

Metodi statistici per gli studi sulla traduzione e l'interpretazione

Adriano Ferraresi – Dipartimento di Interpretazione e Traduzione, Università di Bologna, Campus di Forlì

Citation: Ferraresi, Adriano (2020) “Metodi statistici per gli studi sulla traduzione e l'interpretazione”, in Adriano Ferraresi, Roberta Pederzoli, Sofia Cavalcanti, Randy Scansani (eds.) *Metodi e ambiti nella ricerca sulla traduzione, l'interpretazione e l'interculturalità – Research Methods and Themes in Translation, Interpreting and Intercultural Studies*, *MediAzioni* 29: A27-A51, <http://www.mediazioni.sitlec.unibo.it>, ISSN 1974-4382.

1. La rilevanza dei metodi statistici per gli studi sulla traduzione e l'interpretazione

Negli anni recenti si è assistito a una diffusione sempre maggiore di metodi derivati dalla statistica in vari ambiti della ricerca linguistica. Con il passaggio da un approccio prevalentemente di tipo teorico-speculativo a uno basato su dati empirici, questi metodi si sono rivelati essenziali per tre scopi: descrivere i fenomeni osservati, spiegarli sulla base di un'ipotesi, e utilizzarli per prevedere regolarità in dati non ancora esaminati (Gries 2013: 3). Nelle parole di Mellinger e Hanson, “scholars have looked to statistical methods to make sense of the data” (2017: xii).

Per quanto riguarda gli studi sulla traduzione e l'interpretazione, l'adozione di metodi quantitativi in senso lato è indubbiamente legata allo sviluppo della cosiddetta branca descrittiva della disciplina, o *Descriptive Translation Studies* (DTS). Per allinearsi ad altre scienze empiriche, Toury suggeriva che lo scopo dei DTS fosse “describing, explaining and predicting phenomena pertaining to its object level” (1995/2012: xi), ossia fenomeni osservabili nel suo oggetto di indagine: il prodotto della traduzione (i testi tradotti) e/o il suo processo (i

comportamenti e le strategie adottati nell'atto del tradurre). L'indagine traduttologica doveva quindi iniziare a concepire e validare i propri modelli teorici alla luce di "facts of real life" (Toury 1995/2012: xi), prendendo le distanze dalla speculazione fondata su ipotesi preconcepite o al più supportata da esempi isolati e scelti in base a criteri soggettivi (Steiner 2012: 66). Parte fondamentale di questa "svolta empirica" (Snell-Hornby 2006: 115) fu l'adozione di metodi di analisi incentrati sulle regolarità, piuttosto che l'eccezionalità, dei fenomeni osservati, e che al contempo garantivano intersoggettività e replicabilità dei risultati (Toury 1995/2012: xiii). Si pensi ad esempio alle analisi di corpora di testi tradotti e interpretati (cfr. i contributi di Bernardini e di Spinolo e Russo in questo volume), oppure all'adozione di protocolli sperimentali di stampo psicologico nell'osservazione di traduttrici/traduttori e interpreti al lavoro (Gile 2016), che hanno conosciuto sempre maggiore diffusione a partire dagli anni '90 del secolo scorso.

Tra le novità apportate da questi metodi figura sicuramente l'aver introdotto una dimensione quantitativa nella ricerca su traduzione e interpretazione: le frequenze di dati testuali osservate in un corpus o i tempi di reazione a determinati stimoli registrati in un esperimento sono per loro natura fenomeni quantitativi. Eppure, il livello di complessità dei metodi di analisi applicati ai dati generati negli studi di traduzione e interpretazione non ha sempre e uniformemente tenuto il passo con l'avanzamento generale delle tecniche statistiche. De Sutter e Lefer (2020: 2), ad esempio, parlano di "lack of (advanced) statistical testing" e Ji (2012: 53) del bisogno di sviluppare "statistical concepts within the context of translation studies". In questo senso è significativo che nella terza edizione della *Routledge Encyclopedia of Translation Studies* (Baker e Saldanha 2019) non sia presente una voce dedicata ai metodi statistici per la ricerca empirica, e che anzi essi siano menzionati solo marginalmente in alcune voci come direzioni future della disciplina (ad es. Marin-Lacarta 2019: 484).

Esistono oggi numerosi volumi di introduzione ai metodi statistici, pensati sia in generale per le discipline linguistiche (ad es. Gries 2013; Levshina 2015; Winter 2019) che specificamente per gli studi sulla traduzione e l'interpretazione

(Mellinger e Hanson 2017). Simili volumi, per lo più in lingua inglese, rappresentano punti di riferimento imprescindibili per chi voglia utilizzare metodi statistici in prima persona. Il presente contributo non può evidentemente ambire a tali livelli di completezza, ma mira piuttosto a fornire una panoramica, o, per così dire, un punto di accesso, a questi metodi di ricerca che raramente vengono inclusi nei curricula dei programmi di traduzione e interpretazione (Mellinger e Hanson 2017: xii). Non ci si soffermerà su aspetti matematici se non nei casi in cui siano necessari per effettuare le analisi presentate, e non si potrà fornire più di qualche suggerimento su come implementare i metodi illustrati tramite software per l'analisi statistica.

Traendo esempio da studi pubblicati, si discuteranno invece le fasi di uno studio quantitativo, dalla formulazione dell'ipotesi di ricerca alla presentazione dei dati raccolti tramite statistiche descrittive e grafici, fino alla fase del *testing* statistico vero e proprio. Così facendo, ci si propone di chiarire la logica e i principi, se non i dettagli matematici e le procedure informatiche, che sottendono concetti quali le variabili dipendenti e indipendenti, le misure di tendenza centrale e di dispersione e il concetto di significatività statistica. Una tale introduzione, ci si augura, permetterà se non altro di comprendere più a fondo e valutare gli studi pubblicati che utilizzano metodi derivati dalla statistica. In quest'ottica, si prenderanno come esempi test statistici e design sperimentali tra i più semplici (chiamati monofattoriali), per due ragioni: perché sono tra i più utilizzati nei DTS (De Sutter e Lefer 2020: 5), e perché capire il loro funzionamento aiuta a progredire verso metodi di analisi più complessi. L'ultima sezione proporrà una riflessione sui limiti dei metodi discussi, attualmente al centro di un dibattito metodologico in svariate discipline empiriche, e suggerirà alcune delle direzioni più promettenti per superarli.

2. Le fasi di uno studio quantitativo

2.1. Ipotesi di ricerca e definizione delle variabili

Come negli studi qualitativi, anche negli studi quantitativi l'intuizione iniziale può derivare dall'interesse personale o dall'osservazione diretta di un fenomeno della realtà, quei "facts of real life" di cui parlava Toury (Sezione 1). Tramite una ricerca nella letteratura di riferimento, questa intuizione deve essere contestualizzata all'interno di una teoria: è durante questa fase che si verificherà, ad esempio, se esistono studi simili a quello che si vuole intraprendere, e se l'intuizione è interessante rispetto a un'eventuale lacuna nella letteratura ("research gap") o a un problema sollevato in studi precedenti. Solo grazie a questo passaggio sarà possibile interpretare i risultati ottenuti e dimostrarne la novità e l'interesse per la disciplina o le discipline coinvolte¹.

Definito il problema o la domanda al centro dello studio, si passerà poi a formularli sotto forma di un'ipotesi di ricerca. Nella definizione di Field (2018: 5) un'ipotesi di ricerca è "an informed, theory-driven attempt to explain [...] a fairly narrow phenomenon or set of observations". L'ipotesi di ricerca deriva insomma da una teoria, ma si focalizza sui fenomeni specifici che si intende testare in uno studio. Questi fenomeni vengono concettualizzati come *variabili* dello studio, e l'ipotesi di ricerca è formulata, di norma, in termini di due tipi di variabili: variabili indipendenti e variabili dipendenti. Le variabili indipendenti sono manipolate o controllate da chi svolge la ricerca, mentre le variabili dipendenti sono quelle che vengono misurate e che (si suppone) varieranno al variare delle variabili indipendenti. Per questa ragione le prime vengono anche chiamate "variabili esplicative" o "predittori" e le seconde "variabili risposta" o "variabili spiegate".

Per fare un esempio che verrà richiamato nelle Sezioni 2.3 e 2.4, una parte dello studio di Kajzer-Wietrzny (2012) prende le mosse dalla teoria di Blum-Kulka (1986), ripresa e rielaborata da Baker (1993), secondo cui i testi tradotti tendono a essere più espliciti rispetto ai testi non tradotti (ad es. preferendo strutture lessico-grammaticali esplicite invece che implicite) a causa di meccanismi sociali

¹ Per un approfondimento sul rapporto tra intuizione iniziale e formulazione del problema di ricerca si veda ad es. Field *et al.* (2012: 4-7).

e/o cognitivi che influenzano le scelte di traduttori e traduttrici². A partire da questa teoria, Kajzer-Wietrzny sviluppa l'idea che tendenze simili a quelle osservate nel prodotto della traduzione scritta possano essere individuate in altre forme di mediazione linguistica, e nello specifico nell'interpretazione simultanea. Formula dunque l'ipotesi che il livello di esplicitezza lessico-grammaticale dei discorsi orali vari in base alla modalità di produzione dei discorsi stessi, ossia se sono prodotti da interpreti durante un compito di interpretazione simultanea o se sono invece discorsi "originali" prodotti da parlanti in condizioni monolingue³. In questo scenario, la variabile indipendente è la modalità di produzione dei discorsi, che Kajzer-Wietrzny *controlla* scegliendo di raccogliere trascrizioni di discorsi interpretati e originali; la variabile dipendente, che la ricercatrice *misura* nei due gruppi di discorsi, è il livello di esplicitezza.

Un'ipotesi di ricerca esprime, dunque, una relazione tra una o più variabili dipendenti e una o più variabili indipendenti, postulando che le prime varino come effetto delle seconde. Una volta definita, l'ipotesi non può però ancora essere testata. Per farlo, occorre infatti *operazionalizzare* le variabili, ossia definire precisamente quali dati saranno raccolti e come questi saranno misurati. Nel caso succitato di Kajzer-Wietrzny (2012), la variabile indipendente (modalità di produzione di un discorso orale) è operazionalizzata raccogliendo discorsi al Parlamento europeo prodotti da: a) interpreti simultaneisti che lavorano verso l'inglese da diverse lingue (discorsi interpretati) e b) membri del Parlamento per cui l'inglese è lingua nativa che pronunciano discorsi originalmente in tale lingua (discorsi originali). Per quanto riguarda la variabile dipendente, la ricercatrice riprende invece l'operazionalizzazione di Olohan e Baker (2000), considerando come segnale di esplicitezza lessico-grammaticale la presenza della congiunzione opzionale "that" in combinazione con "reporting verbs" (ad es. "he said **that** he would" vs. "he said \emptyset he would"). L'ipotesi postula dunque che si

² Si rimanda al lavoro di De Sutter e Lefer (2020) per una critica all'idea che questa costituisca una vera e propria teoria.

³ Lo studio originale comprende numerose altre ipotesi di ricerca. In questa sede ci si limita a discuterne una per ragioni di chiarezza espositiva.

osserverà un numero diverso di *that* in queste costruzioni, corrispondenti a un diverso livello di esplicitzza, a seconda del gruppo di discorsi considerato.

L'operazionalizzazione rappresenta un passaggio cruciale di uno studio quantitativo, in quanto determina, oltre ai dati da raccogliere, il tipo di test statistico da effettuare. In particolare, si possono distinguere tre categorie principali di variabili in base alla loro definizione o misurazione:

- **categoriche**: i fenomeni sono assegnati a più categorie distinte; nel caso in cui le categorie siano due si parla anche di variabili binarie. Ad esempio, la distinzione tra discorsi interpretati e originali rappresenta una variabile binaria, mentre la distinzione tra testi interpretati, testi tradotti e testi originali rappresenta una variabile categorica (ad es. Shlesinger e Ordan 2012);
- **ordinali**: i fenomeni sono assegnati a categorie tra cui sussiste una relazione ordinata o gerarchica. La differenziazione degli apprendenti di una lingua in principianti, intermedi e avanzati (ad es. Saito *et al.* 2015) o le scale di risposta a questionari in cui si chiede di assegnare valori come “raramente”, “a volte” e “spesso” (ad es. Picton *et al.* 2015) costituiscono esempi di questo tipo di variabili;
- **numeriche**⁴: al contrario delle precedenti, le variabili numeriche si esprimono su una scala di valori numerici (non categorici), per cui le differenze tra valori sono direttamente interpretabili. Esempi di variabili numeriche sono la velocità di eloquio di un discorso misurata in numero di parole pronunciate al minuto (ad es. Collard e Defrancq 2019), oppure i tempi di fissazione oculare in millisecondi registrati durante esperimenti di *eye-tracking* (ad es. Vanroy *et al.* 2019).

Si noti come una stessa variabile possa essere operazionalizzata in modi diversi. La competenza di traduttori/traduttrici, ad esempio, può essere una variabile binaria (esperti vs. non esperti), ordinale (inesperti, in formazione, professionisti)

⁴ Svatiati autori introducono un'ulteriore differenziazione tra variabili numeriche proporzionali e a intervallo (ad es. Levshina 2015: 17). Per semplicità, e poiché spesso le variabili a intervallo vengono trattate come variabili proporzionali (ad es. Gries 2013: 17), questa ulteriore differenziazione viene tralasciata qui.

oppure numerica (numero di anni spesi nello studio e/o nella pratica professionale della traduzione). La decisione sul modo più appropriato di operazionalizzare le variabili dipende quindi chi svolge lo studio, che si farà guidare da criteri quali la coerenza rispetto alla teoria di riferimento e la disponibilità di dati.

2.2. Presentare i dati tramite statistiche descrittive

Alla definizione dell'ipotesi e la conseguente operazionalizzazione delle variabili, segue la raccolta dei dati (ad es. la creazione di un corpus e l'identificazione delle frequenze dei fenomeni testuali di interesse, la progettazione e lo svolgimento di un esperimento con partecipanti, ecc.). Per ragioni di spazio e perché esistono innumerevoli tipi di dati e altrettanti modi di raccogliarli, questo contributo non si addenterà nelle pratiche legate a questa fase, ma rimanda a Mellinger e Hanson (2017: capitoli 2-3), che discutono approfonditamente i parametri e le procedure da adottare per assicurarsi che il campione (di testi, soggetti, ecc.) sia bilanciato e rappresentativo, nonché gli aspetti etici di cui tenere conto quando si raccolgono dati di varia natura.

Il passaggio di cui ci si occupa in questa sezione consiste nella presentazione del campione di dati tramite statistiche descrittive. Queste possono essere concepite come metodi che permettono di riassumere alcune proprietà essenziali del campione. Forse la più conosciuta tra le statistiche descrittive è la media aritmetica. La media è una cosiddetta misura di tendenza centrale, ossia un valore che può essere considerato come "most typical of a distribution" (Levshina 2015: 43). Viene calcolata sommando tutti i valori osservati di una variabile numerica, e dividendo il risultato per il numero di valori nel campione⁵. Ad esempio, se consideriamo come variabile i voti ottenuti in un esame, la media

⁵ Tutti i software di analisi statistica includono funzioni apposite per calcolare statistiche descrittive. In questa sezione si riporteranno le funzioni del pacchetto Calc di LibreOffice, scelto in quanto esempio di software libero che non richiede competenze informatiche avanzate. La media viene calcolata con la funzione "=MEDIA()".

sarà calcolata sommando i voti ottenuti in tutte le prove e dividendo il risultato per il numero di prove.

Una buona pratica che viene spesso trascurata, anche in studi pubblicati, è quella di riportare, per ogni misura di tendenza centrale, il valore della corrispondente misura di dispersione, ossia una misura che indichi “how good the measure of central tendency is at summarizing the data” (Gries 2013: 119). Nel caso della media, la misura di dispersione da riportare è la deviazione standard⁶. Per comprendere perché sia fondamentale riportare una misura di dispersione si considerino i due grafici in Figura 1, che rappresentano i voti fittizi ottenuti in un esame.

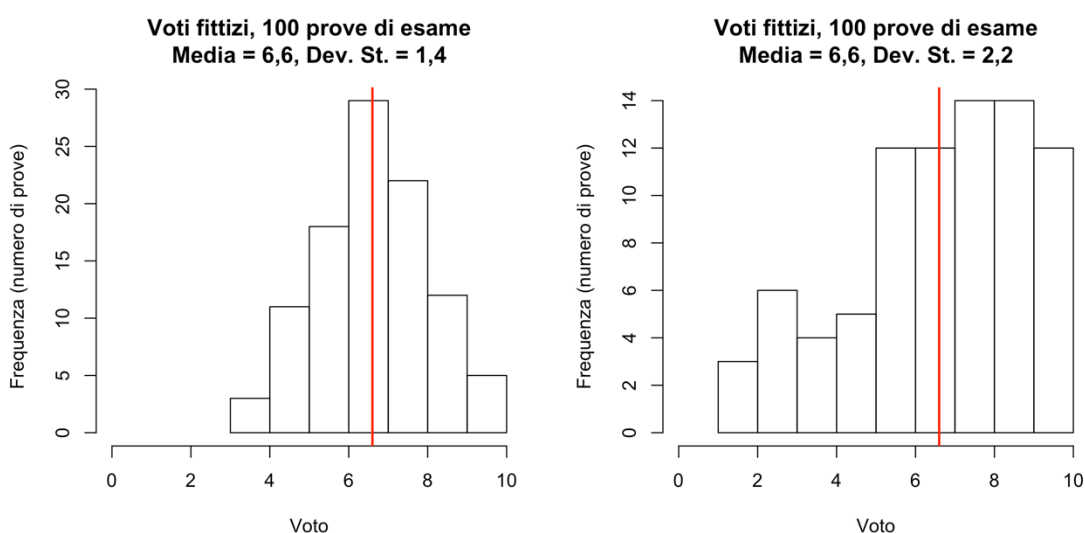


Fig. 1 Distribuzioni con una stessa media ma diversa deviazione standard (dati fittizi); la linea rossa rappresenta la media.

I grafici rappresentano due distribuzioni di dati basate su uno stesso numero di osservazioni (100) e con una stessa media (6,6). Nel grafico di sinistra, tuttavia, i valori sono distribuiti uniformemente vicino alla media stessa (da un minimo di 3 a un massimo di 10, con molti valori compresi tra 5 e 7); nel grafico di destra, invece, i valori sono più dispersi, ossia lontani dalla media (da un minimo di 1 a un massimo di 10, con molti valori compresi tra 5 e 10). Questa differenza è

⁶ Funzione in Calc di LibreOffice: “=DEV.ST()”.

espressa dal valore di deviazione standard, più alto nel secondo caso che nel primo (2,2 vs. 1,4): è in questo senso che la deviazione standard indica quanto la media rappresenti appropriatamente la distribuzione, poiché fornisce informazioni sulla variabilità osservata.

Non è sempre consigliato ricorrere a media e deviazione standard come statistiche descrittive. Infatti, quando la distribuzione dei valori presenta una forma non simmetrica (v. ad es. la Figura 2), oppure si osservano valori estremi (ossia lontani dal centro della distribuzione), la misura di tendenza centrale più appropriata è la mediana⁷. Questa è calcolata ordinando tutti i valori della distribuzione e prendendo il valore osservato a metà della distribuzione stessa⁸: se la distribuzione ha un numero di valori dispari, la mediana coinciderà con il valore centrale; se il numero di valori è pari, la mediana corrisponderà alla media dei due valori centrali. Un esempio di distribuzione per cui è opportuno utilizzare la mediana è rappresentato nella Figura 2, dove si nota come un numero consistente di valori è compreso tra l'8 e il 10, mentre gli altri valori risultano dispersi in modo decrescente tra il 7 e il 2: in questi casi si parla di una distribuzione asimmetrica a sinistra (in inglese "left-skewed"). Trattandosi di 100 valori, la mediana corrisponde alla media tra il 50° valore e il 51° (entrambi uguali a 8; v. anche Figura 3).

Al contrario della media, la mediana è poco sensibile alla presenza di valori estremi: se ad esempio si aggiungessero dieci valori pari a 0 nella distribuzione della Figura 2, la media scenderebbe sensibilmente (da 7,6 a 6,9), mentre la mediana passerebbe da 8 a 7,9. Per questa caratteristica, Mellinger e Hanson (2017: 44) suggeriscono di utilizzare la mediana in tutti i casi in cui ci si aspetta di osservare valori estremi, come le misure legate al tempo registrate in

⁷ In generale, è preferibile riportare la mediana in presenza di distribuzioni cosiddette "non normali" o asimmetriche. La rappresentazione dei dati tramite un grafico permette spesso di intuire se la distribuzione sia normale o meno, e/o se siano presenti valori estremi. Tuttavia, esistono metodi formalizzati per verificarlo: su questo si rimanda a Lee Ling (2015).

⁸ Funzione in Calc di LibreOffice: "=MEDIANA()".

condizioni sperimentali (ad es. il tempo impiegato a svolgere un compito) e dati testuali quali la lunghezza delle frasi.

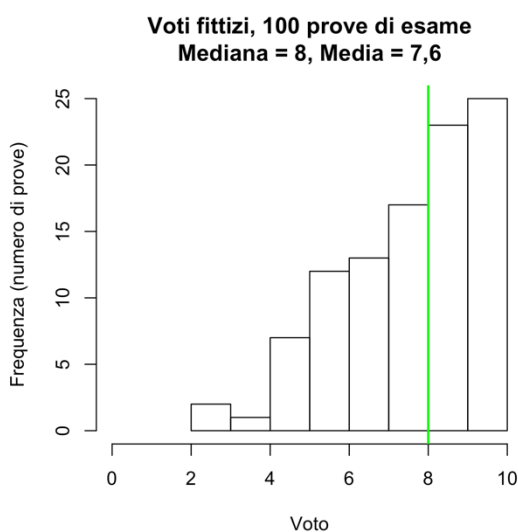


Fig. 2 Distribuzione non simmetrica; la linea verde rappresenta la mediana.

La misura di dispersione associata alla mediana è lo scarto interquartile (IQ). Se la mediana è il punto che si trova a metà della distribuzione, ossia in corrispondenza del 50% dei valori, lo scarto interquartile può essere concepito come la differenza tra il punto che si trova in corrispondenza del 75% dei valori (3° quartile) e il 25% dei valori (1° quartile)⁹. La stessa distribuzione della Fig. 2 è rappresentata nella Figura 3 in modo da illustrare meglio il concetto. Ogni cerchio rappresenta un valore della distribuzione in ordine crescente, ossia i voti di ogni singola prova dal più basso al più alto; le righe orizzontali arancione e viola sono posizionate rispettivamente all'altezza del 25% e del 75% dei dati (ossia circa in corrispondenza dei valori della 25esima e 75esima prova). La differenza tra il valore del 3° quartile (9) e quello del 2° quartile (6.5) rappresenta lo scarto interquartile, ed è il valore da riportare accanto alla mediana.

⁹ In LibreOffice Calc non è presente una funzione che calcoli lo scarto interquartile automaticamente: si rimanda a Lee Ling (2015), che descrive la procedura da adottare.

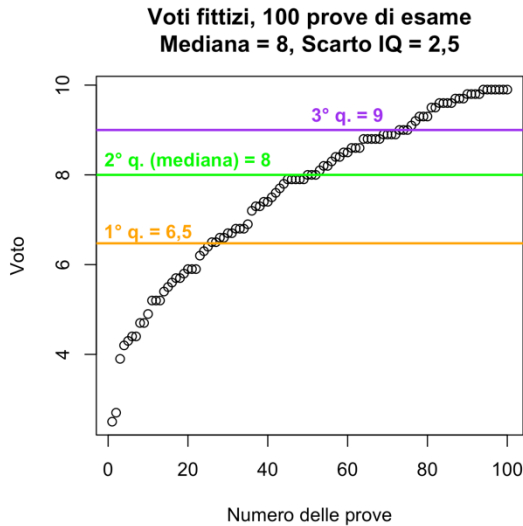


Fig. 3 Distribuzione ordinata dei singoli valori rappresentati anche in Fig. 2, con indicazione del 1°, 2° e 3° quartile.

2.3. Descrivere i dati tramite grafici

Se le statistiche descrittive sono necessarie per fornire informazioni puntuali sul campione di dati raccolto, l'utilizzo di grafici costituisce un metodo spesso più diretto ed efficace per visualizzare tendenze e relazioni tra le variabili considerate (Mellinger e Hanson 2017: 51-52). Tra le molte forme di rappresentazione grafica di dati quantitativi (v. ad es. Gries 2013: cap. 3) questa sezione si concentrerà su due tipi di grafici: primo, il grafico a colonne, particolarmente indicato per rappresentare differenze tra gruppi che coinvolgono variabili categoriche, e tra i più semplici da realizzare; secondo, il cosiddetto diagramma a scatola (o *boxplot*, in inglese), che costituisce una delle forme di rappresentazione più informative per distribuzioni di variabili numeriche.

Il grafico a colonne, anche conosciuto come grafico a barre, permette di visualizzare frequenze di variabili categoriche. Un esempio di tali dati sono quelli raccolti nello studio di Kajzer-Wietrzny (2012: 99), menzionato nella Sezione 2.1. Si ricorderà che lo studio ipotizzava che discorsi originali e interpretati in inglese differissero per quanto riguarda la presenza o assenza di “that” come connettore che segue “reporting verbs”. La Tabella 1 riprende le frequenze osservate nello

studio originale per i due tipi di costruzioni nelle due categorie di discorsi (per i discorsi interpretati si riportano solo i dati relativi alle interpretazioni che hanno come lingua di partenza il francese)¹⁰.

	Discorsi originali EN	Discorsi interpretati FR>EN
<i>that</i>	146	122
∅ connettore	73	35
TOTALE	219	157

Tab. 1 Frequenze di “that” e “∅ connettore” dallo studio di Kajzer-Wietrzny (2012: 99).

Osservando la tabella si nota come il totale di casi osservati nelle due categorie di discorsi sia diverso (219 contro 157): questo fa sì che le frequenze (ad esempio relative ai casi di “that”: 146 contro 122) non siano direttamente paragonabili. In questi casi, prima di rappresentare graficamente i dati è necessario compiere un’operazione di normalizzazione, che consiste nell’esprimere le frequenze osservate come proporzioni rispetto al totale dei casi nella categoria di riferimento delle frequenze stesse. Il calcolo viene dunque effettuato come:

$$\frac{\text{frequenza osservata di } X * \text{costante (ad es. 100/1.000/10.000)}}{\text{totale dei casi nella categoria di } X}$$

La frequenza normalizzata per “that” nei discorsi originali è quindi $(146 * 100) / 219 = 66,66$; lo stesso calcolo viene poi effettuato per le altre celle della Tabella 1. Da notare come la costante possa cambiare in base alla scala dei valori: qui la normalizzazione è effettuata su base 100, ma quando le frequenze sono nella scala delle migliaia, la costante può essere 1.000 e così via¹¹. A questo punto i

¹⁰ Lo studio originale riporta anche dati per discorsi interpretati dallo spagnolo, dal tedesco e dal neerlandese.

¹¹ Negli studi che paragonano frequenze di fenomeni in corpora di dimensioni diverse, il calcolo delle frequenze normalizzate utilizza solitamente le dimensioni dei corpora (numero di parole) come denominatore della frazione (totale dei casi). In questo scenario è buona norma scegliere come costante la scala di grandezza in cui sono comprese le dimensioni del corpus (ad es. 100.000 per corpora tra 100.000 e 1.000.000 di parole).

dati possono essere rappresentati in forma di grafico a colonne: la Figura 4 presenta due possibili forme del grafico utilizzabili.

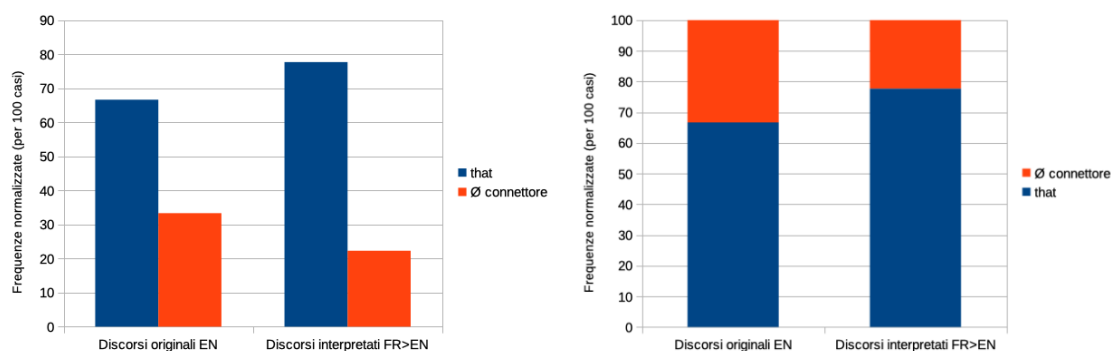


Fig. 4 Frequenze normalizzate di “that” e “Ø connettore” dallo studio di Kajzer-Wietrzny (2012: 99); a sinistra un grafico a colonne raggruppate, a destra un grafico a colonne in pila.

La tendenza che poteva essere intuita sulla base della Tabella 1 emerge molto più chiaramente da queste rappresentazioni grafiche: i discorsi interpretati utilizzano in proporzione più “that” espliciti rispetto ai discorsi originali. In termini di differenze tra i due grafici, il grafico con colonne raggruppate (a sinistra) mette più in evidenza le frequenze delle singole categorie di connettori nei due diversi gruppi, mentre quello con colonne in pila (a destra) rende più chiaramente l’idea della proporzione dei casi all’interno dello stesso gruppo. Come di consueto spetta al(la) ricercatore/ricercatrice scegliere quale sia la rappresentazione grafica più adeguata ad accompagnare la descrizione dei risultati in forma scritta.

Anche i *boxplot* possono essere utilizzati per rappresentare differenze tra categorie, ma in questo caso a essere rappresentate sono distribuzioni di variabili numeriche, invece di singole frequenze, associate a ogni categoria. In particolare, un *boxplot* permette di presentare graficamente molte delle statistiche descrittive illustrate nella Sezione 2.1, ossia mediana, primo e terzo quartile, nonché minimo e massimo di una distribuzione.

Un caso di studio esemplificativo in cui sono stati utilizzati *boxplot* è quello di Ferraresi *et al.* (2018). Come Kajzer-Wietrzny (2012), gli autori partono da un’ipotesi derivata dal lavoro di Baker (2003) secondo cui i testi tradotti tendono a essere più semplici dal punto di vista lessicale rispetto ai testi non tradotti. Il

concetto di semplicità lessicale è operazionalizzato tramite la misura della densità lessicale, ossia il rapporto tra parole lessicali e funzionali: più è bassa tale misura, più i testi sono considerati semplici (v. anche Laviosa 1998)¹². Seguendo una procedura sempre più diffusa negli studi basati su corpora (Lijffijt *et al.* 2016), gli autori calcolano il valore della misura di interesse all'interno di ognuno dei testi che compongono i corpora¹³. Rispetto a una procedura più semplice, che consisterebbe nel calcolare un singolo valore globale per ogni corpus (ad es. la proporzione tra il numero totale di parole lessicali e funzionali in ognuno), questo permette di basare il confronto su informazioni che rispecchiano le tendenze centrali e la variabilità interna dei due corpora: ottenendo un valore per ogni testo è infatti possibile tenere conto di valori medi o mediani e della relativa dispersione, il che non è ovviamente possibile quando il fenomeno di interesse è quantificato tramite un singolo valore “rappresentativo” dei corpora nella loro interezza. La Tabella 2 riporta statistiche descrittive sui risultati ottenuti, e la Figura 5 presenta il relativo *boxplot*¹⁴.

	Testi originali EN	Testi Tradotti IT>EN
Mediana	48,3	49,7
Scarto IQ	5,1	4,9

Tab. 2 Statistiche descrittive relative ai valori di densità lessicale (da Ferraresi *et al.* 2018: 731).

¹² Anche in questo caso viene ripresa solo una parte dei dati dello studio originale, che include altre misure di semplicità lessicale, oltre ad altri paragoni tra discorsi orali e interpretati, e tra interpretazioni e traduzioni in inglese da lingue diverse (italiano e francese).

¹³ In questo caso si tratta della densità lessicale, ma la stessa procedura può essere adottata per altre misure derivabili da un corpus, inclusa la frequenza di singole parole o categorie di parole, di fenomeni lessicali o sintattici, e così via.

¹⁴ La versione corrente di LibreOffice Calc (7.0.4) non comprende una funzione per creare questo tipo di grafico, che invece può essere creato ad es. con Microsoft Excel (V. 16.4).

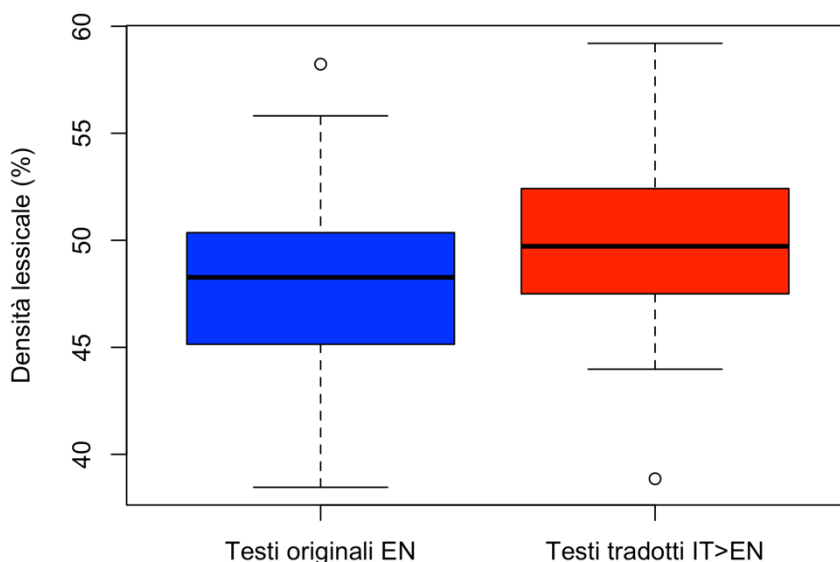


Fig. 5 Boxplot che rappresenta il paragone fra valori di densità lessicale (da Ferraresi *et al.* 2018: 731).

Per comprendere il grafico occorre conoscere alcune delle convenzioni in uso. La linea orizzontale nera che taglia in due ogni rettangolo colorato, o “scatola”, rappresenta la mediana della distribuzione, mentre i suoi limiti inferiore e superiore coincidono rispettivamente con il primo e terzo quartile. La scatola racchiude così il 50% centrale dei dati (dal 25% al 75% dei valori ordinati in scala crescente) e da essa si diramano due righe tratteggiate che rappresentano i valori rimanenti: le due linee orizzontali con cui terminano queste righe verticali corrispondono solitamente al valore minimo e massimo osservato. In casi come quello della Figura 5 sono però presenti valori estremi (“outlier” in inglese), rappresentati come cerchi. Per convenzione, vengono rappresentati come estremi i valori che si discostano per più del valore dello scarto IQ moltiplicato per 1,5 rispetto al limite superiore o inferiore del rettangolo.

In termini di interpretazione del grafico, possiamo osservare come i valori di densità lessicale associati ai testi tradotti sono più elevati rispetto a quelli dei testi originali: non solo la mediana è posizionata più in alto, ma quasi tutta la metà superiore della scatola di destra sovrasta la scatola di sinistra. Questo è di solito interpretabile come segno che le due distribuzioni sono considerevolmente diverse l’una dall’altra. Al contrario di quanto ipotizzato, i testi tradotti sembrano quindi lessicalmente più complessi, invece che più semplici, di quelli non tradotti.

Il passaggio successivo, l'ultimo che verrà illustrato in questo contributo, consiste nell'eseguire un test che permetta di valutare la significatività statistica della differenza osservata. Cosa significhi esattamente significatività statistica e quali siano le procedure per verificarla sono gli argomenti trattati nella prossima sezione.

2.4. Testare la significatività

Come accennato nella Sezione 2.2, i dati raccolti in qualsiasi studio sono necessariamente campioni tratti da una popolazione più grande, e statistiche descrittive e grafici servono a descrivere tali campioni. Tuttavia, la descrizione di un campione non è sufficiente per rispondere alla domanda centrale per uno studio quantitativo, ossia se è probabile che le tendenze emerse (ad esempio le differenze tra due gruppi) si ripetano anche in altri casi non osservati nello studio: in altre parole, se i risultati siano *generalizzabili* alla popolazione che il campione mira a rappresentare. Per rispondere a tale domanda occorre fare ricorso a un'altra famiglia di metodi: le statistiche inferenziali.

La scelta del test statistico più appropriato per vari tipi di dati dipende da numerosi fattori che non sarebbe possibile esaurire nello spazio di questo contributo. Per questo si è deciso di esemplificare i principi che governano la fase del *testing* statistico tramite uno studio ripreso più volte nelle sezioni precedenti, quello di Kajzer-Wietrzny (2012). Il test che verrà presentato, chiamato test del chi quadrato (χ^2), non potrebbe essere utilizzato ad esempio con le variabili coinvolte nello studio di Ferraresi *et al.* (2018; v. Sezione 2.3), dove invece è più appropriato il test U di Mann–Whitney, o test di Wilcoxon (v. Gries 2013: 209-214; Levshina 2015: 108-110). Si consideri tuttavia che molti tipi di test, tra cui i due appena menzionati, sono accomunati da uno stesso approccio, conosciuto come “Null Hypothesis Statistical Testing” (NHST). Illustrare i principi dell’NHST tramite l’esempio del test del χ^2 pone dunque le basi per comprendere altri test statistici anche più avanzati.

Nell'approccio NHST i test prevedono due tipi di ipotesi: un'ipotesi nulla ("null hypothesis", da cui il nome dell'approccio) e un'ipotesi alternativa ("alternative hypothesis"). L'ipotesi alternativa coincide solitamente con l'ipotesi di ricerca (v. Sezione 2.1), e postula dunque quello che si ipotizza di osservare nei dati: nel nostro caso, che vi sia una differenza tra discorsi originali e interpretati per quanto riguarda le frequenze di "that" espliciti o impliciti. L'ipotesi nulla postula il contrario dell'ipotesi alternativa, ossia che non vi sia nessuna differenza. Esplicitare le due ipotesi è fondamentale per interpretare correttamente i risultati dei test basati su NHST: il loro scopo è infatti verificare che l'ipotesi nulla possa essere confutata con un relativo margine di certezza. Questa logica segue uno dei principi fondanti della scienza moderna, basata sulla falsificazione: come afferma Levshina (2015: 9), "[i]t is impossible to prove that something is right, but it is possible to reject the opposite". Confutare l'ipotesi nulla permette quindi di "accettare" indirettamente l'ipotesi alternativa, e dunque l'ipotesi di ricerca iniziale.

Il risultato dei test basati su NHST è solitamente espresso tramite tre valori: il valore p (o probabilità di errore, o " p -value"), il valore della statistica relativa al test (ad esempio il valore χ^2) e i gradi di libertà. Riguardo questi ultimi, basti sapere che possono essere concepiti come un'informazione relativa all'ampiezza del campione, e che in un caso come quello considerato, dove il calcolo viene eseguito sulla base di una tabella con due righe e due colonne (v. Tabella 3), i gradi di libertà sono pari a 1 (v. anche Mellinger e Hanson 2017: 75-76). Il valore della statistica è quello che viene calcolato a partire dai dati del campione secondo la formula prevista dal test, come illustreremo sotto. Infine p è il valore più direttamente associato alla significatività statistica. Tecnicamente, p esprime la probabilità che, data l'ipotesi nulla, il campione presenti un valore della statistica più estremo (più positivo o più negativo) di quello osservato (v. anche Gries 2013: 27-29).

In pratica, il valore p può anche essere pensato come la probabilità che i risultati ottenuti nello studio siano compatibili con l'ipotesi nulla (assenza di una differenza tra gruppi o di un effetto delle variabili considerate): più il valore di p è basso, meno è probabile che l'ipotesi nulla sia confermata. Il valore è espresso come un numero che va da 0 a 1, e, nelle scienze umane e sociali, un valore di

p inferiore a 0,05 (corrispondente a una probabilità del 5%) è generalmente considerato la “soglia critica”, ossia il valore al di sotto del quale è possibile rifiutare l’ipotesi nulla e accettare l’ipotesi alternativa. In altre parole, con $p < 0,05$ si può affermare, ad esempio, che vi è una differenza tra due gruppi e che questa è statisticamente significativa. La soglia dello 0,05 è evidentemente arbitraria, così come lo sono le altre due soglie critiche convenzionali, ossia $p < 0,01$ (1%) e $p < 0,001$ (0.1%). Benché la pratica sia contestata, queste soglie si trovano a volte indicate anche tramite asterischi posizionati accanto ai risultati di un test: un singolo asterisco corrisponde a un valore di p inferiore a 0.05, due asterischi a 0.01 e tre asterischi a 0.001 (v. Mellinger e Hanson 2017: 75). È importante tuttavia ricordare che un valore di significatività statistica più basso, ad esempio $p < 0.001$ invece di $p < 0.05$, non significa che le differenze osservate siano più grandi o più “importanti”. Per giungere a una conclusione simile occorrerebbe calcolare un’ulteriore misura, chiamata dimensione dell’effetto (“effect size”). Si rimanda alla Sezione 3 per una discussione su questo e altri aspetti potenzialmente problematici di cui tenere conto nell’interpretare i risultati ottenuti con un approccio NHST.

Software pensati appositamente per l’analisi statistica come R (R Core Team 2020) o SPSS (IBM Corp. 2020) permettono di calcolare tutti i risultati rilevanti di un test statistico (quasi) automaticamente, ma richiedono uno sforzo di apprendimento considerevole (soprattutto nel caso di R), che può far desistere l’utente alle prime armi. Per i test più semplici, può essere sufficiente conoscere software per la gestione di fogli di calcolo come LibreOffice Calc o Microsoft Excel, ma la procedura richiede un numero maggiore di passaggi¹⁵. Qualunque sia il software selezionato, l’aspetto cruciale è comprendere esattamente quali sono i dati attesi dal software e in quale formato devono essere inseriti, oltre ovviamente a sapere come interpretare i risultati. Illustreremo questo punto

¹⁵ Si noti comunque che i fogli di calcolo supportano un numero molto limitato di test e sono scarsamente configurabili. Ad es., il risultato calcolato con LibreOffice Calc e Excel riportato in questa sezione non coincide con quello di Kajzer-Wietrzny (2012: 99). Infatti l’autrice applica una correzione al test (correzione di Yates), che non è possibile configurare automaticamente in questi software.

tornando all'esempio di Kajzer-Wietrzny (2012) e al test del χ^2 condotto sui suoi dati.

La base di partenza per effettuare il test del χ^2 è una tabella come quella presentata nella Tabella 3 (in cui le frequenze osservate sono riprese dalla Tabella 1). Il calcolo del valore del χ^2 mette a confronto queste frequenze con quelle che ci si aspetterebbe di osservare se l'ipotesi nulla fosse vera, ossia se fosse vero che la proporzione di "that" impliciti ed espliciti è uguale nei discorsi originali e interpretati. Occorre quindi calcolare quali sono le frequenze attese: come suggerito da Mellinger e Hanson (2017: 172), queste devono tenere conto del fatto che non tutte le categorie sono ugualmente frequenti/probabili. Ad esempio, le costruzioni con "reporting verbs" (sia con "that" espliciti che impliciti) sono più frequenti nei discorsi originali che nei discorsi interpretati, e i "that" espliciti (sia nei discorsi originali che in quelli interpretati) sono più frequenti dei "that" impliciti. La probabilità di osservare un "that" esplicito in un discorso originale inglese secondo l'ipotesi nulla è dunque calcolabile come la probabilità di osservare un "that" esplicito (268 casi su 376; $268/376 = 71,28\%$) moltiplicata per la probabilità che una costruzione con "reporting verbs" sia osservata in discorsi originali inglesi (219 casi su 376 = $58,24\%$). Moltiplicando le probabilità risultanti ($71,28\% * 58,24\% = 41,51\%$) per il totale dei casi osservati si ottiene la frequenza attesa ($376 * 41,51\% = 156,1$). Il calcolo può essere semplificato secondo la formula:

$$Frequenza\ attesa = \frac{Totale\ colonna * Totale\ riga}{Totale\ generale}$$

Nel caso appena discusso, si otterrebbe dunque lo stesso risultato moltiplicando il totale delle costruzioni nei discorsi originali (totale della colonna = 219) per il totale di "that" espliciti (totale della riga = 268) e dividendo per il totale dei casi osservati (376): $(219 * 268) / 376 = 156,1$; il calcolo viene poi ripetuto per tutte le altre celle. I risultati sono mostrati nella parte destra della Tabella 3.

	Frequenze osservate			Frequenze attese		
	Orig. EN	Int. FR>EN	Totale	Orig. EN	Int. FR>EN	Totale
<i>that</i>	146	122	268	156,1	111,9	268
Ø connettore	73	35	108	62,9	45,1	108
Totale	219	157	376	219	157	376

Tab. 3 Tabella delle frequenze osservate e attese per il calcolo del χ^2 (dati da Kajzer-Wietrzny 2012: 99).

Una volta stabilite le frequenze osservate e attese è possibile effettuare il test vero e proprio anche con un semplice foglio di calcolo (per la procedura di calcolo completa si rimanda a Gries 2013: 183)¹⁶. Il risultato che si otterrà in questo caso è il valore di p , che qui equivale a 0,019: poiché questo è inferiore a 0,05, l'ipotesi nulla può essere rifiutata. Come menzionato sopra, oltre al valore di p è tuttavia necessario riportare anche il valore del χ^2 , che va calcolato a parte¹⁷.

Terminato questo passaggio, si può finalmente riportare il risultato dello studio. Sulla base dei dati di Kajzer-Wietrzny (2012), al Parlamento europeo i discorsi interpretati dal francese all'inglese dimostrano un livello di esplicitezza lessico-grammaticale più elevata rispetto a discorsi paragonabili prodotti originalmente da parlanti nativi, come dimostrato dalla differenza statisticamente significativa tra le frequenze di "that" impliciti ed espliciti ($\chi^2 = 5,44$; $df = 1$; $p = 0.019$).

3. Conclusioni, limiti e prospettive future

In questo contributo si è cercato di fornire una panoramica, necessariamente limitata, sulle fasi principali che compongono uno studio basato su dati quantitativi e sulla loro analisi con metodi derivati dalla statistica. Partendo dalla

¹⁶ La funzione in LibreOffice Calc è "=TEST.CHI()". Nella parentesi vanno inseriti riferimenti alle celle che contengono i valori della Tabella 3, separati da una virgola: prima le frequenze osservate, poi le frequenze attese.

¹⁷ In LibreOffice Calc: "=INV.CHI()". Nella parentesi va inserito il valore di p ottenuto nella fase precedente e poi, separati da una virgola, i gradi di libertà (qui: 1).

formulazione dell'ipotesi di ricerca e dall'operazionalizzazione delle variabili dipendenti e indipendenti, si è passati a illustrare le statistiche descrittive per esplorare tendenze centrali e dispersione nei dati raccolti. Tramite gli esempi del grafico a colonne e del diagramma a scatola, si è poi sottolineato come la rappresentazione grafica sia un modo non solo per scoprire tendenze del campione di dati, ma anche per presentarle più chiaramente a chi legge lo studio. Nell'ultima sezione, sono stati infine introdotti alcuni dei concetti chiave della statistica inferenziale, tra cui quello di ipotesi nulla e alternativa e di significatività statistica; anche in questo caso, i concetti sono stati illustrati tramite un esempio, quello del test del χ^2 .

Come in tutto il contributo, lo scopo principale non era tanto quello di fornire istruzioni dettagliate su come applicare uno specifico metodo in un possibile nuovo studio. A tal fine si è cercato, ovunque possibile, di inserire riferimenti puntuali a manuali di statistica che permettano di approfondire le questioni a cui si è solo potuto accennare, oppure a ricerche pubblicate che applicano in studi di prodotto e di processo sulla traduzione e l'interpretazione i concetti statistici presentati. Piuttosto, lo scopo prefissato era quello di chiarire i presupposti su cui si fonda uno studio quantitativo e in particolare un test statistico, discutendo anche in parte come funziona il suo calcolo. Così facendo, si è voluto mettere in luce quel livello di complessità che tende a essere oscurato se ci si limita a inserire dati in un software di analisi per ottenere un valore di significatività.

Proprio la tendenza ad attribuire eccessiva importanza ai valori di significatività, specialmente se non accompagnata da un'adeguata comprensione di cosa indichino veramente i valori di p , ha portato a un atteggiamento critico, talvolta radicale, all'approccio NHST (v. ad es. Lakens *et al.* 2018). Queste obiezioni, nate soprattutto all'interno del campo della psicologia, vertono ad esempio sulla natura arbitraria delle soglie di significatività, e sul fatto che ottenere un valore di p considerato significativo porti spesso a tralasciare aspetti altrettanto importanti quali la misurazione dell'effettiva scala delle differenze o degli effetti osservati: una differenza può essere statisticamente significativa anche se la sua grandezza o rilevanza pratica sono minime (v. Baayen 2008: 115).

Visti i limiti dell'approccio NHST, occorre però anche sottolineare che sono state suggerite numerose tecniche per ovviarli, tra cui l'adozione di misure della dimensione dell'effetto, che forniscono appunto informazioni su quanto è grande o rilevante una differenza o un effetto statisticamente significativo (Mellinger e Hanson 2017: 77). A ciò, si aggiunge ovviamente la possibilità di eseguire test più complessi rispetto a quelli che sono stati presentati in questo studio. De Sutter e Lefer (2020) suggeriscono, ad esempio, di adottare design cosiddetti "multifattoriali", che coinvolgono un numero anche elevato di variabili per spiegare i fenomeni osservati, e che possono così tener conto della natura della traduzione "as an inherently multidimensional linguistic activity and product" (*ibid.*: 1). Ivaska et al. (in stampa) dimostrano invece come metodi di rappresentazione grafica quali i diagrammi a scatola possano essere combinati con tecniche di analisi statistica esplorativa, che non coinvolgono cioè un'ipotesi nulla, nell'analizzare i rapporti tra traduzione e altre forme di produzione linguistica multilingue. Padroneggiare le tecniche statistiche coinvolte e i software che permettono di implementarle richiede senza dubbio un percorso di apprendimento non facile. Ma se come afferma Halverson (2019: 69) "[i]ncreasing methodological rigour, for example through the use of advanced statistical methods" è un segno di vitalità negli studi sulla traduzione e l'interpretazione in generale, i risultati ripagheranno lo sforzo.

Bibliografia

Baayen, H. R. (2008) *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*, Cambridge: Cambridge University Press.

Baker, M. (1993) "Corpus Linguistics and Translation Studies: Implications and Applications", in M. Baker, G. Francis e E. Tognini-Bonelli (eds.) *Text and Technology: In Honour of John Sinclair*, Amsterdam/Philadelphia: John Benjamins, 233-250.

Baker, M. e G. Saldanha (eds.) (2019) *Routledge Encyclopedia of Translation Studies*, New York: Routledge.

Blum-Kulka, S. (1986) "Shifts of Cohesion and Coherence in Translation", in J. House e S. Blum-Kulka (eds.) *Interlingual and Intercultural Communication*, Tübingen: Gunter Narr. Verlag, 17-35.

Collard, C. e B. Defrancq (2020) "Disfluencies in Simultaneous Interpreting: A Corpus-based Study with Special Reference to Sex", in L. Vandevoorde, J. Daems e B. Defrancq (eds.) *New Empirical Perspectives on Translation and Interpreting*, New York: Routledge, 264-299.

De Sutter, G. e M.-A. Lefer (2020) "On the Need for a New Research Agenda for Corpus-based Translation Studies: A Multi-methodological, Multifactorial and Interdisciplinary Approach", *Perspectives* 28(1): 1-23.

Ferraresi, A., S. Bernardini, M. Miličević Petrović, M.-A. Lefer (2018) "Simplified or Not Simplified? The Different Guises of Mediated English at the European Parliament", *Meta* 63(3): 716-37.

Field, A. (2018) *Discovering Statistics Using IBM SPSS Statistics. 5th edition*, London: Sage Publishing.

Field, A., J. Miles e Z. Field (2012) *Discovering Statistics Using R*, London: Sage Publishing.

Gile, D. (2016) "Experimental Research", in C. Angelelli e J. B. Baer (eds.) *Researching Translation and Interpreting*, New York: Routledge, 220-228.

Gries, S. Th. (2013) *Statistics for Linguistics with R. A Practical Introduction. 2nd edition*, Berlin and New York: Mouton de Gruyter.

Halverson, S. (2019) "Cognition", in M. Baker e G. Saldanha (eds.) *Routledge Encyclopedia of Translation Studies*, New York: Routledge, 65-70.

IBM Corp. (2020). *IBM SPSS Statistics for Windows, Version 27.0*, Armonk: IBM Corp.

Ivaska, I., A. Ferraresi e S. Bernardini (in stampa) "Syntactic Properties of Constrained English: A Corpus-driven Approach", in S. Granger e M.-A. Lefer

(eds.) *Extending the Scope of Corpus-Based Translation Studies*, London/Oxford: Bloomsbury.

Ji, M. (2012) "Hypothesis Testing in Corpus-based Literary Translation Studies", in M. P. Oakes e M. Ji (eds.) *Quantitative Methods in Corpus-Based Translation Studies*, Amsterdam/Philadelphia: John Benjamins, 53-72.

Kajzer-Wietrzny, M. (2012) *Interpreting Universals and Interpreting Style*. Tesi di dottorato non pubblicata, Poznań: Università Adam Mickiewicz.

Lakens, D., F.G. Adolphi, C.J. Albers *et al.* (2018) "Justify Your Alpha", *Nature Human Behaviour* 2: 168–171.

Laviosa, S. (1998) "Core Patterns of Lexical Use in a Comparable Corpus of English Narrative Prose", *Meta* 43(4): 557-570.

Lee Ling, D. (2015) *Introduction to Statistics Using LibreOffice.org Calc, Apache OpenOffice.org Calc and Gnumeric*, Pohnpei: College of Micronesia, <http://www.comfsm.fm/~dleeling/statistics/text5.html>.

Levshina, N. (2015) *How to do Linguistics with R*, Amsterdam/Philadelphia: John Benjamins.

Marin-Lacarta, M. (2020) "Research Methodologies, Translation", in M. Baker e G. Saldanha (eds.) *Routledge Encyclopedia of Translation Studies*, New York: Routledge, 479-484.

Mellinger, C. D. e T. A. Hanson (2017) *Quantitative Research Methods in Translation and Interpreting Studies*, New York: Routledge.

Olohan, M. e M. Baker (2000) "Reporting *that* in Translated English: Evidence for Subconscious Processes of Explicitation?", *Across Languages and Cultures* 1(2): 141-158.

Picton, A., M. Fontanet e D. Pulitano (2015) "Defining the Notion of 'Corpora' in Translation: Addressing the Gap between Scholars' and Translators' Points of

View”, Working Paper, Università di Ginevra, <https://archive-ouverte.unige.ch/unige:135687>.

R Core Team (2020) *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing, <https://www.R-project.org/>.

Saito, K., S. Webb, P. Trofimovich e T. Isaacs (2016) “Lexical Profiles of Comprehensible Second Language Speech: The Role of Appropriateness, Fluency, Variation, Sophistication, Abstractness and Sense Relations”, *Studies in Second Language Acquisition* 38(4): 677-701.

Shlesinger, M. e N. Ordan (2012) “More Spoken or More Translated? Exploring a Known Unknown of Simultaneous Interpreting”, *Target* 24(1): 43-60.

Snell-Hornby, M. (2006) *The Turns of Translation Studies: New Paradigms Or Shifting Viewpoints?*, Amsterdam/Philadelphia: John Benjamins.

Steiner, E. (2012) “Methodological Cross-Fertilization: Empirical Methodologies in (Computational) Linguistics and Translation Studies”, in O. Czulo e S. Hansen-Schirra (eds.) *Crossroads between Contrastive Linguistics, Translation Studies and Machine Translation*, Berlin: Language Science Press, 65-90.

Toury, G. (1995 [2012]) *Descriptive Translation Studies – and Beyond. Revised edition*, Amsterdam/Philadelphia: John Benjamins.

Vanroy, B., O. De Clercq e L. Macken (2019) “Correlating Process and Product Data to Get an Insight into Translation Difficulty”, *Perspectives* 27(6): 924-941.

Winter, B. (2019) *Statistics for Linguists: An Introduction Using R*, New York: Routledge.