NGHIÊN CỨU DỰA TRÊN KHỐI LIỆU: MỘT SỐ ĐIỂM TƯƠNG ĐỒNG GIỮA BÀI PHÁT BIỂU NGHỊ VIỆN VÀ TÁC PHẨM VĂN HỌC VĂN XUÔI

Trần Thu Trang*

Nghiên cứu mô tả quá trình xây dựng hai khối liệu và mô tả những đặc điểm của hai khối liệu: Các bài phát biểu nghị viện của cựu thủ tướng Italia Mario Draghi trong vòng sáu tháng kể từ khi khủng hoảng chính trị Nga-Ucraina nổ ra (tháng 02-08/2022) và tác phẩm văn học văn xuôi "Il sentiero dei nidi di ragno" của nhà văn Italo Calvino về chủ đề chiến tranh sáng tác trong thế kỷ 20. Hai khối liệu được so sánh thông qua công cụ Google Colaboratory, trên nền tảng ngôn ngữ lập trình Python, kết quả của bước so sánh này giúp tác giả giải đáp một số giả thiết về những điểm chung có thể tồn tại giữa hai loại hình văn bản và phân tích đặc điểm của những cặp kết quả có độ tương đồng cao nhất trong hai khối liệu.

Từ khóa: *nghiên cứu dựa trên khối liệu, phát biểu nghị viện, văn xuôi, Google Colaboratory.*

This study describes a compilation of two corpora and their characteristics. One includes the parliamentary speeches delivered by former Italian Prime Minister Mario Draghi in the span of six months starting from the outbreak of the Russian-Ukrainian political crisis (from February to August 2022). The other is a twentieth-century prose work on the theme of war named "Il sentiero dei nidi di ragno" by Italo Calvino. The two corpora are compared thanks to Google Colaboratory which operates based on Python – a programming language. The results are used to clarify the hypotheses around the possible similarities between the two genres of text and to analyze the characteristics of the most relevant pairs in the two corpora.

Keywords: *corpus-based research, parliamentary speeches, prose, Google Colaboratory.*

ANALISI DELLE SIMILITUDINI FRA I DISCORSI PARLAMENTARI E LA NARRATIVA – UNA RICERCA CORPUS-BASED

Introduzione

Con lo sviluppo rapido dell'informatica, l'uso degli strumenti digitali negli studi e nel lavoro gioca un ruolo sempre più importante. Negli ultimi anni abbiamo testimoniato i progressi senza precedenti del mondo digitale che a sua volta influisce significativamente su ogni aspetto della quotidianità e della sfera professionale della quale l'insegnamento e l'apprendimento di lingue non fa eccezione. Tra i molti strumenti esistenti che sostengono gli insegnanti e i ricercatori

* **ThS., Khoa tiếng Italia, Trường Đại học Hà Nội**
Email: trangtt@hanu.edu.vn

negli studi didattici e linguistici non possiamo negare il corpus e la linguistica dei corpora e innumerevoli ricerche scientifiche che si basano sul corpus (*corpus based*) o che sono indotte dal corpus (*corpus driven*). La presente ricerca scientifica, come tutte le ricerche umane, nasce da una curiosità su una possibile similitudine fra il discorso parlamentare (caratterizzato dallo stile accademico e dal controllo grammaticale) e la narrativa (caratterizzato dall'aspetto estetico, dalle figure retoriche e da una scelta libera del registro linguistico). In particolare, abbiamo scelto una serie di discorsi dell'ex primo ministro italiano Mario Draghi sul tema del conflitto militare fra la Russia e l'Ucraina data la scottante attualità. Di stesso interesse in tale contesto è il secondo corpus composto dall'intero romanzo "Il sentiero dei nidi di ragno" dello scrittore italiano Italo Calvino. Il lavoro di paragonare i due corpora ha come l'obiettivo quello di trovare un possibile modello nell'espressione, che è caratteristico nella letteratura italiana, cioè si scrive prendendo come punto di riferimento uno scrittore che aveva modi particolari di esprimere concetti, emozioni, fatti, ecc. Il paper consiste di altre tre sezioni oltre all'Introduzione, alla Conclusione e Bibliografia. Nella sezione 1 parleremo del processo di compilazione dei due corpora. Nella sezione 2 descriveremo lo strumento digitale che usiamo per fare l'interrogazione dei corpora. Nella sezione 3 riassumeremo i principali risultati conseguiti e indicheremo gli sviluppi futuri del nostro lavoro.

1. Compilazione dei corpora

1.1. Considerazioni sulla progettazione

Corpus viene definito da Francis come "*a collection of texts assumed to be representative of a given language, dialect or other subset of a language, to be used for linguistics analysis*" (1982). Esistono altre definizioni del corpus ma troviamo fondamentale nelle molte definizioni l'elemento della "raccolta dei testi", così come nella definizione data da Sinclair "*a collection of naturally-occurring language text, chosen to characterize a state or variety of language*" (1991); nonché in quella offerta da Biber et al. "*a corpus is a large and principled collection of natural texts*" (1998).

I testi che compongono la raccolta, inoltre, devono essere:

a) leggibili a macchina, cioè in formato elettronico;

b) autentici;

c) rappresentativi di una particolare lingua o una varietà linguistica.

Per la progettazione dei corpora utilizzati per le analisi all'interno del paper, scegliamo i princìpi determinati da Sinclair che in generale possono fungere da linee guida sia per corpora dello scritto sia per quelli del parlato. Ne citiamo quelli più rilevanti:

1. I contenuti di un corpus dovrebbero essere selezionati indipendentemente dalla lingua in cui si trovano, ma secondo la loro funzione comunicativa nella comunità in cui sorgono.
2. Qualsiasi informazione su un testo diversa dalla stringa alfanumerica delle sue parole e la punteggiatura dovrebbe essere memorizzata separatamente dal testo normale e unita quando richiesto nelle applicazioni
3. I campioni di linguaggio per un corpus dovrebbero, ove possibile, essere costituiti da interi documenti o trascrizioni di eventi vocali completi, o dovrebbe avvicinarsi il più possibile questo obiettivo possibile. Ciò significa che i campioni differiranno sostanzialmente in termini di dimensioni.
4. La progettazione e la composizione di un corpus dovrebbero essere documentate integralmente con informazioni sui contenuti e le argomentazioni a giustificazione delle decisioni assunte.
5. Qualsiasi controllo sull'argomento in un corpus dovrebbe essere imposto dall'uso di strumenti esterni, e non interni, criteri.
6. Un corpus dovrebbe mirare all'omogeneità nei suoi componenti mantenendo copertura adeguata e testi vuoti dovrebbero essere evitati.

1.2. Progettazione dei corpora

I corpora che intendiamo costruire servono per due scopi:

i. indagare una eventuale similitudine nello stile e nel lessico tra i discorsi sulla crisi Russia-Ucraina dell'ex Presidente del Consiglio italiano Mario Draghi e il romanzo "Il sentiero dei nidi di ragno" di Italo Calvino.

ii. analizzare il comportamento degli elementi più rilevanti che emergono dall'indagine sopra indicata.

Il primo corpus è compilato con nove campioni che sono discorsi dell'ex premier italiano Mario Draghi nell'arco di sei mesi dal febbraio all'agosto 2022, in particolare sono:

1. L'informativa alla Camera dei deputati (25/02/2022)
2. Il discorso in Parlamento sull'Ucraina (01/03/2022)
3. Il discorso in Senato (19/05/2022)
4. Comunicazioni sul conflitto Russia-Ucraina, la replica in Senato (01/03/2022)
5. Il discorso in occasione della visita a Kiev (17/06/2022)
6. Il discorso all'Europarlamento (03/05/2022)
7. Comunicazioni sul conflitto Russia-Ucraina, la replica alla Camera dei deputati (01/03/2022)
8. Dichiarazioni del Presidente del Consiglio (24/02/2022)
9. Intervento del Presidente Draghi al Meeting di Rimini (24/08/2022)

Il secondo corpus è "Il sentiero dei nidi di ragno", il primo romanzo di Italo Calvino, pubblicato nel 1947 dalla casa editrice torinese Einaudi. La scelta del romanzo di Calvino è dovuta al tema della Seconda guerra mondiale e della Resistenza partigiana che Calvino tratta nel suo romanzo.

Prima di tutto, i tre corpora vengono compilati dalla stessa autrice, seguendo tutti i criteri sopra citati, e vengono salvati in formato .txt il quale favorisce l'interrogazione attraverso il corpus reader senza essere distratti da impostazioni particolari. In secondo luogo, bisogna eseguire un trattamento dei dati perché il corpus reader possa comprenderli e leggerli. Il trattamento dei dati innanzitutto consiste nello scomporre i testi all'interno di ciascun corpus per permettere di estrarre informazioni a seconda dello scopo dell'indagine. I più popolari metodi di scomporre i testi sono: lasciare il testo puro, suddividere il testo in paragrafi, suddividerlo in frasi, *tokenizzare* il testo ovvero suddividerlo in singole parole. Dopo di che serve il "ripulire" dei dati: convertire tutte le lettere maiuscole in quelle minuscole correggere gli errori dove necessario.

I testi del primo corpus vengono suddivisi in frasi singole che sono considerate unità del discorso e sono utili per individuare strutture linguistiche. Normalmente si può utilizzare codice sent_tokernize nel Natural Language Toolkit per questo lavoro, ma abbiamo deciso di farlo manualmente perché non è troppo lungo. Per quanto riguarda il secondo corpus, essendo un romanzo, contiene sia la prosa sia stringhe di dialoghi, e la punteggiatura può essere molto ambigua, il che ci ha costretto a eseguire manualmente anche questa fase. Come il primo, la prosa del romanzo viene suddivisa in frasi, delle quali ciascuna frase occupa una singola riga. Le proposizioni che introducono i discorsi diretti vengono separate e collocate in righe diverse. I segni caratteristici delle battute nella narrativa sono omessi per non confondere l'indagine dei testi.

2. Metodo di indagine e interrogazione dei corpora:

Secondo le stime del settore informatico, più del 80% dei dati esiste in formato non strutturato, i quali possono essere testi, messaggi, audio, video, ecc. Dati, nel senso largo, sono tutto quello che parliamo, scriviamo, twittiamo su piattaforme di social network, sulle applicazioni di messaggi, sulle applicazioni di commercio elettronico e in molte altre attività, e la maggior parte di tali dati è in formato testuale. Perché bisogna analizzare i dati non strutturati? È perché la maggior parte di informazioni utili risiedono all'interno di vari tipi di dati non strutturati. Sbloccarli gioca un ruolo importante nel comprendere idee, opinioni, conoscenze per motivi di ricerca o presa di decisioni. Al fine di liberare il potenziale dei dati testuali, in questo lavoro utilizziamo il *Natural Language Processing*, noto come NLP, accompagnato da *machine learning* e *deep learning*. Innanzitutto, bisogna dare una breve definizione del NLP. Come è ben noto a tutti, i computer o gli algoritmi non

sono capaci di comprendere testi o caratteri, ciò significa che è necessario convertire i dati testuali in formato comprensibile e leggibile dal computer (in numero binario, ad esempio) per il lavoro di indagine più tardi. Le applicazioni di NLP possono svolgere molti tipi di indagine quali analisi di sentimento (capire il sentimento dei clienti verso un prodotto), ricerca di modelli tipici, classificazione di lamentele/e-mail/prodotti dell'e-commerce, ecc. Nel presente lavoro, quello che cerchiamo di fare è il rilevamento della lingua per analizzare i punti in comune fra diversi corpora.

Come abbiamo menzionato in precedenza, il computer non legge i testi come li leggiamo noi esseri umani. Per analizzare i testi a esso forniti, occorre lavorare sugli algoritmi che operano su uno spazio di funzionalità numeriche prevedendo l'input come una matrice a due dimensioni in cui le righe sono istanze e le colonne sono funzionalità. Per eseguire il machine learning sul testo, dobbiamo trasformare i testi e rappresentarli in *vector* tali da poterli applicare in machine learning numerico. Questo processo è chiamato vettorizzazione ed è un primo passo essenziale verso l'analisi linguistica. La rappresentazione numerica dei testi ci dà la capacità di eseguire analisi significative e crea anche le istanze su cui operano gli algoritmi di machine learning. Nella fase analitica, le istanze sono interi testi o solo espressioni che possono variare in lunghezza, da citazioni o tweet a interi libri, ma i cui vettori sono sempre di un'uniforme lunghezza. Ogni proprietà della rappresentazione vettoriale è una caratteristica (*feature* in inglese). Per il testo, le caratteristiche rappresentano attributi e proprietà del testo, inclusi il contenuto e gli attributi, per esempio la lunghezza del testo, l'autore, la fonte, la data di pubblicazione. Quando messi insieme, gli attributi descrivono il testo in uno spazio multidimensionale su cui sono applicati metodi di *machine learning*. Per questo motivo, è necessario che pensiamo al linguaggio in modo diverso: da una sequenza di parole a punti che occupano uno spazio semantico ad alta dimensione. I punti nello spazio possono essere vicini o distanti tra loro, strettamente raggruppati o uniformemente suddivisi. Lo spazio semantico è quindi mappato in modo tale che i testi con significati simili sono più vicini e quelli diversi sono più distanti. Codificando la similitudine con distanza, è possibile derivare le componenti primarie del testo e tracciare confini nello spazio semantico.

Per indagare le possibili similitudini fra i corpora, ricorriamo al Colaboratory, in breve "Colab", un prodotto della Google Research. Colab consente a chiunque di scrivere ed eseguire codice Python arbitrario tramite il browser ed è particolarmente adatto per il *machine learning*, l'analisi dei dati e l'istruzione. Quanto all'ipotesi cui cerchiamo di dare una risposta, Colab permette di calcolare la distanza mettendo a confronto ciascuna di tutte le stringhe del primo corpus e ciascuna stringa del secondo corpus.

Il corpus dei discorsi di Mario Draghi contiene un totale di 774 frasi, il secondo

dell'opera *Il sentiero dei nidi di ragno* di Italo Calvino 3750 frasi che sono collocate secondo la seguente regola:

corpus_1 = ["frase 1",
"frase 2",
...
"frase n"]

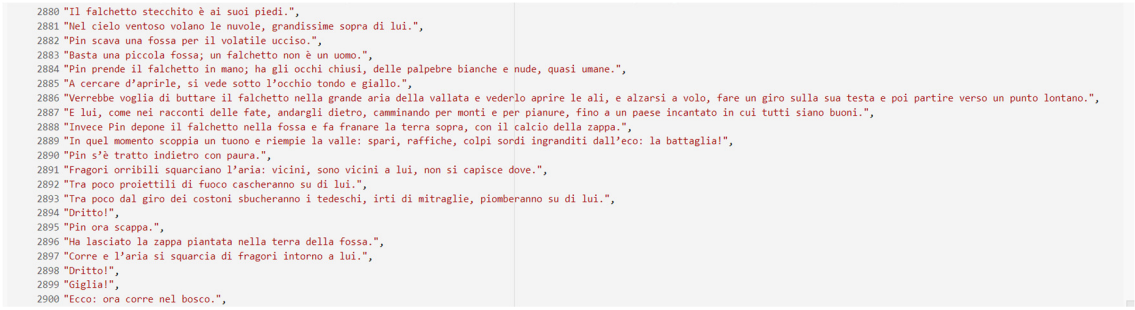

Figura 1. Disposizione delle frasi nei corpus

3. Analisi dei dati:

Il sistema ha eseguito un totale di 2,902,500 di calcoli di distanza fra le frasi mettendo a confronto il *vector* di ciascuna frase del corpus 1 e quello di ciascuna frase del corpus 2. La distanza viene presentata in figure come nell'immagine seguente:

		corpus_2 s1	corpus_2 s2	corpus_2 s3	corpus_2 s4	corpus_2 s5	corpus_2 s6	corpus_2 s7	corpus_2 s8	corpus_2 s9	corpus_2 s10
corpus_1	s1	3.1898441	3.307311	3.26069	3.2203007	3.060756	3.3536947	2.9156997	3.1000583	3.2549982	2.8450985
corpus_1	s2	2.2390983	2.5381072	2.1509824	3.880917	2.6080072	3.5167139	2.847601	3.0911596	2.8759077	2.6522205
corpus_1	s3	2.1241732	2.366501	1.991255	4.035946	2.6651387	3.5321896	2.7722654	3.039426	2.8906558	2.6217632
corpus_1	s4	2.824862	3.100448	2.8896976	3.1159482	2.499878	2.7651362	2.6213503	2.4655795	2.5576837	2.5648806
corpus_1	s5	1.798314	2.0958216	1.9009783	4.078424	2.8180237	3.6907697	2.8180983	3.176356	2.9515772	2.5587308
corpus_1	s6	2.0236719	2.3784122	2.1256115	3.955623	2.7557063	3.6028547	2.8537729	3.130819	2.7748494	2.6087146
corpus_1	s7	2.0363784	2.223336	2.0132098	3.9230103	2.6632545	3.5306642	2.704423	3.0264416	2.7585647	2.5583057
corpus_1	s8	2.1398578	2.3785105	2.197403	3.9474077	2.8221202	3.609073	2.7759202	3.0899975	2.950095	2.621694
corpus_1	s9	2.022441	2.3190043	2.0566673	4.1313267	2.8948247	3.7334032	2.9268942	3.191899	2.8492656	2.716212
corpus_1	s10	2.290627	2.588303	2.3029559	3.9144244	2.8009183	3.568547	2.8671978	3.0171118	2.937882	2.4755938

Figura 2. Risultati dei calcoli di distanza fra i *vector*

In questa fase siamo costretti ad interrogare due volte i corpora. Quando abbiamo analizzato i risultati della prima interrogazione, i risultati sono quelli che seguono.

valore	**totale di risultati**
dallo 0 allo 0.9	531
dal 1 al 1.9	123.239
dal 2 al 2.9	1.695.900
dal 3 al 3.9	956.947
dal 4 al 4.9	110.702
dal 5 al 5.9	12.236
dal 6 al 6.9	1.533
dal 7 al 7.9	1.407
dal 8 al 8.9	5

Tabella 1. Calcolo dei risultati della prima interrogazione suddivisi in gruppi di valore

Minore è la figura, più "vicine" sono le due frasi dal punto di vista sintattico. Il valore più basso è appena superiore allo 0, mentre quello più alto supera 8, precisamente 8.257049.

Dovuto al limite di spazio e di tempo, all'interno del presente paper analizziamo le frasi la cui distanza è minore e maggiore,

cioè le coppie di frasi che hanno la distanza dallo zero allo 0.9 e quelle che hanno la distanza tra l'8 e l'8.9.

Tra i risultati delle coppie di frasi meno distanti, se ne registrano 116 dalla distanza 0.3, 54 dalla distanza 0.4, 20 dalla distanza 0.5, 49 dalla distanza 0.6, 18 dalla distanza 0.7, 10 dalla distanza 0.8 e infine 264 dalla distanza 0.9.

Analizzando le sopra indicate frasi dei due corpus, abbiamo notato che la distanza calcolata è basata sulla forma dell'enunciato, ovvero la struttura formale di questo. Dunque, le frasi brevi e brevissime hanno più possibilità di trovare quelle "vicine". Infatti, nel corpus 1 sono sei le frasi che contengono una sola parola, tra le quali cinque "Grazie." e una "No.". Corrispondono quasi assolutamente alla "No." sono le identiche 'No." nel secondo corpus. Corrispondono maggiormente alle frasi 267, 477, 549, 575, 771 e 774 del corpus 1 dal punto di vista della forma sono le frasi "Basta.", "chiede.", "Così.", "Ben.", "dicono.", "dice.", "Guarda.". Un altro aspetto di cui ci accorgiamo è che tutte le frasi sono affermative e non si differenziano nella parte del discorso. La differenza nel numero delle sillabe fa aumentare un pochino la distanza fra le frasi, mentre il tipo di annuncio definito dal punteggio che lo conclude allontana notevolmente le frasi. Prendiamo in considerazione due annunci: nel primo vogliamo evidenziare la differenza tra la frase 477 "Grazie." del corpus 1 e le frasi 155 "dice." e 155 "dice," e nel secondo la frase 267 "Grazie." del corpus 1 e la mettiamo a confronto con la frase 3262 "Sputa," e quella 3263 "dicono.". Mentre nel primo esempio, la distanza fra le due coppie è rispettivamente 0.9592644 e 1.8635465, la distanza fra la 3263 e la 267 è 0.38764155 e la distanza fra la 2362 e la 267 aumenta quasi di dieci volte (esattamente 3.775938).

Tra i risultati conseguiti, abbiamo notato che le frasi del corpus 2 che sono più distanti con una maggior parte delle frasi del corpus 1 in base alla punteggiatura, cioè un annuncio più distante dall'altro se dall'aspetto formale sono tipi di frase diversi: affermativa, esclamativa e interrogativa oppure contengono punteggiature di natura completamente diversa. In particolare, il valore di distanza supera l'8 si registra fra la frase 104 del primo corpus ("Grazie" - senza alcuna punteggiatura) e le frasi 1436 ("... è perché,"), 2255 ("Uno dei nostri ha tradito,"), 2404 ("Non è così,"), 2406 ("Non è così,"), 2672 ("Tiragli il collo,").

Quindi, riteniamo necessario aggiungere un punto alla frase 104 per segnare non solo la fine della frase ma anche la fine del discorso.

I risultati della seconda interrogazione sono seguenti:

valore	totale di risultati
dallo 0 allo 0.9	620
dal 1 al 1.9	123.342
dal 2 al 2.9	1.696.795
dal 3 al 3.9	959.194
dal 4 al 4.9	110.877
dal 5 al 5.9	11.648
dal 6 al 6.9	23
dal 7 al 7.9	0

dal 8 al 8.9	0

Tabella 2. Calcolo dei risultati della seconda interrogazione suddivisi in gruppi di valore

Emergono nei risultati della distanza dal 6 al 6.9 i tratti che caratterizzano la differenza principale fra le frasi. Vale a dire che rimane sempre nelle tipologie di frasi e anche nel numero dei costituenti delle frasi. Infatti, la frase 2198 del corpus 2 ("Vedrete," - che serve per attirare l'attenzione dell'interlocutore piuttosto che esprimere un'idea) si distingue maggiormente da 23 frasi del corpus 1 per le ragioni sopra indicate. Da una parte si tratta di un elemento che ha la funzione comunicativa, dall'altra parte si tratta di frasi complesse con più proposizioni coordinate o subordinate, più complementi.

L'analisi che abbiamo fatto occupa uno spazio molto modesto rispetto ai risultati che il programma ha calcolato. Prima di tutto riteniamo necessario rivedere la compilazione del corpus 2 in quanto la narrativa in prosa dispone di discorsi diretti spezzati in più parti che sono collegate "dal verbo *dire* o da verbi analoghi come *sostenere, affermare, dichiarare, chiedere, domandare, rispondere*, cui seguono i due punti e le virgolette o i trattini"[1]. Se i discorsi diretti diventassero completi dal punto di vista di forma e di significato, i risultati sarebbero più esatti. In secondo luogo, bisogna analizzare in modo più approfondito i valori che variano dal 1 al 3.9, poiché dal punto di vista formale le coppie di frasi in queste fasce sono abbastanza identiche e potrebbero fornire informazioni interessanti. In terzo luogo, sarebbe opportuno considerare un'indagine dei paragrafi invece che delle singole frasi per affermare se ci fossero idee analoghe fra il politico e lo scrittore o se il primo imparasse ad esprimere una qualche idea dal secondo. In quarto luogo, lo studio potrebbe essere eseguito su una scala maggiore aggiungendo dati al corpus 1 e altre opere letterarie sul tema guerra dello stesso autore e anche di altri autori.

Conclusione

Gli studi effettuati hanno consentito di condurre un'indagine di due tipologie di testo diverse anche se entrambe sono in prosa. Il risultato più rilevante che i risultati sono riusciti a chiarire è trovare il carattere delle frasi vicine e di quelle distanti analizzate sotto l'ottica del linguaggio computazionale. Anche se sull'ipotesi esiste un velo che non ci permette di convalidarne l'esattezza, abbiamo fornito una metodologia valida agli studi di natura comparativa che ricorrono all'uso del corpus. Restano ancora delle lacune da colmare nel futuro prossimo per confermare a pieno l'ipotesi iniziale. Per raggiungerlo, si consiglia di approfondire

[1] https://www.treccani.it/enciclopedia/discorso-diretto_%28La-grammatica-italiana%29/#:~:text=Il%20discorso%20diretto%20riporta%20le,sono%20state%20dette%20o%20scritte.&text=Giulio%20Cesare%20disse%3A%20%C2%ABIl%20dado%20%C3%A8%20tratto!%C2%BB&text=Giulio%20Cesare%20disse%3A%20%C2%ABIl%20dado%20%C3%A8%20tratto!%C2%BB,-%E2%80%93%20in%20un%20inciso&text=%C2%ABIl%20dado%C2%BB%20disse%20Giulio%20Cesare%20%C2%AB%C3%A8%20tratto!%C2%BB

ulteriormente la compilazione dei corpora ed esaminare le frasi dai valori minori.

Nonostante ciò, i lavori svolti possono fungere da modello di metodologia delle ricerche comparative corpus-based, soprattutto nella letteratura. Oltre al Colab Research, esistono altri programmi e siti che consentono di indagare diversi aspetti importanti dei corpora, specialmente quelli di grande dimensione.

Ringraziamenti

Desidero ringraziare i professori che mi ha aperto la strada alla presente ricerca, in particolare la professoressa Dang Thi Phuong Thao (Università di Hanoi) e il professore Franco De Vivo (Università di Tor Vergata). La ricerca è stata condotta in occasione del ventesimo anniversario del Dipartimento di Italianistica dell'Università di Hanoi che segna una pietra miliare non soltanto del nostro Dipartimento ma anche della mia maturità nella professione e nella ricerca scientifica. Ringrazio di cuore i colleghi e gli studenti che mi sostengono.

Riferimento bibliografico

1. Bengford B., Bilbro R. & Ojeda T. (2018). Applied Text Analysis with Python. O'Reilly Media.

2. Biber D., Conrad S. & Reppen R. (1998). Corpus Linguistics. Investigating language structure and use. Cambridge University Press.

3. Calvino I. (1964). Il sentiero dei nidi di ragno. Einaudi.

4. Francis W.N. (1982). Problem of assembling and computerizing large corpus, in Johansson S. (ed.) *Computer Corpora in English Language Research*, Bergen, Norwegian Computing Centre for the Humanities.

5. McEnery T. & Hardie A. (2012). Corpus Linguistics. Method, Theory and Practice. Cambridge University Press.

6. O'Keeffe A. & McCarthy M. (2010). The Routledge Handbook of Corpus Linguistics. Routledge.

7. Sinclair J.M. (1991). Corpus, Concordance, Collocation. Oxford University Press.

Riferimento sitografico

www.treccani.it

dizionario.internazionale.it

https://www.lastampa.it/politica/2022/02/25/news/guerra_in_ucraina_il_discorso_integrale_di_draghi-2862812/

https://www.corriere.it/politica/22_marzo_01/draghi-discorso-integrale-ucraina-44e3e8b2-9941-11ec-9c59-6d8197f09466.shtml

https://www.ilfattoquotidiano.it/2022/05/19/guerra-in-ucraina-il-discorso-integrale-del-presidente-del-consiglio-draghi-in-senato/6597424/

https://www.governo.it/it/articolo/comunicazioni-sul-conflitto-russia-ucraina-la-replica-del-presidente-draghi-senato/19297

https://www.italiaoggi.it/news/kiev-discorso-storico-di-draghi-2566756

https://www.ansa.it/europa/notizie/europarlamento/news/2022/05/03/draghi-a-strasburgo-prima-volta-da-premier-davanti-a-plenaria_5deff6bc-ea1d-46b9-9194-176201450fda.html

https://www.governo.it/it/articolo/comunicazioni-sul-conflitto-russia-ucraina-la-replica-del-presidente-draghi-alla-camera-dei

https://www.governo.it/it/articolo/ucraina-dichiarazioni-del-presidente-draghi/19239

https://www.governo.it/it/articolo/intervento-del-presidente-draghi-al-meeting-di-rimini/20424

(Ngày nhận bài: 25/11/2022; ngày duyệt đăng: 02/02/2023)