

Saša MODERC
(Università di Belgrado)

Produzione e applicazione di corpora bilingui per l'insegnamento dell'italiano L2

Abstract: (Producing and applying bilingual corpora in teaching Italian as a foreign language) Single bilingual texts and bilingual corpora haven't yet found a relevant place in foreign language teaching, although today it is an easy task, even with limited computational skills, to create bilingual texts starting, in our case, from literary works written in one language (Serbian) and translated into another language (Italian). These linguistic materials are also easily manageable and editable; they can be organized dependently on how they are intended to be used in the fields of foreign language teaching, translation studies, contrastive analysis and the like. Single parallel texts can be easily united in one "corpus", and this solution, though rudimental, allows the user to have an elementary insight into all the included literary texts, or into the texts written by one author, or by one translator, or into translated texts published in one specific period. Since these bilingual texts and corpora are produced for research activity and linguistic analysis and since they are conceived for scholars and students working with foreign languages, they may represent, in our opinion, an instrument of enormous potential upon which didactic, linguistic, literary or critic activities can be organized, along with the possibility to use them as a lexicographic basis and repertory, especially for those languages, like Serbian, which still lack good bilingual dictionaries.

Keywords: Italian, Serbian, bilingual corpus, translation studies, contrastive analysis

Riassunto: I corpora elettronici bilingui e i testi bilingui in genere non sembrano aver ancora trovato adeguato spazio nella didattica delle lingue straniere, nonostante il fatto che oggi sia possibile creare materiali bilingui mediante procedimenti informatici richiedenti competenze informatiche elementari, partendo, nel nostro caso, da testi letterari serbi tradotti in italiano. Siffatti materiali bilingui possono essere modificati e gestiti agevolmente e si possono organizzare in modo da rispondere agli usi che se ne desiderino fare, in seno all'analisi contrastiva, alla glottodidattica, alla traduttologia ecc. I testi bilingui prodotti autonomamente possono essere raggruppati e uniti in un "corpus", soluzione che, per quanto rudimentale, permette di analizzare a livelli elementari i testi di un autore, oppure i testi di un traduttore, o testi di traduzioni generate in uno specifico periodo o pubblicato in uno specifico centro (elemento importante per i testi serbo-croati). Trattandosi di testi e corpora che si possono elaborare per attività di ricerca, analisi e riflessione linguistica e destinati a ricercatori e studenti, essi rappresentano uno strumento di enorme potenziale didattico ed ermeneutico intorno al quale è possibile organizzare diverse attività glottodidattiche, linguistiche, letterarie, critiche, nonché sfruttarli come base lessicografica e repertorio di soluzioni per quelle lingue, che, come il serbo, ancora oggi deficitano di dizionari bilingui.

Parole chiave: italiano, serbo, corpus bilingue, traduttologia, analisi contrastiva

Un documento elettronico bilingue contiene due testi di varia lunghezza, uno nella lingua originale e un altro, che ne è la traduzione. I due testi sono disposti in una tabella con due colonne parallele; ciascun segmento del testo originale (di lunghezza non costante, andante da singole frasi a interi paragrafi) è allineato nella stessa riga di tabella, col rispettivo segmento di traduzione. I testi bilingui sono in formato elettronico compatibile con la maggior parte dei programmi di scrittura al computer, quindi sono comodamente consultabili e modificabili. La facilità di accesso a questi testi consente di leggere, confrontare, valutare ed analizzare simultaneamente il testo dell'originale e quello della traduzione (o delle traduzioni, dal momento che è possibile allineare con l'originale traduzioni in più lingue. I testi bilingui vengono definiti anche bitext, o testi paralleli. Nelle edizioni cartacee è di uso comune il termine „testo a fronte“, da evitare in contesti di

linguistica computazionale e testi elettronici onde non creare ambiguità. Un testo bilingue può avere questa forma¹:

Ivo Andrić Na Drini ćuprija Većim delom svoga toka reka Drina protiče kroz tesne gudure između strmih planina ili kroz duboke kanjone okomito odsečenih obala. ...	Ivo Andrić Il ponte sulla Drina Traduzione di Bruno Meriggi Per la maggior parte del suo corso il fiume Drina s'apre la strada attraverso anguste gole tra scoscese montagne o attraverso profondi canyon dai fianchi a picco. ...
--	---

Un testo elettronico contenente più traduzioni può avere questa forma²:

Lewis Carroll	Lewis Carroll	Lewis Carroll	Lewis Carroll
CHAPTER I Down the Rabbit-Hole	Első fejezet Lenn, a Nyuszi barlangjában	Capítulo 1 En la madriguera del conejo	I NELLA CONIGLIERA
Alice was beginning to get very tired of sitting by her sister on the bank, and of having nothing to do:	Alice már elunta, hogy tétlenül üldögéljen nénye mellett az árokparton.	Alicia empezaba ya a cansarse de estar sentada con su hermana a la orilla del río, sin tener nada que hacer:	Alice cominciava a sentirsi assai stanca di sedere sul poggetto accanto a sua sorella, senza far niente:

Uno dei vantaggi offerto da questo elementare formato elettronico, compatibile con Microsoft Office, è la possibilità di tagliare il testo, copiarlo, modificare il carattere, variare le dimensioni o le linee della tabella, ovvero adeguare alle proprie necessità la forma e il contenuto della tabella: nel caso della tabella multilingue qui sopra, per motivi di spazio abbiamo dovuto omettere le colonne con le traduzioni in portoghese, francese, tedesco ed esperanto.

Il primo tentativo di creare un testo elettronico bilingue è di alcuni decenni posteriore ai primi testi e corpora elettronici monolingui; l'iniziativa risale al 1987 (Véronis 2000, 1) e deve la sua origine all'idea di sfruttare gli allora emergenti computer personali per agevolare ed accelerare il lavoro dei traduttori. Il risultato di questi sforzi è stato porre le basi teoriche e pratiche per la creazione di programmi CAT (Computer Aided/Assisted Translation), con memorie di traduzione (Translation Memory; in altre parole, glossari elettronici) che, una volta definite, memorizzate e inserite nella banca dati del programma stesso aiutano il traduttore a reperire rapidamente la terminologia specialistica, riducendo i tempi di lavoro del traduttore ed aumentandone la precisione e la coerenza terminologica. Da quanto esposto si conclude che ai testi paralleli è stata dedicata grande attenzione, ma con finalità precipuamente commerciali (produzione e vendita di software per traduttori) o teoriche (risolvere i problemi matematici relativi ai corpora), trascurando il campo delle traduzioni letterarie che, come vedremo più avanti, riserva notevoli potenziali per la didattica e la ricerca linguistica.

Per analizzare più testi bilingui è necessario organizzarli in un corpus. Nella linguistica dei corpora esistono dei criteri ben precisi in base ai quali si definisce che cosa è un corpus. Un corpus, per essere tale, deve essere elettronico, finito (circoscritto a un

¹ Il testo bilingue serbo-italiano del "Ponte sulla Drina" è opera nostra. Abbiamo ritenuto sufficiente in questa sede citare solo il nome dell'autore, senza riportare dati bibliografici precisi e completi sulle edizioni in serbo e su quelle in italiano dei romanzi inclusi nel nostro corpus (vedi nota 8).

² Il testo è stato scaricato dal sito www.farkastranslations.com, dove sono a disposizione testi letterari paralleli che includono fino a sette lingue-target.

numero preciso di testi, in modo da porgere dati statistici attendibili), tokenizzato e markuppato (BARBERA 2013, 43). Numerosi documenti ufficiali dell'UE sono attualmente disponibili in formato bitext³; per una prima informazione sull'argomento possono essere utili i diversi link verso corpora bilingui, disponibili all'indirizzo http://en.wikipedia.org/wiki/Parallel_text o in fonti più documentate, come Gandin (2009, 135-136) per i corpora bilingui e Barbera (2013) per una concisa ma preziosa rassegna sulle problematiche relative ai corpora monolingui e bilingui. In questo contesto, segnaliamo due realtà interessanti per il serbo: il progetto dell'Università di Graz, che ha dato vita a un corpus multilingue costituito da testi letterari in tedesco e nelle lingue slave (il Gralis Korpus⁴, accessibile online, sviluppato dallo slavista Branko Tošović) o i corpora bilingui (serbo-francese, serbo-inglese) sviluppati presso l'Università di Belgrado⁵, contenenti diverse tipologie di testi.

Per quanto riguarda i testi non tecnico-specialistici, ovvero i corpora di testi letterari bilingui, l'applicazione nella didattica e nella ricerca di questo strumento, pur non nuovo, pare non aver trovato ancora lo spazio che merita. I corpora letterari bilingui non sembrano essere numerosi, oppure non sono accessibili online: questo stato di cose ci ha indotto a considerare la possibilità di creare un corpus bilingue di testi letterari serbi e delle loro traduzioni in italiano, da usare come base per ricerche linguistiche e come fonte di materiali didattici da mettere a disposizione degli insegnanti di italiano L2 presso l'Università di Belgrado. Ora, nella nostra visione, non tecnica, di "costruttori ingenui" (BARBERA 2013, 42), consideriamo corpus bilingue un insieme di testi bilingui prodotti e selezionati in base a determinati criteri. Nel nostro caso, il corpus contiene testi letterari serbi e le loro traduzioni in italiano, ma risulta privo della elaborazione tecnica richiesta per un corpus vero e proprio (non è tokenizzato e markuppato). Il nostro corpus, viste le sue "imperfezioni tecniche" può essere definito anche "pseudocorpus", o "protocorpus", poco togliendo al rilievo del materiale bilingue preparato per la ricerca e l'ulteriore uso⁶.

È importante sottolineare il fatto che i corpora elettronici sono ben lungi dall'essere una risorsa riservata a una ristretta cerchia di adepti, cui si accede solo in veste di fruitori, nella speranza di reperire materiali bilingui confacenti al percorso didattico o di ricerca che si intende intraprendere. La creazione e l'uso di testi bilingui, seppure in forma elementare, senza l'ausilio di appositi programmi di consultazione, è un procedimento semplice, alla portata di chiunque abbia competenze informatiche di base. Per creare un testo bilingue è sufficiente avere in formato elettronico il testo originale e la sua traduzione (scannerizzando le edizioni cartacee, oppure acquistando i testi direttamente in formato elettronico). Una volta procurati, i testi monolingui vanno emendati da eventuali imperfezioni tecniche e grafiche originate durante la scannerizzazione e la conversione nel formato desiderato, come Microsoft Office (più avanti, M.O.) o formati alternativi compatibili. In questa fase di lavoro è importante avere a disposizione i moduli per il controllo ortografico delle lingue di lavoro (nel nostro caso, per il serbo e per l'italiano), al fine di rimuovere gli errori ortografici generati durante il lavoro di preparazione dei testi (a volte ci è successo di incorrere in ovvi

³ Vedi <http://ipsc.jrc.ec.europa.eu/index.php?id=198>, oppure: <http://www.statmt.org/europarl/>, due siti contenenti corpora paralleli di testi giuridici prodotti dal Parlamento Europeo.

⁴ http://www-gewi.uni-graz.at/gralis/korpusarium/gralis_korpus.html.

⁵ <http://korpus.matf.bg.ac.rs/>.

⁶ Il collega Miloš Utvić, matematico ed esperto di linguistica computazionale presso la Facoltà di Filologia di Belgrado, in una comunicazione personale ha confessato che uno dei maggiori problemi per lui e i suoi colleghi matematici e programmatori è procurarsi i testi bilingui su cui lavorare.

errori di stampa, prontamente rimossi in quanto non espressione di intenzioni stilistiche dello scrittore o del traduttore). L'ultima fase del lavoro consiste nella parallelizzazione automatica dei due testi, ovvero nella disposizione "a fronte", in una tabella, dei segmenti di testo corrispondenti. L'allineamento dei testi si può effettuare con vari programmi elaborati appositamente⁷; gli errori di allineamento, inevitabili ma relativamente pochi, sono dovuti alle differenti soluzioni grafiche presenti nei testi da parallelizzare, in primo luogo alle differenze di punteggiatura tra l'originale e la traduzione. Infatti il programma LF Aligner allinea il testo segmentandolo in frasi, le quali vengono identificate in base a segni grafici come il punto fermo, il punto esclamativo e il punto interrogativo: i segmenti dell'originale e della traduzione che divergono nella punteggiatura saranno allineati erroneamente e richiederanno una correzione manuale dell'allineamento, correzione che anche in testi di notevole lunghezza richiede poco tempo e conoscenze di base relative al funzionamento delle tabelle di M.O. o programmi affini.

Il risultato finale di queste semplici operazioni è un documento bilingue in formato M.O., come illustrato nei due esempi sopra, con l'incipit del "Ponte sulla Drina" e di "Alice nel Paese delle Meraviglie". Quali sono i vantaggi di questo formato e di questo materiale bilingue? Il formato di M.O. è compatibile con altri programmi di scrittura elettronica, i quali si possono usare alternativamente, anche senza l'obbligo di acquisto (come nel caso di OpenOffice e LibreOffice, per citarne solo due). L'uso di questi programmi di scrittura comunemente usati consente di intervenire sui testi (per effettuare correzioni, aggiungere commenti ecc.), copiarli, inserirne segmenti in altri documenti, variare la larghezza delle tabelle, convertirli in altri formati ecc. Il formato M.O. e la sua componente Excel danno, inoltre, la possibilità di unire in un unico file più testi bilingui, creando in tale maniera un corpus (con le limitazioni tecniche di cui sopra) "su misura". Nel nostro caso, in pochi mesi di lavoro non intenso siamo riusciti ad allineare 18 romanzi serbi⁸ e 19 traduzioni in italiano (per un romanzo di Andrić, "Prokleta avlija", sono state prese in considerazione ambedue le traduzioni: ne abbiamo ottenuto un "bitext doppio"). Quindi abbiamo inserito i testi bilingui in un documento di Excel, copiando le singole tabelle dei bitext e incollandole in un unico foglio di lavoro. Il corpus formato in tale modo conta più di 2.500.000 parole; la sua realizzazione ha richiesto risorse minime e conoscenze informatiche elementari. Attualmente il corpus è in fase di ampliamento e includerà, nel giro di pochi mesi, grazie all'aiuto di collaboratori, una cinquantina di titoli serbi tradotti in italiano. Esaurita questa fase, ci proponiamo di formare un altro corpus, inverso, di opere letterarie italiane tradotte in serbo(croato): tale corpus, considerato il numero di opere in prosa tradotte dall'italiano sarà ben più voluminoso del primo. Il documento-corpus ottenuto, nonostante la mole, risulta perfettamente consultabile anche su computer meno sofisticati e moderni e costituisce, già in questa sua forma "ingenua", uno strumento di lavoro prezioso, convertibile nei formati necessari per una elaborazione più avanzata, in seno alla linguistica computazionale.

Indipendentemente dal lavoro su testi letterari e dalla ricerca linguistica, saper applicare le tecniche per formare un testo elettronico bilingue e un corpus bilingue

⁷ Nelle nostre attività abbiamo usato il programma LF Aligner 4.05, elaborato da Andras Farkas e abbiamo conseguito risultati soddisfacenti nella creazione di testi paralleli partendo da formati comuni (come .doc, .txt, .pdf ecc.). Il programma è liberamente scaricabile dal sito <http://sourceforge.net/projects/aligner>.

⁸ Il corpus attualmente comprende sette opere di Ivo Andrić (romanzi e raccolte di racconti) e due opere di David Albahari, due di Svetislav Basara e una, rispettivamente, di Aleksandar Gatalica, Aleksandar Tišma, Emir Kusturica, Danilo Kiš, Zoran Živković, Goran Petrović e del classico croato Miroslav Krleža.

costituisce una competenza importante per il futuro professionale di quanti studino lingue straniere e intendano specializzarsi nell'interpretariato. Infatti, tale competenza permetterà loro di creare autonomamente corpora bilingui contenenti i testi integrali dei documenti tradotti (funzione di corpus-archivio), a cui accedere per reperire gli equivalenti terminologici e fraseologici (funzione di corpus-dizionario) e scambiare i risultati del proprio lavoro con i colleghi interpreti (funzione di corpus-“esperienza professionale”).

Una volta prodotti, i testi bilingui possono essere impiegati in diversi modi nella didattica delle lingue straniere. A seconda degli specifici obiettivi glottodidattici, è possibile estrarne brani da usare in esercizi aventi come obiettivo identificare e/o fissare vari elementi linguistici, come l'uso delle preposizioni, l'uso dell'articolo, dei tempi verbali, dei pronomi clitici, la concatenazione dei tempi, o applicarli in esercizi di completamento del testo e di lavoro sul lessico, nonché in altre attività ideate dall'insegnante, a seconda dei percorsi didattici che si intendono proporre. In ultima istanza i testi bilingui si possono usare come strumenti per facilitare la lettura di testi in lingua straniera, specialmente ai corsi di base o nel caso di autori dalla lingua particolarmente difficile.

La tipologia e la varietà delle attività didattiche basate sui testi paralleli dipendono dalla creatività dell'insegnante. Improvvisando, possiamo proporre attività incentrate su punti nevralgici dell'insegnamento dell'italiano L2 - l'uso delle preposizioni semplici e l'uso dell'articolo:

1. Inserisci la preposizione semplice:

Naime, Gec, ili Majer, oduvek je želeo da bude pilot ratnog aviona.
Nemam nikakvih dokaza da je on to stvarno želeo, ali dopada mi se pomisao da on ulazi u kabinu svog kamiona kao da stupa u bombarder, na sebi ima kožnu jaknu, a pilotsku kapu ne stavlja samo zato što mu je pomalo neprijatno da to čini u prisustvu suvozača.

Infatti, Goetz, o Meyer, ha sempre desiderato essere pilota _____ un aereo militare.
Non ho nessuna prova che lo desiderasse davvero, ma mi piace l'idea _____ lui che sale _____ cabina _____ suo camion come se entrasse _____ un bombardiere, indossa un giubbotto _____ pelle, e non si mette il casco _____ pilota solo perché gli secca un po' farlo _____ presenza del suo compagno _____ guida.

2. Inserisci l'articolo determinativo:

Posle toga, Gec i Majer nemaju više šta da rade, osim da voze, naravno.
Kamion sa stvarima odavno ih je napustio.

Duše u karoseriji nisu.

One će uzleteti sve zajedno, tek kada kamion stigne na odredište.

Dopo di che, _____ Goetz e Meyer non hanno più _____ niente da fare, tranne guidare, naturalmente.

Il camion con _____ bagagli li ha lasciati da _____ tempo.

_____ anime a _____ interno del cassone no.

Loro voleranno via tutte assieme, quando _____ camion arriverà a _____ destinazione.

Bastano questi due brevi brani paralleli tratti dal romanzo “Goetz e Meyer”, di David Albahari, per notare l'abbondanza e la varietà del materiale linguistico messo in evidenza dal testo parallelo. Sta all'insegnante segnalare i nodi sintattici e lessicali che emergono dal materiale bilingue; nel caso della tabella presentata qui sopra, possiamo identificarne alcuni, senza andare troppo nel dettaglio, proponendo tra parentesi gli equivalenti, non felici, che ci si aspetterebbe considerato il testo serbo:

- il perfetto “ha sempre desiderato” (non: “desiderava sempre”, quantunque il verbo serbo sia imperfettivo),
- “pomisao” : “idea” (non “pensiero”),
- “na sebi ima” : “indossa” (non un improbabile “ha su di sé”),

- d. “jaknu” : “giubbotto” (non: “giacca”),
 e. “pilotsku kapu” : “casco” (non: “cappello”),
 f. “mu je (...) neprijatno” : “gli secca” (non: “sente imbarazzo”),
 g. “uzleteti”, è riferito alle anime, ma è collegato anche con le ambizioni aeronautiche del tedesco, ovvero con l’idea del “decollare”, “prendere il volo”. La semantica di “uzleteti” consente il gioco di parole, mentre i termini tecnici “decollare” e “prendere il volo” lo negano; “volare via” è una soluzione non precisa, ma economica ed accettabile.

Per quanto riguarda l’analisi contrastiva, un esempio di attività è stato offerto qui sopra. È possibile organizzare percorsi di analisi dei testi al fine di identificare le somiglianze e le divergenze lessicali, morfologiche, sintattiche e culturali in essi presenti. Tale attività richiede una particolare competenza linguistica del docente. Nel caso del corpus bilingue serbo-italiano, il confronto tra le due lingue, reso più immediato dal testo a fronte, mette in risalto le caratteristiche più vistose delle due lingue, l’aspetto verbale nel serbo e la temporalità del sistema verbale italiano, due fonti di tante incertezze sia per italofofoni che serbofofoni. Intorno a queste differenze si possono organizzare diverse attività incentrate sulle peculiarità dei due sistemi verbali, lavorando con brani di testo di lunghezza sufficiente affinché le differenze di uso dei tempi verbali e le diverse strategie di strutturazione del testo diventino visibili. Come illustrazione proponiamo un brano tratto dal romanzo “Secolo”, di Aleksandar Gatalica:

3. Analizza l'uso dei tempi verbali:

Bacao je kačket uvis, vojnički salutirajući „gospodi Liberti“ i nije slutio da će uskoro, tu na ostrvu Elis, oostati bez svog majmuna.

Njega su pregledali i dopustili mu da prođe kapiju Amerike, ali životinja nije mogla u Novi svet.

Hladno su mu kazali da će se oni pobrinuti za majmuna i izvestiti ga o tome.

Dantea više nije video.

Buttava il cappello per aria, presentando il saluto militare alla "signora Liberty" e neppure immaginando che di lì a poco, sull'isola di Ellis, sarebbe rimasto senza la sua scimmia.

Lo visitarono concedendogli il permesso di passare il portone dell’America, ma l'animale non poteva entrare nel Nuovo mondo.

Gli avevano detto freddamente che si sarebbero occupati loro della scimmia e che lo avrebbero tenuto informato.

Ma Dante, non lo vide più.

In questo brano desta curiosità il trapassato prossimo “avevano detto”, usato in corrispondenza di un perfetto indicativo serbo (cui in genere corrisponde il passato remoto; cfr. nella riga superiore “su pregledali” e “visitarono”). Usi particolari del trapassato prossimo sono dettagliatamente illustrati in Bertinetto (2003); tali usi non hanno equivalenti morfosintattici in serbo e per questa ragione il testo italiano del romanzo “Secolo” rappresenta un caso interessante. Infatti, ai circa 1400 trapassati prossimi della traduzione non corrisponde nessun tempo trapassato serbo⁹. Poiché solo parte dei trapassati della traduzione sono usati per eventi cronologicamente antecedenti, questo tempo verbale merita di essere studiato nei suoi rapporti di (non) equivalenza con il perfetto serbo.

Abbiamo potuto presentare solo alcune delle possibili applicazioni dei testi paralleli nella glottodidattica e nella ricerca linguistica. Per sfruttare a pieno i testi e i corpora bilingui è indispensabile avvicinarsi a quello che sono i corpora linguistici veri e propri ed effettuare, separatamente per ciascuna lingua, la tokenizzazione e il markupaggio dei testi. Accanto al

⁹ I tempi trapassati serbi, anche se registrati dalle grammatiche, sono disusati. Nel testo serbo in questione vi sono un centinaio di trapassati di verbi di stato e non vengono percepiti come forme verbali che esprimono azioni verbali antecedenti.

lavoro, importante, ma meccanico e ripetitivo ("robotico", a detta di una collaboratrice), di formazione del nostro corpus serbo-italiano, accanto al proposito di popolarizzare la tecnica di generazione di testi paralleli, riteniamo sia necessario ideare un software in grado di gestire testi bilingui e consentire interrogazioni di vario genere, lavorando con due lingue e dando risultati relativi ad ambedue le lingue del corpus.

In conclusione, andrebbero sicuramente menzionati i sensibili aspetti legali relativi all'uso di testi protetti da diritti d'autore. L'accessibilità ai testi bilingui deve essere limitata alle forme previste dalla legislazione locale e dalla legislazione dei paesi in cui sono state pubblicate le traduzioni dei testi letterari. L'eventuale accesso online ai testi bilingui andrebbe limitato a porzioni di testo e a modalità di ricerca definite in accordo con gli editori; è possibile ipotizzare anche una specie di gradazione di accessibilità al materiale bilingue, imponendo limitazioni alla consultazione online in termini di lunghezza del brano di testo visionabile, riservando a ricercatori provvisti di password le opzioni di ricerca su segmenti di testo più ampi, come nel caso dell'analisi dei tempi verbali o nella linguistica testuale, oppure consentire la ricerca senza limitazioni solo in loco, offline, nella sede dell'istituto preposto allo sviluppo e alla gestione del corpus.

Riferimenti bibliografici

- Andorno-Rastelli (a cura di). 2009. *Corpora di Italiano L2: Tecnologie, metodi, spunti teorici*. Perugia, Guerra Edizioni.
- Arnold, Balkan, Lee, Meijer, Sadler. 1994. *Machine Translation: An Introductory Guide*. NCC Blackwell, Oxford.
- Barbera, Manuel. 2013. *Linguistica dei corpora e linguistica dei corpora italiana. Un'introduzione*. Qu.A.S.A.R. s.r.l., Milano.
- Bertinetto, Pier Marco. 2003. *Tempi verbali e narrativa italiana dell'Otto/Novecento*. Edizioni dell'Orso, Alessandria.
- Gandin, Stefania. 2009. "Linguistica dei corpora e traduzioni: definizioni, criteri di compilazione e implicazioni di ricerca dei corpora paralleli". *Annali della Facoltà di Lingue e Letterature Straniere*. Vol. 5, 2005 (2009), pp. 133-152.
- Laviosa, Sara. 2002. *Corpus-based Translation Studies: Theory, Findings, Applications*. Rodopi, Amsterdam and New York.
- Laviosa, Sara. 1999. "Come studiare e insegnare l'italiano attraverso i corpora", in *Italica*, Vol. 76, No. 4. *Linguistics and Pedagogy* (Winter, 1999), pp. 443-453.
- Melamed, Dan. 2001. *Empirical Methods for Exploiting Parallel Texts*. The MIT Press. Cambridge, Massachusetts. London, England, 2001.
- Moderc, Saša. 2014. "I testi letterari paralleli e la valutazione della traduzione: il caso dell'interpunzione". *Nasleđe. Journal of Language, Literature, Arts and Culture*. Volume XI, Issue 29, 2014. Faculty of Philology and Arts Kragujevac, pp. 203-216.
- Moderc, Saša. In stampa. "Paralelizovani tekstovi i učenje na daljinu" (Testi paralleli e distance learning). *Naučni skup Digitalne biblioteke i digitalni arhivi*. Univerzitet u Beogradu, Univerzitet u Novom Sadu. Beograd-Novu Sad, 7-8. april 2014.
- Santos, Diana. 2000. "The translation network. A model for a fine-grained description of translations", in: Véronis (2000), pp. 169-186.
- Véronis, Jean (a cura di). 2000. *Parallel Text Processing. Alignment and Use of Translation Corpora*. Springer Science+Business Media Dordrecht
- Véronis-Langlais. 2000. "Evaluation of parallel text alignment systems. The ARCADE project". In: Véronis (2000), pp. 370-388.