

GRAZIELE ALTINO FRANGIOTTI

VERSO UN'AGENDA *DATA DRIVEN* NELL'INSEGNAMENTO DELL'ITALIANO IN BRASILE

1. INTRODUZIONE

Negli ultimi anni, e ancor più intensamente dopo la crisi sanitaria del Covid-19 nel biennio 2020-2021, l'istruzione di base in Brasile sta attraversando una profonda crisi. Quotidianamente emergono notizie riguardanti livelli di alfabetizzazione ancora insoddisfacenti nelle scuole, difficoltà legate alla capacità di risolvere problemi e sviluppare forme di ragionamento logico-matematico, nonché crescenti problemi di concentrazione da parte degli studenti. A questo scenario si aggiunge una vera e propria invasione delle tecnologie digitali e di quelle basate sull'intelligenza artificiale che, non di rado, vengono utilizzate in maniera poco consapevole e scarsamente critica, contribuendo a rendere ancora più complessa una questione che, del resto, non è mai stata semplice nel paese.

Dalla constatazione della situazione attuale del paese emerge con chiarezza che l'università, in quanto spazio privilegiato di produzione della conoscenza e di riflessione critica sulla società, non può sottrarsi alla necessità di elaborare strategie capaci di affrontare tali sfide. Questa responsabilità appare particolarmente rilevante nei corsi di Lettere e, più in generale, nei percorsi di formazione dei futuri docenti, nei quali la lingua costituisce al tempo stesso oggetto di studio, strumento di costruzione del sapere e mezzo fondamentale di partecipazione alla vita sociale. In tale quadro, l'università è chiamata a riflettere non soltanto sulle proprie pratiche di insegnamen-

to, ma anche sul ruolo che può svolgere nella formazione di individui, che devono essere capaci di osservare, interpretare e analizzare criticamente le informazioni e i fenomeni con cui entrano quotidianamente in contatto. Diventa quindi necessario individuare percorsi di innovazione che favoriscano non soltanto lo sviluppo della competenza comunicativa degli apprendenti, nella lingua madre e nelle lingue straniere, ma anche il rafforzamento delle capacità di ragionamento, senza trascurare le profonde trasformazioni tecnologiche che caratterizzano il nuovo millennio.

In questo contesto, l'insegnamento delle lingue straniere rappresenta un terreno particolarmente fertile, dato che i processi di osservazione, interpretazione e analisi possono infatti trovare un punto di partenza nelle conoscenze che gli apprendenti già possiedono della propria lingua per poi estendersi all'esplorazione di lingue e culture altre. Tale processo favorisce già di per sé lo sviluppo del confronto, della riflessione metalinguistica e della presa di coscienza delle similitudini e delle differenze tra sistemi linguistici e culturali, contribuendo alla formazione intellettuale e alla maturazione degli apprendenti in quanto soggetti più consapevoli.

Bisogna sottolineare che nel presente contributo il concetto di innovazione non intende focalizzarsi semplicemente sull'introduzione in classe di tecnologie o risorse di ultima generazione — sebbene tali strumenti possano eventualmente far parte del processo —, bensì sulla necessità di ripensare le modalità attraverso cui vengono svolte attività tradizionali, come appunto l'insegnamento. L'innovazione viene qui concepita — in linea con quanto sostenuto da Plonski (2017), come un insieme strutturato di azioni e operazioni orientate al raggiungimento di un determinato obiettivo. In questa ottica, con il termine “*innovazione*” ci si riferisce a un processo — e, al tempo stesso, al suo risultato — che implica la creazione di nuove possibilità rispetto a scenari storicamente consolidati, attraverso soluzioni, tecnologiche o meno, che possono essere sperimentate, valutate, adattate e progressivamente riformulate alla luce dei risultati ottenuti.

Inserita in questa prospettiva, la presente proposta cerca di riflettere sul possibile sviluppo di un'agenda *data driven* nell'insegnamento dell'italiano in Brasile. Più specificamente, si discute come l'integrazione sistematica di dati linguistici autentici possa contribuire non soltanto all'innovazione delle pratiche didattiche, ma anche allo sviluppo della consapevolezza (meta)linguistica degli apprendenti.

2. DAL BIG DATA AL DATA-DRIVEN LEARNING

Viviamo in una società sempre più orientata ai dati, tanto che l'attuale fase storica viene definita da molti come l'era dei *Big Data* (Cfr. Mayer-Schönberger, Ramge, Maugeri, 2018; Leonelli, 2018). In questo contesto, la cosiddetta *datafication* (o datificazione), ossia il processo attraverso il quale aspetti sempre più ampi della vita individuale e collettiva vengono trasformati in dati quantificabili e analizzabili mediante strumenti computazionali (Mayer-Schönberger & Cukier, 2013: 78), rappre-

senta una delle trasformazioni più significative della contemporaneità. Secondo gli autori, tale fenomeno avrà un impatto paragonabile a quello prodotto dalla rivoluzione industriale, incidendo profondamente su come gli individui pensano, producono conoscenza, prendono decisioni, interagiscono con l'ambiente e organizzano la vita sociale.

Le trasformazioni generate dall'enorme volume di dati oggi disponibile non si limitano ad un singolo ambito disciplinare. Nell'informatica, ad esempio, stanno diventando sempre più centrali settori quali l'intelligenza artificiale e il *machine learning*, nei quali il valore del dato risiede nella possibilità di essere trasformato in conoscenza utile alla comprensione di fenomeni complessi e al processo decisionale. Nell'economia e nella finanza, i dati favoriscono l'aumento dell'uso di modelli predittivi sempre più sofisticati, capaci di supportare investitori e organizzazioni nell'identificazione di rischi, opportunità e tendenze di mercato. Nel marketing, la datificazione delle abitudini e dello stile di vita dei consumatori ha trasformato le strategie di studio dei comportamenti di consumo, consentendo la progettazione di prodotti e servizi sempre più mirati a specifici profili. In ambito medico, i dati svolgono un ruolo fondamentale sia nei processi diagnostici sia nella produzione di evidenze utili all'interpretazione di fenomeni biologici e clinici. Infine, anche nel settore educativo, si osserva una crescente valorizzazione dei dati, che vengono impiegati per monitorare i processi di apprendimento, progettare percorsi didattici e accompagnare lo sviluppo dell'apprendimento degli studenti.

Nonostante la crescente centralità dei dati negli ambiti sopra menzionati, il loro impiego come principio guida nella progettazione dei percorsi glottodidattici rimane ancora relativamente limitato, soprattutto nel contesto brasiliano, dove le esperienze sono ancora circoscritte all'ambito dell'insegnamento dell'inglese e si configurano più come iniziative isolate o sperimentali piuttosto che come pratiche sistematiche.

Appare pertanto particolarmente importante avviare questa riflessione anche nel campo dell'insegnamento della lingua italiana in Brasile, poiché, oggi più che mai, la *data literacy*, intesa come la capacità di osservare, interpretare e analizzare criticamente i dati, rappresenta una competenza necessaria per qualsiasi percorso formativo che intenda dialogare con la realtà contemporanea e confrontarsi con le molteplici dimensioni della sua complessità.

La crescente disponibilità di traduttori automatici e di strumenti di intelligenza artificiale capaci di generare in pochi secondi testi integrali su pressoché qualsiasi argomento sta determinando cambiamenti profondi nei corsi universitari di Lettere. Tali trasformazioni non riguardano soltanto il rapporto degli apprendenti con le lingue (sia la lingua materna sia la lingua oggetto di apprendimento), ma investono anche il modo stesso in cui gli attori coinvolti concepiscono le nozioni di competenza comunicativa, di lezione e di valutazione.

In questo scenario, tecniche tradizionali di presentazione grammaticale, in particolare quelle di tipo deduttivo, nelle quali il docente continua a svolgere un ruo-

lo centrale, tendono a rivelarsi – se impiegate da sole – sempre meno adeguate a stimolare la motivazione e i processi di ragionamento complessi. Diventa, quindi, necessario individuare percorsi didattici innovativi capaci non solo di coinvolgere gli studenti in pratiche più attive di osservazione, interpretazione e costruzione autonoma della conoscenza linguistica, ma anche di promuovere un apprendimento più significativo e duraturo.

A questo scopo l'idea di utilizzare dati autentici nell'insegnamento delle lingue non costituisce certamente una novità assoluta. Già negli anni 1990, infatti, Tim Johns si interrogava su come portare in classe dati linguistici autentici senza sovraccaricare le risorse cognitive e linguistiche degli apprendenti di una lingua seconda. Come possibile risposta a questa sfida, propose allora il Data-Driven Learning (DDL), una prospettiva didattica che si avvale del lavoro con corpora digitali interrogabili attraverso strumenti computazionali specifici¹ e che si fonda sulla premessa che gli apprendenti debbano assumere un ruolo più attivo nel processo di apprendimento attraverso l'osservazione diretta dei dati linguistici. In questa prospettiva il vantaggio centrale del DDL risiede nel:

«to give the learner direct access to the data, the underlying assumption being that effective language learning is a form of linguistic research, and that the concordance printout offers a unique way of stimulating inductive learning strategies – in particular the strategies of perceiving similarities and differences and of hypothesis formation and testing» (Johns, 1994: 30-1).²

Di conseguenza i dati non vengono utilizzati come supporti illustrativi alle regole grammaticali, bensì come punto di partenza per l'indagine su frequenze, regolarità distributive e contesti d'uso linguistici, aspetti che raramente trovano spazio nei corsi tradizionali di lingua. L'insegnamento viene così inteso come un processo di scoperta guidata, nel quale l'osservazione dei dati precede e orienta la formulazione di ipotesi sul funzionamento della lingua.

Come si evince dalla citazione riportata sopra, a causa delle limitazioni tecnologiche dell'epoca, nelle prime formulazioni di Johns, le attività del DDL si basavano prevalentemente sul lavoro con le linee di concordanza presentate in formato carta-

1 Se negli anni Novanta tali strumenti erano rappresentati da software come *MicroConcord*, oggi docenti e ricercatori dispongono di risorse molto sofisticate, tra cui Sketch Engine, WordSmith Tools e Antconc), capaci di generare liste di concordanze, lista di parole, tra tante altre possibilità.

2 «offrire all'apprendente un accesso diretto ai dati, partendo dal presupposto che un apprendimento linguistico effettivo costituisca una forma di ricerca linguistica e che le linee di concordanza cartacee rappresentino uno strumento privilegiato per stimolare strategie di apprendimento induttivo, in particolare quelle legate alla percezione di somiglianze e differenze, nonché alla formulazione e al test di ipotesi» (Johns, 1994).

ceo. Con il progressivo sviluppo degli strumenti informatici, che ha consentito una maggiore democratizzazione dell'accesso a internet e a *corpora* digitali, le possibilità applicative di questo approccio si sono ampliate notevolmente, favorendo un incremento dell'uso dei corpora nell'insegnamento delle lingue straniere.

In questa direzione, Boulton (2017) individua almeno due filoni principali delle ricerche empiriche sul DDL. Da un lato, numerosi studi hanno indagato atteggiamenti e percezioni degli apprendenti durante l'insegnamento attraverso i corpora. Tali ricerche – orientate a metodologie di tipo qualitativo – analizzano strumenti quali interviste, registri di apprendimento e questionari. Dall'altro lato, gli studi di carattere sperimentale hanno analizzato gli effetti del DDL sui risultati di apprendimento, sia quando il *corpus* viene utilizzato come supporto all'acquisizione di nuove conoscenze linguistiche, sia quando viene impiegato come risorsa di consultazione. In questo secondo filone, i procedimenti metodologici prevedono frequentemente il confronto tra *pré-test* e *post-test* in gruppi sperimentali e di controllo. Pur in presenza di differenze metodologiche e di risultati non sempre uniformi, emerge una tendenza generale a riconoscere il potenziale del lavoro con corpora nello sviluppo della consapevolezza linguistica, della capacità di osservazione dei fenomeni di lingua e dell'autonomia degli apprendenti nella costruzione della conoscenza.

Nonostante tutta l'evoluzione del DDL registrata negli ultimi anni, uno degli aspetti ancora rilevanti riguarda la definizione di “uso di *corpora*” in contesto didattico. Esistono infatti molteplici livelli di integrazione del dato nell'effettiva pratica didattica. Si può, ad esempio, promuovere l'osservazione di esempi autentici, guidando gli studenti in processi inferenziali nei quali devono notare frequenze e regolarità d'uso – come tra l'altro è proposto dallo stesso Johns. In altri casi, l'attività può prevedere la costruzione di piccoli corpora linguistici e la successiva annotazione manuale dei dati, una pratica che tende a favorire la riflessione metalinguistica e la presa di coscienza delle relazioni tra forma e funzione. Analogamente, gli studenti possono essere coinvolti nella revisione di annotazioni automatiche oppure nell'esplorazione autonoma dei corpora mediante la generazione e l'analisi di linee di concordanza, orientando l'indagine verso fenomeni linguistici di loro interesse. Al di là delle differenze, ciò che accomuna comunque tutte queste diverse tecniche è l'idea che la lingua debba essere osservata nel suo funzionamento concreto, superando così una concezione normativa e astratta del sistema linguistico.

3. PROPOSTA METODOLOGICA

Sebbene la revisione della bibliografia internazionale metta in evidenza un aumento del numero di ricerche condotte nell'ambito del DDL (Boulton, Vyatkina, 2021; Boulton, 2017), concentrando l'attenzione sul contesto brasiliano, è possibile individuare due lacune principali: una di natura linguistica e un'altra di natura metodologica. Per quanto riguarda la prima, le esperienze didattiche ispirate al DDL risultano ancora prevalentemente circoscritte all'insegnamento dell'inglese (Pinto, Marques, Garcia, 2025, Tagnin, Murakami, 2014; Santos, 2011; Dayrell, 2011), con alcune eccezioni rappresentate da studi dedicati al francese (Zavaglia, Celli, 2011), allo spagnolo (Pagano; Figueredo 2011), al tedesco (Aquino, 2018, 2019; Finatto, Zilio, Migotto, 2011;), al portoghese (Oliveira, 2005; Zhang, You, 2024 Martins, 2022) e all'italiano (Frangiotti, 2021). Dal punto di vista metodologico, molti di questi contributi si concentrano sulla presentazione di alternative su come integrare il DDL nelle lezioni di lingua. Risultano invece meno frequenti gli studi che descrivono e analizzano esperienze didattiche effettivamente realizzate in classe, discutendone i processi di implementazione, le difficoltà operative e i risultati ottenuti (Mareco; Maracci, 2010; Pinto; Souza, 2011; Frangiotti, 2021; Oliveira, Ottaiano, 2017; Carcia, 2002; Zuppari, 2017).

Alla luce di questo scenario, emerge che nell'insegnamento dell'italiano in Brasile non si è ancora consolidato un impiego sistematico dei corpora. Questa constatazione invita a formulare iniziative che, oltre ad ampliare la presenza del DDL nell'insegnamento dell'italiano, documentino e analizzino gli effetti dell'uso dei corpora sull'apprendimento linguistico. Proprio per rispondere alle lacune individuate, il presente contributo propone una prospettiva nella quale il dato linguistico assume il ruolo di principio organizzatore dell'intero percorso di ricerca e intervento didattico. La proposta, sintetizzata nell'immagine seguente, si fonda sull'idea che l'osservazione sistematica dei dati linguistici possa costituire il punto di partenza per la descrizione dei fenomeni di lingua, la progettazione di materiali didattici e la successiva verifica della loro efficacia in contesto educativo.

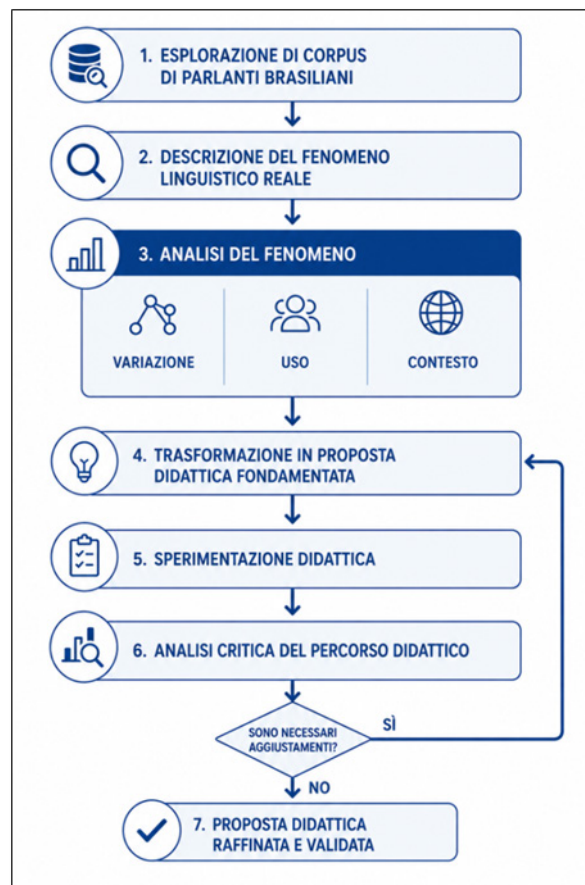


fig. 1. Flusso della proposta di agenda *Data Driven*

Si parte dal presupposto che molte delle difficoltà osservabili nell'apprendimento dell'italiano da parte di parlanti brasiliani non possano essere comprese soltanto attraverso l'intuizione, per cui fenomeni quali l'uso degli articoli determinativi e indeterminativi, dei pronomi clitici, dei tempi verbali e delle strutture pragmatiche e discorsive richiedono una maggiore attenzione. È in questa direzione che la prima fase dell'agenda di ricerca qui proposta prevede la costruzione di un corpus di apprendenti brasiliani di italiano, a partire dal quale si condurranno analisi degli errori degli studenti, con l'obiettivo di individuare quegli aspetti linguistici che presentano maggiore criticità e che, pertanto, meritano maggiore attenzione didattica.

Successivamente, i dati raccolti saranno sottoposti a una fase di analisi linguistica orientata all'identificazione di frequenze, distribuzioni e *pattern* di uso. In questa fase, particolare importanza assumerà il confronto tra il funzionamento dell'interlingua degli apprendenti e l'uso dell'italiano contemporaneo osservabile in corpora di parlanti nativi quali Coris/Codis e Paisà, entrambi dell'Università di Bologna. L'obiettivo consiste nel comprendere i meccanismi attraverso cui gli apprendenti costru-

iscono il proprio sistema linguistico e nel verificare in che misura questo processo converga verso o si discosti dagli usi attestati nella produzione dei parlanti italiani.

A partire dai risultati ottenuti grazie a queste osservazioni, sarà quindi possibile elaborare materiali didattici sensibili all'uso reale della lingua italiana. Invece di limitarsi alla presentazione di esempi artificialmente costruiti, le attività saranno sviluppate secondo i principi del DDL. In particolare, potranno prevedere l'esame di linee di concordanza e liste di frequenza di parole, la realizzazione di annotazioni manuali e la revisione di annotazioni automatiche. Queste attività saranno formulate come veri *task*, nei quali lo scopo fondamentale sarà stimolare processi inferenziali e forme di lavoro collaborativo, favorendo una comprensione più profonda delle caratteristiche linguistiche in differenti contesti d'uso.

La fase successiva prevede la sperimentazione dei materiali in contesto reale di insegnamento. Attraverso l'osservazione sistematica in classe, l'analisi delle produzioni degli studenti e la raccolta di dati relativi alle loro percezioni ed esperienze, diventerà possibile valutare gli effetti delle pratiche *data-driven* sul processo di apprendimento. Con queste informazioni si avrà una base importante per la riflessione critica sull'efficacia del percorso svolto e per eventuali riformulazioni dei materiali e delle tecniche didattiche applicate.

Le diverse tappe descritte daranno origine a un processo ciclico e dinamico che integrerà ricerca, sperimentazione in classe e revisione continua. Lo sviluppo di una simile agenda di ricerca richiede necessariamente la costruzione di reti collaborative internazionali. La formazione e l'analisi di *corpora* di apprendenti, ad esempio, con la condivisione di metodologie di annotazione, la costruzione di strumenti di analisi e la formulazione di pratiche didattiche orientate ai dati costituiscono infatti attività che possono trarre grande vantaggio dal dialogo tra istituzioni e gruppi di ricerca differenti.

In questo senso, la collaborazione tra università brasiliane e italiane appare particolarmente promettente. Da un lato, il Brasile rappresenta un contesto privilegiato per l'osservazione dell'interlingua nell'apprendimento dell'italiano e per l'implementazione di tecniche didattiche orientate al DDL; dall'altro, numerose istituzioni italiane possiedono una lunga tradizione di ricerca nell'ambito dell'italiano L2/LS, della linguistica dei corpora e della formazione docente.

La costruzione di progetti condivisi può favorire non soltanto la mobilità accademica e la produzione scientifica internazionale, ma anche la creazione di infrastrutture di ricerca ancora relativamente poco sviluppate, come corpora sistematici di apprendenti di italiano. Parallelamente, tali collaborazioni possono contribuire alla formulazione di pratiche didattiche più informate ed efficaci, maggiormente sensibili al funzionamento reale della lingua e alle specificità linguistiche degli apprendenti brasiliani.

4. CONSIDERAZIONI FINALI

In un momento storico sempre più orientato ai dati, anche gli studi linguistici e l'insegnamento delle lingue sono chiamati a ripensare il ruolo dell'osservazione empirica nella costruzione della conoscenza linguistica. Nel caso dell'italiano in Brasile, una prospettiva *data-driven* potrà rappresentare non soltanto una possibilità di innovazione metodologica, ma anche un'importante opportunità per integrare descrizione linguistica, ricerca sull'interlingua e pratica didattica.

Partire dal dato linguistico significa riconoscere la dinamicità, la complessità e la variabilità della lingua in uso, aspetti che non devono essere esclusi dalla lezione di lingua perché ritenuti "difficili" da insegnare. Anzi, il compito docente nel terzo millennio si colloca proprio come quello di stimolare la riflessione e di guidare gli studenti perché diventino sempre più capaci di osservare e leggere in modo critico la propria realtà. Il lavoro con i dati allora intende alimentare un circuito virtuoso tra descrizione che genera comprensione e comprensione che porta a procedimenti didattici più efficaci. Più che proporre semplicemente nuove tecniche didattiche, una simile agenda mira a promuovere pratiche capaci di avvicinare gli apprendenti non soltanto alla lingua italiana, ma anche alla pratica stessa della riflessione scientifica sulle lingue.

RIFERIMENTI BIBLIOGRAFICI

- AQUINO M., *A contribuição da ferramenta de corpus para o ensino das partículas modais denn, doch, halt e wohl*, in *Diálogo das Letras*, 8(1), 2019, pp. 144–158, <<https://10.22297/dl.v8i1.3500>>.
- BOULTON A., VYATKINA N., *Thirty years of data-driven learning: Taking stock and charting new directions over time*, in *Language Learning & Technology*, 25(3), 2021, pp. 66–89.
- BOULTON A., *Research timeline: Corpora in language teaching and learning*, in *Language Teaching*, 50(4), pp. 483–506, <<https://10.1017/S0261444817000167>>.
- DAYRELL C., *Corpora no ensino do inglês acadêmico: padrões léxico-gramaticais em abstracts de pós-graduandos brasileiros*, in Viana V., Tagnin S. (a cura di), *Corpora no ensino de línguas estrangeiras*, São Paulo, Hub Editorial, 2010, pp. 137–171.
- FINATTO M. J. B., ZILIO L., MIGOTTO E. J. *Artigos de cardiologia em português e alemão: contribuições da pesquisa em corpus para o ensino de leitura instrumental*, in Viana V., Tagnin S. (a cura di), *Corpora no ensino de línguas estrangeiras*, São Paulo, Hub Editorial, 2010, pp. 205–234.
- FRANGIOTTI G. A. *O efeito da Linguística de Corpus na formulação de material didático e na promoção de output significativo em aulas de italiano/LE. Tradterm*, 37(1), pp. 119–147. <<https://doi.org/10.11606/issn.2317-9511.v37p119-147>>.
- JOHNS T. *From printout to handout. Grammar and vocabulary teaching in the context of Da-*

- ta-driven Learning*, in *Perspectives on Pedagogical Grammar*, Cambridge, Cambridge University Press, 1994, pp. 28-45.
- LEONELLI S., *La ricerca scientifica nell'Era dei Big Data*, Meltermi, Milano, 2018.
- MARECO R. T. M.; MARACCI, C. C., *Linguística de corpus e ensino de inglês instrumental*, in *Colloquium Humanarum*, 7(1), 2010, pp. 1-7. <<https://10.5747/ch.2010.v07.n01.h072>>.
- MARTINS F. *Corpus computadorizado contribuindo para o ensino e pesquisa no campo da língua portuguesa e literatura*, *Brazilian Journal of Development*, 8(9), 2022, pp. 64888-64902, <<https://doi.org/10.34117/bjdv8n9-297>>.
- MEJIAS U. A., COULDRY N., *Datafication*, in *Internet Policy Review*, 8(4), 2019, <<https://10.14763/2019.4.1428>>.
- OLIVEIRA E. N. *A Linguística de Corpus no ensino-aprendizagem de línguas: um estudo de caso das práticas de professores da educação básica de Dourados (MS)*, tesi di dottorato, Universidade Federal de Santa Catarina: Florianópolis, 2005.
- OLIVEIRA E. C., OTTAIANO A. O., *Harry Potter e a Linguística de corpus em aulas de língua inglesa para o ensino fundamental I*, in *Revista do Programa de Pós-Graduação em Letras da Universidade de Passo Fundo*, 13(3), 2017, pp. 609-627.
- PAGANO A. S., FIGUEREDO G. P. *Gramaticalização da dor em português e espanhol: uma abordagem comparada com subsídios da Linguística de corpus e da linguística sistêmico-funcional*, in Viana V., Tagnin S. (a cura di), *Corpora no ensino de línguas estrangeiras*, São Paulo, Hub Editorial, 2010, pp. 173-187.
- PINTO P. T., MARQUES S. M. K., GARCIA D. N. M. *Corpus e escrita acadêmica: uma estratégia data-driven para a internacionalização da produção científica*, in *Trabalhos de Linguística Aplicada*, 64, 2025, pp. 1-11.
- PINTO M., SOUZA R. C. *Filmes comerciais: uma perspectiva da aplicação de pesquisa em corpus na sala de aula*, in Viana V., Tagnin S. (a cura di), *Corpora no ensino de línguas estrangeiras*, São Paulo, Hub Editorial, 2010, pp. 173-187.
- PLONSKI G. A., *Inovação em transformação*, *Estudos Avançados*, São Paulo, 31 (90), 2017, pp. 7-21.
- MAYER-SCHÖNBERGER V., RAMGE T., MAUGERI Giuseppe (a cura di), *Reinventare il capitalismo nell'era dei big data*, Milano, Egea, 2018.
- MAYER-SCHÖNBERGER, V., CUKIER, K. *Big Data: A revolution that will transform how we live, work and think*, London, John Murray, 2013.
- SANTOS A. G. *Working closely with corpus: análise de colocações adverbiais em inglês para negócios*, in Viana V., Tagnin S. (a cura di), *Corpora no ensino de línguas estrangeiras*, São Paulo, Hub Editorial, 2010, pp. 97-136.
- TAGNIN S., MURAKAMI, D., *Linguística de Corpus e ensino: a compilação de um corpus de especialidade para preparação e implementação de um curso preparatório rápido para exame de proficiência*, in *Letras e Letras*, 30 (2), 2014, pp. 349-366.
- ZAVAGLIA A.; CELLI, M., *Aplicação do conceito de transcategorialidade ao ensino de francês como língua estrangeira: o caso da marca 'mas'*, in Viana V., Tagnin S. (a cura di), *Corpora no ensino de línguas estrangeiras*, São Paulo, Hub Editorial, 2010, pp. 235-267.
- ZHANG J., YOU, M., *Corpus de aprendizes de português da Universidade de Macau e ensino de português L2*, in *Texto Livre*, Belo Horizonte, 17, 2024, pp. 1-16. <<https://doi.org/10.1590/1983-3652.2024.47754>>.
- ZUPPARDI M. C. *O uso de corpus especializado no ensino público de língua inglesa*, *Caderno Seminal Digital*, 1, 2017, pp. 304-338.