

ESERCITAZIONI SKETCH ENGINE

Chiara Ghezzi

LINGUISTICA PER LA VALORIZZAZIONE DEL PATRIMONIO SCRITTO E ORALE a.a. 2023/2024

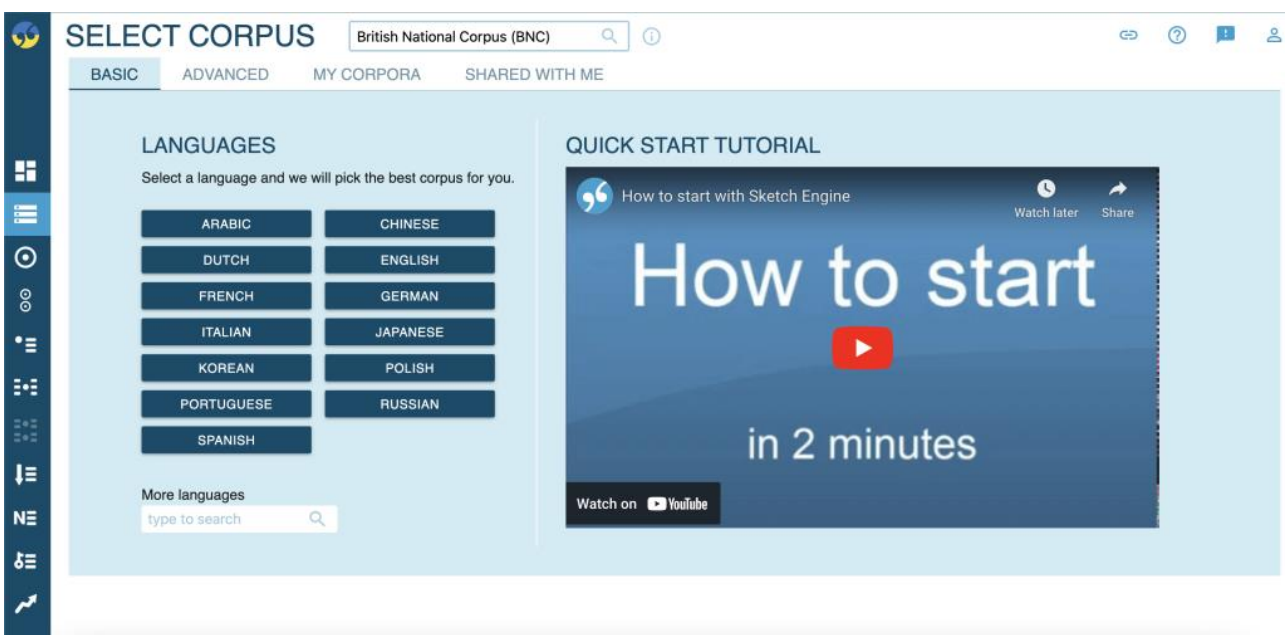
Esercitazione su Sketch Engine (con il British National Corpus)

1 ACCESSO E NAVIGAZIONE NEL SITO

Per utilizzare Sketch Engine, accedere a <https://www.sketchengine.eu> sul browser e fare clic sul pulsante LOG IN nell'angolo superiore sinistro del sito.

Una volta effettuato l'accesso, il sito vi porterà alla pagina SELECT CORPUS. In questa pagina si dovrà selezionare la lingua con cui si desidera lavorare, in modo da restringere le ricerche ai soli corpora in quella lingua.

Per questa esercitazione, ci concentreremo sull'uso dell'inglese nel British National Corpus (BNC). Fare clic su "English" nella pagina iniziale per restringere la ricerca ai corpora in inglese.



Sketch Engine porta quindi alla pagina DASHBOARD, dove il corpus predefinito per l'inglese sarà il corpus English Web 2020 (enTenTen20).

Per utilizzare il British National Corpus, digitate nella casella di ricerca dove è visualizzato "English Web 2020 (enTenTen20)" e digitate "British National Corpus". Fare clic sul primo risultato indicato con il nome corrispondente.

In questo modo verrà caricato il corpus corrispondente su cui effettuare le ricerche. Le funzioni di ricerca elencate nel menu DASHBOARD saranno spiegate nella sezione seguente.

2 FUNZIONI DI RICERCA SU SKETCH ENGINE

La Dashboard di Sketch Engine fornisce un elenco di funzioni che è possibile utilizzare per esplorare un corpus e fornire dati rilevanti con caratteristiche che restringono la ricerca. L'elenco delle funzioni si trova nella pagina DASHBOARD nel primo rettangolo azzurro della pagina con il nome del corpus selezionato come titolo.

Le sezioni seguenti forniscono informazioni su come eseguire ricerche di base con ciascuna delle funzioni di ricerca illustrate.

2.1 CONCORDANZA

La funzione di concordanza di Sketch Engine consente agli utenti di visualizzare una parola chiave in un determinato contesto. Per eseguire una ricerca di concordanza, fare clic sul pulsante "Concordanza" nella pagina della dashboard. In questo modo si accede a una pagina in cui è possibile eseguire semplici ricerche nel corpus selezionato.

2.1.1 Ricerca di base

Per restringere i risultati della ricerca, è possibile fare clic sul cursore accanto a "Text Types" (Tipi di testo) che si trova sotto la barra di inserimento del testo della ricerca semplice. In questo modo vengono elencate le varie caratteristiche del testo/corpus che è possibile specificare per restringere la ricerca.

L'opzione Tipo di testo (prima opzione della sezione Tipi di testo) consente di selezionare il tipo di testo da cui è possibile effettuare la ricerca, ad esempio "Libri scritti e periodici". La barra di ricerca in questa sezione consente di cercare tra tutti i tipi di testo che si possono aggiungere, compresi quelli non elencati immediatamente sotto come predefiniti. Per aggiungere ai criteri di ricerca una qualsiasi caratteristica del tipo di testo elencata, è sufficiente fare clic sull'opzione prevista. È inoltre possibile utilizzare questa opzione per trovare i risultati che iniziano, finiscono, contengono o corrispondono esattamente a una parola, a un morfema, a una lettera, ecc. facendo clic sul simbolo della freccia che punta a sinistra e a destra accanto alla barra di ricerca. Altre opzioni della sezione Tipi di testo sono "Data di pubblicazione", "Luogo di pubblicazione", "Anno di pubblicazione" e altre ancora.

Una volta inserita la parola chiave nella barra di ricerca semplice e selezionati i criteri di ricerca nella sezione Tipi di testo, fare clic sul pulsante "Cerca" in fondo alla pagina per visualizzare i risultati.

I risultati vengono visualizzati per impostazione predefinita nella vista Keyword in Context (KWIC), che mostra la parola cercata evidenziata in rosso all'interno di un contesto sinistro e di un contesto destro. Ciò consente agli utenti di visualizzare la parola in vari contesti e di fare osservazioni in base al lato destro e sinistro della parola. Per vedere il tipo di fonte da cui proviene ogni esempio, controllate il testo dopo il simbolo di informazione, mentre il testo evidenziato in grigio è il tag della fonte:

Un'altra alternativa alla visualizzazione KWIC è la visualizzazione Sentence, che si può attivare facendo clic sul pulsante "KWIC" evidenziato in blu nell'angolo in alto a destra della pagina. In questo modo si attiva il menu a tendina, che presenta "Sentence (frase)" come opzione alternativa a KWIC. Fare clic su questa opzione e apparirà la vista risultante:

Questa vista si concentra sulla costruzione della frase completa che contiene la parola così com'è, piuttosto che allineare la parola cercata in una riga con frasi incomplete.

2.1.2 Ricerca avanzata

Per restringere ulteriormente la ricerca nelle righe della Concordanza, possiamo passare dalla scheda Ricerca di base alla scheda Ricerca avanzata, che si trova nella pagina principale della funzione Concordanza. Direttamente sotto il titolo "CONCORDANCE", si trovano le schede per la ricerca di base e avanzata, intitolate rispettivamente "BASIC" e "ADVANCED". Fare clic sulla scheda "AVANZATA" per avviare la ricerca avanzata.

Sul lato sinistro dello schermo è presente un elenco denominato "Tipo di interrogazione" che consente di scegliere la forma della voce di ricerca che si desidera visualizzare nei risultati. L'opzione "semplice" è la stessa della ricerca predefinita della scheda BASE. L'opzione "lemma" restituisce i risultati per tutte le forme verbali della parola nella barra di ricerca. L'opzione "frase" consente di trovare una frase di più parole digitata nella barra di ricerca. Non vengono considerati lemmi per nessuna delle parole di questa opzione. L'opzione "parola" cerca una forma di parola

così come viene digitata nella barra di ricerca, senza considerare alcun lemma alternativo. L'opzione "carattere" cerca i token che contengono un carattere o un gruppo di caratteri specifici digitati nella barra di ricerca. Infine, l'opzione "CQL" consente agli utenti di effettuare ricerche con il Corpus Query Language, che ha caratteristiche tali da consentire ricerche più complesse per quanto riguarda le strutture lessicali e la formazione grammaticale. In questa guida non si parlerà di CQL, ma alla fine della guida verranno fornite le risorse necessarie per saperne di più.

Sotto l'elenco dei tipi di query si trova il menu a tendina "Subcorpus", che ci permette di scegliere la parte (o le parti) del corpus (se suddiviso in sottocorpora) da cui vogliamo recuperare i risultati. Ad esempio, nel corpus BNC, possiamo restringere la ricerca alle voci del corpus in base a determinati intervalli di anni e alla modalità di raccolta (ad esempio, "frasi audio mp3" o "scritti accademici").

Al di sotto di questa funzione si trova l'opzione "Filtra il contesto", che ci permette di filtrare i risultati in base al numero di lemmi presenti nella nostra ricerca e alle loro parti del discorso. Esistono tre opzioni per questa funzione. L'opzione "NON FILTRARE" è l'opzione predefinita, che consente alla ricerca di produrre risultati come di consueto. L'opzione "CONTESTO LEMMA" restituisce i risultati delle sole righe che contengono o non contengono determinati lemmi se nella barra di ricerca sono stati inseriti uno o più lemmi. È anche possibile specificare la distanza di queste parole dalla KWIC. L'opzione "CONTESTO PARTE DEL PARLATO" restituisce solo le righe che contengono o non contengono determinate parti del discorso a una determinata distanza dal KWIC.

Esercizio 1: Utilizzando la ricerca avanzata Concordance nel BNC, cercate "kick the bucket" senza considerare altri lemmi, ma solo nelle trascrizioni parlate del corpus. Quanti risultati trovate per questa frase?

2.2 LISTA DI PAROLE (WORD LIST)

La funzione Wordlist consente di visualizzare la frequenza di un determinato componente linguistico (parole, lemmi, verbi, ecc.) in un determinato corpus. Per utilizzare questa funzione, fare clic sul pulsante "Wordlist" nella dashboard e apparirà la seguente pagina:

2.2.1 Ricerca di base

Per visualizzare le parole più frequenti nel BNC, selezionare l'opzione "parole" nel primo menu a scorrimento e "tutte" nel secondo menu a scorrimento. Fare clic sul pulsante "VAI" per visualizzare la seguente pagina di risultati:

Questa pagina presenta le parole più utilizzate nel corpus in base alla frequenza. La classifica è riportata nel grafico, insieme alla frequenza totale della parola all'interno del corpus, a destra della parola stessa. Il numero predefinito di parole per pagina è di 50 righe, ma può essere regolato con l'opzione Righe per pagina nell'angolo in basso a destra della pagina.

Questa funzione ci permette anche di restringere ulteriormente la ricerca con le opzioni "che inizia con", "che finisce con" e "che contiene" nel secondo menu a scorrimento della pagina di ricerca principale. Facendo clic su una di queste opzioni, apparirà un'ulteriore barra di ricerca che ci permetterà di inserire il componente desiderato nella nostra ricerca.

2.2.2 Ricerca avanzata

Per restringere la ricerca, è possibile utilizzare le funzioni aggiuntive disponibili nella scheda Ricerca avanzata. Accanto alla scheda di ricerca "BASE" nella parte superiore della pagina, fate clic sulla scheda "AVANZATA" per visualizzare altre funzioni e filtri di ricerca.

Come per la ricerca avanzata della Concorde, è possibile utilizzare il primo elenco di opzioni da sinistra per garantire che i risultati della ricerca riguardino solo determinati criteri (ad esempio, solo lemmi, congiunzioni, sostantivi, ecc.). È possibile selezionare solo un'opzione alla volta da questo elenco. Oltre a questo elenco, l'elenco a destra consente agli utenti di restringere ulteriormente la ricerca a istanze più specifiche del corpus, in base alle indicazioni dell'utente.

Ad esempio, se vogliamo trovare le parole che iniziano con "th" nel corpus, dobbiamo fare clic su "word" nel primo elenco e poi su "starting with" nel secondo elenco. Facendo clic su questa opzione nel secondo elenco, si aprirà una barra di ricerca a destra in cui sarà possibile inserire il delimitatore "th".

È inoltre possibile aggiungere più criteri utilizzando le opzioni del secondo elenco facendo clic sul pulsante "AGGIUNGI CRITERI MULTIPLI" sotto la barra di ricerca. Facendo clic su questo pulsante, si accede a una schermata che mostra un pulsante "+" accanto al primo criterio inserito. Facendo clic su questo pulsante più, la schermata tornerà alla pagina originale con le opzioni del primo e del secondo elenco.

Va notato che l'uso di alcuni criteri può limitare l'uso di altri criteri (ad esempio, "a partire da" può essere usato solo una volta e vieta l'aggiunta di altri criteri come "tutti", "regex corrispondente" e "dall'elenco"). Una volta terminata l'aggiunta di un altro criterio, fate nuovamente clic su "AGGIUNGI CRITERI MULTIPLI" per tornare alla pagina che elenca tutti i criteri scelti.

Sotto i criteri elencati è presente un'opzione che consente di restringere i risultati alle voci del corpus che includono tutti i criteri o solo uno di essi. L'impostazione predefinita di questa opzione è di restituire le voci che corrispondono a tutti i criteri. Per modificare l'opzione in modo da consentire l'inserimento di voci con almeno uno dei criteri, fate clic sul menu a tendina "tutti" e modificalo in "qualsiasi".

A destra dei criteri sono presenti ulteriori funzioni di restringimento. La funzione "Escludi queste parole:" consente di elencare le parole da omettere nei risultati della ricerca, che possono essere inserite nella casella di testo sotto il titolo dell'opzione. Fare clic sulla casella di controllo per assicurarsi che l'elenco di parole da omettere venga applicato alla ricerca. Per questo elenco, inserire una parola per riga affinché Sketch Engine riconosca correttamente ogni parola dell'elenco.

Sotto questa opzione si trova l'opzione "Includi non parole", una casella di controllo che determina se i risultati della ricerca includono parole che non hanno un significato definito nel dizionario. Quando è selezionata, queste "non parole" sono incluse nei risultati della ricerca.

Sotto l'opzione "non parole" si trova l'opzione "A=a", che non tiene conto della maiuscola quando si ottengono i risultati della ricerca. Se questa opzione è disattivata, i risultati della ricerca verranno restituiti indipendentemente dalla loro forma di capitalizzazione (ad esempio, "Eat" e "eat" verranno restituiti entrambi quando si cerca "Eat").

È inoltre possibile regolare le parole che appaiono nei risultati determinando i loro minimi e massimi di frequenza del corpus. Sotto l'opzione "A=a" si trovano le opzioni "Frequenza min" e "Frequenza max", in cui è possibile specificare le volte minime e massime in cui una parola può comparire nel corpus per essere considerata un risultato corrispondente alla nostra ricerca.

L'impostazione predefinita per queste opzioni è un minimo di 5 e un massimo di 0. Per regolarle, è sufficiente digitare il numero di frequenza per ciascun criterio.

Il formato dei risultati può essere regolato dalla funzione sottostante denominata "formato dei risultati". Questa funzione ha due opzioni. "Elenco semplice" mostra i risultati della ricerca nel formato predefinito dell'elenco di frequenza (dal più frequente al meno frequente). La funzione "Visualizza come" mostra i risultati in base al tipo di parola (ad esempio parola, lemma, tag o lemma). È possibile scegliere fino a tre di queste opzioni in una ricerca e selezionare se si desidera che i caratteri maiuscoli e minuscoli siano trattati allo stesso modo.

Infine, è possibile regolare le parti del corpus che si desidera cercare con il menu a discesa "Subcorpus". Come per l'opzione Subcorpus della funzione Concordance, è sufficiente fare clic sul menu a tendina per scegliere quali sottocorpora si desidera considerare come parte della ricerca.

Esercizio 2: Utilizzando la funzione di ricerca avanzata Wordlist, trovare i lemmi più frequenti contenenti "in" ma escludendo le parole "pint" e "point". Escludere le non parole e trattare allo stesso modo le parole maiuscole e minuscole. Mantenere tutte le altre opzioni come predefinite. Quali sono le tre parole più frequenti con questi criteri?

2.3 PAROLE CHIAVE

La funzione Parole chiave consente agli utenti di confrontare due corpora diversi per trovare tratti unici e comuni tra i due, in base a una ricerca.

2.3.1 Ricerca di base

Per utilizzare questa funzione, fare clic sul pulsante "Parole chiave" nella dashboard per accedere a una pagina che spiega brevemente questa funzione. Facendo clic sul pulsante "SO COSA STO FACENDO: VAI" in fondo alla spiegazione, si accede alla pagina seguente.

L'impostazione predefinita di questa funzione è quella di confrontare il corpus correntemente in uso con un altro corpus correlato (in genere basato sulla lingua). Questo corpus di riferimento sarà elencato accanto al pulsante con le frecce blu scuro dopo "corpus di riferimento". La scheda predefinita "SINGLE-WORDS" mostra le 50 parole più comuni tra i due corpora confrontati. Facendo clic sull'altra scheda, "MULTI-WORD TERMS", è possibile visualizzare i cluster di due parole più comuni condivisi tra i corpora.

I risultati di questa ricerca variano a seconda del corpus di riferimento. Nella schermata precedente, il corpus di riferimento è il corpus English Web 2013 (enTenTen13). Il corpus attualmente caricato, il BNC, viene confrontato con questo corpus di riferimento e i risultati vengono visualizzati di conseguenza. Per passare dal corpus di riferimento al corpus da noi scelto (BNC), è sufficiente fare clic sul pulsante con le frecce blu scuro a sinistra del testo "corpus di riferimento:".

2.3.2 Ricerca avanzata

La scheda di ricerca avanzata di Parole chiave ci permette di regolare alcuni aspetti della nostra ricerca. Nella pagina iniziale di Parole chiave, fate clic sulla scheda di ricerca "AVANZATA" accanto alla scheda di ricerca "BASE".

In alto a sinistra della pagina di ricerca avanzata si trovano le opzioni relative al corpus. Con l'opzione "Focus subcorpus", possiamo selezionare quali parti del nostro corpus principale (il corpus selezionato e visualizzato nella barra di ricerca in alto) vogliamo confrontare con il corpus di riferimento. L'opzione "Corpus di riferimento" ci permette di cercare e selezionare dal database di corpora di Sketch Engine un corpus di confronto. L'opzione "Sottocorpus di riferimento" consente di selezionare le parti del corpus di riferimento da confrontare con il corpus principale.

A destra delle opzioni relative al corpus si trovano le opzioni relative alla frequenza. L'opzione "Concentrati su" è una barra scorrevole tra "raro" e "comune", il che significa che se facciamo scorrere la barra più vicino a un lato rispetto all'altro, riceveremo più risultati che sono più frequenti (comuni) o meno frequenti (rari) se la barra è più vicina a un'opzione rispetto all'altra. La frequenza o la rarità sono visualizzate anche numericamente sotto la barra di scorrimento. Quando si regola la barra scorrevole, anche la frequenza viene visualizzata numericamente: la più rara ha una frequenza di 0,001 e la più comune ha una frequenza di 1 milione. Questo strumento cambia semplicemente le parole su cui si concentrerà l'attenzione nei risultati in base alla frequenza del corpus principale rispetto al corpus di riferimento.

È inoltre possibile regolare la frequenza minima e massima che una parola deve avere per essere considerata nell'elenco con le opzioni "Frequenza minima" e "Frequenza massima" sotto la barra "Focus on". La regolazione di questa voce avverrebbe nello stesso modo descritto nella sezione 2.2.2.

L'opzione "Voci massime" sotto questi minimi e massimi ci permette anche di regolare il numero di parole chiave che possono essere estratte per i nostri risultati. Il valore predefinito per questa opzione è di 1000 parole.

A destra delle opzioni "Focus on" si trovano le opzioni generali che abbiamo visto anche nella ricerca avanzata della funzione Wordlist. Attraverso queste opzioni possiamo selezionare se vogliamo: 1) trattare le parole maiuscole e minuscole allo stesso modo ("A=a"), 2) includere i risultati con uno o più caratteri alfanumerici ("Almeno un alfanumerico"), 3) includere solo i risultati alfanumerici ("Solo risultati alfanumerici"), 4) "Includi le non parole" (si veda la sezione 2.2.2), 5) Escludi le non parole. 2.2), 5) escludere un elenco di parole (vedere sezione 2.2.2) e 6) includere alcune parole che vogliamo compaiano nell'elenco se esistono nel corpus principale ("Da elenco"). Quando si seleziona l'opzione "Da elenco", viene visualizzata una casella di testo in cui elencare l'elenco di parole, come nel caso dell'elenco di parole escluse.

Verso la fine della pagina di ricerca avanzata delle parole chiave si trovano altre quattro opzioni. L'opzione "Identifica parole chiave" ci permette di cambiare il tipo di parole che riceviamo come risultati (ad esempio lemmi, tag, parole, ecc.). Per utilizzarla, nella sezione "Impostazioni delle parole chiave", fare clic sul menu a discesa Attributo per visualizzare tutti i tipi di parole e fare clic su quello che si desidera visualizzare nei risultati. Sotto di essa si trova la barra di testo "Matching regex", che consente di recuperare i risultati che corrispondono alla regex digitata dall'utente. Le risorse su come utilizzare le regex per Sketch Engine sono disponibili alla fine dell'esercitazione. A destra, l'opzione "Identifica termini" offre all'utente la possibilità di usare la regex per restituire i risultati in base ai limiti della regex che l'utente digita nella barra di testo sotto la sezione "Impostazioni termini".

All'estrema destra di questa riga si trova la sezione "Identifica n-grammi", che consente agli utenti di recuperare n-grammi o cluster comuni di parole in base ai nostri criteri. Possiamo regolare il tipo di parola utilizzando il menu a discesa Attributo e scegliere la lunghezza dell'n-gramma con

la barra numerica accanto a "Lunghezza dell'n-gramma" in fondo a questa sezione. In questa sezione si possono anche utilizzare le regex per restringere ulteriormente la ricerca.

Infine, la sezione "Tipi di testo" funziona allo stesso modo della pagina di ricerca della Concordezza nella sezione 2.1. È sufficiente fare clic sul cursore accanto a "Tipo di testo". È sufficiente fare clic sul cursore accanto a "Tipi di testo" per accedere a queste opzioni e regolarle di conseguenza.

Esercizio 3: Utilizzando la scheda di ricerca avanzata Parole chiave, trovare le parole chiave con riferimento al corpus English Web 2013 (enTenTen13). Includere le non parole, regolare la frequenza "Focus on" a 1000 e mantenere tutte le altre voci come predefinite. Deselezionare le funzioni "Identifica parole chiave" e "Identifica termini". Regolate la ricerca in modo da includere i bi-grammi e i trigrammi delle parole. Quali sono i primi tre risultati di questa ricerca?

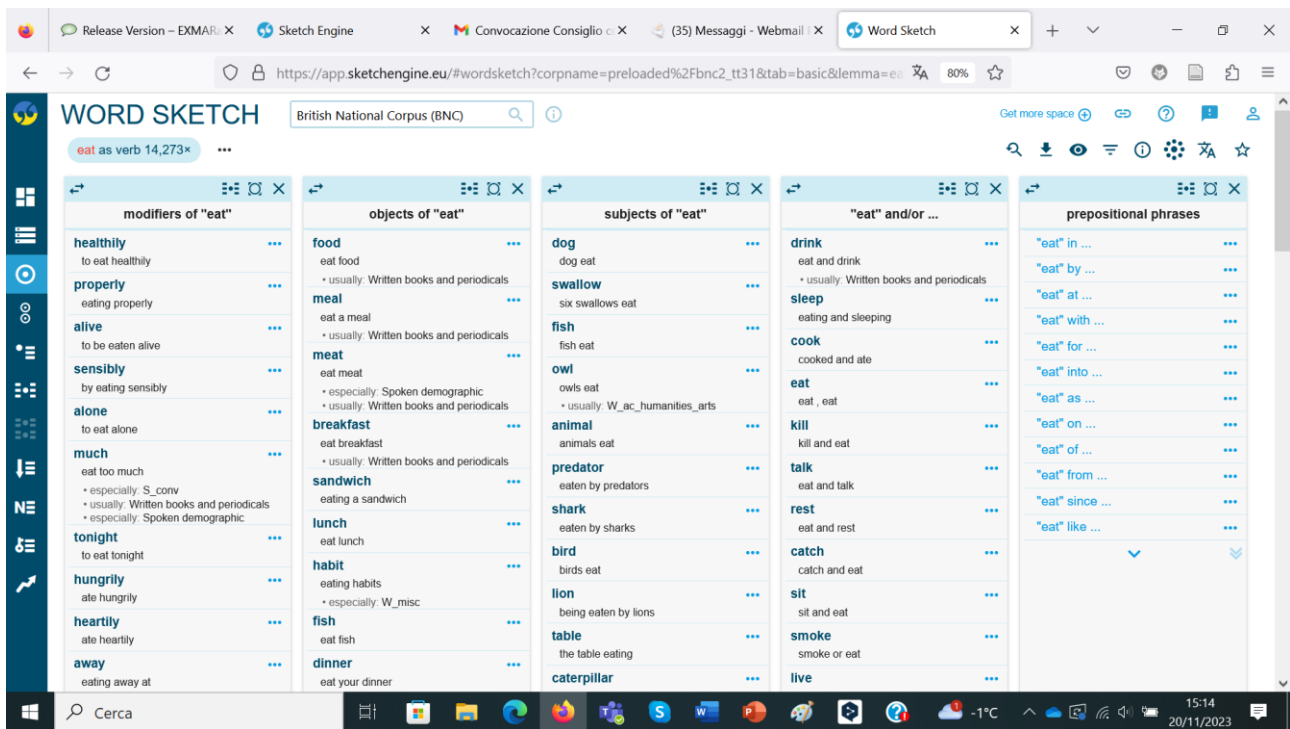
2.4 SCHIZZO DI PAROLE (WORD SKETCH)

La funzione Schizzo di parole consente di cercare collocazioni e combinazioni di parole per un dato lemma. È uno strumento utile per osservare l'uso di una parola che può assumere varie forme e parti del discorso a seconda del contesto e dei requisiti grammaticali.

2.4.1 Ricerca di base

Per effettuare una ricerca con questa funzione, fare clic sul pulsante "Schizzo di parole" nella dashboard. Digitare un lemma qualsiasi nella barra di ricerca e fare clic sul pulsante "GO" in basso a destra. Si accede così a una pagina in cui è possibile visualizzare i diversi lemmi della parola in base alle diverse funzioni grammaticali nel corpus scelto.

I risultati della ricerca per il lemma "eat" (mangiare) come verbo sono mostrati di seguito:



Nell'immagine qui sopra, vengono fornite le collocazioni del verbo mangiare in vari contesti, compresi i suoi modificatori comuni, gli oggetti, le frasi, le particelle e altro ancora.

A volte il lemma scelto può assumere la forma di diverse parti del discorso. Ad esempio, il lemma "giocare" può essere usato come verbo e come sostantivo. Per questa parola, Work Sketch sceglierà automaticamente l'uso come verbo, ma per visualizzare i risultati di "play" come sostantivo, è sufficiente fare clic sulla bolla blu in alto a sinistra degli elenchi che dice "parola come verbo [numero di volte in cui questa forma ricorre nel corpus]". Quindi, fare clic sulla forma di parte del discorso che si desidera visualizzare:

2.4.2 Ricerca avanzata

Word Sketch dispone anche di un'opzione di ricerca avanzata, accessibile dalla pagina iniziale di Word Sketch. Accanto alla scheda di ricerca "BASE", fare clic sulla scheda di ricerca "AVANZATA" per visualizzarne le funzioni.

Sotto la casella di testo della ricerca, possiamo scegliere quale parte del discorso vogliamo che sia considerata dal lemma digitato. Va notato che è possibile scegliere una sola parte del discorso dall'elenco fornito. Nella sezione "Parte del discorso", è sufficiente fare clic sulla parte del discorso che si desidera sia ricercata nel corpus.

Accanto all'opzione Parte del discorso si trova l'opzione Sottocorpus, che consente di selezionare la parte del corpus selezionata da considerare per la ricerca. Fare clic sul menu a discesa di questa opzione per scegliere il sottocorpus che si desidera includere. L'opzione "Frequenza minima" consente di specificare il numero minimo di volte in cui un lemma deve comparire nel corpus per essere considerato un risultato. L'opzione predefinita è "auto", ma può essere sostituita da un numero.

L'opzione "Punteggio minimo" sotto "Frequenza minima" consente di scegliere se visualizzare o meno i collocati con un punteggio di tipicità inferiore al limite specificato. Questo numero può essere regolato con la barra di testo sotto il titolo dell'opzione.

Infine, la ricerca avanzata consente anche di tradurre un lemma in una lingua diversa sulla base di un altro corpus di una lingua diversa. Per utilizzare questa funzione, è necessario selezionare la casella accanto a "Traduci", che si trova sotto l'opzione "Punteggio minimo", e selezionare il corpus da cui si desidera recuperare la traduzione.

La ricerca avanzata comprende anche le opzioni "Tipi di testo" viste nelle sezioni precedenti. Si trovano in fondo alla pagina, verso il pulsante "GO". Per regolare le impostazioni di questa categoria, fare clic sul cursore accanto al titolo, come per le altre funzioni.

Esercizio 4: Utilizzando la ricerca avanzata di Word Sketch, cercate il lemma *walk* "camminare" come verbo. Quali sono i tre soggetti più comuni di "camminare" secondo i risultati?

2.5 WORD SKETCH DIFFERENCE

Come la funzione Word Sketch descritta in precedenza, la Word Sketch Difference offre le stesse funzioni ma per due lemmi diversi a scelta dell'utente. Questo ci permette di visualizzare facilmente le collocazioni e le combinazioni di parole di ciascun lemma affiancate.

2.5.1 Ricerca di base

Facendo clic sul pulsante "Word Sketch Difference" nella Dashboard, si accede alla pagina di ricerca.

Sotto il tag "primo lemma" si trova la barra di ricerca, dove è possibile inserire il primo lemma scelto. Il secondo lemma può essere inserito nella barra di ricerca sotto "secondo lemma". Una volta completate entrambe le barre di ricerca, fare clic sul pulsante "GO" sotto di esse.

Come per i risultati di Word Sketch, vengono forniti esempi dell'uso contestuale dei lemmi. Tuttavia, in questa funzione, ogni lemma sarà indicato con i propri colori (si veda l'esempio precedente). La barra sfumata nella parte superiore della pagina indica il colore con cui ogni lemma è rappresentato e le corrispondenti occorrenze nel corpus sono visualizzate accanto al lemma (ad esempio, "talk" ricorre 28.857 volte nel corpus). Questa codifica dei colori è stata implementata per aiutarci a distinguere le occorrenze contestuali più comuni per ogni lemma che abbiamo fornito. Le tonalità più scure e opache del colore indicano i contesti più frequentemente utilizzati per quel dato lemma, mentre le tonalità più chiare del colore rappresentano usi meno frequenti o leggermente condivisi del contesto tra entrambi i lemmi.

In ogni tabella contestuale, viene indicato anche il numero di occorrenze in cui ogni parola compare nel contesto dato. Il numero più a sinistra in una determinata riga rappresenta il numero

di occorrenze del primo lemma, mentre il numero più a destra rappresenta le occorrenze del secondo lemma.

Oltre alla caratteristica unica del gradiente di questa funzione, Word Sketch Difference ha anche una funzione di visualizzazione che ci permette di confrontare le distribuzioni dei due lemmi in un determinato contesto senza i numeri delle tabelle nel risultato iniziale. Per vedere la distribuzione in un determinato contesto (ad esempio, una tabella nei risultati), fare clic sull'icona in alto a sinistra nell'angolo destro della tabella con più punti. Quando si passa il mouse su questa icona, i punti diventano multicolore e un'etichetta a comparsa visualizza "Mostra visualizzazione". Facendo clic su questa icona, si accede a una pagina con un grafico a gradiente.

In questa visualizzazione, possiamo vedere quali sono le parole che ricorrono più frequentemente con un lemma piuttosto che con l'altro, osservando le parole che appaiono più vicine a quel lemma sul grafico. Le parole che vengono utilizzate con entrambi i lemmi in una distribuzione più o meno uguale appaiono spesso verso il centro del grafico, dove i colori dei due lemmi si fondono. Anche le parole con un numero elevato di occorrenze nel corpus avranno una bolla più grande per indicare questa maggiore frequenza. Le parole con una frequenza più bassa avranno bolle di dimensioni più piccole. Per le parole che si trovano al centro del grafico, accanto alle parole appariranno due bolle, ognuna delle quali indica il colore del lemma corrispondente. La bolla più grande delle due indica che questa parola ricorre più spesso con il lemma del colore della bolla stessa.

Infine, questo grafico di visualizzazione ci permette anche di controllare il numero di collocazioni che vogliamo vedere nel grafico, regolando la barra "Numero di collocazioni" sulla destra del grafico. Basta far scorrere la barra verso sinistra se si preferisce concentrarsi su un maggior numero di parole che ricorrono con il primo lemma e si vuole eliminare un maggior numero di collocati tra i due. Fate scorrere la barra verso destra se volete avere più collocazioni elencate nel grafico per il confronto. Entrambe le visualizzazioni, tabella e grafico, di Word Sketch Difference consentono agli utenti di scegliere la loro preferenza per la visualizzazione dei risultati del confronto tra lemmi, oltre a darci l'opportunità di visualizzare i dati in modi diversi.

2.5.2 Ricerca avanzata

Con la ricerca avanzata di Word Sketch Difference, possiamo utilizzare alcune opzioni in più per restringere ulteriormente i risultati della ricerca. Dalla pagina iniziale di Word Sketch Difference, fare clic sulla scheda di ricerca "AVANZATA" per visualizzare queste funzioni.

La prima caratteristica che possiamo regolare è il tipo di confronto. Nella sezione "confronta" è possibile scegliere se confrontare "Lemmi", "Forme di parole" o "Sottocorpora". L'opzione "Lemmi" confronta i collocati dei due lemmi indicati nelle caselle di testo sottostanti, come nella ricerca di base.

L'opzione "Forme di parola" confronta i collocati di due forme di parola diverse che appartengono allo stesso lemma. Facendo clic su questa opzione, viene visualizzata una barra di testo "lemma" in cui si inserisce il lemma che si desidera controllare. Al di sotto di questa, viene richiesto di indicare due diverse forme verbali per questo lemma. Ad esempio, possiamo inserire "go" come lemma, seguito da "goes" come prima forma verbale e "went" come seconda forma verbale.

L'opzione "Sottocorpora" confronta i collocati dello stesso lemma trovati in diversi sottocorpora del corpus scelto. Questa opzione richiede di inserire un lemma e di selezionare due diversi sottocorpora per il confronto.

Sotto queste opzioni di confronto si trova l'elenco "Part of speech", che ci permette di specificare quale parte del discorso vogliamo osservare nel corpus per il lemma dato.

Infine, l'opzione "Frequenza minima", in fondo a questa pagina, permette di inserire un numero minimo di volte in cui un lemma può comparire nel corpus per essere considerato un risultato. L'opzione predefinita è "auto", ma può essere sostituita da un numero minimo.

Esercizio 5: Utilizzando la ricerca avanzata Word Sketch Difference, cercate il lemma "be" e trovate la differenza tra le sue forme verbali "was" e "are". Lasciare tutte le altre funzioni alle loro selezioni predefinite. Quale soggetto è più frequentemente associato a "was"? Quale soggetto è più frequentemente associato a "are"?

2.6 N-GRAMMI

L'ultima caratteristica degna di nota della dashboard di Sketch Engine è la funzione N-grammi.

Gli n-grammi sono gruppi di parole che si trovano comunemente insieme, con "n" che indica il numero di parole nella sequenza. Ad esempio, i bi-grammi sono gruppi di due parole, i tri-grammi sono gruppi di tre parole e così via. Questa funzione consente agli utenti di visualizzare gli n-grammi più comuni nel corpus selezionato, che possono essere controllati dalle opzioni di lunghezza degli n-grammi nel menu di ricerca.

2.6.1 Ricerca di base

Per accedere agli n-grammi, fare clic sul pulsante "n-grammi" nella dashboard e si accederà alla pagina di ricerca.

Questa funzione ci permette di esaminare gli N-grammi da 2 a 6, il che significa che è possibile ricercare un massimo di 6 gruppi di parole. Per selezionare il numero di n-grammi che si desidera cercare, cliccare due volte sul numero corrispondente nel menu "Lunghezza n-grammi". Ad esempio, se si desidera trovare i bi-grammi, si deve fare clic sul quadratino "2". I risultati di questa ricerca sono i seguenti:

Questa pagina elenca gli n-grammi dal più frequente al meno frequente. La frequenza di ogni n-gramma si trova a destra della voce. È possibile scorrere l'elenco utilizzando le frecce in basso a destra.

La funzione n-grammi ci permette anche di cercare più di un tipo di n-gramma. Se vogliamo cercare i bi-grammi, i tri-grammi e i 4-grammi più frequenti, andiamo alla pagina iniziale di questa funzione, facciamo clic sul quadrato "2" e poi sul quadrato "4". In questo modo i riquadri "2", "3" e "4" verranno evidenziati in blu scuro per confermare la selezione.

Facendo clic sul pulsante "GO", appariranno i seguenti risultati per il corpus scelto:

2.6.2 Ricerca avanzata

Possiamo restringere ulteriormente i risultati della ricerca con la ricerca avanzata di n-grammi. Dalla pagina iniziale di n-grammi, fate clic sulla scheda di ricerca "ADVANCED" per visualizzare le opzioni aggiuntive.

Come per la ricerca di base, possiamo regolare i tipi di n-grammi che vogliamo vedere nei nostri risultati regolando l'opzione "Lunghezza n-grammi". Al di sotto di questa opzione si trova l'opzione "Attributo", che ci permette di selezionare il tipo di parole che vogliamo vedere nei nostri risultati. Queste opzioni includono "parola", "lemma", "tag" e "lemma". È sufficiente fare clic sul menu a discesa per visualizzare queste opzioni e fare clic sul tipo di parola che si desidera visualizzare nei risultati.

Accanto a questa, sulla destra, si trovano le funzioni già viste in altre ricerche avanzate: "A=a", "Includi le non parole" e "Escludi queste parole". Queste tre opzioni funzionano nello stesso modo in cui sono state spiegate nelle precedenti sezioni sulla ricerca avanzata. Tuttavia, in questa sezione viene presentata un'opzione aggiuntiva, la casella di controllo "Nest n-grams". Se spuntata, questa opzione prende gli n-grammi che fanno parte di un n-gramma più grande e li raggruppa con l'n-gramma più grande. È un'ottima funzione da spuntare per evitare n-grammi identici che potrebbero già esistere in un altro n-gram, soprattutto quando si seleziona un intervallo di numeri di n-grammi.

Come le altre opzioni di ricerca avanzata menzionate in precedenza, esiste anche un'opzione "Frequenza min" e "Frequenza max" per specificare le frequenze minime e massime che una parola deve avere per essere considerata un risultato. Accanto a questa opzione si trova l'opzione "Sottocorpus", che consente di selezionare un sottocorpus del corpus scelto da cui recuperare gli n-grammi.

Sotto le opzioni di frequenza si trova la casella di controllo "Key n-grams" che, se selezionata, identifica le collocazioni tipiche del corpus di interesse rispetto al corpus di riferimento. L'utilizzo di questa opzione richiede la selezione di un corpus di riferimento e di un sottocorpus di riferimento (se desiderato) per effettuare i confronti tra le collocazioni. La funzione "Focus on" vista nella ricerca avanzata di Keywords è presente anche nella funzione "Key n-grams" e funziona allo stesso modo di Keywords.

Successivamente, i "Criteri aggiuntivi" che si trovano sotto gli "N-grammi chiave" consentono di specificare altri attributi che un risultato deve avere. Ad esempio, si può fare clic sull'opzione "inizia con" in questo elenco e verrà richiesto di inserire il carattere o i caratteri con cui devono iniziare i risultati.

Infine, le opzioni "Tipi di testo" sotto "Criteri aggiuntivi" funzionano come nelle ricerche avanzate descritte in precedenza. Fare clic sul cursore accanto al titolo di questa sezione per visualizzare queste opzioni e adattarle alle proprie preferenze.

Esercizio 6: Utilizzando la ricerca avanzata N-grammi, cercare parole con 3-4 grammi. Queste parole devono contenere le lettere "ing". Lasciare tutte le altre opzioni come predefinite. Qual è l'n-gramma più frequente per questa ricerca?

3 RISORSE PER IMPARARE ALTRE FUNZIONI

I seguenti link sono suggerimenti video per gli utenti che desiderano approfondire l'uso di Sketch Engine:

Playlist del canale YouTube di Sketch Engine [YouTube] (include tutorial approfonditi per Subcorpus, CQL, costruzione del corpus e altro)
<https://www.youtube.com/@SketchEngine/playlist>

Brevi tutorial sull'uso degli strumenti del corpus per insegnanti (in formazione), studenti, linguisti e chiunque sia interessato alla linguistica del corpus e ai metodi del corpus. - di Ellen Le Foll [YouTube, Playlist video 5-8]
https://www.youtube.com/watch?v=TMAXswf6l_Q&list=PLAq6uhS_0brxTW99jeZjxdsIQ5BRy1Eb5&index=5
