

Evidenze quantitative sull'italiano tradotto in un corpus giornalistico ^{1*}

Stefano Ondelli ¹, Matteo Viale ²

¹ Scuola Superiore di Lingue Moderne per Traduttori e Interpreti dell'Università degli Studi di Trieste – Via Fabio Filzi 14 – 34132 Trieste – Italia

² Dipartimento di Romanistica dell'Università degli Studi di Padova – Via Beato Pellegrino 1 – 35127 Padova – Italia

Riassunto

Da qualche tempo la letteratura linguistica ha messo in luce che la lingua dei testi tradotti presenta una serie di caratteristiche legate al fenomeno dell'“interferenza linguistica” e riconducibili a “universali traduttivi”. L'effettiva presenza delle peculiarità del cosiddetto “traduttese” è stata però verificata e approfondita solo saltuariamente attraverso sistematiche analisi quantitative di corpora testuali. Concentrandosi in particolare sull'italiano, l'intervento proposto indaga le peculiarità linguistiche dei testi tradotti nella prosa giornalistica. Per realizzare questo obiettivo, l'analisi si basa su un corpus appositamente costituito di testi giornalistici tradotti messo a confronto con un corpus paragonabile di testi giornalistici redatti direttamente in italiano. L'analisi è portata avanti secondo tecniche della linguistica dei corpora e attraverso appositi software di analisi testuale (*TALTAC2* e *TreeTagger*). I rilevamenti lessicometrici realizzati con questi strumenti informatici forniscono primi elementi di studio che consentono di far emergere usi lessicali tipici dell'italiano tradotto che vanno al di là degli specifici argomenti trattati. L'analisi del corpus non si limita agli aspetti lessicali, ma è tesa a cogliere i principali tratti morfosintattici e stilistici attraverso appositi strumenti linguistico-informatici. Il quadro d'insieme così delineato consente, sulla base di evidenze empiriche, di caratterizzare con maggiore precisione gli usi linguistici legati all'italiano tradotto nei vari livelli in cui si manifestano nella prosa giornalistica e di offrire nuovi spunti di ricerca.

Abstract

Recent research has highlighted a series of specific features in translated texts deriving from “linguistic interference” and the influence of “translation universals”. However, the characteristics of “translationese” have been only sporadically identified and studied through the systematic quantitative analysis of text corpora. This article is focussed on Italian and investigates a corpus of newspaper articles collected *ad hoc*. These are compared and contrasted with a similar corpus of newspaper articles originally written in Italian. The analysis was conducted following the principles of corpus linguistics and used dedicated text analysis software (*TALTAC2* and *TreeTagger*). The lexicometric data provided by the software supplied the raw material allowing us to investigate the lexical features of translated Italian, irrespective of the subject matter treated in the articles. In addition, the analysis identifies the main morphological, syntactic and stylistic traits by using computerised language analysis tools. The results obtained from empirical evidence provide new insights into the linguistic traits of newspaper articles translated into Italian and reveal new prospects for further research.

Keywords: corpus linguistics, language interference, language of the press, translation, translation universals

¹ * La redazione dei paragrafi 1.1., 1.3., 2.2., 3.2. e 4.1. è dovuta a Stefano Ondelli; la redazione dei paragrafi 1.2., 2.1., 3.1., 3.3., 4.2. e 5 a Matteo Viale.

1. Presentazione della ricerca

1.1. *Il traduttese e l'italiano contemporaneo*

A partire dagli anni '80 gli studi di traduttologia hanno assegnato sempre minore importanza all'opposizione tra fedeltà e infedeltà del testo di arrivo rispetto a quello di partenza, concentrandosi piuttosto sulla ricezione della traduzione nel sistema culturale e linguistico di accoglienza. Il testo tradotto viene quindi considerato come messaggio autonomo che entra a far parte del sistema culturale di arrivo, modificandolo e influenzandolo.

A livello linguistico, le conseguenze possono essere importanti. Alfieri (1994) ipotizza che l'italiano tradotto, in particolare quello del doppiaggio televisivo, possa avere ripercussioni sulle abitudini linguistiche delle fasce più giovani del pubblico, che potrebbero far proprie costruzioni ricorrenti e percepite come moduli fissi, del tipo “il mio nome è” al posto di “mi chiamo”. Su questa scia si collocano gli studi raccolti in Garzone e Cardinaletti (2004) e in Cardinaletti e Garzone (2005): senza alcun intento censorio, l'obiettivo di questo approccio è di individuare regolarità e tendenze, per avanzare ipotesi sulla grammatica mentale dei traduttori e inquadrare le peculiarità della lingua tradotta sullo sfondo dell'evoluzione dell'italiano contemporaneo.

Per quanto concerne gli aspetti più strettamente linguistici, in linea generale questi studi condividono l'assunto in base al quale il testo tradotto, anche da professionisti esperti, tende a presentare tratti marcati, soprattutto a livello sintattico, che spesso sfuggono alla percezione immediata e non sono causa di disturbo nella ricezione del destinatario. Il fenomeno sarebbe riconducibile a quelli che nella letteratura specializzata sono noti come universali traduttivi (cfr., tra gli altri, Baker, 1998), cioè tendenze e strategie adottate da qualsiasi traduttore a prescindere dalla coppia di lingue in gioco. In sintesi, si tratta di:

- *Semplificazione ed evitamento delle ripetizioni*: le scelte lessicali tendono ad appiattirsi su varietà poco connotate e a prediligere iperonimi, approssimazioni culturali e parafrasi esplicative al posto di termini specifici. A livello stilistico, oltre all'eliminazione delle ripetizioni, la semplificazione prevede anche la preferenza accordata a strutture sintattiche più semplici e brevi di quelle presenti nel testo originale.
- *Esplicitazione*: i nessi logici tra le parti del testo sono esplicitati, talvolta in maniera ridondante, anche là dove nel testo originale erano lasciati impliciti. Nello sforzo di garantire una migliore comprensione da parte del lettore, il traduttore tende a inserire informazioni aggiuntive (tramite aggettivi, glosse esplicative, la precisazione di informazioni vaghe o sottaciute) e allunga la traduzione.
- *Normalizzazione*: il traduttore cerca di rispettare la norma codificata della lingua di arrivo. Eventuali peculiarità stilistiche del testo di partenza (per es., punteggiatura irregolare, strutture sintattiche incomplete, l'imitazione della varietà orale della lingua, l'impiego di espressioni originali o arcaiche) possono venire “normalizzate” nella traduzione, che risulta più convenzionale in seno alla cultura di arrivo.
- *Interferenza*: il testo tradotto tende a ricalcare la strutturazione del testo fonte a tutti i livelli. Questo atteggiamento può variare molto in base all'esperienza del traduttore e al maggiore o minore prestigio assegnato al testo originale nella cultura in cui sarà accolto.

1.2. *Le traduzioni nei giornali italiani*

La forte diffusione di testi tradotti che non rendono esplicita la propria natura esogena (oltre all'editoria, si pensi solo al doppiaggio di film e telefilm) potrebbe avere un certo impatto sulla comunità dei parlanti che, pur non conoscendo o non utilizzando la lingua straniera dell'originale, viene esposta a fenomeni di interferenza linguistica. Con le notevoli eccezioni viste sopra, mancano però studi di ampio respiro atti a definire le caratteristiche dell'italiano tradotto.

Oggi la lingua della stampa periodica è la varietà che, a torto o a ragione, ha rimpiazzato l'italiano letterario in qualità di “barometro” e modello di riferimento, per lo meno nell'uso quoti-

diano (Dardano, 1994: 344). In questo contesto, la netta impressione è che negli ultimi anni si sia registrato un aumento del numero degli articoli tradotti da lingue straniere e pubblicati dalle principali testate nazionali ². Tra l'altro, la rapidità e serialità della redazione sembrano favorire l'adozione e l'applicazione di strategie traduttive standard ricorrenti e finiscono per indebolire lo sforzo teso a eliminare l'influsso della lingua di partenza sul testo di arrivo.

1.3. Obiettivi della ricerca

In questo articolo si illustrano i dati relativi all'analisi di alcuni tra i tratti più macroscopici che riguardano l'assetto linguistico di un corpus di articoli tradotti appositamente costituito. La ricerca proposta comprende indagini e misurazioni condotte a partire da un corpus di ampie dimensioni e basate su strumenti atti al trattamento automatico di testi (il software a cui ci si è affidati per lo studio è soprattutto *TALTAC2*). L'obiettivo è duplice: in primo luogo, si intende stabilire se i frammenti scritti ordinariamente in italiano e quelli tradotti presentino tratti peculiari e ricorrenti che permettono di evidenziare differenze significative tra i due gruppi; in secondo luogo, si desidera verificare se i dati lessico-testuali così ottenuti possono fornire riscontri oggettivi a conferma degli assunti della teoria degli universali traduttivi e, più in generale, degli studi linguistici sull'argomento.

A livello della ricerca in linguistica italiana, i risultati di questo studio possono contribuire a colmare una doppia lacuna: oltre ad approfondire la descrizione della varietà giornalistica della lingua ³, si offrono dati utili alla definizione dell'italiano tradotto, un ambito carente di sistematiche analisi quantitative basate su corpora testuali.

2. Il corpus

2.1. Il corpus studiato: costituzione, variabili e problemi

Per la ricerca qui proposta sono stati sottoposti a classificazione e analisi due corpora comprendenti articoli pubblicati tra il 2001 e il 2008 nel «Corriere della Sera» e nell'«Unità». La costituzione del corpus e la classificazione degli articoli è avvenuta tenendo conto delle seguenti variabili: testata; anno di pubblicazione dell'articolo; lingua originale a partire dalla quale l'articolo è stato tradotto; traduttore; genere dell'articolo.

² Il giudizio è il risultato della consultazione degli archivi online delle testate qui considerate ed è stato confermato dagli addetti ai lavori. A tale proposito si ringraziano Marco Del Corona della redazione del «Corriere della Sera» e Angelo Rinaldi di «Repubblica» per le preziose informazioni fornite.

³ Pur dedicando l'appendice II all'influsso dell'inglese sulla lingua dei quotidiani, nell'edizione del 1986 del *Linguaggio dei giornali italiani* di Dardano non si fa accenno alla pratica di tradurre articoli ricavati da testate straniere. Tra gli studiosi che più di recente hanno condotto studi sulla lingua dei giornali (per es., Gualdo, 2007; Bonomi, 2002; 2003), non risulta che qualcuno si sia occupato in modo specifico delle traduzioni. In ambito traduttologico, i testi giornalistici entrano a far parte dei compositi corpora d'analisi utilizzati in nei vari saggi che compongono Garzone and Cardinaletti (2004) e Cardinaletti and Garzone (2005), ma non sono oggetto di considerazione specifica. Per sondare quanto il «traduttese» sia identificabile attraverso l'analisi computerizzata, Baroni and Bernardini (2006) e Bernardini and Baroni (2005) selezionano articoli tratti dalla rivista «Limes», dunque scarsamente assimilabili ai testi della stampa quotidiana o periodica, mentre le indagini proposte in Mazzoleni and Baroni (2005) si limitano ai toponimi stranieri. Probabilmente la ricerca caratterizzata dall'approccio più simile a quello qui proposto è stata condotta da Musacchio (2005), anche se l'interesse principale era dato dal processo traduttivo e lo spoglio ha riguardato corpora paralleli e paragonabili di dimensioni ridotte ai fini dell'analisi statistica (cfr. Tuzzi, 2003), comprendenti articoli di argomento economico in inglese (con le relative traduzioni) e in italiano.

L'indicazione della testata e l'anno di pubblicazione servono a identificare gli articoli e non presentano problemi di sorta. L'informazione sulla lingua di partenza è evidentemente volta all'eventuale raffronto tra testi che potrebbero presentare fenomeni di interferenza diversi proprio perché derivanti da lingue di partenza diverse. L'indicazione del traduttore (o, nel caso del corpus di controllo, dell'autore) si giustifica per ragioni di bilanciamento: è palese che una maggioranza di testi tradotti dalla stessa persona farebbe emergere scelte individuali piuttosto che tendenze derivanti dalla pratica del tradurre in sé. Occorre notare che gli articoli tradotti non appartengono quasi mai (e solo parzialmente) alla cronaca politica e cittadina, ovvero i settori che secondo Dardano (1986: 20-23) sarebbero più rappresentativi della lingua dei giornali. Si tratta, invece, per lo più di editoriali e articoli di commento (sporadici gli esempi 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 questa tipologia, i tratti specifici legati alla personalità e alla lingua dell'autore prevalgono sulle caratteristiche tipiche della lingua dei giornali.

Una volta costituito e classificato il corpus di prosa giornalistica tradotta, si è proceduto alla costituzione di un corpus di controllo adatto comprendente articoli scritti originariamente in italiano. Tale operazione si è resa necessaria per garantire la massima omogeneità e, in prospettiva, per caratterizzare una sotto-varietà specifica dell'italiano giornalistico (quella, appunto, dell'articolo di commento) finora poco studiata, probabilmente perché ritenuta inadatta a descrivere la varietà nel suo complesso. La raccolta di un corpus di testi italiani tipologicamente analoghi agli articoli tradotti è resa necessaria dalla metodologia di indagine che prevede il ricorso a corpora paragonabili⁴.

2.2. Dimensioni e caratteristiche del corpus

Tab. 1 esplicita il numero di articoli esaminati per ciascun subcorpus. Poiché le coppie di subcorpora sono state costruite in modo da avere la stessa dimensione, emerge il maggior numero di articoli che compongono il corpus "Corriere originale" (e la conseguente lunghezza media inferiore: cfr. Tab. 3).

Per quanto riguarda le lingue di partenza degli articoli tradotti, non sorprende la netta predominanza dell'inglese. L'alto numero di testi di cui non è stata specificata la lingua originale è una conseguenza della modalità di definizione della stessa: 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 amor di precisione, nel caso delle agenzie di stampa che distribuiscono testi a livello internazionale non è stata attribuita alcuna lingua sorgente, anche se si può presumere che la stragrande maggioranza dei testi sia stata tradotta dall'inglese.

Per quanto concerne la definizione dei fenomeni da indagare, un corpus così costituito permette di valutare alcuni aspetti generali (per es., la distribuzione delle categorie grammaticali) e di concentrarsi sui tratti che caratterizzano il "traduttese" a prescindere dalla coppia di lingue coinvolte (per es., la tendenza all'esplicitazione), come pure su eventuali fenomeni di interferenza

⁴ Si vedano, per es., le considerazioni svolte in Korzen (2007) per un'analisi interlinguistica. Si distingue tra "corpora paralleli", necessari per un raffronto interlinguistico (originali vs traduzioni), e "corpora monolingui paragonabili", in cui il corpus oggetto di analisi viene messo a confronto con un corpus di controllo (cfr. Olohan, 2004: capp. 3 e 4).

ascrivibili a tutte (o quasi) le lingue di partenza (per es., fatto salvo lo spagnolo, l'esplicitazione del pronome soggetto). Se si intendessero valutare aspetti legati a una lingua specifica (per es., la frequenza dell'anteposizione dell'aggettivo al nome, tipico esempio di influenza dell'inglese sull'italiano) occorrerebbe procedere alla selezione di subcorpora appositi.

	<i>Unità tradotto</i>	<i>Corriere tradotto</i>	<i>Subtotale tradotto</i>	<i>Unità originale</i>	<i>Corriere originale</i>	<i>Subtotale originale</i>	<i>Totale</i>
Articoli	255	291	546	252	385	637	1.183
autori / traduttori	24	33	57	37	29	66	123
ling. or. italiano	–	–	–	252	385	637	637
ling. or. inglese	141	136	277	–	–	–	277
ling. or. francese	4	37	41	–	–	–	41
ling. or. spagnolo	5	12	17	–	–	–	17
ling. or. russo	0	1	1	–	–	–	1
ling. or. tedesco	0	3	3	–	–	–	3
ling. or. non det.	105	102	207	–	–	–	207

Tabella 1: Numero di articoli del corpus in base alle principali variabili prese in considerazione

Le misure lessicometriche (Tab. 2) consentono di evidenziare alcune peculiarità linguistiche dell'italiano tradotto rispetto ai testi redatti direttamente in italiano. In particolare, il corpus tradotto presenta una ricchezza lessicale minore, come mette in luce, tra l'altro, la minore percentuale di parole diverse ($V/N \cdot 100$) e di parole usate una sola volta (hapax %) e la più elevata frequenza media delle forme (N/V). A questo livello, gli universali traduttivi individuano due tendenze contrastanti: da una parte l'evitamento delle ripetizioni, dall'altra il ricorso a una normalizzazione stilistica che potrebbe espungere scelte lessicali poco convenzionali e determinare, *ceteris paribus*, il sovrasfruttamento di un minor numero di forme. Lo scarto che emerge dai dati, pur sfavorevole al subcorpus tradotto in maniera omogenea, è estremamente contenuto e non consente di trarre conclusioni univoche.

	<i>Unità tradotto</i>	<i>Corriere tradotto</i>	<i>Subtotale tradotto</i>	<i>Unità originale</i>	<i>Corriere originale</i>	<i>Subtotale originale</i>	<i>Totale</i>
N	268.516	294.870	563.409	268.738	295.366	564.144	1.124.686
V	24.291	29.653	39.654	27.100	31.591	43.553	60.099
$V/N \cdot 100$	9,05	10,06	7,04	10,08	10,70	7,72	5,34
Hapax %	49,85	50,62	47,46	49,93	51,19	48,04	45,37
N/V	11,05	9,94	14,21	9,92	9,35	12,95	18,71
V / \sqrt{N}	46,88	54,61	52,82	52,28	58,13	57,99	56,67
$a = \log N / \log V$	1,22	1,22	1,25	1,22	1,22	1,24	1,27

Tabella 2: Misure lessicometriche del corpus

Accanto alle misure lessicometriche, altri strumenti rendono possibile il recupero di dati che forniscono ulteriori elementi per valutare la complessità linguistica dal punto di vista sintattico e lessicale (Tab. 3).

Anche per quanto concerne la lunghezza media delle frasi, gli universali traduttivi presentano tendenze in contraddizione reciproca: la maggiore brevità favorita dalla semplificazione si scontra con le aggiunte derivanti dalle eventuali esplicitazioni. A ciò si somma l'effetto dell'interferenza, per cui il traduttore raramente si discosterebbe dai confini sintattici del testo originale. Alla luce della preponderanza di testi originali angloamericani nel corpus studiato, ci si sarebbe po-

tuti attendere frasi più brevi nei testi tradotti, secondo lo stereotipo che vede l'inglese prediligere strutture sintattiche più snelle rispetto all'italiano. Come si evince da Tab. 3, ciò non avviene: la spiegazione potrebbe trovarsi nella particolarità dello stile giornalistico italiano, sia in senso intralinguistico (l'italiano dei giornali sarebbe caratterizzato da uno stile marcatamente franto – si giustifica così la prevalenza di frasi molto brevi, tra 1 e 7 parole) sia in senso interlinguistico (la stampa anglosassone di qualità adotta uno stile più sobrio rispetto ai giornali italiani).

La distribuzione del lessico tra *VdB* e non-*VdB* sembra confermare le conclusioni relative alla ricchezza lessicale (Tab. 2). Oltre che in riferimento agli universali traduttivi (tendenza alla normalizzazione), il fenomeno potrebbe trovare spiegazione nel fatto che i giornalisti italiani adottino anche dal punto di vista lessicale (come pare avvenire a livello sintattico) uno stile più brillante dei loro colleghi stranieri, selezionando risorse non comprese nel *Vocabolario di base*.

	<i>Unità tradotto</i>	<i>Corriere tradotto</i>	<i>Subtotale tradotto</i>	<i>Unità originale</i>	<i>Corriere originale</i>	<i>Subtotale originale</i>	<i>Totale</i>
lunghezza media articoli (in parole)	1.053,00	1.013,30	1.031,88	1.066,42	767,18	885,63	950,71
lunghezza media frasi (in parole) ⁵	24,37	22,75	23,49	22,11	24,67	23,37	23,43
da 1 a 7 parole	8,91%	12,51%	10,86%	16,99%	10,09%	13,57%	12,22%
da 8 a 20 parole	38,48%	39,58%	39,08%	39,44%	37,25%	38,36%	38,72%
oltre 20 parole	52,61%	47,91%	50,06%	43,57%	52,65%	48,07%	49,07%
<i>Vocabolario di base</i> ⁶	71,35%	71,13%	71,24%	71,22%	69,13%	70,17%	70,71%
fondamentale							
alto uso	8,28%	7,78%	8,03%	7,27%	7,7 %	7,48%	7,76%
alta disponibilità	10,45%	10,01%	10,23%	10,34%	10,04%	10,19%	10,21%
non-VdB	9,92%	11,08%	10,50%	11,17%	13,13%	12,15%	11,33%

Tabella 3: Alcune variabili linguistiche tese ad analizzare la complessità linguistica (sintattica e lessicale) del corpus

3. Analisi basate sulle liste di frequenza

3.1. L'espressione del pronome personale

L'esame delle liste di frequenza consente di isolare e quantificare alcuni aspetti particolarmente collegati ai fenomeni di interferenza della lingua di partenza sulla lingua di arrivo. Un primo aspetto che vale la pena di indagare è dato dall'espressione del pronome; con l'eccezione dello spagnolo, infatti, le lingue di partenza del subcorpus tradotto prevedono tutte l'esplicitazione obbligatoria del soggetto. A questo livello non "fine" dell'analisi, i dati maggiormente significativi riguardano i pronomi (Tab. 4) *io, tu, egli, ella, esso, essa, essi, esse*, per lo più utilizzati, appunto, come soggetto (ma – come notato in Rovere, 2005: 58-59 e 63-64 – sono possibili anche espressioni del tipo *chi per esso*), dato che *lui, lei, loro, noi* e *voi* sono utilizzati anche

⁵ La lunghezza media delle frasi è stata calcolata con il software *Èulogos autogulp* allegato a Piemontese (1996).

⁶ La percentuale di lemmi del *Vocabolario di base* nei diversi subcorpora è stata ricavata per campionamento con l'ausilio del software allegato a De Mauro (1997).

per esprimere l'oggetto diretto o indiretto ⁷. Si rileva una generale preponderanza dei pronomi personali in grado di ricoprire il ruolo di soggetto nei testi tradotti. Nel caso della prima persona singolare (oltre il doppio) l'evidenza del fenomeno può in parte essere spiegata in relazione alla centralità assegnata all'autore dell'articolo, selezionato appunto in virtù della fama dell'opinionista. Tra le forme della terza persona, la dominanza dei testi non tradotti della terza persona indicante soggetto non-umano (*esso* ed *essa*) potrebbe essere collegata al più ampio ricorso ai dimostrativi (Tab. 6) nei testi tradotti ai fini di una maggiore esplicitazione dei rimandi indesiderabili. Gli altri squilibri rilevati per le pronomi *tu* ed *esse* sono riconducibili a uno solo dei subcorpora (rispettivamente in "Unità originale" e in "Corriere tradotto").

	<i>Unità tradotto</i>	<i>Corriere tradotto</i>	<i>Subtotale tradotto</i>	<i>Unità originale</i>	<i>Corriere originale</i>	<i>Subtotale originale</i>	<i>Totale</i>
<i>io</i>	65	145	210	58	46	104	314
<i>tu</i>	7	6	13	18	8	26	39
<i>egli</i>	13	63	76	26	19	45	121
<i>ella</i>	2	0	2	0	1	1	3
<i>esso</i>	10	28	38	21	31	52	90
<i>essa</i>	23	41	64	54	32	86	150
<i>lui</i>	60	160	220	116	173	289	509
<i>lei</i>	46	106	152	67	62	129	281
<i>loro</i>	923	787	1.710	503	538	1.041	2.751
<i>noi</i>	246	180	426	247	159	406	832
<i>voi</i>	23	31	54	10	6	16	70
<i>essi</i>	29	52	81	36	23	59	140
<i>esse</i>	7	18	25	14	18	32	57
Totale	1.454	1.617	3.071	1.170	1.116	2.286	5.357
Densità per 1000 forme	5,41	5,48	5,45	4,35	3,78	4,05	4,76

Tabella 4: Frequenza assoluta dei pronomi personali nel corpus

3.2. Frequenza di pronomi e aggettivi possessivi

Un discorso analogo a quello dei pronomi soggetto riguarda i possessivi, in particolare gli aggettivi, utilizzati più frequentemente in inglese e francese rispetto all'italiano (che ricorre maggiormente all'articolo determinativo ed eventualmente alla forma pronominale del verbo). L'aumentata frequenza dei possessivi in italiano è uno dei primi tratti che i linguisti hanno ricondotto a influenze esogene (cfr. Morgana, 1994).

Anche se l'analisi condotta sul corpus non opera alcuna distinzione tra aggettivi e pronomi, tutti i possessivi risultano più frequenti nei testi tradotti (Tab. 5).

3.3. Frequenza dei pronomi e aggettivi dimostrativi

Anche per quanto riguarda la frequenza di pronomi e aggettivi (in questa fase iniziale della ricerca, non ancora distinti) dimostrativi, inglese e francese (ma anche tedesco), cioè le lingue di partenza più rappresentate nel subcorpus tradotto, fanno un uso più largo del dimostrativo (si pensi all'impiego di *ça* o *that/this* come riprese neutre, in italiano di norma sostituiti dal clitico *lo* – cfr. Burigo, 2007). Accanto all'interferenza linguistica, secondo gli universali traduttivi i

⁷ Per una discussione sulla possibile influenza dell'inglese sull'impiego di *loro* dativale – o finanche di *lui* – cfr. Cardinaletti (2004; 2005).

testi di arrivo dovrebbero risentire del maggiore sforzo di esplicitazione (che si realizza attraverso una maggiore frequenza di riprese anaforiche) compiuto dal traduttore. In effetti ciò pare confermato dalla costante preponderanza dei dimostrativi di prossimità (*questo* e varianti), utilizzabili – anche in luogo dei pronomi personali neutri (Tab. 4) – in senso endoforico.

	<i>Unità tradotto</i>	<i>Corriere tradotto</i>	<i>Subtotale tradotto</i>	<i>Unità originale</i>	<i>Corriere originale</i>	<i>Subtotale originale</i>	<i>Totale</i>
<i>mio</i>	115	172	287	48	48	96	383
<i>mia</i>	108	145	253	41	49	90	343
<i>miei</i>	39	78	117	20	22	42	159
<i>mie</i>	27	25	52	8	2	10	62
<i>tuo</i>	14	11	25	11	6	17	42
<i>tua</i>	14	7	21	8	7	15	36
<i>tuoi</i>	6	5	11	2	2	4	15
<i>tue</i>	8	2	10	4	2	6	16
<i>suo</i>	368	584	952	400	542	942	1.894
<i>sua</i>	379	590	969	413	476	889	1.858
<i>suoi</i>	246	329	575	221	244	465	1.040
<i>sue</i>	153	180	333	101	210	311	644
<i>nostro</i>	197	119	316	117	133	250	566
<i>nostra</i>	193	144	337	102	120	222	559
<i>nostri</i>	136	109	245	79	88	167	412
<i>nostre</i>	76	65	141	56	53	109	250
<i>voostro</i>	5	8	13	2	2	4	17
<i>vostra</i>	13	13	26	4	3	7	33
<i>voostri</i>	7	1	8	3	0	3	11
<i>vostre</i>	4	4	8	0	1	1	9
<i>loro</i>	923	787	1.710	503	538	1.041	2.751
Totale	3.031	3.378	6.409	2.143	2.548	4.691	11.100
Densità per 1000 forme	11,29	11,46	11,38	7,97	8,63	8,32	9,87

Tabella 5: Frequenza di pronomi e aggettivi possessivi nel corpus

	<i>Unità tradotto</i>	<i>Corriere tradotto</i>	<i>Subtotale tradotto</i>	<i>Unità originale</i>	<i>Corriere originale</i>	<i>Subtotale originale</i>	<i>Totale</i>
<i>questo</i>	728	901	1.629	554	556	1.110	2.739
<i>questa</i>	553	458	1.011	354	331	685	1.696
<i>questi</i>	239	234	473	160	207	367	840
<i>queste</i>	190	203	393	110	144	254	647
<i>quello</i>	253	314	567	250	356	606	1.173
<i>quella</i>	237	272	509	318	359	677	1.186
<i>quell'</i>	28	45	73	53	40	93	166
<i>quest'</i>	59	95	154	40	44	84	238
<i>quei</i>	59	61	120	77	49	126	246
<i>quelle</i>	97	101	198	129	146	275	473
<i>quelli</i>	111	116	227	128	125	253	480
<i>ciò</i>	180	216	396	222	127	349	745
Totale	2.734	3.016	5.750	2.395	2.484	4.879	10.629
Densità per 1000 forme	10,18	10,23	10,21	8,91	8,41	8,65	9,45

Tabella 6: Frequenza di pronomi e aggettivi dimostrativi nel corpus

4. Analisi realizzate su tagging grammaticale

4.1. Distribuzione delle categorie grammaticali

Per la lemmatizzazione ci si è avvalsi del software *TreeTagger*, lemmatizzatore automatico su base probabilistica. I dati così ottenuti hanno naturalmente un valore sommario, anche se, dal momento che l'errore è statisticamente distribuito, consentono confronti in termini relativi tra i due subcorpora di testi giornalisti italiani originali e tradotti.

A prescindere dalle precisazioni del caso ⁸, non emergono differenze significative nella distribuzione tra subcorpus tradotto e subcorpus originale. Ulteriori indagini potranno gettare luce sui mancati fenomeni di interferenza concernenti la frequenza di articoli e preposizioni (sovra-sfruttati in italiano rispetto all'inglese, lingua di partenza più utilizzata nel corpus in esame).

CAT	Numero di unità lessicali	Originale		Tradotto		Occorrenze totali
		N	%	N	%	
Parole piene						
Aggettivi	14.802	62.726	11,18	61.727	11,03	124.453
Nomi	27.432	163.449	29,14	160.312	28,65	323.761
Verbi	21.864	87.161	15,54	91.627	16,37	178.788
Parole vuote						
Avverbi	723	35.645	6,36	32.204	5,75	67.849
Congiunzioni	87	28.446	5,07	25.902	4,63	54.348
Articoli	32	47.392	8,45	49.557	8,86	96.949
Preposizioni	118	93.064	16,59	93.011	16,62	186.075
Pronomi	275	37.474	6,68	41.113	7,35	78.587
Altro	1.154	5.483	0,98	4.196	0,75	9.679
Totale	66.487	560.840	100,00	559.649	100,00	1.120.489

Tabella 7: Frequenza delle principali categorie grammaticali nel corpus lemmatizzato

4.2. Distribuzione dei modi e tempi verbali

L'esame del *tagset* completo di *TreeTagger*, così come rielaborato da *TALTAC2* secondo la procedura suggerita dagli autori del manuale in linea, consente di trarre alcune ulteriori informazioni. Si nota innanzitutto la notevole omogeneità tra il subcorpus tradotto e quello originale.

Oltre a confermare quanto osservato a proposito di uso dei dimostrativi (cfr. § 3.3.) e dei possessivi (cfr. § 3.2.), l'analisi consente di trarre alcune informazioni sulla morfologia verbale, anche se non si deve dimenticare il carattere probabilistico dei dati e il fatto che le forme composte (cioè formate da uno o più ausiliari e un participio passato) vengono analizzate come verbi distinti (basti pensare al passato prossimo, analizzato come una forma di presente e una forma di participio passato). La somma delle occorrenze di congiuntivo presente, congiuntivo imperfetto e condizionale mostra una leggera prevalenza nell'italiano tradotto di questi modi, tipici di una prosa più rispettosa dei dettami della grammatica standard, in linea con quanto ci si attende dal "traduttese" (tendenza alla "normalizzazione"); in quest'ottica si può anche interpretare la minima prevalenza del passato remoto (in quanto tempo verbale tipico dei testi narrativi scritti)

⁸ Per es., l'assegnazione degli avverbi alla classe delle parole grammaticali.

nei testi tradotti, anche se il dato andrebbe confrontato con la frequenza del perfetto composto. Tra l'altro, resta da valutare l'influsso del *past simple* inglese (che ancora una volta dovrebbe caratterizzare la gran parte dei testi nella lingua originale di partenza, favorendo nel traduttore la selezione del passato remoto).

		Unità				Occorrenze	
		lessicali	Originale		Tradotto		totali
		N	N	%	N	%	N
ABR	abbreviazione	13	869	0,15	609	0,11	1.478
ADJ	Aggettivo	14.802	62.726	11,18	61.727	11,03	12.4453
ADV	Avverbio	723	35.645	6,36	32.204	5,75	6.7849
CON	Congiunzione	87	28.446	5,07	25.902	4,63	5.4348
DET:def	articolo determinativo	19	34.100	6,08	35.949	6,42	70.049
DET:indef	articolo indeterminativo	13	13.292	2,37	13.608	2,43	26.900
INT	Interiezione	89	202	0,04	176	0,03	378
NOM	Sostantivo	26.216	151.256	26,97	149.700	26,75	300.956
NPR	nome proprio	1.216	12.193	2,17	10.612	1,90	22.805
NUM	numerales (agg. o pron.)	1.046	3.986	0,71	3.214	0,57	7.200
PRE	Preposizione	68	58.887	10,50	59.181	10,57	118.068
PRE:det	preposizione articolata	50	34.177	6,09	33.830	6,04	68.007
PRO	Pronome	15	88	0,02	114	0,02	202
PRO:demo	pronome dimostrativo	30	5.417	0,97	6.439	1,15	11.856
PRO:indef	pronome indefinito	51	3.162	0,56	3.642	0,65	6.804
PRO:inter	pronome interrogativo	18	501	0,09	557	0,10	1.058
PRO:pers	pronome personale	44	5.625	1,00	5.585	1,00	11.210
PRO:poss	pronome possessivo	54	4.524	0,81	6.199	1,11	10.723
PRO:refl	pronome riflessivo	14	4.083	0,73	3.933	0,70	8.016
PRO:rela	pronome relativo	49	14.074	2,51	14.644	2,62	28.718
SYM	Simbolo	6	426	0,08	197	0,04	623
VER:cimp	congiuntivo imperfetto	481	788	0,14	1.219	0,22	2.007
VER:cond	Condizionale	486	2.007	0,36	2.596	0,46	4.603
VER:cpre	congiuntivo presente	1.100	2.354	0,42	2.887	0,52	5.241
VER:futu	Futuro	1.005	2.224	0,40	2.389	0,43	4.613
VER:geru	Gerundio	1.397	2.267	0,40	2.273	0,41	4.540
VER:impe	Imperativo	520	495	0,09	418	0,07	913
VER:impf	Imperfetto	1.428	3.922	0,70	5.480	0,98	9.402
VER:infi	Infinito	4.316	17.379	3,10	19.145	3,42	36.524
VER:pper	participio perfetto	5.937	18.468	3,29	20.482	3,66	38.950
VER:ppre	participio presente	298	280	0,05	311	0,06	591
VER:pres	Presente	3.894	35.785	6,38	32.357	5,78	68.142
VER:refl:infi	riflessivo infinito	20	18	0,00	15	0,00	33
VER:remo	Passato remoto	982	1.174	0,21	2.055	0,37	3.229
Totale		66.487	560.840	100,00	559.649	100,00	1.120.489

Tabella 8: Frequenza delle categorie grammaticali di TreeTagger nel corpus

5. Prospettive di ricerca e conclusioni

L'analisi qui proposta intende rappresentare un primo saggio delle possibilità offerte ai linguisti da un approccio alla lingua delle traduzioni basato su metodologie quantitative. Se è indubbio che le misurazioni illustrate riguardano esclusivamente fenomeni molto generali, è pur vero che, incrociando i dati, sembra possibile caratterizzare l'italiano giornalistico tradotto sia senza tener

conto dell'interferenza della lingua di partenza (le percentuali relative al *Vocabolario di base*; la frequenza d'uso di congiuntivo e condizionale), sia considerando la forte dominanza di testi tradotti dall'inglese e dal francese (esplicitazione del pronome soggetto; frequenza dei possessivi), ma anche presupponendo una doppia causa (l'impiego dei dimostrativi può essere ascritto alla caratteristiche proprie delle lingue di partenza come pure alla tendenza all'esplicitazione).

Ampliando il corpus d'analisi con altre testate e sfruttando più a fondo le notevoli potenzialità di *TALTAC2* (ad esempio con la ricerca di combinazioni di elementi individuati in base alla categoria grammaticale, il confronto tra liste ecc.), si potranno individuare e misurare ulteriori possibili spie dell'interferenza linguistica (a titolo di esempio, senza distinguere tra le diverse lingue, sarà utile calcolare la frequenza di frasi scisse e topicalizzate, perifrasi progressive, concordanze *ad sensum*, anteposizioni del soggetto ai verbi inaccusativi e dell'aggettivo al sostantivo ecc.).

Sarà anche possibile ricercare eventuali divergenze nelle linee tendenziali individuate in base alla lingua del testo sorgente, ad esempio per studiare come l'italiano tradotto dall'inglese differisce rispetto a quello tradotto da altre lingue. Ulteriori indagini saranno possibili a più livelli e da prospettive diverse: per esempio, si potranno incrociare i dati relativi a singoli traduttori più rappresentati (quindi indagare lo "stile" del traduttore). Inoltre, in senso intralinguistico, poiché è presumibile che i tipi testuali considerati possano evidenziare peculiarità proprie in seno alla varietà di appartenenza, si può procedere a un confronto tra il corpus comprendente articoli scritti originariamente in italiano e l'italiano giornalistico *tout court* (*TALTAC2* dispone di una raccolta di 12 anni di «Repubblica»), come pure tentare valutazioni comparative in riferimento ad altre liste di frequenza e altri corpora disponibili per l'italiano scritto e parlato.

Riferimenti bibliografici

- Alfieri G. (1994). La lingua di consumo. In Serianni, L. and Trifone, P., editors, *Storia della lingua italiana*, vol. 2, *Scritto e parlato*, Torino: Einaudi, pp. 161-238.
- Baker M. (editor) (1998). *Routledge Encyclopedia of Translation Studies*. London-New York: Routledge.
- Baroni M. and Bernardini S. (2006). A New Approach to the Study of Translationese: Machine-learning the Difference between Original and Translated Texts. *Literary and Linguistic Computing*, vol. 21, n. 3: pp. 259-274.
- Bernardini S. and Baroni M. (2005). Spotting Translationese. A corpus-driven approach using support vector machines. In *Proceedings of Corpus Linguistics 2005*, <http://www.corpus.bham.ac.uk/PCLC/>.
- Bonomi I. (2002). *L'italiano giornalistico. Dall'inizio del '900 ai quotidiani on line*. Firenze: Franco Cesati editore.
- Bonomi I. (2003). La lingua dei quotidiani. In Bonomi, I., Masini, A. and Morgana, S., editors, *La lingua italiana e i mass media*, Roma: Carocci, pp. 127-164.
- Burigo S. (2007). *L'italiano tradotto dal francese: il caso di «Internazionale»*, prova finale del corso di Laurea in Traduzione e Interpretazione, Trieste, SSLMIT - Università degli Studi di Trieste.
- Cardinaletti A. (2004). La traduzione dei pronomi: interferenza sintattica e cambiamento linguistico. In Garzone e Cardinaletti, pp. 129-150.
- Cardinaletti A. (2005). La traduzione: un caso di attrito linguistico, in Cardinaletti e Garzone, pp. 59-84.
- Cardinaletti A. and Garzone G. (editors) (2005). *L'italiano delle traduzioni*. Milano: Franco Angeli.
- Dardano M. (1986). *Il linguaggio dei giornali italiani*. Roma-Bari: Laterza.

- Dardano M. (1994). Profilo dell'italiano contemporaneo. In Serianni, L. and Trifone, P., editors, *Storia della lingua italiana*, vol. 2, *Scritto e parlato*, Torino: Einaudi, pp. 343-432.
- De Mauro T. (1997). *Guida all'uso delle parole*. Roma: Editori riuniti (XII edizione).
- Garzone G. and Cardinaletti A. (editors) (2004). *Lingua, mediazione linguistica e interferenza*. Milano: Franco Angeli.
- Gualdo R. (2007). *L'italiano dei giornali*. Roma: Carocci.
- Korzen I. (2007). Mr. Bean e la linguistica testuale. Considerazioni tipologico-comparative sulle lingue romanze e germaniche. In Barbera, M., Corino, E. and Onesti, C., editors, *Corpora e linguistica in rete*, Perugia, Guerra edizioni, pp. 209-224.
- Mazzoleni M. and Baroni M. (2005). I toponimi stranieri nella stampa italiana: primi risultati di una ricerca sul corpus de la Repubblica. In *Atti di XXII ICOS: Congresso Internazionale di Scienze Onomastiche*.
- Morgana S. (1994). L'influsso francese. In Serianni, L. and Trifone, P., editors, *Storia della lingua italiana*, vol. 3, *Le altre lingue*, Torino: Einaudi, pp. 671-719.
- Musacchio M.T. (2005). The Influence of English on Italian: The Case of Translations of Economic Articles. In Anderman, G. and Rogers, M., editors, *In and Out of English: for Better, for Worse?*, Clevedon: Multilingual Matters, pp.71-96.
- Olohan M. (2004). *Introducing Corpora in Translation Studies*. London-New York: Routledge.
- Piemontese M.E. (1996). *Capire e farsi capire. Teorie e tecniche della scrittura controllata*. Napoli: Tecnodid.
- Rovere G. (2005). *Capitoli di linguistica giuridica. Ricerche su corpora elettronici*. Alessandria: Edizioni dell'Orso.
- Tuzzi A. (2003). *L'analisi del contenuto: introduzione ai metodi e alle tecniche di ricerca*. Roma: Carocci..