rapporto di qualità* (QA report); *Estrazione monolingue* (Monolingual extraction) per estrarre i termini solo dalla fonte o dalla prova di traduzione; *Lavorare con software esterno* (Third-party software access); *Impostazione del numero massimo di termini estratti* (Select Maximum number of extracted terms); Allineamento bilingue per abbinare la migliore traduzione di un termine dalla lingua di partenza alla lingua di destinazione; *Possibilità di effettuare ricerche con espressioni regolari* (Regex search); *Struttura concettuale* (Conceptual structure) per rappresentare un albero concettuale.

Dalle nuove funzionalità descritte, Sketch Engine ha ora l'estrazione contestuale, la visualizzazione delle informazioni linguistiche e un'opzione per escludere i termini; SDL e Rainbow supportano l'impostazione di un numero minimo e massimo di traduzioni; Translated offre la possibilità di scegliere un numero minimo e massimo di parole nella composizione dei termini.

II. Metodologia per la creazione semiautomatica di una database terminologica nel campo della terminologia informatica

Esiste un gran numero di strumenti per l'estrazione della terminologia, ma il bulgaro non è supportato dalla maggior parte di loro. La situazione è ancora più complicata quando si tratta di un campo specializzato: inoltre, non ci sono molte risorse linguistiche disponibili per il bulgaro. Queste limitazioni sono la ragione per proporre una metodologia per la creazione semiautomatica di una database terminologica nel campo della terminologia informatica.

Come è già chiarito, il primo passo per creare una database terminologica è trovare una fonte affidabile. Date le limitazioni e le caratteristiche sopra descritte, per quanto riguarda **l'estrazione di termini**, la risorsa più appropriata è **il corpus bilingue**.

Allo scopo di sviluppare la tesi, è stato selezionato il corpus parallelo inglese-bulgaro nel campo della terminologia informatica QTLeap (QTLeap V1.2)⁴⁴, che contiene 4.000 coppie di domande e risposte nel campo della risoluzione dei problemi informatici (sia

44

<http://metashare.ilsp.gr:8080/repository/browse/qt leap-corpus-v12/0176c39ae9cd11e4a2aa782bcb074135ba7d767f645a48dca1d50ee3c9504253/> [20.07.2022]

hardware che software). Va notato che ci sono diversi svantaggi: il corpus è piccolo, i testi e i termini in contenuti al suo interno non sono diversi, il numero di termini specializzati non è elevato. La licenza d'uso del corpus è per uso accademico e non commerciale, richiede attribuzione (CC-BY-NC-SA). È stato realizzato nel periodo 2014-2015 con finanziamenti europei. L'ultima modifica risale al 2016. Il corpus è stato scelto perché è bilingue e corrisponde all'area terminologica prescelta. Va notato che ci sono diversi inconvenienti: il corpus è piccolo e i testi e i termini in esso contenuti non sono vari. Va inoltre notato che il numero di termini specializzati non è elevato, ma ci sono piuttosto termini che, attraverso la diffusione della tecnologia nella vita di tutti i giorni, sono entrati in uso.

Per essere selezionato uno strumento per **l'estrazione di termini da un corpus**, deve soddisfare i seguenti requisiti: supportare il bulgaro e la lingua di origine da/a cui è tradotto (obbligatorio); essere gratuito (consigliato); essere un programma autonomo o una funzionalità integrata di un software di traduzione assistita (gratuito) (richiesto per le persone senza competenze di programmazione aggiuntive); avere un'interfaccia intuitiva (consigliato); disporre di impostazioni per la scelta del numero minimo e massimo di parole di cui dovrebbe essere composto un determinato termine (consigliato); offrire i termini estratti in un formato che consenta di convertirli facilmente (consigliato).

Dopo la ricerca, è stato riscontrato che Fivefilters soddisfa i criteri specificati. Il programma gestisce l'estrazione dei termini inglesi. Offre risultati insoddisfacenti per il bulgaro, anche se è linguisticamente indipendente. Il programma riconosce solo i nomi di prodotti e marchi scritti in alfabeto latino.

Una componente chiave del fatto che un programma di estrazione di termini troverà impiego è se supporta il bulgaro, che non è comune al momento (soprattutto per gli strumenti open source). Programmi come Multiterm supportano un numero enorme di lingue, ma sono a pagamento e il loro utilizzo è direttamente legato alle licenze Trados. Pertanto, l'uso di Multiterm è limitato a quei traduttori che hanno una licenza.

Nonostante l'esistenza di un numero enorme di strumenti di estrazione della terminologia, trovarne uno che soddisfi i requisiti dichiarati è un compito piuttosto difficile.

IAKE⁴⁵ (Yet Another Keyword Extractor) utilizza molteplici funzioni statistiche ed è un programma indipendente dalla lingua. Soddisfa anche i criteri stabiliti e per questo motivo

⁴⁵ <http://yake.inesctec.pt/demo/user> [20.07.2022]

si è cercato di estrarre termini con esso. Anche in questo caso i risultati non sono buoni, perché il programma riconosce solo i nomi di prodotti (Facebook, Google Chrome) e alcuni verbi bulgari.

Considerato quanto sopra, l'approccio di estrazione dei termini è stato modificato, con i criteri descritti volti a lavorare con una combinazione di programmi piuttosto che uno strumento singolo per estrarre i termini in entrambe le lingue. A tale scopo è stato scelto un programma che supporta l'estrazione di termini in inglese: Termostato Web 3.0⁴⁶. Il programma estrae termini singoli e termini di parole composte, ne mostra la frequenza, il grado di specificità e le forme delle parole che ricorrono nel testo (Drouin 2003).

Candidat de regroupement	Fréquence	Score (Spécificité)	Variantes orthographiques	Matrice
facebook	136	145.96	facebook	Nom
laptop	123	136.52	laptop laptops	Nom
internet	110	130.55	internet	Nom
iphone	91	119.17	iphone	Nom
password	90	114.7	password passwords	Nom
wireless	93	113.3	wireless	Nom
router	80	111.65	router routers	Nom
ipad	74	107.33	ipad	Nom
email	73	106.59	email	Nom
file	182	104.81	file files	Nom
computer	258	99.66	computer computers	Nom
pc	109	96.86	pc pcs	Nom
google	58	94.84	google	Nom
folder	58	85.6	folder folders	Nom
itunes	43	81.41	itunes	Nom
wireless network	42	80.44	wireless network wireless networks	Nom Nom
hard drive	37	75.37	hard drive hard drives	Adjectif Nom
powerpoint	36	74.32	powerpoint	Nom
website	34	72.16	website	Nom activate Windows
wifi	33	71.06	wifi	Nom to Settings to activate Windows.
antivirus	32	69.94	antivirus	Nom
usb	31	68.81	usb	Nom

Figura 2: Parte dei termini estratti dal corpus con il programma Termostat

Come si può vedere nella Figura 2, il programma estrae i termini, ne visualizza la frequenza, la specificità, determina a quale parte del discorso appartengono ed estrae le loro forme che ricorrono nel testo. I criteri stabiliti per l'estrazione automatica dei termini sono: estrarre solo i sostantivi (comprese le parole composte), estrarre il numero massimo di termini riconosciuti dal programma. Naturalmente, nell'elenco dei termini inglesi estratti erano presenti anche parole che non sono termini e ciò ha richiesto la loro rimozione manuale. Ad esempio: *wont*, *much ram*, *doesnt*, *someone*, *much space* e *old file*.

⁴⁶ <http://termostat.ling.umontreal.ca/interfaceTermostat.php> [20.07.2022]

La possibilità di estrarre contemporaneamente la terminologia bilingue può essere aggiunta alle funzionalità per lo sviluppo di un programma rivolto al bulgaro. Esiste una tale possibilità (sviluppata dalla società lettone Tilde⁴⁷), che richiede competenze di programmazione, il che limita la possibilità di un'applicazione più ampia⁴⁸.

Dopo l'identificazione dei termini inglesi, sono state trovate le loro traduzioni dalla parte bulgara del corpus utilizzato. A tale scopo, i testi inglese e bulgaro sono allineati con MateCat⁴⁹, un programma che soddisfa anche i criteri stabiliti per la scelta di uno strumento appropriato. Dopo che i testi sono stati visualizzati nel programma MateCat in modo che le frasi siano una accanto all'altra, i termini inglesi vengono cercati come parola chiave in modo da poter trovare il loro equivalente bulgaro⁵⁰.

Dopo aver determinato i termini equivalenti per il bulgaro, i risultati vengono inseriti in una tabella excel. Sebbene Termostat consenta l'estrazione di file in formato .txt, è stato scelto Excel per la sua accessibilità e perché i formati dei file possono essere importati direttamente in qualsiasi programma di traduzione assistita. Il numero di termini inglesi estratti corrisponde a quello dei termini bulgari. Le abbreviazioni in inglese vengono lasciate in bulgaro.

E' fatta un'**ispezione manuale e il layout della database terminologica**. A seconda delle esigenze del cliente o delle capacità tecniche del programma di traduzione assistita, i formati della database terminologica possono variare. Esistono anche numerosi strumenti autonomi mediante i quali la database terminologica estratta può essere convertita da un formato all'altro. Un esempio di tali sono: TBX Convert⁵¹, Glossary Converter⁵², ApSIC Xbench⁵³. Per il presente esperimento, è stato usato il programma Olifant⁵⁴.

Dopo la conversione nel formato desiderato (in questo caso in tmx), la database terminologica può essere importata nel progetto di traduzione. MateCat è stato scelto come strumento per l'allineamento e la traduzione.

⁴⁷ <https://www.tilde.com/>

⁴⁸ <https://github.com/tilde-nlp/taws>

⁴⁹ <https://www.matecat.com/> [20.07.2022]

⁵⁰ L'allineamento può essere utilizzato per qualsiasi tipo di ricerca di equivalenti di traduzione con uno scopo diverso, ad esempio stilistico.

⁵¹ <https://www.tbxconvert.gevterm.net/glossary/index.html> [20.07.2022]

⁵² <https://appstore.sdl.com/language/app/glossary-converter/195/> [20.07.2022]

⁵³ <https://www.xbench.net/index.php/download> [20.07.2022]

⁵⁴ <http://okapi.sourceforge.net/downloads.html> [20.07.2022]

Sono necessarie le seguenti precisazioni:

1) La conversione da un formato all'altro per creare un termbase non è sempre necessaria. È sufficiente che il documento sia correttamente formattato (in due colonne adiacenti, con il codice della lingua in testa a ciascuna colonna e i termini della lingua di partenza e di destinazione, rispettivamente);

2) Il metodo proposto non è ottimale e non risolve tutti i problemi. Mostra solo uno dei modi possibili per risolvere il problema nelle condizioni di mancanza di risorse e programmi sufficienti per l'elaborazione in lingua bulgara. Un obiettivo per ulteriori ricerche è lo sviluppo di uno strumento che possa rispondere in modo ancora più adeguato alle esigenze dei traduttori.

III. Descrizione delle caratteristiche della database terminologica creata nel campo della terminologia informatica

La nuova database terminologica contiene i termini estratti dal corpus in inglese e i loro equivalenti in bulgaro. Vengono aggiunti tre campi aggiuntivi (per sinonimi, note e parti del discorso), ma non vengono fornite altre informazioni come definizioni, immagini, riferimenti incrociati tra termini o fonti esterne. Al momento, solo un termine nella lingua di destinazione corrisponde a un termine di origine.

Da un punto di vista morfologico, nella base sono presenti parole semplici e composte, e tutti i termini sono sostantivi. I termini estratti variano da una frequenza massima nel corpus parallelo di 136 a una frequenza minima di 3.73.

Quando si esaminano i risultati estratti, è possibile effettuare la seguente analisi linguistica: in inglese ci sono parole composte che vengono trasmesse in bulgaro come due parole separate (ad esempio *cellphone* – *мобилен телефон*); abbreviazioni dall'inglese o sono rese allo stesso modo (ad esempio *Wifi* – *Wifi*⁵⁵), o sono trasmesse con l'intera ortografia in bulgaro (ad esempio *pc* – *personal computer*). Vengono annotate le parole complesse in inglese, che in bulgaro vengono rese come parole composte (es. *antivirus* - programma

⁵⁵ E' necessario chiarire che gli errori di ortografia sono possibili anche nella lingua di partenza, che vengono poi trasferiti nella lingua di destinazione. C'è il rischio che influenzino la qualità complessiva della traduzione del testo.

antivirus). Dal bulgaro all'inglese, non ci sono molti di questi esempi (*URL – indirizzo URL*). Ci sono termini formati dalla calchi, per esempio *playlist – плейлист*, oltre a salvare le abbreviazioni come *ram – рам*.

Va chiarito che la base appena creata non offre sinonimi per i termini e che non ci sono definizioni o spiegazioni per le parti del discorso. La ragione di ciò è che non è buona pratica usare sinonimi quando si traducono termini, a meno che non ci sia un motivo o un'istruzione specifica per farlo. Ad esempio, il termine *рутер - router* ha un sinonimo nella parola bulgara *маршрутизатор*, il termine *изтегляне - scaricare* ha un sinonimo ma a seconda dello scopo della traduzione, può essere utilizzato il termine più comunemente usato o il termine preferito nel discorso letterario. Sebbene i sinonimi mostrino la ricchezza e lo sviluppo del linguaggio, non sono stati presi in considerazione durante la creazione di una database terminologica, perché testi standardizzati e uniformi rendono obbligatorio osservare e utilizzare coerentemente gli stessi termini per un progetto e un testo specifici.

La metodologia proposta in questo capitolo per lo sviluppo delle databasi terminologiche non offre una soluzione universale per migliorare e facilitare il lavoro dei traduttori in relazione all'uso appropriato della terminologia. Ha i suoi limiti (la selezione manuale dei corpora e la ricerca manuale degli equivalenti tradotti) e non ha la pretesa di essere completo, soprattutto se confrontato con le capacità dei programmi di estrazione dei termini per altre lingue. Si tratta comunque di un primo tentativo in questa direzione con l'obiettivo di supportare lo sviluppo di software specializzati per l'estrazione di termini e la creazione di database terminologiche bilingue, in cui una lingua è il bulgaro.

CAPITOLO QUATTRO

Misurare la qualità nei programmi di traduzione assistita e nelle loro parti

Misurare la qualità della traduzione è un compito estremamente difficile perché non esiste un'unica traduzione ideale per un dato testo. Ad esempio, una traduzione legale ha requisiti molto diversi in termini di accuratezza e conformità a norme legislative specifiche rispetto a un testo pubblicitario o un manuale d'uso. Le specificità del testo tradotto dipendono non solo dal traduttore, dal testo da tradurre o dalla sua interpretazione soggettiva, ma anche dai motivi della traduzione, dai potenziali lettori e dalla politica di pubblicazione e marketing, ovvero di fattori che migliorano la traduzione come procedura linguistica. Ciò include fattori sociali, nonché vincoli socioculturali, politici o ideologici, che possono avere un impatto significativo sulla traduzione (House 2015: 153). Secondo S. Colina, «La mancanza di consenso su come definire la qualità di una traduzione deriva dai disaccordi sul concetto di traduzione e dalla natura contraddittoria e relativa della qualità...» (Colina 2020: 458). Secondo K. Chunyu e Wong Tak-ming, la valutazione della qualità della traduzione richiede la comprensione del contenuto del testo da utilizzare per determinare i diversi tipi di relazioni di equivalenza e l'identificazione degli errori di traduzione (Chunyu e Wong Tak-ming 2015: 220-221). Gran parte della critica ai vari approcci di valutazione riguarda la loro dipendenza dalla comprensione dell'"equivalenza". Già alla fine degli anni '90 H. Hönig affermava che accettare l'equivalenza come misura a priori della qualità non è corretto, perché il tipo e il grado di equivalenza variano a seconda delle condizioni in cui si effettua la traduzione (Hönig 1997: 9).

Un'altra difficoltà è che la valutazione della qualità della traduzione è effettuata da persone, quindi la componente soggettiva del "fattore umano" è ancora più pronunciata (Zehnalova et al. 2013: 43). Pertanto, la valutazione della traduzione non si basa su una teoria specifica, ma su punti di vista personali (Colina 2020: 458). La qualità della traduzione può anche essere vista come la misura in cui la traduzione segue le specifiche concordate⁵⁶. Questa escursione segue effettivamente le raccomandazioni degli standard internazionali di

⁵⁶ <https://www.communicaidinc.com/a-10-strategic-translation.php>

qualità. Pertanto, la qualità della traduzione dovrebbe essere definita come relativa (piuttosto che assoluta) rispetto a un quadro concordato in anticipo tra il cliente e il traduttore. Il traduttore non può terminare la traduzione senza prima eseguire un controllo di qualità (a seconda delle impostazioni e del progetto di traduzione) e risolvere eventuali errori e incongruenze. Il controllo di qualità viene svolto prima dalla persona che ha eseguito la traduzione, il traduttore stesso, e poi dal redattore, dal cliente o da altri operatori (Gouadec 2007: 74).

I. Misurazione dell'accuratezza della traduzione con programmi di traduzione assistita

Va notato che per ora non esiste un programma per computer che possa sostituire un redattore professionista. La tecnologia linguistica può solo facilitare il lavoro dei redattori. Se il controllo della qualità della traduzione non ha rilevato errori, ciò non garantisce che la traduzione sia (completamente) corretta. Tuttavia, non tutti gli errori sono sempre veri "errori". È possibile che i segmenti che il software riconosce come contenenti un errore siano effettivamente corretti. Il traduttore (o il redattore) deve decidere a seconda del caso. Poiché gli errori nella lingua di destinazione possono avere un impatto significativo sulla qualità della traduzione, la loro valutazione include violazioni nella struttura della lingua di destinazione e della fraseologia, che dovrebbero essere considerate come "errori di traduzione", vale a dire problemi comuni osservati nella creazione di testi (Hansen 2010: 386).

Al fine di verificare la misurazione dell'accuratezza della traduzione con databasi terminologiche, nonché dimostrare le possibilità di controllo della qualità nei programmi di traduzione assistita (con particolare attenzione alla lingua bulgara), sono stati condotti due esperimenti: per il primo, sono stati selezionati 10 testi specialisti (con un volume di circa 15.000 parole) di regolamenti comunitari⁵⁷, tradotto da esperti dall'inglese al bulgaro, per la cui traduzione viene utilizzata una base già pronta; per il secondo sono stati utilizzati testi del corpus (circa 20.000 parole) utilizzati per creare la database terminologica descritta nel terzo capitolo.

⁵⁷ <https://eur-lex.europa.eu/homepage.html>

Per quanto riguarda il primo esperimento, i testi in inglese e le loro traduzioni in bulgaro sono stati allineati con lo strumento Trados aligner e la terminologia è stata estratta manualmente (50 termini in totale). I termini vengono poi rappresentati nella loro forma di base sia nella lingua di partenza che in quella di destinazione perché la lemmatizzazione è un metodo utilizzato nella creazione di dizionari e databasi terminologiche. I seguenti errori sono stati introdotti nel corpus utilizzato:

- cancellazione di un termine (come simulazione del caso in cui un termine non viene tradotto dal traduttore);
- rappresentazione parziale di un termine (quando il termine è una parola composta), consistente nella cancellazione di una delle parole parte del termine;
- traduzione incoerente (uso di termini diversi per trasmettere lo stesso concetto).

Tutti gli errori inseriti sono possibili in una situazione reale.

Trados e memoQ (i programmi più diffusi per la traduzione assistita) sono stati utilizzati per controllare gli errori introdotti artificialmente nell'uso della terminologia. Per comodità, tutte le altre impostazioni di controllo qualità di entrambi i programmi sono disattivate ad eccezione del controllo terminologico. Con le loro impostazioni predefinite, i programmi rilevano solo il termine nella sua forma base. L'esperimento è stato eseguito da due traduttori con Trados e memoQ su computer diversi.

Di seguito sono riportati i risultati dei rapporti sulla qualità estratti dai programmi testati. Gli errori inseriti vengono rilevati con alcune eccezioni. Con Trados, ad esempio, nella lingua di partenza il programma riconosce solo "наказателно производство - procedimento penale" e ignora il plurale del termine, che è anche un comportamento previsto⁵⁸. La stessa situazione si ripete nella lingua di destinazione. Ogni strumento di traduzione assistita ha una funzionalità di controllo della qualità (**misurazione dell'accuratezza della traduzione senza databasi terminologiche**). Il controllo di qualità viene svolto prima dalla persona che ha eseguito la traduzione, il traduttore stesso, e poi dal redattore, dal cliente o da altri operatori (Gouadec 2007: 74).

Controllare **la misurazione dell'accuratezza della traduzione con databasi terminologiche**, oltre che per dimostrare le possibilità di controllo della qualità nei programmi di traduzione assistita (soprattutto per quanto riguarda la lingua bulgara), sono

⁵⁸ Questo è anche un problema comune in altre lingue come lo svedese e il finlandese.

stati condotti due esperimenti: per il primo sono stati selezionati 10 testi specialistici (15.000 parole circa) dai regolamenti dell'UE⁵⁹, tradotto da esperti dall'inglese al bulgaro, in cui viene utilizzata una base già pronta; per il secondo, sono stati utilizzati testi del corpus parallelo (20.000 parole circa) per creare la database terminologica descritta nel terzo capitolo.

I seguenti errori aggiuntivi sono stati introdotti nei test per entrambi gli esperimenti: **eliminare un termine** (come simulazione del caso in cui un dato termine non viene tradotto dal traduttore); **rappresentazione parziale di un termine** (quando il termine è una parola composta), consistente nella soppressione di una delle parole parte del termine; **traduzione incoerente** (utilizzando termini diversi per trasmettere lo stesso concetto).

Tutti gli errori inseriti sono possibili in una situazione reale. I programmi di traduzione assistita da computer Trados e memoQ sono stati utilizzati per controllare gli errori introdotti artificialmente. Gli errori inseriti vengono rilevati con alcune eccezioni. L'analisi mostra che Trados riconosce il termine nella lingua di partenza solo al singolare, che è un comportamento previsto⁶⁰. La stessa situazione si ripete nella lingua di destinazione. Ad esempio:

Lingua di origine: ... *this Directive aims to strengthen the trust of Member States in each other's criminal justice systems and thus to improve mutual recognition of decisions in criminal matters.*

Lingua di destinazione: *настоящата директива има за цел да засили взаимното доверие на държавите членки в техните системи за наказателно правосъдие и по този начин да спомогне за постигане на по-широко взаимно признаване на решенията по наказателни производства.* - ... la presente direttiva mira a rafforzare la fiducia reciproca degli Stati membri nei loro sistemi di giustizia penale e quindi a contribuire a un più ampio riconoscimento reciproco delle decisioni in materia procedimento penale.

Il secondo esperimento mira ad una verifica tecnica della metodologia per la creazione della database terminologica per la terminologia informatica per vedere se può essere utilizzata dai programmi di traduzione assistita senza difficoltà tecniche. Il controllo dei termini è stato modellato sul primo esperimento e copri gli stessi errori introdotti artificialmente come la cancellazione di un termine, la cancellazione parziale di una delle

⁵⁹ <https://eur-lex.europa.eu/homepage.html>

⁶⁰ Questo è anche un problema comune in altre lingue come lo svedese e il finlandese.

parole, parte del termine e l'uso consecutivo di termini diversi per trasmettere lo stesso concetto. I termini sono stati visualizzati correttamente e il risultato è stato lo stesso dell'esperimento precedente. Il programma rileva l'assenza dei termini o la loro parziale presenza in tutti gli errori inseriti artificialmente. Si ripete però la situazione sopra descritta, con mancato riconoscimento del plurale dei termini, indicando che ogni forma deve essere presente come voce nella base

Gli esperimenti condotti verificano le funzionalità di controllo della qualità nei programmi di traduzione assistita per la lingua bulgara, nonché il comportamento della database terminologica appena creata. Aggiungendo artificialmente gli errori riscontrati in un ambiente di lavoro reale, è stato possibile dimostrare che la database terminologica può essere utilizzata con successo.

II. Confronto dell'efficacia dei programmi di valutazione della traduzione

I programmi di controllo della qualità esterni agli strumenti di traduzione assistita svolgono la stessa funzione di quelli integrati, con la differenza che dispongono di una gamma molto più ampia di funzioni. I seguenti programmi sono tra i più comuni: QA Distiller⁶¹, Xbench⁶² Verifika, ErrorSpy e Linguistic toolbox o Ltb⁶³. Il vantaggio principale di questi programmi è che i possibili errori relativi alla terminologia e all'allineamento sono contrassegnati in modo diverso. Funzioni dei programmi sopra elencati: *Segmenti vuoti* (Empty segments); *Il testo di destinazione corrisponde al testo di partenza* (Target text matches the source text) per la corrispondenza tra la lingua di arrivo e quella di partenza; *Incoerenza nei tag* (Tag mismatch); *Discrepanza nei numeri* (Number mismatch); *Errore grammaticale* (Grammar); *Incoerenza nei link* (URL mismatch); *Errori ortografici* (Spelling); *Incoerenza nella corrispondenza alfanumerica* (Alphanumeric mismatch); *Doppio carattere mancante* (Unpaired symbols) per la presenza di un simbolo di apertura o di chiusura, ad esempio virgolette; *Traduzione parziale* (Partial translation) per un numero minimo di parole consecutive non tradotte; *Carattere vuoto ripetuto* (Double blanks); *Parole ripetute* (Repeated words); *Coerenza nel testo di partenza* (Source consistency) per la traduzione uniforme di diversi segmenti di origine; *Coerenza nel testo di destinazione* (Target

⁶¹ <http://www.qa-distiller.com/en>

⁶² <https://www.xbench.net/>

⁶³ <http://autoupdate.lionbridge.com/LTB3/>

consistency) per segmenti identici nella lingua di partenza con traduzione diversa; *Rapporto sulle modifiche apportate* (Change report); *Elaborazione di più file contemporaneamente* (Multiple files); *CamelCase* per la corrispondenza delle abbreviazioni; *Terminologia* (Terminology) per l'uso coerente dei termini in tutto il testo; *Lista di controllo* (Checklist) – un elenco precedentemente concordato tra il traduttore e il cliente, in cui sono descritte le fasi per la verifica della qualità; *PowerSearch* – una modalità di ricerca che utilizza espressioni regolari; *Profili* (Profiles) – impostazioni personalizzate per il controllo di qualità, ortografia e grammatica selezionate per un utente specifico; *Rapporto* (Report) per la visualizzazione degli errori rilevati dal programma nel testo; *Pannello dei comandi* (Command line) per lavorare senza l'interfaccia utente grafica del programma; *Elenco di parole da non tradurre* (Do Not Translate List o DNT List).

Rispetto alle capacità di controllo della qualità dei programmi di traduzione assistita, i programmi specializzati hanno un numero significativamente maggiore di funzioni. È stato effettuato un confronto delle capacità tecniche dei programmi menzionati sulla base della loro documentazione tecnica, con ciascuna funzione elencata.

Tutti i programmi hanno funzioni come ortografia, discrepanze numeriche, punteggiatura e tag. Tutti i programmi rilevano anche la presenza di segmenti vuoti nella lingua di destinazione, l'assenza di un simbolo particolare nella lingua di destinazione e l'uso coerente della terminologia. Tuttavia, altre funzioni sono inerenti solo a un programma specifico, come la possibilità di recuperare un rapporto sulle modifiche apportate (solo Etb c'è l'ha). QA Distiller, Xbench e ErrorSpy non riconoscono gli errori grammaticali; ErrorSpy non registra la traduzione parziale, il carattere vuoto ripetuto, l'elaborazione di più file contemporaneamente e non offre una lista di controllo. QA Distiller non rileva le parole ripetute, Ltb non ha l'opzione di PowerSearch. A parte QA Distiller, nessuno dei programmi offre un pannello di comando e Xbench e ErrorSpy non offrono la possibilità di personalizzare le impostazioni. Tutte queste differenze devono essere prese in considerazione quando si utilizzano programmi di controllo qualità per selezionare quello più appropriato per gli scopi di ogni progetto di traduzione.

Lavorando con i programmi di controllo qualità esterni e con i programmi di traduzione assistita, è stata notata un'altra cosa: ognuno di loro ha classificato gli errori rilevati in modo diverso. Ad esempio, un errore di battitura è un'omissione molto più grave di un doppio carattere vuoto. In Trados gli errori si dividono in: errore (Error), attenzione

(Warning) e nota (Note), con espressioni numeriche mancanti o errate appartenenti ad un errore, punteggiatura diversa alla fine della frase all'attenzione, e doppio blank simbolo alle note. In Xbench, la divisione è tra errori di base (Basic) - segmento non tradotto o discrepanza nel testo, errori nel testo (Content) - discrepanze di punteggiatura, doppi caratteri in bianco, numeri mancanti o errati e altri, lista di errori (Checklist) e controllo ortografico (correttore ortografico). QA Distiller determina la gravità dell'errore in base a una scala numerica. Ad esempio: un doppio carattere vuoto riceverà Gravità 1, un errore di punteggiatura riceverà Gravità 2, mentre un'omissione o incoerenza nella traduzione riceverà Gravità 5.

III. Classificazione degli errori

Il fatto che ogni programma attribuisca un'importanza diversa agli errori che rileva è un motivo per suggerire la seguente **classificazione degli errori**, nel tentativo di analizzare gli errori e le loro implicazioni nel mondo reale per la qualità della traduzione, la classificazione è unificata e basata sui possibili effetti negativi che l'errore dato può avere sulla qualità della traduzione.

Scala	Descrizione	Tipo di errore
1	Ha un impatto minimo o nullo sulla qualità della traduzione; raramente può portare a confusione	Mancata corrispondenza della punteggiatura: <ul style="list-style-type: none"> - simboli di chiusura (parentesi, virgolette) - assenza di un carattere, assenza di entrambi i caratteri; assenza di uno o entrambi i segni e presenza di un altro segno; presenza di caratteri diversi intorno a una parola; - un segno di punteggiatura diverso, mancante o ridondante alla fine della frase; - caratteri vuoti ridondanti o mancanti (spazi, tab);

		<ul style="list-style-type: none"> - ridondante (mancanza di) nuova linea.
2	Può creare confusione, a seconda del contesto	<p>Incongruenze grafiche</p> <ul style="list-style-type: none"> - discrepanza nelle espressioni numeriche: espressioni numeriche mancanti, confuse o spostate (incoerenze in numeri, date, ore, frazioni, numerazione, ecc.). Tuttavia, in determinate situazioni, i programmi possono confondere qualcosa che è vero (conversione da gradi Celsius a Fahrenheit, localizzazione in diversi sistemi metrici, ecc.); - discrepanza nei collegamenti internet, e-mail, ecc.; - mancata corrispondenza nelle combinazioni di numeri e lettere.
3	Rischio di sostituire seriamente il contenuto del testo	<p>Incoerenze nella completezza della traduzione</p> <ul style="list-style-type: none"> - segmenti vuoti; - i segmenti nella lingua di destinazione e nella lingua di partenza corrispondono. Viene visto anche come testo non tradotto nei vari strumenti; - mancata corrispondenza delle maiuscole all'inizio del segmento o della frase; - lettere omesse, aggiunte, scambiate; - presenza o assenza di simboli specifici; - presenza di caratteri di un altro alfabeto nella parola.

4	Errore inaccettabile	<p>Controllo ortografico:</p> <ul style="list-style-type: none"> - parole ripetute; - parole omesse (rilevabili solo quando si fa riferimento ad abbreviazioni o parole tra virgolette, parentesi o in combinazione con numeri); - incoerenza nelle abbreviazioni (per rilevare tali incongruenze è necessario inserirle preventivamente nelle impostazioni); - mancanza di parti tradotte della frase (tale mancanza si riflette molto spesso confrontando la lunghezza delle frasi in %); - uso incoerente di parole e termini specifici nel testo. <p>Grammatica</p> <p>Solo alcuni strumenti sono dotati di tale funzione, e per le lingue più comuni. I programmi selezionati per l'esperimento non dispongono di tale funzione e questa non sarà verificata.</p>
---	----------------------	---

Tabella 1: Classificazione degli errori di traduzione e loro impatto sulla corretta traduzione del contenuto

Gli errori variano in importanza rispetto alle conseguenze per il corretto trasferimento e comprensione del contenuto durante la traduzione da una lingua all'altra.

IV. Un confronto dei risultati delle misure automatiche di qualità

È stato effettuato un confronto tra il modo in cui funzionano diversi programmi e il loro tasso di successo nel rilevare errori nella traduzione dall'inglese al bulgaro. A tale scopo sono stati utilizzati i programmi specializzati QA Distiller ed Xbench e le funzionalità di controllo qualità di Trados 2019 e memoQ 9.2. Sono stati utilizzati i testi con errori introdotti artificialmente utilizzati per il primo esperimento.

Gli errori di gravità 1 vengono sempre rilevati da tutti i programmi. Poiché si tratta di un carattere di chiusura mancante o di caratteri vuoti mancanti, tali errori hanno conseguenze minime sulla qualità della traduzione.

Vengono sempre rilevati anche gli errori di Gravità 2 (incoerenza tra cifre, collegamenti e combinazioni di cifre). Qui è necessario sottolineare che il contesto può dipendere dal fatto che si tratti di errori reali o fittizi. A questo gruppo appartiene anche una parte significativa dei “falsi errori” riscontrati:

Lingua di partenza: *It will be launch on March 17, 2020 at 2pm Bulgarian time, as the local authorities...*

Lingua di destinazione: *Той ще бъде отворен на 17 март 2020 г. в 14.00 ч. българско време като преди това местните власти... - Sarà inaugurato il 17 marzo 2020 a 14.00 h. Ora bulgara come in precedenza le autorità locali...*

Un esempio di "falso errore" con valore 3. Il programma ha rilevato l'assenza di parentesi nella lingua di destinazione e l'ha contrassegnata come un'omissione:

Lingua di partenza: ...territory of the Member State(s) shall not...

Lingua di destinazione: ...територията на държава/и-членка/и не - ...il territorio di uno o più Stati membri/S No

Quando si rilevano errori caratterizzati da un valore 4, si notano differenze nel comportamento dei singoli programmi. Xbench e QA Distiller non segnalano problemi di ortografia, mentre Trados e memoQ se la cavano decisamente meglio. Tuttavia, le parole corrette si riflettono come errori: *правоприлагане* - *forze dell'ordine*, *дихлоробензен* - *diclorobenzene*, *ароматизатори* - *aromi*, *договорка* - *accordo*, *киберсигурност* - *sicurezza informatica* e altri.

Ciò che è innegabile ed estremamente prezioso nella traduzione è il vantaggio di questi programmi quando si tratta di errori di tipo 4: la capacità di rilevare incoerenze nell'uso di parole specifiche che possono portare a perdita o confusione nell'interpretazione del significato. Riguarda l'uso coerente delle stesse parole o termini in tutto il testo della lingua di destinazione.

Il comportamento dei programmi quando si lavora con la lingua bulgara merita un commento a parte. In generale, alcune regole e segni di punteggiatura specifici non vengono presi in considerazione (più precisamente, i programmi non affrontano il corretto calcolo delle virgolette bulgare), che non sono presenti in altre lingue. Per i programmi, le virgolette bulgare sono errate a causa della loro differenza con quelle nella lingua di partenza:

Lingua di partenza: *“Almost 90 % of the inspected products do not contain substances...has an obligation to provide it when requested,”*

Lingua di destinazione: *„Почти 90% от проверените продукти не съдържат вещества...има задължението да я предоставя при поискване“* - "Quasi il 90% dei prodotti testati non contiene sostanze... ha l'obbligo di fornirlo su richiesta"

Il vantaggio dei programmi di controllo della qualità è innegabile. Tuttavia, per ottenere risultati ottimali sono sempre necessarie impostazioni personalizzate e giudizio umano. Nonostante questi programmi trovino errori "falsi", il loro principale vantaggio è la capacità di rilevare incoerenze nella traduzione (ad esempio, se gli stessi termini sono usati in tutto il testo).

Per ottenere i migliori risultati possibili con questi programmi, è imperativo impostare impostazioni specifiche per ogni traduzione. Proponiamo inoltre uno schema aggiuntivo per la valutazione degli strumenti di garanzia della qualità (tabella 2).

<p>Livello 1</p>	<p>interfaccia e funzioni di base</p>	<ul style="list-style-type: none"> - è uno strumento facile da usare - è autonomo o fa parte di un programma di traduzione assistita - quali file può gestire - quali lingue supporta - consente la personalizzazione
<p>Livello 2</p>	<p>funzioni frequentemente usate</p>	<ul style="list-style-type: none"> - come gestisce problemi comuni (caratteri mancanti, numeri, ecc.) - impostazioni predefinite - incoerenza nell'uso della terminologia - quantità di falsi errori
<p>Livello 3</p>	<p>impostazioni personalizzate</p>	<ul style="list-style-type: none"> - come gestisce la localizzazione - come gestisce le impostazioni per una particolare lingua rispetto a un'altra - lingua personalizzata/impostazioni specifiche del cliente: formato della data, elemento non traducibile, ecc. - lettura di caratteri speciali - quantità di errori "falsi" nelle impostazioni personalizzate

Livello 4	benefici aggiuntivi	<ul style="list-style-type: none"> - comandi audio - modifica del file di traduzione - altri vantaggi (ad es. collegamenti diretti tra diversi programmi)
------------------	---------------------	--

Tabella 2: Schema di valutazione per i programmi di gestione della qualità della traduzione

I programmi di controllo automatico della qualità (integrati nei programmi di traduzione assistita o esterni) sono utili perché, in una certa misura, facilitano il lavoro dei traduttori. Tuttavia, non bisogna dimenticare che non possono sostituire le persone. Quest'ultimo è comprovato dalla presenza di "falsi" errori.

V. Standard internazionali per la qualità della traduzione

Tutti i mezzi di misurazione della qualità, la sequenza in cui i file di traduzione devono essere rivisti ed elaborati, le competenze e le qualifiche dei traduttori, la comunicazione con i clienti e i contratti in base alle quali condizioni verrà eseguita la traduzione sono soggetti a standard di qualità internazionali. Sono adottati per unificare i processi di lavoro, la terminologia, le posizioni e definire le competenze dei traduttori e garantire la qualità della traduzione.

Ad oggi, gli attuali standard internazionali per la traduzione di documenti sono: ISO 9001:2015⁶⁴ – gestione della qualità dei sistemi; ISO 17100:2015⁶⁵ - servizi di traduzione; ISO27001:2013⁶⁶ – sistemi di sicurezza delle informazioni; ISO 18587:2017⁶⁷ – modifica della traduzione automatica.

Sebbene gli standard di traduzione esistano da tempo, è solo nel 2006 che le organizzazioni hanno iniziato a lavorare formalmente insieme per standardizzare il modo di

⁶⁴ <https://www.iso.org/iso-9001-quality-management.html>

⁶⁵ <https://www.iso.org/standard/59149.html>

⁶⁶ <https://www.iso.org/standard/54534.html>

⁶⁷ <https://www.iso.org/standard/62970.html>

lavorare e creare documenti standard ufficiali e schemi di certificazione per il campo della traduzione. Il loro obiettivo è la protezione e la consapevolezza degli utenti garantendo un certo livello di qualità della traduzione. Esistono diversi elementi comuni a tutti gli standard di traduzione: il livello richiesto o concordato di raggiungimento della qualità, la necessità di chiare specifiche di progetto concordate, un requisito per l'accordo tra cliente e fornitore di servizi di traduzione, processo di gestione del progetto, capacità tecnica, così come alcuni altri condizioni relative alla fornitura della traduzione (Bendana e Melby 2012: 81).

Conclusione

L'obiettivo principale della tesi è quello di analizzare le tecnologie e i mezzi utilizzati per creare risorse terminologiche nel contesto dei programmi di traduzione assistita e di proporre (senza pretese di esaustività) una metodologia unificata per la loro creazione.

A tal fine sono stati analizzati i programmi di traduzione assistita, le loro caratteristiche tecniche, i loro componenti, nonché i cambiamenti che questi strumenti impongono ai processi di lavoro, all'istruzione e alla formazione dei traduttori. La linea di fondo è che la memoria di traduzione, almeno per ora, è il componente più prezioso nei programmi di traduzione assistita a causa dell'archiviazione dei dati delle traduzioni precedenti. La database terminologica è importante anche per i traduttori, perché risparmia la necessità di cercare i termini appropriati e supporta la scelta del termine. La tecnologia che ribalta completamente le nozioni dell'attività di traduzione è la traduzione automatica, e il suo utilizzo aumenterà sempre di più.

Viene presentato il processo di creazione e gestione delle databasi terminologiche nei programmi di traduzione assistita da computer. I modi principali per creare e gestire databasi terminologiche sono integrati nei programmi di traduzione assistita. Oltre alle funzionalità integrate, sono disponibili numerosi gadget sviluppati che possono essere utilizzati sia insieme a programmi di traduzione assistita che indipendentemente.

Vengono presentati numerosi studi che esaminano le competenze e le preferenze dei traduttori all'estero. È stato creato un nuovo sondaggio, rivolto specificamente ai traduttori professionisti in Bulgaria. L'analisi dei risultati ha mostrato che (con poche eccezioni) i traduttori bulgari conoscono bene gli strumenti di traduzione assistita e si sforzano di usarli, anche se non tutti li trovano utili e accessibili (l'atteggiamento nei confronti della memoria di traduzione e delle databasi terminologiche è simile). I traduttori si aspettano che le funzionalità dei programmi di traduzione assistita siano sempre più intuitive e semplificate. La traduzione automatica neurale solleva alcune preoccupazioni sul fatto che, se continua a svilupparsi e migliorare, potrebbe rappresentare una competizione per il lavoro dei traduttori.

La tesi propone una metodologia per la creazione semiautomatica di un database terminologica multilingue in una data area e per determinate lingue. Nel caso specifico, la

database terminologica è bilingue, le lingue sono il bulgaro e l'inglese, e la direzione della traduzione è dall'inglese al bulgaro, e il campo è la terminologia informatica.

Nell'analisi della database terminologica creata nel campo della terminologia informatica, è dimostrato che il corpus di origine ha un'influenza significativa sulla qualità e completezza dei termini estratti, che, successivamente, possono influenzare la traduzione. Da un lato, poiché la metodologia si basa su corpora linguistici paralleli e sull'estrazione automatica di termini, si può sostenere che si propone un modo unificato di lavorare nella creazione di databasi terminologiche. D'altro canto, la selezione manuale degli equivalenti di traduzione fornisce un controllo di qualità dei termini selezionati e va notato che in alcuni casi ciò può richiedere molto tempo.

In generale, lo studio offre una descrizione, un'analisi e un confronto dei moderni strumenti utilizzati nella traduzione: i programmi di traduzione assistita, i loro componenti: memoria di traduzione, database terminologica, traduzione automatica, gli strumenti per l'estrazione automatica dei termini e per il controllo della qualità la traduzione. Tutto ciò è finalizzato al raggiungimento dell'obiettivo principale: sviluppare una metodologia per la creazione semiautomatica di risorse terminologiche per una determinata area specialistica.

Elenco delle pubblicazioni relative all'argomento della tesi

Strumenti e processi di valutazione della qualità della traduzione in relazione agli strumenti CAT. In *Proceedings of the Human-Informed Translation and Interpreting Technology Workshop (HiT-IT 2019)*, pages 89–97, Varna, Bulgaria. Incoma Ltd., Shoumen, Bulgaria.

Computer Terminology in Bulgarian Related with the Human Translation. In: *Сборник с доклади от Международната годишна конференция на Института за български език „Проф. Любомир Андрейчин“* (София, 2020), Том II, стр. 250–257, издателство на БАН „Проф. Марин Дринов“.

Caratteristiche e funzionamento degli strumenti di garanzia della qualità della traduzione umana. *Сборник с доклади от Международната годишна конференция на Института за български език „Проф. Любомир Андрейчин“* (София, 2021), Том II, стр. 225–233, издателство на БАН „Проф. Марин Дринов“.

BIBLIOGRAFIA

Arenas 2020: Arenas, A. G. Pre-editing and post-editing. – In: E. Angelone, M. Ehrensberger-Dow, & G. Massey (Eds.), *The Bloomsbury Companion to Language Industry Studies* (1 ed., pp. 333-360). (Bloomsbury Companions). Bloomsbury Accademico [<https://research.rug.nl/en/publications/pre-editing-and-post-editing>] (visitato il 20/07/2022).

Baker 2017: Baker, W. *Controlled vocabularies in the digital age: are they still relevant?*, Dissertation at University of North Texas, US [https://digital.library.unt.edu/ark:/67531/metadc1011802/m2/1/high_res_d/BAKER-DISSERTATION-2017.pdf] (visitato il 20.07.2022).

Bendana e Melby 2012: Bendana, L. and A. Melby. Everything you ever wanted to know about Translation, – In: Multi-Languages Corporation, Toronto, Canada [https://www.academia.edu/34926861/Everything_you_ever_wanted_to_know_about_translation_melby_bendana_1_] (visitato il 20.07.2022).

Bowker e Ciro 2019: Bowker L. and J.B. Ciro. *Machine Translation and Global Research: Towards Improved Machine Translation Literacy in the Scholarly Community*, 37–54.

Bulté e Tezcan 2019: Bulté, B. e A. Tezcan. Neural Fuzzy Repair: Integrating Fuzzy Matches into Neural Machine Translation. 10.18653/v1/P19-1175, [https://www.researchgate.net/publication/334745562_Neural_Fuzzy_Repair_Integrating_Fuzzy_Matches_into_Neural_Machine_Translation] (visitato il 20/07/2022).

Colina 2020: Colina, S. Quality, translation. – In: Baker M. and G. Saldanha (Eds.), *Routledge Encyclopedia of Translation Studies*, Third edition by Routledge [<https://www.routledge.com/Routledge-Encyclopedia-of-Translation-Studies-3rd-Edition/Baker-Saldanha/p/book/9781138933330>] (visitato il 20.07.2022).

Costa et al. 2016: Hernani, C., A. Zaretskaya, G. Corpas Pastor and M. Seghiri. Nine Terminology Extraction Tools: Are they useful for translators? – In: *Multilingual*, 27(3),

April/May, pp. 14–20. [<https://wlv.openrepository.com/handle/2436/622550>] (visitato il 20.07.2022).

Chunyu e Wong Tak-ming 2015: Chunyu, K. and Wong Tak-ming B. Evaluation in Machine Translation and Computer-Aided Translation. – In: Chan Sin-wai (Ed.) *The Routledge Encyclopedia of Translation Technology*, pp. 213–236, New York, Routledge. [<https://www.routledgehandbooks.com/doi/10.4324/9781315749129.ch12>] (visitato il 20.07.2022).

De Caluve e Van Santen 2003: De Caluve, J. e A. Van Santen. Phonological, morphological and syntactic specifications in monolingual dictionaries. – In: P. van Sterkenburg (Ed.), *A Practical Guide to Lexicography*, Institute for Dutch Lexicology, Leiden, pp. 71–82. [<https://benjamins.com/catalog/tlrp.6.10dec>] (visitato il 20.07.2022).

De Schryver 2003: De Schryver, G.-M. Lexicographers' dreams in the electronic-dictionary age. – In: *International Journal of Lexicography*, 16(2): 143–199. [<https://academic.oup.com/ijl/article-abstract/16/2/143/925134>] (visitato il 20.07.2022).

Drouin 2003: Drouin, P. Term extraction using non-technical corpora as a point of leverage. – In: *Terminology*. 9. 10.1075/term.9.1.06dro, [https://www.researchgate.net/publication/228683045_Term_extraction_using_non-technical_corpora_as_a_point_of_leverage] (visitato il 20.07.2022).

Faber e Araúz 2021: Faber, P. and Araúz P. Designing Terminology Resources for Environmental Translation. – In: Meng, Ji. and S. Laviosa (Eds.), *The Oxford Handbook of Translation and Social Practices*, pp. 587–615, Oxford Academic [https://www.researchgate.net/publication/350213272_Designing_Terminology_Resources_for_Environmental_Translation] (visitato il 20.07.2022).

Garcia 2015: Garcia, I. Computer-aided translation: systems. – In: In S.-W. Chan (Ed.), *Routledge Encyclopedia of Translation Technology*, University of Western Sydney, Australia, pp. 68–87. Recuperato da <http://UWSAU.ebib.com.au/patron/FullRecord.aspx?p=1843560> (visitato il 20.07.2022).

Gouadec 2007: Gouadec, D. Translation as a profession. University of Rennes, Benjamins translation library, 0929-7316 ; v. 73, 2007 [<https://benjamins.com/catalog/btl.73>] (visitato il 20/07/2022).

Hansen 2010: Hansen, G. Translation ‘errors’. - In: Y. Gamber and Van Doorslaer L. (Eds.) *Handbook of Translation Studies. Volume I*. Amsterdam: John Benjamins, pp. 383–388.

[https://books.google.bg/books?hl=bg&lr=&id=BTwzAAAAQBAJ&oi=fnd&pg=PA385&dq=translation+error+classification&ots=bhrc4Gmjv2&sig=3l5Dpv8GDPsvv1SXo9Z8pp3ioxs&redir_esc=y#v=onepage&q=translation%20error%20classification&f=false] (visitato il 20.07.2022).

Hoisl, K., T. Stelzer, S. Biala. Forecasting technological discontinuities in the ICT industry. – In: *Res. Policy*, 44, pp. 522–532, [<https://www.sciencedirect.com/science/article/abs/pii/S0048733314001796>] (visitato il 20.07.2022).

House 2015: House, J. *Translation Quality Assessment. Past and Present*. London, Routledge.

Hönig 1997: Hönig, H.G. Positions, Power and Practice: Functionalist approaches and translation quality assessment. – In: *Current Issues in Language and Society* 4(1): pp. 6–34. [<https://www.tandfonline.com/doi/abs/10.1080/13520529709615477>] (visitato il 20.07.2022).

Imre 2013: Imre, A. Term Bases Reloaded. – In: *Philologia 14*, Studia Universitatis „Petru Maior”. pp. 204–210. [https://www.researchgate.net/publication/287911571_Term_Bases_Reloaded] (visitato il 20.07.2022).

Kageura e Umino 1996: Kageura, K. and B. Umino. Methods of automatic term recognition. – In: *Terminology* vol 3(2), pp. 259–289, John Benjamins Publishing Co. [http://www.iro.umontreal.ca/~felipe/IFT6010-Automne2011/resources/tp3/gabriel_bc/Kageura_Umino_1996.pdf] (visitato il 20.07.2022).

Kilgarriff 2014: Kilgarriff, A., M. Jakubíček, V. Kovář, P. Rychlý, V. Suchomel. Finding terms in corpora for many languages with the Sketch Engine. – In: *EACL* 2014, 53.

[https://www.researchgate.net/publication/271849661_Finding_Terms_in_Corpora_for_Many_Languages_with_the_Sketch_Engine] (visitato il 20.07.2022).

Kirova 2002: Кирова, Л. Еволюция на българската компютърна терминология и компютърен жаргон. – В: *LiterNet*, № 2 (27), [<https://litenet.bg/publish3/lkirova/evolution.htm>] (visitato il 20.07.2022).

Kirova 2004: Кирова, Л. Компютърната лексика – актуални процеси и тенденции. – В: *LiterNet* № 5 (54), [<https://litenet.bg/publish3/lkirova/lex-procesi.htm>] (visitato il 20.07.2022).

Kis 2005: Kis, Á. Terminusalkotás: a terminológiai helyzet és a terminológiai szerep. Mindent fordítunk, és mindenki fordít” Értékek teremtése és közvetítése a nyelvészetben, 105–112.

Koeva 2007: Koeva, S. Multi-word term extraction for Bulgarian. – In: *Proceedings of the Workshop on Balto-Slavonic Natural Language Processing Information Extraction and Enabling Technologies - ACL '07*. [<https://doi.org/10.3115/1567545.1567556>] (visitato il 20.07.2022).

Kolkovska 2010: Колковска, С. Лексикални неологизми в българския език, възникнали от чужди инициални абривиатури. – В: *Български език*, кн. 4, с. 21–30. [<http://ibl.bas.bg/neolex/wp-content/uploads/2013/Issues/Publications/20%20Sia%20Leksika%20ini%20neologizmi%20ot%20inicialni%20abreviaturi.pdf>] (visitato il 20.07.2022).

Lefever 2009: Lefever, E., L. Macken and V. Hoste. Language-independent bilingual terminology extraction from a multilingual parallel corpus. – In: *Proceedings of the 12th Conference of the European Chapter of the ACL*, pp. 496–504, Athens, Greece [<https://aclanthology.org/E09-1057.pdf>] (visitato il 20.07.2022).

Łukasik 2012: Łukasik, M. Terminological dictionary as a comprehensive cognitive and linguistic tool. – In: *Language in Different Contexts: Research papers = Kalba ir kontekstai*, Volume 5 (1), pp. s.98-108. [https://www.researchgate.net/publication/294836193_Terminological_dictionary_as_a_comprehensive_cognitive_and_linguistic_tool] (visitato il 20.07.2022).

Mitchell-Schuitevoerder 2020: Mitchell-Schuitevoerder, R. *A Project-Based Approach to Translation Technology*, London and New York, Publisher: Routledge.

Newman 1988: Newman, P. *A textbook of translation. Shanghai foreign language education press*. First published by Prentice HaH International.

Ророва 2016: Попова, М. За някои видове терминографски параметри. Институт за български език – БАН, София [https://ispan.waw.pl/ireteslaw/bitstream/handle/20.500.12528/705/Lexikografia-2016-70-78.pdf?sequence=1&isAllowed=y] (visitato il 20.07.2022).

Ророва 2019: Попова, М. Някои въпроси на терминологията във връзка с обучението. Journal: Български език и литература, Issue Year: 61/2019, Issue No: 2, pp. 149–155, [https://www.cceol.com/search/article-detail?id=758654] (visitato il 20.07.2022).

Sager 2001: Sager J. Terminology theory, In: – Baker, M. (Ed.), *Routledge Encyclopedia of Translation Studies*. London and New York, Routledge. pp. 250–261 [https://books.google.bg/books?id=ewBfSBo8rRsC&printsec=frontcover&hl=bg&source=gb_s_ge_summary_r&cad=0#v=onepage&q&f=false] (visitato il 20.07.2022).

Šajatović e altri. 2019: Šajatović, A., M. Buljan, J. Šnajder and B. Bašić. Evaluating Automatic Term Extraction Methods on Individual Documents. – In: *Proceedings of the Joint Workshop on Multiword Expressions and WordNet*, pp. 149–154, Florence, Italy. Association for Computational Linguistics. [https://aclanthology.org/W19-5118/] (visitato il 20.07.2022).

Søgaard e al. 2015: Søgaard et al. Inverted indexing for cross-lingual NLP. – In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pp. 1713–1722, Beijing, China, July 26-31, Association for Computational Linguistics [https://www.aclweb.org/anthology/P15-1165.pdf] (visitato il 20.07.2022).

Tayebeh 2008: Tayebeh, M. Translation Memories and the Translator. – In: *International Journal of Translation*. 20, pp. 97-106. [https://www.researchgate.net/publication/271602313_Translation_Memories_and_the_Translator] (visitato il 20.07.2022).

Trumble e Stevenson 2002: Trumble, W. R., & Stevenson, A. (Eds.), *Shorter Oxford English Dictionary*. Fifth Edition, North Carolina, publisher: Oxford University Press.

Varga 2012: Varga, C. Virtualization of Research in Terminology. Cloud-based Terminology Management Tools. – In: *Transactions on Modern Languages*, Vol. 12, No. 1-2, Scientific Bulletin of the Politehnica University of Timișoara.
[https://www.academia.edu/5631484/Virtualization_of_Research_in_Terminology_Cloud_based_Terminology_Management_Tools?email_work_card=title] (visitato il 20.07.2022).

Warburton 2015: Warburton, K. Terminology management. In Sin-wai, C. (Ed), *The Routledge Encyclopedia of Translation Technology*, pp. 644–661, New York, Routledge,
[<https://api.taylorfrancis.com/content/books/mono/download?identifierName=doi&identifierValue=10.4324/9781315749129&type=googlepdf>] (visitato il 20.07.2022).

Wright 2001: Wright, S. E. Data Categories for Terminology Management. – In: Wright S. E. and G. Budin (Eds), *Handbook of Terminology Management: Volume 2: Application-Oriented Terminology Management*, pp. 552–571, John Benjamins Publishing Company.

Zaretskaya e altri. 2015: Zaretskaya, A., G. Corpas Pastor and M. Seghiri. Translators' requirements for translation technologies: a user survey. – In: *New Horizons in Translation and Interpreting Studies*, pp. 247–254, Geneva, Switzerland. Tradulex.
[https://www.researchgate.net/publication/282658144_Translators'_Requirements_for_Translation_Technologies_a_User_Survey] (visitato il 20.07.2022).

Zaretskaya e altri. 2015: Zaretskaya, A., G. Corpas Pastor, G and M. Seghiri. Integration of Machine Translation in CAT Tools: State of the Art, Evaluation and User Attitudes. – In: *SKASE Journal of Translation and Interpretation*, vol. 8, no. 1.
[https://www.researchgate.net/publication/283667119_Integration_of_Machine_Translation_in_CAT_Tools_State_of_the_Art_Evaluation_and_User_Attitudes] (visitato il 20.07.2022).

Zaretskaya e altri. 2016: Zaretskaya, A., G. Corpas Pastor, G and M. Seghiri and C. Hernani. Nine Terminology Extraction Tools: Are they useful for translators?. – In: *Multilingual*, 27(3), April/May, pp. 14–20.
[<https://wlv.openrepository.com/handle/2436/622550>] (visitato il 20.07.2022).

Zehnalova e altri. 2013: Zehnalová, J., O. Molnár and M. Kubánek. A Comprehensive Survey of Multilingual Neural Machine Translation. – In: *Tradition and Trends in Trans-Language Communication*, Olomouc, Univerzita Palackého. [<https://arxiv.org/pdf/2001.01115.pdf>] (visitato il 20.07.2022).