

Analisi di corpora per la traduzione: una lezione introduttiva

Silvia Bernardini – Dipartimento di Interpretazione e Traduzione, Università di Bologna, Campus di Forlì

Citation: Bernardini, Silvia (2020) “Analisi di corpora per la traduzione: una lezione introduttiva”, 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: A1-A26, <http://www.mediazioni.sitlec.unibo.it>, ISSN 1974-4382.

1. Il perché di questa lezione

In anni recenti i volumi e i saggi introduttivi sull'analisi di corpora all'interno di enciclopedie e manuali sulla traduzione e l'interpretazione si sono moltiplicati, particolarmente in pubblicazioni in lingua inglese (p. es. Laviosa 2013, Bernardini e Russo 2017; Bernardini e Kenny 2019). L'immensa disponibilità di testi in formato digitale rende infatti i metodi per la costruzione e consultazione di corpora (raccolte di testi più o meno organiche, consultabili con strumenti ad hoc) particolarmente rilevanti per tutte le attività legate al reperimento di informazioni da testi. Allo stesso tempo, la necessità di ottenere informazioni affidabili in modo rapido ed efficace su corrispondenze terminologiche e su preferenze fraseologiche fra lingue diverse non si è certo ridotta, nonostante i rapidi sviluppi delle tecnologie linguistiche (legate ad esempio alla traduzione assistita e automatica). Visto il ruolo essenziale che queste competenze e capacità ricoprono nella pratica professionale, la loro acquisizione è fra gli obiettivi principali della formazione delle traduttrici e dei traduttori. Il più recente quadro europeo delle competenze in traduzione pubblicato dalla rete EMT (European Master's in Translation) nel 2017 menziona direttamente gli strumenti di analisi

dei corpora come parte della competenza tecnologica (n. 16: “Make effective use of search engines, corpus-based tools, text analysis tools and CAT tools”) e indirettamente il reperimento di informazioni rilevanti da fonti affidabili come parte della competenza traduttiva in senso stretto (n. 3: “Evaluate the relevance and reliability of information sources with regard to translation needs”)¹.

Benché la componente tecnologica sia inevitabilmente presente, e possa costituire un ostacolo all'utilizzo di strumenti pensati più per la ricerca scientifica che per lo studio o la professione, sarebbe un errore incentrare oggi un'introduzione all'uso dei corpora per la traduzione unicamente sulle caratteristiche dei diversi strumenti o tipi di corpora esistenti. Molte informazioni sono reperibili in rete, sia nei testi menzionati sopra che su siti e blog. Inoltre un po' di pratica con strumenti scaricabili gratuitamente o piattaforme a libero accesso consente un utilizzo delle funzionalità di base sufficiente per la maggior parte dei bisogni di un* utente alle prime armi. Il resto viene con l'esperienza, purché l'investimento di energie appaia commisurato al ritorno. Il principale ostacolo ad un utilizzo efficace dei corpora è invece senza dubbio la carenza di capacità generali di ricerca (testuale), pur declinate in modalità specifiche. Identificare una carenza di informazione o di comprensione durante un compito traduttivo che un corpus possa colmare; effettuare ricerche adeguatamente mirate, o sufficientemente ampie, in funzione del problema da risolvere e delle risorse testuali a disposizione; operare generalizzazioni a partire da dati testuali autentici (e quindi spesso contraddittori): sono solo alcune delle operazioni, complesse ma indispensabili, che un utilizzo intelligente dei corpora presuppone. E viceversa, ad un utilizzo meccanico e pedissequo è preferibile la consultazione di un buon dizionario, monolingue o bilingue, o di una memoria di traduzione, strumenti che danno comunque una risposta, benché non sempre la migliore.

In questo contributo non offrirò dunque alcuna panoramica dei corpora disponibili in rete, né delle funzionalità principali degli strumenti per la loro consultazione. Non ripercorrerò la storia della linguistica dei corpora né riassumerò i risultati e gli approcci ascrivibili ai cosiddetti *Corpus-based Translation Studies*, siano essi

¹ https://ec.europa.eu/info/sites/info/files/emt_competence_fwk_2017_en_web.pdf

di tipo teorico, descrittivo o applicato (seguendo la ben nota mappa di Holmes 1972). Cercherò invece di presentare i contenuti di un'ipotetica lezione introduttiva di un Corso di *Corpus Linguistics* (pensato per traduttori e traduttrici, ma ugualmente rilevante in qualsiasi corso di studio focalizzato sullo studio del testo e del discorso). Nella realtà, i contenuti che seguono occupano ben più di una classica lezione accademica di due ore². L'acquisizione delle capacità di porre le domande giuste e di vedere le regolarità emergere dal caos richiede una consuetudine con lo strumento che si acquisisce solo con il tempo e con la pratica. Come molte delle discipline che formano il syllabo di un Corso magistrale in traduzione, la *Corpus Linguistics* non si può insegnare, ma solo imparare, sotto la guida di un facilitatore/facilitatrice.

2. I principi cardine: relazioni sintagmatiche, competenza comunicativa, lessico e fraseologia

Se l'utilizzo di un corpus richiede uno sforzo cognitivo e un dispendio di tempo superiori rispetto ad altri strumenti e risorse che utilizziamo quotidianamente per capire e per produrre i testi vale la pena chiedersi perché utilizzare un corpus, qual è il suo valore aggiunto per la comprensione e produzione linguistica. Per rispondere a queste domande è necessario richiamare brevemente e in modo del tutto sommario alcuni principi linguistici alla base della *Corpus Linguistics*, tracciando un ideale percorso che va da Saussure, attraverso Firth e Hymes, a Sinclair.

Il primo e più generale principio che troviamo nel *Cours* di Saussure è l'idea che la linguistica sia una delle discipline che studiano la vita dei segni all'interno della vita sociale: "On peut donc concevoir *une science qui étudie la vie des signes au sein de la vie sociale* [...] La linguistique n'est qu'une partie de cette science

² I contenuti descritti in questo saggio si sono venuti delineando negli anni, come parte del modulo di *Corpus Linguistics* del Corso di Laurea Magistrale in Specialized Translation del Dipartimento di Interpretazione e Traduzione, Università di Bologna.

générale” (Saussure 1916 [1971]: 33)³. Potremmo dire, con Halliday (1978), “language as social semiotics”. L’attenzione alla dimensione sociale in relazione a quella individuale si realizza pienamente nella nota suddivisione della lingua in *langue* (il repertorio linguistico di una comunità) e *parole* (l’atto linguistico individuale). La *langue* è una risorsa condivisa, paragonabile ad un dizionario le cui copie, identiche, sono presenti nella testa di ogni individuo. La *parole*, a sua volta, è la somma di quello che le persone dicono; si realizza quindi in modo estemporaneo. I due aspetti entrano però, inevitabilmente, in una relazione dialogica: se non conosciamo la *langue* non capiremo un atto di *parole* (un testo, un discorso), ma allo stesso tempo la *langue* non esisterebbe se non come generalizzazione di molteplici atti di *parole*: la *langue* “est un trésor déposé par la pratique de la parole dans les sujets appartenant à une même communauté” (*ibid.*: 30). Se la valorizzazione degli atti di *parole* è un passaggio importante a sostegno dell’analisi di testi autentici, due altri aspetti dell’approccio saussuriano sono forse ancora più fondamentali. Da un lato, il richiamo al fatto che la separazione fra morfologia e sintassi da un lato, e lessicologia dall’altro, è essenzialmente arbitraria, dato che “tout mot qui n’est pas une unité simple et irréductible ne se distingue pas essentiellement d’un membre de phrase, d’un fait de syntaxe” (*ibid.*: 185). Il continuum tra lessico e sintassi è d’altronde alla base di un ulteriore assunto, fondamentale per la *Corpus Linguistics*, ovvero il ruolo che rivestono nella competenza dei parlanti i rapporti sintagmatici tra parole, accanto ai rapporti paradigmatici. I secondi, ben noti e ampiamente studiati in semantica, logica e filosofia, descrivono le associazioni fra parole che condividono tra loro tratti formali o di significato. Saussure fa l’esempio della parola “enseignement”, che farà pensare a parole di cui condivide la radice (“enseigner”, “renseigner”), il suffisso (“armement”, “changement”), parte del significato (“éducation”, “apprentissage”), ecc.. Queste associazioni non fanno tipicamente riferimento alla *parole*, perché non è probabile che appaiano insieme nei testi. Sono invece parte integrante “de ce trésor intérieur qui constitue la langue chez chaque individu” (*ibid.*: 20). I rapporti sintagmatici, viceversa, si

³ Il testo integrale dell’edizione del 1971 del *Cours de linguistique generale* è reperibile in rete: https://fr.wikisource.org/wiki/Cours_de_linguistique_générale/Texte_entier.

riferiscono ai legami che si creano fra parole per il fatto di apparire insieme nei testi. Due o più parole che appaiono una di seguito all'altra nei testi stabiliscono fra di loro un rapporto di tipo sintagmatico ("re-lire; contre tous; la vie humaine; Dieu est bon; s'il fait beau temps, nous sortirons") (*ibid.*: p. 20). Sarebbe un errore però, secondo Saussure, limitare l'importanza dei rapporti sintagmatici a fatti di *parole*. Se siamo in grado di riconoscere non solo le espressioni fisse (proverbi, espressioni idiomatiche) ma anche la convenzionalità di gruppi di parole (o sintagmi) formati sulla base di regole generali è perché li abbiamo incontrato un numero sufficiente di volte per far sì che siano entrati a far parte della *langue*:

il faut attribuer à la langue, non à la parole, tous les types de syntagmes construits sur des formes régulières. En effet, comme il n'y a rien d'abstrait dans la langue, ces types n'existent que si elle en a enregistré des spécimens suffisamment nombreux". (*Ibid.*: 173)

Se gli aspetti sintagmatici sono parte integrante della nostra competenza linguistica implicita (e a supporto di questo assunto sarebbe sufficiente richiamare giochi che si basano proprio su di esso, come "La ghiottina", parte di un popolare programma televisivo serale della televisione italiana)⁴, non sono però certo di facile descrizione e categorizzazione. Infatti il rapporto dialettico tra *langue* e *parole* è tale che nello studio dei rapporti sintagmatici sia spesso impossibile stabilire un confine netto tra uso collettivo (o convenzionale) e uso individuale:

il n'y a pas de limite tranchée entre le fait de langue, marque de l'usage collectif, et le fait de parole, qui dépend de la liberté individuelle. Dans une foule de cas, il est difficile de classer une combinaison d'unités, parce que l'un et l'autre facteurs ont concouru à la produire. (*Ibid.*: 173)

Non sorprenderà dunque che il reperimento di informazioni da corpora sia attività cognitivamente, piuttosto che tecnicamente, complessa. È l'oggetto di studio a rendere complessa l'attività e non (solo) lo strumento.

⁴ Il legame tra *langue* e *parole* in questo specifico frangente è confermato dalla possibilità di risolvere il gioco tramite un'applicazione informatica che si basa proprio sui rapporti sintagmatici (Sangati *et al* 2018).

Se i principi fondamentali della *Corpus Linguistics* sono dunque già presenti in Saussure, le origini vere e proprie della disciplina (almeno nella sua incarnazione sinclairiana), vengono comunemente fatte risalire alle intuizioni di John Rupert Firth. Non è un caso infatti che John McH. Sinclair venga spesso identificato come il capostipite della scuola linguistica *neo-Firthian* o *British* (Stubbs 1996; McEnery e Hardie 2012, cap. 6). Il nome di Firth è indubbiamente legato in primis al concetto di *collocazione* – sebbene il termine sia già presente in Palmer (1933), un lavoro non a caso incentrato sull'apprendimento della lingua inglese da parte di apprendenti giapponesi. Secondo Firth è errato ritenere che le parole singole abbiano un significato isolato e possano essere usate in combinazioni libere (Firth 1952/3: 18). È in questo senso che va intesa la famosa (e piuttosto oscura) affermazione "[o]ne of the meanings of *night* is its collocability with *dark*, and, of *dark*, of course, collocation with *night*" (Firth 1957: 196)⁵. L'importanza di Firth non si limita però al concetto di collocazione, tanto spesso citato in relazione al linguista britannico. Più in generale, troviamo in Firth l'idea della natura stereotipica dell'uso linguistico. Sebbene le possibilità di ricombinazione delle risorse di una lingua siano praticamente infinite, nella realtà questo potenziale viene sfruttato in modo molto limitato, perché, osserva Firth, per ciascun ruolo sociale che svolgiamo quotidianamente, in ogni situazione, non inventiamo noi le battute. Queste sono invece già date: "stereotyped and narrowly conditioned by our particular type of culture" (Firth 1935: 69). Un terzo aspetto, collegato al secondo e particolarmente fecondo per le applicazioni in ambito traduttivo, è legato al significato contestuale e alle lingue di specialità. Facendo sue le nozioni di contesto d'uso situazionale e culturale (*context of situation* e *context of culture*) introdotte da Bronisław Malinowski, Firth propose il principio secondo il quale lo studio del significato contestuale dovesse essere circoscritto a micro-

⁵ In realtà una ricerca nel corpus ukWaC (accessibile qui: <https://corpora.dipintra.it>; Baroni *et al* 2009) rivela che *dark* non è fra i collocati più significativi di *night*. Limitandoci agli aggettivi che precedono immediatamente la parola, *last*, *late*, *every*, *opening*, *previous* e *first* vengono tutti prima di *dark* nella lista dei collocati di *night* (ordinati per Log-Dice). Uno degli argomenti principali a favore del corpus come fonte di informazioni linguistiche è proprio legato ai limiti dell'introspezione e dell'intuizione. Espressioni fraseologiche molto frequenti nell'uso, come quelle qui riportate, tendono infatti a passare inosservate a favore di usi più concreti e salienti (Sinclair 1991).

grammatiche e micro-lessici specifici di “situations or sets of or series of situations proper to them, e.g. technical languages [...] or indeed any form of speech or writing with specialized vocabulary, grammar and style” (1956: 106). Queste osservazioni sono confermate da numerosi lavori recenti che sottolineano l'importanza fondamentale dei registri (ad es. Biber 2012) e ipotizzano perfino una strutturazione della competenza linguistica dei parlanti in repertori grammaticali multipli, variabili in funzione del registro/genere (Iwasaki 2015).

Questa idea è d'altronde presente anche in un altro importante precursore della *Corpus Linguistics*, ovvero l'antropologo e sociolinguista americano Dell Hymes. In aperta opposizione agli approcci linguistici generativo-trasformazionali e al loro focus sulla competenza *linguistica*, Hymes propose come centrale per la ricerca linguistica il concetto di competenza *comunicativa*. Infatti, “[a] normal child acquires knowledge of sentences, not only as grammatical but also as appropriate. He or she acquires competence as to when to speak, when not, and as to what to talk about with whom, when, where, in what manner” (Hymes 1972: 277). Come in Firth, l'oggetto di studio della linguistica per Hymes è dunque la competenza nell'uso della lingua *in contesti reali*. Il legame con gli studi basati su corpora è d'altronde già esplicito in un'affermazione modernissima, secondo la quale “[t]he capabilities of language users do include some (perhaps unconscious) knowledge of the probabilities and shifts in them as indicators of style, response etc.” (*ibid.*: 286).

L'ampia letteratura sui rapporti fra frequenza d'uso in un corpus e competenza comunicativa implicita o esplicita (ad es. Wray 2002; Durrant e Schmitt 2009, Ellis e Simpson-Vlach 2009; Granger e Bestgen 2014; Siyanova-Chanturia 2015) conferma l'importanza fondamentale di questo passaggio, che non deve sfuggire a chi si avvicina ai metodi della *Corpus Linguistics* per l'insegnamento e la pratica della traduzione. L'assunto primo, che motiva la ricerca sulle frequenze d'uso in un corpus, è infatti il legame che si suppone esista fra frequenza d'uso e rappresentazione mentale di una lingua. La teoria di Hoey sul “lexical priming”, ad esempio, si basa proprio sull'idea che le collocazioni (di cui parleremo più diffusamente nella sezione seguente) siano innanzitutto fenomeni psicolinguistici la cui esistenza può essere studiata attraverso misure statistiche di co-

occorrenza fra parole applicate a corpora di testi (“collocation is a psycholinguistic phenomenon, the evidence for which can be found statistically in computer corpora”, Hoey 2005: 5). Hoey utilizza a questo proposito l’immagine della “concordanza mentale”: proprio come in un corpus possiamo cercare una parola e vedere, alla sua destra e sinistra, quali altre parole la accompagnano in determinati registri, così, a seguito di incontri ripetuti con le parole in situazioni reali d’uso, nella nostra mente rimarrebbe impressa *l’immagine* dei contesti (sia esterni che testuali) più frequenti in cui abbiamo incontrato quelle parole. Se conoscere una parola significa anche conoscere le parole con le quali tipicamente si accompagna (*ibid.*: 8), non sfuggirà come una ricerca in un corpus permetta di accedere a informazioni preziose per diversi utenti. Per un apprendente di una lingua straniera, permette di velocizzare e sistematizzare un processo di “incontro” altrimenti casuale, tipico dell’esperienza di un/a parlante nativo/a. Per un linguista, apre una finestra sui processi attraverso i quali l’esperienza interagisce con la competenza linguistico-comunicativa. Per un/a linguista applicato/a (traduttore o traduttrice, lessicografo/a, terminologo/a), offre una miniera di informazioni sulla base delle quali interpretare singoli eventi linguistici (più o meno tipici, più o meno creativi), o offrire descrizioni di significato e uso basati sulla performance linguistica di un gruppo invece che unicamente sull’intuizione di un singolo individuo. Il punto essenziale, però, non è solo la maggiore affidabilità e ricchezza delle conoscenze acquisibili. Secondo Sinclair (2004: 132), di cui ci occuperemo di seguito, l’analisi di corpora ha posto gli studiosi di fronte all’esistenza di preferenze legate all’uso delle parole che avrebbero dovuto portare ad una radicale revisione del modo di intendere i rapporti fra le tradizionali branche della grammatica (intesa in senso lato): “[t]he need for a more detailed and abstract model for lexical description became clear when lexical information began to be extracted from multimillion word corpora in the early 1980s”. Il contemporaneo emergere e l’affermarsi di approcci linguistici basati sull’uso, come la grammatica cognitiva (Langacker 1987), la grammatica delle costruzioni (Goldberg 1995) e la semantica dei *frame* (Fillmore 1982), oggi fra i più influenti a livello mondiale, confermano la fondamentale correttezza di questa intuizione.

3. Rapporti sintagmatici e uso delle parole

Entrando nel vivo della nostra lezione introduttiva, è opportuno ora presentare alcuni dei fenomeni linguistici osservabili in un corpus a cui abbiamo fatto riferimento nella sezione precedente. Sebbene un testo fondamentale come quello di McEnery e Hardie (2012) ponga sullo stesso piano l'approccio cosiddetto "neo-Firthiano" e diversi altri approcci all'analisi di corpora (in particolare per lo studio della variazione sincronica e diacronica di tipo sociolinguistico e dei registri), ritengo che il primo rivesta un ruolo decisamente più importante degli altri, in particolare per l'introduzione di concetti e categorie che modificano radicalmente il nostro modo di comprendere il discorso e di intendere l'apprendimento e l'uso linguistico. In tal senso, i concetti che descriviamo in questa sezione si applicano ugualmente agli studi che hanno lo scopo di descrivere le varianti d'uso di una parola o struttura linguistica, così come agli studi che hanno lo scopo di descrivere differenze fra testi e varietà testuali, come registri e dialetti (Biber 2012: 12). Nella prossima sezione faremo un breve riferimento anche a questa seconda modalità di utilizzo dei corpora, la cui natura, però, appare di più semplice comprensione, in quanto simile per metodo e per scopo alla classica analisi testuale à la Beaugrande e Dressler (1981) – seppure assistita da strumenti informatici. Ben più innovativo e radicale è proporre un nuovo modo di interpretare e descrivere il significato e l'uso delle parole. In questo senso la linguistica dei corpora di stampo neo-Firthiano ha in sé il proprio quadro teorico di riferimento e non è dunque solo un metodo per acquisire dati empirici. Sebbene secondo McEnery e Hardie (2012: xiv) la disciplina in futuro non sarà indipendente, ma si configurerà come un metodo a supporto di altre discipline ("corpus methods in linguistics' rather than 'corpus linguistics' standing independently"), almeno nell'ambito della linguistica applicata la visione teorica neo-Firthiana dell'uso linguistico appare ancora oggi superiore per profondità e originalità delle intuizioni teoriche ad altre tipologie di studi su corpora, e quindi essenziale *anche* per giustificare l'utilizzo del "corpus come metodo"⁶.

⁶ Su questo tema si veda anche il dibattito, nato sulla Corpora List e poi confluito in un numero speciale dell'*International Journal of Corpus Linguistics* (2010). Di particolare interesse il breve

Secondo Sinclair (di cui citerò qui il testo del 2004, rimandando per approfondimenti almeno all'opera essenziale del 1991, *Corpus, Concordance, Collocation*), la concezione del lessico nella teoria linguistica è rimasta sostanzialmente invariata nel tempo, a differenza delle categorie sintattiche che si sono arricchite di molti livelli di complessità. L'ipotesi di Sinclair è che questa ridotta sensibilità alle questioni lessicali fosse dovuta principalmente alla difficoltà di giungere a generalizzazioni basandosi unicamente sull'intuizione. Laddove le questioni riguardanti la correttezza grammaticale si configurano come "regole" apprese fin dai banchi della scuola, prestandosi quindi ad un giudizio netto e condiviso di un/a qualsiasi parlante, le (dis)preferenze lessicali raramente portano a valutazioni altrettanto nette, a censure o critiche, e quindi più difficilmente emergono nella consapevolezza dei parlanti. Ma una volta che i dati di frequenza e co-occorrenza fra parole hanno cominciato ad essere disponibili grazie all'avvento dei primi grandi corpora (a partire dagli anni '80), la necessità di un modello più dettagliato ed astratto per descrivere il lessico di una lingua è diventato evidente (Sinclair 2004: 132). Sinclair teorizza l'esistenza di due tendenze opposte, che insieme regolano i comportamenti delle parole: da un lato, il principio di scelta *aperta* ("open choice principle"), tipico della terminologia, e in particolare delle nomenclature. Nei casi più netti, questo principio regola l'utilizzo di parole che hanno un riferimento fisso nel mondo, tale per cui "anyone wanting to name [their] referent would have little option but to use [them]" (Sinclair 2004: 29); dall'altro il principio idiomatico ("idiom principle"), tipico della fraseologia, secondo il quale le parole si attraggono a vicenda e derivano parte del proprio significato dalla reciproca interazione, "[they] make meanings by their combination" (*ibidem*), secondo una formulazione di chiara derivazione Firthiana. Se normalmente si ritiene che l'uso lessicale standard sia regolato dal primo principio (libera combinabilità delle parole nel rispetto delle sole regole grammaticali di una lingua) e che il secondo si applichi solo a casi eccezionali (frasi fatte, espressioni idiomatiche, proverbi ecc.), Sinclair capovolge il punto di vista. Di regola non sono le parole singole ad esprimere i significati. Le unità di

intervento conclusivo di Wynne (2010), che riassume in modo terso l'argomentazione di queste righe.

significato (“units of meaning” o “multiword lexical items”) sono invece tipicamente complesse, ovvero formate da più parole che stabiliscono fra loro legami di tipo e forza diversi.

Lo schema proposto da Sinclair (2004: 141-142) per l’unità di significato si compone di quattro tipi di rapporti sintagmatici, stabiliti cioè fra una parola e altre parole presenti nel suo co-testo immediato: la collocazione (“collocation”), ovvero il rapporto con altre parole (principalmente lessicali); la colligazione (“colligation”), ovvero il rapporto con categorie di parole funzionali; la preferenza semantica (“semantic preference”), ovvero il rapporto con parole appartenenti ad uno stesso campo semantico e la prosodia semantica (“semantic prosody”, anche definita in modo più trasparente “discourse prosody” o “evaluative prosody”), ovvero il rapporto con parole che esprimono un’intenzione o un atteggiamento di chi scrive o parla. Da un punto di vista tecnico tutti questi rapporti possono essere studiati attraverso la ricerca di una parola e l’analisi delle parole di co-testo, analisi che a sua volta può essere condotta sia grazie allo strumento delle “collocazioni”, sia direttamente attraverso l’analisi di righe di concordanza (esempi della parola cercata ordinabili a seconda del co-testo di destra o di sinistra). Si noti che in questo caso il termine “collocazione” viene utilizzato in modo generale, a indicare la co-occorrenza di elementi lessicali reperiti da un’applicazione informatica sulla base di parametri quali la distanza fra le parole e le statistiche di co-occorrenza (che favoriranno di volta in volta parole più rare e particolari o viceversa parole più comuni e frequenti). La Tabella 1 mostra i primi dieci collocati della parola *night* nella posizione che precede immediatamente la parola cercata (1 parola a sinistra, o posizione “meno 1”), ottenuti utilizzando due diverse misure statistiche (T-Score e Mutual Information). Il corpus è il Brown Corpus, uno dei primi corpora in formato elettronico, contenente un milione di parole di inglese americano pubblicato nel 1961 (Kučera e Francis 1964).

T-Score	Mutual Information
last	Saturday
at	Wednesday
the	tomorrow

Saturday	Friday
that	Tuesday
one	last
tomorrow	Monday
Friday	Sunday
all	previous
Sunday	dark

Tab. 1 Collocati di *night* nel *Brown Corpus* ordinati in ordine decrescente di T-Score e Mutual Information (posizione -1).

Secondo il modello di Sinclair è possibile innanzitutto distinguere, fra tutte le parole presenti in tabella (tutti collocati in senso lato) le collocazioni, ovvero le singole parole che tipicamente precedono *night*, formando con quest'ultima delle sequenze familiari. La collocazione è quindi il livello base dell'analisi, più superficiale degli altri in quanto fa riferimento alla co-occorrenza di singole parole. La capacità analitica umana in questo caso è necessaria solo per individuare le coppie più interessanti, basandosi sul conteggio delle volte che una determinata parola appare nelle vicinanze di un'altra⁷. Il fatto che il primo collocato della lista ordinata per *Mutual Information* sia "Saturday", ad esempio, può essere culturalmente rilevante ma non sorprendente⁸. A seconda degli scopi dell'analisi si può decidere di includere o escludere dal computo le sequenze formate con singole parole che rispondono a scelte più grammaticali che lessicali (*last, at, the*

⁷ In realtà la procedura è tutt'altro che totalmente automatica. Come si sarà notato, le parole estratte tramite *T-Score* sono molto diverse da quella estratte tramite *Mutual Information* (innanzitutto la prima contiene molte più parole funzionali e solo la seconda contiene l'aggettivo *dark*, menzionato da Firth). Ma mentre nel caso dei successivi livelli di analisi il lavoro dell'utente è a valle dell'estrazione, in questo caso la scelta è operata prioritariamente. Nel decidere il metodo di estrazione delle collocazioni occorrerà dunque riflettere e ove opportuno giustificare esplicitamente la scelta di una o più misure statistiche di co-occorrenza lessicale (per approfondimenti si vedano Rychlý 2008 e Evert 2009).

⁸ Più curioso il fatto che il secondo collocato di *night* sia *Wednesday*. Per scoprire il perché potremmo cercare questa parola nello stesso corpus e analizzarne i collocati, e così via, con una procedura che potremmo definire "serendipitous", facendoci cioè guidare nell'esplorazione dal caso e dalla curiosità (Bernardini 2000).

ecc.)⁹. Il secondo livello di analisi, quello delle colligazioni, consiste nel raggruppare i collocati (in senso lato) che appartengono ad una determinata classe grammaticale¹⁰. Nel caso di *night*, la lista dei collocati ordinati per *T-Score* mostra ad esempio una chiara preferenza per parole appartenenti alla categoria dei “determiner” (*last, the, that, one, all*). Qui, come di seguito, non è sufficiente elencare le parole reperite nel corpus: è necessario uno sforzo di categorizzazione che costringe l’utente a generare “regole” sintagmatiche, ovvero di co-occorrenza (in questo caso grammaticale).

Il livello successivo è invece prettamente semantico: consiste nel raggruppare collocati che condividono l’appartenenza ad un determinato ambito di significato. Per esemplificare questo tipo di rapporto fra parole osserviamo il caso del verbo inglese *unleash* il cui significato figurato, liberamente tradotto, è “liberare repentinamente una forza incontrollabile”¹¹. La Tabella 2 mostra i primi 50 collocati del lemma estratti dal corpus ukWaC, limitando la ricerca al co-testo di destra (posizione +1 o +2) e alle parole che hanno una frequenza di almeno 30 nel corpus e di almeno 15 nella posizione di nostro interesse. La misura statistica è la *Mutual Information*.

⁹ Anche la distinzione fra parole funzionali e parole lessicali (o parole appartenenti a classi aperte o chiuse) è in realtà da considerarsi un gradiente, non un’opposizione netta: *last*, ad esempio, può essere considerato funzionale, in quanto determinatore (*last night*), o lessicale, in quanto aggettivo (*the last person I saw*).

¹⁰ Per semplicità qui facciamo riferimento alle liste di collocazioni, ma le stesse analisi sono possibili anche analizzando le concordanze e registrando manualmente le co-occorrenze di parole co-testuali. In tutti i casi in cui le co-occorrenze sono categorizzate (grammaticalmente, semanticamente o pragmaticamente) anche i casi singoli diventano rilevanti, se riconducibili a categorie generali. Un’analisi talmente dettagliata da prendere in considerazione anche i casi singoli non può essere condotta sulla base delle collocazioni (che escludono a priori le singole co-occorrenze), ma solo utilizzando le concordanze.

¹¹ Libera traduzione della definizione del *Cambridge English Dictionary*, <https://dictionary.cambridge.org/dictionary/english/unleash>.

unstoppable, torrent, ferocious, fury, wrath, mayhem, fierce, creativity, devastating, deadly, entrepreneurial, artillery, terror, awesome, inner, shot, tremendous, storm, hell, wave, powerful, creative, potential, flood, debut, upon, power, talent, destruction, evil, onto, massive, violence, superb, force, nuclear, weapon, true, dog, full, war, attack, energy, against, his, their, new, your, city, another, its, low

Tab. 2 I primi 50 collocati di *unleash* (*Mutual Information*, posizione +1 e +2).

Accanto a parole la cui presenza era prevedibile sulla base della definizione (come *unstoppable* o *powerful*) si possono osservare altre parole che si raggruppano fra di loro per un tratto del significato e la cui presenza non era scontata. Sebbene la categorizzazione lasci ampio margine alla discrezionalità, si può notare la presenza di un gran numero di parole riferite agli ambiti semantici della violenza (*ferocious, fury, wrath, mayhem, devastating*) e un numero di parole più ridotto ma ben delimitato, riferito a eventi atmosferici e alle loro conseguenze (*torrent, storm, wave e flood*); queste ultime sono usate a loro volta in modo figurato (un aspetto osservabile solo dalla concordanza). A queste preferenze semantiche se ne potrebbe aggiungere una terza, diametralmente opposta, per parole riferite alle potenzialità dell'essere umano, come *creativity/creative, entrepreneurial, inner, talent*. Sulla base delle somiglianze semantiche fra collocati è possibile capire meglio in quali situazioni si usa normalmente la parola (in riferimento a forze distruttive ma anche a forze creative) ed estendere la nostra capacità di usarla in modo appropriato, o addirittura creativo. Sulla scorta della generalizzazione relativa agli eventi atmosferici, ad esempio, potremmo optare per l'utilizzo del termine *deluge* ("unleash a *deluge* of complaints") che, sebbene non sia un collocato frequente, è coerente con la preferenza semantica osservata; più dubbio sembrerebbe l'uso di un'espressione come "unleash an *earthquake* of complaints", che non trova supporto nell'analisi appena effettuata (un terremoto non è un evento atmosferico).

L'ultimo e più complesso livello di analisi, quello della prosodia semantica, consiste nel raggruppare i collocati che esprimono indirettamente un punto di vista o un atteggiamento, hanno cioè delle connotazioni particolari. Il termine "prosodia", che si richiama al concetto fonetico, suggerisce infatti l'idea che una

parola “assorba”, per vicinanza, tracce di connotazioni dai propri collocati tipici. La parola non sviluppa connotazioni proprie, ma acquisisce un’aura che ne condiziona l’uso in modo talvolta evidente, talvolta quasi impercettibile. Naturalmente, l’esistenza di questo fenomeno diventa più visibile nel momento in cui viene contraddetto. Partington (2017) discute molti casi di quello che definisce “evaluative clash”, ovvero un contrasto tra un punto di vista espresso in un caso specifico e le aspettative del lettore, che si basano sugli incontri pregressi con la stessa parola. Particolarmente interessante l’esempio del participio passato inglese (*be*) *embroiled in*, utilizzato (in un testo scritto da un linguista) in riferimento ad una situazione di per sé non negativa (*positive-sum games*). Il verbo è sinonimo di “be involved in”, ma a differenza di quest’ultimo, che nel corpus ukWaC co-occorre con parole che indicano situazioni positive, negative e neutre (*accident, project, development*), è quasi immancabilmente utilizzato in riferimento a situazioni negative, spesso riferite a contrasti tra persone (*scandal, controversy, litigations*). È l’aspettativa negata di questo co-testo negativo a generare nel testo un’incongruenza e nel lettore incertezza interpretativa.

In conclusione, secondo Sinclair l’unità di significato non è tipicamente una singola parola ma un’espressione fraseologica più o meno fissa, formata dall’unione della parola e delle sue collocazioni, colligazioni, preferenze e prosodie semantiche. È questo il caso di *budge*, il cui uso non può essere descritto se non tenendo in considerazione tutti i livelli visti sopra, in primis la particolarissima prosodia semantica (“frustration [...] at the refusal or inability of some obstacle to move, despite pressure being applied”, Sinclair 2004: 141). Non necessariamente tutti i tipi di relazione saranno ugualmente rilevanti in tutti i casi, però: l’esperienza insegnerà a identificare rapidamente gli aspetti più salienti. L’importante è sviluppare la sensibilità ai legami sintagmatici che si stabiliscono tra le parole tramite l’uso, imparare a leggerne le implicazioni e tenerne conto quando creiamo i nostri testi/discorsi, in produzione spontanea e ancor più in traduzione. Da un lato l’uso linguistico di tipo “interpretativo” (Gutt 1991) tipico della traduzione restringe la nostra libertà nelle scelte espressive e ci costringe (ancora di più rispetto all’uso linguistico “descrittivo”) a relazionarci con le scelte degli altri. Dall’altro lato la nostra competenza comunicativa nelle due lingue coinvolte è tipicamente (se non necessariamente) diversa, tale per cui

necessitiamo di risorse a sostegno delle nostre intuizioni *in particolare* nella lingua dove risultino meno affidabili. Per questi motivi l'uso di corpora per lo studio dei legami sintagmatici è praticamente inevitabile per un traduttore o una traduttrice.

4. Oltre lo studio dei rapporti sintagmatici nella costruzione del significato

Come sottolineato da McEnery e Hardie (2012), le potenzialità dei metodi di analisi basati su corpora vanno ben oltre il nucleo teorico essenziale dell'approccio neo-Firthiano, estendendosi fra gli altri alla linguistica contrastiva e diacronica e all'analisi (critica) del discorso. Un'utile differenziazione in questo senso, già introdotta in questo contributo, è quella proposta da Biber (2012: 12) tra studi che hanno lo scopo di descrivere le varianti d'uso di una parola o struttura linguistica e studi che hanno lo scopo di descrivere differenze fra testi e varietà testuali come registri e dialetti. I concetti descritti nella sezione precedente rientrano tipicamente nel primo tipo di studio, anche se naturalmente il loro focus può essere ristretto a corpora di specialità che rappresentano varietà linguistiche specifiche. Il secondo tipo di studio utilizza invece proprio quest'ultimo tipo di corpus e cerca di descriverne le caratteristiche tipiche, spesso paragonandole a quelle di altre varietà, rappresentate da corpora "paragonabili" ai primi. Si tratta in pratica di applicare metodi di analisi testuale informatizzati ad un insieme omogeneo di testi. Una volta ben compreso (e adeguatamente esemplificato) il ruolo dei rapporti sintagmatici nella costruzione del significato e nella definizione degli usi lessicali/fraseologici tipici delle parole (la nostra "prima lezione"), il corso di *Corpus Linguistics* può procedere illustrando e approfondendo una o più questioni di ricerca all'interno di uno di questi ambiti, secondo gli interessi della classe.

Le applicazioni di questo tipo sono virtualmente infinite, come infinite sono le varietà linguistiche di interesse degli studiosi. Per menzionare solo alcune macro-categorie, limitandoci alla lingua inglese contemporanea, possiamo studiare/paragonare l'inglese delle diverse varietà geografiche (ad es. Hundt 2020), l'inglese degli apprendenti (ad es. Granger *et al.* 2017), l'inglese lingua

franca (ad es. Prodromou 2008), l'inglese tradotto (ad es. Laviosa 2002), l'inglese accademico (ad es. Connor e Upton 2004) l'inglese giornalistico (ad es. Bednarek and Caple 2017) e così via. Molti corpora di specialità di questo tipo sono disponibili per la consultazione online o per il download e la consultazione in locale¹². Ad un livello ulteriore di dettaglio possiamo definire varietà di interesse più o meno ampie sulla base di specifiche ipotesi di ricerca (discorsi di personalità politiche o della cultura, tweet su argomenti di attualità, sottotitoli di serie tv ecc.). In questo caso sarà probabilmente necessario raccogliere i testi personalmente (un'operazione che richiede riflessione, progettazione e documentazione) e analizzarli con un programma gratuito e di semplice utilizzo come AntConc o con un programma più completo (e complesso) come SketchEngine (il cui utilizzo è al momento gratuito per i membri delle università europee)¹³.

Una volta definita la varietà di interesse sarà necessario decidere su quali aspetti testuali concentrare la nostra attenzione. Alcune caratteristiche che esulano da un'ipotesi di ricerca specifica possono essere, per così dire, "suggerite" dal corpus. Fra queste ci sono gli indici di complessità e di ripetitività come il rapporto type/token, il numero di parole uniche, la densità lessicale, la lunghezza media delle frasi, che permettono di valutare aspetti legati al registro; le liste di frequenza delle parole nel corpus e le liste di parole chiave nel corpus rispetto ad un corpus di riferimento, che permettono di osservare il lessico caratteristico (e dunque i temi principali); le liste di ngrammi, che offrono un punto di accesso alla fraseologia e permettono di delineare meglio i temi e i modi in cui vengono affrontati. Qualora invece lo studio parta da un'ipotesi di ricerca (ad esempio un'intuizione o un'affermazione di uno studio precedente da confermare o smentire), e sia dunque *corpus-based* invece che *corpus-driven* (Tognini Bonelli 2001), gli strumenti di generazione di concordanze e di collocazioni daranno accesso al ricco insieme di informazioni sintagmatiche ampiamente discusso nella sezione precedente.

¹² Si vedano, fra le molte altre, le piattaforme <https://www.english-corpora.org/> e <https://www.clarin.eu/portal>.

¹³ <https://www.laurenceanthony.net/software/antconc/>; <https://www.sketchengine.eu>.

Un ultimo punto degno di attenzione riguarda i corpora plurilingue. Se l'analisi contrastiva di corpora di specialità in lingue diverse (i cosiddetti corpora multilingue paragonabili) non si discosta da quanto appena descritto (anche se sarà necessario tenere a mente le specificità di ciascuna lingua nel trarre conclusioni, ad esempio relativamente agli indici di complessità e ripetitività), un caso particolare è costituito invece dai corpora multilingue paralleli. Questi corpora, contenenti testi legati da un rapporto di traduzione, e quindi allineabili a livello di segmento/periodo/paragrafo, vengono tipicamente utilizzati per raccogliere e categorizzare le scelte traduttive fatte da diverse/i professioniste/i. Ciascuno dei sotto-corpora all'interno di un corpus parallelo può essere sottoposto (indipendentemente) alle analisi di cui abbiamo detto sopra, ma perché possa essere consultato anche in modalità parallela è necessario l'allineamento, che viene effettuato automaticamente (talvolta con correzione manuale), di solito a livello di periodo. La necessità dell'allineamento e di strumenti specifici per la consultazione rende queste risorse più rare (e preziose) rispetto alle altre tipologie di corpora di cui ci siamo occupati in questo contributo¹⁴. D'altro canto però l'utilità di un corpus parallelo per un traduttore o una traduttrice risulta di più immediata comprensione rispetto ai corpora monolingue o bilingue paragonabile. Ignoreremo qui il caso più semplice ed evidente, ovvero l'utilizzo di un corpus parallelo per identificare equivalenti traduttivi scelti precedentemente da altri traduttori e traduttrici. Questo uso non si discosta troppo dall'utilizzo di un dizionario bilingue o di una memoria di traduzione e non richiede capacità analitiche specifiche (al di là dall'ovvia necessità di valutare i contesti e discernere il caso più simile a quello che ci si trova a tradurre).

Illustreremo invece brevemente un possibile uso dei corpora paralleli nel quadro di uno studio delle scelte tipiche dei traduttori e delle traduttrici e delle caratteristiche della lingua tradotta. Il corpus InterCorp "core" contiene testi letterari tradotti tra varie lingue, consultabili in modalità parallela o monolingue.

¹⁴ Fra i corpora paralleli consultabili online ricordiamo InterCorp (https://kontext.korpus.cz/first_form) e EPTIC (<https://corpora.dipintra.it/eptic/>), quest'ultimo contenente anche testi interpretati, oltre che testi tradotti.

Partendo dalla modalità monolingue, una ricerca delle espressioni “camera da letto” e “stanza da letto” nel corpus di testi originali italiani mostra una frequenza media normalizzata di 4,55¹⁵. La stessa ricerca all’interno del corpus monolingue paragonabile mostra una frequenza ottanta volte superiore (82,71). Questo sorprendente risultato potrebbe essere spiegato con il fatto che i testi presenti nel corpus semplicemente trattano di temi diversi (i testi italiani originali inclusi nel corpus fanno riferimento al concetto di camera/stanza da letto di meno rispetto ai tradotti). Un’altra spiegazione però potrebbe essere che un’espressione presente nel testo di partenza viene trasferita nel testo d’arrivo in modo semi-automatico, nonostante l’esistenza di forme alternative. L’analisi parallela del sotto-corpus di testi inglesi originali e traduzioni in italiano rivela che in 154 casi su 389 la parola “bedroom” viene tradotta con “camera da letto” o “stanza da letto”, con una frequenza normalizzata di 21,25. In moltissimi altri casi la parola non è tradotta perché ridondante, oppure è tradotta semplicemente con “stanza” o “camera”. L’osservazione delle concordanze parallele in cui la scelta è caduta su “camera da letto” o “stanza da letto”, suggerisce che la traduzione meno esplicita “stanza” o “camera” sarebbe stata ugualmente accettabile (e forse più naturale) anche in molti di questi casi, come nei tre esempi che seguono.

1. At first he had lots of friends coming to see him in the one time sumptuous **bedroom**, furnished by his third wife, who had left him ten years ago, with an Empire four-poster bed and gilded mirrors, Cupid with his head inside a bow. (Bellow, *The Adventures of Augie March*)

All’ inizio ebbe un sacco d’ amici che venivano a trovarlo nella **stanza da letto** un tempo sontuosa, ammobiliata dalla sua terza moglie, quella che l’aveva lasciato dieci anni prima, con un letto col baldacchino stile impero e specchi dorati, Cupido con la testa infilata dentro un arco. (*Le avventure di Augie March*, tr. V. Mantovani)

2. Back in his **bedroom** he finished dressing hurriedly, whistling tunelessly as he stooped to grease and comb his hair before the mirror inside his wardrobe. (McEwan, *Atonement*)

Tornato in **camera da letto**, finì di vestirsi in gran fretta, fischiettando stonato mentre si piegava per pettinarsi e impomatarsi i capelli davanti allo specchio interno dell’armadio. (*Espiazione*, tr. S. Basso)

¹⁵ La definizione è reperibile qui: <https://www.sketchengine.eu/documentation/average-reduced-frequency/>; in questo caso comunque non importa il valore assoluto ma il rapporto fra originale e tradotto.

3. I keep him upstairs in my mother's **bedroom**. (Rowling, *Harry Potter and the Order of the Phoenix*) Lo tengo di sopra, nella **stanza da letto** di mia madre. (*Harry Potter e l'ordine della fenice*, tr. B. Masini)

Pur senza alcuna pretesa di esaustività, questi esempi potrebbero essere interpretati come casi di interferenza dalla lingua di partenza, probabilmente influenzata anche da una preferenza dei traduttori e delle traduttrici per espressioni più complete, esplicite e *corrette* nella lingua d'arrivo. Il fatto che la frequenza normalizzata nelle traduzioni dall'inglese sia molto più bassa del valore osservato per i testi tradotti in generale (seppure molto più alta rispetto ai testi originali italiani) indurrebbe a pensare che questa seconda possibilità sia plausibile. La combinazione di un corpus monolingue paragonabile e di un corpus, o più sotto-corpora, paralleli apre la possibilità di studiare le specificità della traduzione come forma di comunicazione mediata secondo l'ipotesi proposta nei primi anni 1990 da Mona Baker (1993), valutando inoltre separatamente l'effetto dell'interferenza da lingue specifiche e l'operazione di meccanismi generali tipici della mediazione interlinguistica, come la preferenza per una maggiore formalità (Bernardini 2011).

5. Riflessioni finali

In questo contributo ho provato a riassumere i principi base e le motivazioni a sostegno di un uso consapevole dei corpora, in particolare da parte di professioniste/i delle lingue straniere come i traduttori e le traduttrici. L'idea che l'apprendimento delle competenze e capacità traduttive sia innanzitutto un'attività pratica, che consiste nell'acquisire esperienza in ambiti il più possibile simili a quelli professionali, è attraente ma a ben vedere semplicistica. Scopo di un corso universitario per traduttrici e traduttori non è (solo) formare professioniste/i immediatamente in grado di fornire un servizio linguistico adeguato. In un mondo in rapidissima evoluzione, in cui l'intelligenza artificiale erode poco a poco spazi e attività tipicamente umani, aprendo allo stesso tempo nuove prospettive professionali, è necessario innanzitutto affinare le proprie capacità di apprendere e sviluppare la flessibilità necessaria a riadattare le proprie competenze e

capacità in contesti nuovi. Questa necessità emerge chiaramente dal più recente schema delle competenze dello European Master's in Translation (EMT) Network, con la sua enfasi sulle capacità interpersonali, fra cui la capacità di autovalutare, aggiornare e sviluppare competenze e capacità attraverso strategie personali e forme di apprendimento collaborativo¹⁶. È inoltre costantemente presente nelle nostre interlocuzioni con le parti sociali: i datori di lavoro sottolineano infatti i rischi insiti nel formare persone in grado di svolgere perfettamente un compito ben definito, ma incapaci di andare oltre, di agire sui processi migliorandoli attraverso soluzioni innovative.

Nel valutare il ruolo della *Corpus Linguistics*, e più in generale delle tecnologie linguistiche, nel processo di formazione delle traduttrici e dei traduttori è dunque essenziale ribaltare la prospettiva: la complessità cognitiva insita nelle operazioni di analisi di un corpus (dalla definizione delle ipotesi di ricerca, ai metodi per investigarle, all'interpretazione dei risultati) è il valore principale di questa disciplina, non un ostacolo. Come sottolineato da Mitchell-Schuitevoerder (2014: 241), lo scopo dell'insegnamento universitario non può essere quello di insegnare a usare uno strumento, per quanto complesso: per questo esistono i manuali utente degli sviluppatori. Il punto è invece incoraggiare, tramite l'uso delle tecnologie, il pensiero critico e analitico, perché "the enhancement of critical thinking during their studies will prepare students to make well-founded decisions and choices in their [...] careers". Carriere che saranno composite, variegata e imprevedibili, ma nelle quali di certo non verrà mai meno la necessità di risolvere i problemi e di continuare ad imparare.

¹⁶ "26. Continuously self-evaluate, update and develop competences and skills through personal strategies and collaborative learning", https://ec.europa.eu/info/sites/info/files/emt_competence_fwtk_2017_en_web.pdf.

Bibliografia

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

Baroni, M., S. Bernardini, A. Ferraresi e E. Zanchetta (2009) "The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora", *Language Resources and Evaluation* 43(3): 209–226.

de Beaugrande, R.-A e U. Dressler (1981) *Introduction to Text Linguistics*, London and New York: Longman.

Bednarek, M. e H. Caple (2017) *The Discourse of News Values: How News Organisations Create Newsworthiness*, Oxford: Oxford University Press.

Bernardini, S. (2000) "Systematising Serendipity: Proposals for Concordancing Large Corpora with Language Learners", in L. Burnard e T. McEnery (eds.) *Rethinking Language Pedagogy from a Corpus Perspective*, Frankfurt am Main: Peter Lang, 225–234.

Bernardini, S. (2011) "Monolingual Comparable Corpora and Parallel Corpora in the Search for Features of Translated Language", *Synaps* 26, <https://openaccess.nhh.no/nhh-xmlui/handle/11250/2393975>.

----- e M. Russo (2017) "Corpus Linguistics, Translation and Interpreting", in K. Malmkjær (ed.) *The Routledge Handbook of Translation Studies and Linguistics*, London: Routledge, 342–356.

----- e D. Kenny (2020) "Corpora", in M. Baker e G. Saldanha (ed.) *Routledge Encyclopedia of Translation Studies*, 3rd edition, London: Routledge, 110–115.

Biber, D. (2012) "Register as a Predictor of Linguistic Variation", *Corpus Linguistics and Linguistic Theory* 8(1): 9–37.

Connor, U. e T. A. Upton (ed.) (2004) *Discourse in the Professions. Perspectives from Corpus Linguistics*, Amsterdam: Benjamins.

- Crawford, W. e E. Csomay (2016) *Doing Corpus Linguistics*, London: Routledge.
- Durrant, P. e N. Schmitt (2009) "To what Extent do Native and Non-native Writers Make Use of Collocations?", *International Review of Applied Linguistics* 47: 157–177.
- Ellis, N. e R. Simpson-Vlach (2009) "Formulaic Language in Native Speakers: Triangulating Psycholinguistics, Corpus Linguistics, and Education", *Corpus Linguistics and Linguistic Theory* 5(1): 61–78.
- Evert, S. (2009) "Corpora and Collocations", in A. Lüdeling e M. Kytö (eds.) *Corpus Linguistics. An International Handbook*, Berlin: Mouton de Gruyter, 1212–1248.
- Fillmore, C. (1982) "Frame Semantics", in The Linguistic Society of Korea (ed.) *Linguistics in the Morning Calm*, Seoul: Hanshin Publishing, 111–137.
- Firth, J.R. (1935 [1964]) "On Sociological Linguistics", extracted from J. R. Firth, *The Technique of Semantics – Transactions of the Royal Society*, reprinted in D. Hymes (ed.) *Language in Culture and Society*, New York: Harper International, 66–70.
- Firth, J.R. (1952/3 [1968]) "Linguistic Analysis as a Study of Meaning", in F. R. Palmer (ed.) 12-26.
- Firth, J.R. (1956 [1968]) "Descriptive Linguistics and the Study of English", in F.R. Palmer (ed.) 96-113.
- Firth, J.R. (1957) *Papers in Linguistics 1934-1951*, London: Oxford University Press.
- Francis, W. N. e H. Kučera (1964) *Manual of Information to Accompany A Standard Corpus of Present-Day Edited American English, for Use with Digital Computers*, Providence (RI): Brown University.
- Goldberg, A. (1995) *Constructions. A Construction Grammar Approach to Argument Structure*, Chicago: Chicago University Press.

Granger, S. e Y. Bestgen (2014) "The Use of Collocations by Intermediate vs. Advanced Non-native Writers: A Bigram-based Study", *International Review of Applied Linguistics* 52: 229–252.

-----, G. Gilquin e F. Meunier, (eds.) (2015) *The Cambridge Handbook of Learner Corpus Research*, Cambridge: Cambridge University Press.

Gutt, E. A. (1991) *Translation and Relevance: Cognition and Context*, Oxford: Basil Blackwell.

Halliday, M. A. K. (1978) *Language as Social Semiotics*, London: Edward Arnold.

Hoey, M. (2005) *Lexical Priming: A New Theory of Words and Language*, London: Routledge.

Holmes, James S. (1972) "The Name and Nature of Translation Studies", in J.S. Holmes (ed.) *Translated! Papers on Literary Translation and Translation Studies*, Amsterdam: Rodopi. Reprinted in L. Venuti (ed.) (2000) *The Translation Studies Reader*, London: Routledge, 172–185.

Hundt, M. (2020) "Corpus-based approaches to World Englishes", in D. Schreier, M. Hundt e E. Schneider (eds.) *The Cambridge Handbook of World Englishes*, Cambridge: Cambridge University Press, 506–533.

Hymes, D. (1972) "On Communicative Competence" in J. B. Pride, e J. Holmes (eds.) *Sociolinguistics*, London: Penguin, 269–293.

Iwasaki, S. (2015) "A Multiple-Grammar Model of Speakers' Linguistic Knowledge", *Cognitive Linguistics* 26(2): 161–210.

Jones, C. e D. Waller (2015) *Corpus Linguistics for Grammar*, London: Routledge.

Langacker, R. (1987) *Foundations of Cognitive Grammar Volume 1. Theoretical prerequisites*, Stanford (CA): Stanford University Press.

Laviosa, S. (2013) "Corpus Linguistics in Translation Studies" in C. Millán e F. Bartrina (eds.) *The Routledge Handbook of Translation Studies*, London: Routledge, 228–240.

McEnery, T. e A. Hardie (2012) *Corpus Linguistics*, Cambridge: Cambridge University Press.

Mikhailov, M. e R. Cooper (2016) *Corpus Linguistics for Translation and Contrastive Studies*, London: Routledge.

Mitchell-Schuitevoerder, R. (2014) *A Project-Based Syllabus Design Innovative Pedagogy in Translation Studies*, Durham theses, Durham University, <http://etheses.dur.ac.uk/10830/>.

Palmer, H. E. (1933) *Second Interim Report on English Collocations*, Tokyo: Institute for Research in English Teaching.

Palmer, F.R. (ed.) (1968) *Selected Papers of J.R. Firth 1952-59*, London and Harlow: Longman.

Partington, A. (2017) "Evaluative Clash, Evaluative Cohesion and how we Actually Read Evaluation in Texts", *Journal of Pragmatics* 117: 190–203.

Prodromou, L. (2008) *English as a Lingua Franca: A Corpus-based Analysis*, London: Continuum.

Rychlý, P. (2008) "A Lexicographer-Friendly Association Score" in P. Sojka and A. Horák (eds.) *Proceedings of Recent Advances in Slavonic Natural Language Processing RASLAN 2008*, Brno: Masaryk University, 6–9.

Sangati, F., A. Pascucci e J. Monti (2018) "Exploiting Multiword Expressions to solve 'La Ghigliottina'" in T. Caselli, N. Novielli, V. Patti *et al.* (eds.) *EVALITA Evaluation of NLP and Speech Tools for Italian Proceedings of the Final Workshop 12–13 December 2018*, Torino: Academia University Press, 256–261.

de Saussure, F. (1916 [1971]) *Cours de Linguistique Générale*, C. Bally e A. Sechehaye (eds.) Parigi: Payot.

Siyanova-Chanturia, A. (2015) "On the 'Holistic' Nature of Formulaic Language", *Corpus Linguistics and Linguistic Theory* 11(2): 285–301.

Sinclair, J. McH. (1991) *Corpus, Concordance, Collocation*, Oxford: Oxford University Press.

----- (2004) *Trust the Text*, London: Routledge.

Stubbs, M. (1996) *Text and Corpus Analysis*, London: Blackwell.

Tognini Bonelli, E. (2001) *Corpus Linguistics at Work*, Amsterdam: Benjamins.

Wray, A. (2002) *Formulaic Language and the Lexicon*, Cambridge: Cambridge University Press.

Wynne, M. (2010) "Interdisciplinary relationships", *International Journal of Corpus Linguistics* 15(3): 425–427.