

VORTIT ITALICE. LA LINGUA ITALIANA DELLA TRADUZIONE. LINEE GUIDA PER LA DEFINIZIONE DEL TRADUTTESE DALLE LINGUE CLASSICHE

Laura Bigoni¹, Marco Palma^{2*}

1. INTRODUZIONE

La traduzione in italiano dalle lingue classiche, nonostante polemiche e attacchi, rappresenta ancora oggi un esercizio affrontato da migliaia di studenti³. La resa del testo di partenza solleva molte questioni relative alla natura dell'italiano impiegatovi, poiché esiste uno scarto tra la lingua utilizzata per la produzione scritta dai nativi e quella delle traduzioni⁴. Ronconi (1971: 109) parla di «un linguaggio convenzionale che è diverso da quello che lo scolaro [...] usa nei suoi componimenti o nel parlare di ogni giorno». Per tentare di definire lo scarto in modo rigoroso, questo articolo intende consolidare un approccio bifronte alla questione, fondato sull'apporto combinato di linguistica e statistica⁵. Sulla base di un *corpus* di testi tradotti e di un secondo *corpus* “di controllo” composto da testi redatti in italiano da parlanti nativi, si tenterà di definire le relative differenze qualitative e quantitative, per fornire un quadro generale del fenomeno del cosiddetto “traduttese” in ambito scolastico⁶.

Esistono molti studi sulla “lingua italiana della traduzione”, ovvero quella lingua marcata che risulta dalla trasposizione di un testo in lingua straniera⁷. Molta attenzione è stata rivolta ai calchi e alle storpiature sintattiche che nascono spontaneamente nella traduzione dalle principali lingue moderne europee. Da questa panoramica restano però escluse le lingue classiche, probabilmente per il ridotto volume di testi tradotti da queste lingue all'interno del mercato librario italiano. Esse rappresentano una nicchia linguistica in cui gli studiosi lasciano spesso in sospeso il problema della traduzione per scopi accademici, ma spesso anche didattici⁸. Nell'insegnamento delle lingue antiche in Italia, l'esercizio della traduzione è centrale per la valutazione dell'apprendimento. L'esperienza

¹ Università di Strasburgo (Francia).

² Università di Warwick (Coventry, Regno Unito).

* Il lavoro è stato condotto nel periodo della nostra permanenza presso il Collegio Superiore dell'Alma Mater Studiorum Università di Bologna. Desideriamo ringraziare sentitamente il Professor Federico Condello per il suo affettuoso supporto nel condurre questa ricerca.

³ Per dati approfonditi, si veda Condello (2018).

⁴ In proposito si vedano Condello, Pieri (2011) e Neri, Tosi (2009).

⁵ L'approccio è lodato *e.g.* da Salsnik, 2007: 123s., che riporta anche una breve cronistoria del suo sviluppo.

⁶ Per una breve storia del termine, si veda Condello (2013: 426s, n. 12).

⁷ Si veda *e.g.* Cardinaletti, Garzone (2005).

⁸ Milanese (2012: 69) lamenta la distanza tra le lingue moderne e quelle classiche a livello di apprendimento (e quindi, implicitamente di diffusione) tra gli studenti delle scuole secondarie superiori, imputandola senz'altro al metodo didattico. Sul problema del riconoscimento dei lavori di traduzione a livello accademico, si veda l'articolo di Emma Gee apparso su *Times Higher Education* il 17 marzo 2016, con il titolo *Classics' elitism should be lost in translation* (<https://www.timeshighereducation.com/comment/classics-elitism-should-be-lost-in-translation>), in cui si fa riferimento alla valutazione della ricerca nelle università, dove le traduzioni non vengono considerate nel computo dei lavori di ricerca.

della traduzione si interseca con la didattica in un doppio senso, ovvero quello di verifica dell'apprendimento della lingua di partenza (LP) e di miglioramento della padronanza della lingua di arrivo (LA)⁹. Questa bivalenza della traduzione nella prassi scolastica riguarda le lingue classiche, diversamente – per molti aspetti – da quanto accade per le lingue moderne. Per queste caratteristiche, le lingue classiche hanno effetti diversi sulla «grammatica mentale» del traduttore rispetto alle lingue moderne¹⁰.

2. CONSIDERAZIONI PRELIMINARI

Lo studio rappresenta la prima fase di un progetto più ampio, volto alla raccolta e all'analisi di un *corpus* di testi “originali” effettivamente prodotti da studenti e studentesse delle scuole secondarie superiori. Si presenta qui l'analisi effettuata su un *corpus* pilota, propedeutica all'applicazione diretta in ambito scolastico, con un *corpus* controllabile dal punto di vista della provenienza dei testi. A questo scopo si è creato un questionario da abbinare ai testi raccolti, al fine di contestualizzare meglio la ricerca e trarne quindi informazioni più rilevanti. Parallelamente si deve creare un *corpus* di controllo di testi nativi, con cui effettuare confronti¹¹. L'importanza di costruire *corpora* paragonabili viene espressa da Salsnik (2007: 123) nel senso di creare «un insieme di testi tradotti da una o più lingue accompagnato da un altro gruppo di testi che coprono lo stesso ambito, la stessa varietà di linguaggio e lo stesso periodo storico dei primi»; si comprende che queste caratteristiche non sono verosimili nel caso delle lingue classiche. Si vuole comporre il *corpus* in modo paragonabile almeno per quanto riguarda gli autori, cioè studenti e studentesse della stessa fascia d'età. La creazione del *corpus* prevede la partecipazione tanto dei docenti quanto dei singoli studenti. I primi possono partecipare assegnando una versione a un gruppo ristretto di studenti e compilando il questionario a loro riservato, nel quale si richiedono la regione, la provincia e la tipologia di scuola. Vengono poi richieste altre informazioni per contestualizzare meglio le traduzioni in rapporto al metodo didattico utilizzato e al livello della competenza linguistica degli studenti, in particolare:

- la classe in cui la versione era stata assegnata;
- il metodo di insegnamento (tradizionale, natura, altro da specificare);
- la frequenza media con cui il docente assegna verifiche di traduzione;
- il grado di “letteralità”¹² richiesto o apprezzato nella traduzione (da 1 = molto letterale a 5 = traduzione a senso);
- l'importanza assegnata alla traduzione dal docente ai fini della valutazione finale, rispetto ad altre competenze riguardanti la materia;
- il manuale di grammatica utilizzato per le spiegazioni;
- data della versione.

⁹ Cfr. *e.g.* Santini, 2005: 18.

¹⁰ Negli studi sulle lingue moderne, come Cardinaletti, Garzone (2005), si parla sempre dei traduttori come di “parlanti quasi nativi” della LP; nel caso delle lingue classiche probabilmente una tale definizione risulterebbe impropria, in virtù del fatto che l'apprendimento del latino e del greco avviene attraverso uno solo dei quattro canali tipici delle lingue straniere, ovvero quello ricettivo scritto (cfr. Milanese, 2012: 77).

¹¹ Cfr. Zanettin, 2000.

¹² Si veda in proposito la problematizzazione dei concetti di “letteralità” e “fedeltà” della traduzione in Condello, Pieri (2011).

Il secondo questionario è dedicato agli studenti. In questo caso le domande sono ancora più dettagliate e mirate a comprendere la situazione dello studente al momento della traduzione, oltre ad una serie di informazioni contestuali, ad esempio se la versione sia stata o meno valutata dal docente o quanto lo studente si sia trovato in difficoltà nella comprensione del testo.

È indubbio che ogni lingua porta con sé non solo un codice grammaticale, ma anche un'intera cultura di cui è frutto e veicolo di espressione; non si traduce mai da una semplice lingua, ma da una *linguocultura*¹³. La traduzione dal latino e dal greco comporta uno sforzo notevole in questo senso, poiché la distanza temporale sottende cambiamenti anche radicali di usi, ma soprattutto di concezioni¹⁴. Come ricorda Pasetti (2015: 2), molti errori di traduzione a livello scolastico derivano da «un'immagine poco plausibile del mondo di cui i testi parlano». Per far fronte alla difficoltà di trasmettere simultaneamente lingua e cultura, si incorre spesso in una serie di automatismi nello studio della grammatica e di conseguenza nella traduzione. Un'altra peculiarità della traduzione da queste lingue riguarda la competenza linguistica dei traduttori, non paragonabile ai cosiddetti «parlanti quasi nativi» (Cardinaletti, 2005: 59) di altre lingue da cui si traduce; l'attrito linguistico che si verifica nell'atto traduttorio è alla base dei principali fenomeni di distorsione della grammatica della LA in traduzione ed è riconducibile a quel processo mentale per cui il contatto prolungato con una lingua straniera modifica parzialmente la grammatica mentale del parlante nativo. La chiave di queste variazioni nel caso del latino è probabilmente da scoprirsi nei metodi didattici utilizzati.

Anche uno dei fenomeni più tipici della traduzione merita qui una menzione, ovvero l'*interferenza*, per la quale la grammatica della LA viene forzata sulla base della struttura grammaticale della LP¹⁵. Toury (1995: 271-279) afferma che l'impatto del fenomeno dell'interferenza dipende da fattori come l'esperienza del traduttore, il prestigio della lingua fonte, il livello di tolleranza della lingua d'arrivo¹⁶. Per le lingue classiche si dovrà trattare la questione con più attenzione, per il prestigio della lingua fonte: nel caso in esame, la LA è storicamente derivata dalla LP e nei piani didattici si sottolinea come l'apprendimento della lingua antica debba migliorare le competenze nell'uso dell'italiano¹⁷. All'orecchio del traduttore la LP suona come la vera portatrice dell'italiano alto e non esiste – per parafrasare Milanese (2012: 77) – «sanzione sociale» nell'impiego di termini e costrutti di chiara ascendenza latina; al contrario tali espressioni non solo sono tollerate, ma si mantengono anche quando l'italiano standard le abbia ormai abbandonate¹⁸.

La specificità delle traduzioni scolastiche è legata all'identità dei traduttori e allo scopo della versione. Attraverso i questionari sopra descritti, si cerca di risalire nel modo più preciso possibile alla situazione didattica da cui i testi scaturiscono, nella convinzione che la contestualizzazione possa individuare i tic traduttivi più diffusi. Si tratta di studenti e studentesse con una conoscenza incompleta della lingua, per i quali la traduzione ha per sua natura carattere di esercizio e non di prodotto finito. A proposito dello scopo della

¹³ Cfr. Salmon (2005: 17). In questo contesto si definiscono i tratti ritenuti universali e perciò trasversali rispetto alle linguoculture e quelli invece relativi in ciascuna di esse. Una problematizzazione del concetto di «universale della traduzione» si trova invece in Salsnik (2007).

¹⁴ Cfr. le riflessioni di Pieri (2009: 215-219) a proposito della resa delle metafore legate al contesto della LP e dei *Realien*, tra la volontà di restituire un «effetto equivalente» e quella di «trasmettere quella distanza spazio-temporale».

¹⁵ Per la definizione di interferenza e dei fenomeni di interfaccia, si rimanda a Chomsky (1995); cfr. Cardinaletti, 2005: 68.

¹⁶ Cfr. anche Salsnik, 2007: 102.

¹⁷ In proposito cfr. e.g. Santini, 2005: 18.

¹⁸ Condello (2013: 431) ha sottoposto al vaglio di un *corpus* di controllo di testi italiani alcuni calchi verbali comuni nelle versioni, con risultati impressionanti in termini di frequenze relative.

traduzione, Pasetti (2015: 3) individua le finalità del docente quando assegna l'esercizio di traduzione, dividendole tra competenze passive e attive, e insieme ammettendo che lo scopo didattico prevale sul versante comunicativo: «una traduzione che si inserisce nel processo di apprendimento delle strutture morfosintattiche risponde alla finalità di controllare che tali strutture siano state recepite e può quindi essere più aderente al testo di partenza e (entro limiti ragionevoli) un po' meno funzionale sul piano della comunicazione».

Si noterà che la competenza linguistica, diversamente da quanto avviene per lingue moderne, è per il latino e il greco quasi esclusivamente concentrata sulla conoscenza precisa delle forme grammaticali, a scapito di altre competenze attive¹⁹. In questo contesto la traduzione assume di per sé un valore paradigmatico: «lo studio delle strutture del latino o del greco non si effettua attraverso una descrizione funzionale interna alla lingua che si vorrebbe apprendere, ma attraverso meccanismi traduttivi»²⁰. È dunque in un certo senso la metodologia didattica stessa a imporre l'automatismo tipico di queste traduzioni, indotto ad esempio dalle traduzioni campione contenute nelle grammatiche o dai dizionari. Questo dipende dalla scarsa confidenza dei discenti con la lingua, che li porta a stereotipare il più possibile, allo stesso modo di chi apprende una lingua straniera e inizialmente associa un solo significato ad ogni nuova parola (cfr. Blum-Kulka, 1986: 132). In questo contesto, lo scopo della traduzione è proprio tradurre, nel senso di capire il testo fin nelle sue minuzie grammaticali, e non produrre un testo italiano raffinato; la comprensione avviene per lo più in senso passivo e il testo che ne deriva appare nettamente *source-oriented*²¹. Manca, e per un motivo strutturale, una vera e propria conversione del testo a livello pragmatico. Lo studente che traduce opera dunque in questa situazione “estrema”, in cui tra le due lingue sceglie di privilegiare la LP ogni volta che gli si presenti una difficoltà, e tanto si abitua a questa impostazione da arrivare a non riconoscere più la stranezza del suo italiano. Inoltre, spesso si trova a operare anche in assenza di contesto, cotesto e intertesto²². In virtù dello scopo specifico della traduzione in ambito scolastico, la varietà di testi proposti non è considerevole: si tratta quasi sempre di testi in prosa, con uno stile medio-alto, di forma tipicamente narrativa o argomentativa, spesso selezionati o rivisitati; in generale si prediligono pericopi poco marcate dal punto di vista diafasico, come le narrazioni storiche. In questo modo si vogliono evitare, oltre alle difficoltà di testi marcati nella LP, anche alcuni dei crucci tipici del traduttore letterario, ben esposti da Pieri (2009), tra cui proprio la variazione del livello stilistico. Una certa piattezza nei testi in LA è forse anche imputabile alla monotonia di quanto proposto nella LP. In ogni caso, non si possono confinare tali fenomeni all'ambito scolastico, come del resto rivela già Condello (2013: 432s.) quando afferma che è in atto in tutti gli ambiti di competenza della traduzione da lingue antiche «un sistematico travaso di convenzioni».

Per rivolgere al *corpus* domande che evidenzino lo scarto tra l'italiano della traduzione e lo standard, si prendono in considerazione alcuni aspetti propri delle lingue moderne oltre ad alcuni peculiari del latino²³. Alcuni tratti di interesse nel confronto tra italiano originale e tradotto sono i seguenti:

¹⁹ Salvo alcuni casi, riportati da Pasetti (2015: 1), come la lettura ad alta voce degli esercizi svolti o la produzione autonoma di semplici sintagmi.

²⁰ Milanese (2012: 78), nell'ambito di una critica al metodo didattico «grammaticale-traduttivo», definito «una sciagura» (ivi: 69); riserve sul «metodo pan-grammaticale» si trovano anche in Santini (2005: 21ss.).

²¹ Cfr. Bertazzoli (2006: 78) per la definizione del concetto di *source* e *target oriented*; cfr. Pieri (2009: 215, n. 15).

²² Si veda a tal proposito il recente dibattito sull'impostazione della seconda prova di maturità del Liceo Classico.

²³ Una disamina ampia dei “traduttismi più comuni” si trova già in Condello (2013).

- a) Incidenza dei pronomi personali soggetto e dei dimostrativi; come l'italiano, anche il latino non esplicita il soggetto pronominale, se non per una particolare enfasi, mentre per la terza persona usa i determinativi (*is* o *ipse*) o il dimostrativo *ille*, per ciascuno dei quali esiste una traduzione stereotipata. Il traduttore finisce per usare (e considerare naturali) i pronomi "egli" ed "ella", in luogo dei ben più diffusi "lui" e "lei"²⁴, anaforici che hanno come valore proprio esattamente quello dei pronomi latini, ovvero la ripresa e l'enfasi. Spesso il soggetto è ripetuto nelle frasi subordinate con lo stesso soggetto della principale, dove l'italiano ricorre alla costruzione implicita. In casi come questo il pronome diventa un «segno» più che una traduzione vera e propria²⁵.
- b) Uso smodato del gerundio; si tratta della forma più comune di traduzione di molte costruzioni latine come il *cum* e congiuntivo ed evita al traduttore un'interpretazione netta del testo²⁶.
- c) Uso del participio, specialmente passato, per motivazioni simili.
- d) Presenza di forme trapassate, ormai pressoché scomparse nell'uso corrente della lingua italiana, ma traduzione quasi obbligata del rapporto di anteriorità nel passato tipico della *consecutio temporum* del latino.
- e) Impiego diffuso del congiuntivo²⁷; quando il congiuntivo non è obbligatorio in italiano, il suo impiego è una scelta di registro alto, tratto tipico delle traduzioni dalle lingue classiche. La complessa sintassi del congiuntivo latino, specialmente nei suoi usi indipendenti, lascia spesso nel dubbio gli studenti, che rischiano l'ipercorrettismo.
- f) Tendenza spiccata all'ipotassi: ci si aspetta che la lunghezza dei periodi sia in media maggiore nei testi tradotti rispetto a quelli originali.
- g) Tendenza al calco semantico, specialmente nel caso dei *Wertbegriffe*, tradotti in modo univoco e spesso antiquato, quando non concettualmente errato²⁸.
- h) Tendenza all'arcaismo lessicale, talvolta definito «*jet lag traduttorio*»²⁹.
- i) Particolare uso del *focus*: le frasi tradotte tendono a mantenere l'*ordo verborum* della LP, con esiti marcati in italiano³⁰.
- j) Scarsa varietà del vocabolario; una delle tendenze considerate universali della traduzione è quella alla normalizzazione del testo di partenza, con la connessa semplificazione delle possibilità espressive della LA³¹. Nel caso del latino ci si aspetta che si usino poche parole, ma particolarmente marcate verso l'italiano alto (o ritenuto tale dal traduttore).

²⁴ Sabatini (1985: 159) afferma che l'uso di questi pronomi come soggetto è invalso ormai «in ogni tipo di parlato, anche formale, e nelle scritture che rispecchiano atti comunicativi reali»; cfr. Garzone (2005: 46).

²⁵ Di «segno» parla Condello (2013: 433) a proposito, ad esempio, del classico "o" che precede il vocativo. Si tratterebbe di convenzioni automatiche adottate dai traduttori per far capire al docente che il testo latino è stato compreso nelle sue strutture grammaticali.

²⁶ Come afferma Ronconi (1971: 109-110), «se nella traduzione libera si poteva temere un pretesto a giustificare (o a nascondere) le incomprensioni del testo, la fedeltà pedantesca può diventare scappatoia che dispensa dalla esatta penetrazione».

²⁷ Cfr. l'affermazione di Garzone (2005: 48) secondo la quale nell'italiano della traduzione «non mancano elementi di marcatezza nella gestione del sistema verbale che, come è facilmente prevedibile, si rilevano soprattutto nelle aree di maggiore instabilità di tale sistema»; il congiuntivo ha nell'italiano corrente proprio le caratteristiche di un modo instabile. Per le affermazioni che seguono sul registro, cfr. *ibid.*, p. 49.

²⁸ Alcuni esempi in Pasetti (2015: 3).

²⁹ L'espressione è di Magrelli (2009: 41); cfr. Condello, 2013: 434.

³⁰ Sul *focus* nei testi tradotti, si veda Giusti (2005).

³¹ Cfr. Salsnik, 2007: 106-108.

3. COSTRUZIONE E PRETRATTAMENTO DEL CORPUS

L'analisi del fenomeno della traduzione dalle lingue classiche ha richiesto la compilazione di un nuovo *corpus* pilota monolingua. In questo studio si considera un *corpus* di piccole dimensioni costituito da testi reperiti su Internet in licenza *Creative Commons*.

I testi in italiano tradotti dal latino (testi T, in linea con la notazione proposta in Volansky *et al.*, 2015) corrispondono a traduzioni di capitoli in prosa di letteratura latina. Per evitare eventuali effetti di distorsione dovuti all'autore e agli argomenti dei testi latini sono state escluse le traduzioni di opere troppo estese e sono stati selezionati testi appartenenti a epoche diverse. Il numero complessivo di *token* (punteggiatura compresa) è 259594.

Dal gruppo di testi originali in italiano (testi O) sono state escluse trascrizioni di discorsi concepiti in forma orale e interviste, per garantire una certa omogeneità nel corpus. Anche in questo caso per una esigenza di variabilità delle fonti sono stati selezionati testi di diversi argomenti (con una preponderanza di testi a carattere storico per esaltare la compatibilità con i testi T) e diversi generi. Il numero complessivo di *token* (punteggiatura compresa) è 236873.

Il *corpus* così ottenuto è stato sottoposto ad una prima annotazione strutturale: è stato tokenizzato e segmentato in brani di circa 1500 *token* (compresi i segni di interpunzione), per garantire che le differenze osservate non dipendano dalla dimensione del brano. Tutti i brani terminano in corrispondenza della fine di un periodo. La scelta del numero di *token* per brano è stata effettuata in base ad un compromesso tra il numero indicato in letteratura (2000 *token* in Avner *et al.*, 2016) e l'esigenza, dettata dalla disponibilità dei testi, di ottenere un campione sufficientemente ampio ai fini dell'analisi. I brani sono stati inoltre sottoposti ad una procedura di analisi grammaticale (*Part-of-Speech Tagging*) e lemmatizzazione automatiche.

Tutte le procedure sono state implementate nell'ambiente di lavoro *open-source* R³². Per l'analisi grammaticale e la lemmatizzazione è stato richiamato il software Treetagger³³ con il *tagset* implementato da Marco Baroni³⁴.

4. ANALISI STATISTICA

4.1. Metodi

L'obiettivo di distinguere automaticamente i testi T e O consta di due operazioni:

1. individuare alcuni caratteri linguistici misurabili sui testi che risultino rilevanti nello spiegare le differenze tra le due tipologie di testi;
2. misurare l'entità dell'effetto di ogni carattere linguistico selezionato sulla capacità di predire se un testo non precedentemente analizzato sia T oppure O.

L'informazione portata da ogni testo viene riassunta in un certo numero di caratteri di interesse. Nel problema in esame, la variabile che si cerca di prevedere ("testo tradotto") può assumere solo due valori.

Un approccio classico per lo studio di problemi con variabile dipendente binaria è noto con il nome di classificazione supervisionata: un algoritmo (classificatore) viene

³² R Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.r-project.org/>.

³³ <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>.

³⁴ <http://ssllmit.unibo.it/~baroni/collocazioni/itwac.tagset.txt>.

addestrato a riconoscere la classe di appartenenza di un insieme di unità statistiche per cui questa è conosciuta (*training set*), così da misurare il tasso di errore compiuto. Se l'errore è ridotto, la prestazione del classificatore è buona e quindi questo può essere utilizzato per predire la classe di appartenenza di una nuova unità statistica per cui invece tale classe non è nota a priori (*test set*). Nel caso del traduttese i vettori di valori osservati dei caratteri linguistici in ogni testo del *corpus* per cui è nota anche la classe di appartenenza (T oppure O) fungono da *training set*. Se i caratteri linguistici utilizzati sono rilevanti per la discriminazione dei testi T dai testi O, il classificatore sarà capace di predire se un nuovo testo è tradotto oppure originariamente scritto in LA.

Nei pochi studi quantitativi sul riconoscimento del traduttese, tutti concentrati sulle lingue vive, l'approccio della classificazione supervisionata è stato largamente impiegato (cfr. Baroni, Bernardini, 2006; Ilisei *et al.*, 2010; Volansky *et al.*, 2015). Tali studi hanno il vantaggio di considerare collezioni di testi molto omogenei tra loro per autore e per genere, eliminando da un lato qualsiasi effetto dovuto a tali aspetti, ma dall'altro riducendo le possibilità di generalizzare i risultati. Nel caso di specie, in cui costruire un *corpus* effettivamente comparabile risulta difficile, quantomeno per l'evidente problema di diacronia, si preferisce invece ricorrere alla specificazione di modelli di regressione logistica per variabile dipendente binaria, in modo da controllare tutte le possibili fonti di variabilità che possano condurre all'errata individuazione dei caratteri linguistici più rilevanti.

In particolare, con il modello di regressione logistica si valuta la probabilità (per definizione compresa tra 0 e 1) che un qualsiasi brano sia T in funzione del vettore dei p caratteri linguistici osservati su di esso:

$$P(Y = 1|X) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)} \Leftrightarrow \text{logit}(P(Y = 1|X)) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p.$$

Nella formula, $P(Y = 1)$ indica la probabilità che un brano sia T; ogni carattere osservato sul brano i è indicato con x_i e il rispettivo coefficiente β_i indica la direzione e l'entità del suo effetto sulla probabilità. Se il coefficiente relativo al carattere x_i è negativo vuol dire che all'aumentare di tale valore la probabilità che un testo sia T si riduce; al contrario, se il coefficiente è maggiore di 0, allora all'aumentare di tale numero medio corrisponderà anche un aumento nella probabilità che un testo sia T. Per ogni coefficiente viene data anche una misura della sua variabilità (*standard error*) che serve per verificarne la significatività, ovvero se l'effetto misurato mediante il coefficiente sia statisticamente diverso da 0 (e quindi il carattere linguistico di interesse non mostri alcun effetto sulla probabilità che un testo sia T). La significatività di ogni coefficiente si indica mediante il *p-value*, ovvero la probabilità che, data una certa distribuzione statistica, si osservi nei dati un valore uguale o più estremo di una certa soglia.

Il brano sarà classificato come T se la probabilità stimata è maggiore di 0,5, oppure come O se la probabilità è minore. Questa procedura prevede la possibilità di errore: nella classificazione binaria dei testi, si commette errore quando un brano O viene classificato dal modello come T o viceversa.

Sulla base delle previsioni ottenute e delle classi di appartenenza realmente osservate per i brani si costruisce una matrice di confusione le cui righe contengono le due classi previste e le colonne le classi osservate. Il numero di brani correttamente classificati è uguale alla somma dei valori sulla diagonale principale della matrice; il tasso di corretta classificazione si ottiene dividendo tale numero per il numero complessivo di brani. La massima accuratezza possibile corrisponde al 100% (quando tutti i brani sono classificati

correttamente), mentre il livello minimo è il 50%, corrispondente alla prestazione di un classificatore che assegna ogni brano casualmente ad una classe.

Il modello di regressione logistica può essere riadattato laddove si possano ipotizzare delle strutture di correlazione tra le osservazioni oppure qualora venga meno l'ipotesi di indipendenza tra i brani nel campione. Se questa correlazione fosse rilevante, potrebbe oscurare o amplificare la significatività delle stime ottenute, limitando pertanto la possibilità di generalizzazione delle conclusioni³⁵.

4.2. Descrizione dei caratteri linguistici utilizzati

La scelta dei caratteri linguistici da valutare si è basata sulla letteratura esistente, con particolare attenzione alle ipotesi che tali indici intendono verificare. Essi sono raggruppati in grandi famiglie, ognuna delle quali corrisponde a fenomeni specificamente designati nell'ambito dei *translation studies*. Nella definizione delle ipotesi e delle corrispondenti misure si seguono le direttrici proposte in Volansky *et al.* (2015), eliminando indici che sembrano poco coerenti con il problema della traduzione dalle lingue classiche. In questo modo risulta ancora più interessante il ragionamento intorno all'universalità di tali ipotesi, uno degli aspetti-cardine di diverse teorie linguistiche sul processo della traduzione.

Le ipotesi da tenere in considerazione corrispondono ai concetti di:

- semplificazione degli aspetti complessi nel testo in LP;
- esplicitazione dei riferimenti impliciti in LP;
- normalizzazione e preferenza per la “grammatica convenzionale”;
- interferenza di strutture di LP in LA.

Per caratterizzare queste ipotesi le seguenti misure sono state calcolate su ogni brano:

- a) **Numero di token**: risulta poco variabile in quanto il valore medio è fissato per evitare che emergano evidenze sperimentali troppo legate alla lunghezza dei brani. Viene tenuto in considerazione solo come quantità di controllo per le unità del database.
- b) **Lunghezza media delle parole**: misura il numero medio di caratteri che compongono le parole nel brano. È un indicatore legato all'ipotesi di semplificazione, in quanto si può ritenere che nei testi tradotti si utilizzino parole più semplici e più brevi.
- c) **Lunghezza media dei periodi**: nella letteratura per la traduzione delle lingue moderne si assume che la lunghezza media dei periodi (in numero di parole) nei brani T sia inferiore a quella dei testi originariamente scritti nella lingua di arrivo. Questa misura sarà correlata inversamente al numero di periodi per brano, una volta fissato il numero di token.
- d) **Type-Token Ratio (TTR)**: misura di variabilità lessicale ottenuta dividendo il numero di tipi di *token* per il numero di *token* nel testo. Si ipotizza che il vocabolario utilizzato nei testi T sia ridotto rispetto a quello dei testi O.
- e) **C di Herdan**: simile al TTR, si ottiene dividendo il logaritmo del numero di tipi per il logaritmo del numero di *token*.
- f) **Densità lessicale**: la proporzione di *token* che non sono nomi, aggettivi, avverbi o verbi, potenzialmente sintomo di semplificazione nel brano tradotto.

³⁵ Per approfondimenti si vedano Baayen (2008) e Gries (2013).

- g) **Mean multiple naming**: l'esplicitazione consiste nella spiegazione dei nomi propri che sono noti nella lingua di origine ma non nella lingua in cui si traduce. Diversamente da Volansky *et al.* (2015), si è costruito l'indicatore relativo al numero medio di nomi propri consecutivi.
- h) **Connettivi**: sono spesso utilizzati a fine di esplicitazione.
- i) **PMI medio**: le collocazioni di *token* sono presentate in letteratura come possibili misure di normalizzazione delle traduzioni. Per valutare la rilevanza di questa ipotesi si costruisce l'indice PMI (*pointwise mutual information*) medio sui bigrammi di parole di ogni brano, secondo la formula in Volansky *et al.* (2015).
- j) **Proporzione delle forme passive tra tutti i verbi**: in Volansky *et al.* (2015), si congettura che i traduttori tendano a mantenere il verbo nella diatesi in cui si trova in LP. Il numero di forme passive viene calcolato contando i bigrammi misti composti da qualsiasi forma del verbo essere più un verbo al participio passato. Nonostante talvolta rientrino nel computo anche forme non passive, si ipotizza che esse non differiscano significativamente tra brani T e O.
- k) **Participi presenti**: se ne calcola la presenza nell'ipotesi che il latino ne faccia un uso più largo rispetto all'italiano.
- l) Altre variabili: **ausiliari al participio passato** (*tag* "AUX:ppast"), **avverbi che terminano in -mente** (*tag* "ADV: mente"), **frequenza di pronomi** (*tag* che iniziano per "PRO"), **frequenza di pronomi clitici** (compresi quelli che compaiono come suffisso ai verbi), **frequenza di "forse"**.

4.3. Risultati

Prima di costruire il modello si è proceduto con il calcolo delle correlazioni tra i caratteri linguistici considerati. Le uniche correlazioni rilevanti (ovvero superiori in valore assoluto alla soglia di 0,4) sono attese in quanto dovute alla costruzione delle misure. Ad esempio, fissato il numero di *token* per testo, la lunghezza media dei periodi è negativamente correlata con il numero di periodi. Per questo motivo nei modelli verrà utilizzato solo il numero medio di parole nel periodo. Analogamente, emerge una correlazione positiva tra le due misure di ricchezza lessicale, quindi si considera solo l'indice C di Herdan. In questa fase esplorativa risulta inoltre che i brani T presentano in media un più alto numero di pronomi e clitici rispetto ai brani O; al contrario, la lunghezza media delle parole e il numero di avverbi che terminano in -mente sono maggiori in media per i brani O.

Nel modello di regressione logistica molti dei caratteri impiegati risultano significativamente associati alla variabile indicatrice della traduzione. Tra le variabili che non mostrano significatività si annoverano il numero di connettivi, il PMI medio, la frequenza di participi presenti e il numero di clitici. Deboli invece sono le evidenze per quanto concerne l'importanza dell'indice C di Herdan, l'indice di densità lessicale, il numero di nomi propri consecutivi e i verbi ausiliari al participio passato. Su questo modello è stato applicato un metodo di selezione automatica delle variabili con processo *backward* e sono state inoltre eliminate la misura di densità lessicale e la frequenza di clitici. I risultati sono riportati in Tabella 1.

Tabella 1. Risultati del modello di regressione logistica

	Stima	Std. Error	p-value	
Intercetta	-44,85149	22,50606	0,046276	*
Lung.media.parole	-11,81755	1,96833	1,93e-09	***
Lung.media.frase	0,23834	0,06309	0,000158	***
TTR_herdanC	109,37015	26,72090	4,26e-05	***
NPR_consec	-0,08576	0,03053	0,004960	**
Rapp.ppast.ver	53,58729	12,74992	2,63e-05	***
Pronomi	0,15714	0,03569	1,07e-05	***
ADVmente	-0,50769	0,09356	5,75e-08	***
Freq.forse	-1,40393	0,30727	4,90e-06	***
Significatività: 0 '***' 0,001 '**' 0,01 '*' 0,05				

La probabilità che un qualunque brano sia T è più alta quando la lunghezza media delle parole è più bassa e quando i periodi sono più lunghi; inoltre, all'aumento dell'indice C di Herdan aumenta anche la probabilità di essere T. Anche un rapporto alto di passivi fra tutti i verbi e un'alta frequenza di pronomi sono buoni predittori per identificare brani T; al contrario, l'aumento del numero di nomi propri consecutivi, di avverbi che terminano in -mente e di "forse" fa aumentare la probabilità che un testo sia O.

L'entità in termini assoluti dei coefficienti dell'indice C di Herdan e della proporzione di passivi inducono a ritenere che il modello possa presentare dei problemi di specificazione, probabilmente dovuti alle correlazioni tra brani dello stesso testo o autore. Nonostante ciò, si presta favorevolmente per gli scopi di previsione. La matrice di confusione in Tabella 2 mostra che il tasso di corretta classificazione per questo modello è pari al 92,7%: oltre 9 brani su 10 sono dunque classificati correttamente.

Tabella 2. Matrice di confusione per il modello logistico

<i>Classificati Osservati</i>	<i>T</i>	<i>O</i>	
<i>T</i>	163	15	178
<i>O</i>	9	144	153
	172	159	331

4.4. Discussione

Le variabili inerenti l'ipotesi di semplificazione sono significative ma non sempre in linea con quanto atteso: sebbene i brani T tendano ad avere parole più brevi, i periodi risultano più lunghi rispetto ai brani O e anche la varietà lessicale misurata mediante l'indice C di Herdan, se alta, favorisce la classificazione dei brani come O. Si può ipotizzare dunque che il fenomeno di suddivisione dei periodi lunghi non si presenti nella traduzione delle lingue classiche e dunque che non si possa parlare in questo caso della semplificazione come universale della traduzione.

Anche per quanto concerne normalizzazione ed esplicitazione non si hanno evidenze chiare: da un lato il PMI medio non è significativo, come già rilevato in Volansky *et al.* (2015), ma anche i connettivi non appaiono mai rilevanti ai fini della classificazione.

Altrettanto interessante è la considerazione relativa al carattere del numero medio di nomi propri consecutivi: all'aumentare di questo valore, aumenta la probabilità che il brano sia classificato come O. Anche in questo frangente la deviazione dalla situazione attesa potrebbe essere spiegata banalmente dal fatto che nei testi tradotti possono essere state utilizzate delle sigle che non sono state interpretate come nomi propri dal *POS tagger*.

5. CONCLUSIONI E PROSPETTIVE DI RICERCA

L'analisi qui illustrata presenta alcune limitazioni. Intanto occorre precisare che i testi non sono stati controllati per quanto riguarda la correttezza della traduzione; ci possono essere dunque sia errori di traduzione (ovvero di scorretta interpretazione del testo latino), ma anche errori in italiano, ovvero scelte contrarie alla norma in concomitanza con proprietà strettamente sintattiche³⁶. L'analisi statistica offre una visione complementare di fondamentale importanza: il modello non mira infatti semplicemente a descrivere il fenomeno del traduttese, ma ha anche lo scopo di distinguere un testo dato e mai analizzato prima tra le due categorie di "testo originale" e "testo tradotto". Si tratta di un passo avanti notevole rispetto alla riflessione linguistica teorica, di stampo descrittivo³⁷.

Alcuni degli aspetti che si consideravano interessanti, come ad esempio l'uso del congiuntivo, non possono ancora essere rilevati con questo metodo, anche se i tratti che risultano controllabili danno già un buon orientamento alle considerazioni linguistiche. La natura arcaizzante del lessico non può essere stimata se non con ricerche mirate sull'occorrenza relativa di una lista di termini rappresentativi. La tendenza all'uso di "parole lunghe" non sembra essere rilevante dal punto di vista linguistico: probabilmente l'indice si riferisce a letteratura in lingua inglese, dove forse vale l'equivalenza "parola lunga = parola difficile", che non sembra dimostrato per le lingue qui prese in esame. Quello che semmai si può osservare (ma bisognerebbe confermarlo con studi specifici) riguarda la percezione che un traduttore, specialmente in età scolare, ha delle parole lunghe, ovvero se le ritenga o meno veicolo preferenziale per l'espressione di uno stile alto. Occorre a tal proposito specificare il *target* delle traduzioni che si prendono in esame: spesso la letteratura sull'argomento si riferisce alle traduzioni per l'editoria, dove le tendenze alla semplificazione, alla normalizzazione ed all'esplicitazione sono consuete e preferibili per una migliore fruizione del testo tradotto, mentre in questo studio si considera l'ambito scolastico con le sue specificità. Le traduzioni dalle lingue classiche inoltre tendono a mantenere il più possibile le strutture della LP, ritenute di registro più alto. La particolare tendenza alla subordinazione e al periodo lungo rilevata non rispecchia nel caso specifico delle lingue classiche le teorie linguistiche più diffuse³⁸. Questo indizio rende interessante la prospettiva di esplorare il campo della traduzione da queste lingue considerandola un sistema a sé, connesso ma non identico a quello della traduzione da lingue moderne vive. Gli indirizzi più promettenti riguardano il rapporto tra innovazione e conservazione linguistica, l'influsso delle LP sulla LA e i fenomeni di interferenza.

I risultati positivi ottenuti in questa prima fase su un *corpus* pilota non ideale (per la provenienza eterogenea dei testi, le dimensioni ridotte e l'impossibilità di ottenere metadati) incoraggiano ad applicare le linee guida del progetto in contesti scolastici reali.

³⁶ Cfr. la distinzione di Cardinaletti (2005: 68) tra ciò che è interpretabile a livello di uso della lingua e ciò che invece è contrario al codice grammaticale della lingua stessa, ovvero tra «proprietà strettamente sintattiche» e «proprietà interpretabili dal punto di vista semantico».

³⁷ Cfr Salsnik, 2007: 101-103.

³⁸ In questo caso il *discourse transfer* è molto marcato e la complessità della frase latina viene riportata intatta in quella italiana; cfr. Garzone, 2005: 35.

Tale approccio suscita nuovi spunti di ricerca per lo studio e la didattica delle lingue classiche: si è rilevato infatti che alcune tendenze comuni nella traduzione delle lingue moderne non trovano corrispondenza nel caso di specie. I tratti linguistici qui monitorati sembrano essere un buon punto di partenza per ulteriori indagini sul traduttese dalle lingue classiche. La questione può generare interesse anche negli altri campi dei *translation studies*, fornendo rilevanti evidenze per supportare o confutare le teorie sugli universali della traduzione.

RIFERIMENTI BIBLIOGRAFICI

- Avner E. A., Ordan N., Wintner S. (2016), “Identifying translationese at the word and sub-word level”, in *Digital Scholarship in the Humanities*, 31, 1, pp. 30-54.
- Baayen R. H. (2008), *Analyzing linguistic data: A practical introduction to statistics using R*, Cambridge University Press, Cambridge.
- Baker M. (1993), “Corpus linguistics and translation studies: Implications and applications”, in Francis G., Baker M., Tognini Bonelli E. (a cura di), *Text and technology: in honour of John Sinclair*, John Benjamins, Amsterdam, pp. 233-252.
- Baroni M., Bernardini S. (2006), “A new approach to the study of translationese: Machine-learning the difference between original and translated text”, in *Literary and Linguistic Computing*, 21, 3, pp. 259-274.
- Bertazzoli R. (2006), *La traduzione: teorie e metodi*, Carocci, Roma.
- Blum-Kulka S. (1986), “Shifts of cohesion and coherence in translation”, in House J., Blum-Kulka S. (a cura di), *Interlingual and intercultural communication: Discourse and cognition in translation and second language acquisition studies*, Gunter Narr, Tübingen.
- Canfora L. (2002), “Il fiume si scava il suo letto”, in Dionigi I. (a cura di), *Di fronte ai classici. A colloquio con i Greci e i Latini*, Rizzoli, Milano.
- Cardinaletti A. (2005), “La traduzione: un caso di attrito linguistico”, in Cardinaletti A., Garzone G. (a cura di), 2005, pp. 59-84.
- Cardinaletti A., Garzone G. (a cura di) (2005), *L'italiano delle traduzioni*, FrancoAngeli, Milano.
- Condello F. (2013), “Su qualche caratteristica e qualche effetto del «traduttese» classico”, in Canfora L., Cardinale U. (a cura di), *Disegnare il futuro con intelligenza antica. L'insegnamento del latino e del greco antico in Italia e nel mondo*, il Mulino, Bologna, pp. 423-441.
- Condello F. (2018), *La scuola giusta. In difesa del liceo classico*, Mondadori, Milano.
- Condello F., Pieri B. (a cura di) (2011), *Note di traduttore*, Pàtron, Bologna.
- Garzone G. (2005), “Osservazioni sull'assetto del testo italiano tradotto dall'inglese”, in Cardinaletti A., Garzone G. (a cura di), 2005, pp. 35-57.
- Giusti G. (2005), “La struttura informativa della frase e il *focus* nell'italiano delle traduzioni”, in Cardinaletti A., Garzone G. (a cura di), 2005, pp. 107-122.
- Gries S. T. (2013), *Statistics for linguistics with R: A practical introduction*, De Gruyter Mouton, Berlin.
- Ilisei I., Inkpen D., Pastor G.C., Mitkov R. (2010), “Identification of translationese: A machine learning approach”, in Gelbukh A. (a cura di), *International Conference on Intelligent Text Processing and Computational Linguistics*, Springer, Berlin-Heidelberg, pp. 503-511.
- Magrelli V. (2009), “Finalmente liberi dai Greci e dai Romani”, in Andreotti R. (a cura di), *Resistenza del classico*, Rizzoli, Milano, pp. 38-46.

- Milanesi G. (2012), “Insegnare le lingue antiche, insegnare le lingue moderne. Convergenze e illusioni”, in Oniga R., Cardinale U. (a cura di), *Lingue antiche e moderne dai licei alle università*, il Mulino, Bologna, pp. 67-82.
- Milanesi G. (2014), “Dopo i miti didattici, a partire dal “Documento dei Saggi””, in *Nuova Secondaria Ricerca* 9, pp. 16-24.
- Neri C., Tosi R. (2009), (a cura di), con la coll. di Garulli V., *Hermeneuein. Tradurre il greco*, Pàtron, Bologna.
- Pasetti L. (2015), “Quale traduzione a scuola? Riflessioni su un esercizio difficile”, relazione per il seminario di studio *Insegnare i classici – apprendere competenze*, nell’ambito del progetto “Didattica delle Lingue e delle Letterature classiche”, organizzato daUSR del Veneto e AICC, Padova, 24 settembre 2015 (disponibile su AlmaDL Campus).
- Pieri B. (2009), “La traduzione dalle lingue antiche fra prassi e riflessione: appunti da un esperimento didattico”, in Neri C., Tosi R. (a cura di), con la coll. di Garulli V., 2009, pp. 211-241.
- Ronconi A. (1971), “Traduzione e interpretazione”, in Id., *Interpretazioni grammaticali*, Edizioni dell’Ateneo, Roma, pp. 107-135.
- Sabatini F. (1985), “L’italiano dell’uso medio: una realtà tra le varietà linguistiche italiane”, in Holtus G., Radtke E. (a cura di), *Gesprochenes Italienisch in Geschichte und Gegenwart*, Gunter Narr, Tübingen.
- Salmon L. (2005), “Su traduzione e pseudotraduzione, ovvero su italiano e pseudoitaliano”, in Cardinaletti A., Garzone G. (a cura di), 2005, pp. 17-33.
- Salsnik E. (2007), “Dagli universali traduttivi all’italiano delle traduzioni”, in Montella C., Marchesini G., *I saperi del tradurre. Analogie, affinità, confronti*, FrancoAngeli, Milano, pp. 101-132.
- Santini C. (2005), “L’insegnamento della lingua latina: dalla autarchia alla complementarità ovvero l’*aureolum magnum* di Atalanta”, in Rocca S. (a cura di), *Latina didaxis XX. Atti del Congresso Genova – Bogliasco 8-9 aprile 2005*, Compagnia dei Librai, Genova, pp. 17-38.
- Toury G. (1995), *Descriptive translation studies and beyond*, John Benjamins, Amsterdam.
- Vecchiato S. (2005), “Interferenza e strategie stilistiche nella traduzione dal francese all’italiano”, in Cardinaletti A., Garzone G. (a cura di), 2005, pp. 153-192.
- Volansky V., Ordan N., Wintner S. (2015), “On the features of translationese”, in *Digital Scholarship in the Humanities*, 30, 1, pp. 98-118.
- Zanettin F. (2000), “Parallel Corpora in Translation Studies: Issues in Corpus Design and Analysis”, in Olohan M. (a cura di), *Intercultural Faultlines: research models in translation studies I. Textual and cognitive aspects*, Routledge, Manchester, pp. 105-118.