



La diversità linguistica nell'era digitale: le sfide della lingua serba

di Olja Perišić

ABSTRACT: Il presente contributo è dedicato all'analisi degli strumenti digitali utilizzati nell'ambito dell'insegnamento e della lessicografia del serbo come LS. Considerate le particolari caratteristiche di questa lingua, soprattutto in relazione alla sua ricca morfologia, tra gli strumenti indispensabili nelle prime fasi di studio si annoverano i dizionari morfologici, mentre ai livelli intermedio e avanzato si rendono necessari i dizionari digitali e i corpora linguistici. A differenza dell'inglese, che in tema di tecnologie digitali conosce un aggiornamento e uno sviluppo pressoché continui, le lingue con meno risorse, come il serbo, fanno prevalentemente ricorso a strumenti tradizionali, e solo negli ultimi anni si è affermata la tendenza a orientarsi verso il digitale. La pandemia globale, esplosa all'inizio del 2020, ha accelerato questo processo sollecitando gli insegnanti nella ricerca di percorsi innovativi che si prestino alla didattica virtuale. Le questioni qui enucleate e per le quali in questa sede si cercherà di trovare una risposta riguardano: gli strumenti digitali per la lingua serba come LS, i loro vantaggi rispetto al materiale didattico tradizionale, le possibili strade da seguire nella sperimentazione della metodologia *Data-driven learning*.



ABSTRACT: The paper is dedicated to the analysis of digital tools used in the teaching and lexicography of Serbian as foreign language. Considering the particular characteristics of this language, especially in relation to its rich morphology, the essential tools in the early stages of study include morphological dictionaries, while at the intermediate and advanced levels digital dictionaries and linguistic corpora are necessary. Unlike English, which in terms of digital technologies is being updated and developed almost continuously, languages with fewer resources, such as Serbian, mainly make use of traditional tools, and only in recent years has the tendency to orient themselves towards digital. The global pandemic, which exploded in early 2020, has accelerated this process by urging teachers to search for innovative paths that lend themselves to virtual teaching. The questions listed here and for which we will try to find an answer concern: digital tools for the Serbian as foreign language, their advantages over traditional teaching material, the possible ways to follow in experimenting with the Data-driven learning methodology.

PAROLE CHIAVE: lingua serba; didattica; dizionari; digitale; DDL.

KEY WORDS: Serbian language; didactics; dictionaries; digital; DDL.

INTRODUZIONE

La lingua serba vanta una lunga tradizione di insegnamento nelle università italiane cui si unisce una tradizione lessicografica bilingue ancora più consolidata (Perillo 401; Banjanin 67). Nell'ambito della didattica, negli ultimi anni gli approcci tradizionali stanno ormai cedendo il passo alle tecnologie digitali. In un primo periodo il mondo digitale era inteso sia come una fonte da cui attingere il materiale linguistico autentico sia come una risorsa mediante la quale condividere le conoscenze e comunicare con i parlanti nativi. In questo modo anche le lingue meno diffuse potevano ricavarsi un proprio spazio nel mondo digitale dominato, come si sa, dall'inglese. La rivoluzione digitale che si è ormai affermata a livello globale ha aperto nuove strade nella didattica delle lingue e ha provocato mutamenti e una sorta di integrazione tra le metodologie esistenti, soprattutto in relazione alle lingue più parlate. Secondo una ricerca della Multilingual Europe Technology Alliance (META), che ha compreso trenta lingue europee, l'inglese è l'unica a godere di un buon supporto ("good support") da parte di strumenti digitali come: *Machine Translation, Speech Processing, Text Analysis, Speech and Text*. L'italiano rientra nel gruppo delle lingue alle quali si riconosce un supporto



moderato (“moderate support”) in tutte le categorie tranne che per strumenti specifici come *Machine Translation* (“fragmentary support”). Invece il serbo è stato ritenuto una lingua “weak/no support” per tutte le categorie considerate, fatta eccezione per strumenti quali *Speech and Text Resources* e *Speech Processing* (“fragmentary support”) (META).

A partire dai primi mesi del 2020 l'emergenza sanitaria da COVID-19 ha influito irreversibilmente sulle modalità di utilizzo e fruizione dei contenuti digitali. Non a caso il report sulla comunicazione digitale per il 2021 intitolato *Think forward* (Tecci) parla di “social reset”, cioè di profondi cambiamenti nella quotidianità anche per quanto concerne l'istruzione. L'attuale emergenza sanitaria fa sì che ogni individuo espanda e in un certo senso privilegi la comunicazione virtuale non solo nella vita privata ma anche nel contesto lavorativo e negli studi. Secondo il rapporto *Digital 2021* sull'uso delle tecnologie digitali a livello globale, in Italia, su una popolazione di 60,41 milioni di persone (gennaio 2021) si registrano 50,54 milioni di utenti Internet, con una crescita del 2,2% rispetto all'anno precedente (Kemp). La domanda sull'effettiva necessità delle tecnologie digitali e sulla misura in cui queste dovessero essere utilizzate ha trovato una risposta con l'emergenza pandemica. Le università si sono rapidamente organizzate, cosicché a livello individuale ogni insegnante ha dovuto rimodulare i propri programmi per affrontare le nuove esigenze dettate dalla situazione contingente. Oggi, dopo un intero anno accademico svolto in modalità online, si è più consapevoli dei benefici e dei vantaggi della didattica virtuale che, a detta di molti, andrebbe mantenuta “in una prospettiva [...] sempre più intrecciata, mista e blended” (Magni 21).

Secondo l'esperienza di chi scrive, è emerso che nell'insegnamento del serbo come LS il numero degli strumenti digitali e tradizionali di cui si dispone è nel complesso modesto e che gli studenti conoscono poco e usano altrettanto poco le risorse digitali in ambito universitario. Al tempo stesso, quando si offre loro la possibilità di acquisire nuove conoscenze digitali, i dati risultano incoraggianti ed è da questa consapevolezza che occorre partire per una didattica che orienti gli studenti verso i nuovi dispositivi per diventare sempre più autonomi. Come si è visto, il mondo digitale offre anche agli studenti delle lingue meno parlate la possibilità di usufruire di nuove risorse (ad es. corpora linguistici e dizionari digitali) grazie alla metodologia conosciuta come *Data-driven learning* – DDL (Johns, *Printout* 14), ormai affermatasi nel mondo anglosassone.

Nella prima parte del presente contributo verranno delineate le risorse digitali disponibili per la lingua serba e il campo del loro possibile utilizzo. Saranno prese in considerazione le risorse lessicografiche, monolingui e bilingui, presenti online, ma anche quelle basate sulla collaborazione e l'interazione degli utenti, che in questo caso possono apportare modifiche ai contenuti perfezionandoli o implementandoli. La seconda parte dell'indagine sarà dedicata ai corpora linguistici per la lingua serba e al loro utilizzo nella didattica per gli studenti italofoni. In ultimo saranno esposti i principi della metodologia *Data-driven learning* e il suo possibile impiego nella didattica del serbo come LS.

Nel complesso il contributo si propone di individuare i vantaggi degli strumenti digitali rispetto a quelli tradizionali soffermandosi nello specifico sulle risorse esistenti,



per capire se queste permettano una proficua applicazione della metodologia *Data-driven learning* anche per lo studio della lingua serba.

LE RISORSE DIGITALI PER LA LINGUA SERBA

Una delle maggiori difficoltà che ogni studente di serbo alle prime armi deve affrontare è la ricca quanto complessa morfologia della lingua, tanto più evidente se si pensa che ai 120.000 lemmi di un dizionario corrispondono circa 4,5 milioni di forme grammaticali flesse (Vitas *et al.* 12). Per questo motivo, soprattutto nelle fasi iniziali dell'apprendimento, lo studente ha la necessità di consultare dizionari morfologici, utili per ricavare la forma base (lemma) di una parola a partire da qualsiasi forma flessa. Attualmente la lingua serba dispone di un dizionario morfologico ad accesso libero, *Morfološki rečnik srpskog jezika*, che, seppur non completo, consente tuttavia, la ricerca di parole in caratteri cirillici o latini, a fronte di risultati in alfabeto latino (Fig. 1).

МОРФОЛОШКИ РЕЧНИК СРПСКОГ ЈЕЗИКА

Претрага

Упишите (ћирилицом или латиницом) реч коју тражите:

(Реч можете уписати у основном или у неком изведеном облику, нпр. ако тражите реч *језик*, у поље можете уписати: *језик, језика, језику, језиком, језици* итд.; ако тражите реч *певати*, у поље можете уписати: *певати, певам, певамо, певај, певаше, певаху* итд.)

Потражи!

Ово може бити:

- imenica ž. r. [ruka](#)

Кликните на плаву реч и погледајте све њене облике.

imenica ž. r. **ruka**

	Jd.	Mn.
Nom.	ruka	ruke
Nom. 2		
Gen.	ruke	ruku
Gen. 2		
Dat.	ruci	rukama
Dat. 2		
Ak.	ruku	ruke
Ak. 2		
Vok.	ruko	ruke
Vok. 2		
Ins.	rukom	rukama
Ins. 2		
Lok.	ruci	rukama
Lok. 2		

Fig. 1. Schermata di ricerca della parola *ruka* ("mano") nel dizionario morfologico della lingua serba

Per valutare al meglio lo scenario lessicografico serbo è però indispensabile prestare attenzione a un contesto più vasto, per non dire globale. In proposito si può parlare di due momenti 'rivoluzionari' nella lessicografia, il primo dei quali si lega alla *Corpus revolution* e al *COBUILD project* iniziato nel 1981 presso l'Università di Birmingham, che ha portato alla creazione del primo corpus lessicografico in inglese. Di tale corpus ci si era serviti per la compilazione del dizionario *Collins COBUILD English Dictionary* (1987), che rappresenta un decisivo salto in avanti verso la lessicografia di nuova generazione, dal momento che questo strumento ha modificato radicalmente ogni concezione alla base dei precedenti dizionari. I corpora potevano di fatto svolgere l'intero lavoro in precedenza affidato ai lettori/informatori e se le definizioni nel periodo pre-elettronico spesso proponevano casi insoliti e rari di utilizzo di una parola, in questa



fase più evoluta i computer operavano secondo modalità e criteri oggettivi, per cui oltre a mettere a disposizione una messe di esempi essi mostravano allo stesso tempo la norma e le eccezioni. Già dai primi istanti era apparso evidente come le dimensioni di un simile corpus avessero reso tanto più rappresentativi questi esempi. Un corpus di grandi dimensioni consente infatti di estrarre esempi significativi che vengono recuperati tramite lo strumento delle concordanze. Una criticità connessa alle dimensioni di un corpus si ha ogniqualvolta si renda necessario passare in rassegna un numero molto ampio di concordanze al fine di poter sintetizzare i risultati; tuttavia l'inconveniente è stato risolto grazie alla statistica lessicale (Church e Hanks 22) e alle diverse formule di associazione tra le parole, come ad es. il sistema MI (*Mutual Information*).

Una seconda trasformazione si è avuta a partire dagli anni '90 del secolo scorso con la transizione dai dizionari cartacei a quelli digitali. La prima fase di questo processo è stata puramente formale, dal momento che il contenuto del dizionario cartaceo veniva semplicemente riversato su supporto digitale. Le prime versioni dei dizionari informatici erano pubblicate su CD, modalità che da alcuni anni a questa parte sta scomparendo a vantaggio delle edizioni online. Questo tipo di pubblicazione non solo risulta accessibile a un pubblico più vasto, ma dal punto di vista economico consente anche di diminuire i costi a carico degli editori, che ora dispongono di un diretto controllo sui dizionari e sono in grado di provvedere al loro aggiornamento in qualsiasi momento.

Nella seconda fase, a partire dal primo decennio degli anni Duemila, si è registrato anche in questo campo un massiccio impiego della telefonia mobile e del Web, supporti che hanno reso possibile l'accesso a una nuova generazione di dizionari digitali e il loro utilizzo con benefici di varia natura, se si pensa che si tratta di dispositivi soggetti ad aggiornamento continuo e che si distinguono per la semplicità d'utilizzo per ogni tipo di ricerca e consultazione (ipertestualità); tali dizionari possono inoltre essere multimediali (prevedono cioè funzionalità video e audio), possono contenere link per contenuti esterni e non presentano inconvenienti come limiti di spazio. Come noto, i dizionari informatici sono strumenti aperti e in divenire, ed è grazie a questa loro versatilità che sono in condizione di acquisire e registrare ogni cambiamento della lingua. In sintesi, l'ipertestualità e la multimedialità dei contenuti sono peculiarità che rendono questi strumenti davvero appetibili da parte di ogni tipologia di utenti, soprattutto se nativi digitali.

Dopo il passaggio dai dizionari cartacei a quelli digitali è ora in corso un'ulteriore trasformazione che vede i personal computer cedere il posto alle applicazioni di telefonia mobile (Marello 63). Si fa presente che nei gruppi di lavoro con gli studenti sono ormai rare le occasioni (tranne forse le sedute di esame) in cui chi apprende una lingua consulta un dizionario cartaceo.

Dopo aver offerto una panoramica diacronica della lessicografica digitale occorre ora prendere in considerazione i dizionari serbi monolingui e bilingui che possono essere utilizzati nelle classi di lingua serba come LS per italofoni. A partire dai livelli iniziali dello studio del serbo la traduzione può essere praticata come un'attività di profilo contrastivo che consente un ritorno e un confronto continui con la lingua madre



dello studente, in questo caso l'italiano. In tal modo viene sviluppata la consapevolezza che la traduzione non è mai un processo lineare e che le collocazioni variano da una lingua all'altra. Ai livelli intermedi e avanzati la traduzione viene insegnata come un'abilità indipendente ed è a questo livello di apprendimento che solitamente gli studenti si accorgono dei limiti dei dizionari tradizionali, connessi soprattutto all'anisomorfismo lessicale che si esprime attraverso le collocazioni, i vuoti lessicali ("lexical gaps") e i casi di polisemia. Le lacune riscontrabili a livello di micro-lingua sono dovute spesso allo spazio limitato che i dizionari tradizionali di solito riservano a ogni lemma. Nella lessicografia serba dominano ancora i dizionari cartacei e sono relativamente pochi i dizionari digitali disponibili, anche se il loro numero è destinato a crescere (Kosem *et al.* 95). In questa sede sono presentate alcune piattaforme con materiale lessicografico consultabile online.

Il *Raskovnik* è una piattaforma per ricerche lessicografiche fondata nel 2015 dall'Istituto per la lingua serba (SANU) con il sostegno del Ministero per la cultura e per l'informazione serba. Ha lo scopo di creare un ambiente per la digitalizzazione del patrimonio lessicografico e favorire lo sviluppo degli strumenti per la ricerca in merito alla lingua e alla letteratura serba. Tale risorsa è stata sviluppata dal Centro per Digital Humanities e da DARIAH-RS (The Digital Research Infrastructure for the Arts and Humanities). Per ora sono stati inseriti lo storico dizionario trilingue di Vuk Karadžić (serbo-tedesco-latino) e quattro dizionari dialettali. La ricerca avanzata avviene per mezzo di caratteri jolly che permettono di consultare i dizionari in modo selettivo: per lemma, forma grammaticale, definizione. Di particolare interesse è la mappa del territorio ex iugoslavo, dove a molti toponimi corrisponde un link: cliccando si viene rimandati alla sezione del sito corredata di mappe nelle quali si possono localizzare le singole località.

Il *Rečnik srpskohrvatskog književnog i narodnog jezika* ("Dizionario della lingua serbo-croata letteraria e popolare") ha mantenuto il nome originale, visto che la sua compilazione ha avuto inizio in epoca iugoslava e almeno il 40% del materiale illustrativo è ascrivibile a scrittori non serbi (bosniaci, croati e montenegrini). Il dizionario, tuttora incompiuto, annovera 21 volumi per un totale di circa 250.000 lemmi, ma per giungere alla completa realizzazione dell'opera si prevede di realizzarne almeno altri 14. Nel 2016 è iniziata la digitalizzazione del dizionario SANU, parzialmente già inserito nel dizionario online *srpskijezik.com*, sulla piattaforma *Srpska digitalna biblioteka*, insieme agli ultimi tre volumi del *Rečnik Matice srpske*. Numerosi sono gli esempi di lingua in uso, con estratti dalle maggiori opere letterarie, motivo per cui tale dizionario è assimilabile ai moderni dizionari *corpus-based*. Sulla stessa piattaforma sono poi presenti altre risorse lessicografiche, tra le quali una molto utile per gli studenti italofofoni, ossia il dizionario bilingue serbo-italiano di Mirjana Kosović, con circa 30.000 lemmi.

Negli ultimi anni è stato censito nel Web un alto numero di dizionari basati sul TM System (*translation memory*): tra questi, il più utilizzato dagli studenti per la combinazione linguistica cui si è fatto cenno, è il *Glosbe*, che si configura come risorsa aperta, o meglio, un progetto collaborativo che offre a ogni utente la possibilità di modificare, aggiungere o rimuovere contenuti. Le traduzioni si basano sulle memorie di



traduzione che permettono sia di visualizzare le parole nel loro contesto sia di osservare come queste si comportano. Per certi aspetti si tratta di un dizionario *corpus-based* in cui le traduzioni vengono estrapolate dal contesto, ma il suo punto debole è costituito dal numero e dalla varietà delle traduzioni e dall'assenza di una reale diversificazione dei testi. Ciò significa che i registri linguistici sono numericamente limitati e nella maggior parte degli esempi (tranne che per i sottotitoli dei film) non è possibile risalire alla fonte. In compenso gli utenti possono collaborare attivamente nella creazione dei contenuti inviando traduzioni, registrando la pronuncia di parole, correggendo eventuali errori e aggiungendo significati mancanti o integrando quelli esistenti. Questo tipo di collaborazione è ormai prevista anche per i dizionari consolidati e più conosciuti, come ad es. il *Macmillan English Dictionary (MED)*, che raccoglie i contributi degli utenti in una sezione separata denominata *The Macmillan Open Dictionary (MOD)*, successivamente integrata nel dizionario principale con le note degli utenti che hanno offerto un contributo. Con questa tecnica di *gamification*, che contempla anche una sorta di competizione e di premio, gli utenti non sono più partecipanti passivi ma diventano a tutti gli effetti parte attiva di questo processo (Rundell 15).

La più grande enciclopedia ad accesso libero, Wikipedia, contiene numerosi contenuti in serbo e figura al 21° posto nel mondo per il numero di voci in questa lingua (648.269). Considerata la possibilità di consultare una stessa voce in numerose lingue, questa enciclopedia vale anche come risorsa per le traduzioni della terminologia settoriale (ad es. flora e fauna) o per nozioni strettamente enciclopediche (politica, cultura, geografia ecc.), spesso assenti dai comuni dizionari. Anche se inizialmente sottovalutata soprattutto in ambito accademico, oggi tale risorsa viene vista come più affidabile per l'accuratezza dei contenuti, la verifica dei dati e la trasparenza delle fonti (Rundell 4). Con analoghe modalità è stato realizzato il Wiktionary, dizionario che pur non offrendo una completa panoramica della lingua, permette comunque di risalire ad alcuni significati attraverso definizioni, frasi illustrative (anche se non sempre rappresentative), sinonimi e la traduzione in inglese. Questa tipologia di dizionari gode di grande popolarità tra gli studenti perché permette un accesso diretto, cioè senza registrazione, ed è di semplice consultazione. Se nella didattica si favorisce l'uso di tali risorse lo studente è in grado non solo di valutare la loro adeguatezza e i possibili ambiti d'uso, ma anche di comprenderne gli aspetti positivi e quelli negativi, così come la necessità di integrare questi strumenti con altre risorse che, in determinati casi, offrono traduzioni certo più accurate.

CORPORA LINGUISTICI

I corpora linguistici consistono in raccolte di testi (o parti di testi) autentici e rappresentativi di una data varietà di lingua nella sua versione scritta o parlata. Il loro punto di forza è dovuto al fatto che sono disponibili in formato elettronico, anche se l'interrogazione/consultazione può variare da un software all'altro.

Le prime ricerche nell'ambito della linguistica dei corpora hanno avuto inizio negli anni '80 con il progetto COBUILD e a lungo si sono limitate all'inglese. Oggi questa



ricerca, applicata a tutti gli aspetti della linguistica, è praticata in molte altre lingue, come il serbo, che dispone di corpora molto validi.

Il primo corpus rappresentativo della lingua serba è stato creato all'interno del progetto "Le interazioni tra il testo e il dizionario" (*Interakcije teksta i rečnika*), realizzato nel 2002 dal Dipartimento per la lingua serba della Facoltà di Filologia di Belgrado, dalla Facoltà di Filosofia di Novi Sad e dalla Facoltà di Matematica di Belgrado. La sua prima versione, SrpKor 2003, non era annotata e annoverava 22,2 milioni di parole (Utvić 248). In un secondo momento è stata sviluppata la versione SrpKor 2013 con 122 milioni di parole e annotata morfologicamente: il corpus è accessibile online previa registrazione e include testi pubblicati nell'arco temporale dal 1910 al 2012. La piattaforma, inoltre, si completa di due corpora paralleli: quello inglese-serbo (*Englesko-srpski korpus 2012*) e quello francese-serbo (*Francusko-srpski korpus 2012*). Ancor prima dell'ultima versione gli autori evidenziavano alcuni aspetti peculiari della lingua serba che richiedevano una soluzione, tra cui l'uso parallelo di due alfabeti (cirillico e latino) e l'oscillazione tra le varianti *ecava* e *iecava*¹, previste e utilizzate entrambe nello standard serbo (Krstev e Vitas 2005). Per superare le criticità di alcuni nessi grafici (i digrammi *lj, nj, dž*) e dei segni diacritici (*č, ć, ž, š, đ*), si è dovuto codificare le lettere nel seguente modo (Fig. 2):

Latin	č	ć	ž	š	đ	lj	nj	dž
Cyrillic	ч	ћ	ж	ш	ђ	љ	њ	џ
Corpus encoding	cy	cx	zx	sx	dx	lx	nx	dy

Fig. 2. Codifica dei nessi grafici e delle lettere con segno diacritico.

I risultati delle *query* impostate con codificazione interna sono visualizzati nelle concordanze con le lettere dell'alfabeto latino. Nella ricerca semplice si può inoltre specificare il registro dei testi (amministrativo, letterario, scientifico, giornalistico) e la lingua in cui sono prodotti (cioè se si tratta di originali o traduzioni). La ricerca può essere svolta per *word-form*, lemma o tag e ammette l'opzione maiuscole/minuscole. Per quanto riguarda la visualizzazione, i risultati possono essere aggregati in base al contesto, con la possibilità di indicare la lunghezza di entrambi (a destra e a sinistra della parola chiave). Secondo le impostazioni predefinite vengono visualizzati cento risultati per pagina, ma si può anche scegliere di visualizzare risultati selezionati in modo casuale.

Un'altra piattaforma che contiene i corpora per la lingua serba è Sketch Engine (SE), con 500 corpora in 94 lingue diverse. Sviluppata dal linguista e lessicografo britannico Adam Kilgarriff, nel 2004 è stata resa disponibile dal programmatore ceco Pavel Rychlý per usi commerciali. Il sistema SE consiste di un software che permette a

¹ La distinzione come noto, si basa sulla diversa realizzazione di un'antica vocale *jat*, che nella parte orientale dell'area linguistica serba, è diventata *e* (*mleko*), mentre in quella occidentale *je/ije* (*mlijeko*).



ogni utente di creare, installare e rendere pronti all'uso corpora personali. La piattaforma offre inoltre metodi di statistica per ricavare frequenze, calcolare modelli di co-occorrenze ed esplorare corpora mono e multilingue. I corpora presenti su questa piattaforma sono già caricati (*preloaded*) e gestiti direttamente da SE, mentre i corpora degli utenti (*user corpora*) sono gestiti e condivisi dagli stessi utilizzatori. Per quanto riguarda la lingua serba sono presenti i seguenti corpora:

Tabella 1. Corpora Serbi e loro dimensioni

CORPORA SERBI SU SKETCH ENGINE	
Nome	Consistenza in parole
OPUS2 Serbian	153.237.786
Serbian Web (SrWaC 1.2)	476.888.297
Timestamped JSI web corpus 2014-2020 Serbian ²	519.947.546
Gutenberg Serbian 2020	70.724
Open Subtitles 2018 – Serbian	480.367.760

Fonte: *Sketch Engine*, <http://www.sketchengine.co.uk>.

La maggior parte dei corpora presenti in SE è lemmatizzata, pertanto si può eseguire una ricerca per lemma che come risultato darà tutte le forme flesse di quella parola nel contesto. Nel caso della lingua serba i corpora OPUS2 e Timestamped JSI non sono lemmatizzati, perciò è necessario svolgere la ricerca per word-form. Il corpus OPUS2, dal quale successivamente è stato creato l'Open Subtitles 2018, svolge la funzione di un corpus parallelo, in quanto consente la traduzione di una parola in altre lingue, tra cui l'italiano. Purtroppo il registro dei testi consultabili per questo tipo di ricerca è al momento molto limitato. Si tratta di sottotitoli di film, ma non viene specificato se siano traduzioni dirette da una lingua all'altra oppure eseguite con la mediazione dell'inglese. Il corpus più utilizzato dagli studenti è SrWaC 1.2, che contiene testi presenti nel Web e dunque offre anche gli esempi di lingua informale, usata nella comunicazione online. Questo corpus è disponibile anche nella variante croata e bosniaca e permette ricerche diatopiche tramite le quali si possono individuare somiglianze o differenze a tutti i livelli dell'analisi linguistica. L'ultimo corpus inserito per la lingua serba è Gutenberg Serbian 2020, un corpus creato dagli ebook disponibili nel database del Project Gutenberg ad aprile 2020.

Dello stesso software esiste anche un'altra versione *open source*, NoSketch Engine, che rispetto alla prima presenta tuttavia alcuni limiti. In primo luogo non sono presenti corpora già caricati, per cui l'utente dovrebbe disporre di adeguate conoscenze informatiche per predisporre un corpus da impiegare su questa piattaforma. Al tempo stesso NoSketch Engine è privo di una serie di funzionalità di SE, come a esempio *Word Sketches*, *terminology*, *thesaurus*, *n-grams*, *trands*, inoltre non contempla gli strumenti per la costruzione di un corpus. Questa versione dovrebbe essere scaricata, installata e gestita dall'utente.

² Si tratta di una serie di corpora in continuo aggiornamento per cui il numero di parole è soltanto indicativo. Per approfondire si rimanda all'apposita sezione sul sito Sketch Engine.



Gli strumenti principali per procedere all'analisi dei corpora sono: *concordance*, *word-list*, *keywords*, *Word Sketch*, *Sketch difference*, *thesaurus* e *n-grams*.³

DDL PER LA LINGUA SERBA

La metodologia conosciuta come DDL (*Data-driven learning*) si basa sul procedimento induttivo, *bottom-up*, attraverso il quale lo studente, grazie all'osservazione di dati autentici, perviene a conclusioni circa il funzionamento di determinati aspetti della lingua. La metodologia risulta particolarmente appetibile per ricerche di ambito lessicale, condotte attraverso l'analisi dei collocati, ma anche a livello morfosintattico e sociolinguistico. Scopo finale non è solo l'acquisizione della grammatica o del lessico, ma lo sviluppo di competenze e di un'autonomia che consentano allo studente di operare come un ricercatore (Johns, *Contexts* 101), mentre l'esperienza nel complesso può essere sintetizzata nella metafora del viaggio: *learner as traveller* (Bernardini 22). Un altro vantaggio riguarda lo sviluppo e il perfezionamento delle abilità cognitive, come ad es. la formulazione delle ipotesi e la loro verifica, il ragionamento, l'analisi, l'interpretazione e il confronto dei dati (O'Keeffe 3).

Il ricorso alla metodologia DDL nella didattica del serbo come LS è un campo al momento ancora poco esplorato (Vitaz e Poletanović 409). Le prime sperimentazioni della metodologia DDL nell'ambito della lingua serba come LS sono state svolte all'Università di Torino nell'anno accademico 2017/18 e come obiettivo si proponevano di osservare l'interesse degli studenti verso questa nuova metodologia di apprendimento che si avvale di esercizi lessicali e morfosintattici mirati. A partire dai primi risultati si è subito rilevata una risposta entusiastica da parte dei soggetti coinvolti, anche se sono emerse criticità. Occorreva infatti aumentare il numero di ore e gli studenti, per raggiungere i risultati previsti, avrebbero dovuto garantire una regolare presenza e una costante partecipazione alle lezioni. Un'altra difficoltà ha riguardato la complessità delle interrogazioni informatiche (*query*), che per alcuni ha costituito un vero e proprio blocco, soprattutto nelle fasi iniziali del corso (Perišić, *Upotreba* 103). Per superare l'ostacolo si è deciso di focalizzare le successive ricerche sulle collocazioni, utilizzando gli strumenti presenti in SE che, nel semplificare la ricerca, offrono una rappresentazione schematica dei risultati (collocati) raggruppati in base alla loro funzione sintattica (*Word Sketch* e *Sketch Difference*) (Perišić, *Collocazioni* 88). Nel complesso le sperimentazioni si sono finora mostrate utili soprattutto nell'ambito lessicale, *in primis* nella distinzione tra sinonimi vicini, il cui significato può essere dedotto dalle collocazioni create con determinate parole nel contesto. Le analisi morfosintattiche si sono rivelate utili con studenti principianti che hanno avuto l'occasione di affrontare i temi della morfologia di base attraverso uno strumento e una

³ Per istruzioni dettagliate su ogni funzionalità di questa piattaforma si vedano le apposite sezioni sul sito Sketch Engine.



metodologia del tutto innovativi (Perišić, *Corpora* 126). Con gli studenti di livello intermedio si è aperta una nuova strada nell'acquisizione dell'aspetto verbale, punto solitamente problematico per chi apprende le lingue slave.⁴

Un altro ambito che si presta particolarmente alle ricerche *corpus-based* è quello sociolinguistico, dal momento che la lingua rispecchia anche le specificità culturali e i mutamenti sociali, le cui dinamiche variano da una cultura all'altra e comprendono diversi modi di interpretare i fenomeni e parlare di essi.

Infine l'importanza della didattica DDL consiste nell'accrescere la consapevolezza circa l'importanza della lingua autentica anche tra gli studenti principianti, i quali tendono a pensare alla traduzione come a un processo lineare. Tramite l'accesso a un numero ragguardevole di esempi tratti dalla lingua in uso, ma soprattutto grazie alle collocazioni e alle combinazioni lessicali, essi imparano a discostarsi dalla logica della lingua madre e iniziano a ragionare in modo contrastivo. Si tratta di un passaggio molto importante nello sviluppo delle diverse abilità linguistiche, per giunta applicabili a ogni nuova lingua che si studia.

CONCLUSIONI

Nel presente contributo sono stati presentati gli strumenti digitali per la didattica del serbo come LS e si è cercato di rispondere a domande circa la loro utilità rispetto ai sistemi tradizionali e alla possibile applicazione della metodologia DDL.

Considerata la complessità del lessico serbo dal punto di vista morfologico, a partire dal livello principiante gli studenti avrebbero necessità di un dizionario morfologico per poter risalire alla forma base di una parola. In una situazione ideale ogni dizionario monolingue o bilingue consentirebbe la ricerca attraverso l'inserimento delle forme flesse, ma questa strada non sempre appare praticabile.⁵

Per quanto concerne il serbo, il numero di dizionari online bilingui risulta limitato, costituito per la maggior parte da repertori in formato digitale che sono il risultato di un'acquisizione dal cartaceo. Diversamente i dizionari monolingui solitamente presentano una struttura della microlingua molto più articolata, giacché oltre ai significati essi comprendono un numero assai elevato di esempi tratti dalla lingua autentica, in prevalenza letteraria. Un passo in avanti è costituito dai dizionari online definiti "collaborativi", dove gli utenti possono contribuire alla creazione dei contenuti e alla loro modifica. Basati spesso su memorie di traduzione che estrapolano i significati dal contesto, essi operano come dizionari *corpus-based*. Il loro numero, per la combinazione linguistica di cui ci si è occupati, è per ora limitato.

⁴ I risultati della ricerca sono stati presentati durante il convegno internazionale *51. Međunarodni naučni sastanak slavista u Vukove dane* tenutosi a Belgrado dal 15 al 20 settembre 2021. Il contributo verrà pubblicato negli Atti di convegno con il titolo *Usvajanje glagolskog vida na osnovu jezičkih korpusa*.

⁵ Ci si riferisce solo ai dizionari della variante serba, mentre il croato dispone di un dizionario monolingue (*Hrvatski jezični portal*) che consente di effettuare ricerche morfologiche.



Per quanto riguarda i corpora per il serbo, sono disponibili strumenti distribuiti su due piattaforme importanti che contengono testi diversificati per registro, tali da offrire una panoramica assai vasta sulla lingua e la possibilità di svolgere ricerche lessico-grammaticali e sociolinguistiche di diverso grado di complessità. La presenza di alcune tipologie di corpora per tutte le lingue provenienti dal serbo-croato permette di eseguire ricerche diatopiche e contrastive su vari livelli lessico-grammaticali o anche socioculturali. I corpora paralleli, che rivestono grande importanza nella traduzione, risultano quasi assenti, tranne che per un corpus contenente sottotitoli di film, ossia caratterizzato da un solo registro linguistico, pertanto privo di dati che possano dimostrare che la traduzione dei testi è diretta, senza la mediazione dell'inglese. Per le ricerche nell'ambito dei linguaggi settoriali si potrebbe utilizzare il corpus dei documenti dell'Unione Europea presente per ora solo nella variante croata. Un'altra soluzione per svolgere questo tipo di ricerche potrebbe essere la creazione autonoma di un corpus su SE che, attraverso specifici strumenti, permetta di scaricare dati dal Web tramite parole chiave oppure inserire testi di cui si è in possesso. Mancano anche i cosiddetti *learner's corpora* o corpora degli studenti di serbo come LS, che potrebbero facilitare più ricerche nell'ambito dell'acquisizione e della didattica delle lingue, come l'analisi degli errori, l'analisi contrastiva e altre ancora. Un progetto dedicato alla creazione di un corpus di questo tipo (*KSKS, Korpus srpskog kao stranog jezika*) è stato presentato durante la conferenza "Srpski kao strani jezik u teoriji i praksi III" (Miličević 279), ma fino a oggi non ci sono dati sulla sua realizzazione.

L'importanza e l'utilità dei corpora sono state provate da una lunga fase di ricerche e sperimentazioni, avviate nell'ambito della linguistica dei corpora e dirette soprattutto all'inglese, ma che negli anni si sono estese a molte altre lingue. L'alto numero di esempi autentici non solo permette l'accesso a un'enorme mole di dati da parte degli studenti che consultano i corpora in modalità autonoma, ma può servire agli stessi insegnanti per la creazione di materiale didattico sempre aggiornato e diversificato. Secondo l'esperienza di chi scrive le ricerche *corpus-based* si confermano di estrema utilità, perché permettono di considerare peculiarità e caratteristiche della lingua non avvertite neppure dai parlanti nativi. Le prime ricerche condotte negli ultimi quattro anni presso l'Università di Torino con studenti italofoni hanno messo in luce il loro vivo interesse: essi si sono rivelati aperti a nuovi stimoli, purché sollecitati in modo mirato e seguiti nei casi di eventuali difficoltà e resistenze iniziali. Oltre all'apprendimento delle strutture linguistiche, che sono parte irrinunciabile di ogni programma svolto in modalità tradizionale, gli strumenti digitali presentati durante il corso rappresentano un importante passo in avanti nelle ricerche individuali svolte dagli studenti in forma sempre più autonoma. L'entusiasmo per le conclusioni autentiche, spesso non presenti nei libri di testo, oltre a stimolare curiosità e partecipazione attiva predispone gli studenti a future ricerche in funzione dei loro percorsi di studio. Si auspica così che l'utilizzo dei corpora nella didattica della lingua serba possa continuare in modo uniforme e che il presente contributo risvegli l'interesse degli studenti e degli insegnanti intenzionati a sperimentare alcuni degli strumenti qui delineati. A chi si occupa dello sviluppo dei dizionari digitali o dei corpora va l'invito a cogliere alcuni



suggerimenti in merito a tali strumenti al fine di ampliare le ricerche e aumentare l'efficacia delle metodologie didattiche esistenti.

BIBLIOGRAFIA

Banjanin, Ljiljana. "Un contributo alla storia della serbo-croatistica italiana." *Nasleđe*, vol. XI, no. 29, 2014, pp. 67-81.

Bernardini, Silvia. "Corpora in the Classroom: An Overview and some Reflections on Future Developments." *How to Use Corpora in Language Teaching*, a cura di John Sinclair, John Benjamins, 2004, pp. 15-36.

Church, Kenneth Ward e Patrick Hanks. "Word Association Norms, Mutual Information, and Lexicography." *Computational Linguistics*, vol. 16, no.1, 1990, pp. 22-29.

Glosbe dizionario. <https://it.glosbe.com/>. Consultato il 3 Ago. 2021.

Hrvatski jezični portal. <https://hjp.znanje.hr/index.php?show=search>. Consultato il 07 Ago. 2021.

Johns, Tim. "From Printout to Handout: Grammar and Vocabulary Teaching in the Context of Data-driven Learning." *CALL Austria*, 10, 1990, pp. 14-24.

---. "Contexts: the background, development and trialling of a concordance-based CALL program." *Teaching and Language Corpora*, a cura di Anne Wichmann e Steven Fligelstone, Longman, 1997, pp. 100-115.

Karadžić, Vuk Stefanović. *Srpski rječnik: digitalno izdanje*. <http://raskovnik.org/recnici/VSK.SR>. Consultato il 02 Ago. 2021.

Kemp, Simon. *Digital 2021 Italy, Overview Report*. <https://datareportal.com/reports/digital-2021-italy>. Consultato il 9 Ago. 2021.

Kosem, Iztok, et al. "The image of the monolingual dictionary across Europe. Results of the European survey of dictionary use and culture." *International Journal of Lexicography*, vol. 32/1, 2019, pp. 92-114.

Kosović, Mirjana. *Srpsko-italijanski rečnik*. Prosveta, 2002.

Krsteš, Cvetana, e Duško Vitas. "Corpus and Lexicon-Mutual Incompleteness." *Proceedings of the Corpus Linguistics Conference, 14-17 July 2005*, a cura di Pernilla Danielsson e Martijn Wagenmakers, Birmingham, 2005. <https://www.birmingham.ac.uk/Documents/college-artslaw/corpus/conference-archives/2005-journal/TheLexicon/KrstešVitas.doc>. Consultato il 09 Ago. 2021.

Magni, Francesco. "Ciò che è vivo e ciò che è morto dell'università. Prime riflessioni pedagogiche nel mezzo della pandemia globale." *Formazione lavoro persona*, vol. 33, 2021, pp. 10-26. <https://forperlav.unibg.it/index.php/fpl/article/view/537>. Consultato il 10 Ago. 2021.

Marello, Carla. "Using Mobile Learning Dictionary in an EFL Class." *Proceedings of the XVI EURALEX CONGRESS: The User in Focus*, a cura di Andrea Abel, Chiara Vettori e Natascia Ralli, Eurac Research Press, 2014, pp. 63-83.

META-NET. *White Paper Series: Key Results and Cross-Language Comparison*. <https://bit.ly/373irwJ>. Consultato il 11 Ago. 2021.



Miličević, Maja. "Korpus srpskog kao stranog jezika (KSKS): opis građe i plan izrade." *Srpski kao strani jezik u teoriji i praksi III*, a cura di Vesna Krajišnik, Filološki fakultet, Centar za srpski kao strani jezik, 2016, pp. 279-289.

Morfološki rečnik srpskog jezika. <http://www.lexicom.rs/lexicon/search.aspx>. Consultato il 29 Ago. 2021.

O'Keeffe, Anne. "Data-driven learning – a call for a broader research gaze." *Language Teaching*, 2020, pp. 1-14.

Perillo, Francesco Saverio. "La serbocroatistica italiana: bilancio di un cinquantennio." *La slavistica in Italia. Cinquant'anni di studi (1940-1990)*, a cura di Giovanna Brogi Bercoff et al., Ministero per i Beni culturali e ambientali, Direzione Generale per gli Affari Generali Amministrativi e del personale, Divisione Editoria, 1994, pp. 401-428.

Perišić Arsić, Olja. "Upotreba korpusa u didaktici srpskog jezika kao stranog." *Srpski kao strani jezik u teoriji i praksi IV*, a cura di Vesna Krajišnik, Filološki fakultet, Centar za srpski kao strani jezik, Beograd, 2020, pp. 103-115.

---. "Collocazioni nella didattica del serbo come lingua straniera." *Filolog*, XI/22, 2020, pp. 88-115.

---. "Corpora in the Classroom – the Case of the Serbian Language for Italian Speakers." *New Trends in Slavic Studies*, a cura di Suarez Cuadros, Simon Jose et al., URSS, Mosca, 2021, pp. 126-137.

Raskovnik: srpski leksikografski portal. <http://raskovnik.org/>. Consultato il 29 Ago. 2021.

Rečnik srpskohrvatskoga književnog jezika, vol. I, no. VI, Matica srpska, 1967-1978.

Rečnik srpskohrvatskog književnog i narodnog jezika, Srpska akademija nauka i umetnosti – Institut za srpski jezik, 1959-2019.

Rundell, Michael. "Dictionaries and crowdsourcing, wikis and user-generated content." *International Handbook of Modern Lexis and Lexicography*, a cura di Patrick Hanks e Gilles-Maurice de Schryver, Springer, 2017. https://doi.org/10.1007/978-3-642-45369-4_26-1. Consultato il 09 Ago. 2021.

Sketch Engine. <http://www.sketchengine.co.uk>. Consultato il 29 Ago. 2021.

Utvić, Miloš. *Izgradnja referentnog korpusa savremenog srpskog jezika*. Tesi di dottorato, 2013. <https://fedorabg.bg.ac.rs/fedora/get/o:10061/bdef:Content/get>. Consultato il 09 Ago. 2021.

Srpska digitalna biblioteka. <http://www.srpskijezik.com>. Consultato il 29 Ago. 2021.

Tecci, Bruno. *Think forward 2021: The social reset*. <https://wearesocial.com/it/blog/2020/11/think-forward-2021-the-social-reset/>. Consultato il 10 Ago. 2021.

Vitas, Duško, et al. "Српски језик у дигиталном добу - The Serbian Language in the Digital Age." *Meta-Net, White Paper Series*, a cura di Georg Rehm e Hans Uszkoreit, Springer, 2012. <http://www.meta-net.eu/whitepapers/volumes/serbian>. Consultato il 01 Ago. 2021.

Vitas Duško e Miloš Utvić. *Korpus savremenog srpskog jezika na Matematičkom fakultetu Univerziteta u Beogradu*. <http://www.korpus.matf.bg.ac.rs>. Consultato il 22 Lugl. 2022.



Vitaz, Milica, e Milica Poletanović. "Data-Driven Learning. The Serbian case." *EL.LE*, vol. 8/2, 2019, pp. 409-422.

Wikipedia. <https://bit.ly/3iTZlcG>. Consultato il 29 Ago. 2021.

Wiktionary. <https://bit.ly/3GYp8AO>. Consultato il 29 Ago. 2021.

Olja Perišić è assegnista di ricerca e docente a contratto di Lingua serba e croata all'Università degli Studi di Torino. È laureata in Lingue e letterature straniere e ha conseguito il Dottorato di ricerca in Digital Humanities. Ha tradotto tre volumi della scrittrice croata Dubravka Ugrešić e ha pubblicato diversi articoli e saggi sulla linguistica dei corpora, l'analisi contrastiva e la traduzione. Partecipa attivamente a convegni internazionali in qualità di relatore.

<https://orcid.org/0000-0002-1219-862X>

olja.perisic@unito.it