

Eleonora Gallitelli
Università IULM

Digital Humanities come risorsa per i *translation studies*

Abstract

In Anglo-Saxon countries the digital humanities (known as “informatica umanistica” or “umanistica digitale” in Italian) enjoy a good reputation among scholars and researchers of linguistics and translation studies. In Italy, on the contrary, this area of research is still looked on with some suspicion, especially when its methodologies are applied to literary texts, this despite the fact that software-based corpus linguistics studies have often given interesting results indicating trends of various kinds.

This paper presents as an example one of the many possible applications of digital humanities, a semi-automatic analysis based on a corpus of Italian novels and novels translated into Italian from English. The study sought to establish whether, in different periods of recent Italian history (from national unification to the present day), “translated literature” (Even-Zohar [1978] 1990) has shown peculiar and characteristic traits when compared to fictional texts written directly in Italian. A short description of the main phases of this study will be followed by a reflection on the perspectives opened by digital humanities, which can be seen as a useful tool to get to the heart of language change and dismantle the most enduring stereotypes regarding language and translation.

1. *Digital Humanities: lo stato dell'arte in Italia e nel mondo*

Le *Digital Humanities* si presentano come un contenitore ampio ed eterogeneo che accoglie diverse aree di ricerca e interessa vari media. Si tratta di un campo di studi relativamente giovane, inestricabilmente legato allo sviluppo delle tecnologie informatiche. Dopo il secondo conflitto mondiale, le intuizioni ingegneristiche di Vannevar Bush e poi di Douglas Engelbart spianarono la strada alla futura disciplina, da subito guardata con sospetto e allo stesso tempo con interesse dall'accademia per la conservazione e il rinnovamento degli studi umanistici. Persino il celebre detrattore Stanley Fish, che definisce sbrigativamente l'appellativo «an umbrella term for new and fast-moving developments across a range of topics [...] in short, everything», prevede che l'applicazione

dell'informatica alle scienze umane sarà «the next thing in literary studies» (Fish 2011)¹.

Secondo alcuni, la disciplina avrebbe mosso i primi passi proprio in Italia, con il vicentino Padre Roberto Busa, un gesuita, che nel 1946, per la sua tesi di laurea, ebbe l'idea di sottoporre a un'analisi linguistica il *corpus* di opere di Tommaso D'Aquino giovandosi dell'ausilio del computer. Grazie alla collaborazione con la IBM e dopo varie fasi di rielaborazione, nel 2005 l'opera monumentale da lui avviata venne infine realizzata compiutamente². Agli albori dell'informatica umanistica troviamo dunque uno studio riconducibile alla metodologia che oggi chiamiamo *corpus linguistics*.

Questo primato italiano, tuttavia, è stato presto scalzato dagli studi di area anglo-americana, che hanno indirizzato la ricerca verso il perfezionamento della linguistica computazionale e la creazione di applicazioni per lo studio e la didattica. Le esperienze a tutt'oggi in corso vedono il “digital humanist movement” trainato da Stati Uniti, Regno Unito e Canada, che accolgono sul proprio territorio rispettivamente quarantaquattro, quattordici e undici centri dedicati alle DH su un totale di 114 istituti ospitati da 24 stati nel mondo. Negli ultimi anni l'area umanistica dell'università americana (National Endowment for the Humanities) ha deciso di puntare sulla disciplina, sostenendo dal 2007 al 2011, per citare un solo dato, duecentocinquanta progetti per un totale di \$15,268,130. In Europa la ricerca in questo campo si concentra in tre paesi, escludendo il Regno Unito: Francia, Germania e Italia³.

Arriviamo così all'università italiana. Anche da noi le DH conservano la loro natura pluridisciplinare. Al suo interno convivono insegnamenti dai nomi oscuri, spesso anglicizzanti o direttamente inglesi. Compulsando il sito web dell'università di Genova, che alla disciplina ha dedicato un corso di laurea magistrale, per esempio, si ritrovano denominazioni quali: Ergonomia dell'Interaction Design, Nuovi Media, Fotografia Applicata, Grafica, Reti di Comunicazione e Multimedia, Web Design, ma anche Devianza e Controllo

¹ Per una panoramica sulla *corpus linguistics* si rimanda a McEnery and Wilson.

² È consultabile sul web all'indirizzo <<http://www.corpusthomicum.org/it/>>.

³ Melissa Terras, *Quantifying Digital Humanities*, disponibile al link: <<http://www.ucl.ac.uk/infostudies/melissa-terras/DigitalHumanitiesInfographic.pdf>> (cit. in Monella 7).

Digitale, Future Internet, Interazione Uomo-Macchina e Realtà Aumentata, Interfacce Multimodali e Adattive. Nomi cifrati pressoché inaccessibili ai non iniziati, che rimandano a un mondo più visuale che scritto, e rigorosamente multimediale. Anche Ca' Foscari di Venezia propone un master specifico sull'elaborazione e lo sviluppo di strumenti informatici nell'ambito delle scienze umane. In questo caso le denominazioni dei moduli sono tutte in inglese; alla base dei corsi ci sono le scienze sociali (Quantitative Methods for The Social Sciences), l'archivistica (Historical Data Analysis and Archival Sciences, Digital Library Fundamentals) e complesse metodologie di ricerca (Data Mining and Pattern Recognition) che trovano le più svariate applicazioni (Digital History, Digital Art History, Digital Philology, Digital Epigraphy, Digital Film, etc.). In nessun caso si fa cenno agli studi letterari o ai *translation studies*, due ambiti in cui, in Italia, le DH stentano a decollare.

Eppure, sin dai primi anni Novanta l'università La Sapienza di Roma dispone di un importante centro di studi sulle DH, istituito dal professor Tito Orlandi, tra i pionieri a livello internazionale della disciplina: il CISADU (Centro Interdipartimentale di Servizi per l'Automazione nelle Discipline Umanistiche). Alla fine dello stesso decennio, sempre a Roma, Giuseppe Gigliozzi ha istituito il CRILET (Centro Ricerche Informatica e Letteratura), veicolando la "dura" informatica nel campo degli studi letterari. Gigliozzi e altri giovani ricercatori, tra cui Fabio Ciotti, avviarono in quella sede un lavoro che ha condotto alla digitalizzazione e codifica dei testi letterari e all'applicazione di metodi computazionali all'analisi dei testi (Ciotti 2014). Grazie alla loro iniziativa e a quella di Pasquale Stoppelli, che negli stessi anni creò LIZ (Letteratura Italiana Zanichelli) – un *corpus* di testi di letteratura italiana allora disponibili solo su CD-ROM, oggi consultabili anche sul sito di BibIT (Biblioteca Italiana) – anche l'Italia rende disponibili a lettori, critici e studiosi nazionali e internazionali testi in formato elettronico. Come si spiega allora l'esiguità degli studi di informatica digitale in campo linguistico-letterario?

2. Corpus linguistics *per un ritorno alla centralità del testo (elettronico)*

David Hoover, professore di anglistica presso la New York University, tenta di sciogliere quest'enigma ricercando le cause del relativo disinteresse verso

quest'area di studi negli Stati Uniti, dove pure, come si è detto, le DH sono tenute in massimo conto da un punto di vista culturale (e di conseguenza anche economico). A suo modo di vedere, negli ultimi decenni, quando l'ampia disponibilità di testi in formato elettronico ha reso più facile l'impiego degli strumenti di analisi tradizionali e insieme ha incoraggiato lo sviluppo di strumenti informatici, l'analisi del testo è passata in secondo piano. Colpevoli di tale marginalizzazione sarebbero l'ostilità da parte della linguistica chomskiana e di teorie letterarie influenzate da Jacques Derrida come quella di Jerome McGann, che rivendica il carattere "fondamentalmente soggettivo" della critica (50–51), screditando l'analisi "oggettiva" dei dati quantitativi. L'approccio formalistico e innatista (nonché il rifiuto dei *corpora*) dell'una e la posizione decostruzionista delle altre rispetto all'instabilità del segno e alla tendenziale disintegrazione dei testi sotto la lente della critica tenderebbero in definitiva a ridimensionare l'importanza del testo negli studi letterari.

Dal canto suo, Hoover auspica un ritorno al testo, «specifically the electronic, computable text, to see what scholarly digital editions, corpora, text analysis, statistical stylistics, and text alteration can reveal about meanings and style» (Hoover 2007). Con Sinclair (2004) e in opposizione a Chomsky, McGann e Fish, Hoover ritiene che i *corpora* offrano norme e riferimenti utili all'analisi stilistica e rappresentino una base ben più solida dell'introspezione per gli studi di linguistica, rivelando tratti semantici e sintattici altrimenti difficilmente identificabili.

A partire dai primi anni Novanta un nuovo impulso all'analisi dei testi è giunto dalla diffusione della *corpus linguistics* (in italiano: linguistica dei *corpora*) nei *translation studies*, un campo, com'è ampiamente riconosciuto, tra i più attivi e prolifici negli studi letterari e, più in generale culturali (Susan Bassnett parla di un «translational turn» che ad oggi interesserebbe trasversalmente varie discipline umanistiche – 123-40).

Il primo nome che è d'uopo citare, a questo proposito, è quello di Mona Baker, apripista nell'impiego di *corpora* testuali per spiegare i fenomeni traduttivi, seguita a breve passo da teorici come Gideon Toury, Kirsten Malmkjaer e Miriam Shlesinger.

Baker distingue tre tipi di *corpora* applicabili alla traduzione: paralleli, comparabili e multilingue. I primi, tra i più diffusi nel campo dei *translation studies*,

raccolgono testi originali e loro traduzioni, e permettono di fare luce su singole prassi traduttive poste a confronto. I *corpora* multilingue sono composti da due o più *corpora* monolingue in lingue diverse, e possono essere impiegati negli studi di linguistica contrastiva o di didattica delle lingue. I *corpora* comparabili affiancano invece due diversi gruppi di testi prodotti nella stessa lingua, consistenti l'uno in testi originali, l'altro in traduzioni eseguite nella lingua degli originali, di cui si possono confrontare alcuni tratti linguistici. Quest'ultima tipologia di *corpora* è stata spesso impiegata nelle ricerche sugli universali traduttivi (Olohan 68).

Rispetto al loro ambito applicativo, gli studi di traduttologia basati su *corpora* si possono suddividere in due macroaree: studi teorici e studi pratici. Nei primi, i *corpora* sono impiegati principalmente per lo studio dei processi traduttivi, l'analisi delle modalità in cui avviene il trasferimento di un'idea da una lingua all'altra o il confronto di tratti linguistici e relative frequenze in testi tradotti e originali. Negli studi pratici, invece, i *corpora* costituiscono un valido riferimento per i traduttori, nonché un punto di partenza per lo sviluppo di applicazioni quali i sistemi di traduzione automatica e assistita.

Oltre ad aver elaborato questa utile sistematizzazione, Baker è nota per aver diretto il primo studio applicativo di linguistica dei *corpora* in campo traduttivo, che a tutt'oggi resta ineguagliato per ampiezza e intelligenza nella progettazione del *corpus*. La studiosa di origine egiziana ha elaborato presso lo University of Manchester Institute of Science and Technology (UMIST, Manchester) il Translational English Corpus (TEC), una raccolta di traduzioni contemporanee in lingua inglese consultabile gratuitamente attraverso un software realizzato *ad hoc*. Il *corpus*, che in totale conta circa dieci milioni di parole ed è sottoposto a continuo aggiornamento e ampliamento, consta al momento di testi in inglese tradotti da varie lingue sorgenti, europee e non, organizzati in quattro *subcorpora*: testi di narrativa, biografie, testi giornalistici e riviste di bordo. Anche il cosiddetto *mark-up* è stato eseguito con grande meticolosità e precisione dal team di Manchester: in un documento consultabile separatamente sono raccolte informazioni extralinguistiche relative a ciascun testo quali il genere, la nazionalità e lo stato occupazionale del traduttore, la direzione della traduzione, la lingua fonte, l'editore della traduzione, etc. Questo *corpus* di riferimento è alla base di un'ampia varietà di studi, che convergono in

due tipologie: studi sulle differenze fra testi tradotti e testi scritti direttamente in lingua inglese; studi sulle variazioni stilistiche tra singoli traduttori⁴.

3. *Dal generale al particolare: un caso di corpus linguistics applicata ai translation studies*

In Italia una sollecitazione all'impiego di *corpora* per l'analisi linguistico-letteraria è giunta già negli anni Sessanta dal linguista Tullio De Mauro, che in *Storia linguistica dell'Italia unita* esortava a quantificare e osservare nel vivo i cambiamenti della lingua, perché «i criteri di analisi sistematica e quantitativa fatti valere dalla più recente linguistica generale sono indispensabili per valutare nella sua concretezza e nella sua storicità la portata effettiva di una innovazione» (187)⁵. Più recentemente anche Michele Cortelazzo si è espresso in favore della metodologia della *corpus analysis*, giudicandola «fondamentale per poter disporre di conoscenze sicure sull'italiano contemporaneo, più chiare e conformi alla realtà e, al tempo stesso, scevre dagli abbagli che ci possono venire da false percezioni impressionistiche della realtà linguistica» (2010, xiv).

Proprio da percezioni impressionistiche pubblicate a mezzo stampa o sul web sono partita io stessa per condurre uno studio pilota di linguistica dei *cor-*

⁴ Cfr. Baker 1999 e Kenny. Per una bibliografia più completa si rimanda alla pagina del progetto sul sito della University of Manchester:

<<http://www.alc.manchester.ac.uk/subjects/tis/research/projects/tec/>>.

⁵ De Mauro riporta i risultati di una piccola ricerca da lui condotta su un corpus tanto esiguo quanto eterogeneo ed esorta a intraprendere ricerche statistiche (che lui definisce «“astrazioni” numeriche») su *corpora* più consistenti, per «intendere i fenomeni linguistici nella loro reale e concreta storicità» (215-16); Nadiani rimarca l'interesse scientifico di un approccio di questo tipo e applica procedure analitiche semi-automatiche allo studio contrastivo di due romanzi del tedesco Friedo Lampe e delle relative traduzioni di cui lui stesso è autore (31-57); l'italianista Giuseppe Gigliozzi, il primo a occuparsi di informatica umanistica in Italia, ha progettato e avviato l'elaborazione del software SEB (Sistema esperto analizzatore di brani) e poi se ne è servito, insieme al suo gruppo di studio, per l'analisi di romanzi di vari autori italiani, tra cui Verga e Pirandello; Lorenzo Renzi e Giampaolo Salvi hanno diretto il progetto Italant basato sul corpus delle scritture fiorentine dalle origini al Trecento, finalizzato alla compilazione di una grammatica dell'italiano antico; un'équipe di ricerca facente capo all'Università degli Studi di Padova ha recentemente condotto uno studio su morfosintassi e *corpora* informatici dell'italiano antico.

pora coordinato da Tim Parks nell'ambito del dottorato in Letterature comparate dell'Università IULM.

Come dimostrano le tante iniziative e pubblicazioni dedicate al tema (si veda, tra le più recenti, il volume 73 della storica rivista *Nuovi argomenti* [gennaio-marzo 2016], intitolato significativamente *Che lingua fa?*, che accoglie interventi di autorevoli linguisti quali De Mauro, Beccaria, Coletti, Cortelazzo, Serianni e altri, insieme a contributi di editor, traduttori, poeti e scrittori italiani come Camilleri, Lagioia e Trevi), nell'ultimo decennio è emersa una nuova e stringente "questione della lingua". È opinione diffusa che la lingua d'oggi sia più semplice e più povera rispetto al passato. C'è chi, come il comparatista Daniele Giglioli (2011), ritiene che la responsabilità di tale deriva sia da ricercarsi nell'influenza delle traduzioni e chi, come Roberto Carnero (2011), si sofferma sull'influsso nefasto esercitato da altri media sulla generazione «senza stile» degli scrittori under 40: «la televisione in casa e i ragazzi che si sono formati più sulla tv, sul cinema, sui fumetti, sulla musica leggera, rock e pop, che non sui libri». Sulla stessa linea, Gabriele Pedullà (2011) descrive l'italiano contemporaneo come «una simil-lingua, senza stile e senza musica, che ricorda piuttosto le cattive traduzioni dall'inglese».

Partendo da queste riflessioni, che si potrebbero forse considerare delle *idées reçues* sullo stato attuale della lingua letteraria, mi sono posta l'obiettivo di indagare sul rapporto reale tra italiano degli scrittori e italiano dei traduttori, limitando il campo di ricerca al macrogenere della narrativa, ma andando indietro nel tempo fino al periodo postunitario per seguire i cambiamenti della lingua sin dalla sua istituzionalizzazione sul territorio nazionale. L'ipotesi di partenza, sulla scorta delle posizioni riassunte più sopra, era duplice: si intendeva appurare se l'italiano dei traduttori abbia risentito del contatto con l'inglese dei testi fonte e insieme se l'italiano degli scrittori sia stato "contaminato" dal cosiddetto traduttese, che a sua volta ricalca certe strutture sintattiche dell'inglese. Se si pensa che, stando alle statistiche sull'editoria relative al 2000 (Peresson 2002), la narrativa tradotta copre circa il 70% delle copie stampate dalle case editrici italiane, di cui circa il 63,5% è rappresentato da opere tradotte dall'inglese, in primis dagli Stati Uniti, sembra probabile che vi sia qualche ricaduta (o, per essere più precisi, qualche "interferenza") sull'italiano degli scrittori nativi.

Sin dai primi passi l'impresa si è rivelata estremamente complessa. Già la fase successiva all'ideazione del progetto, ovvero la costruzione del *corpus*, che, concordiamo con Shen Guo-rong (182), rappresenta «the most tedious and important task», ha richiesto un notevole impegno per il reperimento e/o la conversione dei testi in formato elettronico. È qui che il sostegno tecnico dell'università si rivela prezioso. Benché attualmente l'università IULM non disponga di un laboratorio digitale, la generosa collaborazione offerta a livello individuale dal bibliotecario Andrea Montagner e dall'addetto ai servizi multimediali Ibrahima Mbengue è stata utilissima. Attraverso il sistema di riconoscimento ottico OCR abbiamo potuto digitalizzare alcuni romanzi di fine Ottocento non reperibili in formato elettronico, in modo che fossero analizzabili dal software selezionato per la ricerca, WordSmith Tools. Altri testi erano disponibili online in un formato compatibile con il software sui siti <www.liberliber.it> e <www.gutenberg.org>, mentre certi altri siti, come <http://www.bibliotecaitaliana.it/>, offrivano testi consultabili per singoli capitoli o non esportabili in un formato compatibile.

La tappa successiva era il «design del corpus» (Olohan 45-61). Il *corpus* approntato per l'analisi si estende diacronicamente dall'unificazione italiana alla contemporaneità coprendo diversi periodi della storia letteraria per consentire il confronto tra stili, strategie traduttive e specificità linguistiche di testi indigeni e tradotti. I testi scelti appartengono a tre fasi cronologiche assunte come coagulo di momenti dell'evoluzione linguistica italiana: anni 1870-1910 (quindi grosso modo dall'Unità alla prima guerra mondiale); anni 1930-1965 (dal “decennio delle traduzioni” al consolidamento dell'editoria di massa); anni 1989-2011 (la contemporaneità). Per ogni fase sono stati raggruppati due *subcorpora* di dieci romanzi italiani e dieci romanzi tradotti dall'inglese, per un totale di sei gruppi di testi integrali in formato elettronico (cfr. *Bibliografia* per i titoli del *corpus*). L'ampiezza complessiva del *corpus* è di 5.471.600 parole (*tokens*), che, trattandosi di uno studio pilota, è piuttosto consistente⁶.

Per avere un'ampia gamma di stili narrativi, la scelta dei titoli è caduta sia su romanzi di genere che su testi letterari, anche se gli autori più sperimentali

⁶ Altri importanti corpus di testi come il LIP, il lessico di frequenza dell'italiano parlato, e il CoLFIS, il corpus del lessico di frequenza dell'italiano scritto, per esempio, sono assai più limitati, contando rispettivamente 500.000 e 3.800.000 parole.

(come Gadda o Joyce) sono stati evitati. Anche per le traduzioni è stata operata una scelta simile (si è esclusa, per esempio, la traduzione arcaizzante di *Emma* di Jane Austen del 1951, eseguita da Mario Praz, per la sua eccessiva difformità – o “deviazione standard” – dalle altre traduzioni del periodo). Tutti i romanzi scelti sono stati pubblicati da grandi case editrici e molti dei loro autori sono ben noti (a titolo esemplificativo: Dickens, Hemingway, Agatha Christie tra i tradotti, Pirandello, Sciascia, Pavese, Baricco tra gli originali), per cui si possono ritenere a buon diritto “rappresentativi” dell’universo della narrativa italiana e anglofona pubblicata in Italia nei tre suddetti periodi.

A questo punto, era necessario elaborare i parametri di ricerca. Confrontandomi con supervisor e colleghi (Tim Parks innanzitutto, ma anche Francesco Laurenti, Edoardo Zuccato, Francesca Santulli, Paola Carbone), ho individuato le forme e i costrutti da ricercare in concreto nei testi. All’inizio si trattava di una ventina di parametri; poi, una volta raggiunta una chiara visione d’insieme, ho ristretto il focus su nove tratti linguistici riconducibili a tre macrocategorie:

- a) parametri generali (lunghezza media dei periodi, varietà lessicale, pronomi relativi);
- b) tratti peculiari o storici dell’italiano (forme lessicali enfatiche, congiuntivo, passato remoto e trapassato remoto);
- c) tratti suscettibili di interferenza dall’inglese (perifrasi progressiva, pronomi personali).

La fase successiva è stata senza dubbio la più “creativa”, per quanto questo aggettivo possa forse apparire fuori luogo nel contesto di un’analisi quantitativa semi-automatica di corpus testuali. Se, infatti, WordSmith Tools consente facilmente di conoscere la lunghezza media dei periodi – si tratta di una ricerca che richiede pochi secondi, una volta che si è acquisita una certa dimestichezza con il software – o la varietà lessicale dei *subcorpora*, quest’ultima una misurazione che si basa su calcoli complessi, fortunatamente eseguiti anche in questo caso dal software⁷, meno immediata è stata l’individuazione dei para-

⁷ In WordSmith Tools l’indice che misura la varietà lessicale è denominato *standardised type/token ratio* (STTR). Questo indice è ottenuto mediante il ricalcolo da zero ogni mille vocaboli del rapporto in termini percentuali tra il numero delle parole diverse (*type*) e il totale delle parole o occorrenze (*token*) di un dato corpus; successivamente il programma procede

metri per la ricerca dei pronomi relativi, delle forme lessicali enfatiche, delle forme verbali. Un'altra importante decisione riguardava la modalità di presentazione dei risultati, che in ultima istanza è caduta sui valori delle frequenze su 10.000 parole. Successivamente, bisognava capire quali parole o morfemi inserire nel motore di ricerca del programma.

Le soluzioni a cui si è pervenuti sono state le seguenti: per quantificare l'uso dei pronomi relativi nei sei *subcorpora* si è interrogato varie volte il software inserendo *cui* e *quale* preceduti da ciascuna delle preposizioni articolate (maschili, femminili, singolari e plurali), si è poi sommato il numero totale di occorrenze del relativo preceduto da preposizione e si è calcolato il valore su 10.000. Il conteggio è stato più semplice per le espressioni enfatiche, dove si è sommato il numero di occorrenze di *mica*, *meno male*, *senz'altro*, *magari* su 10.000.

Per il congiuntivo, altro parametro di difficile misurazione, si è optato per una ricerca delle occorrenze della prima e della terza persona del verbo *avere* all'imperfetto congiuntivo (*avessi*, *avessi*) e all'imperfetto indicativo (*avevo*, *aveva*). Il verbo *avere* è stato considerato in tutti i suoi valori: come imperfetto indicativo o congiuntivo per esprimere possesso o come ausiliare di altri verbi al trapassato prossimo indicativo e al trapassato congiuntivo. Come per l'indicativo, non si è operata una distinzione tra i diversi valori assunti dal congiuntivo nelle varie occorrenze.

Si è poi calcolato il rapporto tra occorrenze dell'indicativo e del congiuntivo per comprendere se, o in quale misura, nell'italiano letterario contemporaneo si riscontri un'effettiva erosione nell'uso del congiuntivo a vantaggio della forma concorrente dell'uso medio⁸. I risultati finali si riferiscono al numero di occorrenze dell'indicativo per ogni singolo uso del congiuntivo.

Per il passato remoto e il trapassato remoto si è partiti dalla percezione di un calo nell'uso di queste forme, spesso sostituite dal passato prossimo. Si sono ricercate le occorrenze delle forme *ebbe* e *fu*, che in questo caso sono state messe in rapporto con le concorrenti *ha avuto* ed *è stato*. Anche per questo parametro *essere* e *avere* sono stati considerati sia come ausiliari che come verbi

al calcolo della media di tali rapporti percentuali. Pertanto, più alto è il valore dell'indice, maggiore sarà la densità lessicale del *subcorpus*.

⁸ Per la distinzione tra italiano standard e italiano dell'uso medio si rimanda a Berruto.

autonomi, dal momento che il rapporto tra trapassato remoto e trapassato prossimo è analogo a quello tra passato remoto e passato prossimo.

Per quantificare l'uso della perifrasi progressiva nei quattro gruppi di testi si è sfruttata una particolare funzione di WordSmith Tools, che consente di ricercare le occorrenze di singoli morfemi o parti di parole. Si sono ricercate tutte le persone del presente del verbo *stare* seguite da **ando* e **endo* e, dopo aver sommato i dati, si è riportato il valore su 10.000.

I pronomi personali di terza persona sono stati ricercati sia nel loro insieme (*egli/esso; ella/essa; essi/esse*), ipotizzando una strategia di evitamento attuata da scrittori e traduttori per non dover scegliere tra le due forme concorrenti di diverso registro, sia isolando il caso particolare del pronome *esso* preceduto da preposizione.

Dopo questa fase di raccolta, elaborazione e categorizzazione, mi sono dedicata all'analisi dei risultati, che per molti parametri sono stati sorprendenti⁹. Attraverso il confronto dei dati relativi ai sei *subcorpora* comparabili ho potuto soffermarmi su singoli fenomeni linguistici, ma anche inquadrare ciascun fenomeno nel più vasto sistema linguistico italiano, per delineare le tendenze della lingua dei traduttori letterari e dei romanzieri italiani nell'ottica dei processi di standardizzazione e neostandardizzazione linguistica oggi in atto.

È emerso, per esempio, che la tradizionale propensione all'amplificazione e alla complessità sintattica, che si esprime anche nella lunghezza dei periodi, non solo sembra più contenuta nei testi tradotti (come già notato da Cortelazzo 2010, xv), ma è oggi meno accentuata anche nei testi letterari autoctoni. In effetti, il divario che in passato sussisteva per questo parametro fra traduzioni e testi indigeni, oggi sembra essersi appianato. A livello di vocabolario, gli indici TTR delineano una sostanziale uniformità nei sei gruppi di testi: la varietà lessicale resta più o meno costante nel tempo e anche molto omogenea nei due macrogruppi (originali e traduzioni). I risultati, dunque, sfatano il luogo comune dell'impovertimento linguistico: nell'ultima fase le traduzioni contemporanee sono lessicalmente più ricche rispetto alle due fasi precedenti e anche

⁹ Per conoscere in dettaglio i risultati della ricerca, cfr. Gallitelli, Laurenti e Parks. Una prima esposizione dei risultati dei quattro *subcorpora* novecenteschi si trova invece in Laurenti e Gallitelli 13-33.

rispetto agli originali italiani coevi, i quali, a loro volta, sono appena meno variegati sotto il profilo lessicale di quelli della fase a cavallo tra i due secoli, ma non di quelli della fase centrale.

Dai risultati si evince inoltre che gli scrittori italiani contemporanei fanno un uso più parco del congiuntivo rispetto al passato (un'occorrenza ogni quindici circa dell'indicativo contro una ogni cinque nella prima fase), come pure più contenuto è oggi il ricorso al congiuntivo da parte dei traduttori (si passa da un congiuntivo ogni tre indicativi a uno ogni dieci). Non sembra, però, che sia il contatto con l'inglese a far propendere per l'indicativo: tanto nella prima quanto nella terza fase si osservano frequenze del congiuntivo più alte nei testi tradotti che in quelli autoctoni, a testimonianza di una tendenza da parte dei traduttori a favorire l'uso del modo considerato più colto.

Una simile inclinazione si riscontra anche nell'uso del passato remoto e del trapassato remoto, per i quali negli originali italiani si osserva una discontinuità nella seconda fase; se è vero che in questo periodo si rileva un nuovo «orientamento verso una lingua “viva” e moderna a partire dal rapporto tra parlato e scrittura letteraria» (Testa 273), permangono comunque usi più formali come il passato remoto. Quanto alle traduzioni, si riscontra che, pur essendo toccate dal calo graduale nell'uso di questo tempo verbale, nella prima e nella terza fase esse presentano un maggior uso del passato remoto rispetto ai romanzi autoctoni, ulteriore indicazione di una strategia conservatrice e di una predilezione per un'idea tradizionale di “standard letterario”. Anche in questo caso, quindi, si nota che mentre nella fase centrale lo stile delle traduzioni si distaccava da quello dei romanzi autoctoni per un minor livello di formalità, nelle traduzioni contemporanee vi è un relativo recupero di un tempo verbale sempre meno usato negli originali pubblicati nella stessa fase. Ancora una volta vale la pena di ricordare che alcune case editrici insistono sull'uso del passato remoto nelle traduzioni anche laddove il traduttore aveva scelto di impiegare il passato prossimo. Non è raro, infatti, che in fase di revisione tutti i tempi verbali vengano modificati dal passato prossimo al passato remoto.

Diversa appare la tendenza d'uso della perifrasi progressiva. Il dato mostra in primo luogo come nel tempo il ricorso a questa forma sia aumentato in entrambi i tipi di testi, soprattutto nelle traduzioni. Lo scarto tra i due generi te-

stuali è netto nella seconda e nella terza fase, in cui la perifrasi è molto più utilizzata in traduzione.

A tale riguardo, il fatto che i traduttori contemporanei manifestino un atteggiamento conservatore solo in alcuni casi (per esempio nell'uso del congiuntivo e del passato remoto) e non in altri (appunto il progressivo) induce a una riflessione: sembrerebbe che una forma straniera entri con più facilità nelle traduzioni in presenza di un costrutto formalmente equivalente nel patrimonio grammaticale dell'italiano, anche se questo non ha una frequenza d'uso pari a quella del suo corrispettivo inglese, come dimostra il caso della perifrasi progressiva qui esaminato. È quanto afferma anche Cortelazzo quando osserva, *en passant*, che «è naturale che nelle traduzioni si trovi con maggiore frequenza che in testi originali un costrutto, disponibile nella lingua d'arrivo, che ha somiglianze con un costrutto ampiamente diffuso nella lingua di partenza» (2007, 1763n).

Inoltre, la mancata “difesa” del presente indicativo a favore del progressivo può indicare che la tendenza conservatrice dei traduttori dipende da certi cliché sull'italiano convenzionale o tradizionale. Congiuntivo e passato remoto vengono difesi in quanto la loro paventata scomparsa è molto discussa; per contro, ciò di cui non si discute – per esempio, l'aumento nella frequenza del progressivo – non è percepito come problema.

In generale, per tirare le fila del discorso, i dati emersi gettano una nuova luce d'un canto sulla svolta linguistica degli anni Trenta, in cui scrittori e traduttori (figure che allora si sovrapponevano sovente nella stessa persona) si liberano delle convenzioni dell'italiano letterario tardo-ottocentesco con scelte trasgressive individuali, non centralizzate; dall'altro, sulla “biforcazione linguistica” della contemporaneità, quando a un tentativo di ristandardizzazione o neostandardizzazione da parte degli scrittori si oppone la spinta alla conservazione linguistica dei traduttori letterari (a loro volta indirizzati dalla linea della casa editrice e dagli interventi ipercorrettistici dei revisori editoriali). Infatti, se le traduzioni contemporanee sembrano tendere a quella che Baker (1996) chiama «normalizzazione», non può dirsi altrettanto delle traduzioni prodotte nei decenni centrali del Novecento, che, secondo alcuni (Pavese, Ungaretti) dovevano fungere da motore e veicolo del cambiamento linguistico con le loro mescolanze tra lingua “media” e lingua della tradizione letteraria, presen-

tandosi come un terreno di sperimentazione ma anche come un laboratorio dello stile personale degli stessi scrittori-traduttori. Per la maggior parte dei parametri è emerso che nella prima e nell'ultima fase considerata (post-unificazione e contemporaneità), quelle cioè interessate rispettivamente dai processi di standardizzazione e neostandardizzazione linguistica, le traduzioni si configurano come un fattore di stabilità e di ancoraggio all'italiano della tradizione letteraria piuttosto che di innovazione.

Insomma, il ruolo delle traduzioni e i comportamenti traduttivi variano nelle tre fasi individuate a seconda del significato assunto dall'atto traduttivo nelle diverse congiunture storiche. Come nota ancora Cortelazzo, infatti, «se in una lingua, in un determinato momento, si contrappongono opzioni diverse per uno stesso costrutto, le scelte traduttive possono finire per far pendere la bilancia verso una di queste» (2010, xi). I dati raccolti in questo studio smentiscono le ipotesi universaliste formulate da Baker (1996) e obbligano a circostanziare i testi nel sistema e nell'epoca che li ha prodotti o riprodotti per cercare di dare una spiegazione ai mutamenti che tenga conto della «costellazione specifica» (Even-Zohar [1978] 1990, 46) del polisistema storico, linguistico e letterario italiano.

La conclusione porta lo studio qui illustrato, pertanto, è la seguente: la lingua delle traduzioni odierne, lungi dall'essere «livellata su un registro medio» (Coletti 49) è più formale e grammaticalmente “corretta” rispetto a quella dei romanzi italiani; semmai, quindi, appare livellata su un registro alto della lingua. I dati (maggior uso del congiuntivo e del passato remoto, etc.) mostrano con una certa coerenza che il repertorio di riferimento del sottosistema dell'italiano tradotto contemporaneo tende maggiormente alla “conservazione” linguistica rispetto a quello dei testi italiani autoctoni.

4 Idee per possibili applicazioni future della linguistica dei corpora

La linguistica dei *corpora*, nata in Italia ma qui raramente applicata all'indagine letteraria, benché impiegata altrove con grande efficacia sulla scia degli studi di Baker (1995, 1996, 2009) e Laviosa (1998), permette di affrontare secondo una prospettiva nuova il rapporto e lo scambio linguistico e culturale che si realizza attraverso il processo di traduzione, interrogando e “misurando” il cambiamen-

to linguistico indotto dal confronto di diverse modalità espressive e tracciando secondo parametri oggettivi l'evoluzione della lingua letteraria italiana.

Vorrei chiudere questa breve disamina sul potenziale delle DH, e in particolare della linguistica dei *corpora*, per gli studi letterari e di teoria della traduzione con alcune idee per un possibile sviluppo futuro della disciplina. Un primo obiettivo sarebbe che anche l'università italiana, sul modello del TEC dell'Università di Manchester, producesse un ampio *corpus* di traduzioni, non di dieci ma di cento milioni di parole, suddiviso per genere e (per quanto riguarda le traduzioni) per lingua di provenienza. Per avere dati precisi e significativi sarebbe necessario poi annotare il *corpus*, aggiungendo informazioni extra-testuali, come quelle relative alle categorie grammaticali o sintattiche. Questa operazione può essere svolta in maniera semi-automatica, inserendo *tag* o etichette al testo attraverso un software specifico e intervenendo manualmente allorché in fase di controllo si scorgano errori, imprecisioni o ambiguità.

Un secondo obiettivo sarebbe di utilizzare tale *corpus* per uno studio diacronico su vasta scala dell'italiano delle traduzioni letterarie basato sulla metodologia della linguistica dei *corpora*, paragonandolo successivamente all'italiano dei romanzi autoctoni. In questo modo sarebbe possibile, partendo da una base di dati significativa, confermare o smentire le tante ipotesi sulla differenza tra testi scritti in italiano e testi tradotti in italiano da altre lingue in ambito letterario. Sarebbe anche possibile appurare con maggior precisione se e in quale misura il rapporto tra italiano autoctono e italiano delle traduzioni sia cambiato nell'arco del secolo.

Il *corpus* potrebbe essere impiegato dalla comunità scientifica per integrare studi di lingua, letteratura e traduzione nei più diversi ambiti, dalla poesia, al campo delle traduzioni dal mondo ispanofono, francofono e così via. I dati forniti da tali analisi, abbinati a quelli della ricerca storica e archivistica, permetterebbero di uscire dalla genericità delle osservazioni finora fatte in materia, stabilendo con una diversa sicurezza il senso e la direzione di sviluppo della traduzione moderna. I risultati potrebbero sollevare questioni e produrre riflessioni di tipo comparativo fra diverse aree di interesse. Per esempio: le traduzioni dalla lingua spagnola sono tendenzialmente diverse da quelle dalla lingua inglese? È un fatto che si può dimostrare statisticamente? Oppure prevale una più generica

“lingua della traduzione”, una sorta di *koinè* dei traduttori, quale che sia la lingua d’origine?

Tra le tante potenzialità applicative vi sono anche l’analisi dello stile dei traduttori (si tratta, come già detto, della prima applicazione del corpus TEC – Baker) e degli scrittori (già Mahlberg ha impiegato il software WordSmith Tools per studi di *corpus stylistics* sulle opere di Charles Dickens e Jane Austen, adottando metodi e approcci della *corpus linguistics* per esplorare questioni di stilistica e critica letteraria – 2010, 292-98) o di entrambi gli stili posti a confronto (Nadiani ha svolto un’analisi di due romanzi del tedesco Friedo Lampe, individuando, attraverso l’impiego del software dell’IMS Corpus Workbench, quelle che Mahlberg chiama «local textual functions», per poi studiare la resa traduttiva dell’aggettivazione anomala e ipertrofica di Lampe in corpora paralleli costituiti da testi originali e traduzioni, con risultati notevoli – 31-57).

L’utilizzo di questi dati varrebbe anche a integrare studi critici, linguistici, filologici e antropologici. Si potrebbe attingere al *corpus* per studi di ecdotica sull’attribuzione dei testi antichi o di stilometria, o ancora per un’analisi semantica dei testi, (già avviata su singoli testi da Tim Parks per i testi di D.H. Lawrence e sulla trilogia di E. L. James – 2010 e 2013 –, e da Hoover sul romanzo *The Inheritors* di William Golding – Hoover 1999) considerando, per esempio, per tornare al campo letterario italiano, la semantica del coraggio e della paura durante l’era fascista. Il rapporto con quanto emerso dalle ricerche basate su questo vasto «corpus di traduzioni italiane» (CTI) permetterebbe di valutare se i mutamenti osservati nell’ordine diacronico siano in relazione o meno con i fatti che si registrano nelle scelte individuali ed editoriali del periodo storico corrispondente.

Bibliografia del corpus

Opere italiane originali

ABBA G.C. *Le rive della Bormida*. Milano: Tipografia E. Civelli e C., 1875.

CAPUANA L. *Giacinta*. Milano: G. Brigola, 1879.

VERGA G. *I Malavoglia*. Milano: F.lli Treves, 1881.

INVERNIZIO C. *La trovatella di Milano*. Milano: Carlo Barbini Editore, 1889.

- D'ANNUNZIO G. *Giovanni Episcopo*. Napoli: Pierro, 1892.
- SALGARI E. *Il Corsaro Nero*. Genova: A. Donath, 1899.
- PIRANDELLO L. *Il fu Mattia Pascal*. Roma: Nuova Antologia, 1904.
- SERAO M. *Il ventre di Napoli*. Napoli: Francesco Perrella Editore, 1906.
- PALAZZESCHI A. *Il codice Perelà*. Milano: Edizioni futuriste di Poesia, 1911.
- TOZZI F. *Con gli occhi chiusi*. Milano: F.lli Treves, 1919.
- DELEDDA G. *Il paese del vento*. Milano: F.lli Treves, 1931.
- BRANCATI V. *Don Giovanni in Sicilia*. Milano: Rizzoli, 1941.
- VITTORINI E. *Il garofano rosso*. Milano: Mondadori, 1948.
- PAVESE C. *Il carcere*. Torino: Einaudi, 1949.
- FENOGLIO B. *I ventitré giorni della città di Alba*. Torino: Einaudi, 1952.
- SILONE I. *Il segreto di Luca*. Milano: Mondadori, 1956.
- CALVINO I. *Il barone rampante*. Torino: Einaudi, 1957.
- MORANTE E. *L'isola di Arturo*. Torino: Einaudi, 1957.
- CASSOLA C. *La ragazza di Bube*. Torino: Einaudi, 1960.
- SCIASCIA L. *Il giorno della civetta*. Torino: Einaudi, 1961.
- BARICCO A. *Castelli di rabbia*. Milano: Rizzoli, 1991.
- DE CARLO A. *Arcodamore*. Milano: Bompiani, 1996.
- AMMANITI N. *Io non ho paura*. Torino: Einaudi, 2001.
- BENNI S. *Saltatempo*. Milano: Feltrinelli, 2001.
- SAVIANO R. *Gomorra*. Milano: Mondadori, 2006.
- VOLO F. *Un posto nel mondo*. Milano: Mondadori, 2006.
- CAMILLERI A. *La pista di sabbia*. Palermo: Sellerio, 2007.
- WU MING. *Manituana*. Torino: Einaudi, 2007.

FALETTI G. *Io sono Dio*. Milano: Baldini Castoldi Dalai, 2009.

SCARPA T. *Corpo*. Torino: Einaudi, 2011.

Traduzioni dall'inglese¹⁰

CARROLL L. *Alice nel paese delle meraviglie*. Trad. it. di T. Pietrocòla-Rossetti. Torino: Loescher (1865/1872).

RADCLIFFE A. *I misteri del castello d'Udolfo*. Trad. anon. Milano: Oreste Ferrario (1794/1875).

DICKENS C. *La piccola Dorrit*. 3 voll. Trad. it. di F. Verdinois. Milano: F.lli Treves (1855-1857/1879).

BEECHER H. Stowe. *La capanna dello zio Tom*. Trad. anon. Firenze: Salani (1852/1891).

BRONTË C. *Jane Eyre*. Trad. anon. Milano: F.lli Treves (1847/1904).

WILDE O. *Il dovere del delitto*. Trad. it. di Federigo Verdinois. Napoli: Società Editrice Partenopea (1887/1908).

GOLSWORTHY A. *Un grido nella notte*. Trad. anon. Firenze: Adriano Salani Editore (1899/1909).

LOCKE W.J. *Idoli*. Trad. it. di E. Vecchi. Milano: F.lli Treves (1905/1911).

DICKENS C. *Davide Copperfield*. Trad. it. di S. Spaventa Filippi. Milano: Istituto Editoriale Italiano (1849-1850/1914).

MAY ALCOTT L. *Piccole donne*. Trad. it. di A. Mazzoni. Firenze: Bemporad (1868-1869/1915).

HUXLEY A. *Il mondo nuovo*. Trad. it. di L. Gigli. Milano: Mondadori (1932/1933).

¹⁰ In parentesi sono indicate le date dell'edizione originale e di quella italiana utilizzata per l'analisi quantitativa.

CHRISTIE A. *Assassinio sull'Orient Express*. Trad. it. di A. Pitta. Milano: Mondadori (1934/1935).

STEINBECK J. *Furore*. Trad. it. di C. Coardi. Milano: Bompiani (1939/1940).

HEMINGWAY E. *Per chi suona la campana*. Trad. it. di M. Napolitano Martone. Milano: Mondadori (1940/1945).

ORWELL G. *1984*. Trad. it. di G. Baldini. Milano: Mondadori (1949/1950).

GREENE G. *Il nostro agente all'Avana*. Trad. it. di B. Oddera. Milano: Mondadori (1958/1959).

KEROUAC J. *Sulla strada*. Trad. it. di F. Pivano. Milano: Mondadori (1957/1959).

SALINGER J.D. *Il giovane Holden*. Trad. it. di A. Motti. Torino: Einaudi (1951/1961).

DICK P.K. *I simulacri*. Trad. it. di R. Minelli. Piacenza: La Tribuna (1964/1965).

WELLS H.G. *La guerra dei mondi*. Trad. it. di A. Motti. Milano: Mursia (1898/1965).

RUSHDIE S. *I versi satanici*. Trad. it. di E. Capriolo. Milano: Mondadori (1988/1989).

MCCARTHY C. *Cavalli selvaggi*. Trad. it. di I. Legati. Torino: Einaudi (1992/1993).

NAIPAUL V.S. *Una via nel mondo*. Trad. it. di M. Dallatorre. Milano: Adelphi (1994/1995).

FOLLETT K. *Il terzo gemello*. Trad. it. di A. Raffo. Milano: Mondadori (1996/1996).

CORNWELL P. *Morte innaturale*. Trad. it. di R. Pera. Milano: Mondadori (1997/1998).

DELILLO D. *Underworld*. Trad. it. di D. Vezzoli. Milano: Mondadori (1997/1999).

MCEWAN I. *Espiazione*. Trad. it. di S. Basso. Torino: Einaudi (2001/2002).

BROWN D. *Il codice Da Vinci*. Trad. it. di R. Valla. Milano: Mondadori (2003/2003).

AUSTER P. *Viaggi nello scriptorium*. Trad. it. di M. Bocchiola. Torino: Einaudi (2007/2007).

SMITH W. *La legge del deserto*. Trad. it. di G. Hirzer. Milano: Longanesi (2011/2011).