

Stefano Ondelli
Floriana C. Sciumbata

Capitoli sul traduttese



EUT

Lingua italiana e società

6

Secondo la teoria degli universali traduttivi, tutte le traduzioni presenterebbero delle caratteristiche linguistiche che le differenziano dai testi scritti direttamente in una determinata lingua. Oltre all'interferenza del testo fonte, sarebbe il processo traduttivo stesso a spingere a i traduttori a comportamenti specifici e costanti, che si concretizzano in tratti linguistici non immediatamente percepibili dai lettori e dalle lettrici, bensì rilevabili tramite l'analisi con strumenti automatici delle tendenze di fondo di grandi quantità di testi. Ma quali sono le caratteristiche del traduttese in lingua italiana a livello di lessico, morfosintassi e organizzazione testuale? È possibile distinguere nettamente le traduzioni dai testi scritti direttamente in italiano? E, tra i vari fattori in gioco (la penna dell'autore, lo stile del traduttore, le caratteristiche della lingua di partenza), quale risulta dominante nel determinare l'assetto linguistico delle traduzioni? I tre capitoli in cui consiste questo volume cercano di fornire delle risposte a queste domande: dopo aver illustrato la teoria degli universali traduttivi e alcuni principi e metodi della linguistica dei corpora e dell'analisi statistica dei dati testuali, ci si concentra sullo studio degli articoli giornalistici e della prosa letteraria e paraletteraria per valutare – anche nel corso del tempo – come e quanto l'italiano delle traduzioni risulti diverso da quello dei testi non tradotti.

STEFANO ONDELLI insegna linguistica italiana presso l'Università degli Studi di Trieste. Le sue ricerche si concentrano su diversi aspetti dell'italiano contemporaneo: variazione sociolinguistica, strutture morfosintattiche, tipi e generi testuali, contatto linguistico. Si è occupato di italiano L2, della produzione linguistica di interpreti e traduttori, di numerose varietà specialistiche e dei problemi della scrittura professionale, in particolare adottando i metodi della linguistica dei corpora e dell'analisi statistica dei dati testuali.

FLORIANA C. SCIUMBATA è assegnista di ricerca in Linguistica italiana presso il Dipartimento di Scienze Giuridiche, del Linguaggio, dell'Interpretazione e della Traduzione dell'Università degli Studi di Trieste. Le sue ricerche riguardano la scrittura e la semplificazione linguistica (con particolare attenzione al linguaggio facile da leggere e da capire), la linguistica dei corpora, l'applicazione di metodi quantitativi e automatici di estrazione dei dati, la realizzazione di strumenti informatici per l'analisi linguistica, i metodi di *distant reading* e le *digital humanities*.



Lingua italiana e società

Finalità e ambito

La collana *Lingua italiana e società* offre una serie di studi puntuali che riguardano il ruolo dell'italiano in diversi momenti della vita del cittadino nella odierna società della comunicazione, per esempio la comunicazione istituzionale e le difficoltà che pone all'utente medio; la gestione dei rapporti sociali (il rispetto delle donne e delle minoranze); i problemi del mondo dell'informazione (fake news e violenza verbale); la comunicazione politica, soprattutto tramite i social network ecc. I volumi, di dimensioni contenute, adottano una lingua efficace e diretta con taglio divulgativo. Oltre all'analisi di fenomeni precisi viene offerta una panoramica degli studi più recenti tramite bibliografie ragionate. La collana si rivolge a un pubblico di lettori non esperti ma interessati ai fatti linguistici e curiosi di indagare il ruolo svolto dall'italiano nella società contemporanea, pur non escludendo qualche tuffo all'indietro nella storia della nostra lingua.

Garanzia scientifica

Fermo restando l'intento eminentemente divulgativo della collana, la solidità scientifica delle pubblicazioni sarà garantita dal Comitato editoriale, comprendente studiosi internazionali qualificati che hanno una competenza negli ambiti della linguistica italiana pertinenti all'iniziativa. Qualora contributori esterni sottoponessero proposte di pubblicazione i cui argomenti eccedano gli ambiti di competenza dei componenti del Comitato editoriale, questi si riservano di contattare esperti esterni per avere un parere motivato e affidabile sulla qualità dei contributi.

Comitato editoriale

Federico Baricci (Università di Trieste), *Michele A. Cortelazzo* (Università di Padova), *Stefano Ondelli* (Direttore, Università di Trieste), *Fabio Romanini* (Università di Ferrara), *Roman Sosnowski* (Università di Cracovia), *Matteo Viale* (Università di Bologna).

Modalità di pubblicazione

La collana, in linea con le politiche editoriali di EUT e dell'Università degli Studi di Trieste, prevede la pubblicazione online nella modalità ad accesso aperto, allo scopo di rendere liberamente disponibile la comunicazione scientifica. Accanto a questa modalità sono poi previste le più consuete pubblicazioni a stampa a tiratura limitata e/o su ordinazione degli autori.

Lingua italiana e società

Aims and Scope

The series *Lingua italiana e società* includes a range of studies focussing on the role played by the Italian language in our modern society dominated by communication and the new media, e.g. plain language in the public administration and corporate communication; linguistic challenges to the social inclusion of groups such as women and minorities; recent developments in mass media communication (fake news and hate speech); political discourse, especially in social networks etc. In terms of their size, contents and language, the volumes of the series address readers who are not specialists but wish to gain insight into the use of Italian both today and in the past: the analysis of specific linguistic aspects is complemented with a detailed bibliography to help the reader know more about the topic under discussion.

Although the series is addressed to the general public, its scientific value is guaranteed by the Editorial Board members, whose qualifications include the range of topics dealt with in the volumes. If contributors wish to submit studies that exceed the competencies of the Editorial Board, external experts may be contacted to provide their qualified opinion on the scientific quality of the contributions being submitted.

Editorial Board

Federico Baricci (University of Trieste), *Michele A. Cortelazzo* (University of Padova), *Stefano Ondelli* (Director, University of Trieste), *Fabio Romanini* (University of Ferrara), *Roman Sosnowski* (University of Krakow), *Matteo Viale* (University of Bologna).

Publication procedure

In line with the publishing guidelines of EUT and the University of Trieste, the series is published online according to the Open Access policy to make scientific publications freely available. In addition to a certain number of paper copies distributed to the authors and the most important libraries, more copies may be ordered on demand.

Impaginazione
Elena Tonzar

© Copyright 2024 EUT

EUT Edizioni Università di Trieste
via Weiss 21, 34128 Trieste
eut@units.it
<https://eut.units.it>
<https://www.facebook.com/EUTEdizioniUniversitaTrieste>

Proprietà letteraria riservata.
I diritti di traduzione, memorizzazione elettronica,
di riproduzione e di adattamento totale e parziale
di questa pubblicazione, con qualsiasi mezzo
(compresi i microfilm, le fotocopie e altro)
sono riservati per tutti i paesi.

ISBN 978-88-5511-591-9 (print)
ISBN 978-88-5511-592-6 (online)

Capitoli sul traduttese

Stefano Ondelli

Floriana C. Sciumbata

Sommario

11 **Premessa**

15 **1. Le basi teoriche, i metodi e i materiali**

15 1.1. Il traduttese

16 1.1.1. Dalla fedeltà alla ricezione

18 1.1.2. Il livello linguistico

21 1.1.3. L'interferenza

24 1.1.4. Le regolarità del traduttese: gli universali

25 1.1.5. Termini di paragone

26 1.1.6. Problemi teorici e prospettive di ricerca

28 1.2. L'approccio quantitativo: la linguistica dei corpora

30 1.2.1. La ricchezza lessicale

31 1.2.2. La densità lessicale

32 1.2.3. La distanza intertestuale

33 1.3. I corpora esaminati

34 1.3.1. Il primo corpus giornalistico

38 1.3.2. Il secondo corpus giornalistico: bilanciamento e campionamenti

43 1.3.3. Il corpus letterario

57	2. Il traduttese nei giornali
57	2.1. Obiettivi e strumenti della ricerca
59	2.2. Il lessico
59	2.2.1. Le misure lessicometriche
63	2.2.2. Il vocabolario di base
64	2.2.3. Categorie del lessico
68	2.2.4. Forestierismi non adattati
76	2.2.5. Le forme plurali dei forestierismi: il caso della s
79	2.2.6. Parole italiane derivate o connesse con lingue straniere
81	2.2.7. Interferenza di lingue straniere nell'uso di alcune forme italiane
83	2.2.8. Espressioni politicamente corrette
87	2.3. Morfologia
90	2.3.1. Distribuzione di tempi e modi verbali
92	2.3.2. Perifrasi [<i>stare</i> + gerundio]
92	2.3.3. Frequenza di pronomi personali
95	2.3.4. Frequenza di pronomi e aggettivi possessivi
96	2.3.5. Frequenza di pronomi e aggettivi dimostrativi
97	2.3.6. Altri aspetti morfologici
97	2.4. Sintassi: lunghezza dei periodi e leggibilità
100	2.4.1. Costrutti <i>ad sensum</i>
102	2.4.2. Calchi sintattici
110	2.5. Testualità: l'uso dei connettivi
113	2.6. Trattamento del corpus per il calcolo della distanza intertestuale
115	2.6.1. Macrotesti tradotti e macrotesti non tradotti
124	2.6.2. L'effetto della lingua di partenza
132	2.7. Conclusioni
132	2.7.1. Le caratteristiche del traduttese nei giornali
135	2.7.2. Il traduttese alla prova della distanza intertestuale
137	3. Il traduttese nella narrativa
137	3.1. Obiettivi e strumenti della ricerca

139	3.2. Lessico
140	3.2.1. Misure lessicometriche
142	3.2.2. Il vocabolario di base
142	3.2.3. Categorie del lessico
145	3.3. Morfologia
149	3.3.1. Distribuzione di tempi e modi verbali
152	3.3.2. Perifrasi [stare + gerundio]
153	3.3.3. Frequenza di pronomi personali
156	3.3.4. Frequenza di pronomi e aggettivi possessivi
159	3.3.5. Frequenza di pronomi e aggettivi dimostrativi
162	3.4. Sintassi: lunghezza dei periodi e leggibilità
166	3.4.1. Costrutti <i>ad sensum</i>
168	3.5. Testualità: l'uso dei connettivi
170	3.6. Conclusioni: differenze con la stampa periodica e prospettive di ricerca

175 Per approfondire

Premessa

Questo volume fa il punto su quasi 15 anni di ricerche che mi hanno visto impegnato in un filone che, all'epoca in cui ho iniziato a occuparmene, non era particolarmente frequentato dagli studiosi di linguistica italiana: l'italiano delle traduzioni e, in particolare, la sua indagine con le metodologie della linguistica dei corpora e dell'analisi statistica dei dati testuali. Non che non ci fossero stati tentativi precedenti: basti menzionare i due volumi curati da Giuliana Garzone e Anna Cardinaletti intitolati *Lingua, mediazione linguistica e interferenza* (Franco Angeli, Milano, 2004) e *L'italiano delle traduzioni* (Franco Angeli, Milano, 2005). Tuttavia, la mia ambizione era provare a utilizzare il più possibile gli strumenti informatici per mettere a confronto due subcorpora bilanciati di testi scritti originariamente in italiano e tradotti da altre lingue, cercando di disinnescare l'effetto di altre variabili quali il tipo di testi e la loro lunghezza, la mano del traduttore, la lingua di partenza, gli argomenti ecc.

L'occasione giusta si presentò quando mi accorsi che i principali quotidiani italiani, che intanto avevano reso accessibile una versione online perlopiù identica a quella cartacea e (al tempo) mantenevano liberamente consultabile un archivio di una certa ampiezza, avevano iniziato a pubblicare con cadenza quasi settimanale articoli tradotti da (come si usa dire ancora) "prestigiose testate straniere".

Così ho spulciato a ritroso gli archivi di tre testate nazionali alla ricerca di tutte le traduzioni, analizzandone la distribuzione nel tempo in base agli argomenti, al tipo di articoli, agli autori e ai traduttori, alla lingua di partenza. Quando non ho più trovato traduzioni, sulla base della classificazione che avevo fatto del materiale raccolto fino a quel momento sono passato alla compilazione di un corpus che riflettesse la stessa distribuzione, ma fosse composto da articoli scritti direttamente da giornalisti italiani. Le principali riflessioni utili a queste mie ricerche preliminari sono descritte nell'articolo intitolato *Per un'analisi dell'italiano tradotto nei quotidiani: considerazioni preliminari sulla costituzione di un corpus*, in "Rivista internazionale di tecnica della traduzione", n. 10/2008, pp. 81-99.

Contemporaneamente continuavo a frequentare metodologie di ricerca considerate al tempo piuttosto innovative all'interno di un gruppo di lavoro capitanato da Arjuna Tuzzi e Michele A. Cortelazzo, che poi è sfociato nel *Gruppo Interdisciplinare di Analisi Testuale* (www.fisppa.unipd.it/giat-it/home). Grazie ai preziosi consigli dei membri del gruppo, io e Matteo Viale (senza il quale non sarei riuscito a condurre la ricerca) abbiamo analizzato i due subcorpora secondo l'ipotesi degli "universali traduttivi" e, dopo alcune tappe intermedie, ne abbiamo illustrato i risultati in un altro articolo di respiro assai più ampio: *L'assetto dell'italiano delle traduzioni in un corpus giornalistico. Aspetti qualitativi e quantitativi*, in "Rivista internazionale di tecnica della traduzione", n. 12/2010, pp. 1-62.

Terza tappa delle mie ricerche sulle traduzioni nei giornali italiani è stato l'incontro con Dominique Labbé, l'ideatore di una formula (oggi i giornalisti lo chiamerebbero un "algoritmo") in grado di calcolare la somiglianza o la dissomiglianza tra testi o corpora di testi basandosi sulla distribuzione del lessico al loro interno. Dopo aver compilato un ulteriore corpus di articoli giornalistici adatto allo scopo, insieme a un altro componente del GIAT, Paolo Nadalutti, e grazie e alle sue competenze informatiche e statistiche (e alla sua pazienza), abbiamo provato a sfruttare la distanza intertestuale per verificare se riuscisse a distinguere testi tradotti e testi non tradotti e se fosse possibile anche valutare l'impatto della lingua di partenza. I risultati di queste indagini sono raccolti nei contributi: *Distanza intertestuale e lingua fonte: premesse teoriche, compilazione di un corpus e procedure di analisi*, in *Testi, corpora, confronti interlinguistici: approcci qualitativi e quantitativi*, a cura di G. Palumbo, EUT, Trieste, 2017, pp. 27-42 e *Distanza intertestuale e*

lingua fonte: analisi di un corpus giornalistico, sempre nello stesso volume, pp. 44-64.

Nel frattempo avevo spostato il mio interesse verso un'altra varietà della lingua italiana, quella della prosa letteraria e paraletteraria, conducendo indagini su diversi aspetti delle differenze tra opere di autori italiani e opere tradotte da altre lingue, anche in considerazione dello sviluppo temporale dei fenomeni indagati negli ultimi due secoli. Alcuni dei corpora e subcorpora che avevo compilato in questa nuova fase dei miei studi sono stati utilizzati da Floriana C. Sciumbata per replicare (per quanto possibile e con strumenti informatici che intanto si erano parecchio affinati) le medesime indagini che io, Matteo Viale e Paolo Nadalutti avevamo condotto a suo tempo sugli articoli giornalistici. Si tratta di una ricerca di notevole importanza perché permette di valutare (anche in diacronia) il ruolo del genere testuale e della varietà linguistica nello studio del cosiddetto "traduttese".

Per offrire una visione d'insieme ai lettori e alle lettrici, abbiamo deciso di riunire in un unico volume i risultati dei diversi momenti di una ricerca che trova la sua unità nelle premesse teoriche e nell'approccio metodologico adottato. Il primo capitolo contiene, appunto, una sintesi delle teorie che ipotizzano l'esistenza del traduttese e degli universali traduttivi, seguito da un breve excursus sugli assunti di base della linguistica dei corpora e sui principali strumenti di indagine, e dall'illustrazione dei vari corpora utilizzati per le ricerche. Il secondo capitolo riunisce (con le dovute modifiche e gli eventuali aggiornamenti) i contenuti degli articoli a suo tempo scritti da me insieme a Matteo Viale e Paolo Nadalutti. Il terzo capitolo (del tutto inedito) rende conto dell'applicazione delle medesime teorie e metodologie alla prosa italiana nella letteratura "alta" e di consumo prodotta sia da autori italiani sia da traduttori di romanzi e racconti stranieri tra il XIX secolo e l'inizio del XXI secolo. Chiude il volume una panoramica bibliografica che permette, se lo si desidera, di approfondire ulteriormente diversi aspetti degli studi qui presentati.

Ai meri fini dell'attribuzione autoriale, io mi sono occupato della stesura (oltre che di questa premessa) dei capitoli 1 e 2, Floriana C. Sciumbata ha redatto il capitolo 3 e la sezione finale di approfondimento bibliografico.

Trieste, dicembre 2024
Stefano Ondelli

1. Le basi teoriche, i metodi e i materiali

1.1. Il traduttese

Perché talvolta (sempre meno spesso, man mano che le tecnologie migliorano) le traduzioni proposte dai sistemi informatici – come per es. *Google Translate*, *DeepL* o anche *ChatGPT* – non ci soddisfano? Perché “sanno” di traduzione, nel senso che suonano poco naturali in italiano, evidenziando strutture grammaticali e scelte lessicali che un parlante madrelingua non avrebbe mai selezionato, magari chiaramente calcate sulla lingua di partenza. Però siamo anche pragmatici, e utilizziamo questi sistemi di traduzione automatica per conoscere i contenuti di testi scritti in lingue che non conosciamo o mastichiamo poco: meglio avere un testo che suona un po’ strano che non capire niente del tutto. Analogamente, non ci scandalizziamo più di tanto se la versione italiana di un sito web che propone in vendita di capi di abbigliamento o i servizi di un hotel dimostra chiaramente di essere il risultato di una traduzione automatica, mentre non accetteremmo mai una qualità linguistica così bassa nell’ultimo libro pubblicato in Italia dal nostro romanziere straniero preferito. Insomma, nel caso di testi che potremmo chiamare “strumentali”, cioè di cui vogliamo semplicemente conoscere il significato per fare qualcosa di pratico – per es. prenotare un hotel, azionare una la-

vatrice, comprare un paio di scarpe – accettiamo tranquillamente di percepire che ciò che leggiamo è una traduzione; quando invece pensiamo che la qualità della lingua rivesta una particolare importanza – per es. nel caso di prodotti artistici di tipo letterario o cinematografico – allora vorremmo che non ci fosse alcuna differenza linguistica tra una traduzione e un’opera scritta direttamente in italiano. Tuttavia, secondo alcuni, questo non sarebbe possibile: al massimo un bravo traduttore riesce a nascondere alla nostra percezione il fatto che ciò che scrive è derivato da un testo preesistente, ma non può evitare di produrre un italiano diverso da quello che usa quando si esprime direttamente nella nostra lingua. In altre parole, nessun traduttore, seppur bravo, riuscirebbe a evitare del tutto il cosiddetto “traduttese”, cioè una varietà distinta di italiano che in parte risente dell’influenza della lingua di partenza e in parte del processo di passaggio da una lingua all’altra.

1.1.1. *Dalla fedeltà alla ricezione*

L’idea che esistano – in maniera più o meno evidente – delle caratteristiche comuni a tutte le traduzioni e che le distinguono dai testi scritti originariamente in una data lingua è stata proposta qualche decina di anni fa a seguito di un cambio di prospettiva in alcune ricerche relative a contatto linguistico e traduttologia. In effetti, fino all’ultimo quarto del XX secolo, gli studi sulla traduzione erano sempre stati incentrati sul confronto tra il testo di partenza e il testo di arrivo, per cercare di stabilire quanto il secondo dovesse restare “fedele” al primo oppure potesse allontanarsene per venire incontro alle attese linguistiche e culturali dei suoi nuovi destinatari; tale contrapposizione era già nota nei tempi antichi, come dimostrano Cicerone e Orazio prima e San Girolamo poi, quando risponde a chi lo critica per la scarsa aderenza della sua versione latina al testo greco del Vecchio Testamento (“Non verbum de verbo, sed sensum exprimere de sensu”). Se il concetto di fedeltà emerge esplicitamente nella nozione di *belle infidèle* (cioè una traduzione il più possibile vicina al gusto e alle attese del pubblico a cui si rivolge), teorizzata e perseguita particolarmente nella Francia del XVII secolo, anche approcci più recenti e propriamente scientifici all’attività traduttiva adottano etichette diverse che però paiono sempre riconducibili al confronto tra testo di partenza e di arrivo. Per limitarci ai più noti modelli teorici elaborati nella seconda metà

del Novecento, possiamo citare (a mero titolo di esempio) scelte terminologiche piuttosto rivelatrici, come quelle di corrispondenza formale vs. equivalenza dinamica (Eugene Nida); traduzione semantica vs. comunicativa (Peter Newmark); traduzione manifesta vs. nascosta (Juliane House); traduzione straniante vs. addomesticante (Lawrence Venuti) ecc. Il problema, naturalmente, è stabilire che cosa, in che misura e quando sia meglio restare aderenti all'opera originale, col pericolo di offrire ai destinatari una lingua e contenuti suscettibili di risultare poco accettabili, o "tradirla", eliminando gli elementi più estranei alla lingua e cultura di arrivo, ma correndo il rischio di presentare ai destinatari un testo significativamente "diverso" da quello di partenza.

In realtà, negli ultimi trent'anni del XX secolo, alcuni studiosi hanno iniziato a concentrarsi sulla traduzione in quanto eminentemente atto comunicativo nel contesto linguistico, sociale e culturale di arrivo. Per es., secondo la scuola funzionalista, la teoria dello *Skopos* e i successivi sviluppi, è la strategia operativa, determinata in base alla situazione, alla cultura e ai bisogni di chi commissiona una traduzione, che permette di valutare l'efficacia ed, eventualmente, la qualità del lavoro svolto. Analogamente, anche l'approccio descrittivo delle traduzioni esistenti (*Descriptive Translation Studies*) perde interesse nel legame che intercorre tra testo di partenza e testo di arrivo per focalizzarsi sulla traduzione come atto comunicativo inserito nella cultura che lo accoglie, la quale a sua volta ne è influenzata e modificata.

Sulla scorta di studi precedenti, con focus eminentemente letterario, più che del rapporto con il testo di partenza i *Descriptive Translation Studies* si occupano di come una traduzione riesca a contribuire allo sviluppo del sistema culturale in cui si inserisce modificandolo, arricchendolo e, talvolta, perfino rivoluzionandolo, quando vengono soppiantati i modelli preesistenti. Quindi, invece della domanda tradizionale posta fino a quel momento (quanto e in che modo è possibile restare fedeli all'originale?), i nuovi quesiti a cui si cerca una risposta discendono piuttosto dall'osservazione a posteriori dell'intero processo traduttivo, anche dal punto di vista economico: per quale motivo viene tradotto proprio questo testo in questo momento? A quale scopo? In che modo viene tradotto e perché? Quali sono le conseguenze nella cultura e nella società che lo accolgono? Che novità introduce? Inoltre, se si dispone di un numero sufficiente di osservazioni su processi traduttivi diversi,

sarà possibile estrapolare le tendenze di fondo (che Gideon Toury definisce “norme”) che caratterizzano l’azione dei traduttori in seno a una determinata cultura e un determinato periodo storico.

Come si può vedere, si tratta di un cambio di prospettiva notevole che sposta l’attenzione dal testo di partenza (e dal suo autore) al processo traduttivo e alle sue conseguenze (insomma sui risultati) sulla platea dei destinatari. In effetti occorre ricordare che questi ultimi di norma non hanno accesso al testo originale: per definizione, abbiamo bisogno di una traduzione quando non conosciamo (sufficientemente bene) una lingua straniera. Di conseguenza, i lettori (o gli ascoltatori/spettatori, nel caso di testi orali e audiovisivi) conosceranno un dato testo solo nella sua versione tradotta, senza poterne verificare la fedeltà all’originale, come dimostrano non solo i numerosi esempi di “rimaneggiamento migliorativo” (*le belles infidèles* già menzionate sopra) ma anche il caso della censura: per es. le prime traduzioni delle opere di Walter Scott disponibili in Italia, oltre a passare per il tramite del francese, furono depurate di numerosi riferimenti politici e religiosi; inoltre è ben noto che il doppiaggio cinematografico sia stato reso obbligatorio dal regime fascista proprio per poter avere un controllo più agevole non solo sull’ingresso dei forestierismi in italiano, ma soprattutto sui contenuti dei film importati dall’estero.

1.1.2. Il livello linguistico

Se, come abbiamo visto sopra, l’interesse si sposta dal testo di partenza alla produzione e alla fruizione della traduzione presso i suoi destinatari, una delle conseguenze può essere una maggiore attenzione per la lingua che la caratterizza. Come accade per generi letterari, idee e contenuti, le traduzioni possono agire da tramite per l’ingresso di nuovi fenomeni linguistici, oppure come spinta aggiuntiva nei confronti di evoluzioni già in atto, oppure ancora come freno conservatore che si oppone a certe innovazioni linguistiche. Questo soprattutto se, come abbiamo già detto, non si presentano al pubblico in quanto tali ma come se fossero state scritte originariamente nella lingua di arrivo: è il caso delle traduzioni “nascoste” (*covert*, nella terminologia di Juliane House). A livello linguistico, ciò significa (come del resto ci aspettiamo di norma) che non sono caratterizzate da lessico e strutture che possano essere percepiti come “strani” o poco accettabili dai destinatari.

Tuttavia, se ipotizziamo, in accordo con l'approccio descrittivo, che le traduzioni contribuiscono all'evoluzione della cultura di accoglienza, è lecito concludere che ciò sarà vero anche per quanto riguarda la norma linguistica. In altre parole, espressioni e strutture innovative che derivano dal processo traduttivo e che vengono ripetutamente proposte ai destinatari senza che questi si rendano conto della loro origine esogena potrebbero entrare nella lingua di arrivo, arricchendola e modificandola. Un caso emblematico è quello delle cosiddette "routine traduttive" nel doppiaggio cinematografico e televisivo: espressioni come "come posso aiutarla?", "il mio nome è", "ma dici a me?" "stai scherzando o cosa?" oggi sono utilizzate nella comunicazione quotidiana anche da parlanti di lingua madre italiana sull'esempio dei doppiaggi dei film e serie statunitensi. Un discorso simile (non limitatamente al doppiaggio) può essere valido per l'incremento della frequenza e degli ambiti d'uso della perifrasi [stare + gerundio], secondo il modello inglese [to be + -ing].

A livello linguistico, dunque, il fenomeno che è più intuitivo attendersi nelle traduzioni è quello dell'influenza del testo di partenza (soprattutto se è scritto in una lingua che gode di grande prestigio, come capita oggi con l'inglese). Questo fenomeno, ampiamente studiato da chi si occupa del contatto linguistico (cioè dall'impiego di due o più lingue a livello di uno stesso individuo o di una comunità di parlanti), va sotto il nome di "interferenza linguistica", ed emerge quando il testo tradotto presenta delle caratteristiche che sono riconducibili all'influsso del testo fonte. Naturalmente, ci si attenderebbe che un traduttore professionista esperto abbia l'accortezza di sottrarsi a questo influsso e quindi di evitare di scrivere un testo che riveli immediatamente al lettore la sua origine "straniera", proprio perché caratterizzato da una lingua "poco naturale". In effetti, nelle loro esercitazioni preliminari, molti corsi di laurea destinati alla mediazione linguistica si premurano proprio di sottrarre gli studenti a questa tendenza.

Tuttavia non tutti i fenomeni linguistici sono così facili da individuare e controllare. In particolare, nelle persone che hanno pari competenze in due lingue (i parlanti bilingui) o quasi (per es. i traduttori), il passaggio continuo da un idioma all'altro potrebbe renderli meno sicuri – anche solo in qualche occasione – nell'applicazione di alcune regole grammaticali. Ovviamente non si tratta di regole di base, subito riconoscibili a tutti: per rendere l'inglese "I

have often been told to pay attention” nessun traduttore, nemmeno alle prime armi, scriverebbe (o direbbe) mai “Io sono stato spesso detto di fare attenzione”: semplicemente, non si tratta di una costruzione italiana. Però ricordo di un mio conoscente anglofono, ma molto competente in italiano, che spesso mi diceva: “Ma io non capisco perché tu hai fatto questo”, e io sorridevo. Non era una frase indiscutibilmente sbagliata, ma nessun italiano l’avrebbe mai detta: troppi pronomi soggetto espliciti (“io” e “tu”, senza nessuna esigenza di contrasto) e un dimostrativo di troppo (un’alternativa più accettabile potrebbe essere: “non capisco perché l’hai fatto”) rendevano la frase non impossibile ma poco “naturale”.

Naturalmente, il mio amico, per quanto avesse familiarità con l’italiano, era madrelingua inglese, ma lo stesso può accadere con parlanti di lingua madre italiana che siano abituati a esprimersi in inglese: ricordo un famoso economista italiano naturalizzato americano che in televisione continuava a insistere che “il debito italiano è troppo largo” (*large* invece di “grande, ampio”). Attenzione: non è necessario essere bilingui per incorrere in incertezze di questo tipo: oggi gli studenti dicono tranquillamente cose come “Professore, sto capendo che cosa intende dire” invece di “comincio a capire” oppure “mi è sempre più chiaro”, mentre la generazione precedente, meno soggetta all’influsso dell’inglese, avrebbe trovato questo enunciato alquanto bizzarro.

Insomma, anche se, soprattutto in Italia, la scuola ci ha abituato a decidere in maniera indiscutibile che cosa è giusto o sbagliato quando si tratta di regole grammaticali, non è sempre così. Per es. “io non capisco perché hai fatto questo” non è una frase che possiamo definire grammaticalmente inaccettabile: diciamo che è una frase particolare (in linguistica si direbbe che è “marcata”), perché può essere pronunciata solo in certe situazioni e non in altre, per es. se “io” si oppone ad “altri”, che invece capiscono benissimo, e “questo” è un contrasto con “quello” o altre opzioni. Infatti, l’espressione del pronome soggetto (che in italiano non è obbligatorio) o la preferenza per il pronome dimostrativo invece di un pronome debole come “lo” non sono scelte libere ma motivate da esigenze comunicative (per es. il contrasto); la loro presenza non richiesta non rende la frase inaccettabile, perché contraria alle norme che governano la lingua italiana, ma solo più o meno “strana” a seconda delle situazioni. Si tratta di sfumature che riguardano aspetti come l’espressione e la posizione dei prono-

mi soggetto, la posizione degli aggettivi, il sovrasfruttamento dei possessivi, l'uso degli articoli, l'ordine di certi elementi all'interno della frase ecc., e che possono sfuggire non solo all'attenzione del traduttore attento, influenzato dalla lingua del testo su cui sta lavorando, ma anche dei suoi lettori, che non notano queste "stranezze", e anzi ci si abitua.

1.1.3. L'interferenza

In tutti gli esempi visti sopra abbiamo trattato del tema dell'interferenza linguistica: nel caso della traduzione, ci riferiamo all'influsso del testo di partenza sul testo di arrivo, perché il traduttore non riesce a sottrarsi a questo influsso, in (gran) parte anche perché magari il testo di partenza è scritto in una lingua e appartiene a una cultura che godono di grande considerazione da parte dei destinatari della traduzione, che sono ben propensi ad accogliere le novità che ne derivano. Come è evidente, non si tratta solo ed esclusivamente di una serie di variazioni linguistiche determinate dal processo traduttivo: qui si tratta dell'interferenza di una lingua specifica su un'altra, che determina esiti specifici; per es. se la lingua di partenza presenta la caratteristica dell'esplicitazione obbligatoria del pronome soggetto, proprio come la lingua di arrivo, non rileveremo alcuna interferenza; se invece la lingua di partenza impiega il passato remoto secondo una distribuzione diversa rispetto alla lingua di partenza, ecco che potrebbero emergere delle sorprese. In altre parole, dal punto di vista dell'interferenza, potremmo studiare l'influsso dell'inglese, del francese, dello spagnolo o del tedesco sull'italiano, ma non potremmo trarre conclusioni generalmente valide sul "traduttese", cioè sulle differenze tra l'italiano prodotto da traduttori (a partire da qualsiasi lingua) e da parlanti/scriventi di testi originali in italiano.

Resta il fatto che, a prescindere dalle lingue in gioco, i traduttori evidenziano la tendenza a ricalcare le strutture del testo di partenza a tutti i livelli; questa tendenza potrà risultare più o meno accentuata in base all'esperienza del traduttore e al prestigio della lingua (e cultura) del testo fonte. Infatti, se da una parte è lecito attendersi che i traduttori alle prime armi saranno meno propensi a "staccarsi" dal testo di partenza, dall'altra potranno anche subire maggiormente l'influenza di testi provenienti da culture e lingue considerate di grande prestigio nella comunità dei destinatari. In particolare,

questo prestigio può essere particolarmente percepibile in base agli argomenti e ai tipi di testo; per es. può capitare che committenti di documenti di argomento finanziario raccomandino al traduttore di cercare di “essere glamour”, mantenendo e lasciando trasparire tutto l’inglese possibile, mentre in traduzioni che parlano di moda e profumi è piuttosto il francese che garantisce una certa attrattività (anche se bisogna dire che, nel marketing globalizzato e anglofono di oggi, anche settori come il lusso e la cucina ormai vedono il dominio dell’inglese, e non solo nei testi tradotti).

Una conseguenza interessante che è stata ipotizzata dagli studiosi riguarda la minore frequenza nelle traduzioni di strutture linguistiche presenti esclusivamente nella lingua di arrivo. A ben vedere si tratta di un fenomeno speculare rispetto a quanto illustrato finora. Mentre è facile e intuitivo prevedere che la costante presenza di certi costrutti nel testo di partenza lasci tracce nella traduzione (per es. il passato remoto nel passaggio dallo spagnolo in italiano), meno automatico è aspettarsi che, di converso, se il testo originale non presenta certe caratteristiche che sono comuni nella lingua di arrivo, il traduttore potrebbe essere, per così dire, meno “invogliato” a utilizzare queste caratteristiche nella sua resa (per es. il passato remoto dal francese).

Infine, un’ulteriore conseguenza della costante abitudine a tradurre da una certa lingua verso la propria potrebbe essere una riduzione della creatività: i traduttori (ma io mi aspetto qualcosa di analogo da parte di tutti coloro i quali scrivono per professione, per es. i giornalisti) si dotano di un armamentario di soluzioni preconfezionate per rendere parole e strutture ricorrenti nella lingua di partenza. In altre parole, quando il traduttore incontra una certa espressione nel testo di partenza (per es. l’inglese *nevertheless*), forse per risparmiare energia cognitiva la rende automaticamente e con costanza con un solo equivalente nella lingua di arrivo (per es. “ciononostante”), anche se naturalmente esiste un’intera batteria di opzioni più o meno sinonimiche (per es. “nonostante ciò”, “cionondimeno”, “comunque”, “tuttavia”, “e però” ecc.) che, a seconda dei contesti, potrebbero essere altrettanto o addirittura più adatte a rendere l’originale. Come si può vedere, siamo di fronte a un effetto collaterale dell’influenza della lingua di partenza su quella di arrivo, che ha per conseguenza una riduzione della creatività linguistica del traduttore, più propenso a ripetere soluzioni preconfezionate invece di optare per la *variatio*.

1.1.4. Le regolarità del traduttese: gli universali

L'interferenza è un fenomeno a cui i traduttori idealmente non dovrebbero essere soggetti – perché dovrebbero essere in grado di distinguere nettamente tra lingua di partenza e lingua di arrivo, evitando di introdurre in quest'ultima soluzioni ibride che non risultano “naturali”. Eppure si tratta anche della caratteristica più riconoscibile: quando diciamo che un testo “sa di traduzione”, nella maggior parte dei casi siamo in grado di indicare alcune caratteristiche che non appartengono pienamente alla lingua di arrivo e sono – almeno parzialmente – mutate da quella di partenza. Ne consegue che, poiché sarebbero caratterizzate – in maniera più o meno evidente – dalla tendenza a ricalcare le strutture dei testi di partenza a livello lessicale, morfosintattico e anche testuale, tutte le traduzioni avrebbero qualcosa in comune e differirebbero dai testi scritti direttamente nella lingua considerata.

Questa tendenza non è però l'unica regolarità che emerge dall'osservazione delle traduzioni come testi autonomi nella cultura di arrivo: secondo la teoria degli “universali traduttivi” l'operazione del tradurre in sé determinerebbe comportamenti e strategie ricorrenti che si concretizzerebbero in caratteristiche linguistiche che distinguono le traduzioni dai testi non tradotti. Insomma, i traduttori risentirebbero, per es., dei vincoli posti dall'esistenza del testo fonte; della consapevolezza che il pubblico a cui si rivolgono è diverso da quello originariamente ipotizzato per il testo che stanno traducendo; delle convenzioni e attese che riguardano le traduzioni nella loro cultura e nel periodo storico in cui lavorano: tutti questi vincoli lascerebbero tracce linguistiche in maniera più o meno marcata e a livelli diversi. Infatti, i comportamenti adottati dai traduttori e le caratteristiche che ne discendono non sono probabilmente né certi né obbligatori: si tratta di tendenze di fondo che emergono dallo studio di un grande numero di esempi tangibili di testi tradotti, a prescindere dalle lingue in gioco, dalle caratteristiche del singolo testo e delle preferenze personali del singolo traduttore. Tali tendenze, su cui ci soffermeremo nei paragrafi che seguono, sono quattro: esplicitazione, semplificazione, normalizzazione e convergenza. L'interferenza, invece, non apparterebbe pienamente agli universali traduttivi perché non dipende esclusivamente dal processo traduttivo in sé, ma rientra nel fenomeno più ampio del contatto tra lingue diverse.

La tendenza all'esplicitazione è, tra gli universali traduttivi, quello che più ha raccolto il consenso degli studiosi. Si suppone che i traduttori, consapevoli delle differenze culturali che distinguono i loro lettori rispetto al pubblico dell'opera originale, si sforzino di rendere il testo più accessibile, introducendo materiale linguistico aggiuntivo e, talvolta, ridondante, per esprimere in maniera più esplicita le informazioni che veicolano. Così, rispetto agli originali, nelle traduzioni si eviterebbero le ellissi e, oltre ad aggettivi, avverbi e interiezioni, verrebbe inserito un maggior numero di connettivi che esplicitano i collegamenti logici e sintattici tra le frasi, come certe congiunzioni e preposizioni, ma anche aggettivi e pronomi dimostrativi. Per quanto riguarda il lessico, accanto alla preferenza per la ripetizione dei nomi a scapito dei pronomi, le metafore innovative e le espressioni che sono più legate alla cultura di partenza (specie se questa è poco nota in quella di arrivo) possono essere sostituite da soluzioni più neutre o accompagnate da spiegazioni e chiarimenti, magari in strutture parentetiche o in glosse esplicative, e di conseguenza le traduzioni risulterebbero di norma più lunghe degli originali.

Una seconda tendenza universalmente seguita dai traduttori riguarderebbe la semplificazione del lessico e della sintassi del testo fonte. Diverse ricerche hanno dimostrato che, rispetto agli originali, a livello lessicale le traduzioni risultano meno ricche e dense (cfr. il § 1.2 per una spiegazione dei concetti di ricchezza e densità lessicale); registrano un uso più abbondante di parole dal significato generico e approssimativo e di parafrasi esplicative in sostituzione di lessico specifico, specie se radicato nella cultura del testo di partenza; sostituiscono espressioni formali e auliche con equivalenti più moderni e di uso più comune nella lingua di arrivo. A livello sintattico cresce nelle traduzioni il numero di frasi subordinate esplicite a scapito delle implicite, ma la lunghezza delle frasi viene ridotta grazie all'introduzione di segni di punteggiatura forti (come i punti fermi) in sostituzione di quelli più deboli (per es. le virgole). Un'ultima strategia di semplificazione consiste nella cancellazione delle ridondanze presenti nel testo originale.

Di notevole importanza per i traduttori italiani pare essere l'universale traduttivo della "normalizzazione", che prevede la preferenza per una lingua standard e convenzionale, che denota un atteggiamento piuttosto conservativo. In poche parole, i traduttori avrebbero sempre ben presenti le regole che hanno appreso a

scuola e che possono agevolmente recuperare nelle grammatiche di riferimento, anche se tutti i parlanti italiani sono consapevoli che non sempre queste regole vengono seguite nell'uso quotidiano (il cosiddetto "italiano dell'uso medio"). Insomma, oltre a preferire soluzioni conformi alla norma codificata a livello di morfosintassi e punteggiatura, i traduttori sembrerebbero prediligere un lessico convenzionale e collocazioni stereotipiche, e preferiscono evitare i forestierismi (ma molto dipende dal prestigio della lingua di partenza). Oltre ad attenuare l'inventiva e l'originalità del testo fonte, questo atteggiamento conduce a testi orali (per es. le interpretazioni simultanee, ma anche – *mutatis mutandis* – i doppiaggi dei film) da cui vengono espunte molte delle caratteristiche tipiche del parlato (per es. errori, false partenze, autocorrezioni, frasi in cui le parole hanno un ordine particolare) proprio per conformarsi alla norma codificata guardando ai testi scritti.

L'ultimo universale traduttivo è quello meno studiato e probabilmente più difficile da cogliere. Per semplificare, possiamo dire che la convergenza (o *levelling out*) prevede che tutte le traduzioni si assomiglino un po', o comunque siano più simili tra loro di quanto non lo siano i rispettivi testi di partenza. Si tratterebbe di una tendenza che non risente dell'incrocio tra lingua di partenza e di arrivo. All'atto pratico, se confrontiamo un certo numero di traduzioni di romanzi, dovremmo rilevare delle differenze nella densità e ricchezza lessicale e nella lunghezza media dei periodi meno marcate delle differenze che è possibile riscontrare tra gli originali.

1.1.5. Termini di paragone

Per rilevare le tendenze di fondo che definiscono il *modus operandi* di tutti i traduttori (i cosiddetti "universali traduttivi") non basta identificare le caratteristiche linguistiche delle traduzioni: occorre naturalmente avere un termine di paragone per stabilire se queste caratteristiche sono peculiari o comuni ad altri testi, se sono più o meno frequenti ecc. Alternativamente, infatti, finora abbiamo fatto riferimento ai testi fonte (per es. è necessario controllare l'originale per stabilire se la lingua della traduzione risulta semplificata, o se sono state aggiunte informazioni che erano implicite) oppure a un ipotetico comportamento standard dei parlanti della lingua di arrivo (per es. per valutare se le traduzioni sono caratterizzate da una forte aderenza alla norma codificata). Nel primo caso gli studiosi

parlano di S-Universals (dove S sta per *source text*, cioè “testo fonte” in inglese), nel secondo di T-Universals (dove T sta per *target text*, cioè “testo di arrivo” in inglese).

In particolare, sarebbero più facilmente ricavabili dal confronto con i testi fonte il peso dell’interferenza linguistica, la normalizzazione, l’esplicitazione, oltre al fatto che le traduzioni tendono a essere più lunghe degli originali e che traduzioni successive dello stesso testo di partenza dimostrino una fedeltà crescente. Il confronto con testi analoghi redatti direttamente nella lingua di arrivo (per es. tra articoli di giornale o romanzi tradotti e non tradotti), oltre alla convergenza, renderebbe più evidenti alcuni aspetti della semplificazione (come la minore ricchezza e densità lessicale) e degli esiti “in negativo” dell’interferenza linguistica, cioè la minor frequenza di costrutti tipici della lingua di arrivo ma assenti nella lingua di partenza.

Naturalmente nulla impedisce che lo stesso universale linguistico sia corroborato da entrambe le prospettive. Per es., nel caso della normalizzazione, una traduzione potrebbe caratterizzarsi sia per l’assenza di tratti non convenzionali riscontrabili nel testo fonte, sia per una maggior aderenza alle regole della grammatica standard rispetto a testi analoghi scritti direttamente nella lingua di arrivo. Nel caso della semplificazione, si potrebbe riscontrare un maggiore snellimento sintattico (quindi frasi più brevi e con meno proposizioni subordinate) nel confronto col testo fonte come pure con testi analoghi che non sono il risultato di una traduzione.

1.1.6. Problemi teorici e prospettive di ricerca

L’ipotesi degli universali traduttivi apre interessanti prospettive di ricerca da diverse angolazioni, dalla descrizione degli effetti del processo della traduzione sui fenomeni linguistici all’indagine dei processi cognitivi che caratterizzano i traduttori al lavoro; dalle ricadute didattiche sulla formazione degli studenti di traduzione e interpretazione allo studio dell’evoluzione di certe tendenze traduttive, attraverso i secoli ma anche attraverso i vari tipi di testo. Tuttavia, come spesso avviene nella ricerca scientifica, al problema del reperimento dei dati (di cui parleremo tra poco al § 1.2.) si aggiunge quello della loro interpretazione. Per es. abbiamo già avuto modo di accennare (§ 1.1.5.) al fatto che distinguere tra S- e T-Universals non è sempre agevole, a causa di possibili incroci e sovrapp-

posizioni, per cui lo stesso dato linguistico, che può confermare un certo universale traduttivo, potrebbe emergere dal confronto sia con i testi fonte delle traduzioni sia con testi analoghi scritti direttamente nella lingua di arrivo. Se prendiamo l'esplicitazione dei pronomi soggetto in lingue in cui non sono sempre obbligatori, come l'italiano, per valutarne (e spiegarne) la presenza possiamo confrontare una traduzione con l'originale e controllare se i pronomi emergono in traduzione proprio laddove erano presenti nei testi originali, ma possiamo anche controllare se esibiscono una maggiore frequenza in paragone a testi scritti direttamente in italiano. Analogamente, per mantenere lo stesso esempio, la presenza dei pronomi soggetto espliciti potrebbe essere ascritta alla legge dell'interferenza se la lingua da cui si traduce prevede il soggetto esplicito obbligatorio, ma anche all'universale linguistico dell'esplicitazione, perché esplicitando il soggetto il traduttore rende più agevole al lettore ricostruire i riferimenti personali all'interno del testo (in altre parole: è più facile capire chi fa che cosa).

Al di là delle possibili sovrapposizioni, per cui un certo fenomeno può avere diverse spiegazioni, emergono casi più spinosi che tratteggiano vere e proprie contraddizioni. Per es., se dall'analisi di un certo numero di testi tradotti rileviamo che aumenta il numero delle frasi esplicite (non importa se in confronto ai testi fonte o a testi analoghi scritti in italiano), che cosa possiamo concludere? La presenza di verbi di modo finito e di connettivi potrebbe far pensare a un caso di esplicitazione, che rende il testo più facile da capire, ma le frasi secondarie esplicite possono prevedere l'uso di modi e tempi tradizionalmente considerati difficili, come il congiuntivo, mentre le strutture implicite potrebbero essere considerate più semplici (per es. "spiego perché impariate" è un costrutto più esplicito ma meno comune di "spiego per farvi imparare"). Analogamente, la giustapposizione di frasi semplici ed esplicite (per es. "Vado a casa. Sono stanco") determina una sintassi più semplice ma sicuramente meno esplicita e forse meno facile da capire rispetto alla gerarchia imposta dalla subordinazione (per es. "Vado a casa perché sono stanco").

Un altro rischio di contraddizione emerge da altri due atteggiamenti che, secondo la teoria degli universali traduttivi, caratterizzerebbero i traduttori: la tendenza a ricalcare le strutture del testo di partenza a tutti i livelli (o legge dell'interferenza) e la tendenza a rispettare la norma codificata (o normalizzazione). In effetti, la

prima prevede una sorta di spinta innovativa, cioè l'introduzione, nella lingua di arrivo, di soluzioni lessicali, morfosintattiche e persino testuali che appartengono alla lingua fonte; la seconda invece presuppone una sorta di resistenza conservativa, che predilige soluzioni più tipiche e frequenti nella lingua di arrivo.

Ovviamente, molti di questi dubbi permangono anche a causa della relativa scarsità di studi disponibili, condotti su lingue diverse, tipi di testo diversi, e con metodologie diverse (dagli spogli manuali all'analisi automatica). Per es. è lecito attendersi che i testi fonte scritti in una lingua di assoluto prestigio internazionale come l'inglese possano comportare un'influenza assai maggiore rispetto a lingue frequentate da gruppi di parlanti ristretti; anche la prossimità strutturale di idiomi della stessa famiglia linguistica dovrebbe facilitare l'interferenza; le traduzioni di testi considerati molto importanti (per es. la Bibbia o certi pilastri del canone letterario) sono probabilmente realizzate con tempi e criteri diversi rispetto a testi "di consumo"; i testi orali (per es. le interpretazioni simultanee e i doppiaggi) sottostanno a vincoli (la durata temporale, la sincronizzazione labiale) che non impattano sui testi scritti.

Alla luce delle osservazioni condotte sopra è auspicabile che le ricerche che muovono dall'ipotesi degli universali traduttivi esplicino in maniera molto dettagliata alcuni aspetti relativi ai materiali sottoposti ad analisi e alle metodologie dispiegate. Nel primo caso si devono individuare le lingue coinvolte e i tipi di testo (e, eventualmente, le loro parti) selezionati per la ricerca; nel secondo le procedure (automatiche o manuali) di approntamento dei testi e di estrazione e misurazione dei dati, soprattutto nel caso di approcci quantitativi. In tal modo, da una parte diventa possibile replicare gli studi per confermarne i risultati, magari variando una delle variabili in gioco; dall'altra si riesce a contestualizzare i dati estratti e a valutare la portata delle conclusioni delle ricerche.

1.2. L'approccio quantitativo: la linguistica dei corpora

Nel § 1.1. abbiamo definito gli universali traduttivi come tendenze di fondo nel comportamento dei traduttori che lasciano tracce sulle superficie linguistica dei testi. Siccome si tratta, per l'appunto, di tendenze di fondo, va da sé che sarebbe ingenuo pensare di analizzare una singola traduzione per individuare specifici tratti linguistici che comproverebbero questo o quell'universale traduttivo:

infatti, finora, abbiamo sempre fatto riferimento all'analisi di "un certo numero di traduzioni". In altre parole, concludere dall'analisi di un singolo testo che un numero ridotto di pronomi soggetto espliciti (anche quando la grammatica italiana non li richiede) conferma l'universale dell'esplicitazione non è possibile: potrebbe trattarsi della preferenza di un singolo traduttore, oppure all'influenza della lingua di partenza, o potrebbe anche frutto della casualità. Se invece lo spoglio di un numero notevole di traduzioni opera di persone diverse, da lingue diverse, e magari di tipi di testo diversi indica che la frequenza dei pronomi soggetto espliciti è maggiore rispetto a quanto normalmente si riscontra nei testi che non sono traduzioni, potremo ragionevolmente concludere (soprattutto se possiamo contare su test di affidabilità statistica) che i traduttori tendono a esplicitare alcune informazioni.

D'altro canto, analizzare manualmente un numero molto elevato di testi è un compito che comporta un notevole dispendio di tempo ed energia, per non parlare dell'alta probabilità di fare errori sia nell'estrazione dei dati che nella loro classificazione. Per questo motivo, da quando è diventato possibile utilizzare i computer per trattare testi in formato elettronico, ha ricevuto grande impulso un filone di studi che va sotto il nome di linguistica dei corpora (in inglese: *corpus linguistics*). Un corpus può essere definito come una raccolta di testi selezionati in base ad alcune caratteristiche che li rendono rappresentativi di una certa varietà della lingua: per es. se voglio studiare l'italiano contemporaneo escluderò i testi antichi; se il mio oggetto di studio è la prosa letteraria escluderò la poesia, la saggistica e gli articoli giornalistici; dovrò evitare che il corpus sia sbilanciato, per es. con una prevalenza quantitativa di un certo autore rispetto agli altri, o degli uomini rispetto alle donne (a meno che, ovviamente, tale squilibrio non sia giustificato). Inoltre, i testi del corpus devono essere in formato elettronico e devono essere preparati adeguatamente per essere indagati per mezzo di appositi software; per es. può essere utile distinguere le varie parti del discorso, oppure definire e segnalare chiaramente i confini di frase ecc.

Con particolare riferimento agli studi sugli universali traduttivi, la distinzione tra S- e T-Universals (cfr. § 1.1.5.) ci permette di individuare diversi tipi di corpora utili a indagare certi fenomeni e non altri. Quindi, se vogliamo confrontare le traduzioni con i loro originali, avremo bisogno di quelli che nella terminologia specifica si chia-

mano corpora bilingui (o multilingui) paralleli. Se invece vogliamo indagare le caratteristiche delle traduzioni nel confronto con testi scritti originariamente in italiano, ricorremo a corpora monolingui paragonabili, cioè comprendenti tipi di testo analoghi (per es. articoli di giornale, romanzi, sentenze ecc.). Nel § 1.3. illustreremo in dettaglio la composizione e le caratteristiche dei corpora utilizzati per le ricerche presentate in questo volume; per il momento ci soffermiamo su alcuni concetti chiave dell'analisi quantitativa dei dati testuali: la ricchezza e la densità lessicale (a cui abbiamo già fatto accenno al § 1.1.4), e la distanza intertestuale.

1.2.1. La ricchezza lessicale

Come si contano le parole in un testo? Sembra una domanda banale ma non lo è. Per prima cosa è necessario stabilire che cosa intendiamo per "parola". Se scriviamo "l'amore" in un documento di *MS-Word*, il programma ci dirà che abbiamo scritto una parola sola; noi sappiamo benissimo che sono due (un articolo e un nome), ma il computer considera "parola" qualsiasi insieme di segni grafici compreso tra due spazi bianchi, e l'apostrofo quindi non conta come confine di parola. Analogamente, ci sono parole, come "carta di credito", "ferro da stiro" o "treno merci", che sono composte da più elementi, ma che noi consideriamo unitarie (e infatti non è possibile inserire aggettivi o avverbi al loro interno). Si tratta delle "polirematiche" o "unità lessicali superiori": a seconda di come le consideriamo, il numero di parole di cui è composto il testo può variare. Un problema ulteriore è dato da parole che noi consideriamo uguali (diremmo che "sono la stessa parola"), ma per il computer non lo sono, come avviene per certe varianti grafiche (per es. "cuore" e "cuor", "pronunciare" e "pronunziare") o anche il semplice uso delle maiuscole: in assenza di istruzioni specifiche, nella frase "La ragazza mangia la pasta" il computer considera "La" e "la" due parole diverse. Si tratta di aspetti (tra gli altri) che rientrano nella fase di preparazione del corpus all'analisi automatica.

Dopo aver definito i criteri che permettono di identificare le parole in un testo, resta il problema di come contarle. Se prendiamo una frase come "io so che tu sai che io so", chiunque direbbe che possiamo contare otto parole, ma alcune di esse si ripetono: "io", "che", "so" compaiono due volte. Secondo la terminologia italiana, le otto parole sono le "occorrenze" (in inglese, *tokens*) che deter-

minano la consistenza del corpus (in questo caso, la frase che abbiamo scelto come esempio), che a sua volta viene indicata con la lettera N. Le cinque forme grafiche (“che”, “io”, “sai”, “so”, “tu”; in inglese *types*) che compaiono nel corpus, alcune delle quali ripetute, formano il vocabolario del corpus, a sua volta indicato dalla lettera V. Le forme grafiche ordinate in base alla frequenza formano il vocabolario di frequenza (tabella. 1.1.); le forme che compaiono una sola volta vengono dette “hapax legomena”.

Rango	Forme grafiche	Frequenza
1	che	2
2	io	2
3	so	2
4	sai	1
5	tu	1

Tabella 1.1 – Vocabolario di frequenza del corpus “io so che tu sai che io so”

La ricchezza lessicale di un corpus è data dal rapporto tra forme grafiche e occorrenze (o *Type-Token Ratio* in inglese, abbreviato in TTR): dati due corpora di pari dimensioni (N), quello che ha più forme grafiche che si ripetono meno di frequente avrà una maggiore ricchezza lessicale, che negli studi sulla leggibilità dei testi viene considerata un indice di maggiore complessità.

1.2.2. La densità lessicale

Anche la densità lessicale è considerata un indice di complessità linguistica, calcolato come il rapporto tra le parole piene e vuote (o grammaticali). La prima classe comprende aggettivi, avverbi, nomi e verbi, cioè le parole dotate di un significato proprio, mentre la seconda è costituita da articoli, congiunzioni, preposizioni e pronomi, cioè parole che contribuiscono soprattutto alla struttura sintattica e di norma sono di numero finito (cioè non prevedono la possibilità di creare neologismi). In realtà la distinzione non è così univoca:

alcuni aggettivi (per es. “molto”) o avverbi (per es. “più”) ricoprono soprattutto una funzione grammaticale. Comunque, più è alta la percentuale di parole piene, più il corpus viene considerato semanticamente denso, cioè carico di significato e, probabilmente, difficile da comprendere. Va da sé che, anche in questo caso occorre preparare il corpus in maniera tale che il programma di analisi automatica sia in grado di riconoscere forme che possono avere più classificazioni (come nel caso di “letto” inteso come nome o come participio passato), anche in base al contesto sintattico (per es. “il più è fatto” o “il bello deve ancora venire”).

1.2.3. La distanza intertestuale

Se ricchezza e densità lessicale rendono conto della difficoltà di lettura dei testi, la distanza intertestuale è un'altra misurazione basata sulle parole atta a valutare il grado di somiglianza o divergenza tra due testi o due corpora. Questo sistema riduce l'informazione contenuta nei testi a una sola dimensione quantitativa, basata sulla frequenza delle parole, calcolando la differenza tra le frequenze rivelate in due testi (o due corpora di testi) oggetto di confronto. Nella fig. 1.1. i due testi A e B sono rappresentati da due insiemi, ciascuno contenente parole che possono essere esclusive di A (porzione più scura), esclusive di B (porzione più chiara) o comuni a entrambi (porzione bianca):

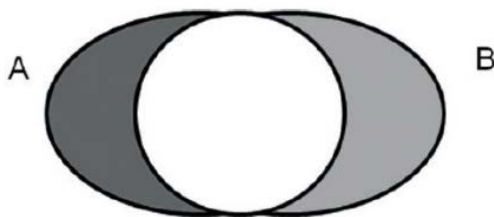


Figura 1.1 - La distanza intertestuale

Se i due testi non presentano nessuna parola in comune, lo spazio bianco sarà assente e la distanza tra i due sarà massima (pari a 1); se invece A e B contengono esattamente le stesse parole e con la

stessa frequenza, si può concludere che i due testi sono identici e la distanza è pari a 0. Ovviamente, nel caso di testi di dimensioni diverse (cioè comprendenti numeri diversi di parole) sarà necessario fare una proporzione secondo la formula riportata qui di seguito:

$$d(A, B) = \frac{\sum_{i \in V_{A \cup B}} |f_{i,A} - f_{i,B}^*|}{2N_A}$$

in cui d è la distanza tra i testi A e B, $f_{i,A}$ e $f_{i,B}$ sono le frequenze delle parole e N_A e N_B sono le dimensioni dei due testi.

Secondo le ricerche, la distanza intertestuale risente di quattro fattori; innanzitutto il tipo di testo, inteso molto largamente come il genere (per es. una commedia risulterà più distante da una tragedia che da un'altra commedia) o la varietà della lingua (per es. la distanza tra due brani di italiano parlato sarà minore della distanza tra un testo orale e un testo scritto). In seconda battuta emerge il periodo storico (testi prodotti negli stessi anni saranno più simili di testi prodotti in secoli diversi); quindi è rilevante l'autore (testi dello stesso autore risulteranno più vicini di testi di autori diversi; per questo motivo la distanza intertestuale è stata spesso utilizzata per verificare l'attribuzione d'autore); infine, anche gli argomenti trattati avranno il loro peso (testi che trattano argomenti diversi risultano più distanti tra loro).

1.3. I corpora esaminati

I paragrafi che seguono dettagliano le caratteristiche delle raccolte di testi che sono state analizzate nelle ricerche illustrate in questo volume. Si tratta di testi che appartengono a due generi diversi: l'articolo di giornale e la prosa letteraria e paraletteraria. Anche i periodi sono diversi: nel primo caso i testi sono stati pubblicati nei primi anni 2000, nel secondo tra il XIX e il XXI secolo. In comune i due corpora hanno il fatto di comprendere esclusivamente testi scritti in lingua italiana e di includere al loro interno due subcorpora: uno composto di testi scritti originariamente in italiano, l'altro di traduzioni da altre lingue. Secondo la classificazione tratteggiata al § 1.2., si tratta dunque di corpora monolingui paragonabili.

Quando si costruisce un corpus, non è sufficiente raccogliere testi che abbiano una qualche caratteristica in comune: come abbiamo già detto al § 1.2. occorre garantire che il corpus sia rappresentativo di una certa varietà linguistica, il che comporta anche un bilanciamento interno. Immaginiamo che io voglia studiare l'italiano degli studenti universitari: potrei limitarmi a raccogliere un certo numero di esami scritti dai miei studenti, ma questo mi permetterebbe di avere solo una visione molto parziale del fenomeno, per una serie di ragioni; eccone alcune:

- non tengo conto della lingua orale;
- non tengo conto del passato, cioè di come può essere cambiata la lingua degli studenti universitari negli ultimi decenni o secoli;
- la maggior parte dei miei studenti sono in realtà studentesse, e il genere potrebbe incidere sulle caratteristiche della lingua;
- la maggior parte delle mie studentesse proviene dalle regioni settentrionali, quindi usa presumibilmente un italiano diverso dagli studenti del sud Italia;
- tutte le mie studentesse studiano lingue, quindi hanno una competenza specifica diversa da chi studia altre discipline;
- le prove d'esame sono esempi di registro formale, quindi non saprei nulla di come gli universitari si esprimono in contesti informali.

Se diamo per scontato che non esiste il corpus perfetto, la cosa migliore che si può fare è avere chiare in mente le caratteristiche del materiale che siamo riusciti a raccogliere (anche la fattibilità ha il suo peso: non sempre i testi che desidereremmo sono liberamente disponibili in formato digitale), compresi anche gli inevitabili difetti e lacune. Nell'ottica delle ricerche che intendiamo svolgere, potremo dunque valutare fino a che punto il corpus che analizziamo è adatto ai nostri scopi e quali sono invece i limiti di cui tener conto nella valutazione della validità dei risultati.

1.3.1. Il primo corpus giornalistico

Per studiare le caratteristiche del traduttese nell'italiano dei giornali, è stato costituito un corpus di articoli giornalistici frutto di

traduzione da testate straniere, che ha poi funto da modello di riferimento per la compilazione di un corpus paragonabile di articoli scritti originariamente in italiano. Occorre fare subito una precisazione a proposito della nozione di “articolo di giornale scritto originariamente in italiano”. Immaginiamo di avere un articolo scritto in italiano da un giornalista italiano, inviato a Washington per conto di una testata italiana. Il giornalista conosce l’inglese, vive in un contesto anglofono e scrive i suoi pezzi a partire da documenti e lanci di agenzia scritti in inglese: fino a che punto i suoi articoli sono “scritti in italiano”? Che peso potrebbe avere l’interferenza linguistica dell’inglese? Per tener conto di questi fattori, è stata introdotta la nozione di “traduzioni invisibili”, cioè testi che, pur non essendo vere e proprie traduzioni, possono risentire degli effetti del contatto linguistico.

Dobbiamo quindi tener conto del fatto che il subcorpus di controllo comprende in realtà testi che, proprio per gli argomenti trattati (principalmente questioni internazionali), possono essere il frutto della manipolazione di fonti in lingua straniera (principalmente l’inglese) e recare tracce evidenti dell’influsso di queste ultime. Ad ogni modo, questo subcorpus è stato costruito in modo da rispecchiare le principali caratteristiche del subcorpus delle traduzioni per quanto riguarda l’anno di pubblicazione degli articoli e, nei limiti del possibile, il genere testuale e gli argomenti trattati. A questo proposito, è importante notare che gli articoli tradotti (e di riflesso anche quelli scritti direttamente in italiano) non appartengono quasi mai alla cronaca politica e cittadina, cioè ai settori tradizionalmente considerati più rappresentativi della lingua dei giornali e quindi anche più studiati dai linguisti. Nel nostro caso si tratta invece perlopiù di editoriali e articoli di commento (rari i casi di reportage, intervista e lettera aperta) che trattano di grandi problemi globali o che analizzano la situazione politica, economica e culturale di paesi stranieri. In questi casi gli autori degli articoli sono selezionati soprattutto per la loro autorevolezza (in particolare nel caso delle traduzioni); di conseguenza ci si può attendere che i tratti specifici riconducibili all’autore prevalgono sulle caratteristiche tipiche della lingua dei giornali. Per evitare che l’analisi si risolvesse nello studio dello stile di un dato autore o traduttore, responsabile della maggioranza dei testi compresi nei subcorpora, si è quindi cercato di bilanciare il corpus anche in relazione agli autori/traduttori rappresentati:

dalla tabella 1.2 si evince una media di poco superiore alla decina di articoli per autore/traduttore.

Gli articoli raccolti sono stati pubblicati tra il 2001 e il 2008 nei quotidiani *Repubblica*, *Corriere della Sera*, e *Unità*. In questo modo i due subcorpora – che chiameremo per comodità “non tradotto” (o “originale”) e “tradotto” – sono a loro volta suddivisi in ulteriori tre subcorpora per ciascuna delle testate esaminate, per un totale di sei. Le caratteristiche di ciascun subcorpus e del corpus complessivo sono descritte nella tabella 1.2.

Il corpus comprende 1963 articoli, con una lunghezza media di 1.000 occorrenze, ma con dimensioni diverse che riflettono i diversi usi giornalistici e i modi di intendere l'editoriale: si va da un minimo di circa 200 occorrenze a un massimo di circa 4.000 occorrenze. Come si può notare dalla figura 1.2, che mostra la distribuzione degli articoli in base alla dimensione espressa in fasce, la grande maggioranza degli articoli varia tra 1.000 e 2.000 occorrenze, con un numero molto limitato di articoli brevissimi (meno di 500 occorrenze) e lunghissimi (più di 2.000 occorrenze).

Dal momento che le coppie di subcorpora sono state costruite in modo da avere la stessa dimensione, si nota che la lunghezza media degli articoli è variabile a seconda delle testate: gli articoli non tradotti del *Corriere della sera* sono più numerosi (e quindi mediamente più brevi), mentre sono meno numerosi (e quindi più lunghi) gli articoli tradotti di *Repubblica*. In generale, a parità di dimensione, gli articoli tradotti sono meno numerosi e dunque più lunghi.

La tabella 1.2 riporta dettagliatamente le lingue di partenza degli articoli tradotti, in cui spicca la netta predominanza dell'inglese. Le redazioni delle testate consultate hanno specificato che la lingua sorgente è sempre quella della fonte editoriale (quindi, per es., l'inglese per *The Independent*, il francese per *Le Monde*, lo spagnolo per *El País*, ecc.) e non quella dell'autore (quindi un ispanofono che ha pubblicato su un giornale tedesco viene tradotto dal tedesco). Per un elevato numero di testi, le fonti (cioè le banche dati online delle testate consultate) non specificavano la lingua di partenza, né citavano il nome del quotidiano che aveva originariamente pubblicato l'articolo. Nella maggior parte dei casi si tratta di testi acquisiti da agenzie stampa internazionali, per cui si può presumere che la lingua fonte fosse l'inglese. Nel dubbio, quando necessario le analisi riportano l'etichetta “lingua non determinata” o “n.d.”.

	Corriere tradotto	Corriere non tradotto	Repubblica tradotto	Repubblica non tradotto	Unità tradotto	Unità non tradotto	Subtotale tradotto	Subtotale non tradotto	Totale
Articoli	291	385	411	377	253	246	955	1.008	1.963
autori / traduttori	33	29	36	27	24	37	93	93	186
ling. or. italiano	-	385	-	377	-	246	-	1.008	1.008
ling. or. inglese	136	-	170	-	141	-	447	-	447
ling. or. francese	37	-	9	-	4	-	50	-	50
ling. or. spagnolo	12	-	25	-	6	-	42	-	42
ling. or. russo	1	-	-	-	-	-	1	-	1
ling. or. tedesco	3	-	7	-	-	-	10	-	10
ling. or. non det.	102	-	200	-	105	-	405	-	405

Tabella 1.2 – Numero di articoli del corpus in base alle principali variabili prese in considerazione

Il corpus di testi tradotti e quello di controllo sono stati costruiti in modo da risultare il più possibile omogenei dal punto di vista delle variabili prese in considerazione. Per es., per quanto riguarda la distribuzione cronologica, la figura 1.2 mostra come si sia raggiunto un sostanziale equilibrio nel numero di articoli per anno nei due sub-corpora, con differenze legate alla necessità di tenere conto della diversa lunghezza media degli articoli tradotti rispetto ai non tradotti.

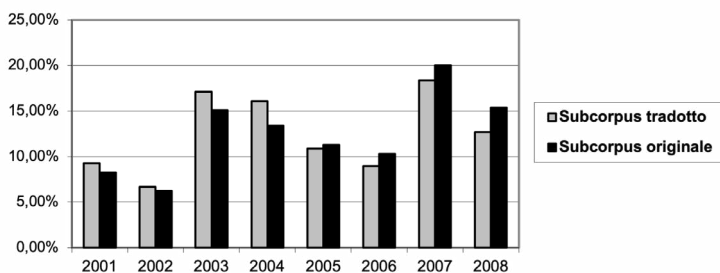


Figura 1.2 – Distribuzione degli articoli per anno. Valori percentuali

Al di là delle differenze rilevate nella lunghezza media degli articoli, per quanto possibile i due subcorpora sono stati bilanciati anche per quanto riguarda il numero di articoli per fasce di dimensione, come risulta dalla figura 1.3.

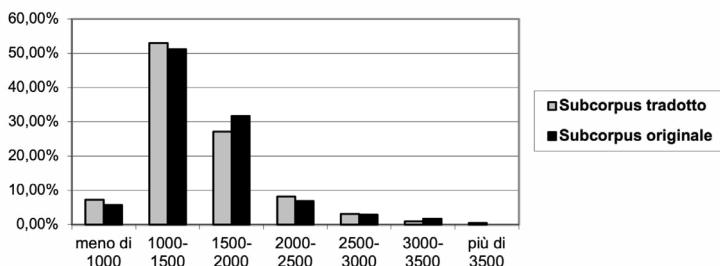


Figura 1.3 – Distribuzione degli articoli in base alla dimensione espressa in fasce. Valori percentuali

1.3.2. Il secondo corpus giornalistico: bilanciamento e campionamenti

Gli articoli di giornale sembrano offrire un ottimo banco di prova per utilizzare la distanza intertestuale come metodo per testare l'esistenza del traduttese. In italiano sono stati fatti dei tentativi di applicare questo metodo a testi di tipo letterario, che tuttavia sono forse i meno adatti a rilevare le conseguenze linguistiche del processo traduttivo in sé. È probabile, infatti, che un'opera letteraria, soprattutto se di prestigio tale da meritare più traduzioni successive, imponga la propria "impronta" individuale; il traduttore si sforzerà di renderne lo stile, magari dandone la propria resa in-

terpretativa, ma più difficilmente si abbandonerà agli automatismi traduttivi che invece potrebbe applicare nel caso di traduzioni più “dozzinali” e di minor prestigio.

Per tentare, invece, di cogliere le tendenze del traduttese, pare più opportuno considerare gli esiti di una produzione più rapida e “automatica”, come potrebbe essere la letteratura di consumo o il giornalismo. In particolare la stampa periodica, proprio per la rapidità di esecuzione di traduzioni che non puntano a esibire la loro origine esogena (*covert*), per il ruolo attualmente ricoperto di modello di riferimento per un italiano scritto di media formalità e per l’ampio pubblico raggiungibile, sembra offrire il materiale più adatto ad analisi intese a confermare o smentire la teoria degli universali traduttivi.

Si rende dunque necessario un corpus costruito in maniera tale da permettere il confronto tra i testi che lo compongono per controllare se la distanza intertestuale può rispondere ai seguenti quesiti di ricerca:

- se applichiamo il metodo della distanza intertestuale a un corpus comprendente testi tradotti in italiano a partire da diverse lingue e testi scritti direttamente in italiano, riusciamo a distinguere le traduzioni dai testi non tradotti?
- Con la distanza intertestuale riusciamo a raggruppare chiaramente i testi compresi in un corpus di sole traduzioni secondo la loro lingua fonte?

Il corpus illustrato nel § 1.3.2. si presta alle nuove analisi limitatamente alla componente dei testi non tradotti, mentre le traduzioni sono viziate dalla dominanza quantitativa quasi assoluta dell’inglese come lingua fonte. Si è reso dunque necessario compilare un nuovo subcorpus di 516 articoli tradotti da 22 lingue straniere e pubblicati sulla rivista *Internazionale* tra il 1993 e il 2009. Al termine della selezione alcune lingue di partenza, come naturalmente l’inglese, risultavano comunque più presenti e ne abbiamo limitato l’incidenza a circa 50 articoli (tabella 1.3):

Lingua	Numero di testi	Lingua	Numero di testi
francese	50	ungherese	22
spagnolo	49	norvegese	16
inglese	49	sloveno	12
russo	48	bulgaro	12
tedesco	48	romeno	10
polacco	38	coreano	9
ceco	37	finlandese	8
neerlandese	26	albanese	7
danese	25	estone	2
svedese	23	lituano	2
cinese	22	lettone	1

Tabella 1.3 – Lingue fonte del subcorpus di traduzioni: totale 516

Anche per quanto riguarda il periodo di pubblicazione (figura 1.4), la distribuzione dei testi tradotti presenta discrepanze, soprattutto nel confronto con i testi non tradotti. Tuttavia, data la limitatezza dell'arco temporale (17 anni), difficilmente si può pensare che si siano verificati degli sviluppi significativi della lingua; tutt'al più si potrebbero ipotizzare delle differenze importanti in termini contenutistici. In effetti, anche se gli articoli non tradotti sono stati selezionati in modo tale che i contenuti fossero quanto più possibile analoghi alle traduzioni (società, costumi, economia, eventi internazionali ecc.), poiché si concentrano tutti nel secondo decennio del periodo considerato, con tutta probabilità presenteranno argomenti assai diversi rispetto alle traduzioni degli anni 1993-2009. Ciononostante, con le adeguate procedure di bilanciamento e di campionamento illustrate tra breve, è possibile mitigare l'effetto dei contenuti sulla misurazione della distanza intertestuale.

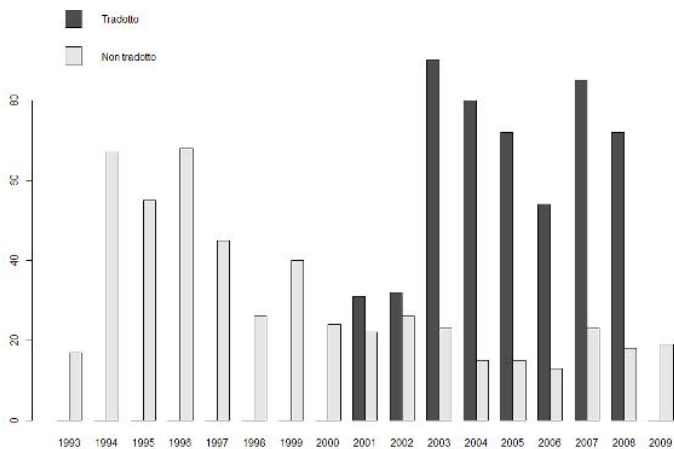


Figura 1.4 - Distribuzione temporale dei subcorpora: articoli per anno

Dopo queste integrazioni, il subcorpus degli articoli non tradotti comprende 1.008 testi composti da 93 autori per un totale di 997.047 occorrenze, mentre quello delle traduzioni conta 516 testi a nome di 67 traduttori per un totale di 632.059 occorrenze. Naturalmente vi sono sia autori sia traduttori che hanno firmato più articoli (il limite massimo è 65 per i giornalisti e 93 per i traduttori, il minimo è un solo testo). Tuttavia, nel caso dei traduttori, il fatto che ad alcune sigle siano attribuiti numerosi testi tradotti da un notevole numero di lingue straniere diverse (per es. le iniziali CP identificano 93 testi tradotti da ben 17 lingue che vanno dal cinese all'ungherese) impedisce di pensare che le iniziali identifichino una sola persona fisica: si tratta evidentemente di un'agenzia che si avvale di collaboratori diversi, con tutte le ovvie conseguenze sulla possibilità di stabilire la preminenza dell'autore originale o del traduttore nel calcolo della distanza intertestuale. Infatti, poiché per svariati testi non siamo in grado di identificare con certezza il traduttore, il nostro corpus non si dimostra adatto allo scopo. Restano comunque casi in cui la stessa persona traduce verosimilmente da lingue diverse ma imparentate: è il caso di Gronberg, che firma 13 articoli dal danese e 2 ciascuno da norvegese e svedese.

Alla luce della diversità delle dimensioni complessive risulta inutile fare confronti tra i subcorpora delle traduzioni e dei testi non tradotti. Emergono infatti grandi differenze a livello dei singoli testi per quanto riguarda sia la lunghezza (da un minimo di 216 occorren-

ze a un massimo di oltre 6.000) sia la ricchezza lessicale. Sebbene preveda un correttivo per tenerne conto, la formula del calcolo della distanza intertestuale resta sensibile alle dimensioni dei testi considerati e, di conseguenza, non sarà possibile procedere a confronti diretti tra i testi bensì occorrerà procedere a campionamenti.

Per realizzare un confronto tra testi non tradotti e traduzioni occorre procedere a un campionamento in grado di sterilizzare l'effetto dell'autore e del traduttore, della lingua di partenza e dei contenuti. Consideriamo quindi tutti i 516 testi tratti da *Internazionale* (tradotti dunque da 22 lingue diverse) e una selezione casuale di 516 articoli non tradotti, effettuata tramite un campionamento stratificato per testata e autore per evitare che si verifichi una qualsiasi preminenza e assicurare al tempo stesso la rappresentatività di tutte le categorie coinvolte. Si è così deciso di riunire i testi tradotti e non tradotti rispettivamente in 30 + 30 sottoinsiemi per ottenere dei macrotesti composti da 16 o 17 articoli ciascuno, cercando di evitare che si verificassero concentrazioni anomale di articoli dello stesso autore o traduttore o tradotti dalla stessa lingua di partenza. Poiché la lunghezza media degli articoli è di poco superiore alle mille occorrenze, ogni macrotesto risulta sufficientemente grande da permettere di eseguire 200 campionamenti di segmenti (*chunks*) di 3.500 occorrenze ciascuno.

Per semplificare, è come se confrontassimo 30 opere di un astratto "giornalista non tradotto modello" con 30 opere di un altrettanto astratto "traduttore modello". Per evitare che i contenuti e le diverse dimensioni dei testi incidano sul calcolo della distanza intertestuale, confrontiamo tra loro dei brani di 3.500 parole estratti casualmente da ciascuna opera, ripetendo l'estrazione per 200 volte in ciascuna opera. L'assunto è che, se la distanza intertestuale è in grado di accoppiare i testi (o gli estratti) di uno stesso autore, nel nostro caso dovrebbe essere in grado di distinguere nettamente testi non tradotti e traduzioni.

Il secondo dei quesiti della ricerca riguarda invece la possibilità di individuare la lingua di partenza misurando la distanza intertestuale tra diverse traduzioni. A questo scopo abbiamo selezionato dal nostro corpus tratto da *Internazionale* solo le lingue fonte che presentavano almeno 20 articoli (tabella 1.4), ottenendo un totale di 436 testi e 554.052 occorrenze che permette di coprire una discreta varietà di famiglie linguistiche:

Lingua	Frequenza	Lingua	Frequenza
inglese	49	neerlandese	26
spagnolo	49	danese	25
russo	48	cinese	23
tedesco	48	svedese	23
polacco	38	ungherese	22

Tabella 1.4 – Corpus per il confronto tra le lingue fonte

Sempre per ovviare al possibile impatto dovuto alla diversa dimensione, ai contenuti e al traduttore, gli articoli attribuiti a ciascuna lingua di partenza sono stati fatti confluire in 3 macrotesti in cui l'ordinamento dei singoli articoli è casuale: ciò significa che le lingue più rappresentate producono raccolte di oltre 15 articoli, l'ungherese di appena 7. Quindi procediamo al calcolo della distanza intertestuale secondo il metodo di campionamento già visto sopra nel confronto tra traduzioni e testi non tradotti. In altre parole, in questo caso è come se avessimo tre opere di dodici “traduttori modello” corrispondenti alle lingue fonte (tre opere del traduttore francese, tre opere del traduttore inglese e così via), quindi calcoliamo le distanze intertestuali all'interno del corpus per vedere se i campioni estratti dai testi tradotti dalla stessa lingua fonte risultano più vicini.

1.3.3. *Il corpus letterario*

Per verificare l'effetto del tipo di testo sulle caratteristiche del traduttese, abbiamo deciso di rivolgere la nostra attenzione alla prosa letteraria e paraletteraria e abbiamo compilato un corpus che ci permettesse di eseguire le stesse analisi di tipo quantitativo a cui abbiamo sottoposto il corpus giornalistico illustrato al § 1.3.1. Oltre che per le considerazioni già svolte al § 1.2.3., non è stato invece possibile adottare il metodo della distanza intertestuale dato il numero ridotto di romanzi e racconti, come pure di autori e traduttori; inoltre, questi testi non si prestano al campionamento perché al loro interno sono molto meno omogenei degli articoli di

giornale: se estraiamo *chunks* di qualche migliaio di parole da un romanzo potremmo ritrovarci con segmenti testuali assai diversi tra loro, per es. dialoghi e descrizioni, che risulterebbero molto distanti pur essendo frutto della stessa penna, mentre potrebbero essere classificati come simili ad altri estratti da opere diverse. Tuttavia, è stato possibile considerare un ulteriore fattore che restava escluso dall'analisi degli articoli giornalistici: i cambiamenti nel corso del tempo. Infatti, è risultato piuttosto agevole raccogliere testi prodotti negli ultimi due secoli circa, in modo tale da poter confrontare le caratteristiche del traduttese (ma anche della lingua non risultante dal processo traduttivo) in diversi periodi storici.

È stato quindi raccolto un corpus di testi in prosa letteraria e paraletteraria, sia tradotti che scritti direttamente in italiano, pubblicati in volume tra il 1811 e il 2005. La scelta di includere esempi di paraletteratura (anche detta "letteratura di consumo") nell'analisi ha due ragioni. In primis, dal punto di vista stilistico, si tratta di un tipo di prosa caratterizzata da una lingua meno sorvegliata e più vicina alla lingua d'uso. In seconda battuta, come abbiamo avuto modo di ribadire più volte, è probabile che il processo traduttivo lasci maggiori tracce (perché più rapido e oggetto di minor riflessione stilistica) in testi considerati meno prestigiosi dal punto di vista culturale e destinati a un consumo massificato.

Per questioni di bilanciamento, in gran parte determinate dalla disponibilità delle opere in formato elettronico, è stata adottata la seguente periodizzazione: il XIX secolo è stato considerato come un blocco unitario (circa 860.000 parole di opere in traduzione e altrettante di opere non tradotte) per avere materiali sufficienti per l'analisi (stante la difficoltà di reperimento dei testi) e in considerazione del predominio linguistico e culturale francese su un'Italia ancora frammentata e scarsamente alfabetizzata. I testi editi nel XX secolo sono stati invece suddivisi secondo tre fasi cronologiche: il periodo comprendente le due guerre mondiali (1900-1945; circa 1.200.000 parole per ciascun subcorpus), quando alle influenze esogene (ma anche alle variazioni regionali) si è opposto gradualmente il purismo fascista; il periodo dello sviluppo economico postbellico, caratterizzato dall'incremento dell'alfabetizzazione e dell'italianizzazione della popolazione, dai flussi migratori, dall'avvento della televisione e dalle nuove questioni linguistiche (1946-1975; circa 915.000 parole per ciascun subcorpus); un periodo conclusivo, comprendente anche il primo quinquennio del XXI se-

colo (1976-2005; circa 1.200.000 parole per ciascun subcorpus), caratterizzato dalla progressiva oralizzazione dell'italiano in un contesto di crescente globalizzazione anglicizzante.

Naturalmente, non solo i criteri storico-sociali alla base di questa partizione cronologica sono piuttosto opinabili, ma il corpus stesso presenta degli squilibri significativi (cfr. le tabelle dalla 1.5 alla 1.14; le date delle traduzioni si riferiscono alla pubblicazione in Italia), sempre a causa della difficoltà di reperimento dei testi in formato elettronico. Tali squilibri tra i diversi subcorpora riguardano le dimensioni (da un minimo di circa 860.000 parole a un massimo di oltre 1.200.000), la composizione interna (da un minimo di 8 opere di 8 autori diversi a un massimo di 19 opere di 15 autori diversi; inoltre i testi sono di lunghezze diverse e alcuni autori sono presenti con più di una sola opera, anche in diversi subcorpora, per es. Dickens, Hugo, Scott e Turgenev) e la distribuzione cronologica (la prima opera non tradotta risale al 1860, mentre la prima traduzione è del 1811; l'inglese è di gran lunga la lingua fonte più presente, ma non sempre con la stessa dominanza in tutti i periodi considerati). Si tratta comunque di un corpus che può essere considerato più bilanciato di quelli utilizzati in ricerche precedenti e presenta anche il vantaggio di includere traduzioni da lingue diverse (francese, russo e tedesco nel XIX secolo e giapponese, spagnolo e tedesco nel XX secolo) e opere di letteratura di consumo, anche se definire precisamente quali titoli e autori appartengano al canone letterario non è semplice, soprattutto nel caso delle traduzioni e man mano che ci si avvicina ai nostri giorni: se Verga e Dickens possono essere sicuramente considerati esempi di letteratura "alta", non si può però dire che Salgari e Verne non ricoprano un ruolo importante nella storia letteraria italiana ed europea. A questo proposito, poiché il subcorpus non tradotto relativo al periodo 1946-1975 risultava composto esclusivamente da opere e autori appartenenti al canone, mentre il subcorpus corrispondente di traduzioni (peraltro esclusivamente dall'inglese) sembra piuttosto appartenere alla paraletteratura (perlopiù romanzi gialli), per verificare l'impatto del genere letterario sull'assetto linguistico del traduttese abbiamo analizzato anche un ulteriore subcorpus compilato da Marianna De Biasi (che ringraziamo), accuratamente bilanciato, e composto esclusivamente di letteratura di consumo (romanzi gialli, rosa, di fantascienza e del terrore) pubblicata tra il 1950 e il 1999.

Autore	Titolo	Anno	Occorrenze
I. Nievo	<i>Il barone di Nicastro</i>	1860	54.453
G. Verga	<i>Storia di una capinera</i>	1871	31.529
C. Collodi	<i>Pinocchio</i>	1881	40.670
R. Fucini	<i>Le veglie di Neri: paesi e figure della campagna toscana</i>	1882	43.829
E. De Amicis	<i>Cuore</i>	1886	84.055
E. De Marchi	<i>Il cappello del prete</i>	1888	57.486
G. D'Annunzio	<i>Il piacere</i>	1889	105.195
L. Capuana	<i>Profumo</i>	1892	59.565
F. De Roberto	<i>I viceré</i>	1894	145.781
A. Fogazzaro	<i>Piccolo mondo antico</i>	1895	112.456
E. Salgari	<i>I pirati della Malesia</i>	1896	58.934
G. Deledda	<i>La via del male</i>	1896	67.091

Tabella 1.5 – Opere di autori italiani 1800-1899. N= 861.044

Autore	Titolo	Anno	Occorrenze	Traduttore
J,P. Claris de Florian	<i>Numa Pompilio secondo re di Roma (vol. 1)</i>	1811	19.795	n.n.
L. Sterne	<i>Viaggio sentimentale di Yorick lungo la Francia e l'Italia</i>	1813	43.500	D. Chierico
F.A. Chateaubriand	<i>Atala ovvero Gli amori di due selvaggi nel deserto</i>	1814	23.900	n.n.
F. Schiller	<i>L'oste del sole ovvero Il delinquente per onore perduto</i>	1831	7.891	C. Belligoni
W. Scott	<i>Ivanhoe (vol. 3)</i>	1832	34.612	A. Clerichetti
H. de Balzac	<i>Papà Goriot</i>	1835	85.448	L.M.
V. Hugo	<i>Le ore estreme di un sentenziato a morte</i>	1835	39.367	G.B. Carta
W. Scott	<i>Il castello pericoloso</i>	1843	38.537	L.M.
A.R. Lesage	<i>Storie di Gil Blas di Santillano (vol. 1)</i>	1853	137.639	Q. Viviani
A. Dumas	<i>I tre moschettieri (vol. 1)</i>	1855	43.844	C. Coriani
C. Dickens	<i>Le campane ovvero Il capo d'anno</i>	1855	50.924	P. Bettoni
W. Goethe	<i>I dolori del giovane Werther</i>	1858	52.044	R. Ceroni
E.A. Poe	<i>Storie incredibili (L'uomo della folla)</i>	1869	55.365	B.E. Maineri
L. Carroll	<i>Alice nel paese delle meraviglie</i>	1872	22.687	T. Pietrocola Rossetti
I.S. Turgenev	<i>Acque di primavera</i>	1872	45.718	S. de Gubernatis-Besobrasoff
J. Verne	<i>Dalla terra alla luna. Tragitto in 97 ore e 20 minuti</i>	1872	52.248	n.n.
V. Hugo	<i>Novantatré (vol. 2)</i>	1874	40.412	C. Pizzigoni
I.S. Turgenev	<i>Il primo amore e Assia</i>	1876	39.276	E. Zucchelli
C. Dickens	<i>Cantico di Natale</i>	1888	26.815	F. Verdinois

Autore	Titolo	Anno	Occorrenze
A. Fogazzaro	<i>Piccolo mondo moderno</i>	1901	92.355
G. D'Annunzio	<i>Le novelle della Pescara</i>	1902	50.123
L. Pirandello	<i>Il fu Mattia Pascal</i>	1904	75.010
E. Salgari	<i>Gli ultimi filibustieri</i>	1908	90.030
E. Salgari	<i>Il figlio del Corsaro Rosso</i>	1908	109.410
Vamba	<i>Il giornalino di Gian Burrasca</i>	1912	74.340
E. Salgari	<i>La rivincita di Yanez</i>	1912	80.225
G. Deledda	<i>Canne al vento</i>	1913	60.430
L. Pirandello	<i>I vecchi e i giovani</i>	1913	109.430
F. Tozzi	<i>Bestie</i>	1917	15.624
F. Tozzi	<i>Con gli occhi chiusi</i>	1919	47.445
F. Tozzi	<i>Tre croci</i>	1920	28.272
I. Svevo	<i>La coscienza di Zeno</i>	1923	145.433
L. Pirandello	<i>Uno nessuno centomila</i>	1926	46.414
I. Svevo	<i>Una burla riuscita</i>	1926	21.330
L. Pirandello	<i>Una giornata</i>	1937	30.319
C. Levi	<i>Cristo si è fermato a Eboli</i>	1945	81.122
E. Vittorini	<i>Uomini e no</i>	1945	48.668

Autore	Titolo	Anno	Occorrenze	Traduttore
L. Tolstoj	<i>Anna Karenina</i>	1901	327.334	E. V. Foulques
E. Bronte	<i>Cime tempestose</i>	1926	117.562	R. Binetti
F. Dostoevskij	<i>L'idiota</i>	1927	245.393	F. Verdinois
C. Dickens	<i>Il circolo Pickwick</i>	1930	296.865	F. Verdinois
N. Hawthorne	<i>La lettera scarlatta</i>	1930	80.331	F.M. Martini
R. Stout	<i>La lega degli uomini spaventati</i>	1937	57.799	A. Pitta
A. Christie	<i>Il Natale di Poirot</i>	1940	50.929	E. Piceni
Stendhal	<i>La badessa di Castro</i>	1942	34.392	P.P. Trompeo

Tabella 1.8 – Traduzioni 1900-1945. N= 1.210.605

Autore	Titolo	Anno	Occorrenze
E. Flaiano	<i>Tempo di uccidere</i>	1947	83.113
C. Malaparte	<i>La pelle</i>	1949	111.759
A. Moravia	<i>Il conformista</i>	1951	99.957
I. Calvino	<i>Il barone rampante</i>	1957	71.189
F. Tomasi Di Lampedusa	<i>Il Gattopardo</i>	1958	74.657
O. Ottieri	<i>Donnarumma all'assalto</i>	1959	60.914
P. Chiara	<i>Il piatto piange</i>	1962	45.531
N. Ginzburg	<i>Lessico familiare</i>	1963	65.496
A. Bevilacqua	<i>L'occhio del gatto</i>	1968	86.219
L. Romano	<i>Le parole tra noi leggere</i>	1969	94.795
G. Piovene	<i>Stelle fredde</i>	1970	21.030
D. Maraini	<i>Memorie di una ladra</i>	1972	99.340

Tabella 1.9 – Opere di autori italiani 1946-1975. N= 914.000

Autore	Titolo	Anno	Occorrenze	Traduttore
K. Millar	<i>La città del diavolo</i>	1950	53,264	A. Tedeschi
E. Queen	<i>Il re è morto</i>	1954	57,556	B. Tasso
R. Stout	<i>Nero Wolfe e i ragni d'oro</i>	1954	42,760	I. Omboni
E.S. Gardner	<i>Perry Mason e la sveglia sotterrata</i>	1955	44,599	G. Gogioso
E. Queen	<i>Il villaggio di vetro</i>	1957	78,082	G. Camia
U. Curtiss	<i>Ombre di cera</i>	1958	48,071	P. Soleri
C.E. Maine	<i>Rischio calcolato</i>	1961	45,829	A. Negretti
F. Brown	<i>Gli strani suicidi di Bartlesville</i>	1962	47,756	M. Galli
M. Leinster	<i>Questo è un Gizmo</i>	1962	38,019	B. Russo
R. Stout	<i>Nero Wolfe fa la spia</i>	1964	48,133	L. Grimaldi
J.H. Chase	<i>Canaglia cercasi</i>	1964	53,790	H Brinis
I. Levin	<i>Rosemary's Baby</i>	1967	60,716	A. Veraldi
D. Koontz	<i>Jumbo-10. Il rinnegato</i>	1969	32,488	B. Della Frattina
M. Zimmer Bradley	<i>Il drago di bronzo</i>	1972	45,663	M.B. De Castiglione
C.D. Simak	<i>Pescatore di stelle</i>	1973	75,892	R. Rambelli
M.G. Eberhart	<i>Il giudice buonanima</i>	1974	68,482	C. Merlo
R. Silverberg	<i>Vacanze nel deserto</i>	1975	77,696	G. Tamburini

Tabella 1.10 – Traduzioni 1946-1975. N= 918.796

Autore	Titolo	Anno	Occorrenze
F. Cialente	<i>Le quattro ragazze Wieselberger</i>	1976	67.486
F. Tomizza	<i>La miglior vita</i>	1977	94.441
A. De Carlo	<i>Uccelli da gabbia e da voliera</i>	1982	70.610
C. Sgorlon	<i>L'armata dei fiumi perduti</i>	1985	88.049
U. Eco	<i>Il pendolo di Foucault</i>	1988	205.964
A. Baricco	<i>Castelli di rabbia</i>	1991	53.052
S. Veronesi	<i>Venite venite B-52</i>	1995	95.519
N. Ammaniti	<i>Fango</i>	1996	33.936
M. Buticchi	<i>Le pietre della luna</i>	1997	120.641
G. Faletti	<i>Io uccido</i>	2002	199.268
M. Mazzantini	<i>Non ti muovere</i>	2002	76.093
V.M. Manfredi	<i>L'impero dei Draghi</i>	2005	121.571

Tabella 1.11 – Opere di autori italiani 1976-2005. N= 1.226.630

Autore	Titolo	Anno	Occorrenze	Traduttore
A. Christie	<i>Addio miss Marple</i>	1976	55,688	D. Fonticoli
M. Crichton	<i>Congo</i>	1980	89,529	E. Capriolo
S. King	<i>Pet cemetery</i>	1985	132,829	H. Brinis
P. Süskind	<i>Il profumo</i>	1988	81,071	G. Agabio
R. Dahl	<i>Storie impreviste</i>	1993	90,942	A. Veraldi
H. Murakami	<i>Dance dance dance</i>	1998	116,264	G. Amitrano
H. Corbin	<i>Cadaveri senza volto</i>	2000	104,456	F. Sessi
K. Follett	<i>Le gazze ladre</i>	2001	136,296	A. Raffo
I. Allende	<i>Il regno del drago d'oro</i>	2003	86,493	E. Liverani
C. Ruiz Zafón	<i>L'ombra del vento</i>	2004	134,436	L. Sezzi
T. Yamada	<i>Estranei</i>	2005	45,949	A. Martini
P. Coelho	<i>Lo Zahir</i>	2005	81,235	R. Desti

Tabella 1.12 - Traduzioni 1976-2005. N= 1.155.188

Autore	Titolo	Anno	Occorrenze
L. R. Johannis	<i>C'era una volta un pianeta...</i>	1954	38.662
L. R. Johannis	<i>Quando ero aborigeno</i>	1955	48.470
Liala	<i>Le briglie d'oro</i>	1955	64.272
S. Donati	<i>Il sepolcro di carta</i>	1956	46.258
M. Soldati	<i>Storie di spettri</i>	1962	62.864
U. Malaguti	<i>Il sistema del benessere</i>	1965	45.264
G. Scerbanenco	<i>Europa molto amore</i>	1966	51.417
G. Scerbanenco	<i>I milanesi ammazzano al sabato</i>	1969	41.990
G. Montanari	<i>La sepoltura</i>	1973	39.170
G. De Maria	<i>Le venti giornate di Torino</i>	1977	31.250
L. Aldani	<i>Eclissi 2000</i>	1979	27.262
C. Fruttero & F. Lucentini	<i>A che punto è la notte</i>	1979	102.056
L. Macchiavelli	<i>La strage dei centauri</i>	1980	46.189
L. Aldani e D. Piegai	<i>Nel segno della luna bianca</i>	1985	59.262
S. Casati Modignani	<i>Disperatamente Giulia</i>	1986	91.770
T. Sclavi	<i>Dellamorte Dellamore</i>	1991	26.467
T. Sclavi	<i>Nero</i>	1992	22.953
C. Lucarelli	<i>Falange armata</i>	1993	33.710
N. Vallorani	<i>Dream Box</i>	1997	67.420
C. Lucarelli	<i>L'isola dell'angelo caduto</i>	1999	53.013

Tabella 1.13 – Paraletteratura. Opere di autori italiani 1950-1999. N= 999.719

Autore	Titolo	Anno	Occorrenze	Traduttore
E. Queen	<i>Il re è morto</i>	1954	56.046	B. Tasso
A.C. Clarke	<i>La città e le stelle</i>	1957	75.410	H. Brinis
P. Hoesl	<i>Decidi, mio cuore!</i>	1958	64.822	n.n.
E. McBain	<i>Chiamate Frederick 7- 8024</i>	1961	42.895	A. Negretti
F. Brown	<i>Gli strani suicidi di Bartlesville</i>	1962	46.861	M. Galli
E. McBain	<i>Lungo viaggio senza ritorno</i>	1964	49.301	A. Negretti
I. Levin	<i>Rosemary's Baby</i>	1968	59.068	A. Veraldi
C. Rauch	<i>L'inquietante signora del piano di sopra</i>	1976	76.360	M. Manciola Billi
A. Norton	<i>Gli esuli delle stelle</i>	1978	66.861	R. Rambelli
R. Rendell	<i>Sulle orme di un'ombra</i>	1979	54.874	n.n.
O. Butler	<i>Incidente nel deserto</i>	1985	56.786	B. della Frattina
D. Steel	<i>Due mondi due amori</i>	1985	94.999	G. M. Griffini
R. Crais	<i>Corrida a Los Angeles</i>	1989	48.922	n.n.
R. Bloch	<i>H gusto del fuoco</i>	1991	46.576	M. Tosello
I. Banks	<i>Canto di pietra</i>	1999	68.912	M. Birattari
J. C. Grangé	<i>I fiumi di porpora</i>	1999	91.688	I. Landolfi

Tabella 1.14 - Paraletteratura. Traduzioni 1950-1999. N= 1.000.381

2. Il traduttese nei giornali

2.1. Obiettivi e strumenti della ricerca

In questo capitolo illustreremo le ricerche che hanno avuto come oggetto i corpora e subcorpora descritti nei §§ 1.3.1. e 1.3.2., mentre le tecniche di indagine sono quelle proprie della *corpus linguistics* e dell'analisi statistica di dati testuali a cui abbiamo fatto accenno al § 1.2. La prima parte della ricerca è mirata a identificare le caratteristiche del traduttese e la preparazione dei materiali si è limitata alla pulizia grafica dei testi secondo la funzione della "normalizzazione" (identificazione dei confini di parola e regolarizzazione di apostrofi, accenti ecc.) del software *Taltac2*, mentre il *tagging* (la lemmatizzazione, cioè la riduzione di tutte le parole ai lemmi corrispondenti, e l'assegnazione alle varie parti del discorso, come nomi, verbi, aggettivi ecc.) è stato fatto con il programma *TreeTagger*.

Per agevolare la presentazione dei dati e accorparli secondo categorie omogenee, nelle pagine che seguono i fenomeni sono classificati secondo la tradizionale quadripartizione dell'analisi linguistica: lessico, morfologia, sintassi e testualità. Questo approccio taglia trasversalmente le categorie a cui i *translation studies* attribuiscono per tradizione i tratti indagati (cioè la legge dell'interferenza e i quattro universali traduttivi visti al § 1.1). Per recuperare una visione d'insieme, in un secondo momento si presenteranno i

principali fenomeni individuati secondo le direttrici fornite dagli universali traduttivi (§ 2.7).

A questo proposito, le analisi statistiche saranno dedicate principalmente agli aspetti più legati alla tendenza alla semplificazione, esplicitazione, normalizzazione e convergenza tramite una serie di misurazioni in grado di valutare l'assetto complessivo di un corpus testuale, come le misure lessicometriche, l'incidenza del vocabolario di base, la distribuzione delle parti del discorso e la densità lessicale, la lunghezza dei periodi e l'uso della punteggiatura. A queste possiamo aggiungere le analisi basate su liste che riguardano l'impiego dei connettivi (presumibile spia della tendenza all'esplicitazione). Per misurare invece il conservatorismo linguistico dei traduttori, si è deciso di concentrarsi sui tratti dell'italiano dell'uso medio che più si prestano all'analisi automatica, come la frequenza di allotropi colti, la costruzione del periodo ipotetico con il doppio imperfetto, la frequenza del congiuntivo e del condizionale, lo sfruttamento del *che* esclamativo e interrogativo.

Altre misurazioni, invece, si concentreranno maggiormente sull'interferenza della lingua fonte delle traduzioni, pur non escludendo altri possibili influssi: per il lessico si tratta della presenza di forestierismi, calchi semantici ed espressioni politicamente corrette; a livello morfologico si darà conto della frequenza del perfetto semplice, dei pronomi personali e pronomi e aggettivi possessivi e dimostrativi; per quanto concerne la sintassi, ci si concentrerà sui costrutti *ad sensum* e sui calchi sintattici. Alcune di queste analisi si basano su liste e richiedono un intervento attivo da parte del ricercatore (per es., nel caso dei forestierismi, si pensi al problema degli omografi) e concedono aperture maggiori a un'analisi più marcatamente qualitativa del corpus.

Naturalmente, benché l'interesse principale della ricerca sia l'identificazione del cosiddetto "traduttese", alcuni dei dati ottenuti saranno utili per formulare osservazioni e ipotesi che riguardino le caratteristiche dell'italiano dei giornali nel suo complesso, con particolare riferimento a una tipologia testuale poco studiata in quanto non compresa nel pastone politico o nella cronaca, che rappresentano lo "zoccolo duro" dell'italiano giornalistico. Se si tiene poi conto del fatto che i giornali ormai sono considerati il modello di riferimento per individuare tratti e tendenze dell'italiano d'oggi, per lo meno nell'uso quotidiano, alcune delle osservazioni tratte dal nostro corpus potrebbero dimostrarsi interessanti per co-

gliere alcuni aspetti che riguardano la lingua contemporanea nel suo complesso.

2.2. Il lessico

Il livello lessicale verrà innanzitutto indagato da una prospettiva puramente quantitativa tramite misure lessicometriche ottenute automaticamente, come la ricchezza lessicale e, dopo aver individuato le parti del discorso, la densità lessicale e l'incidenza del Vocabolario di Base, cioè le circa 7.000 parole più usate nell'italiano. La possibile interferenza dei testi fonte verrà invece valutata con riferimento alla presenza di forestierismi, calchi sintattici ed espressioni politicamente corrette.

2.2.1. Le misure lessicometriche

I primi dati in grado di caratterizzare i subcorpora consistono nelle misure lessicometriche (tabella 2.1), Per limitarsi alle misure principali, il corpus di articoli tradotti presenta, rispetto a quello di articoli non tradotti, una minore percentuale di parole diverse (V/N%, cioè il *Type-Token Ratio* visto al § 1.2.1.) e di parole usate una sola volta (Hapax%), oltre a una più elevata frequenza media delle forme (N/V). Questi dati – tanto più significativi in quanto calcolati su subcorpora della stessa dimensione e confermati dal confronto per singoli subcorpora omogenei – mostrano una ricchezza lessicale lievemente inferiore dei testi tradotti, in linea con quanto ci si poteva attendere dal punto di vista teorico per via dell'azione dell'universale traduttivo della semplificazione. In quest'ottica, la minore ricchezza lessicale dei testi tradotti può essere interpretata come il risultato della tendenza dei traduttori a ripetere più spesso un numero ridotto di forme grafiche. L'analisi dei dati sulla ricchezza lessicale per testate giornalistiche (dunque, ancora una volta, tra subcorpora della stessa dimensione) rivela inoltre come lo scarto tra testi originali e tradotti sia maggiore nel caso di *Repubblica*. In assenza di altre cause apparenti, si tratta del probabile segno di uno stile brillante e ricco dal punto di vista lessicale negli articoli scritti in italiano, che contrasta con quello lessicalmente più omogeneo dei testi tradotti più di quanto non avvenga nel *Corriere della Sera* e nell'*Unità* (cfr. anche i dati sui forestierismi al § 2.2.4).

	Corriere tradotto	Corriere non tradotto	Repubblica tradotto	Repubblica non tradotto
N	294.111	292.545	436.098	436.816
V	29.585	32.733	35.965	46.256
V/N	0,10	0,11	0,08	0,11
V/N%	10,06	11,19	8,25	10,59
Hapax%	50,45	53,18	48,51	53,31
N/V	9,94	8,94	12,13	9,44

Tabella 2.1 – Misure lessicometriche del corpus

Un'ulteriore indagine che è stato possibile affrontare è stata l'analisi della variabilità delle principali misure lessicometriche per articolo. Dal momento che, come si è notato, la lunghezza dei singoli articoli è molto variabile (figura 1.3.), si è proceduto a questa misurazione su campioni di 1.000 occorrenze (la lunghezza media degli articoli del corpus) in una scelta di 80 articoli (40 scritti originariamente in italiano confrontati con 40 traduzioni) rappresentativi delle variabili che caratterizzano il corpus (anno, autori/traduttori diversi, lingue di partenza degli articoli tradotti). La necessità di procedere con campioni di uguale lunghezza – escludendo quindi gli articoli più brevi e campionando solamente una porzione di quelli più lunghi della media – è legata al fatto che le misure lessicometriche sono sensibili alla dimensione: se non si operasse per campioni omogenei, i dati rifletterebero la variabilità di lunghezza degli articoli e non sarebbe possibile operare confronti. In altri termini, in questo modo si ha la garanzia che i dati riflettano effettive differenze linguistiche tra gli articoli.

La figura 2.1. presenta sotto forma di *box-plot* la distribuzione dei valori in percentuale del rapporto tra forme e occorrenze (V/N%) degli articoli del campione esaminato. Per interpretare i dati dei grafici *box-plot* si tenga conto che i valori si leggono sull'asse verticale: la riga al centro del box segnala la mediana

Unità tradotto	Unità non tradotto	Subtotale tradotto	Subtotale non tradotto	Totale
260.426	260.900	994.149	997.047	1.991.196
23.884	27.223	54.204	65.019	86.373
0,09	0,10	0,05	0,07	0,04
9,17	10,43	5,45	6,52	4,34
49,92	50,98	45,16	47,59	45,14
10,90	9,58	18,34	15,33	23,05

(ciò significa che al di sotto di questo valore c'è il 50% delle osservazioni e al di sopra il restante 50%); il bordo inferiore del box rappresenta il primo quartile (ciò significa che al di sotto di quel valore c'è il 25% delle osservazioni con i valori più bassi) e il bordo superiore della scatola è il terzo quartile (sopra a quel valore c'è il 25% delle osservazioni con valori più alti); le linee tratteggiate che proseguono in verticale prima e dopo ciascun box evidenziano come si estendono i dati e in questo caso mostrano come la variabilità degli articoli tradotti sia più contenuta di quella degli originali.

Come si può notare, le traduzioni hanno valori di V/N% mediamente inferiori e con minore variabilità rispetto agli articoli non tradotti. In dettaglio, gli articoli originali del campione presentano una media di V/N% pari a 51,41, contro 50,05 degli articoli tradotti. Anche la forbice tra i valori minimi e massimi di V/N% è più ampia negli originali (rispettivamente 44,85 e 59,08) rispetto agli articoli tradotti (rispettivamente 45,55 e 54,15). Questa minore variabilità è confermata dalla *standard deviation*, più bassa nei testi tradotti del campione (2,67) rispetto a quelli originali (3,24).

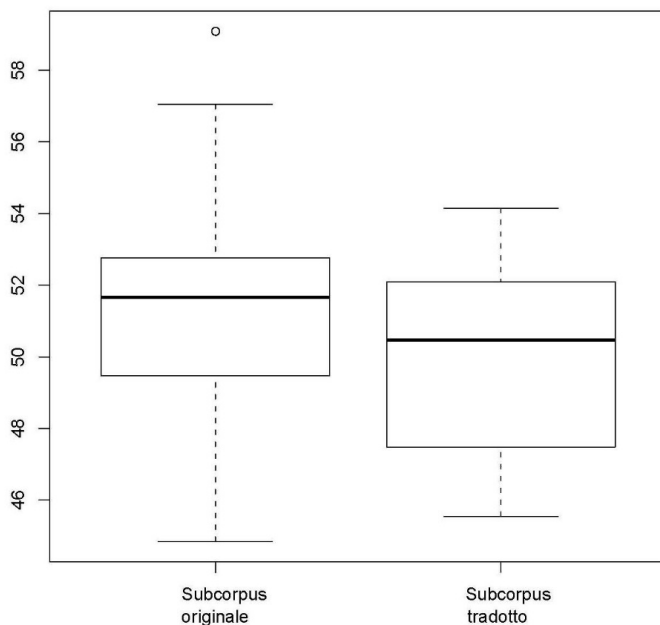


Figura 2.1 – Box-plot con la distribuzione dei valori di $V/N\%$ negli articoli di uguale dimensione del campione

Questi dati possono essere interpretati come un indizio del fatto che il corpus in esame confermi l'universale traduttivo della convergenza, secondo il quale le traduzioni “si assomiglierebbero” tra loro di più, mentre i testi non tradotti presentano delle differenze maggiori. Risultati analoghi si ottengono anche esaminando nei singoli articoli del campione la distribuzione del rapporto tra numero di forme presenti una sola volta nel corpus (Hapax) e totale di forme moltiplicato per 100 (figura 2.2). Anche in questo caso gli articoli tradotti hanno valori di Hapax% mediamente inferiori e con minore variabilità rispetto agli originali. In dettaglio, gli articoli originali del campione presentano una percentuale media di hapax pari a 74,34, contro 73,76 degli articoli tradotti. Anche la forbice tra i valori minimi e massimi di Hapax% è più ampia negli originali (rispettivamente 63,98 e 80,31) rispetto agli articoli tradotti (rispettivamente 67,54 e 77,97). Questa minore variabilità è confermata dalla *standard deviation*, più bassa nei testi tradotti del campione (2,88) rispetto a quelli originali (3,89).

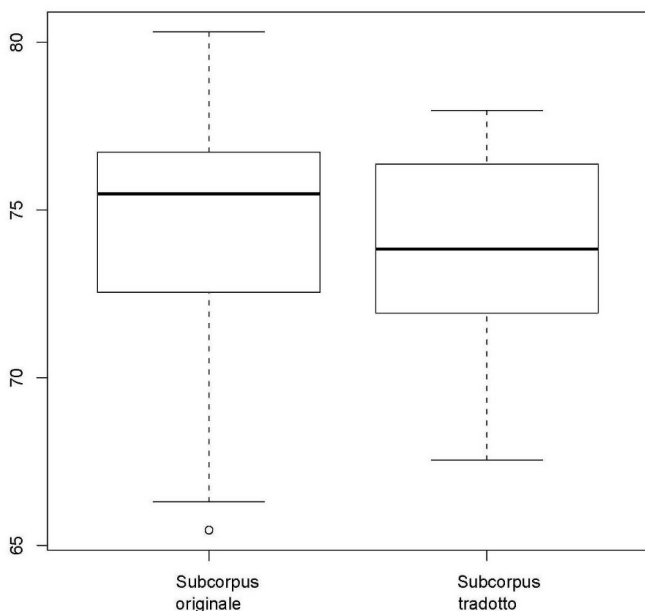


Figura 2.2 - Box-plot con la distribuzione dei valori di Hapax% negli articoli di uguale dimensione del campione

2.2.2. Il vocabolario di base

Se i dati relativi alle misure lessicometriche forniscono indicazioni sull'uso più o meno ricco e variato delle risorse lessicali, uno strumento che consente di far emergere anche qualche informazione sull'effettiva qualità del lessico è il *Vocabolario di base* (o *VdB*). Il presupposto è che la percentuale di lemmi di un testo che rientra tra quelli di base riesce a dare un'idea del livello di complessità del lessico di quel testo.

La distribuzione del lessico tra *VdB* e non-*VdB* mostra una maggiore frequenza di forme di base nelle traduzioni: come si può notare nella figura 2.3, la percentuale di forme appartenenti al *VdB* è di due punti percentuali superiore nei testi tradotti rispetto a quelli non tradotti. Si tratta di una differenza minima, ma che può trovare una spiegazione nel fatto che le scelte lessicali dei traduttori – presumibilmente sia per influsso dello stile dei testi in lingua originale sia per effetto della tendenza alla semplificazione e normalizzazione – si collocano su un piano di maggiore 'medietà' lessicale rispetto allo stile brillante dei giornalisti italiani.

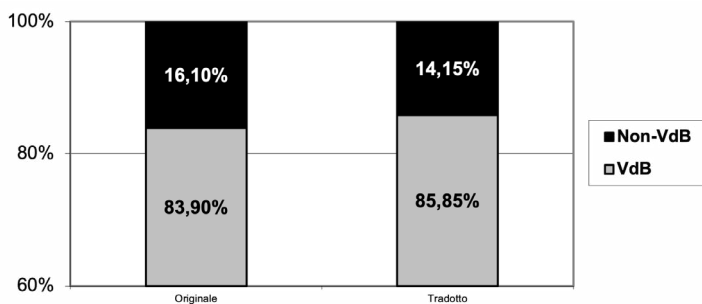


Figura 2.3 – Distribuzione percentuale delle forme appartenenti al VdB nei subcorpora non tradotto e tradotto

Il dato è confermato anche dalla tabella 2.2, che riporta le densità per 1.000 forme di lemmi di base nei vari subcorpora e mostra come, tra coppie di subcorpora della medesima testata, siano sempre gli articoli tradotti quelli con una maggior densità di forme appartenenti al VdB. In questa tabella e nelle successive, indicheremo con la sigla “n” le frequenze assolute dei dati, con “%” le frequenze percentuali e con “d” la densità per 1.000 forme rispetto a ciascuno dei rispettivi subcorpora.

	Corriere tradotto	Corriere non tradotto	Repubblica tradotto	Repubblica non tradotto	Unità tradotto	Unità non tradotto	Subtotale tradotto	Subtotale non tradotto	Totale
N	250.014	245.810	378.161	366.382	225.257	224.335	836.527	853.432	1.689.959
D	850,07	840,25	867,15	838,76	864,96	859,85	839,00	858,45	848,72

Tabella 2.2 – Valori assoluti (n) e densità per 1.000 forme (d) dei lemmi appartenenti alla lista del *Vocabolario di base*

2.2.3. *Categorie del lessico*

La distribuzione del lessico dei diversi subcorpora in base alle parti del discorso individuate dalla grammatica tradizionale consente di caratterizzare ulteriormente dal punto di vista linguistico il corpus in esame. La figura 2.4 riporta la ripartizione

per categorie grammaticali sul totale delle occorrenze nei subcorpora esaminati. Anche se la classificazione automatica col software *TreeTagger* si caratterizza per una certa percentuale di errori, il confronto tra i subcorpora resta significativo. Come si può notare, non emergono differenze apprezzabili nella distribuzione delle parti del discorso tra subcorpus tradotto e subcorpus non tradotto; anche tra i subcorpora delle varie testate le differenze sono minime e non significative (tabella 2.3).

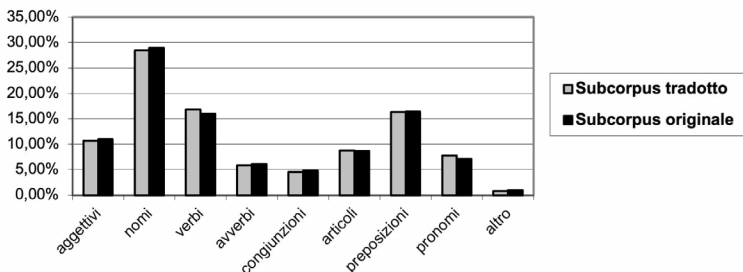


Figura 2.4 - Distribuzione percentuale delle parti del discorso nei subcorpora non tradotto e tradotto

Una situazione analoga si osserva dal punto di vista della densità lessicale (cfr. § 1.2.2.): a differenza di quanto ci si sarebbe potuti attendere in base agli studi precedenti, le parole piene hanno frequenza percentuale analoga in entrambi i subcorpora (56%, contro 44% di parole vuote), segno che in quest'ambito non emergono differenze significative nell'assetto dei testi frutto di traduzione rispetto a quelli originali. Nel leggere i dati, occorre tener conto del fatto che alcune diversità nel totale dei dati assoluti sono legate alle diverse procedure di lemmatizzazione.

	Corriere tradotto		Corriere non tradotto		Repubblica tradotto		Repubblica non tradotto	
	n	%	N	%	n	%	N	%
Aggettivi	33.199	11,29	33.223	11,29	44.554	10,22	49.749	11,35
Nomi	84.080	28,58	86.261	29,33	122.230	28,03	130.684	29,81
Verbi	47.548	16,16	44.865	15,25	75.619	17,34	64.164	14,64
Totale parole piene	164.827	56,03	164.349	55,87	242.403	55,58	244.597	55,80
Avverbi	17.751	6,03	19.215	6,53	26.129	5,99	28.384	6,48
Congiunzioni	13.995	4,76	14.693	5,00	19.734	4,52	22.373	5,10
Articoli	25.996	8,84	24.928	8,47	38.080	8,73	38.585	8,80
Preposizioni	47.655	16,20	48.574	16,51	70.499	16,17	71.363	16,28
Pronomi	21.530	7,32	19.080	6,49	35.905	8,23	27.915	6,37
Totale parole vuote	126.927	43,15	126.490	43,00	190.347	43,65	188.620	43,03
Atro	2.424	0,82	3.301	1,12	3.365	0,77	5.131	1,17
Totale	294.178	100,00	294.140	100,00	436.115	100,00	438.348	100,00

Tabella 2.3 – Frequenza delle principali categorie grammaticali nel corpus lemmatizzato

Unità tradotto		Unità non tradotto		Subtotale tradotto		Subtotale non tradotto		Totale	
n	%	N	%	n	%	n	%	n	%
28.166	10,81	28.992	11,08	105.919	10,69	111.964	11,26	217.883	10,98
75.325	28,92	76.384	29,20	281.635	28,43	293.329	29,51	574.964	28,97
43.306	16,62	41.476	15,85	166.473	16,80	150.505	15,14	316.978	15,97
146.797	56,35	146.852	56,13	554.027	55,92	555.798	55,91	1.109.825	55,91
14.276	5,48	16.078	6,15	58.156	5,87	63.677	6,41	121.833	6,14
11.536	4,43	13.240	5,06	45.265	4,57	50.306	5,06	95.571	4,81
23.003	8,83	21.766	8,32	87.079	8,79	85.279	8,58	172.358	8,68
43.786	16,81	43.351	16,57	161.940	16,34	163.288	16,43	325.228	16,39
19.204	7,37	17.921	6,85	76.639	7,74	64.916	6,53	141.555	7,13
111.805	42,92	112.356	42,95	429.079	43,31	427.466	43,00	856.545	43,15
1.890	0,73	2.416	0,92	7.679	0,78	10.848	1,09	18.527	0,93
260.492	100,00	261.624	100,00	990.785	100,00	994.112	100,00	1.984.897	100,00

2.2.4. *Forestierismi non adattati*

La presenza di forestierismi nelle traduzioni è uno di quei fenomeni che può condurre a esiti contrapposti a seconda che a prevalere sia la tendenza al conservatorismo linguistico (e quindi la resistenza all'introduzione di materiale esogeno) o l'interferenza del testo di partenza (che favorisce l'ingresso di stranierismi e calchi). Nei testi qui considerati possono valere fattori diversi, più o meno riconducibili all'opposizione tradotto/non tradotto. Un certo peso ha la selezione del corpus che, come si è visto, è stato assemblato a partire dalla tipologia e dai contenuti delle traduzioni (principalmente articoli di commento relativi a questioni internazionali). Ne discende che la presenza di forestierismi sia il diretto riflesso degli argomenti trattati, così giustificando la presenza di internazionalismi e riferimenti alla cultura di origine (*realia*). In questo caso, poiché per bilanciare il corpus di controllo sono stati inclusi tra gli articoli non tradotti anche testi che riguardano costume e società in Italia, ci si sarebbe potuti attendere una certa preponderanza dei forestierismi nelle traduzioni. Una simile previsione poteva essere giustificata anche dall'interferenza del testo di partenza su quello di arrivo: come noto, il livello lessicale è quello più ricettivo degli influssi esogeni.

Abbiamo quindi proceduto alla ricerca, nel corpus, dei forestierismi registrati nel *Dizionario delle parole straniere nella lingua italiana*. Per primi ci si è concentrati sui prestiti non adattati, controllando che le forme omografe di lemmi italiani fossero effettivamente presenti nel corpus come forestierismi. È il caso di forme grafiche suscettibili di invalidare completamente i risultati dell'analisi, vista l'incapacità del software di distinguere le varie accezioni: si pensi alla frequenza di *da*, riportato dalla nostra fonte come russismo, ma anche ad altre forme molto frequenti come *in*, *fu*, *pace*, *parole* ecc. Per fare un esempio, l'analisi delle concordanze (cioè dei contesti d'uso delle parole) rivela che solo 4 su 53 occorrenze di *future* sono classificabili come forestierismi non adattati. Naturalmente, gli omografi si concentrano nella parte alta della lista di frequenza e diminuiscono mano a mano che ci si avvicina agli hapax.

Una volta depurato di questi "falsi positivi", l'elenco dei forestierismi non adattati presenta, come ci si poteva aspettare, un netto dominio degli anglicismi, che individuano grossomodo

i tradizionali campi semantici di politica, economia e tecnologia. Da soli, gli anglicismi rappresentano oltre la metà dei forestierismi nel corpus e più del doppio del totale delle voci (in ordine decrescente) francesi, spagnole e tedesche. Inoltre, a fronte di una sostanziale parità delle voci spagnole e francesi, gli anglicismi registrano una leggera prevalenza nel subcorpus non tradotto.

	Subcorpus tradotto		Subcorpus non tradotto		Totale	
	n	%	n	%	n	%
Inglese	3.625	71,65	4.144	76,61	7.769	73,96
Francese	599	11,84	552	10,21	1.151	10,96
Spagnolo	88	1,74	107	1,98	206	1,96
Altre lingue	747	14,77	606	11,20	1.378	13,12
Totale	5.059	100,00	5.409	100,00	10.504	100,00

Tabella 2.4 – Distribuzione dei forestierismi non adattati presenti nel corpus per lingua. Dati assoluti e in percentuale

Per quanto concerne le singole voci, a titolo di esempio, riportiamo in tabella le forme con frequenza superiore a cinquanta. In questa selezione i francesismi si limitano a prestiti ultracentenari che, tra l'altro, non sono riconducibili a settori specifici, mentre gli arabismi riflettono chiaramente gli argomenti trattati nel corpus. Un caso interessante è quello del nipponismo *kamikaze*, la cui maggiore frequenza nel subcorpus non tradotto probabilmente deriva dal fatto che la stampa inglese non si serve di questa espressione per indicare i terroristi suicidi. Lo spoglio manuale indica che contengono *kamikaze* 38 testi italiani, 4 testi tradotti dall'inglese, 1 dal tedesco, 1 dallo spagnolo e 10 testi la cui lingua sorgente è incerta.

Forma grafica	Lingua	Subcorpus tradotto	Subcorpus non tradotto	Totale
<i>leader</i>	inglese	364	225	589
<i>film</i>	inglese	280	184	464
<i>islam</i>	arabo	268	55	323
<i>internet</i>	inglese	85	191	276
<i>media</i>	inglese	129	111	240
<i>computer</i>	inglese	62	95	157
<i>leadership</i>	inglese	84	57	141
<i>intelligence</i>	inglese	96	33	129
<i>standard</i>	inglese	79	46	125
<i>partner</i>	inglese	76	49	125
<i>premier</i>	inglese	41	79	120
<i>rock</i>	inglese	42	61	103
<i>test</i>	inglese	64	34	98
<i>slogan</i>	inglese	47	48	95
<i>video</i>	inglese	41	45	86
<i>club</i>	inglese	37	43	80
<i>camion</i>	francese	45	33	78
<i>manager</i>	inglese	24	53	77
<i>autobus</i>	francese	39	36	75
<i>album</i>	francese	52	22	74
<i>boom</i>	inglese	23	50	73
<i>kamikaze</i>	giapponese	17	55	72
<i>bar</i>	inglese	26	45	71
<i>business</i>	inglese	14	56	70
<i>sport</i>	inglese	24	42	66
<i>on</i>	inglese	22	40	62
<i>web</i>	inglese	28	23	61
<i>hotel</i>	inglese	30	27	57
<i>marine</i>	inglese	39	17	56
<i>status</i>	inglese	34	20	54
<i>élite</i>	francese	42	12	54
<i>mullah</i>	arabo	23	30	53
<i>taxi</i>	francese	22	29	51
<i>dossier</i>	francese	36	15	51
<i>record</i>	inglese	16	34	50
<i>software</i>	inglese	14	36	50
<i>city</i>	inglese	23	27	50
<i>privacy</i>	inglese	11	39	50

Tabella 2.5 – Forestierismi non adattati con frequenza > 50. Distinzione per lingua

In effetti, come già dimostrato dalla tabella 2.4 e ribadito dalla tabella 2.6 più sotto, i dati forniti dallo spoglio non confermano l'atteso influsso delle lingue di partenza: pur entro limiti molto contenuti, i subcorpora comprendenti le traduzioni evidenziano una densità di forestierismi inferiore a quella degli articoli scritti originariamente in italiano. La densità totale per mille occorrenze è pari a 5,40 nel sottotale non tradotto e 5,06 nel sottotale tradotto; determinante risulta il contributo di *Repubblica*: il confronto tra i subcorpora registra un divario nettamente maggiore ai risultati fatti registrare da *Corriere della Sera* e *Unità*, che evidenziano una sostanziale parità tra i rispettivi subcorpora. Difficile fornire una spiegazione: le maggiori dimensioni del subcorpus di *Repubblica* da sole non riescono a spiegare le differenze nella densità dei forestierismi non adattati. Si può ipotizzare che la maggior apertura all'influsso esogeno sia da ricondurre a una generale predilezione di questa testata per uno stile più mosso e brillante rispetto alla concorrenza.

Una spiegazione della leggera prevalenza dei forestierismi negli articoli scritti originariamente in italiano si ottiene dal riferimento all'atteggiamento tenuto dai traduttori, che tendono a sostituire con costanza anche parole straniere ormai entrate nell'uso. In ultima analisi, il fenomeno potrebbe essere ricondotto nell'alveo degli universali traduttivi in quanto conseguenza del maggior conservatorismo linguistico e della tendenza alla normalizzazione. Mentre i giornalisti italiani utilizzano materiale lessicale esogeno (introducendolo di propria sponte o mantenendolo parzialmente quando si trovano a lavorare su fonti in lingua straniera, come interviste o comunicati di agenzie stampa; cfr. il concetto di "traduzioni invisibili" esposto al § 1.3.1.) a fini stilistici ed espressivi, i traduttori cercano di sottrarsi all'influenza del testo di partenza 'italianizzandolo' negli aspetti più evidenti (quelli lessicali, appunto).

	Corriere tradotto	Corriere non tradotto	Repubblica tradotto	Repubblica non tradotto	Unità tradotto	Unità non tradotto	Subtotale tradotto	Subtotale non tradotto	Totale
n	1.577	1.558	2.547	2.813	991	1.049	5.059	5.409	10.504
d	5,33	5,30	5,81	6,40	3,79	4,00	5,06	5,40	5,28

Tabella 2.6 – Distribuzione e densità per 1.000 occorrenze dei forestierismi non adattati. Dati per lemmi

A rigor di logica, la controprova di tale tendenza dovrebbe basarsi sulla frequenza dei prestiti di lusso (cioè che sarebbe possibile tradurre agevolmente in italiano) e dei rispettivi sostituenti italiani; purtroppo, un confronto di questo tipo risulta arduo (se non impossibile), alla luce della non univocità del rapporto tra parola straniera e traduzioni italiane (si pensi alle varie possibilità di resa di parole come *business* o *manager*; *premier* può rendere *capo del governo*, *primo ministro* o *presidente del consiglio*) o degli evidenti vantaggi che il forestierismo comporta in termini di economia grafica (fattore generalmente importante nella scrittura giornalistica: per es. *mass media* si presenta 25 volte nel corpus, contro le 7 di *mezzi di comunicazione di massa*) o ancora di specificità tecnica (*internet* è assai più trasparente di *rete*). Inoltre, a meno di non procedere a uno spoglio manuale, non è possibile stabilire quando certi sostituenti italiani vengono usati nell'accezione specifica prevista dal prestito di lusso (per es. *programma* per *software*). Tra gli esempi della tabella 2.5 (frequenza > 50), l'unico caso che sembra prestarsi a una valutazione univoca è quello di *intelligence* (129 occorrenze nel subcorpus non tradotto, 33 in quello tradotto), la cui traduzione italiana (*servizi segreti*) è quasi due volte più frequente negli articoli tradotti (24 e 42 occorrenze rispettivamente).

Una spiegazione alternativa si basa principalmente sul prestigio linguistico dell'inglese. Si era già notato il dominio degli anglicismi nel subcorpus non tradotto (tabella 2.4), tendenza che si conferma anche andando a controllare i dati per testata. Come si può vedere nella tabella 2.7, gli anglicismi prevalgono sempre nei vari subcorpora originali, mentre il saldo si inverte per i francesismi nel *Corriere della Sera* e nell'*Unità*. Insomma, "l'anglomania" sembra essere una costante dei giornalisti italiani a prescindere dalla te-

stata, mentre le differenze nel computo totale dei forestierismi non adattati sono imputabili alle altre lingue presenti nel corpus.

	Corriere tradotto	Corriere non tradotto	Repubblica tradotto	Repubblica non tradotto	Unità tradotto	Unità non tradotto	Subtotale tradotto	Subtotale non tradotto	Totale
Anglismi									
n	1.080	1.227	1.840	2.145	705	772	3.625	4.144	7.769
d	3,66	4,18	4,20	4,89	2,70	2,95	3,63	4,14	3,90
Francesismi									
n	178	120	270	334	151	98	599	552	1.151
d	0,60	0,41	0,62	0,76	0,58	0,38	0,60	0,55	0,58
Ispanismi									
n	32	43	53	35	14	29	99	107	206
d	0,11	0,15	0,12	0,08	0,05	0,11	0,10	0,11	0,10
Germanismi									
n	13	26	11	39	4	10	28	75	103
d	0,04	0,09	0,03	0,09	0,02	0,04	0,03	0,08	0,05

Tabella 2.7 – Distribuzione dei forestierismi non adattati per lingua.
Densità per 1.000 forme in base alla testata

Un altro dato interessante deriva dal calcolo della differenza delle forme straniere più presenti negli articoli originali e tradotti. Pur consapevoli dei limiti (la differenza tra le 191 e 85 occorrenze di *internet* è meno significativa in termini percentuali della differenza tra le 22 e 2 occorrenze di *zar*), basiamo il confronto sui numeri assoluti e riportiamo le voci con differenza > 15:

Forma grafica	Subcorpus tradotto	Subcorpus non tradotto	Totale	Differenza
<i>internet</i>	85	191	276	106
<i>business</i>	14	56	70	42
<i>premier</i>	41	79	120	38
<i>kamikaze</i>	17	55	72	38
<i>computer</i>	62	95	157	33
<i>manager</i>	24	53	77	29
<i>privacy</i>	11	39	50	28
<i>tunnel</i>	7	35	42	28
<i>boom</i>	23	50	73	27
<i>software</i>	14	36	50	22
<i>deregulation</i>	2	24	26	22
<i>bunker</i>	12	33	45	21
<i>zar</i>	2	22	24	20
<i>rock</i>	42	61	103	19
<i>bar</i>	26	45	71	19
<i>sport</i>	24	42	66	18
<i>on</i>	22	40	62	18
<i>record</i>	16	34	50	18

Tabella 2.8 – Forestierismi non adattati più presenti nel subcorpus non tradotto. Dati assoluti; differenza > 15

Nella stampa italiana il russismo *zar* (peraltro poco numeroso come occorrenze totali) viene spesso utilizzato in senso figurato (del tipo *lo zar del gas*), cosa che avviene più di rado negli articoli tradotti. Per il resto, gli anglicismi dominano e presentano forme ormai entrate stabilmente nella lingua italiana, legate ai settori della tecnologia (*internet*, come è lecito attendersi nel decennio successivo al 2000, domina, mentre *web*, anche nel poliforme *world wide web*, risulta leggermente più frequente nelle traduzioni: 28 contro 23 occorrenze) e dell'economia. *Sport*, *bar*, *rock* sono parole prive di alternative italiane; *tunnel* (al pari dei francesismi legati ai trasporti *camion* e *autobus* visti sopra) prevale su *galleria* (nel senso di 'traforo' solo 16 occorrenze) anche in virtù delle espressioni figurate *entrare/uscire dal tunnel*. Della sfera politica fanno parte *premier* (di cui abbiamo già detto) e *deregulation*: quest'ultima voce, al pari di *privacy* (e *mobbing*: 3 occorrenze, tutte nel subcorpus non tradotto), pur non potendo essere definita un vero pseudoan-

glicismo, in italiano ha incontrato una fortuna che non conosce nella lingua di origine. Con due eccezioni (*off* e *on the record*), *record* compare quasi solo nell'accezione di 'primato' (a sua volta, 22 occorrenze negli articoli italiani, 11 nelle traduzioni); la frequenza di *on* è determinata dal poliforme *on line* e da titoli nomi di istituzioni e iniziative (per es. *Foundation on Economic Trends*). Infine, come *kamikaze* (cfr. sopra), anche *bunker* può essere messo in relazione con gli argomenti trattati (le attività belliche in Medio Oriente e nel Golfo).

Passiamo ora a esaminare i forestierismi non adattati più frequenti nel subcorpus tradotto:

Forma grafica	Subcorpus tradotto	Subcorpus non tradotto	Totale	Differenza
<i>Islam</i>	268	55	323	-213
<i>leader</i>	364	225	589	-139
<i>film</i>	280	184	464	-96
<i>intelligence</i>	96	33	129	-63
<i>jihad</i>	36	0	36	-36
<i>standard</i>	79	46	125	-33
<i>test</i>	64	34	98	-30
<i>album</i>	52	22	74	-30
<i>élite</i>	42	12	54	-30
<i>leadership</i>	84	57	141	-27
<i>partner</i>	76	49	125	-27
<i>college</i>	32	9	41	-23
<i>marines</i>	39	17	56	-22
<i>dossier</i>	36	15	51	-21
<i>khan</i>	30	11	41	-19
<i>robot</i>	22	3	25	-19
<i>media</i>	129	111	240	-18
<i>ayatollah</i>	25	7	32	-18
<i>ramadan</i>	23	7	30	-16
<i>caucus</i>	16	0	16	-16
<i>campus</i>	22	7	29	-15

Tabella 2.9 – Forestierismi non adattati più presenti nel subcorpus tradotto.
Dati assoluti; differenza > 15

In questo caso, sembra possibile concludere che le traduzioni risentono molto di più dell'influenza degli argomenti trattati e, di conse-

guenza, della lingua di partenza. Si giustifica così la forte presenza di prestiti dall'arabo (*islam, jihad, ramadan*) e dal persiano (*ayatollah*), oltre alla forma *khan*, evidentemente un nome proprio molto comune (solo 4 occorrenze del personaggio storico Gengis Khan). Numerosi i francesismi, anche se, come spesso avviene, si ha l'impressione che l'ingresso sia mediato da altra lingua: *élite* e *album* sono parole di uso comune in inglese (specie il secondo in ambito musicale) e infatti lo spoglio indica che *élite* compare in 6 testi tradotti dal francese, ma ben 11 tradotti dall'inglese, 18 di origine incerta e 1 dallo spagnolo, mentre *album* non risulta nelle traduzioni dal francese (9 testi dall'inglese e 8 di origine incerta). Anche *dossier* compare in 3 traduzioni dal francese, 3 dall'inglese, 1 dallo spagnolo e 9 la cui lingua fonte non è specificata. Tra gli anglicismi, oltre a quelli più frequenti in assoluto (*film* e *leader*), si riscontrano forme connesse agli argomenti trattati (*intelligence, marines*) e alla cultura della lingua di partenza (*college, campus* e *caucus*).

2.2.5. Le forme plurali dei forestierismi: il caso della s

Un caso interessante riguarda le forme plurali segnalate dal morfema s, la soluzione più comune nelle tre lingue di partenza più rappresentate nel subcorpus tradotto (inglese, francese e spagnolo, ma anche portoghese). In genere, le grammatiche, che si concentrano per lo più sugli anglicismi, notano che gli esotismi non adattati da tempo acclimatati nella nostra lingua rimangono invariati nel plurale (*il film - i film*), mentre se la parola è di introduzione recente o di uso poco comune, può mantenere la s del plurale.

Lo spoglio del corpus si è basato sulla ricerca delle forme inverse del vocabolario, escludendo titoli, nomi propri e citazioni di intere frasi o sintagmi complessi in lingua originale. Si è invece tenuto conto delle polirematiche (cioè di espressioni fisse composte da più parole come *carta di credito* o *ferro da stiro*) e delle forme segnalate graficamente (per es. virgolettate). Netta la predominanza dei plurali in s nei testi non tradotti rispetto ai tradotti:

	Subcorpus tradotto	Subcorpus non tradotto	Totale
n	152	406	558
d	0,15	0,41	0,28

Tabella 2.10 – Densità dei plurali in s per 1.000 forme

Riportiamo sotto la distribuzione dei plurali in s tra subcorpus tradotto e subcorpus non tradotto in base alla lingua di appartenenza delle parole:

	Subcorpus tradotto		Subcorpus non tradotto		Totale	
	n	%	n	%	n	%
Inglese	101	66,45	231	56,90	332	59,50
Francese	15	9,87	86	21,18	101	18,10
Spagnolo	33	21,71	86	21,18	119	21,33
Altro	3	1,97	3	0,74	6	1,08
Totale	152	100,00	406	100,00	558	100,00

Tabella 2.11 – Plurali in s nel subcorpus non tradotto e tradotto in base alla lingua di origine. Dati assoluti e percentuali

Pur tenendo conto che il repertorio lessicale non corrisponde (la ricerca dei plurali in s evidenzia parole non comprese negli elenchi forniti usati per individuare i prestiti integrali), nel confronto con i dati relativi ai forestierismi non adattati (tabella 2.4), colpisce l'incremento delle occorrenze con plurale in s per quanto riguarda il francese e soprattutto lo spagnolo, mentre naturalmente crolla la percentuale delle altre lingue (compreso il tedesco), che non prevedono il morfema s come marca del plurale: l'unico arabismo, peraltro non attestato nel *Dizionario delle parole straniere nella lingua italiana*, è *madras*, con 2 occorrenze, che nel corpus trova alternative come *madrasas* (1 occorrenza), *madrasse* e *madrassa* invariato. In termini assoluti, i giornalisti sembrano utilizzare la desinenza s più frequentemente dei traduttori in tutte le lingue, ma in percentuale i traduttori utilizzano maggiormente il plurale in s per le parole inglesi e meno frequentemente per quelle francesi, mentre non emergono differenze per le parole di origine spagnola.

Il problema dell'alternanza tra forma in s e plurale invariato non sembra porsi quando il prestito è un ispanismo o un lusitanismo, e la spiegazione risulta ovvia: siccome in italiano, in spagnolo e in portoghese le marche di genere al singolare per lo più corrispondono, utilizzare la forma invariata causerebbe una sorta di contraddizione morfologica: *gli indio* e *le favela* non sembrano accettabili. Questo vale anche per le parole terminanti in e (il corpus contiene *vigilantes*, *teratenientes*, *merengues*

e, sorprendentemente, *coyotes*) e ancor più per i lemmi che al singolare terminerebbero per consonante: *peones*, *perdedores*, *libertadores*, *ferrocarriles*, *señores*, *frejoles*). Al più, come noto, è possibile l'estensione della forma in s anche al singolare: il corpus riporta *silos* ma anche *murales* usati per indicare il singolare (una occorrenza ciascuno). Discorso analogo ma, con un percorso più accidentato, per *commando* (1 occorrenza di *commandos* e 3 di *commando* con significato plurale, tutte in traduzioni), un lusitanismo mutuato dall'inglese e quindi soggetto alle 'regole' di formazione del plurale valide per gli anglicismi.

Passando ai prestiti dall'inglese, per l'appunto, i dati del corpus sembrano più o meno confermare le osservazioni contenute nelle grammatiche. Per le forme di uso più comune, il plurale in s risulta una scelta minoritaria (5 occorrenze di *partners* contro 85 di *partner* con significato plurale; 3 *managers* contro 39 *manager*, 1 *clubs* contro 6 *club*, 2 sole occorrenze di *leaders*, quando abbiamo visto che *leader* è uno dei forestierismi più frequenti e per circa metà delle occorrenze ha significato plurale; lo stesso dicasi per *computers*). Visto il basso numero di occorrenze per voce, è difficile fare un confronto tra articoli originali e tradotti: *cowboys*, *hippies*, *lobbies* e *yankees* compaiono in entrambi i subcorpora; l'unico caso di netta divergenza è *marines* con 39 occorrenze nelle traduzioni e 17 negli articoli originariamente scritti in italiano. Se per alcuni casi la persistenza della s è chiaramente giustificata (ovviamente *news* e *United States*, ma anche *first ladies*, probabilmente per assonanza con lo stereotipo *ladies and gentlemen*), in altri non si comprende bene il motivo per cui parole di uso tutto sommato comune non si siano acclimatate con plurale invariabile: 4 occorrenze di *hooligans*; *fans* registra 11 occorrenze (10 nel subcorpus non tradotto) contro 7 di *fan* in senso plurale. Particolare il caso di *zombie*, il cui plurale può avere o non avere la s, oppure essere italianizzato in *zombi*.

In genere parrebbe confermato che la s caratterizzi termini specialistici, a prescindere da quanto comunemente siano usati (*contractors*, *royalties*, *defaults*, *futures*, *buyers*, *dealers*, *vouchers*, *tycoons*, *neocons* ecc.), specie se parte di polirematiche (*hedge funds*, *asset managers*, *family values*, *instant messages*, *chat-lines*, *venture capitalists*) o stereotipi (*checks and balances*), anche nel caso di pseudoanglicismi (*Papa boys*).

Più complessa la situazione delle parole di origine francese. L'analisi del corpus rivela che l'impiego del morfema *s* per segnalare il plurale è la scelta maggioritaria, soprattutto tra i giornalisti (cfr. tabella 2.11), ma esclusivamente per il peso relativo dei lemmi *élite* e *banlieue*: tutti gli altri registrano il medesimo numero di occorrenze sia nel subcorpus tradotto che in quello non tradotto. Anche una parola di uso comune come *élite* (compresa la grafia senza accento *elite*) fa segnare una situazione di sostanziale parità, comparando con la desinenza *s* 22 volte (14 nel subcorpus non tradotto e 8 nel tradotto) mentre il significato plurale viene veicolato dalla forma senza *s* in 24 casi (solo 4 occorrenze nel subcorpus non tradotto). Con l'eccezione di *enclave* (3 occorrenze con *s*, 4 senza) per gli altri francesismi la forma con la *s* risulta maggioritaria (riportiamo gli esempi con frequenza superiore a 4: *banlieues* 7 contro 3, *philosophes* 5 a 0, *toilettes* 4 a 0, *enfants* 4 a 2).

Spiegare questa preferenza, che tra l'altro rappresenta un mero fatto grafico e non di pronuncia (fatto salvi eventuali casi di *liaison*), non è semplice. A prescindere dalla presenza di stereotipi polimembri, in cui le marche morfologiche possono essere ridondanti (*enfants terribles*, *enfants gâtés*, *bons vivants*) e che vengono percepiti come forme cristallizzate (*tombeurs de femmes*), la spiegazione più probabile riguarda lo status di 'lingua di cultura' assunto oggi dal francese: i francesismi sono spesso prestiti di lusso, segnali della statura culturale dello scrivente (in opposizione all'ubiquitario inglese) e, come tali, degni di una maggiore precisione morfologica.

2.2.6. Parole italiane derivate o connesse con lingue straniere

Un ulteriore elemento che può contribuire a rendere conto dell'influsso della lingua di partenza sull'italiano tradotto è rappresentato dalla presenza di parole straniere che però risultano italianizzate nella morfologia, grafia e pronuncia, che in tempi recenti derivano da altre lingue. Si è dunque proceduto alla ricerca nel corpus degli elenchi riportati nell'*Appendice I* e nell'*Appendice II* del *Dizionario delle parole straniere nella lingua italiana*. Riporiamo in tabella i dati per forme:

	Corriere tradotto	Corriere non tradotto	Repubblica tradotto	Repubblica non tradotto	Unità tradotto	Unità non tradotto	Subtotale tradotto	Subtotale non tradotto	Totale
n	300	332	398	430	189	209	887	971	1.858
d	1,02	1,13	0,92	0,98	0,73	0,80	0,89	0,97	0,93

Tabella 2.12 – Distribuzione e densità per 1.000 occorrenze delle parole derivate o connesse con lingue straniere e delle polirematiche costruite su analoghe espressioni straniere. Subcorpora delle testate; dati per forme

Come si può vedere, anche in questo caso si registra una differenza minima ma costante a favore dei subcorpora originali, possibile indice di un maggiore sforzo, da parte dei traduttori, teso a evitare la possibile influenza del testo fonte. Per quanto riguarda la distribuzione per lingue, gli anglicismi continuano a farla da padrone, con una densità di 0,52 per 1.000 occorrenze e una leggera preponderanza nel subcorpus tradotto (0,53 contro 0,51 per 1.000). La densità dei francesismi è di circa la metà (0,25, ma si inverte il rapporto tra non tradotto e traduzioni: 0,26 e 0,24 rispettivamente), mentre germanismi (0,03) e ispanismi (0,01) sono molto meno presenti.

Naturalmente, i risultati così ottenuti presentano tutti i limiti di una ricerca quantitativa per forme che, dati i numeri coinvolti, non può tenere conto dei contesti d'uso e delle varie accezioni dei lemmi rinvenuti. Così, se si considerano le 10 voci più frequenti (*impatto, etico, genocidio, digitale, fondamentalismo, ricetta, strutturale, tribale, liberalizzazione, riciclare*), i problemi di classificazione emergono con tutta la loro evidenza: mentre si può presumere che nei giornali *digitale* sia utilizzato principalmente nel suo significato tecnico (e quindi sia classificabile come anglicismo), non si può dire lo stesso di *etico* e *ricetta*. Al contrario, le polirematiche costruite su analoghe espressioni straniere non presentano il pericolo della polisemia, ma il loro elenco non sembra rivelatore di particolari tendenze: a parte il solito dominio delle voci di derivazione inglese, possiamo notare l'alta frequenza di due francesismi come *terzo mondo* e *senso unico*.

Polirematica	Occorrenze	Lingua
terzo mondo	55	francese
grande fratello	13	inglese
senso unico	11	francese
gruppo di pressione	5	inglese
profondo sud	4	inglese
ricerca di mercato	4	inglese
intelligenza artificiale	4	inglese
lavaggio del cervello	3	inglese
grande freddo	3	inglese
colletto blu	3	inglese
rampa di lancio	3	francese
pericolo giallo	2	tedesco
stazione di servizio	2	inglese
sala stampa	2	inglese
redattore capo	2	francese
penna a sfera	1	inglese
uomo rana	1	inglese
politicamente corretto	1	inglese
navetta spaziale	1	inglese
nave spaziale	1	inglese
integrazione razziale	1	inglese
giovane leone	1	inglese
senso sicuro	1	inglese

Tabella 2.13 – Polirematiche costruite su analoghe espressioni straniere rinvenute nel corpus

2.2.7. Interferenza di lingue straniere nell'uso di alcune forme italiane

Chiudiamo questa carrellata sul possibile influsso delle lingue straniere sulle scelte lessicali dei giornali italiani con una concessione alla visione più “catastrofista” del fenomeno, spesso propugnata da chi deplora non solo il presunto ingresso di un enorme numero di stranierismi (soprattutto dall'inglese) in italiano, ma anche le modificazioni del significato di parole esistenti come esito della somiglianza morfologica con materiale esogeno: l'esempio lampante è il verbo *realizzare* nel significato di ‘accorgersi, capire’, sul modello di *to realize*. Ovviamente si può trattare di un'innovazione piuttosto recente (questa accezione di *realizzare* è entrata in italiano nel 1938) o del recupero di lemmi o

accezioni poco frequenti (*attitudine* nel senso di *atteggiamento* risale al 1519).

Senza alcuna pretesa di esaustività né di precisione scientifica, solo basandoci su valutazioni impressionistiche e sugli sporadici commenti che si possono leggere su giornali e riviste, abbiamo stilato una lista di verbi, avverbi, aggettivi e nomi potenzialmente interessati dal fenomeno e li abbiamo cercati nel corpus. Risultano assenti: *applicazione* (e *applicante*, nel senso di chi presenta domanda di partecipazione), *in progresso* (per 'in corso'), *suggestione* ('suggerimento'), *confidente* ('fiducioso'), *rumori* ('voci di corridoio'), *dedicato* ('diligente'). Al di là di quelle che sembrano semplici sviste o cattive traduzioni (per es. *libreria* per *biblioteca*), riportiamo in tabella i risultati dello spoglio:

Lemma	Subcorpus tradotto	Subcorpus non tradotto	Totale
<i>realizzare</i>	0	2	2
<i>supportare</i>	0	9	9
<i>confidenziale</i>	4	0	4
<i>intrigante</i>	3	2	5
<i>ansietà</i>	2	3	5
<i>attitudine</i>	4	5	9
Totale	13	21	34

Tabella 2.14 – Presunti calchi semantici. Dati assoluti per lemma

Anche in questo caso risulta difficile trarre delle conclusioni univoche. Per cominciare, sembra comunque che l'incidenza di forme morfologicamente simili a parole straniere (*supportare*, *ansietà*) o di accezioni innovative o preesistenti ma secondarie, probabilmente incoraggiate da influssi esogeni (come in *realizzare* o *attitudine*) risulti piuttosto contenuta, soprattutto nel confronto con alternative ritenute più tradizionali. Per esempio *ansia* compare nel corpus 80 volte, *atteggiamento* 247, ma anche una parola sicuramente non di uso comune come *avvincente* (8 occorrenze nel subcorpus non tradotto, 2 in quello tradotto) risulta più frequente di *intrigante* come calco di *intriguing*. Per quanto concerne il confronto tra giornalisti e traduttori, se si eccettua il caso di *confidenziale* (ma *riservato*, nella stessa accezione, registra 8 occorrenze totali, di cui 5 nelle traduzioni) la situazione di nomi e aggettivi evidenzia un sostanziale pareggio. Solo i due verbi *supportare* e *realizzare*, che probabilmente

rappresentano i due casi più evidenti di calco semantico, hanno una distribuzione netta, che può lasciar pensare a uno sforzo maggiore, da parte dei traduttori, per sottrarsi all'influsso esogeno.

2.2.8. *Espressioni politicamente corrette*

Negli ultimi anni sono stati numerosi gli sforzi compiuti e le raccomandazioni espresse dal legislatore come da parte del mondo accademico per incoraggiare modalità di comunicazione linguistica libere dalla discriminazione, soprattutto da quella di genere. Tuttavia, in Italia si registra una certa resistenza all'introduzione di queste innovazioni, mentre altrove, e soprattutto negli Stati Uniti d'America, si denunciano comunemente gli eccessi del politically correct.

Fatte queste premesse, si può ipotizzare che le espressioni politicamente corrette possano fornire un buon banco di prova per l'ipotesi dell'interferenza linguistica in traduzione. I testi compresi nel corpus, infatti, sono in maggioranza traduzioni da lingue di Paesi con una tradizione più radicata di quella italiana per quanto concerne la censura dei comportamenti linguistici sessisti e potenzialmente offensivi nei confronti delle minoranze. Un esempio per tutti è fornito dal successo della forma femminile *cancelliera* nella stampa italiana, chiaro riflesso del tedesco *Kanzlerin*.

Lo spoglio del corpus è stato ispirato dalle *Raccomandazioni per un uso non sessista della lingua italiana* di Alma Sabatini. In primis, ci si è occupati del maschile generico, inteso come l'uso del maschile come genere neutro quando ci si riferisce a gruppi, categorie, popoli comprendenti anche donne. In particolare, sono state prese in considerazione espressioni di valore generale comprendenti la parola *uomo*, del tipo *diritti dell'uomo* o *i sogni degli uomini*, là dove un atteggiamento più politicamente corretto prevedrebbe l'impiego di forme non marcate dal punto di vista del genere (per es. *diritti umani* o *della persona*).

	Subcorpus tradotto	Subcorpus non tradotto	Totale
espressioni con <i>uomo</i>	49	38	87
espressioni con <i>uomini</i>	25	59	84
Totale	74	97	171

Tabella 2.15 – Espressioni contenenti il lemma *uomo* con significato generico.
Dati assoluti

La tabella 2.15 rende conto delle espressioni in cui *uomo* e *uomini* stanno per *esseri umani* (per es. *l'uomo comune*, *l'uomo bianco*, *la storia dell'uomo*, *al servizio dell'uomo*, *tutti gli uomini*), ma sono state espunti stereotipi quali *caccia all'uomo* o *a memoria d'uomo*, anche se le *Raccomandazioni* prescrivono anche in questi casi l'impiego di parole non connotate (*caccia all'individuo* o *alla persona*). Come si può vedere, il leggero sbilanciamento in favore del subcorpus tradotto nelle occorrenze della forma singolare *uomo* viene più che controbilanciato dal numero delle espressioni che contengono il plurale *uomini*. In particolare, negli articoli scritti originariamente in italiano, godono di particolare fortuna i riferimenti a categorie di persone individuate da un tratto comune, come in *uomini d'affari*, *uomini del Cremlino*, *uomini di buona volontà*, *uomini politici*, *uomini di scienza/di spettacolo/di governo* ecc.). Molto minore (10 occorrenze contro 43) la presenza di questo tipo di costrutti nelle traduzioni.

Più nel dettaglio, se si considera la sola polirematica *diritti dell'uomo* in opposizione alle alternative *diritti umani* e *diritti della persona*, emerge chiaramente la preferenza accordata nelle traduzioni alle varianti politicamente corrette:

	Subcorpus tradotto	Subcorpus non tradotto	Totale
diritti dell'uomo	22	16	38
diritti umani	212	75	287
diritti della/e persona/e	0	5	5

Tabella 2.16 – Confronto tra *diritti dell'uomo/umani/della persona*.
Dati assoluti

Anche se la sensazione è che gli articoli che costituiscono il subcorpus tradotto si occupino maggiormente di argomenti relativi a collettività estese di individui (insomma, forse le traduzioni riguardano maggiormente questioni che riguardano interi popoli, l'essere umano nella sua essenza o l'umanità nel suo complesso, il che potrebbe spiegare il numero più alto di espressioni con valore generico in senso assoluto), resta il fatto che le alternative non marcate come sessiste risultano proporzionalmente più numerose. Se le espressioni con *persona/persone* sono più utilizzate dai giornalisti (per es. *sicurezza* o *dignità delle persone*, *persone comuni*, *persone in carne e ossa*), nel caso di quelle comprendenti

la parola *umanità* (nella storia dell'umanità, la storia dell'umanità; ma escludiamo espressioni fisse quali *patrimonio dell'umanità* o *crimini contro l'umanità*) o l'aggettivo *umano* (*popolazione umana, libertà umana, sicurezza umana* ecc.) lo squilibrio si fa evidente:

	Subcorpus tradotto	Subcorpus non tradotto	Totale
espressioni con <i>persona/e</i>	34	5	99
espressioni con <i>umanità</i>	24	15	39
espressioni con <i>umano</i>	477	189	666
Totale	535	269	804

Tabella 2.17 - Espressioni contenenti i lemmi *persona, umanità* e *umano*.
Dati assoluti

Le traduzioni dimostrano di ricorrere maggiormente anche a costrutti bimembri comprendenti sia la forma maschile che quella femminile, alternandole nell'ordine di presentazione come previsto nelle *Raccomandazioni* onde evitare di dare preminenza al maschile: quindi, non solo *uomini* e *donne* (10 occorrenze nel subcorpus tradotto), ma anche *donne* e *uomini* (per la verità, solo 5 occorrenze).

	Subcorpus tradotto	Subcorpus non tradotto	Totale
espressioni con <i>uomo e donna / uomini e donne</i>	54	21	75
espressioni con <i>ragazzo e ragazza / ragazzi e ragazze</i>	11	3	14

Tabella 2.18 - Espressioni comprendenti sia il femminile che il maschile.
Dati assoluti

Passando agli "agentivi", cioè titoli, cariche, professioni e mestieri, occorre premettere che, poiché l'inglese è la lingua fonte più rappresentata nel subcorpus tradotto e di norma non morfologizza il genere in questa classe di nomi, è difficile attendersi che dallo spoglio emergano differenze significative. Peraltro, alcune delle *Raccomandazioni* (come quella di evitare il suffisso *-essa*) danno adito a qualche dubbio circa la loro accettabilità nell'uso comune. Riteniamo dunque inutile scendere nel dettaglio, limitandoci a no-

tare che *presidentessa* compare solo in traduzione (3 occorrenze), *la presidente* solo nel subcorpus non tradotto (2), quando invece *presidente donna* si divide equamente tra i due subcorpora (4 occorrenze in totale).

A proposito dell'aggiunta del sostantivo *donna* (quasi sempre preposto) per ottenere la forma femminile delle professioni (procedimento possibile anche in inglese, per quanto criticato dai fautori del *politically correct*), si tratta di una scelta esclusiva dei giornalisti, praticamente assente nelle traduzioni. Nel subcorpus non tradotto troviamo *donna* seguito da *capo di Stato*, *premier*, *ministro*, *soldato*, *medico*, *terrorista*, *kamikaze*; due i casi di posposizione (*generale donna* e *apostolo donna*); al plurale abbiamo *donne parlamentari*, *politiche*, *pubbliche* e ancora *soldato*. Le traduzioni si limitano a *candidato donna*, *delegato donna* e *donna poliziotto* (1 occorrenza ciascuna).

Di *cancelliera* si è già detto (15 occorrenze nel subcorpus non tradotto e 3 in quello tradotto, contro 2 e 1 rispettivamente per *il cancelliere* riferito ad Angela Merkel); *la ministra* (3 occorrenze, tutte nei testi originariamente scritti in italiano) risulta meno frequente della forma maschile riferita a una donna (5 occorrenze, di cui 3 in traduzione); 2 le occorrenze di *consigliera* (una per subcorpus) e una forma maschile riferita a donna (in traduzione). Aiutate probabilmente dall'origine deverbale, sono numerose le *candidate* (27, di cui 10 in traduzione) e le *deputate* (11, di cui 6 in traduzione).

Passando ai suffissi, *-essa* (restano escluse naturalmente, forme come *baronessa*, *principessa* ecc.) è presente esclusivamente nel subcorpus tradotto: alla *presidentessa* vista sopra si aggiungono *dottoressa* (22 occorrenze) e *soldatessa* (1) mentre, conformemente alle *Raccomandazioni per un uso non sessista della lingua italiana*, il suffisso *-trice* risulta più utilizzato in entrambi i subcorpora (42 occorrenze nell'italiano non tradotto, 25 nel tradotto). Infine, per quanto concerne "la segnalazione dis-simmetrica di donne e uomini nel campo politico" (Sabatini A. 1987: 106), l'uso dell'articolo (*la Clinton*, *la Palombelli*) o di un qualche titolo (*Frau* o *la signora Merkel*) o anche del nome di battesimo (*Golda Meir*, *Carla Bruni*) di fronte al cognome è la prassi consolidata in tutto il corpus, con rarissime eccezioni ("con Brambilla o con Dell'Utri?" si chiede Furio Colombo sull'*Unità*).

Naturalmente il politicamente corretto non si esaurisce nel

tentativo di evitare il linguaggio sessista, comprendendo anche le sfere della discriminazione razziale, dello svantaggio fisico, mentale ed economico ecc. Tuttavia, lo scarso materiale rinvenuto nel corpus, la mancanza di linee guida ben definite in italiano (*omosessuale* – 6 occorrenze nel corpus non tradotto e 24 nel tradotto – è più o meno politicamente corretto di *gay* – rispettivamente 20 e 6?) e la presenza di usi metaforici (*essere sordo a un appello*) sconsigliano qualsiasi valutazione di tipo quantitativo.

2.3. Morfologia

Per quanto concerne la morfologia lessicale, un aspetto che si è cercato di indagare è la diversa frequenza di varianti della stessa forma, nella presunzione che questa possa essere una spia della tendenza alla normalizzazione postulata dagli universali traduttivi. Come l'aderenza alla grammatica normativa più tradizionale e il respingimento dei tratti più innovativi dell'italiano dell'uso medio, la selezione di forme recessive e tipiche di registri più alti potrebbe gettare luce sul modello linguistico che i traduttori considerano “più sicuro”, in quanto maggiormente condiviso dalla comunità dei parlanti.

Se ci si limita ai casi più significativi, tra le alternative *devo/debbo* e *devono/debbono* i testi tradotti tendono sfruttare maggiormente le varianti più formali, come si può notare dalla tabella 2.19, che riporta dati assoluti, confrontabili alla luce dell'eguale dimensione dei due subcorpora. Più in generale, va notata la minore frequenza del verbo *dovere* nei testi non tradotti rispetto a quelli tradotti, probabilmente nei primi la modalità deontica viene espressa con un più ampio ventaglio di risorse, per es. il passivo costruito con “andare”, o il “futuro iussivo”.

La tendenza al conservatorismo dei testi tradotti non è però confermata dalle alternative *siedo/seggo*, *siedono/seggono* e *sieda/segga*, anche se la scarsa pregnanza del verbo in questione rispetto agli argomenti trattati nel corpus e la sporadicità delle attestazioni di queste forme rendono il caso poco significativo.

	Subcorpus tradotto	Subcorpus non tradotto	Totale
<i>debbo</i>	5	3	8
<i>devo</i>	56	38	94
<i>debbono</i>	75	16	91
<i>devono</i>	246	173	419
<i>segga</i>	0	1	1
<i>sieda</i>	1	0	1
<i>seggo</i>	0	1	1
<i>siedo</i>	0	1	1
<i>seggono</i>	0	1	1
<i>siedono</i>	7	8	15
<i>seggano</i>	0	0	0
<i>siedano</i>	0	0	0

Tabella 2.19 – Forme alternative di *dovere* e *sedere*. Dati assolti nei subcorpora

Un ulteriore aspetto morfologico che potrebbe andare nella medesima direzione è la frequenza della *d* eufonica in quanto, nell'italiano dell'uso medio, le forme *ad*, *ed* vengono limitate ai casi di incontro con la stessa vocale. Tuttavia, osservando la tabella 2.20, si possono notare risultati non univoci nei vari casi.

Per quanto riguarda *ed*, la forma con la *d* eufonica è più frequente nei testi originali, con circa 200 casi in più rispetto alle traduzioni, che quindi sembrerebbero evidenziare un atteggiamento meno conservativo. Inoltre, nei casi di incontro con vocali diverse da *e*, la *d* eufonica è preferita nettamente nei testi originali rispetto a quelli tradotti, che invece preferiscono utilizzare la *d* eufonica quando è prevista dalle grammatiche.

La situazione risulta diversa con *ad*, in assoluto più frequente nei testi tradotti. Tenuto conto degli stereotipi (come *ad ogni modo*, *ad esempio*), anche in questo caso i traduttori sembrano però attenersi in modo più sistematico al principio di ricorrere alla *d* eufonica nei soli casi di incontro con la stessa vocale. I traduttori paiono insomma essersi dati una 'regola' precisa che applicano con costanza, mentre i giornalisti sembrano concedersi una libertà maggiore. Notiamo la, seppur limitata, persistenza di *od*, una forma ritenuta pressoché scomparsa dall'italiano comune di oggi.

	Subcorpus tradotto	Subcorpus non tradotto	Totale
<i>e</i>	24.698	28.165	52.863
<i>ed</i>	910	1.113	2.023
<i>ed e*</i>	497	340	837
<i>ed è</i>	314	323	637
<i>ed + altra vocale</i>	99	450	5 49
<i>o</i>	3.045	3.341	6.386
<i>od</i>	5	2	7
<i>o + o*</i>	16	7	23
<i>a</i>	14.443	13.954	28.397
<i>ad</i>	1.950	1.675	3.625
<i>ad a*</i>	983	688	1.671
<i>ad + altra vocale</i>	827	871	1.698
<i>ad esempio</i>	140	116	256
<i>ad ogni</i>	37	55	92
<i>a ogni</i>	42	44	86

Tabella 2.20 – Frequenza della *d* eufonica secondo i contesti d'uso.
Dati assoluti

Un ulteriore fenomeno che può essere inquadrato nella più generale tendenza alla conservatività e alla ricerca di formalità a livello morfologico è l'uso di *il quale*, *la quale* ecc. come pronome relativo al posto del più comune *che*. Anche in questo caso, i dati relativi alla densità (tabella 2.21) mostrano una preferenza dei testi tradotti per questa forma (compresi i casi obliqui).

	Corriere tradotto	Corriere non tradotto	Repubblica tradotto	Repubblica non tradotto	Unità tradotto	Unità non tradotto	Subtotale Tradotto	Subtotale non tradotto	Totale
n	322	311	480	282	239	260	1.041	853	1.894
d	1,09	1,06	1,10	0,65	0,92	1,00	1,05	0,85	0,95

Tabella 2.21 – Dati assoluti e densità per 1.000 forme del pronome relativo *il quale*, *la quale* ecc.

2.3.1. Distribuzione di tempi e modi verbali

L'analisi della morfologia verbale, in particolare per quanto concerne la distribuzione di modi e tempi, non è possibile in modo automatico per via del limite (al momento non superabile) della lemmatizzazione automatica. La procedura è infatti incapace di riconoscere i verbi composti, formati da ausiliare e participio passato, che tratta erroneamente come forme verbali separate. Ciò limita fortemente le possibilità di esame, anche se, ai fini che ci interessano, dai dati disponibili si possono trarre alcuni spunti parziali.

Un esempio può essere offerto dalla frequenza del perfetto semplice, che mostra una precisa variazione all'interno del corpus considerato: il 57% delle forme riconosciute e assegnate a questo tempo verbale si registra infatti nel subcorpus tradotto. Anche se sarebbero necessari dati sulla frequenza del perfetto composto (non riconoscibile in modo automatico) per trarre indicazioni più sicure, la maggior presenza del perfetto semplice in traduzione può essere spiegata tenendo conto dell'influsso del *past simple* inglese (la lingua fonte per la maggior parte degli articoli tradotti), che ne favorisce la selezione, in parte opponendosi alla ben nota avanzata del perfetto composto a invadere gli ambiti d'uso dell'aoristo nell'italiano contemporaneo. In generale, i dati suggeriscono un radicamento di alcuni modi verbali nei testi tradotti, che potrebbe essere interpretato come segnale di un maggior rispetto della grammatica standard, in linea con la tendenza alla normalizzazione postulata dagli universali traduttivi: ad esempio, delle forme verbali riconosciute come modo condizionale, il 57% appartiene al subcorpus tradotto; una tendenza analoga (59% nel tradotto) si registra anche per le forme del modo congiuntivo. *En passant* notiamo che nell'intero corpus risulta sporadica (solo 10 casi riscontrati, di cui 4 nel subcorpus tradotto) la costruzione [se + imperfetto + imperfetto] (per es.: "se creavi una democrazia in Iraq, potevi tradurla in una democrazia in Libano e via di seguito"), alla quale viene evidentemente preferita la costruzione normativa con congiuntivo e condizionale.

Una conferma indiretta della tendenza alla normalizzazione derivante dalla preferenza accordata al congiuntivo nei testi tradotti proviene anche dall'analisi delle congiunzioni subordinanti che richiedono obbligatoriamente questo modo verbale, già di

per sé indice della ricerca di un registro elevato. Come si può notare dalla tabella 2.22, che riporta le frequenze assolute delle principali congiunzioni, la prevalenza nel subcorpus tradotto è evidente, con 249 casi rispetto ai 143 dei testi redatti originariamente in italiano. Eccezioni a questa tendenza sono *purché* e *benché*, più frequente nei testi non tradotti e di cui si segnala un caso di costruzione con l'indicativo in un articolo tratto da *Repubblica* ("Ma benché le crocifissioni tipo Giacometti incominciano a diffondersi...").

	Corriere tradotto	Corriere non tradotto	Repubblica tradotto	Repubblica non tradotto	Unità tradotto	Unità non tradotto	Subtotale tradotto	Subtotale non tradotto	Totale
<i>affinché</i>	25	7	45	6	25	8	95	21	116
<i>benché</i>	15	21	20	27	3	12	38	60	98
<i>sebbene</i>	24	3	34	2	34	2	92	7	99
<i>malgrado</i>	6	3	4	2	2	1	12	6	18
<i>nonostante (che)</i>	18	11	14	22	14	4	46	37	83
<i>purché</i>	6	7	6	6	0	8	12	21	33
<i>a patto che</i>	2	1	5	0	1	1	8	2	10
<i>a condizione che</i>	2	1	2	2	5	0	9	3	12
<i>nel caso in cui/che</i>	8	2	9	3	15	2	32	7	39
Totale									
n	106	56	139	70	99	38	344	164	508
d	0,36	0,19	0,32	0,16	0,38	0,15	0,35	0,16	0,26

Tabella 2.22 – Frequenza assoluta delle principali congiunzioni subordinanti con congiuntivo obbligatorio

2.3.2. Perifrasi [stare + gerundio]

Un caso di morfologia verbale per il quale è stata postulata l'influenza dell'inglese sull'italiano è la perifrasi progressiva [stare + gerundio], stimata in espansione nell'italiano contemporaneo. I dati sulla frequenza e densità nel corpus giornalistico in esame (tabella 2.23) mostrano in effetti una predominanza abbastanza significativa di questa costruzione nel subcorpus tradotto, interpretabile sulla base dell'interferenza dei testi originali inglesi.

	Corriere tradotto	Corriere non tradotto	Repubblica tradotto	Repubblica non tradotto	Unità tradotto	Unità non tradotto	Subtotale tradotto	Subtotale non tradotto	Totale
n	268	237	627	377	407	307	1.302	921	2.223
d	0,91	0,81	1,44	0,86	1,56	1,18	1,31	0,92	1,12

Tabella 2.23 – Dati assoluti e densità per 1.000 forme della perifrasi stare + gerundio

2.3.3. Frequenza di pronomi personali

Il settore della morfologia pronominale rappresenta un ambito di interesse per l'analisi della traduzione verso l'italiano per l'influenza che esercita da un lato la legge dell'interferenza linguistica e dall'altro l'universale traduttivo della normalizzazione.

Considerato che le lingue di partenza del subcorpus tradotto prevedono l'esplicitazione obbligatoria del soggetto (con l'eccezione dello spagnolo, comunque minoritario), è lecito attendersi un diverso comportamento dei pronomi soggetto nei due subcorpora. In effetti, i dati della tabella 2.24, che riporta la frequenza e la densità per 1.000 forme dei pronomi personali limitatamente a quelli usati in funzione di soggetto, confermano quest'ipotesi. La frequenza nettamente maggiore dei pronomi soggetto nei testi tradotti rispetto a quelli non tradotti

risulta una spia altamente probabile dell'interferenza del testo fonte, per cui si ha l'esplicitazione del soggetto anche nei casi in cui in italiano non sarebbe stato espresso. Si noti che i dati della tabella si riferiscono ai soli pronomi personali usati come soggetto, dopo che sono stati eliminati manualmente tutte le occorrenze in cui questi pronomi sono usati con altre funzioni (ad esempio i casi obliqui, come *con esso*, *tra di noi*, *a voi*, *secondo lei* o formule come *beati voi!* ecc.); è invece compreso l'uso enfatico del pronome (*siamo stati noi a metterli lì*).

Scendendo nel dettaglio, si deve osservare che la tendenza generale appare comunque diversificata in base alle varie forme: il pronome di II persona *tu*, comunque poco sfruttato per il carattere dei testi che compongono il corpus, è più frequente nei testi originali che in quelli tradotti. Tra i pronomi più frequenti (sia in assoluto, sia in termini di maggiore frequenza nel subcorpus tradotto) spiccano quelli di I persona, sia il plurale *noi*, sia il singolare *io*, probabilmente per la centralità assegnata all'autore dell'articolo, spesso opinionista di fama.

Non trova invece riscontro l'azione della normalizzazione, dal momento che i dati confermano la tendenza tipica dell'uso medio, che preferisce le forme *lui*, *lei*, *loro* al posto di *egli*, *ella*, *essa*, *essi*, *esse*. Le prime, in funzione di soggetto, prevalgono nei testi tradotti; si deve tuttavia osservare che anche *egli* e *essi* sono più frequenti nei testi tradotti rispetto a quelli originali, per quanto la frequenza complessiva di queste forme in funzione di soggetto non sia quantitativamente paragonabile agli altri casi.

	Corriere tradotto	Corriere non tradotto	Repubblica tradotto	Repubblica non tradotto	Unità tradotto	Unità non tradotto	Subtotale tradotto	Subtotale non tradotto	Totale
io	145	71	346	118	90	55	581	244	825
n	0,49	0,24	0,79	0,27	0,35	0,21	0,58	0,24	0,41
tu	6	9	29	29	7	18	42	56	98
n	0,02	0,03	0,07	0,07	0,03	0,07	0,04	0,06	0,05
egli	63	22	106	25	15	26	184	73	257
n	0,21	0,08	0,24	0,06	0,06	0,10	0,19	0,07	0,13
ella	0	1	1	0	2	0	3	1	4
n	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00
esso	23	24	33	23	9	18	65	65	130
n	0,08	0,08	0,08	0,05	0,03	0,07	0,07	0,07	0,07
essa	32	25	54	47	15	41	101	113	214
n	0,11	0,09	0,12	0,11	0,06	0,16	0,10	0,11	0,11
lui	79	94	119	139	37	61	235	294	529
n	0,27	0,32	0,27	0,32	0,14	0,23	0,24	0,29	0,27
lei	72	48	202	92	46	41	320	181	501
n	0,24	0,16	0,46	0,21	0,18	0,16	0,32	0,18	0,25
loro	29	33	73	47	37	24	139	104	243
n	0,10	0,11	0,17	0,11	0,14	0,09	0,14	0,10	0,12
noi	106	90	271	164	156	163	533	417	950
n	0,36	0,31	0,62	0,38	0,60	0,62	0,54	0,42	0,48
voi	18	6	22	20	14	8	54	34	88
n	0,06	0,02	0,05	0,05	0,05	0,03	0,05	0,03	0,04
essi	44	18	54	50	24	22	122	90	212
n	0,15	0,06	0,12	0,11	0,09	0,08	0,12	0,09	0,11
esse	16	11	13	17	6	8	35	36	71
n	0,05	0,04	0,03	0,04	0,02	0,03	0,04	0,04	0,04
totale	633	452	1.323	771	458	485	2.414	1.708	4.122
n	2,15	1,54	3,02	1,76	1,76	1,86	2,42	1,71	2,07

Tabella 2.24 – Frequenza assoluta e densità per 1.000 dei pronomi personali con funzione di soggetto

2.3.4. Frequenza di pronomi e aggettivi possessivi

Rientrano in un discorso analogo a quello dei pronomi soggetto appena discusso anche le tendenze riscontrate nel corpus per i possessivi, utilizzati più frequentemente in inglese e francese rispetto all'italiano, lingua che tende a preferire l'articolo determinativo ed eventualmente la forma pronominale del verbo.

Per quanto l'effettivo peso dell'influenza esogena sia stato spesso oggetto di discussione, anche in questo caso si può postulare l'azione dell'interferenza per spiegare la maggiore frequenza dei possessivi nei testi tradotti rispetto a quelli originali che emerge nettamente dalla tabella 2.25 (si tenga conto che l'analisi condotta sul corpus non distingue tra aggettivi e pronomi possessivi). In dettaglio, si nota anche qui una netta prevalenza in entrambi i subcorpora dei possessivi di III persona, sia singolare che plurale, seguiti da quelli di I persona, in particolare plurale. La maggiore frequenza nei testi tradotti è comunque comune a tutti i pronomi presi in esame.

	Corriere tradotto	Corriere non tradotto	Repubblica tradotto	Repubblica non tradotto	Unità tradotto	Unità non tradotto	Subtotale tradotto	Subtotale non tradotto	Totale
mio									
n	427	122	849	282	303	121	1.579	525	2.104
d	1,45	0,42	1,94	0,65	1,16	0,46	1,59	0,53	1,06
tuo									
n	25	17	63	32	42	25	130	74	204
d	0,08	0,06	0,14	0,07	0,16	0,10	0,13	0,07	0,10
suo									
n	1.695	1.480	2.409	1.623	1.145	1.118	5.249	4.221	9.470
d	5,73	5,03	5,49	3,70	4,38	4,27	5,25	4,22	4,76
nostro									
n	440	395	953	557	597	337	1.990	1.289	3.279
d	1,49	1,35	2,18	1,27	2,29	1,29	2,00	1,29	1,65
vostro									
n	26	6	42	13	29	8	97	27	124
d	0,09	0,02	0,10	0,03	0,11	0,03	0,10	0,03	0,06
loro									
n	632	423	1.088	599	696	360	2.416	1.382	3.798
d	2,14	1,44	2,49	1,37	2,67	1,38	2,42	1,38	1,91
totale									
n	3.245	2.443	5.404	3.106	2.812	1.969	11.461	7.518	18.979
d	10,91	8,28	12,24	7,06	10,68	7,49	11,40	7,48	9,53

Tabella 2.25 – Frequenza assoluta e densità per 1.000 dei pronomi e aggettivi possessivi

2.3.5. Frequenza di pronomi e aggettivi dimostrativi

Per quanto riguarda i dimostrativi, la maggiore frequenza di queste forme nei testi tradotti rispetto a quelli non tradotti (tabella 2.26; anche in questo caso l'analisi non distingue tra aggettivi e pronomi) può essere interpretata postulando l'azione dell'interferenza linguistica, dal momento che inglese e francese (ma anche tedesco), cioè le lingue di partenza più rappresentate nel subcorpus tradotto, fanno un uso più largo del dimostrativo.

Accanto all'interferenza linguistica, si può anche ipotizzare un'azione dell'universale traduttivo dell'esplicitazione, che porterebbe i testi tradotti a una maggiore frequenza di riprese anaforiche realizzate anche attraverso pronomi dimostrativi; una conferma viene dalla netta preponderanza del dimostrativo di prossimità *questo*, utilizzabile in senso endoforico. Nel corpus non si riscontra invece nessun caso di *codesto*, a riprova della definitiva scomparsa di questa forma dall'uso giornalistico.

lemma	Corriere tradotto	Corriere non tradotto	Repubblica tradotto	Repubblica non tradotto	Unità tradotto	Unità non tradotto	Subtotale tradotto	Subtotale non tradotto	Totale
questo									
n	1.882	1.292	2.955	2.217	1.761	1.197	6.598	4.706	11.304
d	6,36	4,40	6,73	5,05	6,72	4,57	6,59	4,70	5,68
quello									
n	1.168	1.260	1.981	1.947	930	1.152	4.079	4.359	8.438
d	3,96	4,29	4,52	4,44	3,56	4,40	4,09	4,35	4,24
medesimo									
n	19	4	26	14	25	11	70	29	99
d	0,06	0,01	0,06	0,03	0,10	0,04	0,07	0,03	0,05
ciò									
n	220	163	456	148	240	217	916	528	1.444
d	0,75	0,56	1,04	0,34	0,92	0,83	0,92	0,53	0,73
totale									
n	3.289	2.719	5.418	4.326	2.956	2.577	11.663	9.622	21.285
d	11,06	9,21	12,27	9,81	11,22	9,78	11,60	9,56	10,69

Tabella 2.26 – Frequenza assoluta e densità per 1.000 di pronomi e aggettivi dimostrativi

2.3.6. Altri aspetti morfologici

A livello morfologico, un ulteriore aspetto che si può prendere in considerazione in questo *excursus* è la maggiore presenza di *che cosa* nei testi tradotti rispetto al subcorpus di controllo di testi non tradotti (tabella 2.27), che può con cautela essere assunto come conferma di una maggiore conservatività delle traduzioni. *Cosa* tende infatti a prevalere nell'intero corpus rispetto al più formale *che cosa*, in particolare nei testi tradotti rispetto a quelli scritti originariamente in italiano.

	Corriere tradotto	Corriere non tradotto	Repubblica tradotto	Repubblica non tradotto	Unità tradotto	Unità non tradotto	Subtotale tradotto	Subtotale non tradotto	Totale
<i>che cosa</i>									
n	86	69	174	82	26	57	286	208	494
d	0,29	0,24	0,40	0,19	0,10	0,22	0,29	0,21	0,25
<i>cosa</i>									
n	91	75	159	102	147	99	397	276	673
d	0,31	0,26	0,36	0,23	0,56	0,38	0,40	0,28	0,34

Tabella 2.27 – Frequenza assoluta e densità per 1.000 di *che cosa* e *cosa*

Inoltre, anche se la sporadicità delle attestazioni (rispettivamente 6 e 20 nel corpus) non consente di trarre conclusioni stabili, la preponderanza nei testi tradotti rispetto ai non tradotti di forme come “questo è ciò che...”, ma soprattutto dell'incontro tra dimostrativo di prossimità e di lontananza come in “questo è quello che...”, è probabilmente legata a un fenomeno di interferenza dell'inglese, che prevede comunemente l'impiego di *what* come pronome relativo doppio (per es. *this is what I feel*), mentre *quanto* in italiano risulta piuttosto ricercato.

Infine, la ricerca di *niente* in funzione aggettivale produce 50 occorrenze nel corpus non tradotto contro le 28 del corpus tradotto, lasciando pensare che i giornalisti italiani abbiano adottato questo tratto dell'italiano dell'uso medio con maggior facilità dei traduttori.

2.4. Sintassi: lunghezza dei periodi e leggibilità

Anche per quanto concerne la strutturazione del periodo, è ipotizzabile l'azione combinata di più di un universale traduttivo. Da

una parte l'influenza del testo fonte, per cui i traduttori tendono a ricalcare la punteggiatura dell'originale, dall'altra la tendenza alla semplificazione condurrebbe alla riduzione della lunghezza media dei periodi.

Naturalmente, fare corrispondere lunghezza dei periodi e complessità sintattica rappresenta una forzatura innegabile. Tuttavia, nell'impossibilità di condurre un'analisi così fine tramite strumenti automatici, vista la dimensione del corpus, abbiamo preferito limitarci a misurare la consistenza dei periodi in parole, per poi raggrupparli in classi in base alla di lunghezza (tabella 2.28).

	Corriere tradotto	Corriere non tradotto	Repubblica tradotto	Repubblica non tradotto	Unità tradotto	Unità non tradotto	Subtotale tradotto	Subtotale non tradotto	Totale
media articoli (parole)	1.010,69	759,86	1.061,07	1.158,66	1.029,35	1.060,57	1.040,99	989,13	1.014,36
media periodo (parole)	22,75	24,67	21,07	21,36	24,37	22,11	22,73	22,71	22,72
% da 1 a 7 parole	12,51	10,09	15,57	15,14	8,91	16,99	12,33	14,07	13,20
% da 8 a 20 parole	39,58	37,25	40,70	40,91	38,48	39,44	39,59	39,20	39,40
% oltre 20 parole	47,91	52,65	43,73	43,95	52,61	43,57	48,08	46,72	47,40

Tabella 2.28 – Lunghezza media di articoli e periodi e distribuzione dei periodi in base alla lunghezza

La lunghezza media dei periodi risulta sostanzialmente uguale nei due subcorpora, anche se ci si attendeva un valore inferiore nei testi tradotti: essendo il subcorpus in questione composto per lo più da articoli provenienti da testate angloamericane, era ipotizzabile che la tendenza alla paratassi e alla strutturazione in periodi brevi, propria dell'inglese moderno, avrebbe lasciato traccia nelle traduzioni italiane. Tuttavia, si registrano differenze interessanti nella distribuzione dei periodi in classi di lunghezza: gli articoli scritti originariamente in italiano presentano una percentuale leggermente maggiore di periodi molto brevi (da 1 a 7 parole) e un valore legger-

mente inferiore (sempre nell'ordine dell'1,5%) di periodi composti da 20 parole e più. Una spiegazione potrebbe essere ricercata nella marcata preferenza della stampa italiana per le nominalizzazioni e uno stile franto, che si opporrebbe al periodare più sobrio e piano della stampa di qualità di lingua inglese.

Alla ricerca di eventuali conferme, si è proceduto a verificare la distribuzione dei segnali di interpunzione nel corpus (tabella 2.29). Gli studi sugli universali traduttivi basati sulla *corpus linguistics* indicano che il traduttore tende a intervenire sulla punteggiatura con intenti semplificatori e chiarificatori; in particolare, i segni di interpunzione più deboli sarebbero sostituiti da segni dotati di forza maggiore, per cui la virgola avrebbe la tendenza a trasformarsi in punto e virgola, il punto e virgola in punto fermo ecc.

Segno	Subcorpus tradotto			Subcorpus non tradotto			Totale		
	n	% di colonna	% in rapporto a forme grafiche	N	% di colonna	% in rapporto a forme grafiche	N	% di colonna	% in rapporto a forme grafiche
.	42.202	39,68	4,07	42.296	33,89	4,07	84.498	36,56	4,24
,	50.625	47,60	4,85	66.880	53,60	6,29	117.505	50,84	5,90
-	5.138	4,83	0,51	5.542	4,44	0,55	10.680	4,62	0,54
:	3.810	3,58	0,38	5.831	4,67	0,58	9.641	4,17	0,48
;	1.183	1,11	0,10	1.040	0,83	0,12	2.223	0,96	0,11
?	2.937	2,76	0,29	2.880	2,31	0,29	5.817	2,52	0,29
!	451	0,42	0,05	318	0,25	0,03	769	0,33	0,04
Totale	106.346	100,00	9,66	124.787	100,00	11,12	231.133	100,00	11,61

Tabella 2.29 – Distribuzione dei segni di interpunzione

I calcoli evidenziano una maggiore frequenza dei due punti e della virgola nel subcorpus non tradotto e del punto esclamativo nelle traduzioni. Se quest'ultimo risultato può essere dovuto semplicemente a fattori non dipendenti dal processo traduttivo (per es. scelte individuali degli autori), ci si chiede che cosa possa giustificare il fatto che virgole e due punti siano più numerosi nei testi originariamente scritti in italiano, che pure sono caratterizzati da un numero maggiore di periodi brevi e quindi, presumibilmente, meno complessi dal punto di vista sintattico. Alla luce della più debole presenza dei connettivi nel subcorpus tradotto (cfr. § 2.5.), non è possibile sostenere che nelle traduzioni virgole e due punti si siano trasformati in elementi lessicali per soddisfare i bisogni dell'esplicitazione. Peraltro, questa ipotesi prevedrebbe un maggior numero di segnali interpuntivi 'forti' per sopperire alla sparizione delle virgole in traduzione, cosa che non si verifica, essendo il numero di punti fermi e punto e virgola sostanzialmente identico nei due subcorpora.

2.4.1. *Costrutti ad sensum*

Come spesso avviene in questo studio, ci si attende che nel caso dei costrutti *ad sensum* entrino in gioco due spinte divergenti: da una parte potrebbe farsi sentire l'influenza delle lingue di partenza (con parziale eccezione dello spagnolo), che prevedono il verbo al plurale dopo percentuali e nomi collettivi, anche se formalmente singolari; dall'altra i traduttori, consapevoli del rischio, dovrebbero fare attenzione e rifugiarsi nella norma grammaticale, evitando un tratto considerato neostandard. Abbiamo dunque ricercato espressioni di significato plurale e forma singolare, come *la maggioranza di, una/gran/grande parte di, la metà di* percentuali e frazioni con numeratore singolare (assenti), ma non quando si tratta di frasi specificative, mentre non abbiamo considerato i nomi collettivi veri e propri (per quanto possa esservi un influsso esogeno, il verbo al plurale riferito a soggetti come *squadra* o *partito* ci appare pur sempre per lo meno substandard).

	Subcorpus tradotto	Subcorpus non tradotto	Totale
<i>maggioranza</i>	7	2	9
<i>parte</i>	48	14	62
<i>metà</i>	7	6	13
percentuali	7	6	13
<i>(mezzo) milione</i>	4	7	11
<i>(mezzo) miliardo</i>	0	0	0
<i>decina</i>	0	2	2
<i>(mezza) dozzina</i>	3	2	5
Totale	69	28	97

Tabella 2.30 – Concordanze *ad sensum* con alcuni quantificatori

I numerali ‘singolari’, come *un milione, un miliardo, una decina, dozzina* ecc. (ci siamo limitati a questi quattro), in genere prediligono il verbo al singolare, ma sembrano subire una certa influenza del contesto in cui sono inseriti. Per quanto il numero complessivo delle occorrenze sia così limitato da impedire di giungere a conclusioni certe, va osservato che con queste forme il verbo accordato al plurale sembra essere più numeroso nel subcorpus non tradotto, fatto salvo il caso di *dozzina*, probabilmente per influsso dell’inglese *dozen* (compaiono 13 occorrenze di *dozzina* nel subcorpus non tradotto, 20 nel tradotto; soprattutto, compaiono in 11 testi tradotti dall’inglese e 9 la cui lingua fonte è indefinita, ma in nessuna traduzione da altre lingue). Al contrario, sempre tenendo conto dei numeri ridotti in gioco, in presenza degli altri quantificatori considerati, le concordanze *ad sensum* (che continuano a rappresentare una piccola minoranza rispetto ai casi di verbo al singolare) tendono a essere più frequenti nelle traduzioni, indicando che l’influenza della lingua straniera ha la meglio sulla tendenza a rispettare la grammatica normativa.

Tra i casi notevoli, segnaliamo un (unico) esempio di accordo di genere invece che di numero: “il 15% della capacità produttiva potrebbe essere andata perduta” (in traduzione pubblicata sull’*Uni-*

tà); inoltre, non è rara l'alternanza tra accordi *ad sensum* e rispetto della norma nello stesso testo, anche a breve distanza: "Il 76% dei venezuelani (pro e contro Chavez) ammettono che il braccio di ferro è fallito. Solo il 19% sostiene di essere a un passo dalla vittoria" (*Unità* non tradotto). Ricordiamo tuttavia che nel corpus la concordanza *ad sensum* risulta una scelta minoritaria, anche se la condizione di debolezza del quantificatore in qualità di testa del sintagma è confermata dalle relative e participiali, che selezionano costantemente l'antecedente immediatamente a sinistra:

i sovietici avevano al lavoro nella ricerca spaziale oltre un milione di tecnici e manager, i quali dovevano ricevere i fondi dalla piramide di enti economici organizzati nel Gosplan; un milione di fotografie provenienti dagli archivi di *The Daily Worker's*; un'indicazione della portata di questa repressione è data dall'esodo di oltre un milione di curdi, fuggiti dalle campagne per cercare scampo a Diyarbakir; più di un milione di posti di lavoro persi da quando la recessione è ufficialmente finita; ecc.

2.4.2. Calchi sintattici

L'analisi dei calchi sintattici qui presentata si basa principalmente su uno studio di Daniele Grasso, a cui rimandiamo per una trattazione più approfondita degli aspetti teorici e di cui adottiamo la distinzione preliminare tra sintagmi fissi e calchi sintattici veri e propri: entrambi sarebbero casi specifici di prestito esogeno gradualmente acclimatatosi nella lingua di accoglienza, ma mentre i sintagmi fissi semplicemente si ripetono, i calchi sintattici si distinguono per la loro produttività. Una volta entrato in italiano, lo schema strutturale deve rivelarsi funzionale ai bisogni comunicativi dei parlanti, presentarsi con una certa frequenza e, soprattutto, non esclusivamente in contesti caratterizzabili come "traduzioni invisibili" (cfr. § 1.3.1.). Quest'ultimo requisito pone dei limiti alle eventuali conclusioni che è possibile trarre dal corpus di testi giornalistici su cui stiamo conducendo le nostre indagini: se infatti è palese che il subcorpus tradotto non permette di individuare calchi sintattici veri e propri, sorgono dubbi anche a proposito del subcorpus di testi scritti da giornalisti italiani, non potendo escludersi l'influsso di materiale linguistico esogeno. Tuttavia, in base alla maggiore o minore frequenza dei costrutti,

sia in senso assoluto sia nel confronto tra subcorpora, possono emergere indicazioni utili a definire lo status di calco sintattico “in via di acclimatamento” o calco sintattico maturo della struttura considerata.

Passiamo dunque in rassegna i fenomeni secondo la tripartizione che distingue tra pseudo calchi sintattici (essenzialmente strutture già presenti nei corpora d’italiano antico oppure assenti nell’italiano standard o ancora non derivanti da influssi esogeni), calchi sintattici patrimoniali rinforzati/funzionali (si ha un doppio contributo esogeno ed endogeno all’affermazione della struttura in italiano) e calchi sintattici veri e propri (strutture di chiara derivazione straniera che si sono rivelate produttive e risultano presenti anche nell’italiano non esposto a fenomeni di “traduzione invisibile”). Dell’elenco di costrutti trattati da Grasso, dal nostro corpus risultano assenti l’impiego anomalo dell’avverbiale *dove* (per es. *qui è dove voglio trovarmi*), lo slittamento funzionale di *tu (non) vuoi* (per es. *tu non vuoi andare a piedi così lontano*), l’uso determinato anomalo del sostantivo *destra* (per es. *alla destra nella foto...*; naturalmente non si considerano i casi in cui *destra* e *sinistra* sono determinati ma indicano gli schieramenti politici), il modulo verbo rogandi + *qualcuno* + *di* + infinito (per es. *richiesti di rispondere, scrollano le spalle*), la lessia *grazie di* rivolta ad azioni future (per es. *grazie di rispondermi al più presto*), la lessia *slow* (per es. *cittaslow*), la lessia *al meglio di* (per es. *si gioca al meglio delle 7 partite*), il nuovo impiego delle preposizioni *in* e *out* (per es. *a Ginevra la bici è molto in*), il nuovo introduttore di domanda retorica (per es. *stai scherzando o cosa?*) e il nuovo impiego dell’avverbio *come* (per es. *Terence Hill come Don Matteo*). Di queste strutture, le prime due sono classificate da Grasso come pseudo calchi sintattici (in particolare, nel caso di *tu non vuoi* verrebbe a mancare il requisito della produttività), le due successive come calchi patrimoniali rinforzati e tutti i restanti calchi sintattici veri e propri.

Per quanto riguarda gli pseudo calchi sintattici, la prefissazione con la congiunzione *quasi* seguita da sostantivi astratti risulta scarsamente frequente: compare infatti 6 volte in tutto, ma se si escludono 4 occorrenze di *quasi totalità* (il cui status di quantificatore rende problematica l’attribuzione alla classe dei “termini astratti” – per Grasso questo sarebbe l’unico caso a risentire direttamente dell’influsso dell’inglese) risultano solo una *quasi*

eliminazione nel subcorpus tradotto e una *quasi timidezza* nel subcorpus non tradotto. Per individuare, invece, le “ipotetiche bi-affermative concessive” (del tipo *se Atene piange, Sparta non ride*), si è limitata la ricerca allo stereotipo *se è vero che*, individuando 14 occorrenze in totale (5 nel subcorpus non tradotto, 9 nel subcorpus tradotto), nella maggior parte dei casi (11) caratterizzate da una qualche forma rafforzativa (*se è vero che ..., è altrettanto/ancora/anche/nondimeno/tuttavia/ugualmente vero che ...*).

Nel corpus i casi di comparazione ellittica (per es. “Lei, signor Berque, ha il merito di approfondire meglio il passato”; 19 occorrenze nel subcorpus non tradotto, 11 in quello tradotto) limitatamente alla forma *meglio* (non sono stati considerati gli avverbiali *di più* e *di meno*), risultano particolarmente frequenti in combinazione con i verbi *capire*, *comprendere* e *conoscere* (24 occorrenze in totale). L’impiego della preposizione *su* con i nomi di città (per es. “la ‘posta prioritaria’ in partenza da Napoli registra appena il 53% di puntualità con destinazione Avellino e il 57% su Benevento”) risulta più frequente negli articoli scritti originariamente in italiano (3 contro 1 traduzione), mentre l’estensione funzionale della stessa preposizione (per es. “raggiungerla sul cellulare”) compare solo in una traduzione. Per finire, la lessia *tra le altre cose* (per es. *il Congresso, tra le altre cose, istituiva la nomina di un osservatore*) compare solo in tre traduzioni.

Due strutture indicate da Grasso come pseudo calchi sintattici non sono state prese in considerazione nello spoglio del nostro corpus. La prima è l’impiego dei verbi modali ‘all’inglese’ in completiva (per es. *è normale che il governo possa cambiare orientamento*). Per quanto tale impiego potrebbe riflettere l’uso dei modali *may* e *should*, valutare la ridondanza dei modali ci sembra alquanto complicato perché entrano in gioco fattori contestuali e pragmatici: data l’attestazione del costrutto anche in italiano antico, la scelta tra *è normale che il governo possa cambiare* e *cambi* sembra legata principalmente all’atteggiamento del parlante. L’altra struttura che si è rinunciato a cercare nel corpus riguarda l’omissione della preposizione (per es. *si rivolsero a lui come persona di fiducia*) per l’evidente accettabilità di questo costrutto nell’italiano normativo. Tra l’altro, è l’inserimento di una preposizione (ed eventualmente dell’articolo indeterminativo) a produrre risultati di dubbia accettabilità.

Passando ai calchi sintattici patrimoniali, la struttura che è risultata più frequente nel nostro corpus è stata il modulo *non* + sostantivo (per es. *nonsense*), con un totale di 49 occorrenze (26 nel subcorpus non tradotto e 23 in traduzione). È però utile distinguere i casi in cui la negazione sembra riguardare principalmente un verbo sottostante, come nei participi presenti (per es. *i non credenti* o *non udenti*; si tratta di forme lessicalizzate che registrano 7 occorrenze, tutte in testi non tradotti) e passati (per es. *il non detto*, *i non allineati*; 1 occorrenza nel subcorpus non tradotto, 3 nel tradotto) e le nominalizzazioni (per es. *la non menzione*, *la non approvazione*, *il non acquisto*, *in non invio*; 4 casi nel subcorpus non tradotto e 5 in quello tradotto): in quest'ultima categoria il modulo *non* + sostantivo si propone come alternativa meno frequente alla struttura *mancato* + sostantivo (es. *mancato ripensamento*; 40 occorrenze nel corpus).

Per quanto riguarda i nomi non deverbali, alcuni possono essere considerati calchi di sintagmi fissi o quasi, come *non violenza* e *non luogo* (a cui si aggiunge il deverbale *non proliferazione*), che registrano 2 occorrenze nei testi originali e 2 in quelli tradotti, mentre un'altra struttura consiste nel binomio antitetico del tipo *le differenze e i confronti tra i diritti e i non diritti*, ma anche *il quantificabile e non quantificabile*, *sia il vivente che il non vivente* ecc. (3 testi originali e 1 tradotto). Si tratta della negazione dell'intero carico semantico del sostantivo, che divide il mondo in due: *x* e *non-x*, fenomeno abbastanza comune (anche in assenza dei binomi visti sopra) con i nomi plurali riferiti a nazionalità (*non afgiani*, *non serbi*, *non americani* ecc.) o di altri gruppi altrimenti individuati (*non fumatori*, *non esperti*, *non specialisti* ecc.; 7 occorrenze in non tradotto, 6 in traduzione). Meno comune nel nostro corpus l'impiego del *non* con funzione privativa e accezione vagamente spregiativa e a volte ironica, come in *non realismo*, *non divertimento*, *non memoria*, *non verità* e anche *non vegetariano* ecc. (3 occorrenze in non tradotto, 4 in traduzione).

L'altro calco sintattico patrimoniale che emerge abbastanza numeroso nel corpus è l'estensione funzionale della preposizione temporale *dopo*, come in *anno dopo anno*, *sera dopo sera*, *giorno dopo giorno* (16 occorrenze nei testi originali, 19 in quelli tradotti), mentre l'ordine SV al posto di quello canonico VS con i verbi inaccusativi, una volta espunte le anticipazioni del soggetto dovute a fattori pragmatici e co-testuali (per es. i casi di

SN determinato e progressione tematica, come in *prima ancora che questa guerra sia scoppiata*) risulta fortemente minoritario. Mentre risulta assente l'influsso esogeno sulla flessione di latinismi invariabili (*focus* e *corpus* compaiono solo al singolare, per lo più in polirematiche mutate dall'inglese come *habeas corpus* e *focus group*), poco numerosi sono gli esempi della lesia *la madre di tutte ...* (4 in non tradotto e 3 in traduzione), anche se è notevole il seguente caso di mancato accordo di genere in una traduzione:

Può darsi ci sia ancora molto da attendere prima che Saddam faccia la sua comparsa dinanzi ad un tribunale per il processo avente per oggetto la 'madre di tutti i crimini di guerra'.

Infine, tra quelli che Grasso considera calchi sintattici veri e propri, abbiamo rinvenuto una sola occorrenza della struttura * è *bello* in un articolo originariamente scritto in italiano, anche se si tratta evidentemente del risultato di una traduzione invisibile. Qualche occorrenza in più (5 in non tradotto e 3 in traduzione) si registra per il superlativo relativo all'inglese (per es. "la prima, seconda e quarta società più ricche del mondo"), particolarmente utile nel caso di graduatorie e classifiche, e per le cosiddette "interrogative multiple", del tipo *chi è chi* e *chi ha fatto che cosa* (2 in non tradotto e 4 in traduzione, anche con più di due elementi interrogativi: "chi attribuirà chi a chi?"). Ma il calco sintattico che più ha preso piede nella stampa italiana risulta essere il riflessivo rafforzato, del tipo *autofinanziarsi*; la ricerca nel nostro corpus ha prodotto 41 occorrenze (solo 6 le scritture con trattino, come in *auto-replicarsi*), di cui 20 nel subcorpus non tradotto e 21 in quello tradotto.

Più che sull'imitazione morfologica (in inglese *auto-* è reso con *self-*), l'introduzione del prefissoide *auto-* in italiano sembra rispondere all'esigenza di rafforzare il clitico normalmente utilizzato per la riflessività. *Auto-* sembra utile ad arricchire il paradigma delle possibilità legate alle costruzione riflessiva in italiano. Prendiamo un verbo come *correggere*: usato in senso transitivo, l'alternativa tra forma atona (*lo correggo*) e tonica (*correggo lui*) dipende dalle ben note considerazioni pragmatiche (per es. *correggo lui e non te*). Nel caso della costruzione riflessiva, una frase come *ha sbagliato, ma si è corretto subito*, detta di uno studente che sostiene

un esame, è perfettamente accettabile. La forma tonica *ha corretto se stesso* funziona bene in un contesto che risulti soddisfacente dal punto di vista pragmatico (per es. *ha corretto se stesso, ma agli altri non ha detto niente*). *Si è auto corretto* sembra piuttosto fare le veci di *si è corretto da solo* (per es.: *durante l'esame ha detto una sciocchezza, ma si è subito corretto da solo* oppure *si è subito auto corretto*, ma non *ha subito corretto se stesso*). L'insufficienza del semplice clitico pare legata ai valori anticausativo (il mancato intervento dell'agente nel verificarsi dell'azione, come in *autoaccendersi*) e focalizzante (l'agente non è quello atteso, come nel caso dello studente dell'esempio sopra, che sostituisce il professore). Così, tra le forme individuate nel corpus, oltre ai più comuni *autofinanziarsi* e *autogovernarsi*, tra gli usi focalizzanti spiccano *autorapirsi* e *autoricusarsi*, mentre il valore anticausativo è ben rappresentato da *autoalimentarsi* e *autodistruggersi*.

In un unico articolo originariamente scritto in italiano, il prefissoide *auto-* è stato utilizzato con un verbo intransitivo pronominale: "mi sono interrogato fra angosce e speranze su come potrà autoevolversi l'uomo nei prossimi millenni". Il testo tratta della manipolazione genetica e la forma verbale può essere parafrasata con *evolversi da solo*, cioè senza sottostare all'evoluzione naturale, confermando la funzione focalizzante non limitata alle forme riflessive.

In aggiunta ai costrutti analizzati da Grasso, sono stati presi in considerazione due ulteriori possibili casi di influenza esogena, in particolare dell'inglese: si tratta delle reggenze del verbo e deverbale *domandare/domanda* e dell'aggettivo e sostantivo *responsabile/responsabilità*. Nel primo caso, tutte le fonti consultate confermano che l'unica preposizione ammessa per introdurre l'oggetto della domanda è *di*, quando invece lo spoglio del nostro corpus rivela che in tre testi tradotti il sostantivo regge un argomento introdotto da *per*, come nell'esempio che segue:

il prezzo del petrolio oltre i 100 dollari al barile, assieme a una valanga di sussidi governativi, specie negli Stati Uniti, ha innescato una forte domanda per il mais da trasformare in etanolo, come carburante sostitutivo della benzina.

Di queste tre traduzioni, una è di certo dall'inglese, mentre delle altre due la lingua fonte non è certa, anche se l'argomen-

to (la ripresa economica negli USA) fa presumere che si tratti ancora dell'inglese. In conformità con i criteri seguiti da Grasso, visto che evidentemente non si può parlare di produttività (il modulo si limita al senso economico del termine) e ci troviamo davanti all'esito (incerto) di una traduzione, in mancanza di riscontri in altri tipi di testo, il costruito va classificato come pseudo calco sintattico.

Il caso di *responsabile* e *responsabilità* è più interessante. Il GRADIT riporta la possibilità di introdurre un argomento tramite la preposizione *per* esclusivamente nel caso di *responsabilità civile* ("responsabilità per danni a persone o cose provocati dalla circolazione di veicoli a motore"), e GDLI ammette la stessa preposizione nel caso di *responsabile civile*, *responsabilità civile automobilistica*, *responsabilità politica*, *responsabilità patrimoniale* e, da ultimo nel significato di "funzione di controllo e di coordinamento nell'ambito di un'attività lavorativa, amministrativa, tecnica". Tolti i casi in cui *per* sta *per conto di* (come in "il responsabile sanitario per l'Unione Europea, David Byrne"), abbiamo individuato 5 occorrenze nel subcorpus non tradotto e 19 in quello tradotto. Accanto a casi di probabile traduzione invisibile (come in "il Gramm-Leach-Bliley Act trasferì la responsabilità per la sorveglianza delle banche di investimento dalla Federal Reserve alla Sec"), gli articoli scritti originariamente in italiano presentano anche esempi che non sembrano spiegabili alla luce del contatto interlinguistico (per es. "La relazione [del Ministero per l'ambiente italiano] fornisce anche l'attribuzione numerica delle responsabilità per questi 'eccessi'"): sono solo 8 (tra l'altro tutti in traduzione) i casi riconducibili alla specializzazione semantica di *responsabile*/*responsabilità* (per es. "il responsabile per la politica estera della Ue, Javier Solana"; "un Commissario con responsabilità speciali per i Balcani"). Ad ogni modo, anche se dai dati del nostro spoglio non si può escludere un certo rafforzamento esogeno, in virtù della precedente attestazione in italiano il costruito va annoverato tra gli pseudo calchi sintattici.

Proponiamo una tabella riassuntiva dei fenomeni indagati in questo paragrafo. Data l'esiguità dei riscontri, escludiamo i dati relativi al quantificatore *la quasi totalità* e alla posizione preverbiale del soggetto con verbo in accusativo.

Struttura	Subcorpus tradotto	Subcorpus non tradotto	Occorrenze totali
<i>quasi + n</i>	1	1	2
<i>ipotetiche bi affermative</i>	9	5	14
<i>comparazione ellittica</i>	11	19	30
<i>su + città</i>	1	3	4
<i>sul cellulare</i>	1	0	1
<i>tra le altre cose</i>	3	0	3
<i>domanda per</i>	3	0	3
<i>responsabile/responsabilità per</i>	19	5	24
Totale pseudo calchi sintattici	48	33	81
<i>non + n</i>	23	26	49
<i>t dopo t</i>	19	16	35
<i>madre di tutte...</i>	3	4	7
Totale calchi patrimoniali rinforzati	45	46	91
<i>riflessivo rinforzato</i>	21	20	41
<i>interrogativa multipla</i>	4	2	6
<i>superlativo all'inglese</i>	3	5	8
<i>agg. è bello</i>	0	1	1
Totale calchi sintattici veri e propri	28	28	56

Tabella 2.31 – Tipologia e occorrenze dei calchi sintattici. Dati assoluti

Come si può notare, scarse sono le differenze tra i subcorpora. Tra gli pseudo calchi sintattici, le “ipotetiche biaffermative” risultano più frequenti in traduzione (ma si consideri il numero limitato di occorrenze rinvenute), senza che sia possibile collegare questa struttura all’influsso di una specifica lingua fonte: delle 9 occorrenze prodotte dal subcorpus tradotto, una sola rimanda a una traduzione dal francese, contro due dall’inglese, una dallo spagnolo e cinque di provenienza incerta.

La preponderanza della comparazione ellittica tra i testi originali conferma la diffusione del costrutto in italiano; al contrario, l’uso della preposizione *su* con nomi di città (abbiamo ritenuto inutile procedere allo spoglio relativo a *in* e *a*) e in costrutti del tipo *sul cellulare* (il corpus contiene 4 occorrenze di *al cellulare*, tutte in traduzione, e 29 di *al telefono*) risulta molto rara, come minoritaria è la scelta di *tra le altre cose* in alternativa a *tra l’altro* (63 occorrenze in totale). Le poche occorrenze di *la domanda per* sono classificabili come casi di interferenza nella traduzione, mentre il costrutto *responsabile/responsabilità per* potrebbe sfociare in un ampliamento degli ambiti d’uso di una preposizione che oggi riflette una specializzazione semantica (ma notiamo che l’interrogazione del corpus produce ben 294 occorrenze di *responsabile/responsabilità di*). Di tutti gli altri fenomeni, per i quali secondo Grasso il ruolo delle spinte esogene dovrebbe essere maggiore, l’analisi del corpus non rivela alcuna differenza degna di nota tra materiale non tradotto e materiale tradotto.

2.5. Testualità: l’uso dei connettivi

L’azione dell’esplicitazione porterebbe intuitivamente a presupporre un diverso ruolo dei connettivi nel corpus di testi tradotti rispetto al corpus di controllo di testi scritti originariamente in italiano, con ricadute nella frequenza statistica. In quest’ottica, gli studi di ambito anglosassone sulla “*explicitation hypothesis*” considerano i connettivi una delle risorse che tendono a essere sovrasfruttate dal traduttore allo scopo di mettere in risalto coesione e coerenza del testo.

Tuttavia, i dati non sembrano confermare questa previsione. Come si può notare nella tabella 2.32, che riporta i dati assoluti di alcuni elementi interfrastici che svolgono una funzione di esplicitazione dei nessi logici del testo, i connettivi, pur con oscillazioni di

rilievo e con l'eccezione di *tuttavia*, sono sempre più frequenti nel subcorpus non tradotto rispetto a quello tradotto.

	Corriere tradotto	Corriere non tradotto	Repubblica tradotto	Repubblica non tradotto	Unità tradotto	Unità non tradotto	Subtotale tradotto	Subtotale non tradotto	Totale
<i>ciò</i>	58	95	35	173	23	59	116	327	443
<i>dunque</i>	42	85	75	212	17	92	134	389	389
<i>infatti</i>	69	105	69	351	45	68	182	524	707
<i>perciò</i>	21	32	50	56	5	23	76	111	187
<i>quindi</i>	77	106	158	211	127	65	362	382	744
<i>tuttavia</i>	183	51	110	89	80	72	373	212	585

Tabella 2.32 - Frequenza assoluta di alcuni connettivi interfrastici

La situazione non muta se si prendono in considerazione, al posto di singoli connettivi, intere classi, elaborate sulla base degli elenchi disponibili nelle principali grammatiche attente a questo aspetto della testualità. Contrariamente alle attese, come si può notare nella tabella 2.33, la frequenza assoluta di tutti i connettivi presi in considerazione e la densità per 1.000 forme nei vari subcorpora indicano una maggiore presenza nei testi originali rispetto a quelli tradotti. Questa tendenza è comune alle varie copie di subcorpora e a tutte le categorie di connettivi, con l'unica eccezione dei continuativi (come già registrato per *tuttavia*, che di questa categoria fa parte).

	Corriere tradotto	Corriere non tradotto	Repubblica tradotto	Repubblica non tradotto	Unità tradotto	Unità non tradotto	Subtotale tradotto	Subtotale non tradotto	Totale
Additivi									
n	9.061	10.116	13.085	15.815	7.818	9.169	29.964	35.100	65.064
d	30,81	34,58	30,00	36,21	30,02	35,14	30,14	35,20	32,68
Avversativi									
n	2.230	2.929	3.150	3.640	1.625	2.010	7.005	8.579	15.584
d	7,58	10,01	7,22	8,33	6,24	7,70	7,05	8,60	7,83
Comparativi									
n	1.791	2.429	2.602	3.155	1.330	1.713	5.723	7.297	13.020
d	6,09	8,30	5,97	7,22	5,11	6,57	5,75	7,32	6,54
Esplicativi									
n	158	214	130	545	78	137	366	896	1.262
d	0,54	0,73	0,30	1,25	0,30	0,53	0,37	0,90	0,63
Consecutivi									
n	774	845	1.249	1.573	654	683	2.677	3.101	5.778
d	2,63	2,89	2,86	3,60	2,51	2,62	2,69	3,11	2,90
Continuativi									
n	235	154	196	194	136	134	567	482	1.049
d	0,80	0,53	0,45	0,44	0,52	0,51	0,57	0,48	0,53
Enumerativi									
n	396	411	543	778	256	315	1.195	1.504	2.699
d	1,35	1,40	1,25	1,78	0,98	1,21	1,20	1,51	1,26
Temporali									
n	2.683	3.352	4.634	5.042	2.376	2.768	9.693	11.162	20.855
d	9,12	11,46	10,63	11,54	9,12	10,61	9,75	11,20	10,47
Tutti									
n	14.837	17.208	21.873	26.140	12.322	14.606	49.032	57.954	106.986
d	50,45	58,82	50,16	59,84	47,31	55,98	49,32	58,13	53,73

Tabella 2.33 – Densità per 1.000 forme delle principali classi di connettivi

Questo dato va in controtendenza rispetto alle attese teoriche fondate sull'universale traduttivo dell'esplicitazione. Un'interpretazione può essere legata al genere testuale preso in considerazione nella costruzione del corpus giornalistico esaminato – l'articolo di commento – con una prevalenza di temi di politica internazionale e caratterizzato da testi dal carattere argomentativo, in cui probabilmente il ruolo dei connettivi appare stilisticamente ben consolidato e meno facilmente negoziabile nel processo traduttivo rispetto ad altre tipologie testuali.

Con particolare riferimento alle lingue in gioco (principalmente inglese e italiano), in quest'ottica si deve comunque notare che i risultati qui ottenuti sono in linea con quanto rilevato anche in altri studi. Il confronto tra corpora paragonabili monolingui e corpora paralleli bilingui indica che le eventuali aggiunte esplicative da parte dei traduttori potrebbero anche essere il segnale del tentativo di conformarsi a una convenzione stilistica italiana. In altre parole, in certe tipologie testuali l'italiano prevede un impiego maggiore di connettivi rispetto alla norma vigente in testi paralleli in lingua inglese e, anche se i traduttori possono inserire un certo numero di elementi in conformità con l'universale dell'esplicitazione, non riescono a colmare la differenza. Detto altrimenti, per ricondurre il discorso agli universali traduttivi a cui è improntata la nostra analisi, sembrerebbe che in questo caso la legge dell'interferenza annulli l'effetto della tendenza all'esplicitazione.

2.6. Trattamento del corpus per il calcolo della distanza intertestuale

Per il calcolo della distanza intertestuale, i subcorpora descritti al § 1.3.2. sono stati preparati con *Taltac2* secondo diversi trattamenti suscettibili di avere un impatto diverso sui risultati, come esposto qui di seguito.

a) Normalizzazione leggera

Con questo trattamento il software si limita a trasformare in accenti gli apostrofi posizionati erroneamente (per cui *liber-ta' → libertà*) e a trasformare in minuscole le maiuscole dovute esclusivamente al contesto sintattico, per cui *Non* e *non* verranno considerati un'unica forma grafica se la maiuscola è dovuta alla presenza di un segno di interpunzione forte, men-

tre *Franco* e *franco* non saranno considerati equivalenti se la maiuscola è dovuta al fatto che si tratta di un nome proprio e non al contesto sintattico.

b) Polirematiche

Taltac2 è in grado di riconoscere unità lessicali superiori (es. *forze dell'ordine*), locuzioni varie (*fra l'altro, riguardo a*), nomi propri di vario tipo, che verranno trattati come forme a sé stanti all'interno del vocabolario.

Con i trattamenti *a* e, soprattutto, *b* aumenta l'incidenza dell'argomento dei testi perché aumenta la precisione con cui le forme grafiche individuate riflettono diverse accezioni semantiche: se la forma *forze* può comparire in un testo di fisica o sociologia, l'inserimento nel sintagma *forze dell'ordine* ne rispecchia più chiaramente il significato. A di là del campionamento, che dovrebbe averlo almeno in parte disinnescato, l'effetto dei contenuti non va trascurato.

Per cercare di annullare il più possibile l'impatto degli argomenti dei testi abbiamo pensato di eseguire le procedure per la misurazione della distanza intertestuale prendendo in considerazione esclusivamente le parole vuote estratte dai nostri subcorpora. A questo scopo abbiamo selezionato i trattamenti che seguono.

c) Locuzioni

In questo caso si procede a estrarre le polirematiche classificate come "locuzioni grammaticali" nel database delle risorse statistico-linguistiche disponibili in *Taltac2*. Si tratta di costrutti di vario tipo, per es. congiunzioni come *dato che*, ma anche sintagmi la cui classificazione da parte del software lascia qualche perplessità (per es. *a in rovina* viene assegnata funzione aggettivale, mentre *un insieme di* viene annoverato tra i sintagmi preposizionali), come pure appare talvolta difficile includere tra le parole grammaticali avverbiali come *senza dubbio* o *in verità*. Tuttavia, l'elenco fornito da *Taltac2* comprende grosso modo collocazioni frequenti della lingua italiana suscettibili di avere funzione grammaticale e abbiamo deciso di includere questo trattamento tra le nostre procedure.

d) Grammaticali

Questo trattamento si basa su un elenco di parole grammaticali ottenuto tramite il *tagging* realizzato con *Taltac2* sul corpus di 160 romanzi italiani. L'elenco comprende articoli, preposizioni, congiunzioni e pronomi ma non gli avverbi. Infatti, aggettivi e avverbi sono trasversali alle categorie di parole piene e vuote e presentano casi alquanto dubbi: in questo caso abbiamo preferito attenerci a un approccio massimalista secondo il quale consideriamo grammaticali solo le classi chiuse del lessico italiano.

e) Locuzioni + grammaticali

Con quest'ultimo trattamento estraiamo dai nostri subcorpora i dati linguistici combinati di entrambi i trattamenti *c* e *d*.

2.6.1. *Macrotesti tradotti e macrotesti non tradotti*

La prima domanda a cui cerchiamo di rispondere è la seguente: se il traduttese presenta caratteristiche tendenziali che lo distinguono dall'italiano prodotto direttamente da parlanti non tradotti, se ne può rilevare l'impatto sulla distanza tra i testi? In pratica, avendo a disposizione due subcorpora (traduzioni e testi non tradotti), se calcoliamo la distanza tra i vari testi che li compongono, quando mettiamo a confronto due traduzioni o due testi non tradotti, rileveremo sempre valori inferiori rispetto a quelli ottenuti dal confronto tra una traduzione e un testo non tradotto?

La tabella 2.34 offre i primi dati utili per rispondere al nostro quesito. Ricordiamo che, poiché la distanza intertestuale è sensibile alle dimensioni e ai contenuti dei testi, non è stato possibile utilizzare direttamente i singoli articoli, ma abbiamo confrontato tra loro 30 macrotesti ottenuti aggregando diverse traduzioni (con lingue di partenza e traduttori diversi) e 30 macrotesti ottenuti aggregando diversi articoli non tradotti (di autori diversi) secondo la procedura di campionamento descritta al § 1.3.2. (200 campionamenti di *chunks* di 3.500 occorrenze ciascuno), calcolando poi la distanza sia sull'intero vocabolario sia su sottoinsiemi di parole grammaticali.

Trattamento	Distanza campionata media tra testi del subcorpus tradotto	Distanza campionata media tra testi del subcorpus non tradotto
a) Normalizzazione leggera	0,529	0,523
b) Polirematiche	0,559	0,553
c) Locuzioni	0,558	0,552
d) Grammaticali	0,138	0,138
e) Grammaticali + Locuzioni	0,138	0,132

Tabella 2.34 – Media e deviazione standard della distanza intertestuale campionata

Come possiamo vedere, la media delle distanze intertestuali calcolate tra i macrotesti non tradotti e tradotti risulta sempre (e significativamente) maggiore rispetto alla media calcolata internamente ai due subcorpora, il che indica una maggiore vicinanza reciproca tra le traduzioni (e tra i testi non tradotti) rispetto a quando il confronto avviene tra i due subcorpora (cfr. anche la figura 2.5 sotto). Nel dettaglio, la media delle distanze tra i macrotesti tradotti non si allontana molto dalla media riferita ai macrotesti non tradotti, anche se quest'ultima è sistematicamente minore (sorprendentemente, la differenza è costante in *a*, *b* e *c*, pari a 0,006). Al contrario, le deviazioni standard sono sistematicamente più basse nel subcorpus delle traduzioni (con l'eccezione del trattamento *e*). Da queste osservazioni possiamo concludere che i macrotesti non tradotti presentano una somiglianza reciproca più marcata rispetto ai macrotesti tradotti. Ciò potrebbe essere dovuto all'influenza delle lingue fonte: mentre per i macrotesti non tradotti questo fattore è assente, i campioni estratti dai macrotesti tradotti e sottoposti a confronto potrebbero comprendere di volta in volta lingue di partenza diverse, che determinano una distanza reciproca leggermente maggiore.

Una spiegazione alternativa (o aggiuntiva, poiché i due fattori non si escludono a vicenda) potrebbe fare riferimento ai conte-

Deviazione standard media subcorpus tradotto	Deviazione standard media subcorpus non tradotto	Distanza campionata media tra testi non tradotti e tradotti	Deviazione standard media per distanze tra testi non tradotti e tradotti
0,007	0,009	0,536	0,009
0,007	0,009	0,567	0,008
0,007	0,009	0,566	0,008
0,008	0,010	0,144	0,010
0,024	0,018	0,154	0,012

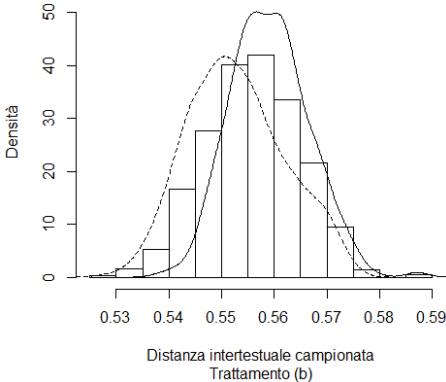
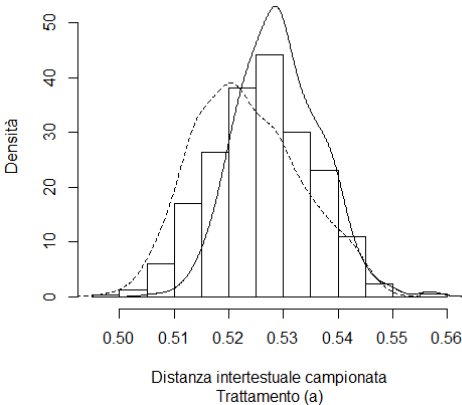
nuti: oltre all'impatto della lingua di partenza, le traduzioni potrebbero essere caratterizzate da contenuti più variabili e condurre a un piccolo incremento della distanza intertestuale. Questa ipotesi sembra essere corroborata dai risultati ottenuti con i due trattamenti che dovrebbero riuscire meglio a limitare l'effetto dei contenuti: *d* (grammaticali) ed *e* (grammaticali + locuzioni). Come si può vedere, la normalizzazione leggera (*a*) e ancor più le polirematiche (*b*) colgono maggiormente la composizione lessicale dei macrotesti che, probabilmente proprio in virtù dell'incidenza dei contenuti, risultano reciprocamente più distanti rispetto a quando calcoliamo la distanza intertestuale considerando solo le parole grammaticali. Anzi, in questo caso traduzioni e testi non tradotti risultano equidistanti all'interno dei propri subcorpora. Le locuzioni (*c*) risultano invece essere un elemento di disturbo: da una parte il relativo trattamento produce distanze intertestuali inferiori solo alle polirematiche (quindi sembrerebbero risentire dell'effetto dei contenuti), dall'altra, unitamente ai grammaticali, ottengono una distanza media pari ai soli grammaticali tra i macrotesti tradotti e addirittura inferiore tra i macrotesti non tradotti. Come già evidenziato, la natura composita di questa componente delle risorse statistico-linguistiche disponibili nel software *Taltac2* non ci permette di offrire spiegazioni valide per un simile comportamento ondivago.

Anche per quanto concerne le deviazioni standard delle distanze intertestuali, notiamo che la differenza tra i valori riferiti alle traduzioni e ai macrotesti non tradotti è molto ridotta, sebbene risulti costantemente maggiore nel secondo dei due subcorpora e, seppure in misura minore, nel confronto tra macrotesti tradotti e non tradotti (con l'eccezione del trattamento e, che addirittura produce il valore più basso nel confronto tra subcorpora). Ciò significa che c'è minore omogeneità tra i valori delle distanze intertestuali tra gli articoli scritti direttamente in italiano, un risultato che parrebbe confermare l'universale traduttivo della convergenza. In altre parole, nei nostri subcorpora i macrotesti tradotti tendono a essere mediamente meno simili tra loro rispetto ai macrotesti non tradotti (tra loro), ma ci sono meno traduzioni che sono molto differenti dalle altre; di converso i macrotesti non tradotti tendono a essere più omogenei, ma presentano alcuni casi che si discostano marcatamente dalla media. Insomma, la più alta deviazione standard tra i macrotesti non tradotti sta a indicare che le rispettive distanze intertestuali sono più "perturbate" rispetto alle distanze intertestuali tra macrotesti tradotti.

Queste conclusioni valgono per tutti i trattamenti a cui è stato sottoposto il corpus, con un leggero incremento della deviazione standard della distanza intertestuale calcolata considerando solo le parole grammaticali, ma anche con la notevole eccezione del trattamento e. In questo caso i valori all'incirca triplicano (per le traduzioni) o raddoppiano (per i macrotesti non tradotti), così ribaltando la situazione descritta sopra: sono le distanze intertestuali delle traduzioni ad avere una distribuzione più perturbata e non è chiaro il motivo per cui questo trattamento conduca a risultati così eccentrici rispetto agli altri.

Finora abbiamo considerato la distanza intertestuale media e la sua variazione standard all'interno di ciascun subcorpus (tradotto e non tradotto), ma ciò che ci preme è, ovviamente, sapere se la distanza intertestuale tra le traduzioni risulti (sempre) minore rispetto alla distanza intertestuale tra una traduzione e un testo non tradotto, e viceversa. Grazie alla figura 2.5 qui sotto possiamo visualizzare la distribuzione della distanza intertestuale per i cinque trattamenti e notare come le distanze tra i macrotesti tradotti (linea continua) abbiano distribuzioni diverse dalle distanze tra macrotesti non tradotti (linea tratteggiata).

Rispetto alla distribuzione generale (i rettangoli al centro), la linea continua tende a formare curve più alte e caratterizzate da andamenti più “ripidi” e basi più “strette” della linea tratteggiata, che però si posiziona quasi sempre “a sinistra” delle linee continue (fa eccezione il trattamento *d*): ciò indica che i macrotesti tradotti sono tutti più distanti tra loro, ma lo sono in maniera costante, mentre i macrotesti non tradotti risultano reciprocamente più vicini ma c’è maggiore variabilità interna. Infine, i grafici dimostrano che le medie riportate in tabella 2.34 non sono solamente frutto del caso o di picchi che inficiano le distribuzioni.



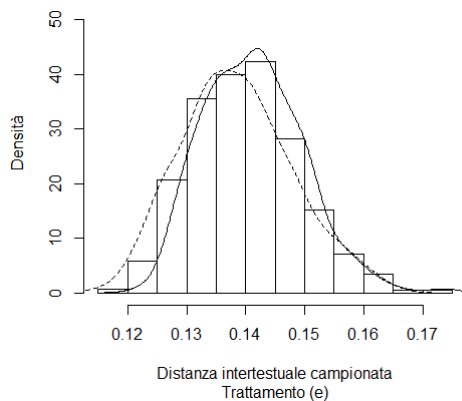
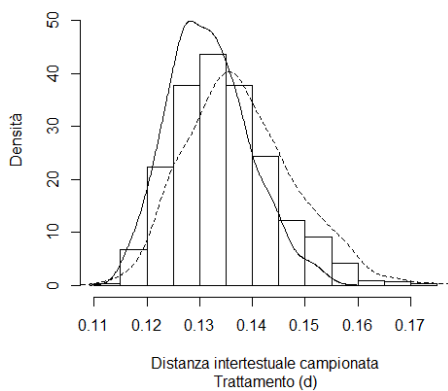
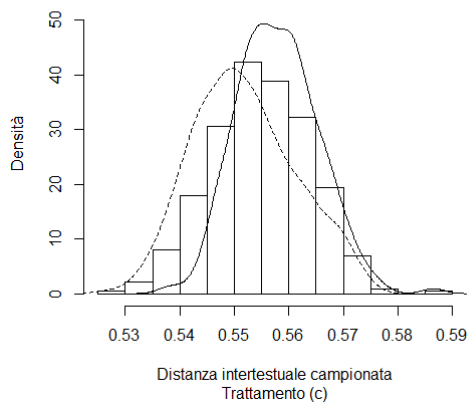
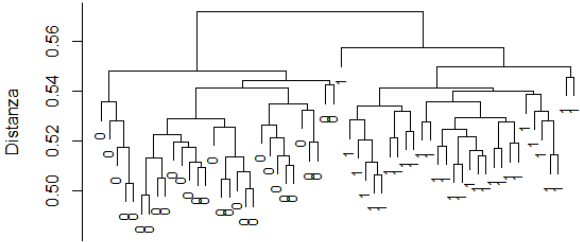


Figura 2.5. Distribuzione delle distanze intertestuali tra macrotesti non tradotti e tradotti

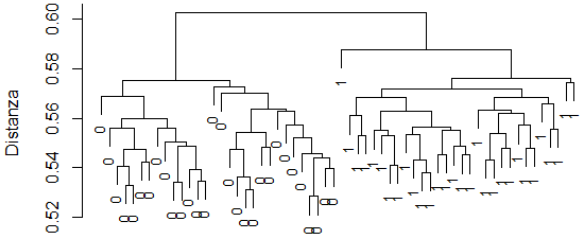
Per illustrare più efficacemente i risultati dei nostri calcoli uno strumento molto utile è il dendrogramma. Il dendrogramma consente di visualizzare in maniera intuitiva gruppi diversi di elementi (nel nostro caso: i macrotesti) e come questi siano collegati tra loro in base a una misura di distanza reciproca (nel nostro caso la distanza intertestuale campionata). Ogni “foglia” dell’albero rovesciato che costituisce il dendrogramma rappresenta uno di questi elementi. Gli elementi stessi sono collegati da linee e finiscono per formare dei raggruppamenti, mentre l’asse delle ordinate indica la distanza a cui due macrotesti vengono collegati.

Trattamento (a) - Normalizzazione leggera



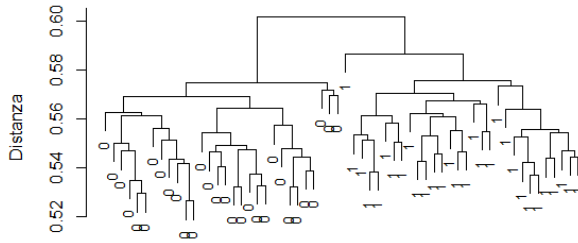
Distanza intertestuale campionata - Soglia 0

Trattamento (b) - Polirematiche



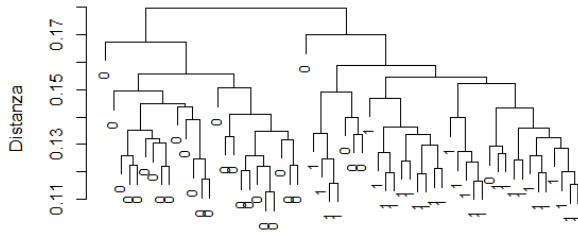
Distanza intertestuale campionata - Soglia 0

Trattamento (c) - Locuzioni



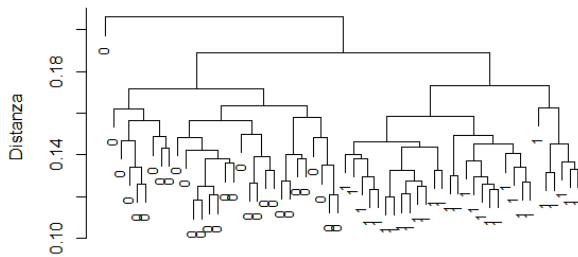
Distanza intertestuale campionata - Soglia 0

Trattamento (d) - Grammaticali



Distanza intertestuale campionata - Soglia 0

Trattamento (e) - Grammaticali+Locuzioni



Distanza intertestuale campionata - Soglia 0

Figura 2.6 - Dendrogrammi per i 60 macrotesti (0 = non tradotti, 1 = tradotti)

La figura 2.6 presenta i dendrogrammi relativi ai diversi trattamenti a cui sono stati sottoposti i nostri subcorpora. I grafici sono stati generati secondo un processo agglomerativo, tramite il metodo del legame completo. Ricordiamo che i processi di *clustering* agglomerativi mirano a raggruppare un insieme di unità statistiche di base (nel nostro caso i macrotesti) in un insieme di gruppi meno numerosi delle unità statistiche di base stesse. Nella fattispecie della tecnica usata in questo studio, il *clustering* viene definito gerarchico, in quanto l'insieme di gruppi individuati è caratterizzato da relazioni di appartenenza univoche a gruppi più ampi. Nel dettaglio, la tecnica usata segue questo processo: individua, nell'insieme di macrotesti, i due macrotesti che hanno la minore distanza reciproca (cioè individua i macrotesti che sono più simili tra loro) e li associa, formando un gruppo (ed è per questo che la tecnica è definita "agglomerativa"). Il nuovo gruppo appena creato viene ora trattato come se fosse una delle unità statistiche di base, perciò bisogna procedere a ricalcolare la distanza tra questo nuovo macrotesto frutto dell'unione di due unità di base e tutte le altre unità di base. Qui entra in gioco il metodo del legame completo: per calcolare la distanza tra il neo-gruppo e tutte le altre unità statistiche di base viene presa la più alta tra le distanze delle unità statistiche che fanno parte del neo-gruppo e le unità statistiche di base. Il processo viene dunque ripetuto fino a quando tutti i gruppi e le unità statistiche di base confluiscono in un gruppo unico.

Come possiamo vedere, con i primi tre trattamenti (*a*, *b* e *c*) i macrotesti tradotti (contrassegnati con il numero 1) sono sempre più vicini agli altri macrotesti tradotti, mentre i macrotesti non tradotti (contrassegnati con 0) sono più vicini agli altri macrotesti non tradotti. A titolo di esempio, per leggere il dendrogramma relativo alla normalizzazione leggera, occorre fare riferimento alla scala riportata sull'asse delle ordinate. I 30 macrotesti tradotti presentano una distanza intertestuale che va da poco meno di 0,50 a poco meno di 0,55, valori deducibili dalla lettura dell'asse delle ordinate in corrispondenza delle singole aggregazioni (linee orizzontali di collegamento tra macrotesti). Per meglio comprendere questi numeri, è utile tornare a consultare la tabella 2.34, da cui rileviamo che la distanza intertestuale campionata media del subcorpus tradotto è pari a 0,529, cioè circa a metà strada tra 0,50 e 0,55.

Il fatto che i macrotesti trattati con le procedure *a*, *b* e *c* vengano posizionati nell'albero secondo una netta divisione tra traduzioni e testi non tradotti è un segnale evidente della reale esistenza di un "effetto traduzione": se dalla tabella 2.34 si possono notare sola-

mente i dati aggregati, con la figura 2.6 invece vediamo come, in maniera sistematica, il macrotesto “più vicino” a un macrotesto non tradotto sia composto da testi non tradotti, e la stessa distribuzione vale per i macrotesti tradotti. Inoltre, tale vicinanza emerge anche per i vari raggruppamenti di macrotesti, fino ad una divisione in due parti del corpus tra traduzioni e testi non tradotti. Occorre infatti ricordare che il metodo agglomerativo, una volta associati due elementi, ricalcola le distanze tra tutti gli altri elementi e la coppia appena formata.

Se la divisione tra macrotesti non tradotti e traduzioni è netta con i primi tre trattamenti, la situazione si complica leggermente quando nel calcolo della distanza tra macrotesti entrano in gioco le parole grammaticali. In combinazione con le locuzioni (trattamento e), c'è un solo macrotesto che risulta totalmente eccentrico, accoppiandosi a grande distanza (superiore a 0,20) con tutti gli altri macrotesti (nativi e tradotti). È difficile dire che cosa renda questo macrotesto così diverso dagli altri; quel che è certo è che il fattore di disturbo risiede nelle parole grammaticali: nel dendrogramma relativo al trattamento d, cinque macrotesti non tradotti “invadono” il campo delle traduzioni a diverse distanze. Anche in questi casi è difficile ipotizzarne la causa: oltre alla procedura di campionamento seguita, è proprio la selezione delle parole grammaticali che dovrebbe garantire il minimo impatto dei contenuti sul calcolo della distanza. Delle due l'una: o in questi cinque campionamenti si è creata qualche combinazione particolare (per es. delle lingue di partenza, ma è molto difficile), oppure si deve concludere che le sole parole grammaticali sono meno precise delle altre risorse linguistiche nel cogliere le specificità del “traduttese” (sul perché torneremo in sede di conclusioni). Resta il fatto che, in ultima analisi, sia all'interno del gruppo degli articoli tradotti sia all'interno degli articoli non tradotti, la distanza tra i macrotesti è tendenzialmente minore rispetto alla distanza con quelli dell'altro gruppo e, dopotutto, nei dendrogrammi si creano due gruppi (*cluster*) di macrotesti ben distinti, pur con qualche *misclassification*.

2.6.2. L'effetto della lingua di partenza

Nelle considerazioni presentate qui di seguito la distanza intertestuale viene utilizzata per verificare se emergano differenze tra macrotesti tradotti da lingue diverse. Come già illustrato nel § 1.3.2., anche in questo caso i testi tradotti sono stati uniti e raggruppati in tre macrotesti per ogni lingua considerata, così da poter gene-

rare segmenti di dimensioni sufficienti a consentire il calcolo della distanza intertestuale secondo il campionamento di segmenti di 3.500 occorrenze ciascuno. Per poter limitare quanto più possibile l'eventuale influenza dello stile individuale del traduttore, i tre macrotesti per ogni lingua sono stati generati in modo da suddividere i testi attribuiti agli stessi traduttori in modo casuale. Come sopra, ci affidiamo a due diversi approcci per verificare se i macrotesti originati dalla stessa lingua di partenza risultino reciprocamente più vicini rispetto agli altri macrotesti tradotti da altre lingue: prima l'analisi aggregata delle distanze intertestuali, poi un'analisi più puntuale illustrata tramite dendrogrammi.

Già nella tabella 2.35 possiamo notare come le distanze tra macrotesti tradotti a partire dalla stessa lingua siano mediamente inferiori rispetto ai valori relativi ai macrotesti tradotti da lingue diverse. Stavolta la differenza è costante in tutti i trattamenti, con uno scarto minimo nel caso di e e massimo quando si prendono in considerazione le polirematiche (che dovrebbe essere più efficaci nel cogliere i contenuti).

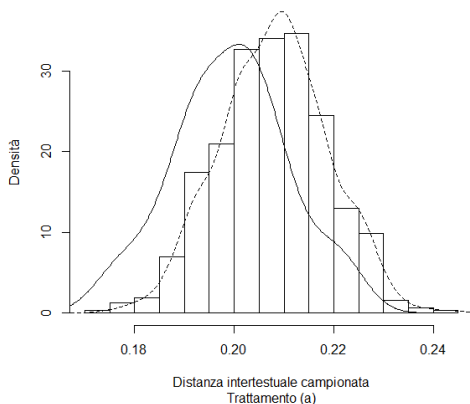
Trattamento	Distanza campionata media infra-lingua	Deviazione standard infra-lingua	Distanza campionata media extra-lingua	Deviazione standard extra-lingua
a) Normalizzazione leggera	0,199	0,012	0,208	0,011
b) Locuzioni	0,196	0,012	0,204	0,011
c) Polirematiche	0,196	0,012	0,204	0,011
d) Grammaticali	0,125	0,009	0,134	0,010
e) Grammaticali + Locuzioni	0,238	0,010	0,239	0,009

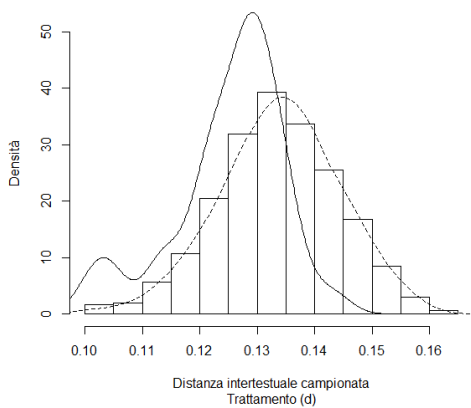
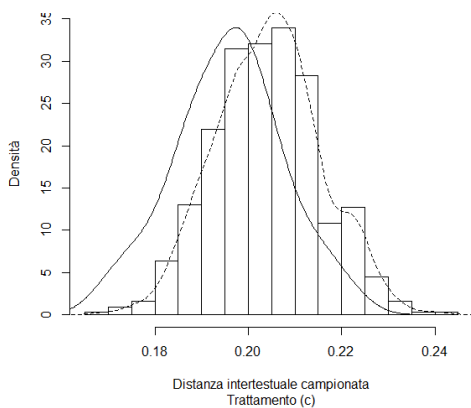
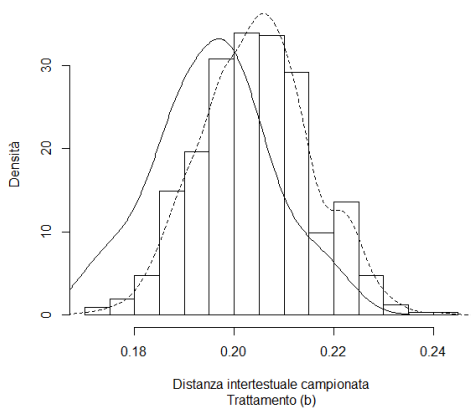
Tabella 2.35 – Media della distanza intertestuale campionata e deviazione standard per macrotesti tradotti dalla stessa lingua e macrotesti tradotti da lingue diverse

Come abbiamo già fatto nel caso del confronto tra testi non tradotti e testi tradotti (figura 2.5), grazie ai grafici possiamo visualizzare nel dettaglio la distribuzione delle distanze intertestuali per ogni trattamento del corpus (figura 2.7). Stavolta le linee conti-

nue rappresentano le distanze tra traduzioni dalla stessa lingua, mentre le linee tratteggiate rappresentano distanze tra macrotesti tradotti da lingue diverse. Analogamente a quanto avveniva in figura 2.5, possiamo vedere come le linee continue si posizionino quasi sempre “a sinistra” delle linee tratteggiate (anche stavolta fa eccezione il trattamento e): ciò significa che nel primo subcorpus la distribuzione delle distanze è sbilanciata verso “il basso”, e che le medie riportate in tabella 2.35 non sono solamente frutto del caso o il risultato di picchi che inficiano la regolarità delle distribuzioni, a conferma della possibilità che esista un “effetto lingua fonte” rilevabile con la misura della distanza intertestuale.

Può essere interessante notare come per i diversi trattamenti siano presenti alcune irregolarità nelle distribuzioni, che non sempre assumono la forma delle classiche “campane” normali. Soprattutto possiamo fare riferimento ai trattamenti *b*, *c* e *d*. Nei primi due è presente una “gobba” alla destra della distribuzione; tali discontinuità è dovuta a un gruppo di distanze particolarmente alto. In questi due casi le distanze maggiori originano da coppie di testi tradotti da lingue diverse, infatti possiamo notare come la gobba non sia presente nella linea continua. Nel caso *d* invece, notiamo una gobba nella parte bassa della distribuzione, e soprattutto che tale gobba è presente sulla linea continua, quella riferita alla distribuzione delle distanze tra testi provenienti dalla stessa lingua. Non sappiamo stabilire perché si verifichino tali perturbazioni, ma si tratta di un indizio del fatto che certi trattamenti del corpus risultano più efficaci di altri nell’evidenziare differenze o similarità tra i testi.





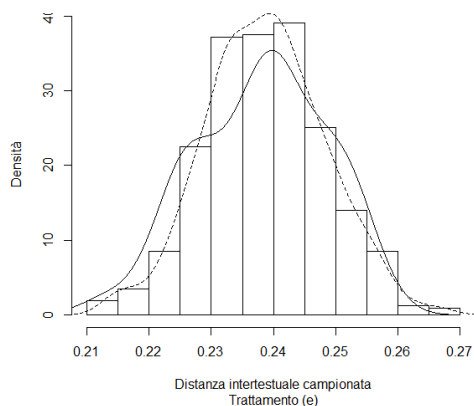
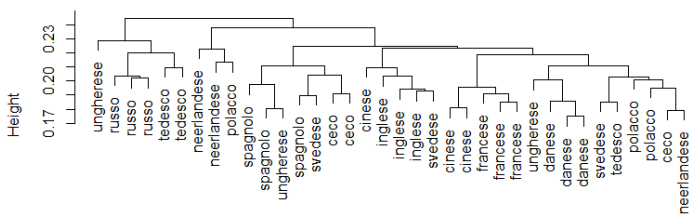


Figura 2.7 – Distribuzione delle distanze intertestuali tra macrotesti tradotti dalla stessa lingua e da lingue diverse

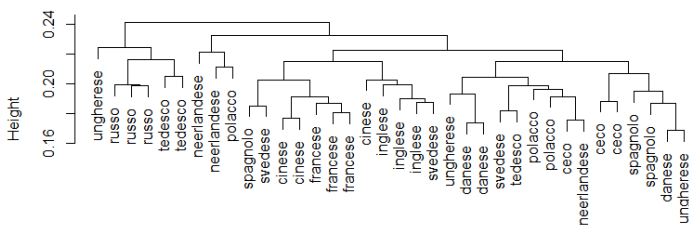
Passando alla rappresentazione dei dati sotto forma di dendrogrammi, la figura 2.8 riporta gli accoppiamenti dei macrotesti in base alle lingue fonte secondo i diversi trattamenti a cui abbiamo sottoposto il subcorpus. È possibile notare una tendenza dei macrotesti ad aggregarsi per lingua di partenza. Tali accorpamenti mostrano una certa coerenza con le lingue fonte e non sono frutto del caso: infatti a partire dalle 12 lingue straniere presenti in questo subcorpus con 3 macrotesti ciascuna, una volta selezionato un macrotesto, la probabilità che, se viene estratto un altro macrotesto in modo casuale, questo risulti tradotto dalla stessa lingua è pari a 0,08 (8%). Secondo tale ragionamento, la probabilità che nel dendrogramma venga identificata “al primo livello” almeno una terna (come succede nel caso della normalizzazione leggera per danese, francese e russo) è pari a 0,007 (lo 0,7%), ben al di sotto di una soglia ragionevole. La probabilità che al primo livello venga identificata almeno una coppia invece è pari a 0,12 (12%) ma, come possiamo vedere dai grafici, con l’eccezione del trattamento e, il numero di coppie correttamente identificate è ben superiore (fino a 6 nel caso del trattamento con i grammaticali).

Trattamento (a) - Normalizzazione leggera



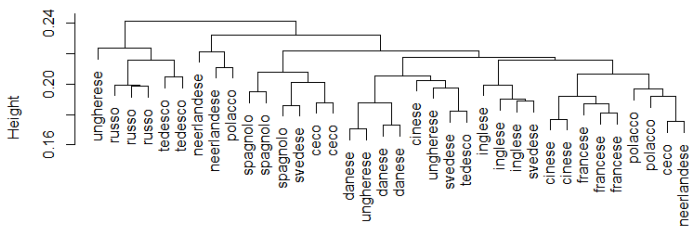
Distanza intertestuale campionata (n=3500,k=200)

Trattamento (b) - Polirematiche



Distanza intertestuale campionata (n=3500,k=200)

Trattamento (c) - Locuzioni



Distanza intertestuale campionata (n=3500,k=200)

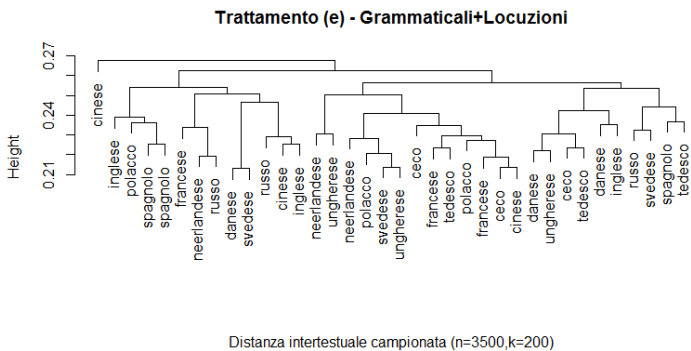
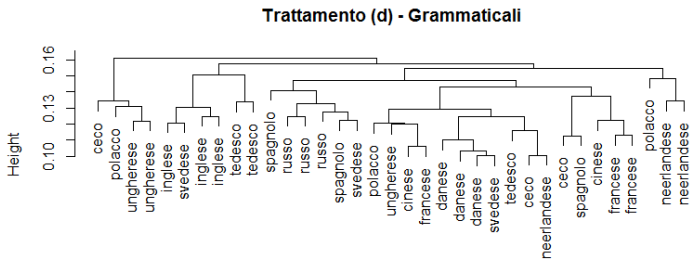


Figura 2.8 - Dendrogrammi dei raggruppamenti di traduzioni dalla stessa lingua

Dai grafici risulta chiaro che il calcolo della distanza intertestuale non riesce a individuare l'effetto della lingua di partenza se vengono considerate le parole grammaticali e le locuzioni individuate da *Taltac2* (trattamenti *d* ed *e*). Per il resto (ricordiamo ancora che il metodo agglomerativo, una volta associati due elementi, ricalcola le distanze tra tutti gli altri elementi e la coppia appena formata), il trattamento *a* individua correttamente in base alle lingue fonte tre terne di macrotesti al secondo livello (danese, francese e russo) e al primo livello altre tre coppie (ceco, cinese e tedesco); il trattamento *b* due terne (francese e russo) e quattro coppie (ceco, cinese, spagnolo e tedesco); il trattamento *c* due terne (francese e russo) e cinque cop-

pie (ceco, cinese, danese, spagnolo e tedesco); infine, il trattamento *d* non individua nessuna terna ma ben sei coppie (francese, inglese, neerlandese, russo, tedesco e ungherese).

Notiamo come, con i primi tre trattamenti, le lingue di partenza individuate si ripresentino con costanza: francese e russo sono sempre contenuti nelle terne, il danese compare in una terna e in una coppia; ceco, cinese e tedesco in tre coppie e spagnolo in due. Occorre inoltre notare che una possibile terna di macrotesti tradotti a partire dall'inglese non riesce a formarsi in tutti e tre i casi perché al primo livello c'è sempre un macrotesto tradotto dallo svedese che interviene: poiché si tratta costantemente dello stesso macrotesto, deve esserci qualcosa che lo avvicina particolarmente alle traduzioni dall'inglese.

Tra le coppie "malriuscite", non si evidenziano tendenze legate alle famiglie linguistiche. In effetti, data per buona l'ipotesi dell'interferenza linguistica, se la distanza intertestuale avesse funzionato perfettamente come metodo per cogliere l'influenza della lingua di partenza sulle traduzioni, non solo al primo e secondo livello si sarebbero create coppie e poi terne formate dai tre macrotesti per ogni lingua compresi nel nostro subcorpus, ma ai livelli superiori si sarebbero anche dovuti realizzare ulteriori accoppiamenti in base alle famiglie linguistiche (per es. il francese con lo spagnolo; il tedesco con il neerlandese; il ceco con il polacco e il russo). Anche in caso di funzionamento parziale, ci si sarebbe potuti aspettare che il calcolo della distanza intertestuale si sarebbe fatto "ingannare" dai macrotesti tradotti tra lingue della stessa famiglia, per es. accoppiando traduzioni dal tedesco con traduzioni dal neerlandese. Questo invece non si verifica: al primo livello, come a quelli superiori, emergono accoppiamenti tra lingue slave, neolatine e germaniche senza ordine apparente: per es. la terna tradotta dal russo viene collegata con tedesco e ungherese, una coppia di macrotesti tradotti dal cinese con la terna dal francese ecc. Particolarmente problematici risultano inoltre polacco e svedese che, in quanto lingue di partenza, non hanno mai creato accoppiamenti, mentre coppie di macrotesti tradotti dall'inglese e dall'ungherese emergono soltanto a seguito del trattamento *d* (del trattamento *e* si è già parlato in precedenza).

Nonostante tutto, in conclusione, è innegabile che, soprattutto i primi tre grafici riportati in figura 2.8, seppure non ideali nella loro distribuzione al fine della conferma dell'ipotesi dell'interferenza

della lingua di partenza, forniscono dati che non possono essere considerati frutto del caso e che confermano l'effetto dell'interferenza linguistica sulle traduzioni.

2.7. Conclusioni

Alla luce degli studi illustrati in questo capitolo, cerchiamo di rispondere, per quanto possibile, ai quesiti di ricerca iniziali: innanzitutto, se la lingua degli articoli di giornale tradotti da altre lingue presenti con una certa costanza caratteristiche diverse dagli articoli non tradotti; in seconda battuta, se il metodo della distanza intertestuale riesce a confermare l'esistenza del traduttese.

2.7.1. Le caratteristiche del traduttese nei giornali

Per quanto riguarda il primo dei corpora analizzati, occorre ammettere che un limite è dato dall'assoluta preponderanza dell'inglese come lingua fonte. Con 447 articoli sicuramente tradotti dall'inglese e altri 405 che probabilmente hanno la stessa origine (totale: 852 articoli) su un corpus di 955, è ovvio che non si può parlare di "italiano delle traduzioni" *tout court* tanto che, anche in presenza di forestierismi originari di altre lingue (per es. *élite*), è ipotizzabile il 'transito' attraverso l'inglese. In pratica, tutto ciò si traduce in un riferimento costante alla legge dell'interferenza e, più specificatamente, dell'interferenza dell'inglese sull'italiano. Persino i risultati delle misurazioni più generali e che meno dovrebbero risentire della coppia di lingue in gioco, come le misure lessicometriche, potrebbero essere interpretati in questo senso: per esempio, oltre che spia della tendenza alla semplificazione, la minore ricchezza lessicale delle traduzioni potrebbe essere attribuita al fatto che l'inglese, rispetto all'italiano, ricerca meno la *variatio*, non teme le ripetizioni delle stesse parole a breve distanza nel testo e in generale, a parità di condizioni in diafasia, opta per un registro meno elevato. Ne discende che anche una misura apparentemente 'tarata' sull'italiano come il *Vocabolario di Base* potrebbe riflettere la situazione lessicale dei testi fonte. In effetti, nel corso della nostra analisi, la legge dell'interferenza è stata invocata nel caso dei forestierismi, delle espressioni politicamente corrette, del perfetto semplice, della perifrasi *stare* + gerundio, di praticamente tutti i fatti a livello della morfologia pronominale, della lunghezza dei pe-

riodi, delle concordanze *ad sensum*, dei calchi sintattici, ma anche (*in absentia*) dei connettivi.

Il riferimento all'interferenza può naturalmente sovrapporsi al possibile influsso di uno degli universali traduttivi (è il caso dei pronomi soggetto espressi, che paiono dimostrare la tendenza all'esplicitazione), ma può anche giustificare il mancato riscontro di determinati fenomeni, come è avvenuto per i connettivi: la loro maggiore presenza nei testi originali può essere giustificata alla luce di convenzioni stilistiche diverse in italiano e in inglese tenendo conto dello specifico tipo testuale considerato nell'analisi. A ben vedere, conclusioni analoghe sono state raggiunte per la lunghezza dei periodi: l'universale traduttivo della semplificazione non trovava riscontro in una lunghezza media minore dei periodi nei testi tradotti, che tra l'altro era attesa anche in virtù dell'interferenza con l'inglese. Per giustificare tale assenza si è fatto ricorso alle abitudini stilistiche dei giornalisti italiani, che prediligono costrutti franti o addirittura nominali per ottenere uno stile maggiormente brillante e concitato.

Una prima conclusione può dunque essere questa: nello studio delle conseguenze del processo traduttivo sul testo di arrivo occorre prestare particolare attenzione alle tipologie testuali considerate; solo così, infatti, sarà possibile valutare correttamente i dati raccolti. Ciò sembra particolarmente importante per quanto riguarda l'interferenza linguistica, che verrebbe a comprendere, accanto alla nozione tradizionale di "interferenza a livello sistemico", anche aspetti stilistici legati alle tipologie testuali. Tornando al nostro corpus, per fare un esempio si potrebbe concludere che, se è opinione comune che in generale l'inglese predilige un periodare meno complesso rispetto all'italiano, nel caso specifico degli articoli di giornale questo non risulta vero.

Questa prima conclusione va a complicare ulteriormente il quadro delle influenze reciproche tra i diversi universali traduttivi e la legge dell'interferenza, quadro che era e rimane estremamente complesso. Accanto all'interferenza di cui si è parlato sopra, nella nostra analisi risultano attribuibili alla tendenza alla semplificazione la minore ricchezza lessicale e la frequenza lievemente maggiore del *Vocabolario di Base*, mentre alla tendenza all'esplicitazione possono essere invece ricondotti i fenomeni legati alla morfologia pronominale. La normalizzazione (qui evidentemente intesa come conservatorismo in senso lato) risulta la tendenza che annovera il

maggior numero di fenomeni ma che dà origine a conclusioni poco significative alla luce dello scarso peso numerico dei dati, come nel caso della preferenza accordata nelle traduzioni ad allotropi di registro più elevato, a certi modi verbali e, al contrario, dell'espulsione dei tratti innovativi dell'italiano dell'uso medio e della minor presenza di forestierismi nelle traduzioni. L'universale della convergenza è responsabile dei risultati delle misurazioni appositamente condotte sulla distribuzione dei valori di V/N e H/N per fasce percentuali dei singoli articoli, con il subcorpus tradotto caratterizzato da una minore variabilità interna. Non risultano invece conformi agli universali traduttivi i dati su densità lessicale, lunghezza dei periodi e rafforzamento dei segni interpuntivi e sui connettivi.

Un ulteriore 'fattore di disturbo' nel tentativo di rendere conto delle eventuali differenze tra testi tradotti e testi non tradotti è dato dalle "traduzioni invisibili". Risulta infatti arduo stabilire in che misura i testi del subcorpus non tradotto siano stati esposti a influssi esogeni tali da giustificare, per esempio, la maggiore frequenza di forestierismi o una distribuzione in fin dei conti equivalente tra i due subcorpora per quanto concerne i calchi sintattici (solo gli pseudo-calchi risultano leggermente più numerosi nei testi tradotti).

A questo proposito, la seconda conclusione del nostro studio è che, probabilmente per il ruolo che ricoprono e per il tipo di compito a cui sono chiamati, i traduttori dimostrano una forte consapevolezza per quanto attiene allo strumento linguistico e si danno 'regole' più o meno esplicite e di applicazione più o meno costante che possono essere basate sul rispetto della norma tradizionale (per es. limitando i fenomeni più innovativi dell'italiano dell'uso medio), puntare a una certa formalità di registro (per es. con l'uso del congiuntivo ogni qual volta la grammatica lo renda possibile), essere improntate a un atteggiamento vagamente puristico (per es. nell'evitamento dei forestierismi) o semplicemente offrire riferimenti certi là dove l'italiano non offre linee guida sicure (per es. tramite l'impiego della *d* eufonica di fronte alla stessa vocale). Questa conclusione deve far riflettere sull'ipotesi che considera le traduzioni vere e proprie un canale di accesso privilegiato per alcune innovazioni di origine straniera, mentre sotto certi aspetti le traduzioni presentano un italiano che sembra più normativo rispetto a quello delle traduzioni invisibili. Dai dati in nostro possesso emerge invece che, almeno nello scritto, i traduttori dimostrano un maggior attaccamento alla norma grammaticale e tengono un

atteggiamento più fortemente censorio nei confronti di eventuali influssi esogeni, in particolare quelli più evidenti: questa conclusione è corroborata dall'impiego di forestierismi e dalla frequenza del congiuntivo, mentre secondo una simile linea di pensiero le concordanze *ad sensum* non rappresentano un tratto a cui i traduttori si oppongono in quanto spia dell'influsso del testo fonte.

2.7.2. *Il traduttese alla prova della distanza intertestuale*

Il metodo di campionamento per il calcolo della distanza intertestuale qui proposto sembra in grado di cogliere la distinzione tra macrotesti tradotti e non tradotti. Non emerge alcun errore di classificazione per i trattamenti *a*, *b* e *c*, mentre il trattamento *e* posiziona un solo macrotesto fuori schema, accoppiandolo a tutto il resto del corpus, e il trattamento *d* colloca 5 macrotesti non tradotti nel ramo del dendrogramma relativo alle traduzioni (figura 2.6). In parte, tali discrepanze nei risultati ottenuti tramite gli ultimi due metodi di trattamento del corpus emergono anche dal confronto tra i valori medi e la deviazione standard della distanza intertestuale relativa al subcorpus tradotto e non tradotto; tuttavia appare evidente che il processo traduttivo comporta delle conseguenze sull'assetto dei testi che viene colto dal calcolo della distanza intertestuale. In qualche modo, sembra dunque che il traduttese effettivamente esista: per paragonare i nostri risultati agli studi sull'attribuzione d'autore, è come se fossimo riusciti ad attribuire una parte dei testi del nostro corpus a un "Traduttore astratto" e una parte a un "Autore nativo astratto", riconoscendone gli stili individuali.

Passando invece all'interferenza linguistica, i risultati che abbiamo ottenuto delineano una situazione più sfumata. Da una parte le traduzioni dalla stessa lingua appaiono reciprocamente più simili di quanto non lo siano nel confronto con traduzioni da lingue diverse (tabella 2.35), tuttavia i dendrogrammi che abbiamo ottenuto (figura 2.8) riescono solo in parte a collegare tutti e tre i macrotesti tradotti dalla stessa lingua e (di conseguenza) non sembrano in grado di riconoscere le diverse famiglie linguistiche comprese nel corpus. Ancora una volta, i trattamenti *a*, *b* e *c* ottengono risultati migliori e coerenti tra loro, tanto da poter confermare l'esistenza di un effetto della lingua fonte sul testo tradotto, mentre in particolare il trattamento *e* è risultato del tutto inaffidabile.

Resta da capire il perché delle somiglianze e delle differenze nei risultati dei trattamenti. In teoria i trattamenti *a* e *b* sono i più sensibili alle scelte lessicali in genere, e quindi anche ai contenuti, e infatti hanno in genere prodotto risultati simili. Non è chiaro perché il trattamento *c*, teoricamente meno sensibile ai contenuti, non si discosti molto dai primi due; per trovare una spiegazione, ci proponiamo un'analisi approfondita del materiale linguistico che *Taltac2* prende in considerazione nell'individuazione automatica delle "locuzioni grammaticali". In questo modo sarà possibile anche ipotizzare le ricadute sul funzionamento (in parte mancato) del trattamento e da noi adottato, che mirava ad arricchire il novero delle parole grammaticali in modo da disinnescare l'impatto dei contenuti sulla distanza intertestuale.

Rimane il fatto che anche il trattamento *d* non ha dato risultati particolarmente soddisfacenti. Si potrebbe pensare che le dimensioni dei nostri campionamenti non forniscano materiale sufficiente a permettere un confronto significativo a livello dei grammaticali, che naturalmente sono meno numerosi dell'insieme delle forme grafiche. Negli esperimenti che ci proponiamo di condurre in futuro sarà possibile verificare la variazione dei risultati delle misurazioni in base alle diverse dimensioni dei campioni considerati.

In alternativa, non è improbabile che i grammaticali risentano particolarmente delle idiosincrasie individuali del traduttore, quindi facendo aggio sulle tendenze generali del traduttore. In questo studio, l'influenza dello stile individuale del traduttore è effettivamente il convitato di pietra: purtroppo, per i motivi già esposti più volte, il nostro corpus mal si presta a tenere conto di questa variabile e sarà necessario assemblarne un altro ad hoc per procedere alle dovute rilevazioni. Le relative misurazioni, anche su testi non tradotti, potranno gettare luce su questo importante aspetto del rapporto che si instaura tra testo, lingua fonte, autore e traduttore.

3. Il traduttese nella narrativa

3.1. Obiettivi e strumenti della ricerca

In questo capitolo approfondiremo le caratteristiche delle raccolte di narrativa descritte nel capitolo 1.3.3., sempre con i metodi di indagine della linguistica dei corpora e dell'analisi statistica di dati testuali (§ 1.2.). La struttura e l'approccio metodologico seguono in gran parte quelli adottati per l'analisi dei giornali nel § 2, di cui questo lavoro condivide finalità e metodi, con alcune variazioni dovute alle specificità del corpus esaminato.

L'oggetto è l'assetto dell'italiano di opere tradotte e non tradotte di narrativa pubblicate tra il 1811 e il 2005. L'obiettivo è quello di contribuire a delineare il profilo dell'italiano della traduzione letteraria, di tracciare la sua evoluzione nel corso degli ultimi due secoli e di rilevare eventuali differenze tra letteratura alta e letteratura di consumo della seconda metà del Novecento. I dati e il confronto di diversi periodi storici permetteranno inoltre di ricostruire non solo gli andamenti del traduttese, ma anche quelli dell'italiano della narrativa in generale, che potrebbero dimostrarsi interessanti anche per cogliere gli eventuali influssi dell'italiano frutto di traduzione sull'italiano prodotto direttamente da madrelingua. In ultimo, i dati permetteranno di confrontare le caratteristiche dell'italiano tradotto e non tradotto della narrativa e della stampa.

Gli strumenti usati in questa sede sono gli stessi impiegati nella ricerca sull'italiano dei giornali: anche se ci sono alternative più

aggiornate, tale scelta ha permesso di ottenere dati comparabili per un ulteriore confronto con l'italiano della stampa. I programmi includono software per l'estrazione e l'analisi automatica di dati testuali, tra cui *Taltac2*, usato principalmente per la gestione, la pulizia e la normalizzazione del corpus; il software di lemmatizzazione e *PoS-tagging TreeTagger*, integrato in *TagAnt*, che permette di estrarre in automatico le parti del discorso. Infine, viste le dimensioni considerevoli del corpus, le ricerche sono state condotte anche con l'ausilio del software *Bran*, che, tra le varie funzioni come lemmatizzazione, *PoS-tagging* avanzato ed estrazione di informazioni sintattiche, offre anche la possibilità di analizzare in automatico i dati e di svolgere ricerche complesse, come quella di liste lessicali (per esempio, il Vocabolario di Base) o la disambiguazione della funzione dei pronomi soggetto, operazione molto dispendiosa da condurre manualmente su un corpus di grandi dimensioni.

Le caratteristiche del corpus letterario oggetto di studio presentano alcune differenze rispetto a quello giornalistico, soprattutto in termini di bilanciamento. Come evidenziato nel § 1.3.1, gli articoli di giornale sono più omogenei per genere, lunghezza e contenuti, grazie anche all'esclusione di alcuni tipi di articolo, come le interviste.

Al contrario, le opere di narrativa presentano una grande varietà di generi (avventura, thriller, per ragazzi, horror, rosa ecc.), e le opere incluse hanno caratteristiche anche molto diverse, come la quantità molto variabile di discorso diretto. Tale disomogeneità rende molto complicate operazioni come la campionatura, necessaria però per ottenere stralci di dimensioni comparabili per l'analisi delle convergenze, proposta invece per i quotidiani. Se si estraggono due campioni da due romanzi, è infatti probabile che questi abbiano caratteristiche molto diverse, come la presenza di dialoghi, che possono inficiare i dati. Anche la lunghezza non costante delle opere che compongono il corpus può porre alcuni problemi di rappresentatività, così come la presenza considerevole di alcuni autori: diventa ovvio che le scelte e gli stilemi di un autore (o di un traduttore) molto rappresentato possono pesare considerevolmente sui risultati, perché dipendono più dallo stile personale che da tendenze generali.

Un'altra differenza risiede nella rappresentatività dal punto di vista diacronico: gli articoli di giornale riguardano un periodo di tempo ristretto (2001-2008), mentre la distribuzione dei romanzi è molto

più dilatata e poco omogenea nel corso del tempo (come nel caso dei romanzi dell'800), con conseguenze sulla lingua, sullo stile e sul genere che compongono i sottoperiodi. La ricostruzione diacronica si baserà quindi sulla divisione in sottoperiodi già descritta nel capitolo 1 e non su una osservazione per anni o per decenni.

Le differenze tra narrativa e stampa e la dimensione considerevole del corpus letterario impongono anche altre variazioni sulle scelte dei fenomeni da trattare rispetto al lavoro sulla lingua dei quotidiani. Oltre alla convergenza, esclusa per la difficoltà di campionare i testi, non saranno ricercati i forestierismi: nello studio sui giornali sono infatti state ricercate le parole straniere raccolte del *Dizionario delle parole straniere nella lingua italiana*. Si tratta però di una lista che prende in considerazione parole straniere non adattate di un periodo limitato (fine del '900) e non può essere applicata a periodi precedenti. In questo senso, le opere di narrativa pongono anche difficoltà che si ricollegano alle ambientazioni straniere e quindi all'incidenza di *realia* o culturemi, cioè parole che richiamano fenomeni, oggetti e concetti tipici di una cultura, come cibo, strumenti musicali, bevande, luoghi, aspetti sociali, usanze ecc., che si risolvono spesso con dei prestiti integrali nei testi tradotti.

La ricerca delle espressioni politicamente corrette può essere interessante per misurare l'interferenza nei giornali in microdiacronia, soprattutto in base alle indicazioni sull'uso di un linguaggio non sessista. I giornali trattano spesso temi inerenti a diritti o alla collettività, cioè gli argomenti considerati nello studio sulla stampa, meno comuni però nella narrativa. Sono stati esclusi anche i calchi sintattici, che non saranno ripresi dal lavoro sugli articoli di giornale poiché quelli analizzati sono a loro volta estratti da un corpus di testi trecenteschi o del ventennio 1980-2000, si tratta quindi di tratti antichi o molto recenti e comunque difficili da individuare in modo automatico.

In ultimo, sarà tralasciata l'alternanza delle forme *cosa/che* cosa nelle frasi interrogative, perché sono molto comuni per la presenza di dialoghi e non facilmente estraibili con gli strumenti automatici a disposizione.

3.2. Lessico

L'analisi dell'aspetto lessicale sarà condotta da un punto di vista quantitativo, prima di tutto attraverso l'osservazione delle misure

lessicometriche, che sono state calcolate automaticamente. Verranno considerati indicatori come la ricchezza lessicale e, dopo l'identificazione delle parti del discorso, la densità lessicale e la frequenza d'uso del Vocabolario di Base.

3.2.1. Misure lessicometriche

Alcune differenze tra i subcorpora tradotti e non tradotti emergono in prima istanza dall'osservazione delle misure lessicometriche (tabella 3.1.; per comodità di esposizione i corpora saranno denominati con il primo anno del periodo coperto seguito da T per le traduzioni e NT per i testi non tradotti; i corpora di paraletteratura saranno invece indicati con le etichette *para T* e *para NT*). Nella maggior parte dei casi, i dati dei subcorpora letterari confermano, anche se in maniera meno accentuata, quanto già rilevato per la stampa periodica, cioè una minore ricchezza lessicale nelle traduzioni, che potrebbe suggerire l'azione della semplificazione. Il rapporto di parole diverse (cioè il V/N%, la *type-token ratio* vista nel § 1.2.1.) nei giornali vede infatti una differenza di circa l'1% in più negli articoli non tradotti, mentre nei subcorpora letterari la differenza oscilla tra lo 0,6 e l'1% circa. Anche il numero di *hapax*, parole usate una sola volta, è solitamente più elevato nei testi non tradotti, anche se in misura minore rispetto ai giornali. Tale risultato rientra negli effetti della normalizzazione: nelle traduzioni il numero di *hapax* è più ridotto perché chi traduce impiega parole di uso più comune che tendono a ripetersi.

	1800NT	1800T	1900NT	1900T	1946NT
N	861.044	860.022	1.205.980	1.210.605	914
V	53.377	59.4	59.964	52.613	55.05
V/N*100	6,199	6,907	4,972	4,346	6,023
Hapax*100	46,661	50,382	44,282	43,947	45,528

Il subcorpus non tradotto dell'800 costituisce un'eccezione notevole. Le traduzioni presentano infatti una ricchezza lessicale sensibilmente maggiore, ma anche valori anomali rispetto ai subcorpora di dimensioni più simili, cioè quelli 1946-1975, in cui lo scarto tra le percentuali di *hapax* è inferiore (1,4%). Si riscontra inoltre un rapporto tra V e N molto alto, soprattutto in relazione alla dimensione più ridotta del corpus, indice di un elevato grado di variazione lessicale. Un'ipotesi vagliata per spiegare la ricchezza lessicale delle traduzioni ottocentesche è che il subcorpus sia composto da un alto numero di opere (19), che potrebbero determinare una variazione maggiore di generi e temi e quindi un lessico più vario. Tuttavia, il numero di testi raccolti è vicino a quello delle traduzioni 1946-1975 e alla paraletteratura non tradotta (rispettivamente 17 e 20), che non presentano valori inaspettati. Le cause restano quindi ancora da verificare in studi futuri.

1946T	1976NT	1976T	ParaNT	ParaT
918.796	1.226.630	1.155.188	999.719	1.000.881
46.11	70.027	57.418	57.233	51.035
5,079	5,709	4,97	5,725	5,099
44,092	45,988	44,665	45,244	44,898

3.2.2. Il vocabolario di base

La frequenza di lemmi appartenenti al Vocabolario di Base, riportata nel grafico in Figura 3.1, permette di confrontare i livelli di difficoltà dei diversi subcorpora e di rilevare eventuali differenze tra testi tradotti e non tradotti e divergenze tra letteratura e stampa periodica.

Diversamente da quanto emerso dai giornali, in cui le traduzioni hanno mostrato più parole del VdB, in quelle del corpus letterario si rileva una percentuale di forme di base anche sensibilmente inferiore fino al 1946, poi leggermente superiore, anche se con scarti minimi (circa 0,5%), anche nella paraletteratura. Nella letteratura tradotta non sembra quindi esserci una tendenza alla medietà e l'universale traduttivo della semplificazione non trova conferma.

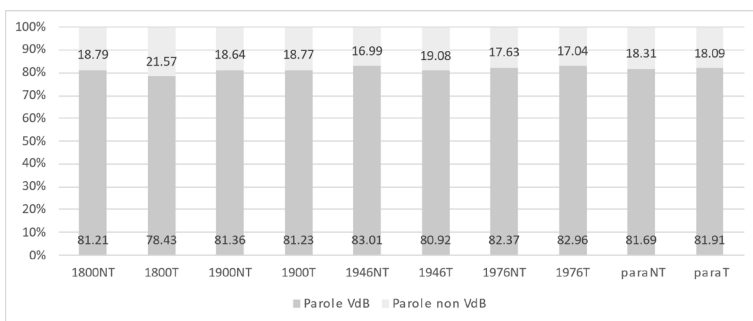


Figura 3.1 - Distribuzione percentuale delle forme appartenenti al VdB nei subcorpora tradotti e non tradotti

3.2.3. Categorie del lessico

L'osservazione delle singole categorie lessicali della grammatica tradizionale è un'altra caratteristica che permette di tracciare il profilo di un corpus. Dalle frequenze delle parti del discorso, estratte con il software *TreeTagger* e riportate nella tabella 3.2, risulta tuttavia difficile trarre conclusioni nette: le percentuali oscillano infatti a seconda del periodo e sono spesso piuttosto ravvicinate. Anche nei giornali non sono state osservate variazioni significative.

Alcune categorie mostrano comunque tendenze più chiare: più frequenti nei testi scritti originariamente in italiano sono gli aggettivi (tranne nell'800), gli avverbi (tranne nel 1946-75 e nella paraletteratura), gli articoli, le congiunzioni (tranne nel 1900-1945) e le preposizioni (tranne in paraletteratura). La presenza di aggettivi e avverbi indica una lingua più ricca di specificazioni e suggerisce la tendenza all'esplicitazione solo nelle traduzioni del primo periodo. Le congiunzioni sono invece spia di complessità sintattica, così come le preposizioni, legate alla preferenza per una sintassi composta da sintagmi complessi.

Sostantivi e verbi sono più frequenti nelle traduzioni, fuorché nel sottoperiodo 1900-1945, in cui verbi e pronomi appaiono meno spesso in queste ultime. La quantità di sostantivi e di verbi potrebbe ricollegarsi alla presenza di frasi brevi e composte da pochi elementi (il soggetto, il verbo, qualche complemento), che possono risultare dall'universale traduttivo della semplificazione. I sostantivi, così come i pronomi, possono anche essere frutto di interferenza delle lingue di partenza delle traduzioni, in cui il soggetto può essere obbligatorio (per esempio, in inglese), ma anche della tendenza dei traduttori all'esplicitazione. L'aumento dei sostantivi nel corso del tempo potrebbe confermare la preferenza per la ripetizione del sostantivo in luogo dell'uso del pronome o dell'ellissi, come dimostra anche la diminuzione dei pronomi registrata nelle traduzioni. In generale, quindi, la distribuzione delle parti del discorso è un'ulteriore prova degli universali traduttivi della semplificazione e dell'esplicitazione.

In ultimo, i dati mostrano un risultato inaspettato riguardo ai pronomi, che rappresentano una caratteristica della lingua delle traduzioni, proprio a causa di interferenze delle lingue a pronome obbligatorio e di eventuali esplicitazioni dei traduttori. La loro prevalenza non trova però conferma nel corpus: in tre sottoperiodi su cinque (tutto il '900, escludendo la paraletteratura), i pronomi sono infatti più frequenti, anche se con uno scarto minimo, nei testi non tradotti.

	1800NT	1800T	1900NT	1900T	1946NT
<i>Aggettivi</i>	8,18	9,19	7,97	7,92	8,32
<i>Sostantivi</i>	25,40	25,95	23,71	25,51	24,15
<i>Verbi</i>	19,38	19,87	20,43	20,35	21,10
<i>Avverbi</i>	7,16	6,93	8,37	8,13	7,89
<i>Congiunzioni</i>	5,35	5,21	5,49	5,62	5,64
<i>Articoli</i>	9,48	7,74	8,75	8,03	8,32
<i>Preposizioni</i>	14,83	13,26	14,29	13,66	13,91
<i>Pronomi</i>	10,21	11,85	11,00	10,77	10,68

Tabella 3.2 – Distribuzione percentuale delle parti del discorso

La distribuzione delle parti del discorso può essere ulteriormente elaborata per estrarre la densità lessicale. I risultati (rappresentati nel grafico in Figura 3.2) vanno in controtendenza rispetto all’universale traduttivo della semplificazione: le traduzioni sono infatti costantemente più ricche di parole piene, a eccezione del 1800. I risultati sulla stampa periodica non hanno invece dimostrato differenze: le parole piene si attestano intorno al 56% e quelle vuote intorno al 44% sia negli articoli nativi sia nelle traduzioni.

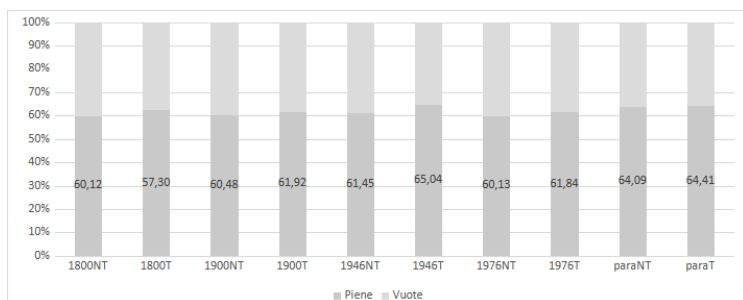


Figura 3.2 – Distribuzione parole piene/vuote

1946T	1976NT	1976T	ParaNT	ParaT
7,28	7,94	7,74	8,76	7,59
27,75	25,42	25,80	26,63	27,12
22,68	19,58	20,91	20,84	22,25
7,32	7,20	7,39	7,86	7,45
4,97	5,14	4,98	5,19	4,83
7,77	9,10	8,93	8,37	8,03
12,86	15,73	14,56	13,07	13,11
9,36	9,89	9,70	9,28	9,62

3.3. Morfologia

Per la morfologia lessicale, gli aspetti indagati includono le forme alternative di *sedere* e *dovere*, al diverso uso della *d* eufonica e del pronome *il quale*, *la quale* ecc. in luogo di *che*.

Per quanto riguarda *sedere* e *dovere*, i dati sono riportati nella tabella 3.3. Tra le forme alternative di registro più elevato, solo *debbo* e *debbono* presentano almeno 10 occorrenze, mentre le altre forme sono più sporadiche e l'analisi è poco significativa. *Debbo*, sempre minoritario rispetto a *devo*, è usato più spesso dai traduttori almeno fino al 1945: nell'800 compare quasi il doppio delle volte nelle traduzioni in confronto ai lavori non tradotti. In seguito la distribuzione si inverte e gli autori italiani lo usano molto più spesso (quasi sette volte in più nel 1976-2005 e in paraletteratura). Anche *debbono*, meno comune rispetto alla forma alternativa *devono*, è più utilizzato dai traduttori nel primo periodo e dagli italiani dal '46 in poi: le differenze sono generalmente meno marcate, al di là del caso del 1976, in cui *debbono* è quasi scomparso nelle traduzioni, ma ancora presente nelle opere di autori italiani. Entrambi i risultati non confermano del tutto quanto già osservato per la stampa, in cui sia *debbo* sia *debbono* sono più frequenti nelle traduzioni. Nella letteratura, i traduttori fanno ricorso alla variante più formale, quindi sono più conservatori, solo nell'800, mentre le traduzioni successive adottano forme meno formali e più innovative.

	1800NT	1800T	1900NT	1900T	1946NT
<i>debbo</i>	39	76	76	77	53
<i>devo</i>	75	67	193	254	116
<i>debbono</i>	10	15	12	9	16
<i>devono</i>	19	30	92	48	30
<i>segga</i>	5	4	5	2	0
<i>sieda</i>	0	0	1	0	2
<i>seggo</i>	0	6	2	0	7
<i>siedo</i>	1	0	1	2	8
<i>seggono</i>	0	0	2	0	0
<i>siedono</i>	1	3	2	4	6
<i>seggano</i>	1	1	0	0	0
<i>siedano</i>	0	0	0	1	0

Tabella 3.3 – Occorrenze di forme alternative di *dovere* e *sedere*

Anche l'impiego della *d* eufonica è indicativo di un eventuale conservatorismo dei traduttori. I dati relativi alla letteratura (tabella 3.4) si sovrappongono a quanto già osservato nella stampa, in cui *ed* è più usato negli articoli non tradotti anche prima di *e*, mentre *ad* è in assoluto più frequente nei testi tradotti, nei valori assoluti ma soprattutto nei casi di incontro con una *a*.

Per quanto riguarda il corpus letterario, le traduzioni contengono più occorrenze di *ed* e *ad* – che hanno andamenti simili – nel primo sottoperiodo, mentre negli altri compaiono più di frequente nei testi non tradotti tranne che in paraletteratura. *Ed* seguito da altra vocale è più usato nelle traduzioni dell'Ottocento e la proporzione s'inverte successivamente, mentre la regola di usare *ed* solo quando incontra la stessa vocale è applicata più spesso dai traduttori in tutti i periodi. Tendenze simili si registrano per *ad*: i traduttori lo usano sempre più spesso davanti ad *a*, con una crescita dopo il 1945; meno davanti ad altra vocale rispetto alle opere in italiano. *Od* è invece poco frequente in generale: anche questo è più utilizzato nelle traduzioni ottocentesche, poi

1946T	1976NT	1976T	ParaNT	ParaT
24	39	6	34	5
340	206	302	175	155
7	31	1	3	2
76	44	58	50	52
0	1	1	0	0
4	2	9	3	0
0	0	1	0	0
1	25	10	4	6
0	0	0	0	0
1	3	0	2	0
0	0	0	0	0
0	0	0	0	0

tende a oscillare e a ridursi man mano, soprattutto dopo il 1945, a parziale conferma della sostanziale scomparsa di *od* dall'italiano contemporaneo.

Dall'analisi si può rilevare che i traduttori sembrano introdurre la *d* eufonica solo se seguita dalla stessa vocale con più costanza, mentre gli autori italiani dimostrano un atteggiamento più tradizionale anche nelle opere più recenti.

	1800NT	1800T	1900NT	1900T	1946NT
e	26.154	24.124	35.170	38.952	28.247
ed	1.314	3.025	2.734	2.129	1.035
ed e	528	604	811	958	419
ed è	65	119	119	173	104
ed + altra vocale	689	2.241	1.767	1.018	432
o	1.621	1.963	2.349	2.349	2.451
od	27	77	63	13	7
o o	2	6	8	13	3
a	14.574	13.193	19.751	19.872	15.332
ad	1.266	2.001	2.203	2.073	1.328
ad a	544	615	1.045	1.099	552
ad + altra vocale	723	1.347	1.154	945	689
ad esempio	3	11	7	15	8
ad ogni	153	133	105	105	65
a ogni	48	28	66	65	27

Tabella 3.4 – Frequenza della *d* eufonica secondo i contesti d'uso

Un'altra indicazione sulla formalità dei testi è data dalla densità di pronomi relativi come *il quale*, *la quale* ecc. al posto del più comune *che*. Dalla densità per 1.000 forme nella Tabella 3.5 si possono prima di tutto osservare alcuni cambiamenti in diacronia al di là della distinzione testi non tradotti/tradotti: i pronomi relativi composti si dimezzano a partire dal sottoperiodo 1946-1975 in tutte le opere, sia italiane sia tradotte da altre lingue. In queste ultime, le varianti composte sono più comuni che nei testi non tradotti dell'800 e della prima metà del '900, così come succede negli articoli di giornale, che hanno registrato una densità per 1.000 forme di 0,85 in quelli non tradotti e di 1,05 in quelli tradotti. Invece, i risultati per la letteratura dei periodi post 1945 non avvalorano il conservatorismo dei traduttori.

	1800NT	1800T	1900NT	1900T	1946NT
Tot.	891	1473	1422	1611	621
d	1,034	1,712	1,179	1,330	0,679

Tabella 3.5 – Dati assoluti e densità per 1.000 del pronome relativo *il quale*, *la quale* ecc.

1946T	1976NT	1976T	ParaNT	ParaT
25.938	35.132	30.658	31.123	27.262
799	870	801	735	751
541	587	662	402	473
148	144	143	113	134
177	105	22	183	105
1.960	2.982	2.463	2.144	2.324
10	2	2	5	3
9	5	9	4	6
15.858	21.849	20.747	16.559	17.801
1.128	1.204	1.049	998	1.180
853	959	1.004	705	897
267	175	27	281	278
15	6	12	2	17
24	31	1	48	27
61	99	102	40	55

3.3.1. Distribuzione di tempi e modi verbali

Alcuni tempi e modi sono indicatori dell'aderenza alla grammatica tradizionale, quindi del conservatorismo e della normalizzazione. Uno dei tempi verbali a cui prestare attenzione è il futuro dell'indicativo, il cui uso è limitato solo ad alcuni ambiti nell'italiano neostandard. Nel corpus letterario, i cui dati delle densità per 1.000 forme sono riassunti nella Ttabella 3.6, le differenze sono minime, ma il futuro è più comune nelle traduzioni: ciò potrebbe confermare una maggiore fedeltà dei traduttori alla grammatica tradizionale. Tale conclusione non tiene tuttavia conto degli specifici usi del futuro (epistemico, dubitativo, concessivo ecc.), su cui dovrebbero basarsi le considerazioni sull'aderenza alla grammatica o all'italiano neostandard. Lo studio

1946T	1976NT	1976T	ParaNT	ParaT
430	584	453	462	461
0,468	0,476	0,392	0,462	0,460

sulla stampa periodica non riporta dati sul futuro dell'indicativo: un'ipotesi è che non sia stato considerato perché i giornali contengono quantità trascurabili di verbi al futuro, visto che si concentrano più che altro sul commentare eventi del passato.

Emerge inoltre una differenza nell'uso dell'imperfetto dell'indicativo, molto più utilizzato nei testi non tradotti che nelle traduzioni fino al 1946; ciò potrebbe essere imputato all'evitamento degli usi modali (descrittivo, attenuativo, fantastico, potenziale, epistemico ecc.), come sembrerebbe confermare la maggior frequenza (seppure di poco) di congiuntivo e condizionale nelle traduzioni.

Differenze più nette si riscontrano nel perfetto semplice dell'indicativo, che conferma inoltre i dati rilevati per quanto riguarda la stampa periodica: il suo impiego nelle traduzioni si può ricollegare anche all'influsso dell'inglese, la lingua più rappresentata del corpus, ma anche a un maggiore conservatorismo. Inoltre si può notare che le traduzioni lo usano più spesso lungo tutto il periodo preso in analisi, con uno scarto massimo tra tradotti e non tradotti nel periodo 1946-1975.

Anche le congiunzioni subordinanti che richiedono obbligatoriamente il congiuntivo, la cui frequenza assoluta è riportata nella tabella 3.7, comprovano la preferenza dei traduttori per questo

	1800NT	1800T	1900NT	1900T	1946NT
Indicativo pres.	3,19	3,87	3,56	3,68	4,64
Indicativo fut.	0,49	0,56	0,45	0,46	0,26
Indicativo imperf.	3,81	2,53	3,73	3,14	4,75
Indicativo perf. Semp.	2,87	3,09	2,67	3,16	1,73
Imperativo	0,58	0,99	0,48	0,41	0,89
Congiuntivo	0,77	1,02	0,91	1,03	0,72
Condizionale	0,32	0,40	0,46	0,49	0,36
Gerundio	0,94	0,78	0,87	0,90	0,74
Participio perf.	2,97	2,84	3,46	3,47	3,50
Participio pres.	0,06	0,07	0,07	0,05	0,05
Infinito	3,37	3,71	3,78	3,57	3,45

modo verbale e la loro tendenza a conservatorismo, con l'eccezione del periodo 1946-1975. La stessa preferenza è già stata rilevata nella stampa, in cui la densità delle congiunzioni si attesta a 0,35 nelle traduzioni e a 0,16 negli articoli non tradotti. È interessante anche sottolineare che le opere di consumo in italiano del subcorpus paraletterario hanno una frequenza inferiore di congiunzioni subordinanti rispetto al sottoperiodo del 1976-2005: le congiunzioni subordinanti che richiedono il congiuntivo segnalano infatti un registro più aulico e sono tipici della letteratura alta, mentre sono meno frequenti nella paraletteratura. In ultimo, così come osservato nei giornali, anche nel corpus letterario prevale la congiunzione avversativa *benché* su *sebbene*, *malgrado* e *nonostante* (*che*). Tuttavia, diversamente da quanto riscontrato nei giornali, in cui *sebbene* e *malgrado* compaiono di rado, nelle opere letterarie registrano un numero non trascurabile di occorrenze. Si osserva anche l'aumento in diacronia dell'uso di *nonostante*, che diventa la congiunzione avversativa più usata. Tale uso sembra trainato dalle traduzioni, in cui aumenta considerevolmente a partire dal 1900. In questa sede, tuttavia, non sono stati distinti i casi di *nonostante* seguiti da un verbo o da un sostantivo.

1946T	1976NT	1976T	ParaNT	ParaT
4,27	4,07	3,48	4,68	3,90
0,51	0,31	0,36	0,34	0,36
3,45	3,87	4,06	3,47	4,15
3,34	1,98	3,26	2,87	3,04
0,72	0,34	0,36	0,66	0,60
0,89	0,74	0,85	0,67	0,91
0,62	0,39	0,51	0,39	0,59
0,72	0,66	0,72	0,75	0,76
4,01	3,70	3,62	3,49	3,94
0,05	0,05	0,05	0,07	0,05
4,09	3,46	3,64	3,45	3,97

3.3.2. Perifrasi stare + gerundio

Dai dati sulla letteratura (tabella 3.8), emerge un uso costantemente più frequente nelle traduzioni della perifrasi composta da *stare* + gerundio, con la forbice che aumenta nel corso del tempo: nel sottogruppo paraletterario, la densità è doppia nelle traduzioni rispetto ai testi non tradotti, a conferma dell'incidenza di questo fenomeno nella paraletteratura. Anche in questo caso si può osservare un graduale aumento in diacronia, soprattutto nei testi non tradotti, che arrivano a utilizzare più spesso la perifrasi progressiva nel sottogruppo più recente. Tale aumento conferma che la diffusione di traduzioni dall'inglese a partire dal secondo dopoguerra ha rinforzato ed esteso, almeno nella narrativa, una struttura preesistente nella lingua ricevente. Inoltre, la maggiore frequenza nelle traduzioni corrisponde a quanto rilevato per la stampa (1,31 per 1.000 forme contro 0,92 nei testi non tradotti).

	1800NT	1800T	1900NT	1900T	1946NT
<i>affinché</i>	17	37	15	16	23
<i>benché</i>	59	161	110	213	144
<i>sebbene</i>	54	73	64	149	71
<i>malgrado</i>	43	74	84	139	17
<i>nonostante (che)</i>	56	42	18	97	63
<i>purché</i>	36	60	38	33	11
<i>a patto che</i>	2	1	6	4	33
<i>a condizione che</i>	2	1	5	5	1
<i>nel caso in cui/che</i>	1	6	16	16	4
tot.	270	455	356	672	367
d	0,314	0,529	0,295	0,555	0,402

Tabella 3.7 – Frequenza assoluta delle principali congiunzioni subordinanti con congiuntivo obbligatorio

	1800NT	1800T	1900NT	1900T	1946NT
d	0,081	0,172	0,154	0,377	0,454

Tabella 3.8 – Densità per 1.000 della perifrasi *stare* + gerundio

3.3.3. Frequenza di pronomi personali

Così come rilevato nello studio degli articoli di giornale, in cui l'esplicitazione dei pronomi soggetto è più frequente nelle traduzioni (densità di 1,71 per 1.000 dei non tradotti contro 2,42 dei tradotti), anche nel corpus letterario ci si potrebbe aspettare questa prevalenza, soprattutto per la composizione della raccolta, che nella parte delle traduzioni include per lo più lavori tradotti dall'inglese o da altre lingue con soggetto obbligatorio. Tuttavia, i dati estratti (tabella 3.9) non confermano sempre la preferenza per l'esplicitazione dei traduttori, che si registra solo a fasi alterne (nell'800 e nel 1976-2005). In tutti gli altri sottoperiodi, e nella paraletteratura, i pronomi sono presenti più di frequente nei testi italiani.

Andando nel dettaglio dei singoli pronomi, occorre innanzitutto tenere a mente che le differenze con la stampa sono ovvie: i pronomi di prima e seconda persona singolare e plurale si registrano molto

1946T	1976NT	1976T	ParaNT	ParaT
12	20	45	9	7
73	74	110	48	14
24	53	105	40	219
20	42	36	32	30
114	200	170	92	133
30	15	27	17	17
1	6	0	2	0
2	0	0	0	2
12	2	13	2	16
288	412	506	242	438
0,313	0,336	0,438	0,242	0,438

1946T	1976NT	1976T	ParaNT	ParaT
1,614	1,968	1,803	1,073	2,093

più spesso nella letteratura perché risentono della presenza massiccia di dialoghi. *Io*, che nella stampa ha una frequenza per 1.000 forme di 0,24 nei testi italiani e 0,58 nelle traduzioni, in letteratura è più sfruttato dai traduttori dell'800 e torna più frequente nel 1976-2005, mentre negli altri subcorpora prevale nei testi non tradotti, seppure con variazioni ridotte. Sono invece minime le differenze nell'uso di *tu*, per il quale non è possibile tracciare una tendenza precisa: la prevalenza è infatti altalenante e le differenze sono sempre poco spiccate. *Noi* e *voi*, entrambi più presenti nelle traduzioni dei giornali, hanno percorsi diversi: il primo mantiene il primato nelle traduzioni dell'800 e poi è più comune nei testi non tradotti; il secondo è invece caratteristico delle traduzioni, con l'eccezione delle opere letterarie più recenti. Il fattore più influente nel sovrautilizzo nelle traduzioni è senza dubbio quello dell'interferenza, soprattutto nel caso delle traduzioni dall'inglese, in cui *you* può indicare sia *tu* sia *voi*; o dal francese, che usa *vous* come allocutivo di rispetto.

I casi di pronomi di terza persona sono interessanti da osservare in diacronia, in quanto l'espansione di *lui*, *lei* e *loro* con funzione di pronome soggetto è una delle caratteristiche cardinali dell'italiano dell'uso medio: sono ormai la norma in ogni tipo di parlato, anche formale, e nelle scritture che rispecchiano atti comunicativi autentici. La frequenza di *egli* è più elevata nelle opere tradotte fino al 1946, mentre nei periodi successivi se ne registra la graduale scomparsa, che si vede anche per *ella*, più presente nelle traduzioni solo nella prima metà del '900. Sia *egli* sia *ella* non corrispondono ai dati della stampa, in cui il primo è più usato dai traduttori (0,07 nei non tradotti e 0,19 nelle traduzioni), mentre il secondo è assente. Tali frequenze non combaciano neppure con i dati del periodo più recente (1976-2005), cioè il più vicino a quello degli articoli di giornale. *Lui* ha invece un andamento piuttosto altalenante: mentre nella stampa prevale nei testi non tradotti (0,29 contro 0,24 forme per 1.000 delle traduzioni), nel caso delle opere letterarie non c'è una tendenza chiara. Per quanto riguarda *lei*, invece, più usato nella stampa tradotta (0,32 e 0,18 in quella non tradotta), si rileva più spesso nei lavori tradotti, tranne che nel sottogruppo paraletterario.

Esso, *essa*, *essi* ed *esse*, per cui lo studio sulla stampa non ha rilevato differenze significative, hanno andamenti molto simili, con densità maggiore nelle traduzioni solo nell'800 e poi una progressiva diminuzione nell'impiego. *Loro* compare con funzione di soggetto già nell'800, periodo in cui è più presente nelle traduzioni, mentre negli altri periodi prevale nei testi non tradotti.

I pronomi di terza persona mettono in luce anche cambiamenti che riguardano l'italiano in generale, e non solo le differenze tra testi tradotti e non tradotti, sotto l'aspetto diacronico: sia nei testi italiani sia in quelli tradotti si nota infatti la diminuzione di *egli* ed *ella* ed *essi* e l'aumento di *lui* e *lei* e *loro*. In questo senso, le traduzioni sono più innovative: i cambiamenti sono più netti e forme come *lui* e *lei* si affermano più velocemente a discapito dei pronomi soggetto previsti dalla tradizione grammaticale. Si può anche notare una progressiva riduzione dell'uso del pronome in generale, dovuta alla preferenza per omissione o ripetizione del sostantivo. Per esempio, osservando le differenze di uso tra *egli* e *lui*, si nota che la frequenza ottocentesca di *egli* non viene del tutto occupata da *lui*, a conferma del fatto che l'eredità di *egli* non è interamente raccolta da *lui*, ma va in parte a *lui* e in parte a \emptyset . Osservando *voi*, invece, si può notare un crollo nel corso del tempo anche nei testi non tradotti, probabilmente anche a causa della sua dismissione come forma di rispetto, soprattutto a partire dal 1945, dopo la caduta dell'imposizione fascista di usare il *voi* come allocutivo di cortesia al posto del *lei*. Può essere gradualmente venuto meno anche l'influsso diatopico del *voi*, tipico dell'italiano di alcune regioni. Un censimento degli autori del sottoperiodo italiano più recente, cioè quello in cui i testi non tradotti registrano il minimo di *voi*, fa emergere che la maggior parte degli autori sono del Nord o del Centro Italia. Fa eccezione Fausta Cialente, sarda.

In conclusione, sebbene l'analisi dei pronomi non permetta sempre di tracciare un quadro preciso, non si conferma l'esplicitazione e le traduzioni sembrano non risentire eccessivamente dell'influenza della lingua fonte, con l'eccezione dell'800. Non emerge neppure la tendenza al conservatorismo, se non nell'800: nei periodi successivi, come abbiamo visto, i traduttori recepiscono prima le tendenze dell'italiano dell'uso medio. Inoltre, hanno imparato a governare meglio l'esplicitazione dei pronomi in funzione soggetto col passare del tempo, soprattutto hanno iniziato a limitarne l'impiego nei casi in cui in italiano risulterebbero superflui. In ultimo, si può notare anche una differenza tra opere letterarie alte e di consumo: in queste ultime, infatti, i pronomi soggetto sono usati più spesso rispetto all'ultimo sottoperiodo preso in analisi. Un'altra differenza risiede nell'uso di *egli*: pur trattandosi di un tratto tipico di uno stile elevato, che ci si aspetterebbe nella raccolta non tradotta 1976-2005, è più comune nella paraletteratura.

3.3.4. Frequenza di pronomi e aggettivi possessivi

Il caso dei pronomi e degli aggettivi possessivi è simile a quello dei pronomi personali. Esistono infatti differenze tra italiano e inglese o francese, che usano un possessivo laddove l'italiano preferirebbe invece un articolo determinativo: ciò potrebbe quindi provocare un'interferenza nelle traduzioni. È anche nota la tendenza dell'italiano contemporaneo, almeno a partire dagli anni Quaranta, a

	1800NT	1800T	1900NT	1900T	1946NT
<i>io</i>	1.809	5.387	4.134	3.211	2.303
<i>d</i>	2,101	6,264	3,428	2,652	2,520
<i>tu</i>	817	952	1.052	856	571
<i>d</i>	0,949	1,107	0,872	0,707	0,625
<i>egli</i>	2.311	2.526	1.636	1.989	506
<i>d</i>	2,684	2,937	1,357	1,643	0,554
<i>ella</i>	1.543	958	725	1.071	211
<i>d</i>	1,792	1,114	0,601	0,885	0,231
<i>esso</i>	37	150	75	59	50
<i>d</i>	0,043	0,174	0,062	0,049	0,055
<i>essa</i>	85	698	498	92	139
<i>d</i>	0,099	0,812	0,413	0,076	0,152
<i>lui</i>	289	99	630	936	1.480
<i>d</i>	0,336	0,115	0,522	0,773	1,619
<i>lei</i>	269	398	659	769	810
<i>d</i>	0,312	0,463	0,546	0,635	0,886
<i>noi</i>	202	718	689	374	352
<i>d</i>	0,235	0,835	0,571	0,309	0,385
<i>voi</i>	326	881	459	1.201	184
<i>d</i>	0,379	1,024	0,381	0,992	0,201
<i>loro</i>	16	56	74	69	94
<i>d</i>	0,019	0,065	0,061	0,057	0,103
<i>essi</i>	141	414	182	161	124
<i>d</i>	0,164	0,481	0,151	0,133	0,136
<i>esse</i>	27	78	82	51	65
<i>d</i>	0,031	0,091	0,068	0,042	0,071
<i>tot.</i>	7.872	13.315	10.895	10.839	6.889
<i>d</i>	9,142	15,482	9,034	8,953	7,537

Tabella 3.9 – Frequenza assoluta e densità per 1.000 dei pronomi personali con funzione di soggetto

impiegare i possessivi in contesti dove non sarebbero previsti: tale uso è stato attribuito a influenze inglesi e francesi.

Dalla tabella 3.10, che riassume i dati assoluti e la densità per 1.000 forme di pronomi e aggettivi possessivi nelle loro forme maschili e femminili, singolari e plurali, si può desumere che le traduzioni fanno più ricorso ai pronomi in tre sottogruppi ('800, 1900-1945 e paraletteratura), mentre nella maggioranza degli altri casi sono più frequenti nei testi non tradotti.

1946T	1976NT	1976T	ParaNT	ParaT
1.955	1.296	1.532	2.019	1.607
2.128	1.057	1.326	2.020	1.606
789	633	646	848	865
0,859	0,516	0,559	0,848	0,864
8	177	65	316	128
0,009	0,144	0,056	0,316	0,128
2	9	3	119	81
0,002	0,007	0,003	0,119	0,081
11	49	41	32	26
0,012	0,040	0,035	0,032	0,026
20	87	47	51	33
0,022	0,071	0,041	0,051	0,033
1.447	1.381	1.526	1.246	1.214
1.575	1.126	1.321	1.246	1.213
1.036	1.301	1.852	1.510	1.345
1.128	1.061	1.603	1.510	1.344
288	348	262	350	318
0,313	0,284	0,227	0,350	0,318
472	133	101	205	234
0,514	0,108	0,087	0,205	0,234
84	132	89	110	107
0,091	0,108	0,077	0,110	0,107
18	126	48	150	39
0,020	0,103	0,042	0,150	0,039
9	37	21	34	20
0,010	0,030	0,018	0,034	0,020
6.139	5.709	6.233	6.990	6.017
6.682	4.654	5.396	6.992	6.012

Nel dettaglio, si possono osservare delle tendenze ricorrenti: l'800 mostra infatti che i possessivi sono più usati dai traduttori, così come nel subcorpus paraletterario, con alcune oscillazioni isolate (*mio* nelle traduzioni del 1976-2005, *suo* nel 1900-1945).

Un'eccezione notevole è *vostro*, che invece è costantemente più frequente nelle traduzioni. Ciò si può spiegare con quanto già detto per *voi* nel paragrafo precedente: l'influenza più rilevante è quella esogena della lingua di partenza, che può determinare un uso della seconda persona plurale come forma di cortesia, rinforzato anche da residui dell'uso fascista e dall'italiano regionale.

Anche in questo caso non emerge dunque l'universale traduttivo dell'interferenza (se non nell'Ottocento e nella letteratura di

	1800NT	1800T	1900NT	1900T	1946NT
<i>mio</i>	3.409	6.480	6.495	4.550	4.678
<i>d</i>	3,959	7,535	5,386	3,758	5,118
<i>tuo</i>	839	939	759	692	513
<i>d</i>	0,974	1,092	0,629	0,572	0,561
<i>suo</i>	7.086	7.982	7.606	9.502	6.124
<i>d</i>	8,230	9,281	6,307	7,849	6,700
<i>nostro</i>	493	1.327	1.038	741	1.044
<i>d</i>	0,573	1,543	0,861	0,612	1,142
<i>vostro</i>	534	1.144	636	1.713	261
<i>d</i>	0,620	1,330	0,527	1,415	0,286
<i>loro</i>	649	1258	1482	969	977
<i>d</i>	0,754	1,463	1,229	0,800	1,069
<i>Tot.</i>	13.010	19.130	18.016	18.167	13.597
<i>d</i>	15,110	22,244	14,939	15,007	14,876

Tabella 3.10 – Frequenza assoluta e densità per 1.000 di pronomi e aggettivi possessivi

consumo) come accade invece nella stampa periodica, che presenta una differenza netta tra traduzioni (11,40 forme per 1.000) e articoli italiani (7,48).

3.3.5. Frequenza di pronomi e aggettivi dimostrativi

Nei giornali, l'uso di pronomi e aggettivi dimostrativi è prevalente nei testi tradotti (9,56 dei non tradotti e 11,6 forme per 1.000 nelle traduzioni). Il totale dei dimostrativi nelle opere letterarie (riportato nella tabella 3.11, anche stavolta in base ai lemmi) conferma solo in parte i risultati dei giornali: i dimostrativi sono più utilizzati fino al 1945 e poi nelle opere di consumo.

1946T	1976NT	1976T	ParaNT	ParaT
2.703	3.006	4.039	2.640	2.845
2.942	2.451	3.496	2.641	2.842
547	1.058	830	770	801
0,595	0,863	0,718	0,770	0,800
5.287	9.527	8.007	6.195	6.820
5,754	7,767	6,931	6,197	6,814
699	1.245	701	765	1.060
0,761	1,015	0,607	0,765	1,059
1.006	226	198	266	518
1,095	0,184	0,171	0,266	0,518
677	1769	1114	779	1084
0,737	1,442	0,964	0,779	1,083
10.919	16.831	14.889	11.415	13.128
11,884	13,721	12,889	11,418	13,116

Negli articoli di giornale è emerso l'impiego di *questo*, preponderante nelle traduzioni, in cui compare 6,598 volte ogni 1.000 forme, contro la densità di 4,706 dei testi non tradotti. Lo stesso si osserva nel corpus letterario, tranne che nel periodo 1946-1975. Tuttavia, la presenza di *questo* va riducendosi nel corso del tempo.

Codesto e *cotesto*, scomparsi nell'italiano dell'uso medio con la semplificazione del sistema dei dimostrativi, non mostrano grandi differenze: la densità è più bassa nelle traduzioni, da cui scompaiono dopo il '46, mentre nei testi non tradotti e in tutto il sottogruppo della paraletteratura si registra ancora qualche residuo anche nei periodi successivi. *Cotesto* è presente nei primi due sottoperiodi e scompare successivamente. La concentrazione più alta si registra nelle traduzioni del 1900-1945 (202 occorrenze), in cui risulta più usato della forma *codesto*. Non è invece stato conteggiato nella stampa perché si tratta di una forma ormai desueta.

	1800NT	1800T	1900NT	1900T	1946NT
<i>questo</i>	2.547	4.639	3.768	6.222	3.897
<i>d</i>	2,958	5,394	3,124	5,140	4,264
<i>codesto/cotesto</i>	100	102	80	226	2
<i>d</i>	0,116	0,119	0,066	0,187	0,002
<i>quello</i>	2.996	2.625	4.900	4.381	3.589
<i>d</i>	3,479	3,052	4,063	3,619	3,927
<i>medesimo</i>	77	175	59	95	36
<i>d</i>	0,089	0,203	0,049	0,078	0,039
<i>ciò</i>	296	1034	333	446	283
<i>d</i>	0,344	1,202	0,276	0,368	0,310
<i>Tot.</i>	5.916	8.473	9.060	11.144	7.805
<i>d</i>	6,871	9,852	7,513	9,205	8,539

Tabella 3.11 – Frequenza assoluta e densità per 1.000 di pronomi e aggettivi dimostrativi

Quello si trova più spesso nei testi italiani, con l'eccezione del sottogruppo paraletterario: anche i giornali contengono una densità maggiore di *quello* nei testi non tradotti (4,359 contro 4,079 degli articoli non tradotti). *Medesimo*, che nei giornali ha una densità di 0,03 e 0,07 rispettivamente nei non tradotti e nelle traduzioni, è più frequente nelle traduzioni fino al 1945, con un picco nell'Ottocento.

In ultimo, *ciò* è sempre più usato nelle traduzioni letterarie, soprattutto nel primo sottoperiodo, un aspetto che suggerisce una maggiore aderenza alla grammatica: l'uso di *ciò* è infatti rarefatto nell'italiano dell'uso medio, in cui è solitamente sostituito da *questo* o *quello*.

1946T	1976NT	1976T	ParaNT	ParaT
3.113	3.190	3.814	2.866	3.328
3.388	2.601	3.302	2,867	3,325
0	2	0	5	2
0,000	0,002	0,000	0,005	0,002
3.299	5.218	3.822	4.072	4.437
3,591	4,254	3,309	4,073	4,433
36	34	32	29	20
0,039	0,028	0,028	0,029	0,020
334	488	630	266	618
0,364	0,398	0,545	0,266	0,617
6.782	8.930	8.298	7.233	8.403
7,381	7,280	7,183	7,235	8,396

3.4. Sintassi: lunghezza dei periodi e leggibilità

Correlare la lunghezza dei periodi con la complessità sintattica è un'operazione discutibile e forzata. Tuttavia, quantificare il numero di parole per frase è utile per avere una misura delle differenze tra italiano non tradotto e traduzioni a scopo di confronto.

La lunghezza media delle frasi, riportata nella tabella 3.12, fa rilevare che le traduzioni letterarie impiegano frasi più lunghe fino al 1945 e nel subcorpus tradotto di paraletteratura. La proporzione si inverte nei periodi successivi, in cui le traduzioni mostrano un accorciamento delle frasi, che restano invece più costanti nelle traduzioni. Nella paraletteratura, testi tradotti e non tradotti sono invece pari. Le lunghezze medie differiscono da quelle degli articoli di giornale, che nel capitolo 2 sono sostanzialmente uguali e si attestano a 22,71 parole. La lunghezza mediamente inferiore in gran parte dei periodi del corpus letterario è ovviamente dovuta all'influenza dei dialoghi, assenti nella stampa.

1800NT	1800T	1900NT	1900T	1946NT	1946T	1976NT	1976T	ParaNT	ParaT
17,744	23,060	16,775	19,247	17,654	13,559	16,795	13,635	13,600	13,434

Tabella 3.12 - Lunghezza media delle frasi

Alla complessità e lunghezza sintattica si riconduce anche l'uso della punteggiatura, un altro aspetto su cui i traduttori intervengono a scopo chiarificatore, semplificando o normalizzando il testo. Le operazioni riguardano soprattutto la riduzione dei segni di punteggiatura più deboli in favore di quelli più forti. Inoltre, è un aspetto particolarmente interessante nell'osservazione dei cambiamenti in diacronia, in quanto il periodo storico preso in analisi è stato particolarmente vivace dal punto di vista degli esperimenti sui segni di interpunzione, che hanno visto un'intensificazione del punto, usato non più come segno di interpunzione logico, ma come segno stilistico per aumentare ritmo ed enfasi, che ha messo in secondo piano altri segni come la virgola.

I dati sulla stampa fanno emergere una maggiore frequenza dei due punti e della virgola nel subcorpus non tradotto e del punto esclamativo nelle traduzioni. Tuttavia, il confronto con la letteratura è difficile: la presenza di discorso diretto incide inevitabilmente anche sull'uso della punteggiatura.

Le frequenze assolute e la densità per 1.000 forme del corpus letterario (tabella 3.13) mettono innanzitutto in luce che le traduzioni fanno ricorso a più punteggiatura solo nel 1946-1975. Da questo periodo in poi, i segni di interpunzione forti diventano invece considerevolmente più utilizzati nei lavori tradotti. Andando nel dettaglio, la stessa proporzione si può rilevare per il punto, che, in generale, aumenta in diacronia anche nella narrativa italiana. I traduttori vi fanno ricorso più spesso a partire dal 1946 e si può notare una particolare concentrazione di punti, indice di un'elevata frammentazione sintattica, nelle opere di consumo rispetto alla letteratura alta, soprattutto in quelle tradotte.

Per contro, la virgola diminuisce leggermente: si tratta di un segno di punteggiatura generalmente preferito dagli autori italiani, a eccezione delle traduzioni della prima metà del '900. Ciò conferma quanto già detto a proposito dell'indebolimento dei segni di interpunzione forti, considerata anche la tendenza inversa dei punti, ma può essere anche dovuto a una semplificazione del periodo nelle traduzioni, già osservata in precedenza.

Per quanto riguarda i due punti, i dati evidenziano che sono un tratto caratteristico delle opere italiane fino al 1975: nel 1946-1975 sono usate con frequenza più che doppia rispetto alle traduzioni. Più di recente, invece, i traduttori hanno iniziato a impiegarli più spesso. Il subcorpus di paraletteratura presenta un dato inaspettato perché si registrano più due punti rispetto alla letteratura alta: una possibile spiegazione si ricollega alla maggiore presenza di dialoghi nelle opere di letteratura di consumo, che evidentemente utilizzano i due punti come demarcatore di discorso diretto.

Il punto e virgola è un segno di punteggiatura sempre meno comune in italiano: il suo crollo in diacronia è confermato anche dalle misurazioni della tabella 3.13. Nel confronto tra opere non tradotte e tradotte, è quasi sempre più frequente nelle prime, tranne nelle traduzioni di consumo. È interessante osservare che i traduttori mostrano una tendenza più innovatrice che conservatrice, poiché la diminuzione è più rapida a partire dal 1946 in poi, mentre nelle opere italiane si registra con la stessa proporzione solo nel periodo successivo.

Sia il punto interrogativo sia quello esclamativo sono inizialmente più comuni nei testi italiani, poi in quelli tradotti. La tendenza si inverte nel 1946-1975 nel caso dei punti interrogativi e nel sottoperiodo successivo per gli esclamativi. Entrambi

trovano largo uso nella letteratura di consumo, un aspetto che potrebbe confermare la maggiore presenza di discorso diretto, già ipotizzata per i due punti.

I tre punti di sospensione sono molto comuni nella narrativa italiana dell'Ottocento, mentre in seguito i risultati sono altalenanti. Come ci si può attendere, la paraletteratura ne fa un uso maggiore rispetto alle opere di letteratura alta del 1976-2005: anche questo dettaglio potrebbe confermare l'incidenza di discorso diretto, perché i puntini di sospensione sono frequenti nella narrativa che simula il parlato.

	1800NT	1800T	1900NT	1900T	1946NT
.	37.173	36.313	59.295	56.601	46.639
<i>d</i>	43,172	42,223	49,167	46,754	51,027
,	75.361	70.941	97.641	101.322	77.609
<i>d</i>	87,523	82,487	80,964	83,695	84,911
:	5.777	5.121	8.053	4.003	8.616
<i>d</i>	6,709	5,954	6,678	3,307	9,427
;	8.192	7.161	7.588	6.955	5.225
<i>d</i>	9,514	8,327	6,292	5,745	5,717
?	8.616	4.644	10.759	9.252	4.656
<i>d</i>	10,006	5,400	8,921	7,642	5,094
!	10.565	6.352	10.738	8.469	3.723
<i>d</i>	12,270	7,386	8,904	6,996	4,073
...	6.934	1.840	5.378	4.711	2.382
<i>d</i>	8,053	2,139	4,459	3,891	2,606
<i>Tot. complessivo</i>	152.618	132.372	199.452	191.313	148.850
<i>d</i>	177,248	153,917	165,386	158,031	162,856
<i>Tot. punteggiatura forte</i>	63.288	49.149	86.170	79.033	57.400
<i>d</i>	73,501	57,149	71,452	65,284	62,801

1946T	1976NT	1976T	ParaNT	ParaT
67.251	67.477	80.601	67.917	75.239
73,195	55,010	69,773	67,936	75,173
63.256	86.263	75.096	80.660	67.002
68,847	70,325	65,008	80,683	66,943
3.973	4.152	4.290	6.442	4.141
4,324	3,385	3,714	6,444	4,137
1.535	2.393	1.558	1.090	1.930
1,671	1,951	1,349	1,090	1,928
9.005	6.250	8.719	9.396	10.807
9,801	5,095	7,548	9,399	10,797
3.235	1.836	2.230	2.981	3.574
3,521	1,497	1,930	2,982	3,571
3.753	3.968	3.092	5.113	5.170
4,085	3,235	2,677	5,114	5,165
152.008	172.339	175.586	17.3599	16.7863
165,443	140,498	151,998	173,648	167,715
83.244	79.531	94.642	85.407	94.790
90,601	64,837	81,928	85,431	94,707

3.4.1 Costrutti ad sensum

Diverse lingue di partenza dei testi tradotti presi in analisi prevedono l'accordo *ad sensum* dopo i quantificatori, anche se questi sono formalmente al singolare: la loro influenza potrebbe quindi indurre i traduttori a ricalcare la struttura della lingua straniera. D'altro canto, si tratta di un aspetto dell'italiano neostandard che può indurre chi traduce a stare più attento e a mantenere un atteggiamento più rispettoso della grammatica tradizionale.

Dalle scelte di autori italiani e traduttori (tabella 3.14) emerge prima di tutto che si tratta di strutture piuttosto rare, nella letteratura più che nella stampa. Come nei giornali, si conferma l'azione dell'interferenza più che del conservatorismo: il controllo sulle concordanze *ad sensum* è più debole per i traduttori, che le impiegano di più rispetto agli scrittori italiani, tranne nell'Ottocento, in cui lo scarto è comunque molto ridotto.

Si conferma inoltre il ruolo debole del quantificatore con le relative e le participiali, che tendono a stabilire l'accordo con l'antecedente immediatamente a sinistra. Risaltano in questo senso soprattutto le scelte di Salgari, che fa spesso ricorso alla concordanza *ad sensum* nelle casistiche appena riportate:

	1800NT	1800T	1900NT	1900T	1946NT
<i>maggioranza</i>	0	0	0	2	0
<i>parte</i>	0	0	1	13	4
<i>metà</i>	1	1	1	2	0
<i>percentuali</i>	0	0	0	1	0
<i>(mezzo) milione</i>	1	0	0	0	0
<i>(mezzo) miliardo</i>	0	0	0	0	0
<i>decina</i>	0	1	5	2	1
<i>(mezza) dozzina</i>	5	4	10	16	1
<i>Tot.</i>	7	6	17	36	6

1. Con qualche decina di piastre, appoggiate dal mio nome, vi darà quella che crederete la migliore per la vostra impresa. (E. Salgari, *Il figlio del corsaro rosso*, 1908)
2. seguiti da una dozzina di archibugieri spagnuoli i quali portavano delle torcie. (E. Salgari, *Il figlio del corsaro rosso*, 1908)
3. ritto di fronte ad una tavola occupata da una mezza dozzina di meticci, i quali stavano vuotando un grosso boccale di mezcal. (E. Salgari, *Gli ultimi filibustieri*, 1908)
4. Tutti e tre, seguiti a breve distanza da una mezza dozzina di malesi i quali, udendo le fucilate non avevan più potuto trattenersi, si erano slanciati a gran corsa lungo la riva del fiume nero. (E. Salgari, *La rivincita di Yanez*, 1913)

1946T	1976NT	1976T	ParaNT	ParaT
0	1	0	0	0
4	1	16	2	7
1	1	1	0	0
0	0	1	0	1
0	2	0	0	0
0	0	1	0	0
7	2	4	4	3
5	4	8	5	6
17	11	31	11	17

3.5. Testualità: l'uso dei connettivi

I connettivi sono una delle risorse più sfruttate dai traduttori per azione dell'esplicitazione. Tuttavia, nei giornali si trovano più connettivi negli articoli non tradotti, con l'eccezione di *tuttavia*. Anche nella narrativa l'analisi dei connettivi non permette di rilevare la tendenza all'esplicitazione: come si può notare nella tabella 3.15, si registra una prevalenza rilevante nelle traduzioni nel subcorpus ottocentesco e una meno marcata in quello più recente. Tra i due periodi, si assiste a una riduzione dei connettivi: le traduzioni diventano quindi sempre meno esplicite, fino a dimezzare la frequenza di connettivi rispetto al primo periodo preso in considerazione dal corpus. Inoltre, non si notano particolari differenze tra paraletteratura e letteratura alta. In ultimo, il dettaglio dei singoli connettivi conferma la loro prevalenza nei testi tradotti dell'800, a eccezione di *infatti*, caratteristico invece dei testi non tradotti. In conclusione, i dati non comprovano la tendenza all'esplicitazione dei traduttori.

	1800NT	1800T	1900NT	1900T	1946NT
<i>cioè</i>	49	98	166	207	164
<i>dunque</i>	271	567	402	199	153
<i>infatti</i>	97	49	168	85	237
<i>perciò</i>	75	161	342	190	183
<i>quindi</i>	107	170	125	183	187
<i>tuttavia</i>	80	141	223	219	162
<i>tot.</i>	679	1.186	1.426	1.083	1.086
<i>d</i>	0,789	1,379	1,182	0,895	1,188

Tabella 3.15 – Frequenza assoluta di alcuni connettivi interfrastici e densità per 1.000 del loro totale

Un'analisi di più ampio spettro, che prende in considerazione intere classi di connettivi (tabella 3.16), non conferma però quanto appena detto, al contrario di quanto rilevato negli articoli di giornale non tradotti, che hanno una densità per 1.000 forme di 58,13, contro le 49,32 degli articoli tradotti. Per quanto riguarda invece il corpus letterario, le differenze sono più ridotte e, se si considera la somma di tutte le categorie di connettivi, si può notare che sono adoperate più spesso dai traduttori dal '900. Oltre al subcorpus dell'Ottocento, fa eccezione anche quello di letteratura di consumo, in cui le traduzioni presentano più additivi, ma meno connettivi delle altre categorie. Le altre classi sono invece caratteristiche delle traduzioni in gran parte dei periodi, con alcune oscillazioni. Si distinguono gli additivi, più usati dagli autori italiani (tranne nel 1900-1945), enumerativi e temporali, che si registrano più spesso nelle traduzioni dal 1946 in poi.

1946T	1976NT	1976T	ParaNT	ParaT
59	84	49	120	39
65	141	76	58	106
49	102	74	68	59
151	56	128	88	121
180	298	351	257	197
68	123	171	140	140
572	804	849	731	662
0,623	0,655	0,735	0,731	0,661

	1800NT	1800T	1900NT	1900T	1946NT
additivi	30.385	28.393	43.362	43.622	31.286
d	35,289	33,014	35,956	36,033	34,230
avversativi	6.016	6.101	9.547	10.278	7.378
d	6,987	7,094	7,916	8,490	8,072
comparativi	4.982	5.059	7.629	9.230	5.818
d	5,786	5,882	6,326	7,624	6,365
consecutivi	2.659	3.414	3.511	4.697	2.102
d	3,088	3,970	2,911	3,880	2,300
continuativi	208	427	344	674	454
d	0,242	0,496	0,285	0,557	0,497
enumerativi	1.567	1.016	2.465	1.838	1.803
d	1,820	1,181	2,044	1,518	1,973
esplicativi	195	271	447	618	494
d	0,226	0,315	0,371	0,510	0,540
temporali	8.458	7.739	13.161	12.711	10.098
d	9,823	8,999	10,913	10,500	11,048
Tot.	54.470	52.420	80.466	83.668	59.433
d	63,260	60,952	66,722	69,113	65,025

Tabella 3.16 - Densità per 1.000 delle principali classi di connettivi

3.6. Conclusioni: differenze con la stampa periodica e prospettive di ricerca

Alla luce dei dati quantitativi illustrati nei paragrafi precedenti, è possibile trarre alcune considerazioni sul profilo dell'italiano letterario tradotto, sugli andamenti dell'italiano letterario negli ultimi due secoli e sulle differenze con la stampa periodica. In primo luogo, occorre però sottolineare i limiti imposti dalla composizione del corpus: nelle conclusioni sugli articoli di giornale del capitolo 2, si fa soprattutto riferimento alla preponderanza dell'inglese dei testi,

1946T	1976NT	1976T	ParaNT	ParaT
30.498	38.067	35.710	36.452	32.452
33,193	31,034	30,913	36,462	32,423
7.695	8.166	9.238	8.505	8.477
8,375	6,657	7,997	8,507	8,470
6.725	6.684	8.106	7.219	7.056
7,319	5,449	7,017	7,221	7,050
2.146	2.574	3.086	2.852	2.708
2,336	2,098	2,671	2,853	2,706
295	471	617	440	426
0,321	0,384	0,534	0,440	0,426
2.212	2.123	2.393	2.516	2.406
2,407	1,731	2,072	2,517	2,404
306	454	453	323	244
0,333	0,370	0,392	0,323	0,244
10.377	12.123	13.089	11.595	11.955
11,294	9,883	11,331	11,598	11,944
60.254	70.662	72.692	69.902	65.724
65,579	57,607	62,927	69,922	65,666

a cui in questo studio si aggiungono limiti legati alla distribuzione cronologica delle pubblicazioni, a quella degli autori e dei generi, molto più variegati nella prosa letteraria, che possono influire in modo considerevole sui risultati. Tali differenze nella natura dei corpora, oltre ad aver determinato differenze metodologiche, rendono alcuni tratti difficilmente comparabili: è il caso, ad esempio, della punteggiatura e dei costrutti *ad sensum*.

Nei paragrafi precedenti non sono sempre emerse tendenze definite e l'influsso della legge dell'interferenza e degli universali traduttivi non è sempre corroborato dai risultati, al contrario di quanto

rilevato per i giornali, in cui diversi tratti degli articoli tradotti sono riconducibili a interferenza, semplificazione, convergenza e, in misura minore, alla normalizzazione. Tali differenze possono essere dovute in primis alle ovvie differenze stilistiche che intercorrono tra stampa periodica e narrativa, poi anche alla diversa modalità di lavoro dei traduttori giornalistici e di quelli letterari: se i primi devono lavorare di fretta su notizie da pubblicare il prima possibile, i secondi hanno (almeno in linea teorica) più tempo a disposizione e maggiori possibilità di curare la lingua e mitigare gli effetti di interferenza e universali traduttivi sulle loro proposte.

Per quanto riguarda la narrativa, molti risultati oscillano a seconda del periodo analizzato e spesso vanno in controtendenza rispetto a quanto ci si aspetterebbe in base alle premesse teoriche sul traduttese. Per esempio, mentre la ricchezza lessicale dei vari subcorpora grosso modo conferma i risultati della stampa periodica (pur con l'eccezione del XIX secolo), così come risulta poco significativo il dato relativo alla densità lessicale, l'incidenza di vocabolario di base è minore nei testi narrativi tradotti rispetto agli originali italiani, e l'uso dei connettivi non comprova sempre la tendenza all'esplicitazione, a differenza di quanto avviene negli articoli di giornale.

Un altro aspetto interessante riguarda le varianti di registro più elevato: anche in questo caso, i dati sulla narrativa divergono da quelli della stampa, dove forme come *debbo* e *debbono* risultano più comuni nelle traduzioni rispetto ai testi non tradotti.

Inoltre, nel corpus di prosa narrativa si aggiunge la dimensione diacronica. Per es. la lunghezza media delle frasi nelle traduzioni letterarie segue un andamento particolare: fino al 1945 e nella paraletteratura le traduzioni presentano frasi più lunghe rispetto ai testi originali italiani, mentre nei periodi successivi la tendenza si inverte. Questo risultato si discosta da quanto osservato nella stampa periodica, in cui la lunghezza media delle frasi si mantiene più costante tra articoli tradotti e non tradotti. Tale inversione si può spiegare tenendo in considerazione le sperimentazioni letterarie, che si sono riflesse anche su punteggiatura e ritmo, che hanno avuto luogo nella letteratura italiana nella seconda metà del XX secolo.

Alcuni dati convalidano invece quanto già osservato per la stampa, come la maggiore frequenza della perifrasi *stare* + gerundio nelle traduzioni e la preferenza modi verbali come il condizionale e il congiunti-

vo, rilevata anche attraverso l'uso delle congiunzioni subordinanti che richiedono obbligatoriamente quest'ultimo modo verbale.

In generale, si può applicare, almeno per alcuni aspetti, la conclusione del § 2.7.1.: “i traduttori dimostrano una forte consapevolezza per quanto attiene allo strumento linguistico e si danno ‘regole’ più o meno esplicite e di applicazione più o meno costante”, che possono basarsi sul rispetto della norma tradizionale e sulla preferenza per un registro più formale (per esempio con l'uso di certi tempi o modi verbali) o su riferimenti certi quando l'italiano non offre linee guida precise.

D'altra parte, si nota che i traduttori fanno spesso da “apripista” per caratteristiche che negli autori italiani trovano minore spazio nei primi periodi. Un caso significativo è quello di *lui, lei, loro* che si affermano come pronomi soggetto prima nelle traduzioni e poi nella narrativa italiana nativa. Lo stesso si verifica anche per altri tratti, come l'impiego della *d* eufonica, in cui i traduttori adottano più precocemente e costantemente la regola proposta dalle grammatiche di riferimento odierne. Anche le forme alternative meno formali per *dovere* e *sedere* sono più frequenti nelle traduzioni. Ciò non conferma pienamente l'idea che la traduzione sia incline a resistere all'integrazione precoce di nuovi elementi linguistici e che si attenga alle forme più tradizionali e consolidate.

I risultati che abbiamo presentato qui sono da considerare come dati preliminari: saranno necessari studi ulteriori per approfondire alcuni aspetti, soprattutto con l'ausilio di corpora più ampi e variegati (ma soprattutto meglio bilanciati) dal punto di vista di autori, generi e lingue fonte, utili a ponderare l'effetto della lingua di partenza e quindi dell'interferenza. È auspicabile anche una distribuzione temporale più regolare delle opere, fondamentale per una ricostruzione più precisa dei contributi del traduttore all'italiano letterario. Tra le caratteristiche da approfondire rientrano sicuramente aspetti che non sono stati presi in considerazione come i calchi sintattici o i forestierismi e altri tratti tipici dell'italiano dell'uso medio, per misurare in modo più sistematico la maggiore apertura dei traduttori che è emersa da questo studio.

Per approfondire

Capitolo 1

Google Translate, *DeepL* o anche *ChatGPT* sono servizi di traduzione automatica ampiamente sfruttati sul web.

Il concetto di fedeltà è stato affrontato da Gianfranco Folena a un convegno triestino sulla traduzione, oggi riedito nel volume, a cura di Gianfelice Peron, *Volgarizzare e tradurre: con altri scritti sulla traduzione*, Cesati, Firenze 2021. A Juliane House si deve la distinzione tra traduzioni che dichiarano esplicitamente la loro derivazione esogena (*overt*) e traduzioni “nascoste” (*covert*), esposta nel saggio *A Model for Translation Quality Assessment*, Gunter Narr, Tübingen 1981, e poi in, *Translation Quality Assessment: A Model Revisited*, Gunter Narr, Tübingen, 1997.

La svolta descrittivista negli studi sulla traduzione è riassunta da Giuliana Garzone, in *Traduzione e interferenza: il punto di vista della traduttologia*, in *Lingua, mediazione linguistica e interferenza*, a cura di G. Garzone, A. Cardinaletti, Franco Angeli, Milano, 2004, pp. 105-127, come pure da Erika Salsnik in *Dagli universali traduttivi all'italiano delle traduzioni*, in *I saperi del tradurre*, a cura di C. Montella, G. Marchesini, Franco Angeli, Milano, 2007, pp. 101-132. Una panoramica in italiano è offerta da Stefano Ondelli

in *L'italiano delle traduzioni*, Carocci, Roma, 2020, insieme a un excursus generale sul traduttese in Italia in diversi tipi di comunicazione scritta e orale. Diverse trattazioni teoriche di tipo enciclopedico sono disponibili in inglese, come quelle curate da Mona Baker e Gabriela Saldanha (*Routledge Encyclopaedia of Translation Studies*, Routledge, London and New York, 2009) o Kirsten Malmkjaer e Kevin Windle (*The Oxford Handbook of Translation Studies*, Oxford University Press, Oxford, 2011) e la serie degli *Handbook of Translation Studies* curata da Yves Gambier e Luc Van Doorslaer per John Benjamins, Amsterdam/Philadelphia, 2010.

La descrizione degli universali traduttivi esposta in questo capitolo è il risultato dello spoglio di diversi testi di consultazione e studi specifici, tra cui principalmente quelli di Mona Baker (*Corpus-based Translation Studies: The Challenges that Lie Ahead*, in H. Somers (ed.), *Terminology, LSP and Translation. Studies in Language Engineering in Honour of Juan C. Sager*, J. Benjamin, Amsterdam, 1996, pp. 175-186), Anna Mauranen (*Universal Tendencies in Translation*, in G. Anderman, M. Rogers (eds.), *Incorporating Corpora. The Linguist and the Translator*, Multilingual Matters, Clevedon/Buffalo/Toronto, 2008, pp. 32-48) e Kirsten Malmkjaer (*Translation Universals*, in Malmkjaer K., Windle K. (eds.), *The Oxford Handbook of Translation Studies*, Oxford University Press, Oxford, 2011, pp. 83-94).

Il principale esponente degli studi descrittivi è Gideon Toury, che definisce la legge dell'interferenza e altri principi cardine di questa teoria in *Translation Studies and Beyond*, John Benjamins, Amsterdam/Philadelphia, 1995. Tra i corollari della legge dell'interferenza, l'ipotesi dei tratti specifici (*Unique Items Hypothesis*) è avanzata da Sonja Tirkkonen-Condit in *Unique Items – Over- or Under-Represented in Translated Language?*, in A. Mauranen, P. Kujamäki (eds.), *Translation Universals: do they exist?*, John Benjamins, Amsterdam/Philadelphia, 2004, pp. 177-184), mentre Kirsten Malmkjær propone il *first translational response* (prima risposta traduttiva) in *Translation Universals*, in Malmkjaer K., Windle K. (eds.), *The Oxford Handbook of Translation Studies*, Oxford University Press, Oxford, 2011, pp. 83-94: si ipotizza che, i traduttori potrebbero scegliere in maniera più o meno costante una certa soluzione per una data parola, espressione o struttura nel testo di partenza perché è quella che gli viene in mente per prima. L'esistenza del traduttese (definito "il terzo codice") è sostenuta, tra gli altri, da William Frawley in

Prolegomenon to a Theory of Translation, in Id. (ed.), *Translation: Literary, Linguistic and Philosophical Perspectives*, University of Delaware Press, Newark, NJ, 1984, pp. 159-75. Un contributo al concetto di “convergenza” è offerto da Anna Mauranen che, nel saggio *Corpora, universals and interference*, in A. Mauranen, P. Kujamäki (eds.), *Translation Universals: do they exist?*, John Benjamins, Amsterdam/Philadelphia, 2004, pp. 65-82, sostiene che, nel corpus da lei considerato, le traduzioni in finlandese, sia dal russo sia dall'inglese, anche se hanno profili diversi, restano comunque più simili tra loro rispetto ai testi scritti originariamente in finlandese. Ad Andrew Chesterman si deve la distinzione tra S- e T-Universals, illustrata nel contributo *Beyond the Particular*, in A. Mauranen, P. Kujamäki (eds.), *Translation Universals: do they exist?*, John Benjamins, Amsterdam/Philadelphia, pp. 33-49.

Tra gli studi di linguistica comparata, indubbiamente utili per tenere conto dell'interferenza, segnaliamo le ricerche di Jean-Paul Vinay e Jean-Louis Darbelnet su francese e inglese, esposte in *Stylistique comparée du français et de l'anglais: méthode de traduction*, Didier, Paris, 1977, mentre per l'italiano ricordiamo lo studio comparativo con l'inglese di Marcella Bertucelli Papi in *Prima di tradurre: note sui vincoli strutturali, concettuali e culturali nella traduzione dall'inglese in italiano*, Pisa University Press, Pisa, 2016.

La fortuna di Walter Scott in Italia e le modalità di traduzione delle sue opere nell'Ottocento sono gli argomenti del saggio di Anna Benedetti, *Le traduzioni da Walter Scott e i loro anglicismi*, Olshki, Firenze, 1974. Fabio Rossi, invece, offre una dettagliata panoramica della storia e delle caratteristiche linguistiche del doppiaggio cinematografico in *Il linguaggio cinematografico*, Aracne, Roma, 2006.

Anna Cardinaletti fornisce una descrizione dei fenomeni di interfaccia che caratterizzano l'interferenza in *La traduzione: un caso di attrito linguistico*, in *L'italiano delle traduzioni*, a cura di A. Cardinaletti, G. Garzone, Franco Angeli, Milano, 2005, pp. 59-84; Michele A. Cortelazzo si occupa invece della perifrasi progressiva come possibile risultato dell'influsso dell'inglese in *La perifrasi progressiva in italiano è un anglicismo sintattico?*, in AA.VV., *Studi in onore di Pier Vincenzo Mengaldo per i suoi settant'anni. A cura degli allievi padovani*, vol. II, SISMEL, Edizioni del Galluzzo, Firenze, 2007, pp. 1753-1764. Francesco Sabatini è l'autore del saggio, *L'italiano dell'uso medio: una realtà tra le varietà*

linguistiche italiane, in G. Holtus, E. Radtke (hrg.), *Gesprochenes Italienisch in Geschichte und Gegenwart*, Gunter Narr Verlag, Tübingen, 1985, pp. 154-184, in cui individua e descrive le caratteristiche dell'italiano di uso quotidiano che divergono dalla norma codificata di stampo letterario.

Le nozioni di base e una panoramica su metodologie e strumenti della linguistica dei corpora sono disponibili nei volumi di Emanuela Cresti e Alessandro Panunzi (*Introduzione ai corpora dell'italiano*, Il Mulino, Bologna, 2013), Maria Freddi (*La linguistica dei corpora*, Carocci, Roma, 2014) e, soprattutto, Arjuna Tuzzi (*Fondamenti di analisi dei dati testuali*, Carocci, Roma, 2024). La definizione di "corpus" esposta in questo paragrafo è derivata dalla proposta di Manuel Barbera contenuta nel volume *Linguistica dei corpora e linguistica dei corpora italiana. Un'introduzione*, Qu.A.S.A.R., Milano, disponibile all'indirizzo www.bmanuel.org/man/Barbera_IntroduzioneCL_2013=Ver1-54.pdf, consultato il 15 febbraio 2025. Per quanto riguarda criteri di compilazione del corpus per garantire rappresentatività e bilanciamento, è possibile consultare il capitolo di Stefano Ondelli intitolato *Treat Texts as Data but Remember They Are Made of Words: Compiling and Processing Corpora*, in A. Tuzzi (ed.), *Tracing the Life Cycle of Ideas in the Humanities and Social Sciences*, Springer, Berlin, 2018, pp. 133-150. Riflessioni analoghe, ma in lingua italiana e focalizzate sul caso specifico degli articoli di giornale, sono disponibili nell'articolo, dello stesso autore, dal titolo *Per un'analisi dell'italiano tradotto nei quotidiani: considerazioni preliminari sulla costituzione di un corpus*, in "Rivista internazionale di tecnica della traduzione", 10, 2008, pp. 81-99.

Esistono diversi metodi per calcolare la distanza intertestuale. Quello utilizzato in questo volume è stato sviluppato da Dominique Labbé, ed è illustrato in lingua italiana nell'articolo *Corneille nell'ombra di Molière. Come identificare un autore?*, traduzione di Irene Borsato, in "Rivista internazionale di tecnica della traduzione", 12, 2010, pp. 117-138.

Il concetto di "traduzioni invisibili" è stato elaborato da Daniele Grasso nella tesi dottorale *Innovazioni sintattiche in italiano (alla luce della nozione di calco)*, Université de Genève, Genève, disponibile online all'indirizzo <https://archive-ouverte.unige.ch/unige:475>, consultato il 15 febbraio 2025.

Uno studio di riferimento sulla lingua dei giornali e sulla centralità della cronaca politica continua senza dubbio a essere

quello di Maurizio Dardano: *Il linguaggio dei giornali italiani*, Bari, Laterza, 1986. Sulle caratteristiche linguistiche della paraletteratura è possibile consultare invece il saggio di Gabriella Alfieri *La lingua di consumo*, in *Storia della lingua italiana*, vol. II: *Scritto e parlato*, a cura di L. Serianni, P. Trifone, Einaudi, Torino, 1994, pp. 161-235, e il volume di Laura Ricci dal titolo *Paraletteratura. Lingua e stile dei generi di consumo*, Carocci, Roma, 2013.

La distanza intertestuale è stata messa alla prova sui testi letterari italiani da Michele A. Cortelazzo, Arjuna Tuzzi e Paolo Nadalutti, anche con il metodo dei campionamenti utilizzato in questo volume, nell'articolo *Improving Labbé's Intertextual Distance: Testing a Revised Version on a Large Corpus of Italian Literature*, in "Journal of Quantitative Linguistics", 20:2, 2013, pp. 125-152. Un tentativo analogo è stato condotto nel 2016 da Mattia Bernardini nel contributo *Originalità della traduzione letteraria: una questione di distanze*, disponibile online all'indirizzo www.treccani.it/magazine/lingua_italiana/speciali/traduttese/Bernardini.html, consultato il 15 febbraio 2015. Quest'ultimo, però, effettua le sue misurazioni su un numero ridotto di testi (48 traduzioni di 16 romanzi da 4 lingue diverse) che risultano poco utili a valutare efficacemente l'incidenza dei fattori "traduttore" e "lingua di partenza": è inevitabile che le diverse traduzioni di una stessa opera risultino meno "distanti". Altre ricerche di impianto almeno parzialmente quantitativo su corpora letterari e paraletterari sono state condotte da Mario Squartini, (*Contributo per la caratterizzazione aspettuale delle perifrasi italiane andare + gerundio, stare + gerundio, venire + gerundio. Uno studio diacronico* in "Studi e saggi linguistici" 30, 1990, pp. 117-212), Pier Marco Bertinetto (*Tempi verbali e narrativa italiana dell'Otto/Novecento*, Edizioni dell'Orso Alessandria, 2003) e Chiara Degano (*Influssi inglesi sulla sintassi italiana: uno studio preliminare sul caso della perifrasi progressiva*, in *L'italiano delle traduzioni*, a cura di A. Cardinaletti, G. Garzone, Franco Angeli, Milano, 2005, pp. 85-105), con particolare attenzione per la perifrasi progressiva.

Marianna De Biasi (che ringraziamo) ha compilato il corpus di paraletteratura del XX secolo nella sua tesi di laurea triennale dal titolo "*I verbi procomplementari. Analisi di corpora letterari tradotti e nativi della seconda metà del Novecento*", Università di Trieste, 2018.

Capitolo 2

Per quanto riguarda il software utilizzato per le analisi, *Taltac2* è disponibile online all'indirizzo www.taltac.com, mentre *TreeTagger* può essere scaricato dalla pagina www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/ (ultima consultazione in entrambi i casi il 15 febbraio 2025). L'incidenza del Vocabolario di Base è stata calcolata con il programma allegato alla *Guida all'uso delle parole* di Tullio De Mauro (Editori riuniti, Roma, 1997), mentre per determinare la lunghezza dei periodi è stato utilizzato il programma *Èulogos autogulp*, allegato al volume *Capire e farsi capire. Teorie e tecniche della scrittura controllata*, Napoli, Tecnodid, 1996, a firma di Maria Emanuela Piemontese.

Per l'analisi dei forestierismi è stato consultato principalmente il *Dizionario delle parole straniere nella lingua italiana* di De Mauro e Mancini (Milano, Garzanti, 2001), la cui fonte è il *Grande dizionario italiano dell'uso*, cura di T. De Mauro, Torino, Utet, 1999. Altri dizionari utilizzati per verificare datazioni e accezioni dei lemmi comprendono Zingarelli, Devoto Oli e Treccani. Per valutare i fenomeni morfosintattici sono state consultate la *Grande grammatica italiana di consultazione*, a cura di L. Renzi, G. Salvi e A. Cardinaletti, Bologna, il Mulino, 1988-2001; la *Grammatica italiana. Italiano comune e lingua letteraria. Suoni forme costrutti* a cura di Luca Serianni (Torino, UTET, 1988); e la *Nuova grammatica della lingua italiana*, a firma di Maurizio Dardano e Pietro Trifone (Bologna, Zanichelli, 1997). In particolare, da quest'ultimo testo e dal volume *Lingua italiana. Una grammatica ragionevole* (Padova, Unipress, 2006) di Maria Giuseppa Lo Duca e Rosa Solarino, sono state estratte le liste di connettivi su cui si basano le elaborazioni statistiche proposte in questo volume.

Per quanto riguarda il ruolo dell'italiano dei giornali come modello di riferimento per la lingua del XXI secolo, rimandiamo al contributo *Lingua*, in *Modernità italiana. Cultura, lingua e letteratura dagli anni Settanta a oggi*, a cura di A. Afribo, E. Zinato, Carocci, Roma, 2011, pp. 15-52, in cui Giuseppe Antonelli aggiorna il grafico a suo tempo proposto da Gaetano Berruto per descrivere le varietà dell'italiano contemporaneo. Sulla lingua dei giornali, e in particolare l'uso espressivo della punteggiatura, si veda Riccardo Gualdo, *L'italiano dei giornali*, Carocci, Roma, 2007.

Sulla tendenza al conservatorismo, che spinge i traduttori a sostituire anche forestierismi di uso comune, Stefano Ondelli, nel

saggio *Inglese e 'eurocratese', in Italiano e inglese a confronto: problemi di interferenza linguistica*, a cura di A. V. Sullam Caimani, Franco Cesati, Firenze, 2003, pp. 177-195, ricorda che, fino a qualche tempo fa circolavano vademecum per i traduttori dell'UE in cui si indicavano sostituenti italiani anche per forestierismi d'uso comune come *lobby* o *premier*.

Sulla lingua politicamente corretta, e in particolare sul rispetto della parità di genere, il riferimento principale resta senza dubbio quello relativo a *Il sessismo nella lingua italiana*, Commissione Nazionale per la realizzazione della parità tra uomo e donna, Roma, Presidenza del Consiglio dei Ministri, Roma, 1987, a firma di Alma Sabatini. La questione fuori d'Italia è affrontata da Edoardo Crisafulli nel volume *Igiene verbale. Il politicamente corretto e la libertà linguistica*, Vallecchi, Firenze, 2004. Per una bibliografia aggiornata sull'argomento è possibile consultare il volume curato da Stefano Ondelli dal titolo *Le italiane e l'italiano*, EUT, Trieste, 2020.

Interessanti osservazioni sull'alternanza tra perfetto semplice e composto (cioè passato remoto e prossimo), in italiano come in traduzione, sono offerte dai seguenti lavori: Pier Marco Bertinetto e Mario Squartini, *La distribuzione del Perfetto Semplice e del Perfetto Composto nelle diverse varietà di italiano*, in "Romance Philology", 49, 4, 1996, pp. 383-419; Michele A. Cortelazzo, *Perfetto semplice e perfetto composto in italiano*, in *Grammatica. Studi interlinguistici*, a cura di C. Taylor Torsello, Unipress, Padova, 1997, pp. 199-208; Mario Squartini e Pier Marco Bertinetto, *Simple and Compound Past in Romance languages*, in *Tense and aspect in the languages of Europe*, a cura di O. Dahl, Mouton de Gruyter, Berlin-New York, 2000, pp. 403-440, 2000. La prevalenza del perfetto semplice viene rilevata nel doppiaggio cinematografico da Fabio Rossi (*Il linguaggio cinematografico*, Aracne, Roma, 2006) e da Maurizio Viezzi (*Written Translation and Simultaneous Translation Compared and Contrasted: A case study*, in "The Interpreter's Newsletter", 5, 1993, pp. 94-100).

Per quanto riguarda l'effetto della legge dell'interferenza sui pronomi in italiano, i lavori di Anna Cardinaletti (*La traduzione dei pronomi: interferenza sintattica e cambiamento linguistico*, in *Lingua, mediazione linguistica e interferenza*, a cura di G. Garzone e A. Cardinaletti, Franco Angeli, Milano, 2004, pp. 129-150 e *La traduzione: un caso di attrito linguistico*, in *L'italiano delle traduzioni*,

a cura di A. Cardinaletti e G. Garzone, Franco Angeli, Milano, 2005, pp. 59-84) e Giuliana Garzone (*Osservazioni sull'assetto del testo italiano tradotto dall'inglese*, in *L'italiano delle traduzioni*, a cura di A. Cardinaletti e G. Garzone, Franco Angeli, Milano, 2005, pp. 35-58), basati sul confronto tra testi in lingua originale a soggetto obbligatorio e relative traduzioni italiane, hanno messo in luce la frequente presenza di un pronome esplicito là dove un pronome soggetto vuoto sarebbe perfettamente possibile.

La tendenza nell'italiano moderno a ricorrere a possessivi anche in contesti in cui questi non sono strettamente necessari è stata spesso ricondotta a influenze esogene, inizialmente francesi (come notato da Silvia Morgana nel saggio dal titolo *L'influsso francese*", in *Storia della lingua italiana*, vol. 3, *Le altre lingue*, a cura di L. Serianni e P. Trifone, Einaudi, Torino, 1994, pp. 671-719) o, soprattutto in tempi più recenti, inglesi.

A livello della punteggiatura, non solo emergono spinte contrastanti tra la tendenza a ricalcare la punteggiatura dell'originale, la semplificazione (che condurrebbe alla riduzione della lunghezza media dei periodi) e l'esplicitazione (che, tramite segni di punteggiatura forti, aiuterebbe il lettore nell'interpretazione del testo): Federica Scarpa (*Corpus-based Specialist Translation Quality Assessment: A Study Using Parallel and Comparable Corpora in English and Italian*, in *Insights into Specialized Translation*, a cura di S. Sarcevic e M. Gotti, Peter Lang, Bern, 2006, pp. 155-172) nota che le traduzioni in italiano di testi inglesi, pur essendo più lunghe degli originali, tendono a presentare un numero minore di periodi. Entrano quindi in gioco tradizioni stilistiche diverse, come ben sanno gli studenti (almeno i migliori) che, nel tradurre dall'inglese verso l'italiano, producono testi con periodi più lunghi e dotati di maggiore ricchezza lessicale rispetto all'originale allo scopo di essere valutati più positivamente.

Davide Mastrantonio illustra usi e strutture delle frasi specificative (*Sul ruolo delle frasi specificative all'interno del testo* in "Lingua e stile", vol. 58 (1), 2024, pp. 63-87) e sostiene che in frasi come "il problema sono le zanzare" o "il mio piatto preferito sono gli spaghetti", il soggetto è lo specificatore perché determina l'accordo verbale. Ne consegue che strutture che contengono un quantificatore grammaticalmente singolare ma dal significato plurale, come in "la maggioranza degli studenti sono ragazzi del luogo", non possono essere considerati casi

di concordanza *ad sensum* (e infatti il verbo al singolare non sarebbe accettabile).

La descrizione e classificazione complessiva dei calchi sintattici proposta in questo volume è derivata dalla tesi dottorale di Daniele Grasso già menzionata sopra, mentre i valori specifici dei costrutti verbali con il prefissoide *auto-* sono descritti da Claudio Iacobini nel contributo *Due casi di interferenza dall'inglese sulla morfologia derivazionale dell'italiano*, in *Italiano e inglese a confronto: problemi di interferenza linguistica*, cura di A. V. Sullam Caimani, Cesati, Firenze, 2003, pp. 43-56.

L'impatto dal genere testuale sull'impiego dei connettivi in traduzione alla luce dell'universale dell'esplicitazione è stato indagato da Giuseppe Palumbo e Maria Tera Musacchio (*When a Clue is not a Clue. A corpus-driven study of explicit vs. implicit signalling of sentence links in popular economics translation*, in "Rivista internazionale di tecnica della traduzione", 10, 2010, pp. 63-76) e José Francisco Medina Montero (*La traducción de los marcadores del discurso del español al italiano: análisis de un corpus de textos*, sempre nella stessa rivista, pp. 77-92).

Oltre a quello illustrato in questo volume, un altro tentativo di calcolare la distanza intertestuale prendendo in considerazione le parole grammaticali è stato fatto da Mattia Baroni e Silvia Bernardini in *A New Approach to the Study of Translationese: Machine-learning the Difference between Original and Translated Text*, in "Literary and Linguistic Computing", 21(3), 2006, pp. 259-274. L'elenco delle parole grammaticali usato per la ricerca illustrata in questo volume è lo stesso utilizzato da Michele A. Cortelazzo, Arjuna Tuzzi e Paolo Nadalutti (che qui ringraziamo) nel saggio già menzionato sopra.

Capitolo 3

Il capitolo condivide gran parte dei riferimenti con quello precedente, vista la sovrapponibilità della loro struttura e di obiettivi e metodi.

Ai software utilizzati si aggiungono *TagAnt* (<https://laurenceanthony.net/software/tagant/>) e *Bran* (<https://github.com/zorba-project/Bran>), entrambi gratuiti e liberamente scaricabili.

Per approfondire l'argomento dei *realia* o *culturemi*, rimandiamo al lavoro del 1993 di Sider Florin, *Realia in Translation* (in *Translation as Social Action. Russian and Bulgarian Perspectives*, a

cura di Palma Zlatevam, Routledge, Londra/New York, 1993, pp. 122-128). Più specificamente, dei *realia* in italiano e nell'ambito della didattica della traduzione si è occupata Lorenza Rega nel contributo *Realia e didattica della traduzione*, in *Testo e traduzione*, a cura di Fabiana Fusco e Monica Ballerini, Francoforte, Peter Lang, 2010, pp. 245-256.

Nel capitolo si fa riferimento al ruolo delle congiunzioni come spia di complessità sintattica, rilevato nelle ricerche di Miriam Voghera pubblicate nel 2004, che approfondiscono le differenze tra lingua parlata e lingua scritta (*La distribuzione delle parti del discorso nel parlato e nello scritto*, in *La variabilité en langue*, vol. 1° (*Langue parlée et langue écrite dans le présent et dans le passé*), a cura di Rika Van Deyck, Rosanna Sornicola, Johannes Kabatek, Gent, *Communication & Cognition*, (*Studies in Language*, 9), pp. 261-284).

Per quanto riguarda la punteggiatura nella narrativa, che ha rappresentato uno spazio di sperimentazione nella letteratura del XX secolo, due approfondimenti interessanti sono *Ritmo e melodia nella prosa italiana: studi e ricerche sulla prosa d'arte* di Gianluigi Beccaria (Firenze, Olschki, 1964) e *Sintassi e stile della narrativa italiana dagli anni Sessanta a oggi* di Giuseppe Antonelli (in *Storia generale della letteratura italiana, Volume XII*, a cura di Nino Borsellino e Walter Pedullà. Milano, Motta, 1999). Per le tendenze in diacronia dell'italiano, rimandiamo invece al lavoro fondamentale di Bice Mortara Garavelli (*Prontuario di punteggiatura*, Laterza, Roma-Bari, 2003) e alle ricerche di Angela Ferrari, Letizia Lala e Filippo Pecorari (*La punteggiatura italiana attraverso i corpora. Teoria, sincronia e diacronia*, in *Corpora e studi linguistici. Atti del LIV convegno internazionale di studi della Società di Linguistica Italiana. Online, 8-10 settembre 2021*, a cura di Emanuela Cresti e Massimo Moneglia, Società di Linguistica Italiana, Roma, 2022).

A conferma del ruolo di innovatori dei traduttori, si vedano anche le considerazioni di Ondelli in *Le perifrasi gerundivali nel romanzo italiano 1800-2000: aspetti quantitativi* (in *Sulle amate sponde. Scritti di storia della lingua italiana per Gabriella Cartago*, a cura di Giuseppe Sergio, Rosa Argenziano e Jacopo Ferrari, Biblion, Milano, 2021, pp. 109-131). Dallo stesso articolo sono tratti i dati sulla diffusione di *stare* + gerundio riportati nel § 3.3.2