

L'ANALISI LINGUISTICA E LETTERARIA

FACOLTÀ DI SCIENZE LINGUISTICHE E LETTERATURE STRANIERE
UNIVERSITÀ CATTOLICA DEL SACRO CUORE

2

ANNO XXVI 2018

EDUCATT - UNIVERSITÀ CATTOLICA DEL SACRO CUORE

L'ANALISI
LINGUISTICA E LETTERARIA

FACOLTÀ DI SCIENZE LINGUISTICHE
E LETTERATURE STRANIERE

UNIVERSITÀ CATTOLICA DEL SACRO CUORE

2

ANNO XXVI 2018

PUBBLICAZIONE QUADRIMESTRALE

L'ANALISI LINGUISTICA E LETTERARIA
Facoltà di Scienze Linguistiche e Letterature Straniere
Università Cattolica del Sacro Cuore
Anno XXVI - 2/2018
ISSN 1122-1917
ISBN 978-88-9335-391-5

Comitato Editoriale

GIOVANNI GOBBER, Direttore
MARIA LUISA MAGGIONI, Direttore
LUCIA MOR, Direttore
MARISA VERNA, Direttore
SARAH BIGI
ELISA BOLCHI
ALESSANDRO GAMBA
GIULIA GRATA

Esperti internazionali

THOMAS AUSTENFELD, Université de Fribourg
MICHAEL D. AESCHLIMAN, Boston University, MA, USA
ELENA AGAZZI, Università degli Studi di Bergamo
STEFANO ARDUINI, Università degli Studi di Urbino
GYÖRGY DOMOKOS, Pázmány Péter Katolikus Egyetem
HANS DRUMBL, Libera Università di Bolzano
JACQUES DÜRRENMATT, Sorbonne Université
FRANÇOISE GAILLARD, Université de Paris VII
ARTUR GAŁKOWSKI, Uniwersytet Łódzki
LORETTA INNOCENTI, Università Ca' Foscari di Venezia
VINCENZO ORIOLES, Università degli Studi di Udine
GILLES PHILIPPE, Université de Lausanne
PETER PLATT, Barnard College, Columbia University, NY, USA
ANDREA ROCCI, Università della Svizzera italiana
EDDO RIGOTTI, Università degli Svizzera italiana
NIKOLA ROSSBACH, Universität Kassel
MICHAEL ROSSINGTON, Newcastle University, UK
GIUSEPPE SERTOLI, Università degli Studi di Genova
WILLIAM SHARPE, Barnard College, Columbia University, NY, USA
THOMAS TRAVISANO, Hartwick College, NY, USA
ANNA TORTI, Università degli Studi di Perugia
GISÈLE VANHESE, Università della Calabria

*I contributi di questa pubblicazione sono stati sottoposti
alla valutazione di due Peer Reviewers in forma rigorosamente anonima*

© 2018 EDUCatt - Ente per il Diritto allo Studio universitario dell'Università Cattolica
Largo Gemelli 1, 20123 Milano | tel. 02.7234.2235 | fax 02.80.53.215
e-mail: editoriale.dsu@educatt.it (*produzione*); librario.dsu@educatt.it (*distribuzione*)
web: www.educatt.it/libri

Redazione della Rivista: redazione.all@unicatt.it | web: www.analisinguisticaeletteraria.eu

Questo volume è stato stampato nel mese di settembre 2018
presso la Litografia Solari - Peschiera Borromeo (Milano)

INDICE

Metafora e ideologia in Hamlet: il discorso mercantilista <i>Renato Rizzoli</i>	5
Esotismo, antischiavismo, colonialismo: <i>Adventures in Borneo</i> di Catherine Gore <i>Luca Brezzo</i>	29
“La parola di un uomo onesto significa ciò che dice” Romano Guardini lettore di Rilke. <i>Lucia Mor</i>	45
La narrazione lecléziana e il pluralismo dei procedimenti espressivi <i>Marilena Genovese</i>	67
Come scrivono i politici italiani su Facebook Appunti per un’analisi linguistica comparativa <i>Yahis Martari</i>	81
La <i>corpus revolution</i> russa e il <i>corpus</i> parallelo italiano-russo: storia, criteri di compilazione e usi <i>Valentina Noseda</i>	115
<i>The Home of the Brave</i> . Sezione monografica a cura di G. Segato <i>Hard, Stoic, Isolate, and a Killer</i> . Appunti sul carattere americano <i>Giulio Segato</i>	133
<i>A Man of Honor</i> . Note sulle origini dell’eroe del romanzo poliziesco americano <i>Giulio Segato</i>	137
Eroismo e femminile, un binomio difficile: il caso di Margaret Fuller <i>Anna De Biasio</i>	145
“Una silenziosa litania operaia”. L’America di Carl Sandburg <i>Franco Lonati</i>	155
Recensioni	165

LA CORPUS REVOLUTION RUSSA E IL CORPUS PARALLELO ITALIANO-RUSSO

VALENTINA NOSEDA

Il presente contributo ha lo scopo di presentare il corpus parallelo italiano-russo, ampliato nel 2015 e ora disponibile all'interno del Corpus Nazionale della Lingua russa (*Nacional'nyj Korpus Russkogo Jazyka*). In particolare, nel saggio si descrivono la storia, i criteri di progettazione, le fasi di compilazione e le potenzialità offerte da questo strumento per gli studi linguistico-letterari della slavistica italiana.

This paper aims at presenting the Italian-Russian parallel corpus, expanded in 2015 and now available within the Russian National Corpus (*Nacional'nyj Korpus Russkogo Jazyka*). In particular, the paper describes the history, the design criteria, the phases of construction and the potential offered by this tool for language and literary studies in the field of Italian Slavistics.

Keywords: Corpus linguistics, parallel corpora, corpus design, contrastive linguistics, applied linguistics

1. Introduzione

La cosiddetta *corpus revolution*, segnalata all'attenzione della comunità scientifica all'inizio degli anni Novanta¹, rappresenta l'esito dello sviluppo tecnologico sugli studi linguistici, iniziato in Inghilterra già negli anni Sessanta. Proprio in quegli anni la riflessione metodologica sul modo di studiare una lingua ha determinato la nascita della Linguistica dei Corpora (LC), che definiremo qui come un approccio² adottato nell'analisi linguistica e basato sull'osservazione dei dati empirici estrapolati da uno o più corpora elettronici³.

Questo modo di trattare il dato linguistico deriva da una base teorica ben precisa, formulata in particolar modo da J.R. Firth⁴, M. Halliday e J. Sinclair, in parte in contrapposi-

¹ M. Rundell – P. Stock, *The corpus revolution*, "English Today", 8, 1992, 3, pp. 21-32.

² Vige tuttora un dibattito sullo status della LC, considerata da alcuni semplicemente come un metodo e da altri, in particolare dai linguisti legati alla corrente Neo-Firthiana, come una disciplina a se stante (G. Gilquin, *Corpus, Cognition and Causative Constructions*, John Benjamins, Amsterdam/Philadelphia 2010, pp. 5-6; T. McEnery – A. Hardy, *Corpus Linguistics. Method, Theory and Practice*, Cambridge University Press, New York 2012, pp. 1-147).

³ Un corpus, a sua volta, è definibile come un insieme di testi, selezionati secondo precisi criteri, raccolti su un portale e consultabili grazie a un software di interrogazione.

⁴ Si ricorda la celebre affermazione di Firth "you shall know a word by the company it keeps", con la quale il linguista voleva sottolineare l'importanza del contesto per comprendere il senso di una data parola (J.R. Firth, *Papers in linguistics 1934-1951*, Oxford University Press, London 1957, p. 11).

zione ad un'altra corrente sorta nello stesso periodo: il generativismo di Noam Chomsky⁵. Si può addirittura affermare che la LC si sia sviluppata e consolidata proprio grazie al dibattito con i generativisti, che basano le loro riflessioni e lo studio della lingua esclusivamente sull'intuizione dei parlanti nativi e sull'osservazione di esempi inventati⁶. La LC, al contrario, ha dimostrato, da un lato, che solo un'analisi supportata da esempi di lingua autentica può riflettere l'uso effettivo di una lingua, e, dall'altro, che l'intuizione dei parlanti spesso è insufficiente per individuare e comprendere a fondo tutti i fenomeni linguistici.

Riassumiamo dunque brevemente i principi su cui si fonda la LC e che hanno contribuito a ridefinire la descrizione linguistica tradizionale: i) la linguistica è una scienza sociale e applicata; ii) lo studio della lingua deve basarsi su esempi autentici; iii) il linguista dovrebbe considerare intere porzioni di testo inserite in un contesto specifico e non proposizioni, sintagmi o parole isolate; iv) è bene che i testi siano confrontati tra loro per mezzo di corpora compilati con criterio; v) la classica concezione dualistica tra forma e significato⁷ è superata a favore di una teoria linguistica che considera forma e significato inseparabili, e la LC è in grado di fare luce su questa connessione⁸.

A partire dagli anni Sessanta, vengono quindi creati i primi corpora elettronici, tra cui il *Brown Corpus* e il *LOB Corpus*, e poco più tardi, presso l'Università di Birmingham, in particolare grazie al contributo di John Sinclair, nascono i primi dizionari e le prime grammatiche *corpus-based*, quali la *Cobuild English Grammar*⁹ e il *Cobuild Dictionary of Idioms*¹⁰. Nei decenni successivi la LC si sviluppa considerevolmente arrivando a coprire un gran numero di lingue. In Italia si considerano gli anni Novanta come periodo di diffusione di questo approccio, in particolare con la compilazione del PIXI Corpus, un corpus di lingua parlata in italiano e inglese¹¹, mentre in Russia la *corpus revolution* approda all'inizio del nuovo millennio, con l'apertura del Corpus Nazionale della Lingua Russa (*Nacional'nyj Korpus Russkogo Jazyka* – NKRJa)¹². Negli ultimi 15 anni, grazie al rapido ampliamento e sviluppo del NKRJa, la LC russa ha recuperato alcune tappe importanti: sempre più dizio-

⁵ Cfr. N. Chomsky, *Syntactic Structures*, Mouton, The Hague 1957.

⁶ T. McEnery – A. Wilson, *Corpus Linguistics*, Edinburgh University Press, Edinburgh 2001.

⁷ Si veda la distinzione saussuriana tra *signifiant* e *signifié*, abbracciata anche da Chomsky, secondo il quale grammatica e significato sono indipendenti (J. Sinclair, *Trust the Text. Language, corpus and discourse*, Routledge, London/New York 2004, p. 2).

⁸ M. Stubbs, *Text and Corpus Analysis*, Blackwell, Oxford 1996, pp. 23-28; S. Johansson, *On the role of corpora in cross-linguistic research*, in *Corpora and Cross-linguistic Research: Theory, Method and Case Studies*, S. Johansson – S. Oksefjell ed., Rodopi, Amsterdam/Atalanta 1998, pp. 3-24.

⁹ J. Sinclair et al., *Collins Cobuild English Grammar*, HarperCollins, London 1990.

¹⁰ J. Sinclair et al., *Collins Cobuild Dictionary of Idioms*, HarperCollins, London 1995.

¹¹ M.T. Prat Zagrebelsky, *L'introduzione della corpus linguistics o linguistica dei corpora nelle università italiane: una ricostruzione 'personale' dagli anni '60 a oggi*, "Quaderni del CIRSIL", 4, 2005, www.lingue.unibo.it /cirsil (ultima consultazione gennaio 2017).

¹² Per un quadro più esaustivo sulle caratteristiche e le applicazioni del NKRJa, cfr. i contributi raccolti in *Nacional'nyj korpus russkogo jazyka: 2006-2008. Novye rezul'taty i perspektivy*, V.A. Plungjan ed., Nestor-Istorija, Sankt-Peterburg 2009.

nari vengono compilati sulla base del NKRJa¹³ e nasce la *Korpusnaja grammatika*, con lo scopo di descrivere parte della grammatica russa in base ai dati del NKRJa¹⁴. Infine, sebbene quest'ultimo rimanga lo strumento favorito dai russisti per le ricerche *corpus-based*, nuovi web-corpora¹⁵ della lingua russa hanno fatto gradualmente la loro comparsa: citiamo, ad esempio, *RuTenTen* o i corpora dell'*Aranea family*¹⁶.

In questo quadro, alla fine degli anni Novanta, comincia a diffondersi, inizialmente in area anglosassone, anche la cosiddetta *parallel corpus linguistics*¹⁷ che grazie all'uso dei corpora paralleli¹⁸ applica i principi sopraelencati all'analisi contrastiva tra le lingue. Sebbene la compilazione dei corpora paralleli risulti più lunga e difficoltosa, e nonostante lo scetticismo iniziale di alcuni studiosi, riluttanti ad inserire nei corpora testi tradotti¹⁹, è ormai innegabile che i corpora paralleli, seppur con qualche limite²⁰, svolgano una funzione fondamentale nella didattica di una L2, negli studi sulla traduzione umana e automatica e in numerosi ambiti della ricerca linguistica²¹, in chiave contrastiva e non solo. Anna Zaliznjak

¹³ Al link <http://dict.ruslang.ru/> (ultima consultazione: 19 gennaio 2018) sono disponibili i primi dizionari compilati interamente sulla base del NKRJa: E.A. Grišina – O.N. Ljaševskaja, *Grammatičeskij Slovar' novych slov russkogo jazyka*, 2008; O.N. Ljaševskaja – S.A. Šarov, *Novyj Častotnyj slovar' russkoj leksiki*, 2008; G.I. Kustova, *Slovar' russkoj idiomatiki. Sočetačija slov so značeniem vysokoj stepeni*, 2008; O.L. Burjuk – V.Ju. Gusev – E.Ju. Kalinina, *Slovar' glagol'noj sočetačivosti nepredmetnyh imen russkogo jazyka*, 2008.

¹⁴ La grammatica è consultabile interamente online al link <http://rusgram.ru/> (ultima consultazione: 19 gennaio 2018).

¹⁵ I web-corpora sono ingenti riserve elettroniche formate da testi estratti in maniera automatica da internet.

¹⁶ Cfr. V. Benko – V.P. Zacharov, *Very large Russian corpora: new opportunities and new challenges*, "Dialog", 15, 2016, pp. 83-98.

¹⁷ L. Borin, ... *And never the twain shall meet?*, in *Parallel corpora, parallel worlds. Selected papers from a symposium on parallel and comparable corpora at Uppsala University, Sweden, 22-23 April 1999*, L. Borin ed., Rodopi, Amsterdam/New York, pp. 1-43.

¹⁸ Raccolte di testi originali in una lingua (*Source Language*), allineati – generalmente a livello di frase – con le rispettive traduzioni in una o più lingue diverse (*Target Language*) (cfr. S. Gandin, *Linguistica dei corpora e traduzione: definizioni, criteri di compilazione e implicazioni di ricerca dei corpora paralleli*, "Annali della Facoltà. Università degli Studi di Sassari", 5, 2009, pp. 133-152).

¹⁹ Come accenneremo anche nel paragrafo 2, le traduzioni sono spesso influenzate dal testo fonte e sono caratterizzate da proprietà universali (universali traduttivi) che le rendono un tipo di testo a sé stante. Ciò ha fatto sì che molti linguisti escludessero inizialmente la possibilità di condurre analisi contrastive sulla base dei corpora paralleli (cfr. M. Olohan, *Introducing Corpora in Translation Studies*, Routledge, London/New York 2004, p. 13; R. von Waldenfels, *Polish tea is Czech coffee: advantages and pitfalls in using a parallel corpus in linguistic research*, in *Methods in contemporary Linguistics*, A. Ender – A. Leemann – B. Wälchli ed., Mouton/De Gruyter, Berlin 2012 ("Trends in Linguistics, 247"), pp. 263-281; F. Zanettin, *Translation-driven corpora*, Routledge, New York 2012, p. 12). Tuttavia, come spiegheremo nel secondo paragrafo, è possibile ovviare a questo problema disponendo di un corpus parallelo bi-direzionale.

²⁰ Oltre ai problemi sopra accennati, relativi all'influsso del testo fonte e agli universali della traduzione, R. von Waldenfels cita anche la questione della non equivalenza di significato nel passaggio da L1 a L2 e della libera interpretazione di cui ogni traduttore può avvalersi nel produrre un testo di arrivo: "translation is a creative process of language production like others. There are more than one way to say and conceptualize things" (R. von Waldenfels, *Polish tea is Czech coffee*, p.).

²¹ Si vedano, a questo proposito, M. Baker, *Corpus Linguistics and Translation Studies. Implications and Applications*, in *Text and technology. In honour of John Sinclair*, M. Baker – G. Francis – E. Tognini-Bonelli ed., John Benjamins, Philadelphia/Amsterdam 1993, pp. 233-250; S. Granger, *The corpus approach: a common way for-*

e Aleksej Šmelev²², infatti, dimostrano come i corpora paralleli possano essere impiegati, oltre che a scopi contrastivi, per individuare cioè le differenze tra due lingue, anche per approfondire la semantica di date unità linguospecifiche di una lingua di interesse²³.

In russistica i primi corpora paralleli compaiono nel 2005, quando all'interno del NKRJa viene creato il corpus parallelo russo-inglese, che al momento conta più di 20 milioni di parole. Con il tempo la sezione dei corpora paralleli si è ampliata notevolmente arrivando a contare 15 coppie di lingue²⁴, tra cui quella russo-italiano.

Nel contesto sopra accennato si inserisce dunque il presente saggio; avendo infatti partecipato, insieme ad altri, all'ampliamento del corpus parallelo italiano-russo (CP ita-ru) presente oggi nel NKRJa, vogliamo qui descriverne lo stato dell'arte e le prospettive future, illustrando i criteri e i procedimenti adottati per la sua compilazione. Evidenzieremo inoltre le possibilità di ricerca che questo innovativo strumento offre agli studi contrastivi linguistico-letterari della slavistica italiana.

Il CP ita-ru nasce come esperimento pilota pochi anni dopo il Corpus Nazionale della Lingua Russa (aperto alla pubblica consultazione nel 2004): in tutto cinque testi – tre originali italiani con la traduzione russa e due originali russi con la traduzione italiana²⁵ –, per un totale di sole 706.246 parole²⁶. Il corpus presentava evidenti lacune che lo rendevano inadatto ad essere uno strumento di ricerca sia per l'esiguo numero di parole e di testi, sia per il suo carattere non sistematico, ossia l'assenza di chiari e dichiarati criteri di progettazione e costruzione²⁷.

ward for contrastive Linguistics and Translation Studies?, in *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*, S. Granger – J. Lerot – S. Petch-Tyson ed., Rodopi, Amsterdam/New York 2003, pp. 17-29; S. Johansson, *Contrastive linguistics and corpora*, in *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*, S. Granger – J. Lerot – S. Petch-Tyson ed., Rodopi, Amsterdam/New York 2003, pp. 31-45.

²² A.A. Zaliznjak, *Lingvospecificičnye edinicy russkogo jazyka v svete kontrastivnogo analiza*, "Dialog", 14, 2015, 1, pp. 683-695; A.D. Šmelev – A.A. Zaliznjak, *Reversivnyj perevod kak instrument lingvističeskogo analiza*, "Dialog", 16, 2017, 2, pp. 394-406.

²³ Particolarmente efficace è il metodo della "traduzione inversa" impiegato dai due autori, che sfrutta le traduzioni russe di testi originali in inglese o francese per analizzare quali elementi della lingua di partenza stimolano l'uso della struttura linguistica indagata nella lingua d'arrivo (in questo caso il russo). Esempi di questo tipo si trovano anche in linguistica inglese. Cfr. ad esempio, D. Noël, *Translation as evidence for semantics: an illustration*, "Linguistics", 41, 2003, 4, pp. 757-785.

²⁴ Tutti i corpora paralleli del NKRJa sono disponibili a questo link: <http://ruscorpora.ru/search-para-en.html> (ultima consultazione: 21 gennaio 2018).

²⁵ M.A. Bulgakov, *Master i Margarita* (prima e seconda parte), 1929-1940; C. Collodi, *Pinocchio*, 1883; F.M. Dostoevskij, *Belye Noči*, 1848; U. Eco, *Il nome della rosa*, 1980; L. Pirandello, *La signora Frola e il signor Ponza, suo genero*, 1917.

²⁶ In linguistica dei corpora per numero di parole si intendono le occorrenze effettive (denominate *token*), che vanno distinte dai *types*, corrispondenti ai lessemi.

²⁷ N. Nesselhauf, *Corpus linguistics: a practical introduction*, 2011, versione elettronica disponibile al link <http://www.as.uniheidelberg.de/personen/Nesselhauf/files/Corpus%20Linguistics%20Practical%20Introduction.pdf> (ultima consultazione: 2 febbraio 2016).

La carenza di corpora paralleli russo-italiani²⁸ e la necessità di non procrastinare ulteriormente l'impiego delle tecnologie informatiche per l'analisi contrastiva di queste due lingue, per gli studi traduttivi e la didattica della lingua russa, ne imponevano dunque un ampliamento, iniziato nel gennaio 2014 grazie alla collaborazione fra l'Istituto di Lingua Russa di Mosca (*Institut Russkogo Jazyka imeni V.V. Vinogradova*), la cattedra di lingua e letteratura russa dell'Università Cattolica di Milano²⁹ e la Scuola Superiore di Lingue Moderne per Interpreti e Traduttori di Forlì (Università di Bologna). Si è così giunti, nel dicembre 2015, a un sostanziale incremento del corpus (ora disponibile al link: <http://ruscorpora.ru/search-para-it.html>), grazie al quale non solo è superata l'iniziale fase pilota, ma sono state messe le basi per assicurarne uno sviluppo sistematico³⁰.

2. Criteri di compilazione del corpus

Per progettare l'ampliamento del CP ita-ru abbiamo in primo luogo analizzato i più significativi corpora esistenti (paralleli e non) e preso in considerazione la letteratura scientifica sviluppatasi intorno alla loro compilazione³¹. La nostra attenzione si è concentrata sui seguenti punti: i) finalità e tipologia; ii) generi testuali; iii) dimensione del corpus; iv) lunghezza dei testi; v) periodo di redazione dei testi; vi) selezione delle traduzioni; vii) rappresentatività e bilanciamento; viii) sistematicità.

Abbiamo innanzitutto stabilito che il corpus sarebbe stato bi-direzionale, caratteristica fondamentale non solo per consentire un maggior numero di ricerche, ma anche per rendere il corpus comparabile, oltre che parallelo. La differenza tra questi due tipi di corpora sta

²⁸ Il portale Opus raccoglie alcuni testi allineati in russo e italiano, tuttavia non sono chiari i criteri di compilazione e non viene data alcuna informazione extralinguistica sui singoli testi inseriti (<http://opus.lingfil.uu.se/> ultima consultazione 12 gennaio 2018). Si parla in questo caso di collezioni di testi, per distinguerle dai corpora tradizionali, ovvero compilati secondo precisi criteri e con lo scopo di condurre ricerche linguistiche. I testi raccolti su portali come Opus hanno come fonte risorse multilingue specifiche e preesistenti, come ad esempio *Europarl*, che raccoglie i documenti prodotti in più lingue all'interno del Parlamento Europeo.

²⁹ Più precisamente, per quanto riguarda l'Università Cattolica, l'ampliamento è stato realizzato nell'ambito del Dottorato di Ricerca dell'autore di questo contributo, svoltosi presso l'Università Cattolica del Sacro Cuore di Milano (gennaio 2014-gennaio 2017) e sfociato nella dissertazione dottorale: V. Nosedà, *Corpora paralleli e linguistica contrastiva: ampliamento e applicazioni del corpus italiano russo nel Nacional'nyj Korpus Russkogo Jazyka*, A.A. 2015/16, Università Cattolica del Sacro Cuore, Milano 2017.

³⁰ Analogamente all'intero NKRJa, infatti, il nostro corpus parallelo non nasce come statico, ma verrà continuamente ampliato e monitorato.

³¹ In particolare si è scelto di prendere in esame i seguenti corpora: tre corpora della lingua inglese, da sempre considerati come esempi validissimi in termini di progettazione: il Brown Corpus (BrC), il British National Corpus (BNC), il Corpus of Contemporary American English (COCA); uno dei maggiori corpora di lingua italiana CORIS/CODIS; e due corpora paralleli, ovvero l'English-Norwegian Parallel Corpus (ENPC), considerato uno dei migliori corpora paralleli esistenti per il rigore e la sistematicità dei criteri di compilazione, e CEXI (the English Italian Translational Corpus), in quanto unico corpus parallelo per la lingua italiana che rispondesse ai criteri di sistematicità e bilanciamento, e sebbene non sia mai stato portato a termine, la descrizione dei criteri in base a cui avrebbe dovuto essere compilato si è rivelata molto utile per definire le caratteristiche del nostro corpus (cfr. V. Nosedà, *Corpora paralleli e linguistica contrastiva*, pp. 50-51).

nella relazione che intercorre tra i testi che li compongono: un corpus parallelo raccoglie testi in una lingua 1 allineati con la rispettiva traduzione in lingua 2; un corpus comparabile è invece formato da testi dello stesso genere ma in due lingue diverse, senza che vi sia tra di essi un rapporto traduttivo, oppure può essere composto da traduzioni e originali redatti nella stessa lingua. Quest'ultimo tipo è particolarmente utile sia negli studi traduttivi (in particolare per le ricerche sui cosiddetti universali traduttivi)³², sia, come spiega Johansson³³, per verificare se le tendenze linguistiche riscontrate nei testi tradotti siano attribuibili solo alle traduzioni o anche alla maggior parte dei testi prodotti in una data lingua. Decidendo di inserire sia originali italiani sia originali russi appartenenti a stessi tipi testuali (romanzi, memorie, saggi) è dunque possibile usare il CP ita-ru per confrontare non solo gli originali in russo o italiano con le rispettive traduzioni, ma anche testi russi originali e traduzioni russe, come pure testi originali italiani e traduzioni italiane.

Abbiamo infine deciso di creare quello che, secondo la terminologia di Merkel, viene definito “target-variant corpus”³⁴: un corpus parallelo che contenga, ove presenti, diverse traduzioni di uno stesso testo. La possibilità di disporre di diverse traduzioni di una stessa opera è infatti estremamente utile in diversi ambiti di ricerca: nella linguistica contrastiva, per analizzare in modo più scrupoloso le caratteristiche e gli usi di una data struttura linguistica; negli studi traduttivi, per condurre analisi comparate delle diverse traduzioni; nella ricerca letteraria e negli studi sulla storia della traduzione, per approfondire la recezione di un autore o di una sua opera³⁵.

Trattandosi di un corpus contenente testi originali con le rispettive versioni in una lingua 2, la scelta tra i vari tipi testuali è più limitata rispetto a un corpus monolingue, dal momento che non tutti i generi vengono tradotti: da questo punto di vista, in un CP le opere letterarie sono il genere testuale dominante, seguite dalla saggistica³⁶. Alla luce di quanto

³² L'idea che i testi tradotti subiscano l'interferenza del testo fonte è di Toury (G. Toury, *Descriptive Translation Studies and beyond*, John Benjamins, Amsterdam/Philadelphia 1995, pp. 267-279; si pensi in particolar modo alle due leggi – *growing standardization* (standardizzazione crescente) e *interference* (interferenza) – enunciate dallo studioso), ma è Baker (M. Baker, *Corpus-based Translation Studies: the Challenges that lie ahead*, in *Terminology, LSP and Translation. Studies in language engineering in honour of Juan C. Sager*, H. Somers ed., John Benjamins, Amsterdam/Philadelphia 1996, pp. 175-186) a precisare questo concetto, rendendolo universale mediante l'affermazione che, indipendentemente dalla lingua di partenza e da quella di arrivo, esistono fenomeni caratteristici del processo traduttivo di per sé, denominati, appunto, universali della traduzione (o universali traduttivi), ossia: *simplification* (semplificazione), *explicitation*, (esplicitazione), *normalization* (normalizzazione) e *levelling out* (convergenza). Le ricerche sugli universali caratterizzano gli studi traduttivi descrittivi (*descriptive translation studies*), che confrontano testi originali in una data lingua A con testi tradotti nella stessa lingua A senza prendere in considerazione né i testi fonte, né la lingua in cui questi ultimi sono redatti.

³³ S. Johansson, *On the role of corpora in cross-linguistic research*, e., l. 1998.

³⁴ Cfr. L. Borin, ... *And never the twain shall meet?*

³⁵ Cfr. Biber sui “*corpus stylistic studies*”, ovvero le indagini corpus-based che analizzano lo stile di un autore (D. Biber, *Corpus linguistics and the study of literature. Back to the future?*, “Scientific Study of Literature”, 1, 2011, 1, pp. 15-23).

³⁶ È stata presa in considerazione anche l'idea di inserire in un secondo momento articoli di giornale, ma il progetto è ancora al vaglio. Gli articoli pubblicati da alcuni giornali online sono infatti spesso traduzioni di pezzi apparsi su testate estere, e il loro inserimento in un corpus parallelo è problematico per almeno due motivi: 1) le traduzioni vengono nella maggior parte dei casi rivisitate o accorciate dalla redazione prima di essere pubbli-

detto, per differenziare il più possibile i testi del CP ita-ru, si è deciso, per la sezione di saggistica, di scegliere brani che trattassero argomenti diversi, utilizzando la classificazione degli argomenti proposta dal motore di ricerca dell'*Index Translationum*³⁷.

Per quanto riguarda le dimensioni, abbiamo stabilito come obiettivo preliminare di raggiungere almeno 3 milioni di parole, pari all'entità dell'*English-Norwegian Parallel Corpus*, adottato come modello nella progettazione del nostro come di diversi altri corpora paralleli, per via della sistematicità e del bilanciamento che lo caratterizzano³⁸. Il criterio della dimensione è influenzato chiaramente dalla lunghezza dei testi inseriti nel CP, che possono essere in versione integrale o ridotta. Per non ledere il copyright e non allungare oltremodo i tempi necessari per il lavoro di allineamento³⁹ dell'originale con la sua traduzione, abbiamo deciso di inserire per ogni titolo un estratto che andasse dalle 10.000 alle 15.000 parole⁴⁰, partendo dall'inizio di ogni opera, ma non si esclude in futuro la possibilità di aggiungere estratti ulteriori di uno stesso testo.

Per consentire studi di tipo diacronico non abbiamo limitato la scelta dei testi ad un solo periodo, come nel caso di alcuni corpora sincronici quali il BrC o il BNC⁴¹: tenendo in considerazione l'evoluzione della lingua russa, sono state dunque utilizzate opere redatte a partire dal XVIII secolo, e pertanto anche i testi italiani, con la loro traduzione russa, non sono anteriori a questo periodo.

Un criterio importante nella progettazione di un corpus parallelo è la scelta delle traduzioni: abbiamo stabilito che i traduttori dei testi selezionati dovessero essere noti (ovvero non anonimi) e parlanti nativi della lingua del testo di arrivo. Nel caso di testi tradotti più volte da diversi traduttori, siamo partiti dalle traduzioni considerate più autorevoli, senza escludere che altre possano essere inserite in un secondo momento (cfr. par. 3.1 sulla scelta delle traduzioni de *Il Cappotto* di N. Gogol').

Tutte le decisioni sopracitate si ricollegano ai criteri della rappresentatività e del bilanciamento, due concetti tanto importanti quanto difficili da realizzare nella compilazione di un corpus. Per rappresentatività Leech intende la possibilità di estendere ad una lingua le caratteristiche che si riscontrano in un corpus di testi redatti in quella stessa lingua⁴². Per

cate, risultando quindi poco fedeli al testo fonte, ed è difficile reperire i testi originali e le traduzioni prima della manipolazione. 2) Rimangono ignote le identità degli autori degli articoli originali e dei traduttori.

³⁷ Un database dell'UNESCO che registra in maniera quasi completa i libri tradotti dal 1979 al 2012 e suddivide i titoli nelle seguenti categorie: arte-giochi-sport, bibliografie, filosofia-psicologia, letteratura, giurisprudenza-scienze sociali-istruzione, scienze naturali-scienze esatte, scienze applicate, storia-geografie-biografie. [<http://www.unesco.org/xtrans/> (ultima consultazione: settembre 2017)].

³⁸ Zanettin, ad esempio, lo identifica come modello per la progettazione del corpus CEXI. Cfr. F. Zanettin, *CEXI. Designing an English-Italian Translational Corpus*, "Language and Computers", 42, 2002, 2, pp. 329-343.

³⁹ Sul processo di allineamento di un testo si veda il paragrafo 3.2.

⁴⁰ L'allineamento di testi integrali, infatti, oltre a creare più problemi per quanto riguarda i diritti d'autore, richiedeva una mole di lavoro ingestibile nel periodo di tempo stabilito per il primo ampliamento del CP ita-ru (tre anni).

⁴¹ I due corpora raccolgono testi risalenti, rispettivamente, agli anni Sessanta e Novanta.

⁴² G. Leech, *The state of the art in corpus linguistics*, in *English corpus linguistics: Studies in Honour of Jan Svartvik*, K. Aijmer – B. Altenberg ed., Longman, London/New York 1991, pp. 8-29.

essere rappresentativo un corpus dovrà pertanto non solo essere vasto, ma anche bilanciato e variegato. ovvero contenere, in ugual misura, testi appartenenti a vari generi e riguardanti temi diversi. Inoltre, nel caso di corpora paralleli bi-direzionali, la quantità dei testi in lingua 1 non dovrebbe discostarsi eccessivamente da quella dei testi in lingua 2. Da ultimo, per evitare di inserire un gran numero di testi appartenenti allo stesso stile⁴³ – il che comprometterebbe i risultati dell'analisi linguistica – abbiamo deciso di inserire inizialmente una sola opera per autore, facendo in modo che anche i traduttori variassero il più possibile.

Infine, per rispondere al criterio della sistematicità, abbiamo annotato (e continueremo ad annotare) ogni fase del nostro lavoro, così da rendere le informazioni accessibili ai futuri utenti. Riteniamo, infatti, che sia estremamente importante per un ricercatore conoscere le caratteristiche dei testi contenuti nel corpus che sta utilizzando, poiché informazioni quali autore, anno di redazione, titolo dell'opera, possono rivelarsi decisive per interpretare i dati dell'analisi.

3. *Le fasi della compilazione*

3.1. La selezione dei testi

Una volta stabiliti i criteri per la creazione del nostro corpus sono stati scelti i testi. Per dare continuità al corpus pilota già presente sul NKRJa è stato deciso di partire dall'ampliamento della sezione di prosa letteraria. A tale scopo ci siamo affidati ai database delle maggiori biblioteche nazionali⁴⁴, all'*Index Translationum* e ai cataloghi delle case editrici, così da stabilire, in primo luogo, i testi tradotti in una e nell'altra direzione e, in seguito, quante traduzioni fossero state pubblicate per ciascuna opera. Esiste una notevole discrepanza tra il numero di versioni italiane di una stessa opera russa e il numero di versioni russe di una stessa opera italiana (nettamente più esiguo), tuttavia, tale sbilanciamento caratterizza solo i classici dell'Ottocento e dei primi del Novecento; per la maggior parte delle opere più recenti, in molti casi meno popolari, è stata pubblicata una sola traduzione, sia che si tratti di originali russi, sia che si tratti di originali in italiano. Pertanto, durante questa fase, è stato necessario distinguere tra opere russe dei secoli XIX e XX, tradotte più volte in italiano (che chiameremo gruppo A), e opere russe più recenti, dei secoli XX e XXI, tradotte una sola volta (gruppo B)⁴⁵: per il gruppo B ci si è basati su quanto effettivamente presente nei cataloghi sopraindicati, mentre per il gruppo A abbiamo scelto innanzitutto gli autori più noti e influenti nel panorama letterario del Paese di arrivo. Per ognuno di essi è stato poi necessario scegliere l'opera da inserire, seguendo i criteri della recezione e della popolarità, solitamente testimoniate dal numero di traduzioni pubblicate in lingua 2⁴⁶.

⁴³ Dove per 'stile' qui si intende quello proprio di un autore.

⁴⁴ <http://opac.sbn.it/opacsbn/opac/iccu/free.jsp><http://opac.sbn.it/opacsbn/opac/iccu/free.jsp>, sito che raccoglie i cataloghi di tutte le biblioteche d'Italia (ultima consultazione: agosto 2017) e il catalogo della Biblioteca Statale Russa di Mosca <http://www.rsl.ru/ru> (ultima consultazione: agosto 2017).

⁴⁵ Si precisa che questa distinzione non emerge sul Corpus, ma ha riguardato solamente la fase di selezione.

⁴⁶ Ad esempio, una volta scelto Tolstoj come autore da inserire necessariamente nel nostro corpus, è stato selezionato il romanzo *Anna Karenina*, in quanto tradotto un maggior numero di volte rispetto a *Guerra e Pace*

Per quanto riguarda il gruppo A, abbiamo acquisito anche alcune opere allineate in modo autonomo dal nostro progetto⁴⁷, che hanno permesso di raggiungere in breve tempo le dimensioni da noi prestabilite per il corpus. Tuttavia, alcuni dei criteri di progettazione da noi previsti (cfr. paragrafo 2) non sono stati rispettati. In particolare, in questo sotto-corpus⁴⁸:

- sono state allineate opere integrali, anziché estratti;
- l’acquisizione dei testi è avvenuta in un’unica direzione – russo-italiano – compromettendo il bilanciamento auspicato tra lingua 1 e lingua 2;
- Il criterio usato per selezionare i testi e le loro traduzioni non ha seguito parametri qualitativi ma ragioni di pura praticità, ossia la facile reperibilità. Di conseguenza per alcuni autori sono state allineate fin da subito più opere (perché disponibili nel web), e dove erano presenti più traduzioni, sono state semplicemente scelte quelle disponibili in formato elettronico.

Pertanto, durante la prima fase di ampliamento del CP la riflessione sui criteri di selezione delle traduzioni ha riguardato solo il racconto di Gogol’ Šinel⁴⁹ (tradotto il per lo più come *Il Cappotto*⁵⁰), su cui è necessario soffermarsi brevemente. Le prime traduzioni italiane de *Il Cappotto*⁵¹ sono state spesso considerate dalla critica poco valide e infedeli, poiché presumibilmente tradotte dal francese anziché direttamente dal russo, come precisato da Inkova⁵² e Pala⁵³. La prima versione che abbiamo dunque voluto inserire è quella del 1922 di C. Rebora, unica nel suo genere grazie “all’impronta personale che hanno tutte le traduzioni del Rebora”⁵⁴. La seconda traduzione selezionata, in ordine cronologico, è quella di E. Carafa D’Andria (1937): nonostante alcune scelte traduttive opinabili e oggi obsolete, la popolarità della duchessa D’Andria negli ambienti editoriali della sua epoca (testimoniata anche da Lo Gatto) ci ha indotto ad inserire anche la sua versione intitolata inizialmente *Il Pastrano* e ripubblicata successivamente come *Il Cappotto*. Per quanto riguarda le traduzioni successive, abbiamo scelto innanzitutto quelle svolte da traduttori noti nel mondo della slavistica italiana: T. Landolfi (1941), L. Pacini-Savoj (1957), G. Pacini

o ad altre opere. Si precisa che le informazioni relative al numero di traduzioni sono state reperite grazie agli strumenti precedentemente citati (database delle biblioteche e *Index Translationum*), i quali possono presentare delle lacune. Ad esempio, nel caso del Cappotto di N. Gogol’, un approfondimento (V. Nosedà, *Corpora Paralleli e linguistica contrastiva*, p. 202) evidenzia che N. Marcialis, nei commenti alla sua traduzione del racconto, cita versioni non menzionate da altri e nemmeno presenti nei database consultati.

⁴⁷ In particolare si tratta dei testi elaborati da Ksenja Balakina dell’Università di Bologna.

⁴⁸ Nel paragrafo 4 presenteremo l’elenco di queste opere e una nostra proposta per il futuro ampliamento del corpus che tenga conto di questo intervento.

⁴⁹ Inserito dall’unità di Milano.

⁵⁰ Vi sono traduttori che hanno proposto titoli alternativi: *Il mantello* (C. Alvaro e T. Landolfi), *La mantella* (N. Marcialis e S. Prina), *Il pastrano* (E. Carafa D’Andria) e *l’uniforme* (D. Ciampoli).

⁵¹ Vale a dire le traduzioni di G. Loschi (1903 e 1919), D. Ciampoli (1916) e C. Alvaro (1920).

⁵² O. Inkova, *Le prime traduzioni romanze del “Cappotto” di Gogol’: precisazioni cronologiche e traduttologiche*, “Diversité et identité culturelle en Europe”, 13, 2016, 2, pp. 93-120.

⁵³ V. Pala, *Tommaso Landolfi traduttore di Gogol’*, Bulzoni Editore, Roma 2009.

⁵⁴ E. Lo Gatto, Recensione a N.V. Gogol’, *Il cappotto*, traduzione di C. Rebora, “L’Italia che scrive”, 5, 1922, 8, p. 151.

(1963), P. Zveteremich (1967), E. Bazzarelli (1980), N. Marcialis (1991)⁵⁵, S. Prina (1994), E. Guercetti (1995). Le restanti traduzioni, seppur di traduttori meno noti, sono state inserite per rispondere al criterio della rappresentatività. Riteniamo infatti che una lingua storico-naturale sia formata da tutti i testi prodotti in quella lingua, anche se ritenuti meno validi dalla critica e dai lettori. Trattandosi poi di traduzioni, abbiamo voluto in parte raccogliere quanto affermato nei *Descriptive Translation Studies*⁵⁶, secondo cui l'analisi dei testi tradotti non dovrebbe tener conto del testo di partenza, superando così definizioni quali 'fedele' o 'infedele'.

Alle opere letterarie è seguita la selezione della saggistica, al momento ancora in fase di allineamento. Come accennato, sono state scelte opere su vari argomenti, secondo la classificazione tematica e per generi proposta dall'*Index Translationum*⁵⁷, e per mantenere il bilanciamento abbiamo scelto non più di uno o due testi per categoria.

3.2. Acquisizione e allineamento dei testi

Dopo la conversione dei testi in formato digitale (quando non già disponibile) e la loro revisione⁵⁸, una volta ottenuti dalle case editrici i permessi per utilizzarli⁵⁹ (momento problematico per la compilazione di qualsivoglia corpus elettronico), tramite il software Euclid è stato eseguito l'allineamento degli originali alle loro traduzioni. Ciò avviene in modo automatico, ma richiede tuttavia un intervento di correzione manuale nel frequente caso in cui vi siano differenze di punteggiatura tra testo fonte e testo di arrivo. Euclid suddivide infatti i due testi in periodi delimitati da un punto fermo: se in traduzione quest'ultimo è sostituito da un

⁵⁵ Per un problema tecnico in fase di allineamento, la versione di Marcialis non è comparsa sul CP a dicembre 2015.

⁵⁶ Cfr. G. Toury, *Descriptive Translation Studies*.

⁵⁷ Fanno eccezione alcuni casi in cui abbiamo denominato le categorie in modo diverso, per esempio nel caso degli epistolari e delle memorie.

⁵⁸ Il programma di allineamento legge esclusivamente file in formato txt. Se si utilizza l'OCR (riconoscimento ottico dei caratteri), che permette di convertire direttamente l'immagine scansionata in un file di testo modificabile, il passaggio al formato txt è possibile, ma un controllo degli errori sarà anche in questo caso necessario.

⁵⁹ L'utilizzo dei testi sarebbe regolato dalla legge sul copyright, valida in tutti i casi in cui un autore non sia deceduto da più di 70 anni. La richiesta alle case editrici che detengono i diritti il più delle volte non è rifiutata, bensì ignorata. Per ovviare a questo problema Zanettin suggerisce di accompagnare alla richiesta una dichiarazione in cui si afferma che il consenso sarà dato per scontato in caso di mancata risposta entro un certo periodo (F. Zanettin, *Translation-Driven Corpora: Corpus Resources for Descriptive and Applied Translation Studies*, Routledge, New York 2012, pp. 52-55). Altri compilatori di corpora si appellano invece alla legge sul *fair use* (uso legittimo) (cfr., ad esempio, M. Davies, *The 385 + million-word Corpus of Contemporary American English (1990-2008+)*. *Design, architecture and linguistic insights*, "International Journal of Corpus Linguistics", 14, 2009, 2, pp. 159-190), ritenendo che l'utilizzo dei testi in un corpus sia legittimo per almeno due motivi: innanzitutto ha come scopo esclusivamente la ricerca e, in secondo luogo, i testi presenti in un corpus non sono mai visualizzabili per intero (nel NKRJa è possibile visualizzare contemporaneamente un massimo di 500 caratteri). In Italia le questioni legate alla richiesta dei permessi potrebbero essere risolte grazie all'articolo 70 della legge n. 633 del 22 aprile 1941, comma 1, secondo il quale: "Il riassunto, la citazione o la riproduzione di brani o di parti di opera e la loro comunicazione al pubblico sono liberi se effettuati per uso di critica o di discussione, nei limiti giustificati da tali fini e purché non costituiscano concorrenza all'utilizzazione economica dell'opera; se effettuati a fini di insegnamento o di ricerca scientifica l'utilizzo deve inoltre avvenire per finalità illustrative e per fini non commerciali".

segno di interpunzione differente, il programma non è in grado di allineare correttamente le due parti⁶⁰.

Abbiamo infine provveduto alla compilazione dei metadati, necessari per l'annotazione meta-testuale del corpus, grazie alla quale è possibile conoscere il titolo di un'opera, l'autore, l'anno di composizione, il traduttore, il genere e la lunghezza dell'estratto (espressa in numero di parole). I programmatori del NKRJa hanno infine curato l'annotazione (cfr. paragrafo 4d) eseguita grazie al supporto di Yandex⁶¹.

4. Stato dell'arte, applicazioni e proposte per un futuro ampliamento

In questo paragrafo daremo alcune informazioni relative ai testi inseriti nel CP ita-ru e alle potenzialità di ricerca di questo strumento. Le nostre riflessioni concerneranno alcuni dei criteri di progettazione considerati nel paragrafo 2. Come accennato, non tutte le decisioni prese in fase di progettazione sono già state realizzate, mentre l'aggiunta di testi da parte di collaboratori esterni ha reso necessari alcuni interventi correttivi che illustreremo.

a) Testi presenti sul corpus

Non ci sembra inutile, per completezza d'informazione, riportare in ordine cronologico i titoli delle opere letterarie rappresentate nel CP ita-ru dopo il primo sostanziale ampliamento (fino a dicembre 2017). Il primo elenco contiene, in ordine cronologico di composizione, le opere del gruppo A, ossia quelle che presentano più di una traduzione italiana, e in particolare: i) le opere inserite integralmente; ii) i due titoli già compresi nel corpus pilota⁶²; iii) 15 versioni italiane del *Cappotto* di Gogol⁶³. A queste seguiranno, in ordine cronologico, anche le opere del gruppo B, ovvero tradotte solo una volta in italiano, attualmente presenti nel corpus. Da ultimo i) l'elenco delle opere letterarie italiane tradotte in russo; ii) le opere italiane che formavano il corpus pilota, inserite in versione integrale; iii) alcuni racconti di D. Buzzati aggiunti dall'unità di Forlì⁶⁴.

Opere ru-ita del gruppo A disponibili nel NKRJa (dicembre 2017):

⁶⁰ Gli errori di allineamento possono essere legati allo stile di autori e traduttori, oppure al tipo sintattico delle due lingue considerate, più o meno ipotattiche. Su questo tema cfr. D.O. Dobrovolskij – A.A. Kretov – C.A. Šarov, *Korpus parallel'nyh tekstov. Arhitektura vozmožnosti ispol'zovanija*, in *Nacional'nyj korpus russkogo jazyka: 2003-2005*, Indrik, Moskva 2005, pp. 263-296.

⁶¹ Una delle maggiori società informatiche russe.

⁶² Allineate e inserite integralmente.

⁶³ Quest'opera gogoliana è stata in assoluto in Italia la più tradotta dal russo. Secondo un recente studio di O. Inkova, si contano almeno 32 versioni, sebbene Inkova riporti due versioni di E. Carafa d'Andria, una del 1937 e una del 1960 (che corrisponde alla versione da noi acquisita), ma confrontando parte dei due testi si evince che a cambiare è solo il titolo: *Il Pastrano* del '37 e *Il Cappotto* del '60 (O. Inkova, *Tradurre il titolo: le traduzioni italiane del "Cappotto" di Gogol*, "Kwartalnik neofilologiczny", 61, 2014, 1, pp. 41-56). Inoltre, come già accennato, N. Marcialis menziona nelle sue note due ulteriori versioni: una di Nerina Martini Bernardi Carrescia (1975) e un'altra di Anjuta Gančikov (1981).

⁶⁴ Questi testi sono in versione integrale, ma non raggiungono le 10.000 parole.

i)

- N.V. Gogol', *Mertvyje duši* (1835-1852) / *Le anime morte* (P. Nori, 2009)
- I. Gončarov, *Oblomov* (1859) / *Oblomov* (A. Michettoni, 2000)
- I.S. Turgenev, *Otcy i deti* (1862) / *Padri e figli* (F. Verdinois, 1908)
- F.M. Dostoevskij, *Prestuplenie i nakazanie* (1866) / *Delitto e Castigo* (G. Kraiski, 1983)
- L.N. Tolstoj, *Anna Karenina* (1875-1877) / *Anna Karenina* (M. B. Luporini, 1961)
- F.M. Dostoevskij, *Velikij inkvizitor* (1879) / *Il grande inquisitore* (S. Vitale, 2010)
- A.P. Čechov, *Rasskazy* (1885) / *Racconti* (F. Malcovati, 2004)
- A.P. Čechov, *Palata N. 6* (1892) / *La corsia N. 6* (F. Malcovati, 2004)
- M. Gor'kij, *Byvsie ljudi* (1897-1898) / *Gli ex-uomini* (E. W. Foulques, 1905)
- M. Gor'kij, *Ozornik* (1897-1898) / *Il burlone* (E. W. Foulques, 1905)
- M. Gor'kij, *Suprugi Orlovy* (1897-1898) / *I coniugi Orlof* (E. W. Foulques, 1905)
- M. Gor'kij, *Toska* (1897-1898) / *L'angoscia* (E. W. Foulques, 1905)
- B.L. Pasternak, *Doktor Živago* (1945-1955) / *Il dottor Živago* (P. Zveteremich, 1957)
- V.V. Nabokov, *Lolita* (1967⁶⁵) / *Lolita* (G. Arborio Mella, 1993)
- V.T. Šalamov, *Kolymskie rasskazy* (1973) / *I racconti di Kolyma* (M. Binni, 1999)⁶⁶

ii)

- F.M. Dostoevskij, *Belye noči* (1848) / *Le notti bianche* (P. Biloni, 2006)
- M.A. Bulgakov, *Master i Margarita* (1929-1940) / *Il Maestro e Margherita* (V. Drisdo 1967)

iii)

- N.V. Gogol', *Šinel' 1842* / *Il Cappotto*⁶⁷ [C Rebora (1922) / *Il Pastrano*. Duchessa D'Andria (1937) / *Il Mantello*. T. Landolfi (1941) / O. Del Buono (1949) / L. Pacini-Savoj (1957) / G. Pacini (1963) / A Julovic (1964) / P. Zveteremich (1967) / E. Bazzarelli (1980) / S. Beffa (1986) / F. Mariano (1986) / L. De Nardis (1993) / *La Mantella*. S. Prina (1994) / E. Guercetti (1995) / F. Legittimo (2001)]⁶⁸.

Opere ru-ita del gruppo B disponibili nel NKRJa (dicembre 2017):

- I.A. Bunin, *Mitina ljubov'* (1923) / *L'amore di Mitja* (R. Kufferle, 1934) – cap. 1-14
- V.Ja. Zazubrin, *Šepka* (1923) / *La scheggia* (S. Vitale, 1990) – cap. 1-4
- A.P. Platonov, *Ščastlivaja Moskva* (1933-1936) / *Mosca felice* (O. Discacciati – S. Vitale, 1996) – cap. 1-6

⁶⁵ Si intende l'anno di pubblicazione della versione russa, che seguì la pubblicazione in inglese nel 1955.

⁶⁶ Se precisa che di quest'opera sono stati allineati, a parte, anche 6 capitoli con due traduzioni (quella di Binni del 1999 e di S. Rapetti del 1998).

⁶⁷ I casi in cui un traduttore abbia scelto un titolo alternativo sono segnalati nell'elenco.

⁶⁸ Il racconto conta poco più di 10.000 parole, pertanto è stato inserito in versione integrale.

- V.V. Nabokov, *Dar* (1937) / *Il dono* (S. Vitale, 1998) – parte del cap. 1
- Č.T. Ajtmatov, *Proščaj, Gul'sary!* (1966) / *Addio Gul'sary* (C. Di Paola – S. Leone, 1988) – cap. 1-5
- S. Sokolov, *Škola dlja durakov* (1973) / *La scuola degli sciocchi* (M. Crepax, 2007) – cap. 1
- S.D. Dovlatov, *Inostranka* (1987) / *Straniera* (L. Salmon, 1991) – cap. 1-4
- S.D. Dovlatov, *Filial* (1987) / *La filiale* (L. Salmon, 2010) – versione integrale⁶⁹
- Ju.I. Druznikov, *Angely na končike igly* (1988) / *Angeli sulla punta di uno spillo* (F. Aceto, 2006) – cap. 1-4
- V.O. Pelevin, *Žizn' nasekomych* (1993) / *La vita degli insetti* (V. Piccolo, 2000) – cap. 1-4
- L.E. Ulickaja, *Veselye pochorony* (1997) / *Funeral Party* (E. Guercetti, 2004) – cap. 1-7
- B. Akunin, *Azazel'* (1998) / *La regina d'inverno* (P. Pera, 2000) – cap. 1-4
- V.V. Erofeev, *Chorošij Stalin* (2004) / *Il buon Stalin* (L. Montagnani, 2008) – cap. 1- e parte del cap. 2
- A.Ju. Kurkov, *Zakon ulitki* (2005) / *I pinguini non vanno in vacanza* (D. Osimo, 2006) – cap.1-8

Opere ita-ru allineate, disponibili nel NKRJa (dicembre 2017):

i)

- C. Goldoni, *La locandiera* (1753) / *Traktirščica* (A. Dživelegov, 1933) – I atto, scene 1-22
- E. De Amicis, *Cuore* (1886) / *Serdce* (V. Davidenkova, 1958) – cap. 1-2
- L. Pirandello, *Il fu Mattia Pascal* (1904) / *Pokojnyj Mattia Paskal'* (G. Rubcova, N. Rykova, 1967) – cap. 1-5
- I. Calvino, *Il Barone rampante* (1957) / *Baron na dereve* (L. Veršinina, 1965) – cap. 1-5
- G. Bassani, *Il Giardino dei Finzi-Contini* (1962) / *Sad Finci-Kontini* (I. Soboleva, 2008) – cap. 1-6
- U. Eco, *Il pendolo di Foucault* (1988) / *Majatnik Fuko* (E. Kostjukovič, 2006) – cap. 1-6
- A. Baricco, *Novecento* (1994) / *Legenda o pianiste* (N. Kolesova, 2005)⁷⁰
- A. Camilleri, *Il cane di terracotta* (1996) / *Sobaka iz terrakotty* (A. Kondjurina, 2005) – cap. 1-5
- V.M. Manfredi, *Alexandros, III, Il confine del mondo* (1998) / *Aleksandr makedonskij. Predely mira* (M. Kononov, 2003) – cap. 1-7
- N. Ammaniti, *Io non ho paura* (2001) / *Ja ne bojus'* (V. Nikolaev, 2005) – cap. 1-2
- G. Faletti, *Io uccido* (2002) / *Ja ubivau* (I. Kostantinova, 2005) – cap. 1-4

ii)

⁶⁹ Inserita dalla Scuola Superiore di Lingue Moderne per Interpreti e Traduttori di Forlì.

⁷⁰ Testo allineato in versione integrale perché conta in tutto 11.700 parole.

- C. Collodi, *Pinocchio* (1883) / *Priključenija Pinokkio* (E. Kazakevič, 1959)
- L. Pirandello, *La signora Frola e il signor Ponza, suo genero* (1917) / *Gospoža Frola i ee zjat' gospodin Ponca* (M. Gasanova-Mijat, 2006)
- U. Eco, *Il nome della rosa* (1980) / *Imja rozy* (E. Kostjukovič, 1989)
- iii)
- D. Buzzati, *Eppure battono alla porta* (1942) / *I vse že stučat v dver'* (I. Smagin, 2010)
- D. Buzzati, *I sette messaggeri* (1942) / *Sem' goncov* (F. Dvin, 2010)
- D. Buzzati, *Il borghese stregato* (1942) / *Očarovannyj buržua* (P. Chlodovskij, 2010)
- D. Buzzati, *Il mantello* (1942) / *Plašč* (G. Kiselev, 2010)
- D. Buzzati, *L'assalto del Grande Convoglio* (1942) / *Napadenie na bol'šoj konvoj* (P. Chlodovskij, 2010)
- D. Buzzati, *L'uccisione del drago* (1942) / *Kak ubili drakona* (F. Dvin, 2010)
- D. Buzzati, *Sette piani* (1942) / *Sem' etažej* (G. Kiselev, 2010)

Per quanto riguarda la saggistica, fra gli autori dei saggi russi selezionati ricordiamo M.M. Bachtin (per la categoria 'critica letteraria'), N.A. Berdjaev (religione/teologia), P.A. Florenskij (epistolario), L.D. Laundau, (scienze esatte e naturali), A.S. Politkovskaja (storia/biografia), E.N. Trubeckoj (arte), V.S. Solovev, (filosofia/psicologia), A.I. Solženicyn (scienze sociali), K.S. Stanislavskij (memorie).

Per la sezione di saggistica italiano-russo, invece, sono stati selezionati saggi di P. Grassi (arte), F. De Sanctis (critica letteraria), E. Lo Gatto (critica letteraria), U. Eco (filosofia/psicologia), M. Montessori (scienze sociali/formazione), A. Donini (religione/teologia), T. Regge (scienze esatte e naturali), G. Boffa (storia/biografia)⁷¹. Solo pochi di questi, tuttavia, sono già disponibili sul corpus (dicembre 2017)⁷², e precisamente:

- M.M. Bachtin, *Problemy poetiki Dostoevskogo* (1929) / *Dostoevskij: Poetica e stilistica* (G. Garritano, 1968) – cap 1
- L.K. Cukovskaja, *Zapiski ob Annoj Achmatovoj* (1976) / *Incontri con Anna Achmatova, 1938-1941* (G. Moracci, 1990)⁷³
- V.S. Solovev, *Krizis zapadnoj filosofii* (1874) / *La crisi della filosofia occidentale* (A. Dell'Asta, 1980) – cap. 1
- S.N. Bulgakov, *U sten Chersonisa* (1922) / *Presso le mura di Chersoneso* (M. Campatelli, 1998) – cap. 1-2

⁷¹ Per un elenco completo si rimanda a V. Nosedà, *Corpora Paralleli e linguistica contrastiva*, pp. 97-101, e al sito del NKRJa, dove è possibile consultare l'elenco delle opere di volta in volta inserite (http://search1.ruscorpora.ru/search.OPERE_RU-ITA) (http://search1.ruscorpora.ru/search.OPERE_ITA-RU) (ultima consultazione: gennaio 2018).

⁷² La maggior parte è ancora in fase di allineamento.

⁷³ Manca una suddivisione in capitoli, perciò sono state inserite le prime pagine fino a contare circa 10.000 parole.

- A.S. Politkovskaja, *Putinskaja Rossija* (2004) / *La Russia di Putin* (C. Zonghetti, 2010)⁷⁴

b) Dimensione del corpus e lunghezza dei testi

Grazie al primo ampliamento qui descritto il CP ita-ru è passato da 700.000 a 4.429.600 parole. L'intervento dei collaboratori esterni (cfr. par. 3) è stato prezioso per raggiungere tale entità, ma ha creato una discrepanza, non solo tra il sotto-corpus dei testi in direzione russo-italiano e quello dei testi in direzione italiano-russo, ma anche tra testi in versione integrale ed estratti (sebbene alcune delle opere intere siano brevi racconti al di sotto delle 15.000 parole). Una possibile soluzione per rimediare a questo sbilanciamento è creare due sezioni separate del corpus, separando appunto i testi integrali dagli estratti. Tuttavia, per un ampliamento futuro continueremo a privilegiare l'inserimento di estratti, avendo potuto constatare⁷⁵ che per la ricerca linguistica disporre di un numero elevato di esempi tratti dalla stessa opera è di scarsa utilità (non così, certamente, per l'analisi letteraria), poiché le tendenze linguistiche riscontrate dal ricercatore potrebbero dipendere dallo stile individuale di un determinato autore o traduttore. La presenza di opere in versione integrale non crea invece alcun sbilanciamento quantitativo del corpus, se queste non superano le 15.000 parole, come per *Il Cappotto* di N. Gogol', *Il grande inquisitore* di F. Dostoevskij e *Novecento* di A. Baricco, da noi inserite.

c) Bilanciamento

Come già accennato, il CP ita-ru nella versione attuale presenta un certo sbilanciamento di cui indichiamo qui sotto i termini quantitativi:

- 33 sono i testi letterari russi, mentre quelli italiani sono 21;
- 5 i testi di saggistica russi e nessuno italiano;
- la sezione ru-ita conta 3.705.044 parole, contro le 724.556 parole della sezione ita-ru;
- 17 opere sono in versione integrale (di entità superiore alle 15.000 parole) mentre della lunghezza prestabilita (non oltre le 15.000 parole) sono 42 (racconti integrali o estratti di opere).

Il futuro ampliamento dovrà pertanto tener conto di questi dati, così da ristabilire un corpus il più possibile bilanciato. In particolare, considerando l'esiguo numero di titoli di saggistica al momento disponibili, e visto lo sbilanciamento generale tra originali russi e originali italiani, gli obiettivi che ci proponiamo per il futuro in merito alla selezione dei testi sono i seguenti:

- inserire le seguenti opere letterarie italiane selezionate e in parte già allineate⁷⁶:
 - A. Manzoni, *I promessi sposi* (1827) / *Obručennnye* (N. Georgievskaja, 2011)

⁷⁴ Anche in questo caso manca una suddivisione in capitoli.

⁷⁵ Cfr. V. Noseda, *Corpora Paralleli e linguistica contrastiva*, p. 139.

⁷⁶ Tutte queste opere sono in fase di allineamento e appariranno sul corpus nel 2018.

- R. Giovagnoli, *Lo Spartaco* (1873-1874) / *Spartak* (S. Stepnjak-Kravčinskij, 1881 / L. Frenkel', 1936 / Jasnaja, 1954)
- E. Salgari, *Il corsaro nero* (1898) / *černyj korsar* (G. Smirnov, 1991)
- I. Svevo, *La coscienza di Zeno* (1923) / *Samopoznanie Dzeno* (S. Buševa, 1972)
- A. Moravia, *Gli indifferenti* (1929) / *Ravnodušnye* (L. Veršinina, 2010)
- G. Guareschi, *Don Camillo: Mondo piccolo* (1948) / *Malyj mir* (O. Gurevič, 2014)
- G. Rodari, *Le avventure di Cipollino* (1951) / *Priključenija Čipollino* (Z. Potapova, 1955)
- N. Ginzburg, *Lessico familiare* (1963) / *Semejnye besedy* (G. Smirnov, 1898)
- L. Sciascia, *I pugnalatori* (1976) / *Palermskie ubijcy* (E. Solonovič, Z. Potapova, 1982)
- B. Pizzorno, *La casa sull'albero* (1984) / *Dom na dereve* (T. Stamojaja, 2012)
- G. D'Agata, *Il ritorno dei templari* (1997) / *Vozvrašenie tamplierov* (I. Kostantinova, 2005)
- aumentare sensibilmente le opere di saggistica;
- realizzare un *target variant corpus* più completo: per il momento abbiamo nel CP ita-ru solo due opere allineate con più di una versione in lingua 2.

d) La ricerca con il corpus

Concludiamo con alcune osservazioni relative alla potenzialità del CP per la ricerca. Innanzitutto, le ricerche possono essere condotte sul corpus in base ai parametri delle annotazioni. Il CP ita-ru presenta infatti tre tipi di annotazione:

- Meta-testuale: comprende le informazioni extralinguistiche sull'opera inserita (vi si accede cliccando sul titolo dell'opera).
- Morfologica: descrive ogni forma di parola dal punto di vista grammaticale (appartenenza alla classe, genere, transitività o intransitività, aspetto, caso, numero, tempo, persona ecc.).
- Semantica: fornisce informazioni di tipo semantico sui lemmi. Troviamo categorie tassonomiche riferite a nomi concreti (es.: "animali", "cibo"), nomi astratti (es.: "emozioni", "percezioni"), aggettivi (es.: "qualitativi", "quantitativi"), verbi (es.: "moto", "causativo") e via dicendo.⁷⁷

Vogliamo fornire di seguito un esempio di come vengono visualizzati i metadati relativi a un'opera (fig. 1) e le informazioni morfologiche e semantiche relative a un lemma (fig. 2):

⁷⁷ Purtroppo, la sezione dei corpora paralleli non è stata sottoposta ad un processo di disambiguazione, per cui il sistema non distingue gli eventuali omonimi presenti, sia lessicali sia grammaticali. Un obiettivo a lungo termine, realizzabile da un gruppo di lavoro numeroso, potrebbe essere la disambiguazione del CP ita-ru, per ovviare e problemi di omonimia e rendere la ricerca ancora più precisa.

Fig. 1. Esempio di visualizzazione dei metadati relativi ad un'opera

2. Виктор Ерофеев. Хороший Сталин (2004) [омонимия не снята] Все примеры (114)

ru	Правда, отец до сих пор жив и	Автор	Виктор Ерофеев	играл в теннис. [Виктор Ерофе
it	Per la verità, mio padre è ancora	Дата рождения автора	1947	iosava a tennis. [Viktor Erofeev.
ru	Уединившись на втором этаже	Название	Хороший Сталин	кребутся ветви высокого д
it	Se ne sta appartato per ore nel s	Дата создания	2004	то — уже подробности. [Ви
	(forse le sue memorie?): ma que	Сфера функционирования	художественная	estre grattano i rami della gra
		Предложений	1620	alin (Luciana Montagnini,) [омон
ru	Наши суждения о них беспом	Словоформ	24914	редвзятости, неизжитых де
it	I nostri giudizi su di loro sono in	Язык	rus	----->
	[Viktor Erofeev. Il buon Stalin (Luciana	Переводчик	Luciana Montagnini	lle insuperate paure infantili,
		Язык перевода	ita	
ru	Но когда семейная интимность	Сообщить об ошибке...		
	начинаешь думать, вспоминать и анализировать. [Виктор Ерофеев. Хороший Сталин (2004)] [омонимия не сня			скандала, который ставит

Fig. 2. Esempio di visualizzazione delle informazioni morfologiche e semantiche relative a un lemma

ru	Правда, отец до сих пор жив и даже по выходным до недавнего времени играл в теннис. [Виктор Ерофеев. Хороший Стали	ВЫХОДНЫМ		Stalin (Luci
it	Per la verità, mio padre è ancora vivo e fino a poc	Лемма	выходной (см. в словарях)	
ru	Уединившись на втором этаже в своем дачном	Грамматика	прил, м, ед, твор, полн	долго, м
it	Se ne sta appartato per ore nel suo studio al primo		прил, с, ед, твор, полн	феев. Хор
	(forse le sue memorie?): ma questi non sono che d		прил, мн, дат, полн	ercia, e le
			сущ, неод, м, мн, дат	e снята]
			сущ, неод, м, ед, твор	
ru	Наши суждения о них беспомощны, высочайше	Семантика основная	der.v, dt.move, r.abstr, r.rel, t.time:week	трахах, €
it	I nostri giudizi su di loro sono incerti, campati in e	Семантика дополнительная	der.v, dt.move, r.rel	udio tra f
	[Viktor Erofeev. Il buon Stalin (Luciana Montagnini,)] [омон	Доп. признаки	пасс, словарн, ru	
ru	Но когда семейная интимность расширяется до масштабов международного скандала, который ставит семью на пор	Сообщить об ошибке...		
	начинаешь думать, вспоминать и анализировать. [Виктор Ерофеев. Хороший Сталин (2004)] [омонимия не снята] <----->			

È dunque possibile condurre la propria interrogazione su una o più forme di parola specifiche (*poisk točnych form*) (un esempio è fornito dall'analisi di A. Bonola del 'comunque' italiano⁷⁸ che consente di confermare la spiccata polisemia e polifunzionalità di questa parola⁷⁹, nonché di stabilire una discrepanza tra descrizione lessicografica e uso effettivo), oppure in base alle categorie semantiche e grammaticali (*leksiko-grammatičeskij poisk*) evidenziate dall'annotazione, che il ricercatore può combinare in base al suo quesito di ricerca (per esempio, verbi al futuro perfettivo, sostantivi astratti femminili al caso strumentale ecc.). Come per l'intero NKRJa, anche nel CP i risultati possono essere richiesti in formato standard (ogni periodo è seguito dalla traduzione in lingua 2), oppure in formato KWIC

⁷⁸ A. Bonola, *Il corpus parallelo italiano-russo-italiano come strumento per l'indagine contrastiva delle marche discorsive e testuali: il caso di 'comunque' e i suoi traduttivi russi*, "Contrastive Linguistics", 42, 2017, 4, pp. 89-104.

⁷⁹ In particolare, si registrano quattro tratti: demarcativo, conclusivo, modale, generalizzante (*Ibid.*, p. 102).

(*key words in context*). La ricerca lessico-grammaticale (*leksiko-grammatičeskij*) rileverà tutte le forme di parola in cui un dato lessema può manifestarsi. È possibile anche inserire linguaggi di interrogazione speciali, ad esempio **pomnit'* -*napomnit'*, che troverà tutte le parole che terminano in *pomnit'*, tranne *napomnit'*. Questo tipo di interrogazione può essere utile per indagare i verbi prefissali russi e la loro resa in italiano. Il russo, in quanto lingua *satellite-framed*⁸⁰ è ricca di verbi prefissali i cui satelliti (i prefissi) conferiscono a verbi con la medesima base verbale (come ad esempio *pomnit'* [ricordare]) tratti semantici specifici e distintivi. Il confronto *corpus-based* con l'italiano (al contrario annoverabile tra le lingue *verb-framed*) rivela le possibili strategie di resa di verbi quali *vspomnit'* [venire in mente] e *zapomnit'* [memorizzare], concludendo che il mancato impiego di tali strategie porta nella maggior parte dei casi a una perdita di informazione. L'analisi degli esempi, inoltre, consente di registrare tratti semantici non riportati negli strumenti lessicografici tradizionali (che spesso trattano i verbi sopraccitati come sinonimi) e funge dunque da stimolo per integrare e migliorare i dizionari bilingui russo-italiano sulla base dei corpora⁸¹.

Dal motore di ricerca del corpus possono essere estratti anche interi sintagmi. Ne è un esempio l'analisi della costruzione causativa italiana 'fare' + infinito, che si ottiene richiedendo al sistema di cercare tutte le occorrenze del verbo 'fare' (in ogni sua forma) seguito direttamente da qualsiasi verbo all'infinito. Nello specifico tale indagine è utile per registrare le modalità di espressione di una situazione causativa in russo e in italiano⁸².

Infine, è possibile creare sotto-corpora *ad hoc* per effettuare la ricerca solo su una sezione desiderata del corpus (una determinata opera o un gruppo di opere selezionate secondo i parametri cronologici e stilistici con cui è stato annotato l'intero NKRJa): creando un sotto-corpus comprendente esclusivamente *Il Cappotto* di N. Gogol' e le sue traduzioni italiane, è possibile ad esempio concentrare l'indagine sul confronto tra le diverse versioni del racconto, tracciando così la storia delle sue traduzioni⁸³.

In conclusione, avendo coniugato l'ampliamento quantitativo con l'applicazione dei più collaudati criteri di composizione dei corpora paralleli, il CP ita-ru presente oggi nel NKRJa si configura come uno strumento sufficientemente ampio e affidabile per condurre ricerche linguistico-letterarie comparate che abbiano anche un fondamento statisticamente rilevante nella pratica traduttiva scritta, sia letteraria sia saggistica. Inoltre, è stato dato l'avvio alla possibilità di un aggiornamento del corpus che sia continuo e, soprattutto, sistematico, caratteristiche indispensabili per assicurare nel tempo l'efficienza e l'affidabilità di questo strumento di ricerca.

⁸⁰ Sulla distinzione tra lingue *satellite-framed* e lingue *verb-framed* cfr. L. Talmy, *Towards a Cognitive Semantics. Typology and Process in Concept Structuring*, MIT Press, Cambridge/London, 2, 2001.

⁸¹ Cfr. V. Noseda, *Corpora paralleli e linguistica contrastiva*, pp. 120-147.

⁸² *Ibid.*, pp. 175-193.

⁸³ *Ibid.*, pp. 199-246.



FACOLTÀ DI SCIENZE LINGUISTICHE E LETTERATURE STRANIERE
L'ANALISI LINGUISTICA E LETTERARIA

ANNO XXVI - 2/2018

EDUCatt - Ente per il Diritto allo Studio Universitario dell'Università Cattolica
Largo Gemelli 1, 20123 Milano - tel. 02.72342235 - fax 02.80.53.215
e-mail: editoriale.dsu@educatt.it (produzione)
librario.dsu@educatt.it (distribuzione)
redazione.all@unicatt.it (Redazione della Rivista)
web: www.analisiilinguisticaeletteraria.eu

ISSN 1122 - 1917



9 788893 353915