



Sommario

1. Cos'è Sketch Engine (SkE)	1
2. Come accedere a Sketch Engine	2
3. I corpus	5
3.1. Come selezionare un corpus su Sketch Engine	5
3.1.1. Select corpus - BASIC	5
3.1.2. Select corpus - ADVANCED	6
3.1.3. Select corpus – MY CORPORA	7
3.1.4. Select corpus – SHARED WITH ME	7
3.2. Come costruire un corpus su Sketch Engine	8
3.2.1. Costruire un corpus dal web – Find text on the web	9
3.2.2. Come costruire un corpus con testi nostri - I have my own texts	11
3.3. Come creare dei subcorpora	13
4. Gli strumenti di Sketch Engine.....	15
4.1. Concordance	15
4.1.1. CQL – Corpus Query Language	20
4.1.2. Frequency: contare parole, tags e text types nelle concordanze.....	28
4.2. Word Sketch	30
4.3. Thesaurus	34

1. Cos'è Sketch Engine (SkE)

Come si pronuncia Sketch Engine: [skɛtʃ ɛndʒɪn]

Sketch Engine è un programma disponibile **online a pagamento*** per operare ricerche linguistiche; in particolare è un **gestore di corpus** e un **software di analisi del testo** sviluppato da Lexical Computing Limited dal 2003. Attualmente supporta e fornisce corpora (ca. 600 già pronti per l'uso e disponibili a tutti gli utenti) in oltre 90 lingue.

*Nonostante sia a pagamento, SkE offre tutte le funzionalità principali sviluppate nel 2003 (e ampliate da allora) gratuitamente con licenza GPL all'interno della suite NoSketch Engine.

Perché è utile Sketch Engine? Perché SkE **genera automaticamente tutte le informazioni che ci servono** (e che gli chiediamo) **sulla base di un corpus**, rendendo il lavoro di ricerca più rapido e preciso.

→ Cos'è un corpus?

TRECCANI

(<https://www.treccani.it/vocabolario/corpus/>)

s. neutro lat., usato in ital. al masch. (pl. lat. *corpora*, ma in ital. anche invar.). – 1. Corpo, nel sign. di raccolta; è termine usato nella nomenclatura bibliografica, fin dall'antichità classica, per indicare la riunione di più opere, raccolte e pubblicate con lo scopo di fornire serie ordinate e complete degli scritti di uno o più autori o intorno a determinate materie: *Corpus Hippocraticum*; *Corpus iuris civilis*; *Corpus iuris canonici*; *Corpus inscriptionum latinarum*. In senso più ampio: *il c. dei vocaboli* o *delle frasi di una lingua*; *il c. delle poesie popolari di una regione*.

2. Come accedere a Sketch Engine

- a. **Accedere al sito internet:** Aprire il nostro motore di ricerca e digitare “Sketch Engine”. Il link che ci interessa, ovvero quello del sito, dovrebbe essere la prima opzione (Figura 1).

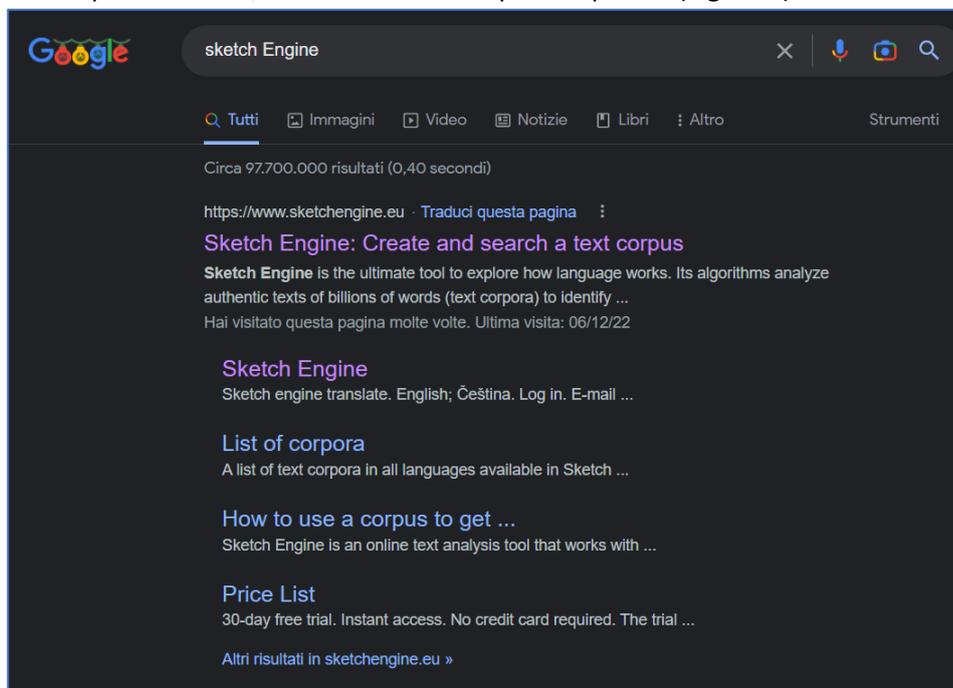


Figura 1 - Sketch Engine su Google

- b. **Effettuare l'accesso al programma online:** Una volta entrati nel sito, dobbiamo selezionare “Login” per poter accedere al programma (Figura 2).

Per **fare il login** abbiamo varie opzioni; nel nostro caso, essendo membri di un'istituzione (UniBg) che ha l'abbonamento al sito, dobbiamo selezionare “Institutional login” (Figura 3); ci verrà quindi richiesto di indicare la nostra istituzione di appartenenza (scrivendo “Bergamo” comparirà già il suggerimento corretto) (Figura 4): selezioniamo “University of Bergamo”. A questo punto verremo rimandati alla nostra solita schermata di accesso a UniBg, dove dovremo inserire i nostri consueti dati di accesso (username e password di UniBg).

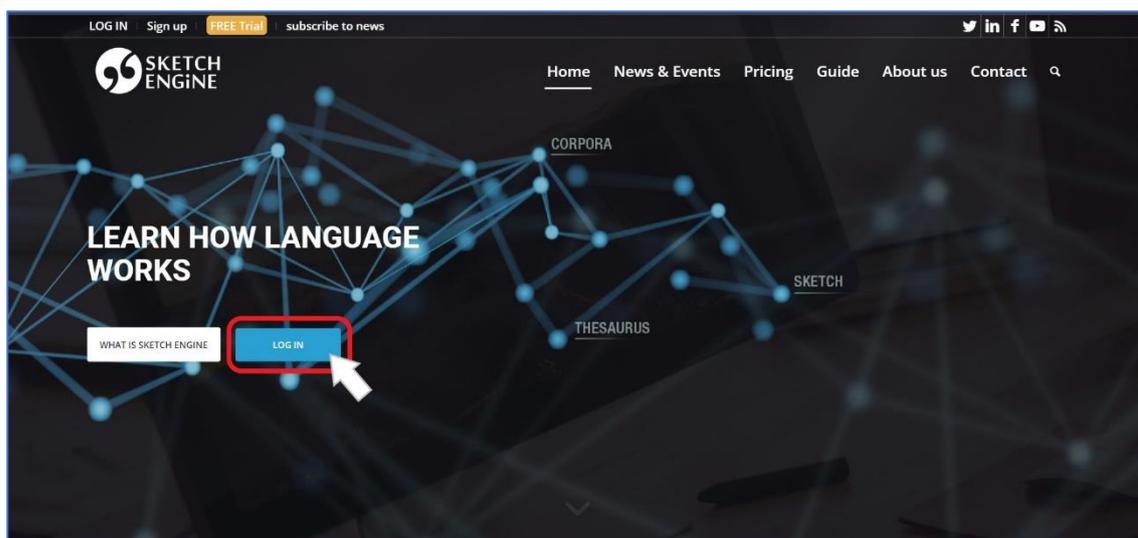


Figura 2 - Schermata home del sito Sketch Engine

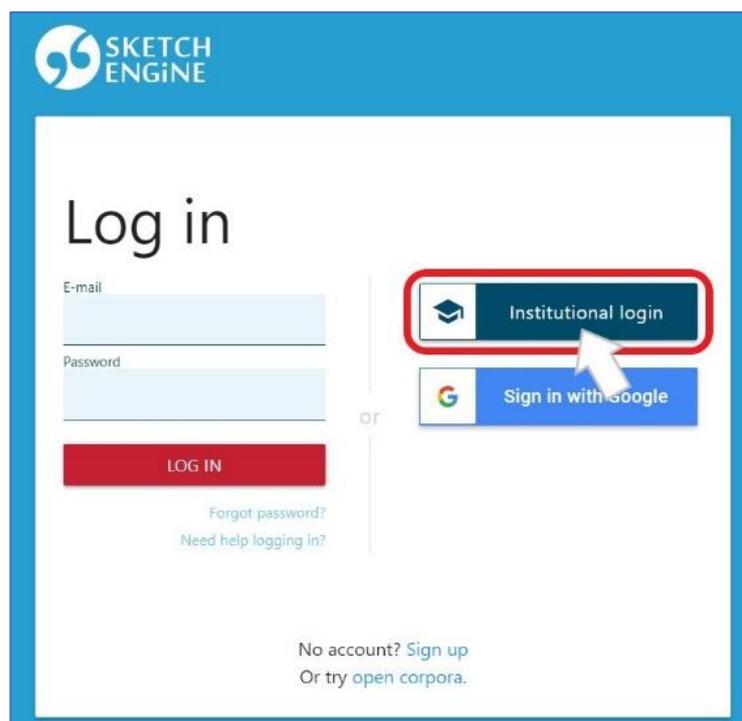


Figura 3 - Schermata Login su Sketch Engine

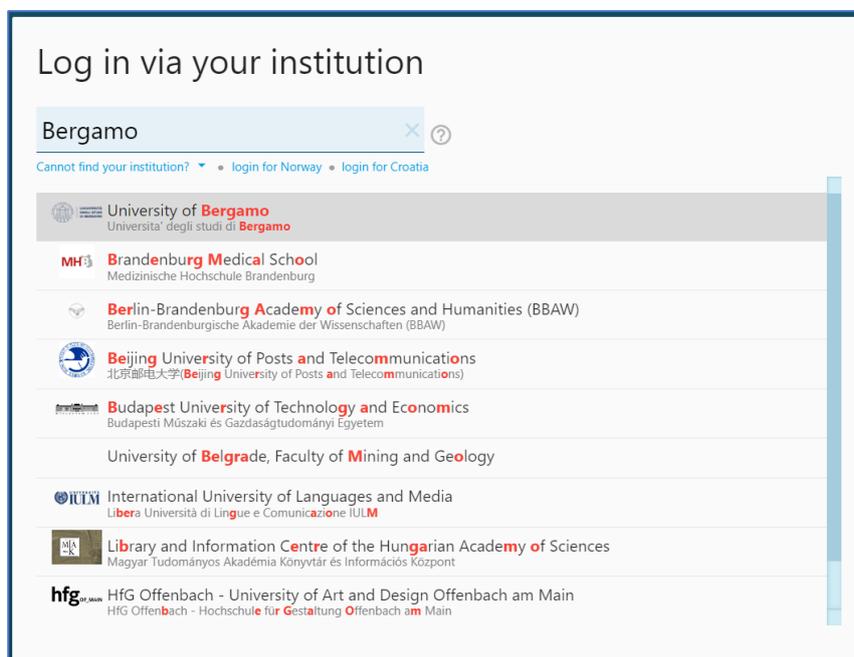


Figura 4 - Ricerca istituzione di appartenenza per Login su Sketch Engine

- c. **Benvenuto su Sketch Engine:** Una volta effettuato il login, questa è la schermata che ci troveremo davanti (Figura 5):

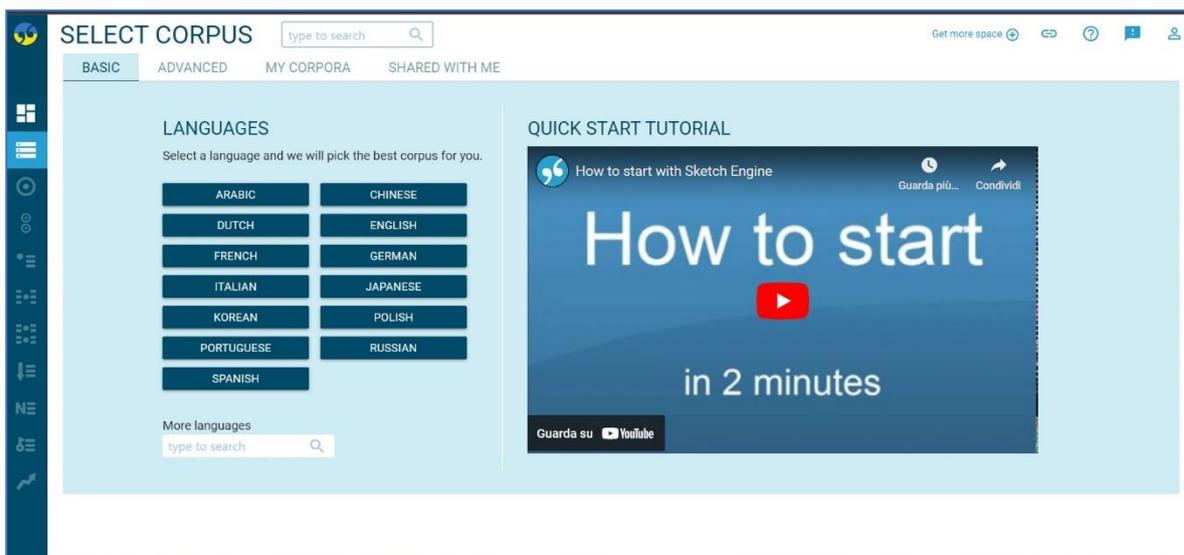


Figura 5 - Schermata home del programma Sketch Engine

3.1 corpus

3.1. Come selezionare un corpus su Sketch Engine

Prima di poter fare qualsiasi cosa su SkE, **dobbiamo selezionare una lingua o un corpus**; infatti la prima azione che ci viene proposta una volta fatto il login è “Select corpus” (Figura 6 o Figura 5). Come possiamo vedere, le opzioni qui disponibili sono diverse: chiaramente dipendono da cosa ci serve fare.

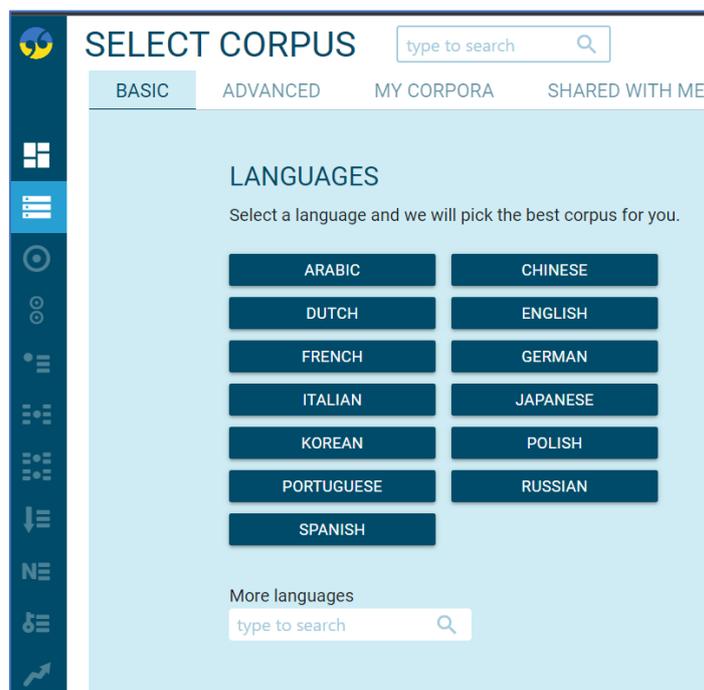


Figura 6 - Select corpus

Chiaramente più è grande il corpus e più risultati otterremo; inoltre questi risultati dovrebbero anche essere tanto più attendibili quanto più è grande il corpus che abbiamo scelto. Detto ciò, ovviamente l’ampiezza del corpus che sceglieremo dipenderà da vari fattori in base alle nostre esigenze e, talvolta, in base alle nostre possibilità (es. ricerche in diacronia = problema di *bad data*).

3.1.1. Select corpus - BASIC

Se non abbiamo già un corpus (né su SkE né sul nostro pc per poterlo caricare sul programma) o non ne abbiamo in mente uno in particolare, possiamo usare l’opzione BASIC: noi scegliamo la lingua che ci interessa, SkE selezionerà il corpus migliore per quella data lingua.

Selezioniamo la lingua che ci interessa tra quelle già proposte nella schermata “Languages” o, se lì non è presente, digitandola nella barra di ricerca in basso “More languages” (Figura 7).



Figura 7 - Select corpus-BASIC

3.1.2. Select corpus - ADVANCED

Se sappiamo già quale corpus usare (tra quelli disponibili sul sito), possiamo usare l'opzione ADVANCED: SkE userà il corpus da noi selezionato.

Selezioniamo il corpus desiderato cercando nell'elenco fornito da SkE oppure cercando nella barra di ricerca in alto (Figura 8).

Come possiamo vedere, in Advanced SkE ci offre un elenco di corpora che possiamo gestire noi: possiamo scegliere di vederli tutti con "Any language" o possiamo digitare la lingua che ci interessa nella barra di ricerca a fianco, in alto a sinistra (che non è la stessa che abbiamo già citato). Si veda l'esempio fornito in Figura 9.

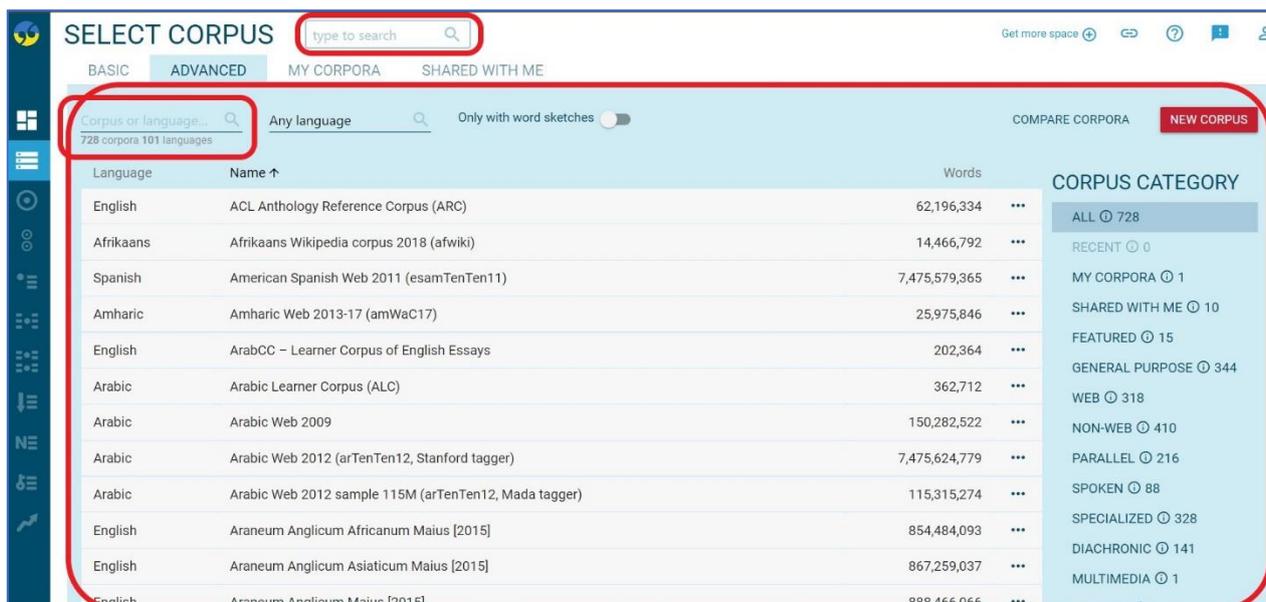
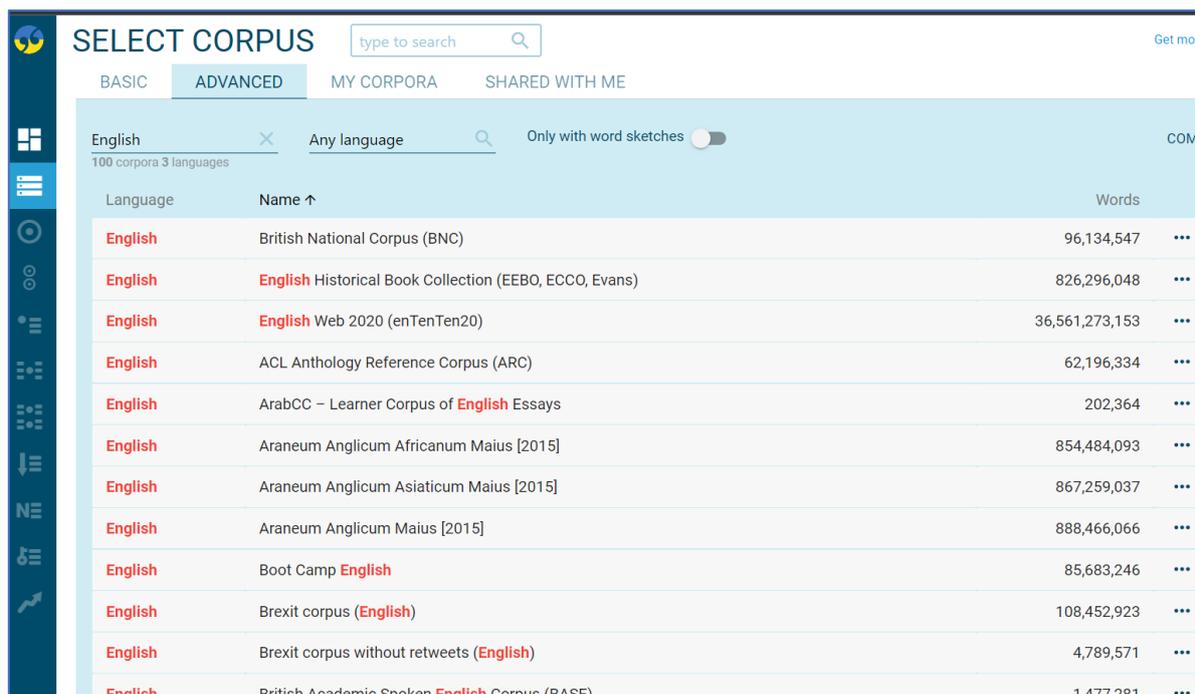


Figura 8 - Select corpus-ADVANCED

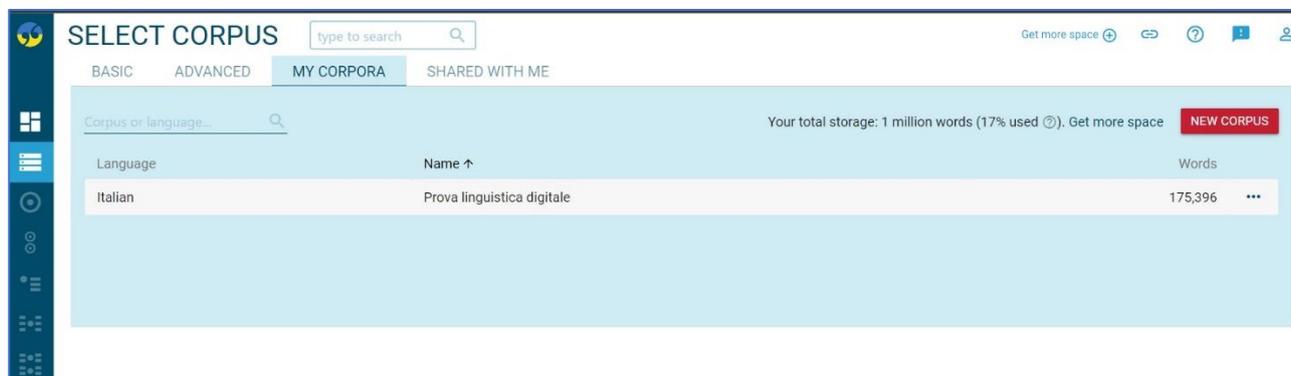


Language	Name ↑	Words
English	British National Corpus (BNC)	96,134,547
English	English Historical Book Collection (EEBO, ECCO, Evans)	826,296,048
English	English Web 2020 (enTenTen20)	36,561,273,153
English	ACL Anthology Reference Corpus (ARC)	62,196,334
English	ArabCC – Learner Corpus of English Essays	202,364
English	Araneum Anglicum Africanum Maius [2015]	854,484,093
English	Araneum Anglicum Asiaticum Maius [2015]	867,259,037
English	Araneum Anglicum Maius [2015]	888,466,066
English	Boot Camp English	85,683,246
English	Brexit corpus (English)	108,452,923
English	Brexit corpus without retweets (English)	4,789,571
English	British Academic Spoken English Corpus (BASE)	1 477 281

Figura 9 - Esempio su Select corpus-Advanced

3.1.3. Select corpus – MY CORPORA

Se abbiamo già un nostro corpus (caricato su SkE o da caricare) possiamo usare l'opzione MY CORPORA (Figura 10).

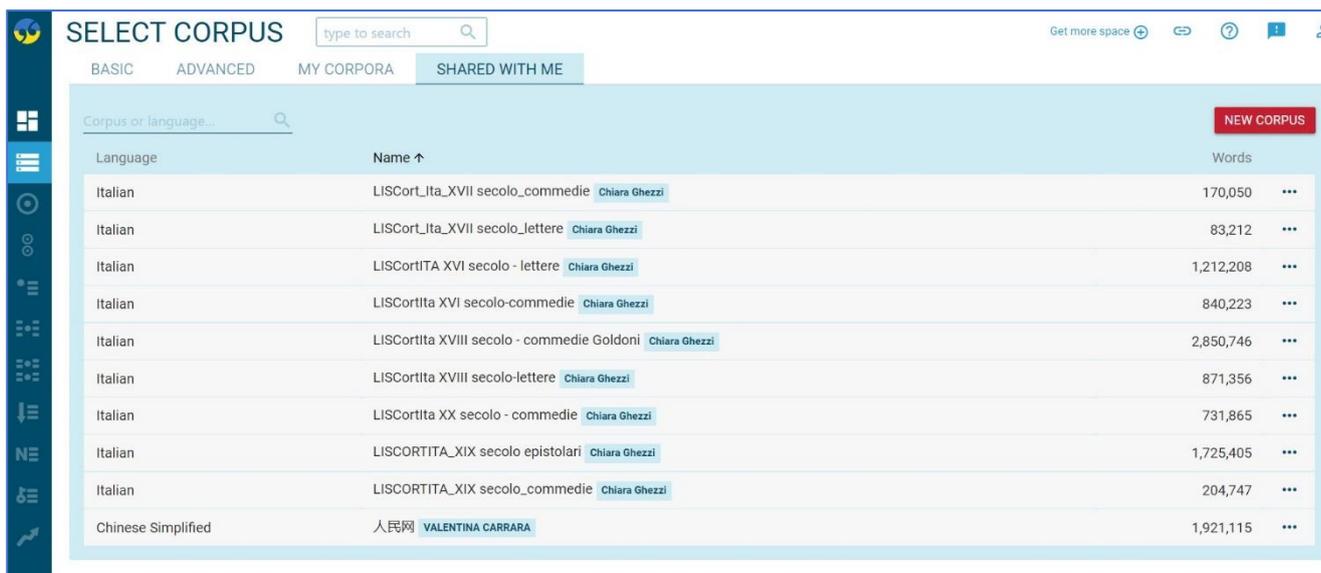


Language	Name ↑	Words
Italian	Prova linguistica digitale	175,396

Figura 10 - Select corpus-MY CORPORA

3.1.4. Select corpus – SHARED WITH ME

se qualcuno ha condiviso con noi un corpus o più corpora e vogliamo usare uno di quelli, possiamo usare questa opzione SHARED WITH ME (Figura 11).



Language	Name ↑	Words
Italian	LISCort_Ita_XVII secolo_commedie Chiara Ghezzi	170,050 ...
Italian	LISCort_Ita_XVII secolo_lettere Chiara Ghezzi	83,212 ...
Italian	LISCORTITA XVI secolo - lettere Chiara Ghezzi	1,212,208 ...
Italian	LISCortIta XVI secolo-commedie Chiara Ghezzi	840,223 ...
Italian	LISCortIta XVIII secolo - commedie Goldoni Chiara Ghezzi	2,850,746 ...
Italian	LISCortIta XVIII secolo-lettere Chiara Ghezzi	871,356 ...
Italian	LISCortIta XX secolo - commedie Chiara Ghezzi	731,865 ...
Italian	LISCORTITA_XIX secolo epistolari Chiara Ghezzi	1,725,405 ...
Italian	LISCORTITA_XIX secolo_commedie Chiara Ghezzi	204,747 ...
Chinese Simplified	人民网 VALENTINA CARRARA	1,921,115 ...

Figura 11 - Select corpus-SHARED WITH ME

3.2. Come costruire un corpus su Sketch Engine

Se i corpora a nostra disposizione su SkE non ci interessano, possiamo crearne uno o più per conto nostro. Per farlo:

1. Usiamo la voce “New corpus” su “My Corpora” oppure la voce “Create corpus” che appare in basso a sinistra della tendina che si apre cliccando sulla barra di ricerca dei corpora (Figura 12).

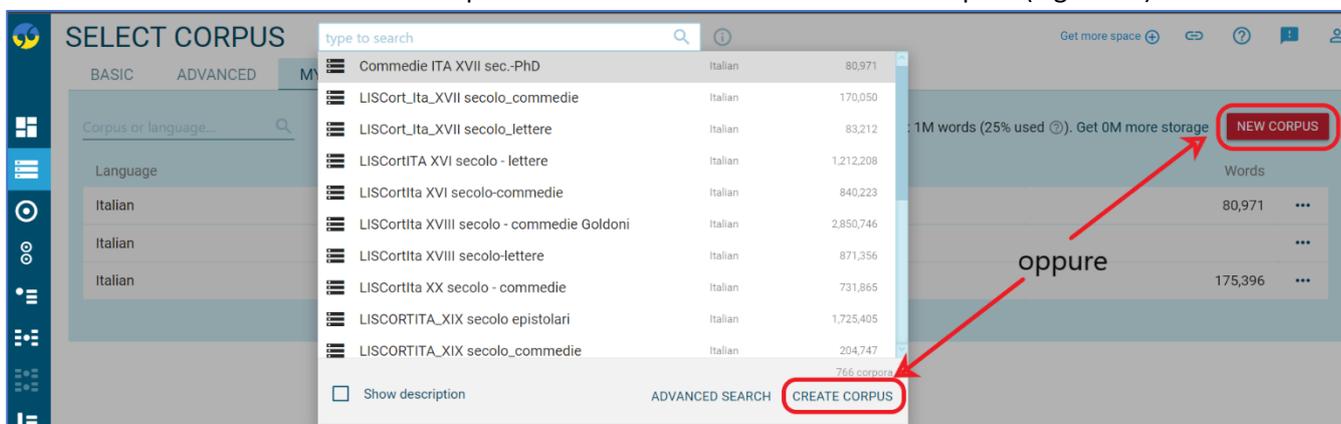


Figura 12 - Creare un nuovo corpus su SkE_1

2. SkE ci chiederà quindi di inserire alcuni dati relativi al corpus che vogliamo creare: il nome/titolo del corpus, se si tratta di un corpus mono- o plurilingue, che lingua contiene (nel caso di un corpus monolingue) e una descrizione del corpus (quest'ultima voce è opzionale). Possiamo inoltre selezionare le funzionalità che avrà il nostro corpus (de)selezionandole dalla voce “Available features”, sotto al riquadro di compilazione appena descritto (Figura 13). Quando abbiamo terminato, clicchiamo su “Next”.

Build your own private corpus from texts on the web or from your own documents.

Name Es. creazione corpus SkE

Corpus type Single language corpus
 Multilingual corpus

Language Italian 🔍

Description Esempio per lezione come creare un corpus di SkE

Storage used: 256,367 of 1,000,000 words (25%)

Available features ▾

<input checked="" type="checkbox"/> Part of speech	<input checked="" type="checkbox"/> wordsketch	<input checked="" type="checkbox"/> GDEX
<input checked="" type="checkbox"/> N-grams	<input checked="" type="checkbox"/> lemma	<input checked="" type="checkbox"/> Thesaurus
<input checked="" type="checkbox"/> Concordance	<input checked="" type="checkbox"/> Frequency distribution	<input checked="" type="checkbox"/> Term extraction
<input checked="" type="checkbox"/> Keywords		<input checked="" type="checkbox"/> Wordlist

Figura 13 - Creare un nuovo corpus su SkE_2

3. A questo punto abbiamo 2 possibilità (Figura 14):
 - a. Attingere a testi dal web → **Find text on the web**: SkE troverà i testi rilevanti da internet per noi.
 - b. Attingere a testi che già possediamo (sul nostro pc, per esempio) → **I have my own texts**: possiamo caricare i nostri testi.

CREATE CORPUS

CORPUS: Es. creazione corpus SkE (Italian)

CREATE CORPUS > ADD TEXTS > COMPILE

 **Find texts on the web**
Automatically find and download relevant texts

 **I have my own texts**
Upload your own files (.txt, .pdf...) or paste text

Figura 14 - Creare un corpus su SkE_3

3.2.1. Costruire un corpus dal web – Find text on the web

Se non abbiamo testi nostri oppure quelli che abbiamo non ci interessano, possiamo creare un corpus utilizzando testi presi dal web. → **Find text on the web**.

Una volta selezionata questa opzione, si aprirà una tenda di compilazione (Figura 15), dove per prima cosa troveremo la voce "Input type", da cui possiamo scegliere le opzioni:

- **Web search:** SkE troverà testi adatti usando un motore di ricerca internet.
- **URLs:** potremo scaricare testi da una lista di indirizzi web forniti da noi (NB: link di pagine rimandanti ad altre pagine non vanno bene!)
- **Website:** potremo scaricare contenuti di un intero sito web, con un limite di 2.000 pagine/sito. ATTENZIONE!!! Questa opzione è molto dispendiosa a livello di tempo, in quanto il tempo di download è di circa 6 pagine/minuto. Tuttavia il processo può continuare, arrivando a concludersi, in background, ovvero anche se l'utente esce (fa il log out) da SkE.

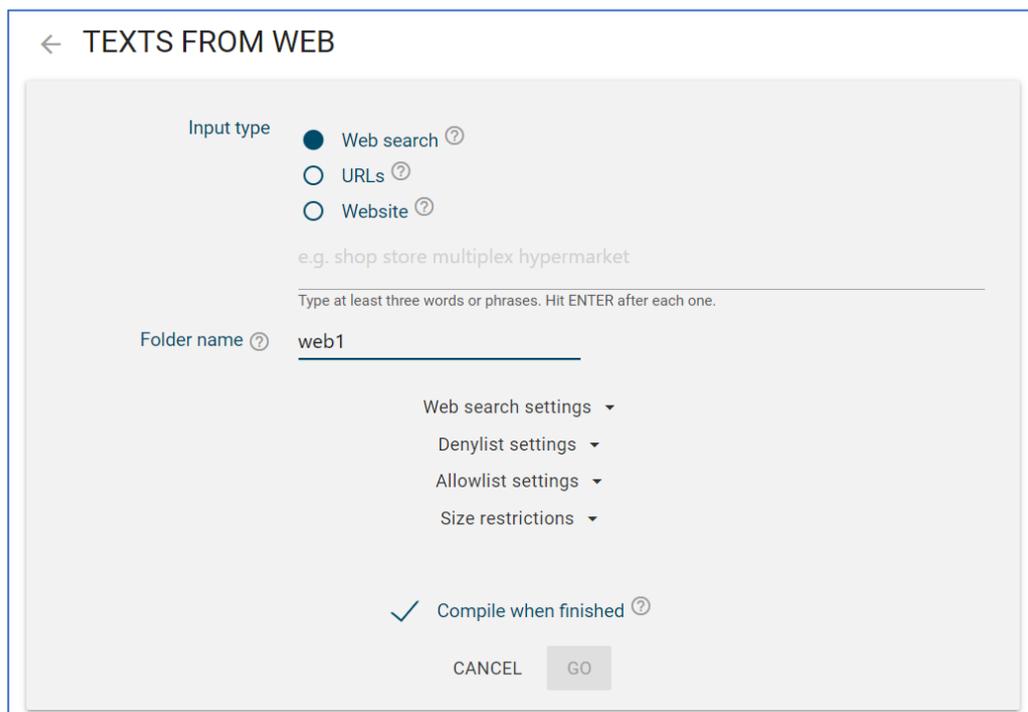


Figura 15 - Creare un corpus dal web

Qui vedremo come usare l'opzione **WEB SEARCH**: prima di tutto dobbiamo fornire alcune parole o sintagmi chiave (*seed words*) che definiscano l'argomento (*topic*) del nostro futuro corpus. SkE combinerà queste parole in vari modi casuali, a gruppi di tante parole quante quelle da noi scritte (es. se ne abbiamo scritte 3, gruppi di 3) e li presenterà al motore di ricerca Bing, il quale cercherà su internet e manderà a SkE indirizzi di pagine web che combaciano con le parole chiave fornite. SkE scaricherà queste pagine web rimuovendo tutti i contenuti linguisticamente irrilevanti (pubblicità, menù di navigazione, ecc.), quindi convertirà i testi in un corpus.

Quante parole chiave è meglio usare? Dipende da cosa ci serve, non c'è una risposta giusta o sbagliata. Dobbiamo sapere che più parole usiamo, più grande e probabilmente generico sarà il nostro corpus; al contrario, meno parole usiamo, più piccolo e quindi preciso e focalizzato sarà il nostro corpus.

Possiamo gestire la ricerca di Bing usando "**Web search settings**" (Figura 16): cliccandoci sopra di apriranno altre voci. Vediamo che di default le opzioni sono già impostate e gestite da SkE, ma possiamo anche decidere di estirle manualmente attivando il flag "Set values manually", dal quale possiamo regolare:

- **Max URLs per search:** questa opzione ci serve per gestire la dimensione del corpus.
- **Seed word in search:** questa opzione ci permette di regolare la quantità di parole da usare per comporre i gruppi randomici per la ricerca di Bing.
- **Sites list:** questa opzione ci serve per limitare la ricerca di Bing solo ad alcuni siti web.

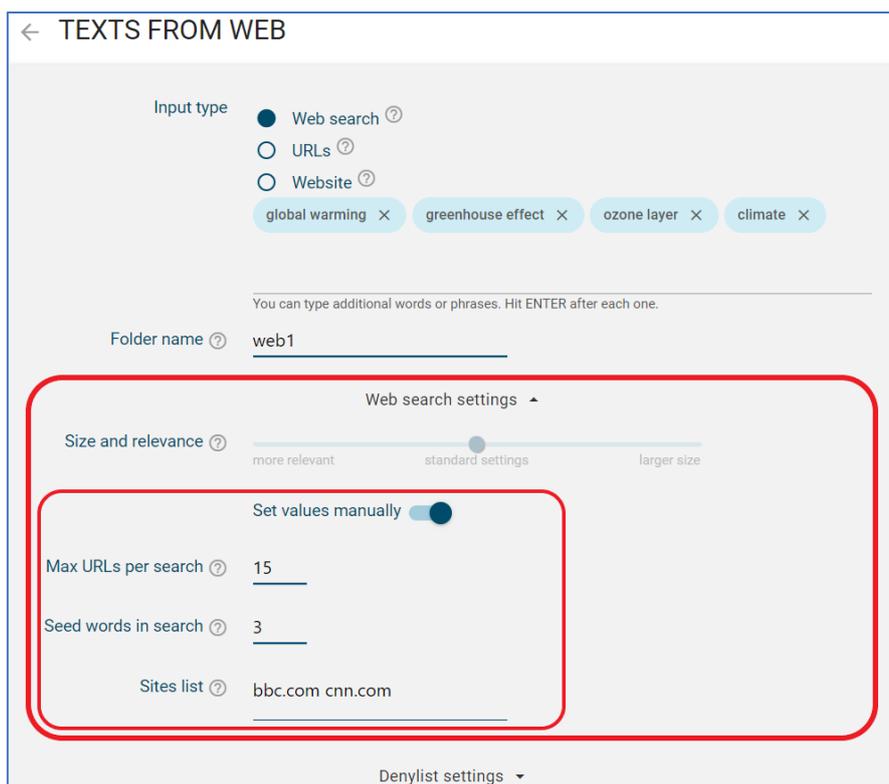


Figura 16 - Creare un corpus dal web – Web settings

Una volta impostata la ricerca per il nostro corpus, clicchiamo su “Go”, in basso: ci apparirà la schermata in Figura 17, dove SkE ci presenta le pagine web che ha trovato, in modo da poter ulteriormente agire sulla compilazione del corpus.

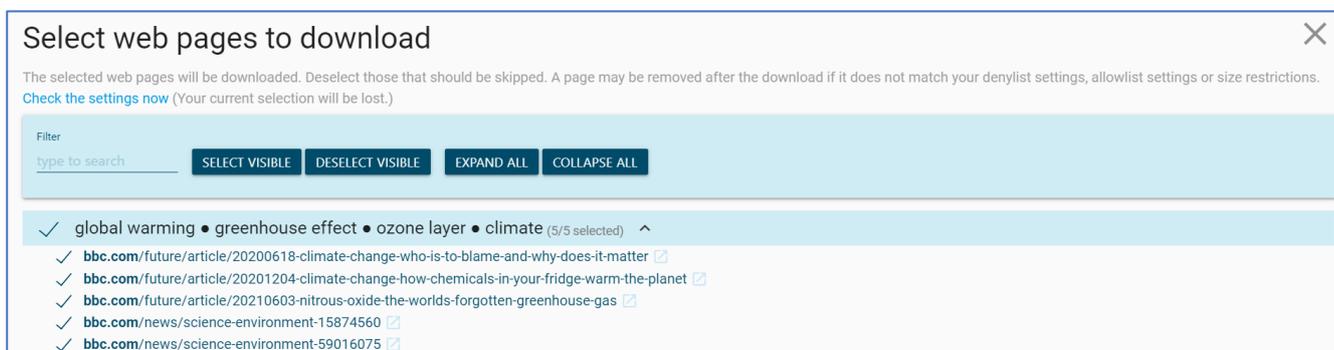


Figura 17 - Creare un corpus dal web – Pagine scaricate

Se va tutto bene (o dopo che abbiamo apportato ulteriori modifiche), clicchiamo nuovamente su “Go”, in basso, e attendiamo che il download delle pagine e la conversione dei relativi testi in corpus sia terminata.

3.2.2. Come costruire un corpus con testi nostri - I have my own texts

Se abbiamo già dei testi nostri (sul nostro pc), possiamo usarli per creare un corpus. → **I have my own text.**

Una volta selezionata questa opzione, SkE ci chiederà di caricare dei files dal nostro pc; tuttavia è anche possibile incollare un testo usando l’opzione “or paste text” in alto a destra (Figura 18).



Figura 18 - Creare un corpus dal nostro PC

NB: SkE non sa processare immagini di alcun tipo, quindi ovviamente queste ultime non possono essere caricate (chiaramente anche se si tratta di immagini di testi). Tuttavia, SkE supporta diversi tipi di file, tra cui i PDF (ovviamente qualsiasi tipo di immagine contenuto in essi viene ignorato).

Se il file ha un layout complesso (es. un poster pubblicitario), l'ordine del testo potrebbe non essere letto correttamente.

Possiamo controllare il processamento dei testi cliccando su "upload": da qui vedremo a che punto di download sono tutti i file che vogliamo caricare. Una volta scaricati, possiamo intervenire ulteriormente sui testi: selezioniamo quello che ci interessa e da "Bulk actions" (in basso) scegliamo se eliminarlo o se **editare i metadata** (questo possiamo farlo anche in un secondo momento, non è necessario farlo in questa fase se non vogliamo o non abbiamo tempo).

Se va bene così o quando abbiamo finito di editare i testi, procediamo cliccando su "Next" e poi "Compile" per convertire i testi scaricati in un corpus.

Che abbiamo scelto di caricare testi dal web o dal nostro pc, **una volta terminato il processo di *compiling*** apparirà la schermata in Figura 19:

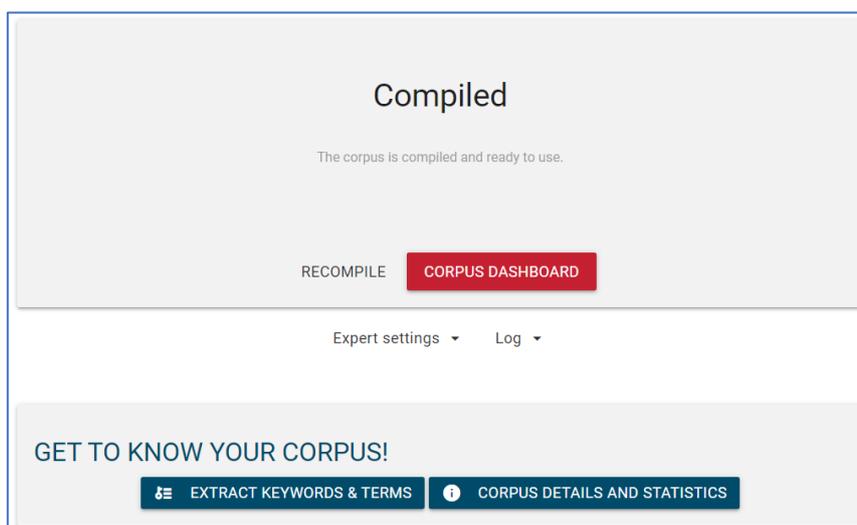


Figura 19 - Creare un corpus dal web - Corpus compilato

Possiamo controllare tutte le informazioni relative al nostro corpus dalla relativa pagina informativa, alla quale possiamo accedere già da questa schermata con **“Corpus details and statistics”**, in basso a destra. Accanto a questa opzione troviamo anche **“Extract keywords & terms”**: da qui possiamo estrarre la terminologia per controllare velocemente se i testi nel corpus sono effettivamente relazionati all’argomento che volevamo.

Ad ogni modo...

Il nostro corpus è ora pronto e disponibile su MY CORPORA!

Una volta creato un corpus su SkE, è possibile gestirlo usando **“Manage corpus”** (NB: prima bisogna selezionare il corpus, come quando dobbiamo usarlo per fare una ricerca), in alto a destra della Dashboard  (vedasi le varie opzioni in Figura 20).

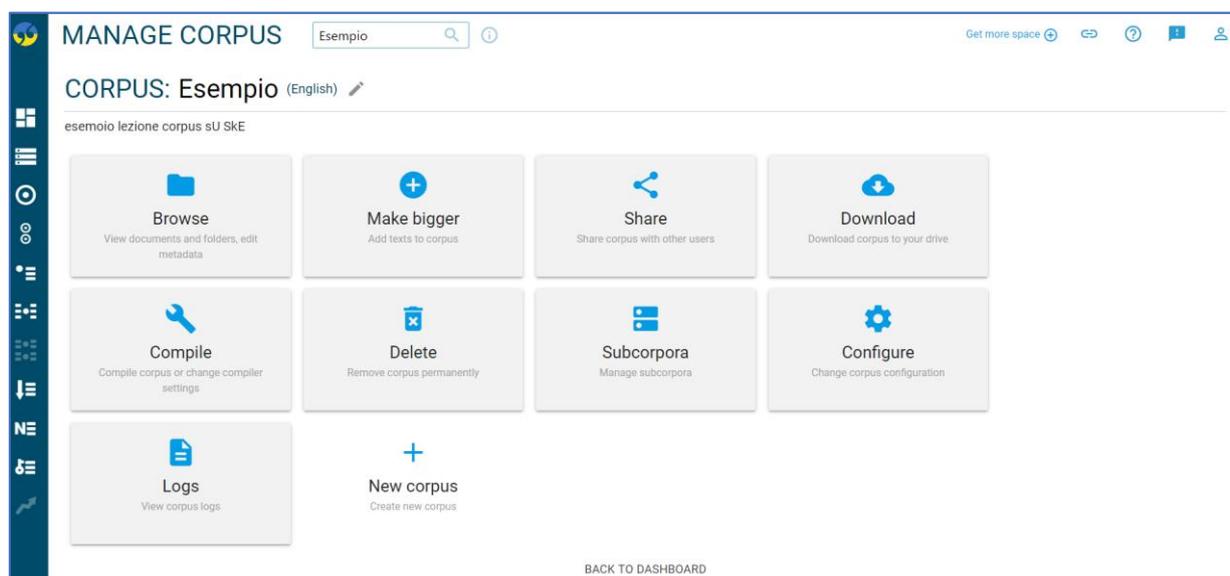


Figura 20 - Manage corpus

3.3. Come creare dei subcorpora

Una volta creato un corpus su SkE è possibile dividerlo in (illimitate) parti più piccole, dette **subcorpora**: queste parti di corpus possono essere analizzate singolarmente o in modo contrastivo.

NB: i subcorpora non usano ulteriore spazio di archiviazione.

Ci sono 3 modi diversi per creare un subcorpus:

- da una concordanza (Concordance):** crea subcorpora da frasi, paragrafi, documenti o altre strutture che contengono il risultato di ricerca su Concordance.
 - Generare una concordanza con basic o advanced Concordance (è indifferente).
 - Aprire il nuovo subcorpus popup.
 - Rinominare il subcorpus.
 - Selezionare la struttura che si vuole usare (doc = documenti; p = phrases, sintagmi; s = sentences, frasi).
- da una combinazione di tipi di testi (text types),** detti anche *metadata* o *annotation* (= annotazione). Per esempio possiamo chiedere a SkE di creare un subcorpus basato su luogo e anno di pubblicazione (es. tutti i testi del nostro corpus pubblicati in Canada nel 2023).

- a. Andare su Manage Corpus oppure usare il pulsante col simbolo + accanto alla voce “Subocorpus” nelle versioni advanced di ricerca.
- b. Rinominare il subcorpus e selezionare i metadata desiderati (es. luogo e anno).

3. da un testo di definizione (definition file)

Alla fine della procedura scelta, cliccare su “create subcorpus” → Il subcorpus è pronto.

NB: non è possibile aggiungere ulteriori dati al subcorpus una volta che è stato creato.

Generalmente i subcorpora si possono trovare nelle versioni avanzate degli strumenti di ricerca di SkE. Ovviamente la selezione di un dato subcorpus restringe la ricerca solo a quel subcorpus.

La lista dei subcorpora si trova nella pagina informativa del corpus principale o su Manage corpus.

Quando visualizziamo le informazioni di un subcorpus, vediamo il **numero di tokens, parole e percentuale**: il numero di tokens è esatto; il numero di parole è una stima; la percentuale indica la parte del corpus rappresentata dal subcorpus.

NB: eliminare un subcorpus non comporta l’eliminazione di quella parte del corpus!

4. Gli strumenti di Sketch Engine

Sketch Engine ha vari strumenti per fare ricerca e che ci permettono di ottenere in modo automatico diversi tipi di informazione in base a quello che ci serve.

Non tutte le funzioni sono sempre disponibili, dipendendo dal corpus o dal nostro abbonamento. Quelle utilizzabili sono evidenziate in bianco, mentre le altre restano in azzurro come la schermata (Figura 21).

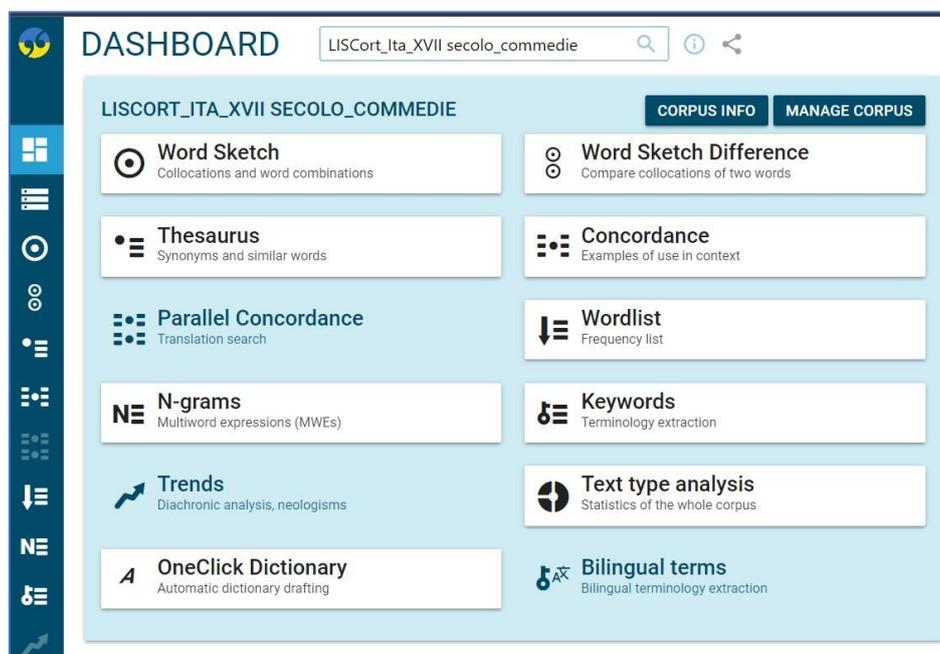


Figura 21 - Schermata funzioni di Sketch Engine

Per ogni funzione è disponibile:

- **Change criteria** (🔍): ci permette di cambiare i criteri di ricerca.
- **Download** (↓): ci permette di scaricare i risultati ottenuti dalla nostra ricerca.
- **View options** (👁): ci permette di modificare cosa o come visualizzare i risultati della ricerca.
- **Filter results** (☰): ci permette di filtrare i risultati, per esempio decidendo di vedere tutte le occorrenze in posizione iniziale o finale, ecc.
- **Show visualization** (🌐): ci permette di modificare la visualizzazione dei dati, per esempio attraverso immagini.
- **Results information** (ℹ): ci permette di vedere alcune i criteri di ricerca usati in quel momento.
- **Add to favourites** (★): ci permette di aggiunge la ricerca fatta ai preferiti, in modo da poterla rivedere quando vogliamo senza bisogno di rifarla.

4.1. Concordance

Se vogliamo vedere **esempi di parole e sintagmi contestualizzati**, possiamo usare la funzione Concordance. Vedremo sia la versione BASIC che quella ADVANCED. Di quest'ultima in particolare vedremo come usare la CQL (Corpus Query Language).

Con la **versione BASIC** di Concordance, digitiamo una parola o sintagma (in forma lemmatica o no) e clicchiamo su "Search": otterremo tutte le occorrenze di quella data parola o sintagma all'interno del nostro corpus, nel loro relativo contesto di occorrenza (Figura 22).

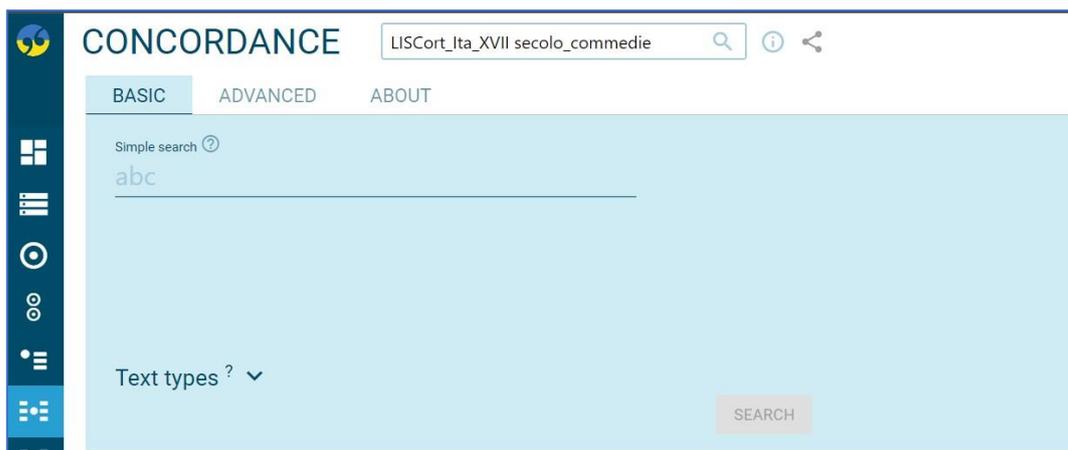


Figura 22 – Concordance BASIC

Nella nostra ricerca su Concordance, possiamo scegliere di inserire:

- forma lemmatica (come da dizionario): in questo caso SkE ci mostrerà tutte le occorrenze di quella parola in tutte le sue forme (es. *andare* -> *va, vai, vanno, andai, andò, andrò, andrai, ecc.*).
- forma non lemmatica: in questo caso SkE ci mostrerà tutti gli esempi contenenti quella forma che abbiamo scelto (es. *andai* -> *andai*).
- un asterisco (*) (sta per qualsiasi numero, parola o carattere non specificato):
 - tra due parole: in questo caso SkE ci mostrerà tutti i risultati in cui una parola compare tra le due da noi specificate (es. *fare * spesa* -> *fare la spesa, fare le spese, fare delle spese, fare una spesa, ecc.*).
 - prima o dopo una parte di parola: in questo caso SkE troverà tutte le parole che iniziano o terminano nel modo che abbiamo indicato o le parole il cui lemma inizia o termina come abbiamo indicato (es. *dir** -> *dire, dirò, detto, disse, ecc.*).
- un punto interrogativo (?) (sta esattamente per un carattere non specificato):
 - tra due parole: in questo caso SkE ci mostrerà tutti i risultati in cui una parola formata da un solo carattere compare tra le due da noi specificate (es. *fare * spesa* -> *fare la spesa, fare le spese, fare delle spese, fare una spesa, ecc.*).
 - prima o dopo una parte di parola: in questo modo SkE ci mostrerà tutte le parole formate da un unico carattere che precedere o segue ciò che abbiamo indicato (es. *?ato* -> *fatto, fatti, andato, andati, grato, grata, ecc.*) o le parole il cui lemma inizia o termina nel modo che abbiamo indicato (es. *?are* -> *fa, fatto, diede, dato, ecc.*) preceduto da un unico carattere.

Figura 23 - Esempio di ricerca con Concordance

Come possiamo vedere dalla Figura 23, nel mostrarci le concordanze, SkE evidenzia le concordanze (**KWIC** = key words in context) in rosso e ci mostra il contesto che le precede (Left context) e che le segue (Right context).

Nei criteri di ricerca di Concordance posso scegliere di limitare la ricerca a solo una parte del mio corpus, selezionando l'opzione Text type (Figura 24).

Figura 24 - Concordance-Text type in Change criteria

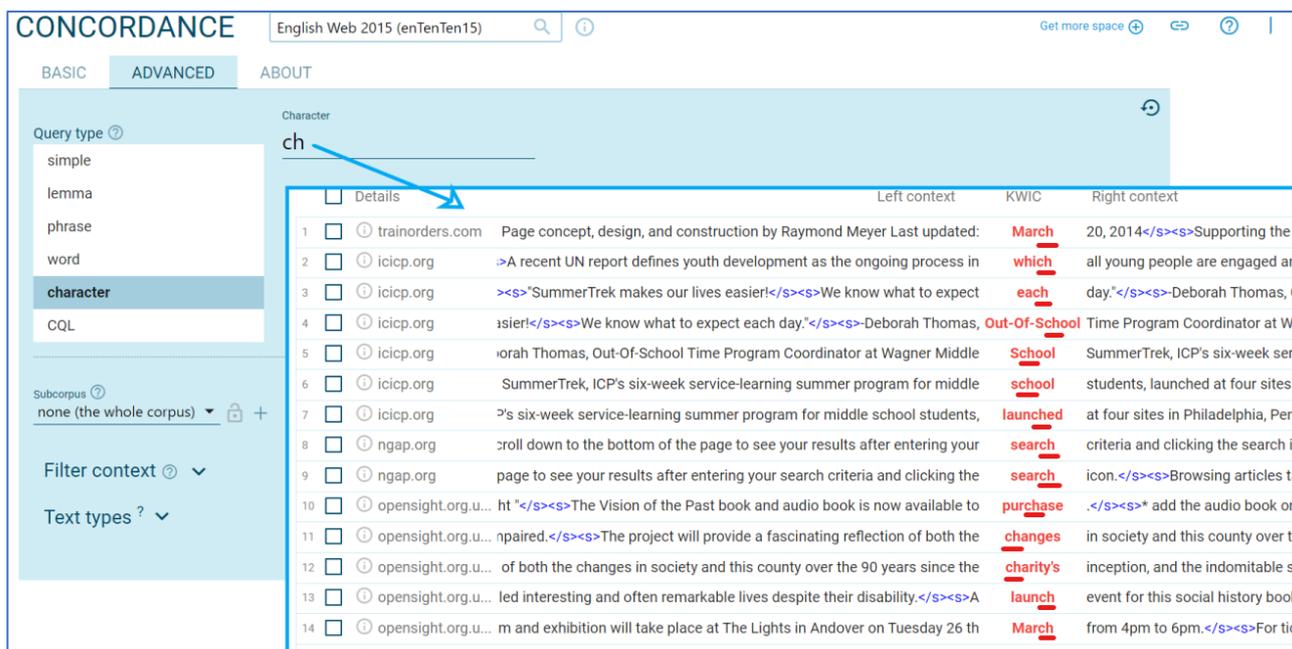
Nella **versione ADVANCED** di Concordance troviamo varie opzioni di ricerca o **Query types** (Figura 25):

- **Simple**: è esattamente lo stesso tipo di ricerca di Concordance BASIC, quindi possiamo scrivere singole parole o interi sintagmi e, se scritti in forma lemmatica, SkE troverà tutte le varie forme attestate nel corpus.
- **Lemma**: troverà tutte le forme appartenenti al lemma che indichiamo.
ATTENZIONE: possiamo indicare solo 1 lemma!
Possiamo anche scegliere di specificare quale tipo di parola desideriamo cercare (es. solo nomi, verbi, aggettivi, ecc.) selezionandolo dalla tabella "Part of speech" accanto all'elenco "Query type".

Query type ?	Part of speech
simple	any
lemma	adjective
phrase	adverb
word	conjunction
character	noun
CQL	preposition
	pronoun
	verb

1. *dis.** → troverà tutti lemmi che iniziano con *dis-*.

- **Phrase:** troverà esattamente il sintagma che abbiamo indicato (es. *go home* → *go home* VS. simple search *go home* → *go home, goes home, went home, ecc.*)
- **Word:** troverà esattamente la parola che abbiamo indicato (es. *go* → *go* VS. simple search *go, goes, went, going, gone*).
- 2. *.et* → troverà tutte le parole di 3 lettere che finiscono con *-et*.
- **Character:** troverà tutte le parole o tokens che contengono il carattere o sequenza di caratteri che abbiamo indicato.



The screenshot shows the Concordance search interface. The search query is 'ch' under the 'Character' type. The results table is as follows:

	Details	Left context	KWIC	Right context
1	trainorders.com	Page concept, design, and construction by Raymond Meyer Last updated:	March	20, 2014</s><s>Supporting the
2	icicp.org	>A recent UN report defines youth development as the ongoing process in	which	all young people are engaged an
3	icicp.org	><s>"SummerTrek makes our lives easier</s><s>We know what to expect	each	day."</s><s>Deborah Thomas, O
4	icicp.org	asier</s><s>We know what to expect each day."</s><s>Deborah Thomas, Out-Of-School		Time Program Coordinator at Wa
5	icicp.org	borah Thomas, Out-Of-School Time Program Coordinator at Wagner Middle	School	SummerTrek, ICP's six-week ser
6	icicp.org	SummerTrek, ICP's six-week service-learning summer program for middle	school	students, launched at four sites
7	icicp.org	's six-week service-learning summer program for middle school students,	launched	at four sites in Philadelphia, Pen
8	ngap.org	roll down to the bottom of the page to see your results after entering your	search	criteria and clicking the search ic
9	ngap.org	page to see your results after entering your search criteria and clicking the	search	icon.</s><s>Browsing articles ta
10	opensight.org.u...	ht "</s><s>The Vision of the Past book and audio book is now available to	purchase	</s><s>* add the audio book on
11	opensight.org.u...	npaired.</s><s>The project will provide a fascinating reflection of both the	changes	in society and this county over th
12	opensight.org.u...	of both the changes in society and this county over the 90 years since the	charity's	inception, and the indomitable s
13	opensight.org.u...	led interesting and often remarkable lives despite their disability.</s><s>A	launch	event for this social history book
14	opensight.org.u...	m and exhibition will take place at The Lights in Andover on Tuesday 26 th	March	from 4pm to 6pm.</s><s>For tid

- **CQL:** la Corpus Query Language è utile per fare ricerche complesse, infatti può combinare tutti tipi di ricerca che abbiamo visto in un'unica ricerca. Questo permette di trovare schemi lessicali o grammaticali senza specificare parole concrete (es. tutti i nomi non preceduti da un articolo ma seguiti da un verbo al tempo passato).

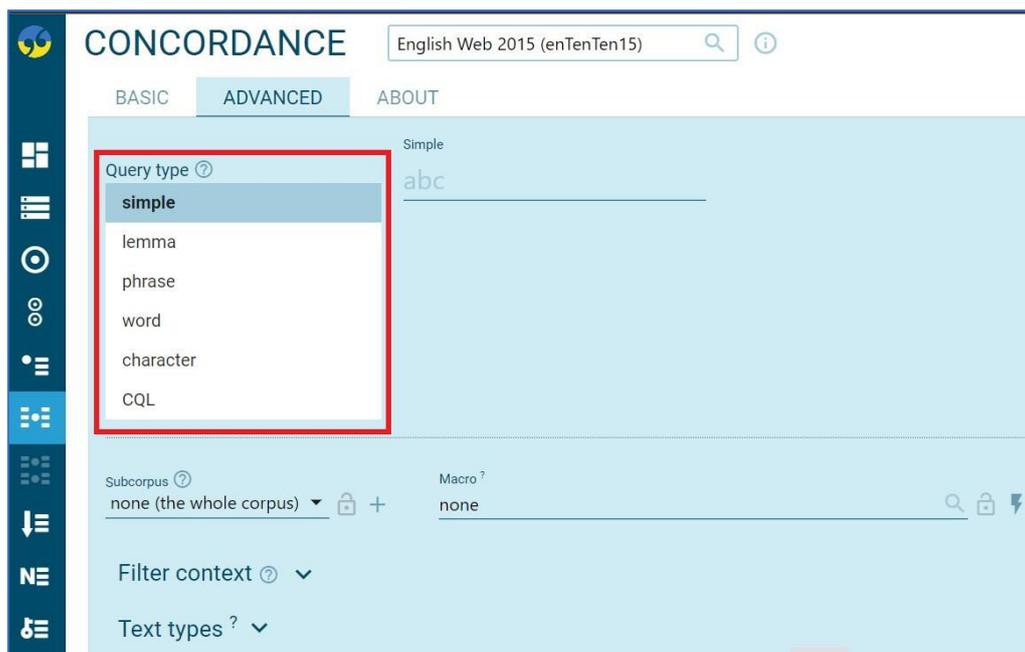


Figura 25 - Concordance ADVANCED

È possibile **limitare la ricerca a una parte del corpus** selezionandola tra i vari subcorpora elencati alla voce **“Subcorpus”** in basso a sinistra (Figura 26). È inoltre possibile definire il nostro subcorpus usando il segno + accanto alla barra di ricerca di “Subcorpus”: da qui possiamo usare la barra di digitazione o le varie tendine per selezionare gli elementi che ci interessano. Lo stesso principio è applicabile alla voce “Text Type” poco più in basso.

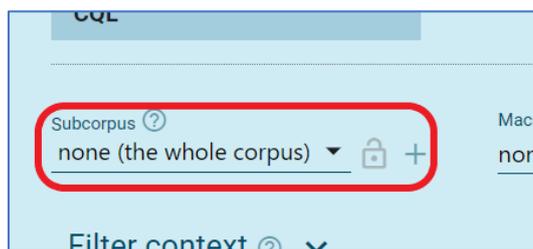


Figura 26 - Scegliere subcorpora

Possiamo impostare filtri per rimuovere linee indesiderate dalle concordanze usando **“Filter context”**, in basso (Figura 27). Quindi se per esempio cerchiamo il lemma *work* (con la ricerca “lemma”) e vogliamo escludere tutti i risultati *home work*, su “Filter context” selezioniamo “Lemma context” e scriviamo sulla barra in basso *home work*. Se serve, possiamo precisare all’interno di quante parole e dove dovrebbero posizionarsi i tokens da escludere.

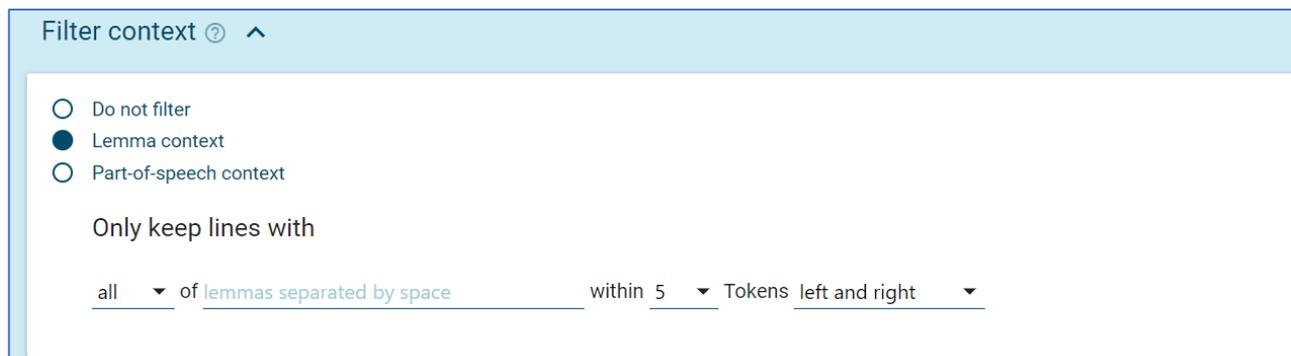


Figura 27 - Filter context – Lemma context

Possiamo anche tenere solo le linee che contengono parti del discorso, nel qual caso in “Filter context” dovremo usare “Part-of-speech context” (Figura 28). Quindi per esempio cerchiamo il lemma *money* (con la ricerca “lemma”) e in “Part-of-speech” scegliamo di mantenere solo le linee con avverbi a sinistra non superiori a 1 token.

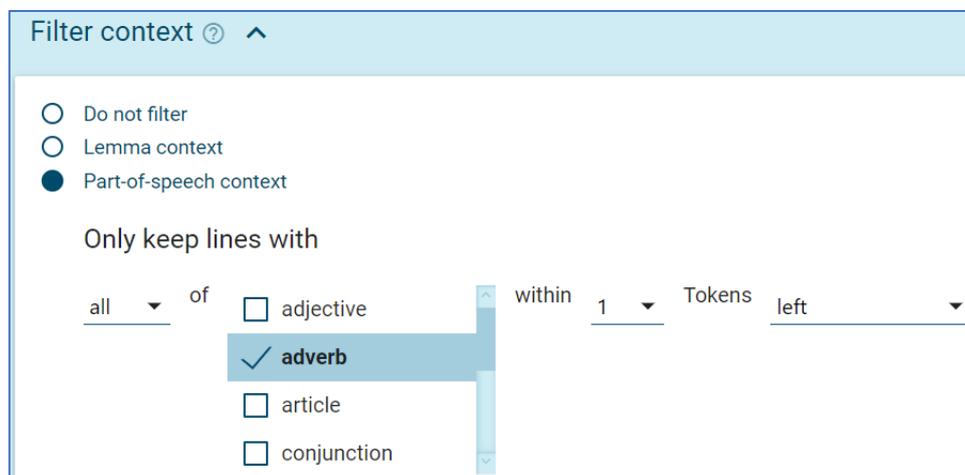


Figura 28 - Filter context - Part-of-speech context

4.1.1. CQL – Corpus Query Language

CQL (Corpus Query Language) è utile per cercare **lessico complesso** (es. Nome + 2 parole + Verbo) o **strutture grammaticali** (es. have + a + Agg + time) in un corpus. CQL è un linguaggio complesso che può combinare in un’unica ricerca parole, tags di parti del discorso, strutture del corpus (es. inizi e terminazioni di frasi o paragrafi), o anche criteri non specificati.

Possiamo utilizzare questo tipo di ricerca su SkE usando la **versione Advanced di Concordance** e selezionando, appunto, la ricerca con CQL (Figura 29).

NB: per fare ricerche simili possiamo usare anche Word Sketch, ma è bene conoscere e saper usare correttamente anche CQL.

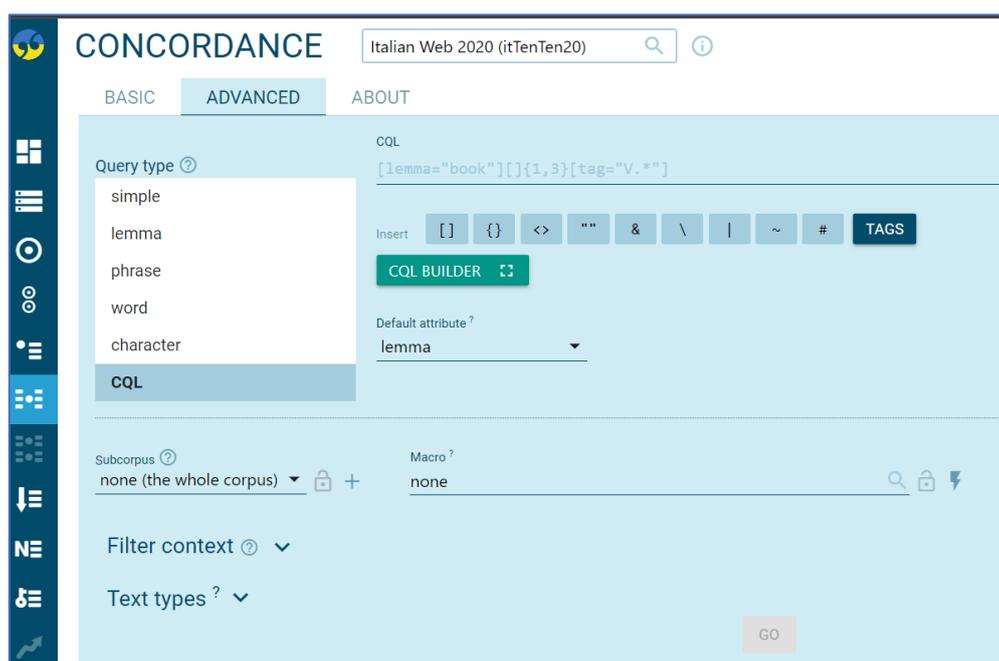


Figura 29 - CQL nella versione Advanced della funzione Concordance

4.1.1.1. CQL – Istruzioni di base

Per scrivere un linguaggio CQL dobbiamo indicare dei **tokens**, ovvero parole e segni di punteggiatura, inserendoli tra parentesi quadrate [].

A meno che non cerchiamo un token qualsiasi, nelle [] dobbiamo sempre indicare cosa stiamo cercando:

- **word**: la forma della parola che indichiamo così come la scriviamo.
- **lemma**: qualsiasi forma della parola che indichiamo.
- **tag**: tipo di parola.

Queste indicazioni vanno seguite dal segno = e ciò che stiamo cercando tra " ". Quindi, per esempio, scriveremo: [word="mind"].

Esempi di **tokens** con CQL:

1. [] → qualsiasi parola.
2. [word="mind"] → troverà solo la forma *mind*.
3. [lemma="mind"] → troverà tutte le forme di *mind*, quindi anche *mind*, *mind*, *mind*.
4. [tag="Pronoun"] → troverà tutti i pronomi.

In questo esempio abbiamo scritto "Pronoun" per facilitare la comprensione, ma ATTENZIONE: in questo caso dobbiamo scrivere correttamente il tag del tipo di parola che stiamo cercando così come viene utilizzato nel corpus che stiamo usando per la nostra ricerca. Se non sappiamo o non ricordiamo i tags usati nel corpus utilizzato, possiamo usare il **pulsante TAGS** sotto alla barra di digitazione e selezionare da lì ciò che ci interessa (Figura 30). Questo vale per tutte le volte in cui dobbiamo indicare un tag.

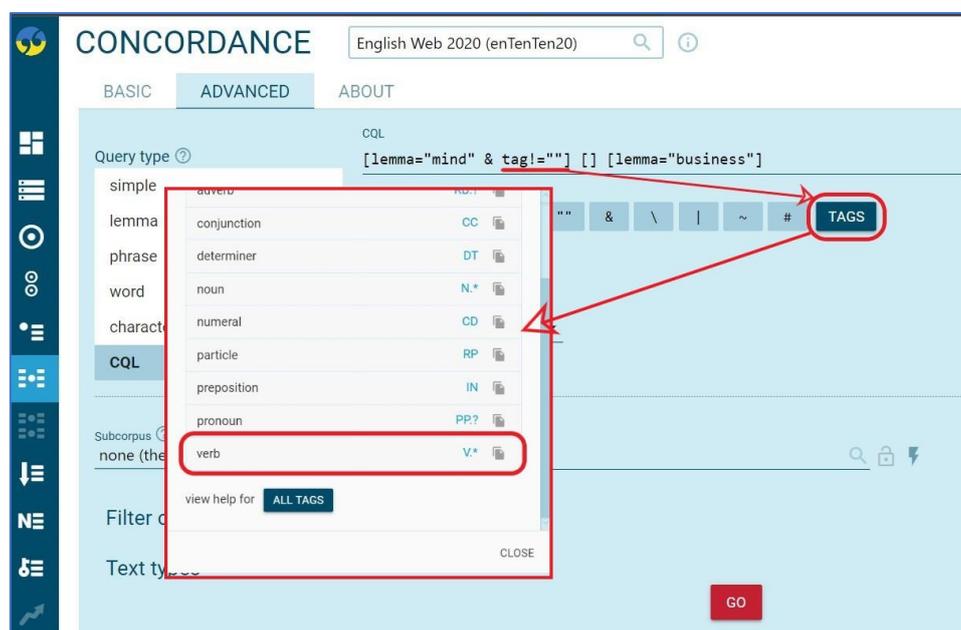


Figura 30 - TAGS per selezionare tags su CQL

5. [lemma="mind" & tag!="Noun"] → troverà tutte le forme di *mind*, ma solo nelle occorrenze in cui è un verbo; infatti la sintassi "& tag!="Noun"" significa "e non è un nome".
6. []{0,3} → troverà 0, 1, 2 o 3 tokens qualsiasi.
7. [lemma="own"]? → troverà *own* in una sequenza di tokens (es. una frase) solo se presente.
8. [lemma="proposal|agreement"] → troverà frasi contenenti *proposal* o *agreement* (vedasi *containing* e *within* più avanti).

NB:

- Possiamo mettere spazi tra una parentesi e l'altra (es. [lemma="mind"] []), ma non nelle citazioni indicate nel CQL (es. " mind ").
- Quando usiamo le virgolette con CQL dobbiamo usare quelle dritte, non quelle curvate.



Esempi di sintassi in CQL per fare ricerche su Concordance, andando da una sintassi più semplice e quindi piuttosto generica a una più complessa e quindi maggiormente precisa:

CQL

- (1) `[lemma="mind"] [] [lemma="business"]` = Trova una sequenza di 3 tokens dove: il primo è una qualsiasi forma di *mind*; il terzo è un qualsiasi token; il terzo è una qualsiasi forma di *business*.

Left context	KWIC	Right context
p-to-date tips on	minding your business	manners, The S
site - so you can	mind your business	./s><s>The·Te
l the hard way to	mind my business	, without judgin
/hen I told her to	mind her business	, she went to yo
ic institutions to	mind the business	of daily govern
ou are quiet and	mind your business	, no one would
even when I was	minding my business	somehow.</s><
Seconds</s><s>	Minding Hollywood's Business	</s><s>After a
sta, pay your bill,	mind your business	" tend to your t

CQL

- (2) `[lemma="mind" & tag!="N.*"] [tag="PP.?"] [lemma="business"]` = Trova una sequenza di 3 tokens dove: il primo è una qualsiasi forma di *mind* e NON è nome; il secondo è un pronome; il terzo è qualsiasi forma di *business*.

Left context	KWIC	Right context
p-to-date tips on	minding your business	manners, The S
site - so you can	mind your business	./s><s>The·Te
the hard way to	mind my business	, without judgin
hen I told her to	mind her business	, she went to yo
ou are quiet and	mind your business	, no one would l
even when I was	minding my business	somehow.</s><
sta, pay your bill,	mind your business	" tend to your ta
he while and not	minding his business	./s><s>So gre

CQL

- (3) `[lemma="mind" & tag!="N.*"] [tag="PP.?"] [lemma="own"]? [lemma="business"]` = Trova una sequenza di 4 tokens dove: il primo è una qualsiasi forma di *mind* e NON è nome; il secondo è un pronome; il terzo è *own*, ma solo se presente, altrimenti ignoralo; il quarto è qualsiasi forma di *business*.

Left context	KWIC	Right context
just sitting there,	mind ing their own business	, waiting to be d
keep you place,"	mind your own business	," and basically d
spirit wants us to	mind our own business	and look the oth
he other way and	mind ed their own business	./s><s>They ke
w that I ought to	mind my own business	but out of conce
ted to tell him to	mind his own business	./s><s>I also fe
ver mind that, sir;	mind your own business	; -I'll have no rel
everyone knows:"	mind your own business	./s><s>The To
gress rather than	mind our own business	./s><s>Is that t
man tells him to	mind his own business	./s><s>As he l

CQL

- (4) `[lemma="mind" & tag!=N.*] []{0,3} [lemma="business"]` = Trova una sequenza di 3 o più tokens dove: il primo è una qualsiasi forma di *mind* e NON è nome; il secondo può essere 0, 1, 2 o 3 random tokens; il terzo è qualsiasi forma di *business*.

Left context	KWIC	Right context
ying the day and	mind ing our own business	, when this small d
erwise I MYOB --	mind my own business	-- and leave it alone
shop owner was	mind ing his business	, doing his job, earn
meet with a friend,	mind ing her own business	./s><s>There's no
were all at least	mind ing their own business	and behaving them
the paper should	mind its business	, because, like any
ne tom, who was	mind ing his own business	as he padded past
still exists, never	mind such an anachronistic business	./s><s>As a bonus
/s><s>Remley is	mind ing his own business	on the street when
nt our country to	mind its own business	./s><s>The averag

CQL

- (5) `[word!="a|an|the"] [tag="N.*"] [tag="V.D"]` = Trova tutti i nomi non preceduti da un articolo ma seguiti da un verbo al tempo passato.

Left context	KWIC	Right context
nd colonialism as	racial categorisation played	a large part in the
ed the idea that a	racialist hierarchy existed	that placed those
also the idea that	White Europeans wished	to continue their
second World War	British politics reinforced	have the idea tha
black immigrants	entering Britain needed	to be restricted a
n was reinforced,	as people began	to see immigrant
9)./s><s>By the	1970s legislation had	been bought in w
92)./s><s>This	political racism persisted	after 1979, with

Default attribute: serve per attribuire autenticamente ai tokens il modo in cui devono essere cercati (es. come lemma, word, ecc.), ma viene applicato solo ai tokens che non vengono messi tra parentesi quadre [] e che non hanno altre indicazioni di attributo (es. in Figura 31). Può essere utile per velocizzare la scrittura in CQL.

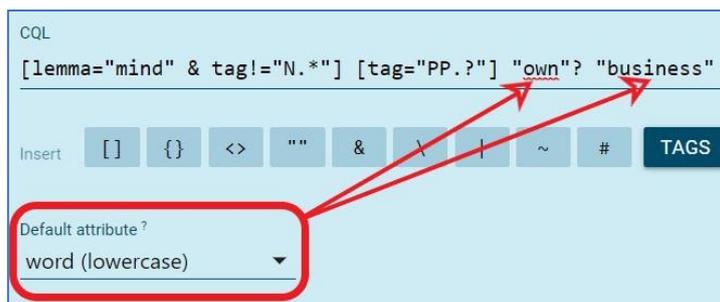


Figura 31 - Default attribute in CQL

Vediamo ora come trovare frasi che iniziano con un nome, tokens alla fine di frasi, come cercare documenti, paragrafi e altre strutture in cui un corpus può essere diviso.

Normalmente le **strutture** che possiamo trovare in un corpus sono frasi, paragrafi e documenti, che in CQL vengono indicati con simboli appositi dentro ai simboli di maggiore e minore <>:

- **frasi** (*sentences*): <s> (inizio), </s> (fine), <s/> (intero).
- **paragrafi** (*paragraphs*): <p> (inizio), </p> (fine), <p/> (intero).
- **documenti** (*documents*): <doc> (inizio), </doc> (fine), <doc/> (intero).

Attenzione!!! I simboli qui indicati sono quelli tipicamente usati per indicare le strutture nei corpora, ma non sono universali. Inoltre anche le strutture disponibili possono cambiare da corpus a corpus (alcuni possono averne di più, altri di meno; alcuni possono averne cert'une ed altri cert'altre). Le strutture disponibili per ogni corpus, insieme ad altre informazioni, sono elencate nella **pagina informativa del corpus** (*corpus information page*), facilmente consultabile come mostrato in Figura 32.

TEXT TYPES	TEXT TYPE ANALYSIS
<bncdoc> (28)	4,054
<note> (6)	17,155
<div1> (4)	83,771
<div2> (4)	72,678
<div3> (4)	38,120
<s> (3)	6,052,190
<head> (3)	222,818
<caption> (3)	89,990
<gap> (3)	87,369
<corr> (3)	16,681
<div4> (3)	12,495
<sic> (3)	7,796
<div> (3)	3,783
<stage> (3)	508

Figura 32 - Scheda info corpus

Per visualizzare le strutture nei risultati di ricerca, nel caso non fossero già visibili, usare View options (👁).

Come trovare:

(6) parole all'inizio di una frase (che appaiono come il primo token di una frase), per esempio nomi.

CQL
<s>[tag="N.*"]

Left context	KWIC	Right context
aval Discussion Forums.</s><s>	Clientcopia	: Stupid Client Qu
nt?</s><s>Sure we can...</s><s>	Clientcopia	was created to gi
at a nice local restaurant.</s><s>	Bob	likes to plan ahead
mother, wife and liberal,...</s><s>	Editor	, Noblesville has
d Trace to Hague Road...</s><s>	Commentary	by Ward Degler A
fourth grades.</s><s>It...</s><s>	Questions	and Comments f
daho's anadromous fish.</s><s>	Electricity	is almost always
om the Pacific Northwest</s><s>	ApacheCon	North America 20
s across the foundation </s><s>	OpenSym	2018, the 14th In

(7) parole alla fine di una frase (che appaiono come il ultimo o penultimo token di una frase).

CQL
[]</s>

NB: In questo caso meglio indicare che cerchiamo 2 tokens a fine frase perché uno di questi (l'ultimo) potrebbe essere un simbolo di punteggiatura (. ! ?).

Left context	KWIC	Right context
<s>Welcome to KBismarck.org! </s><s>		This is a community of friend
Naval Discussion Forums. </s><s>		Clientcopia : Stupid Client Qu
opia : Stupid Client Quotes... </s><s>		There's no getting around it.
no getting around it. </s><s>		At some point in your career
and the future of mankind. </s><s>		You may have dealt with one
you like a deer in headlights. </s><s>		Dumbfounded to utter anyth
hing but an "uhhh ..." </s><s>		.</s><s>Some clients have r
ng but an "uhhh... "</s><s> </s><s>		Some clients have no conce
ave no concept of reality. </s><s>		They make up their mind, jus

(8) la prima parola di tutti i documenti pubblicati in una certa data e contenuti nel nostro corpus.

CQL
<bncdoc year="1993">[]

(bnc = British National Corpus)

	Details	Left context	KWIC	Ri
1	<input type="checkbox"/>	1993 market.</s><s>TOP 30 WORLDWIDE DUTY FREE SPIRITS BRANDS IN 1991</s><s>	Longcase	clo
2	<input type="checkbox"/>	1993 sponsoring smaller groups and provide a useful cross cultural experience.</s><s>	Western	De
3	<input type="checkbox"/>	1993 :d for payment, thus avoiding forfeiture of the lease.</s><s>PRECEDENTS</s><s>	CHAPTER	ON
4	<input type="checkbox"/>	1993 nd the lens to translate the language of photography.</s><s>Photography</s><s>	CHAPTER	ON
5	<input type="checkbox"/>	1993 sponsibility to identify, ultimately for himself, as his community's scheme.</s><s>	The	int
6	<input type="checkbox"/>	1993 r the headlease acting in the capacity in which he is appointed thereunder</s><s>	General	Co
7	<input type="checkbox"/>	1993 <s>Industrial action</s><s>Trade unions</s><s>Independent contractors</s><s>	1	Pre
8	<input type="checkbox"/>	1993 1 this 5th edn, was originally written by Mr Alan Hughes, Solicitor, in 1979.</s><s>	3	Da
9	<input type="checkbox"/>	1993 fecting his tax liability (Phipps v Orthodox Unit Trust Ltd [1958] 1 QB 314).</s><s>	The	Pla
10	<input type="checkbox"/>	1993 r whether it requires any continuing services to be provided by the vendor.</s><s>	2	Da

(9) la prima parola di tutti i documenti appartenenti a un determinato genere presente nel nostro corpus.

CQL

`<bncdoc genre="W_fict_prose">[]`
(bnc = British National Corpus)

Left context	KWIC	Right context
nes 6 February</s><s>	TOO	NOSEY</s><s>My name
of Antiquaries.</s><s>	At	first I wasn't sure at all v
ale at the time.</s><s>	LONDON	was misted over.</s><s>
>PRECEDENTS</s><s>	CHAPTER ONE</s><s>	'Good grief
: Arab masses.</s><s>	CHAPTER ONE</s><s>	'Well...?</s>
>Photography</s><s>	CHAPTER ONE</s><s>	'Hi, need an
ion redundant.</s><s>	CHAPTER ONE</s><s>	Shiona was
ata File 8.6.3.1</s><s>	CHAPTER ONE</s><s>	It was like c

NB: Come possiamo notare dagli esempi 7 e 8, possiamo usare anche **attributi/metadati assegnati alle strutture**, come la data, il nome del corpus, come indicare che stiamo cercando il genere, come si chiama il genere che stiamo cercando, ecc. Tutte queste informazioni sono contenute nella **scheda informativa del corpus** (Figura 33).

The screenshot shows a CQL search interface. At the top, the query `<bncdoc genre="W_fict_prose">[]` is entered. Below the query, a list of attributes is shown with their frequencies. The attribute `David Lee's classification, bncdoc.genre` has a frequency of 71. A table below this list shows the structure frequency for different genres:

Attribute value	Structure frequency ⁷
1 W_misc	500 ...
2 <u>W_fict_prose</u>	432 ...
3 W_pop_lore	211 ...
4 W_ac_polit_law_edu	186 ...
5 S_conv	153 ...
6 W_ac_soc_science	138 ...
7 S_meeting	133 ...

Figura 33 - Come ricercare info strutture e metadati per CQL nelle informazioni del corpus

Infine vediamo come usare nella sintassi CQL **within** e **containing**:

- **CONTAINING** si usa per trovare tutte le strutture che contengono qualcosa.
- **WITHIN** serve per trovare parole, sintagmi, strutture grammaticali o lessicali all'interno di frasi o documenti. È particolarmente utile per queries che contengono tokens non specificati.

CQL

(10) `<s/> containing [word="an?"][tag="N.*"][word="of"]` = Trova una frase che contenga una sequenza composta da un pronome indefinito, un nome e la parola *of*.

s	Left context	KWIC
84	t (pl.37).</s><s>	Equally, to begin consideration of the archaeological (as opposed to the specifically epigraphic) evidence,
84	on style.</s><s>	Given this limitation therefore, one has to rely on an interpretation of the relationships
84	mselves.</s><s>	Such an interpretation involves us, immediately, with a consideration of style and, by implicat
84	mosaics.</s><s>	In general, one notable but not fully comprehensive style (or two or more styles which do not correspond) is in evidence: another, "indefinite" style accom
84	helpers.</s><s>	Moreover, the relationship between mosaics need not necessarily be "direct": a group of mosaics can be included at this level when the style of its m
84	e others.</s><s>	In such cases a number of craftsmen appear to be working together (to the exclusion of oth
84	vements.</s><s>	A predominant or composite style (an association of styles) seems to orientate itself in respect to a group
84	al stone.</s><s>	Although most mosaics comprise three to six basic colours, a work of good quality

Se vogliamo visualizzare effettivamente la frase e non l'intera KWIC, selezioniamo "sentence" nel menù a tendina in altro a destra (Figura 34).

Figura 34 - Visualizzare una frase nella ricerca con containing

(11) `CQL`
`<doc/> containing [lemma="contract|agreement|proposal"]` = Trova tutti i documenti che contengono le parole *contract* o *agreement* o *proposal*.

(12) `CQL`
`<doc tld="eu"/> containing [lemma="contract|agreement|proposal"]` = Trova tutti i documenti con dominio europeo (.eu) che contengono le parole *contract* o *agreement* o *proposal*.

	Details	Left context
1	<input type="checkbox"/> yanisvaroufakis... ggle Maps</s><s>	
2	<input type="checkbox"/> europa.eu ilSender.</s><s>	
3	<input type="checkbox"/> ehealthnews.eu out 9am</s><s>	
4	<input type="checkbox"/> ipex.eu / Verizon</s><s>	
5	<input type="checkbox"/> get-ehealth.eu to enjoy.</s><s>	This project has r
6	<input type="checkbox"/> europa.eu e movie)</s><s>	Docum
7	<input type="checkbox"/> ipex.eu 10, 2014</s><s>	Home / A
8	<input type="checkbox"/> europa.eu mp.org.</s><s>	
9	<input type="checkbox"/> votewatch.eu :75-0626</s><s>	

CQL

- (13) `[tag="N.*"] []{1,3} [tag="V.*"] within <s/>` = Trova all'interno della stessa frase una sequenza composta da un nome seguito da un verbo con 1, 2 o 3 tokens nel mezzo (tra il nome e il verbo).

NB: se usassimo la sintassi `[tag="N.*"] []{1,3} [tag="V.*"]`, che abbiamo visto prima, SkE troverebbe anche casi in cui il nome e il verbo appartengono a frasi diverse, divisi da un punto o una virgola e simili (es. ...*cane. Andarono*...). Questi casi non vanno bene, perciò usare "within" ci è utile.

Notiamo dai seguenti esempi (13 e 14) che talvolta la stessa informazione può essere ottenuta con CQL differenti (confrontare i due risultati di ricerca riportati). La differenza risulta in cosa viene evidenziato da SkE come KWIC:

CQL

- (14) `[tag="J.*"]{3} containing [word="international"]` = Trova una sequenza di tre aggettivi, ma solo quelli che contengono la parola *international*.

Left context	KWIC	Right context
Convention and	other applicable international	instruments.</s>
res reports : The	international medical humanitarian	organization D
Development and	many other international	institutions to
Scotiabank is the	largest Canadian-based international	bank, with bran
need.</s><s>The	praiseworthy international juridical	framework of t
Woods System, a	post-war international regulatory	mechanism th

CQL

- (15) `[word="international"] within [tag="J.*"]{3}` = Trova la parola *international* all'interno di una sequenza di tre aggettivi.

Left context	KWIC	Right context
ention and	other applicable international	instruments.</s><s>A nu
ans Frontières reports : The	international medical humanitarian org	
velopment and	many other international	institutions to help them
the largest Canadian-based	international	bank, with branches in m
d.</s><s>The	praiseworthy international	juridical framework of the
Woods System, a	post-war international	regulatory mechanism th

4.1.2. Frequency: contare parole, tags e text types nelle concordanze

Quando le concordanze sono numerose, può essere utile usare lo strumento di frequenza, **Frequency** (Figura 35), in alto verso destra, per calcolare varie statistiche.

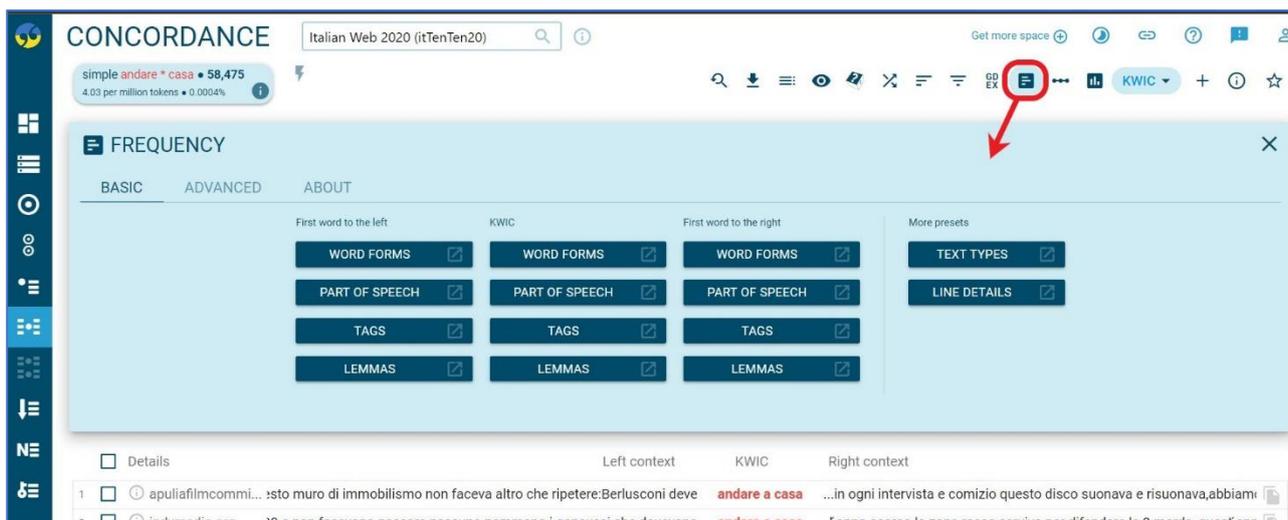


Figura 35 - Funzione Frequency: menù

Per esempio, per **contare il numero di occorrenze** di ogni forma della parola che abbiamo cercato dobbiamo cliccare su Frequency e poi, nella finestra che si apre e restando su Basic, Word Forms sotto la voce KWIC. → SkE ci mostrerà un elenco delle varie forme in cui occorre la nostra parola chiave nel corpus, messo in ordine di frequenza (Figura 36).

Il Menù (•••) sulla sinistra dell'elenco (vicino alla frequenza) ci permette di vedere solo gli esempi di quella forma o tutti gli altri esempi tranne questa forma (Figura 37).

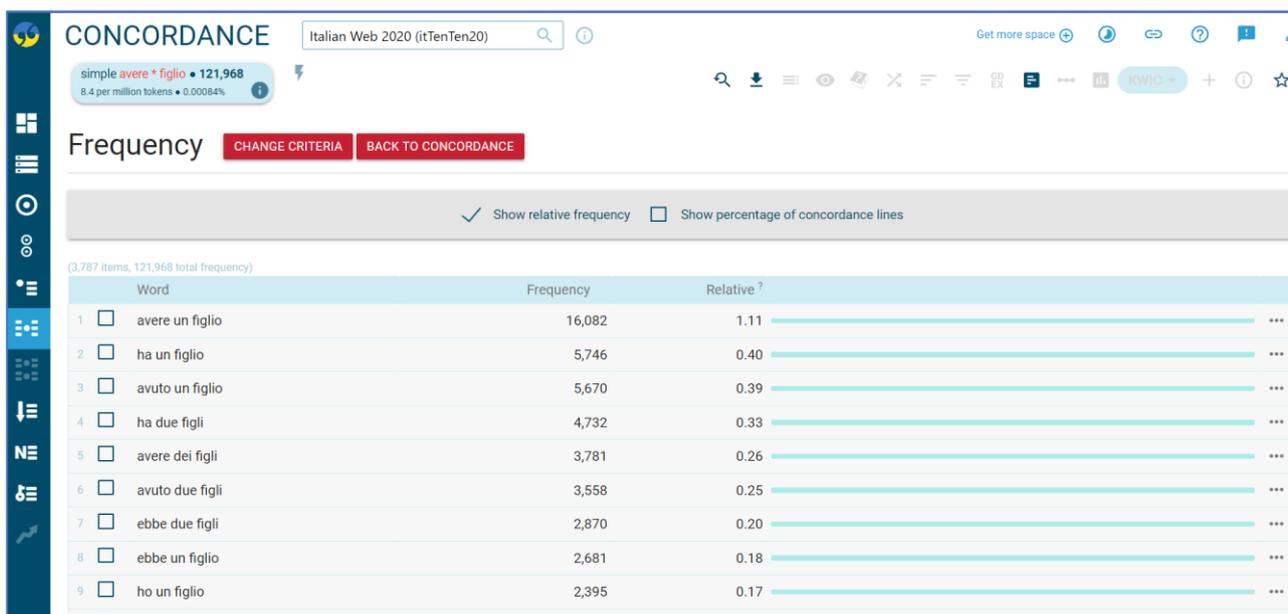


Figura 36 - Funzione Frequency > KWIC- Words Forms

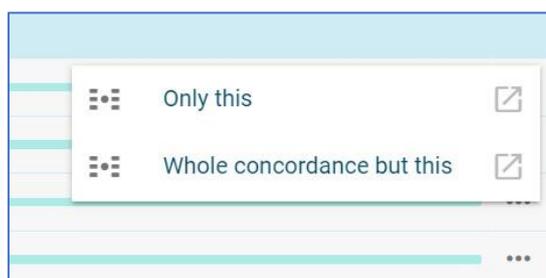


Figura 37 - Funzione Frequency: Menù vicino alla frequenza di una forma

Possiamo anche calcolare le parole che si trovano una posizione a sinistra o a destra della parola chiave (First word to the left/right) (Figura 35).

La versione Advanced di Frequency serve quando vogliamo **calcolare la frequenza di parole che si trovano ad una certa distanza dalla KWIC**. Qui possiamo scegliere l'attributo che preferiamo della KWIC (word, lemma, tag, ecc.) e aggiungere più posizioni nel calcolo usando il + in basso.

4.2. Word Sketch

Se vogliamo vedere le **collocazioni** di una parola, ovvero le tipiche combinazioni che quella data parola ha con altre parole, possiamo usare la funzione Word Sketch.

→ **collocazioni**: rapporti fra lessemi fondati su cooccorrenze regolari nel discorso.

Es. *ringraziare/caldamente, bandire/concorso, saluti/cordiali, rassegnare/dimissioni*.

ATTENZIONE: collocazioni ≠ **solidarietà semantiche**: rapporti di compatibilità semantica sull'asse sintagmatico: basate sull'occorrenza obbligatoria, o fortemente preferenziale, di un lessema con un altro, nel senso che la selezione dell'un termine è dipendente dall'altro, e la possibilità di essere usato in combinazione con altri lessemi è fortemente ridotta, se non assente.

Es. *miagolare/gatto, leccare/lingua, rafferma/pane*.

Quindi, mentre il rapporto di *solidarietà* è basato sulle proprietà e restrizioni semantiche previste dal sistema linguistico, il rapporto di *collocazione* riflette convenzioni e idiosincrasie tipiche dell'uso della singola parola.

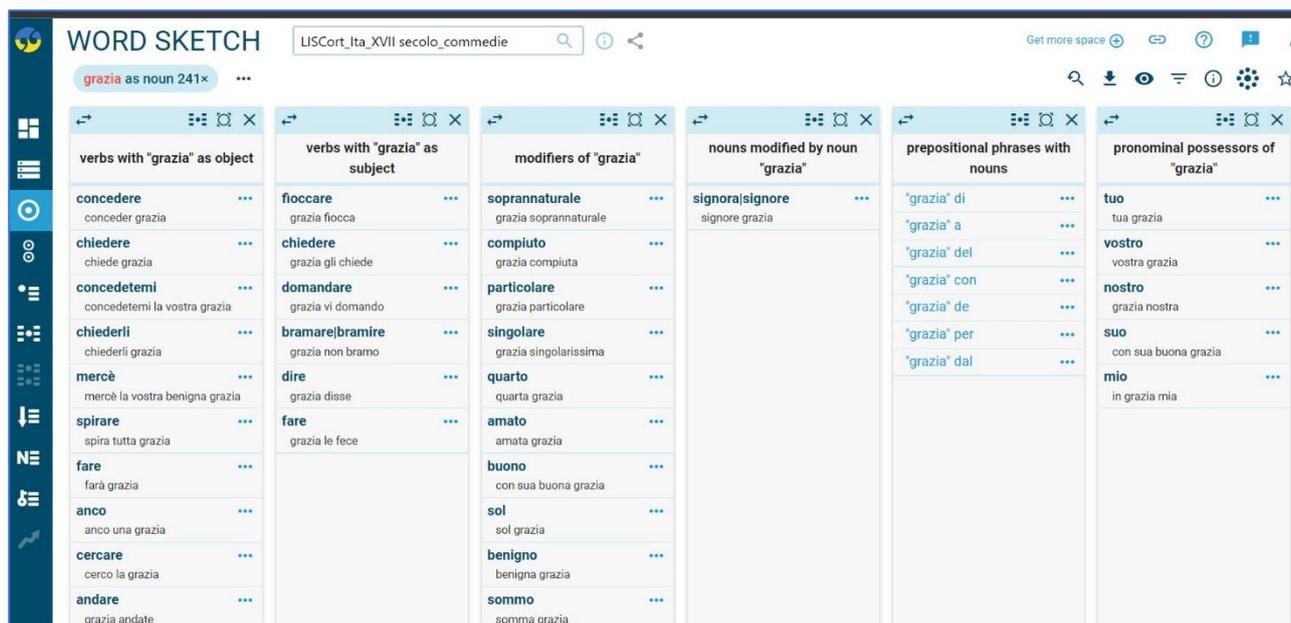
(Berruto & Cerruti 2017: 205-206)

Digitiamo il lemma (la forma della parola come da dizionario) che ci interessa e clicchiamo su "Go" (Figura 38): otterremo un elenco delle collocazioni di quella data parola all'interno del corpus che abbiamo selezionato (Figura 39).

Possiamo anche inserire un sintagma inserendo la parola chiave e la parola con la quale sarebbe in rapporto di collocazione (sempre in forma di lemma). Facendo questo e selezionando poi l'opzione Concordance (vedi più avanti), potremo ottenere tutte le concordanze che ci interessano in modo piuttosto semplice.



Figura 38 - Word Sketch



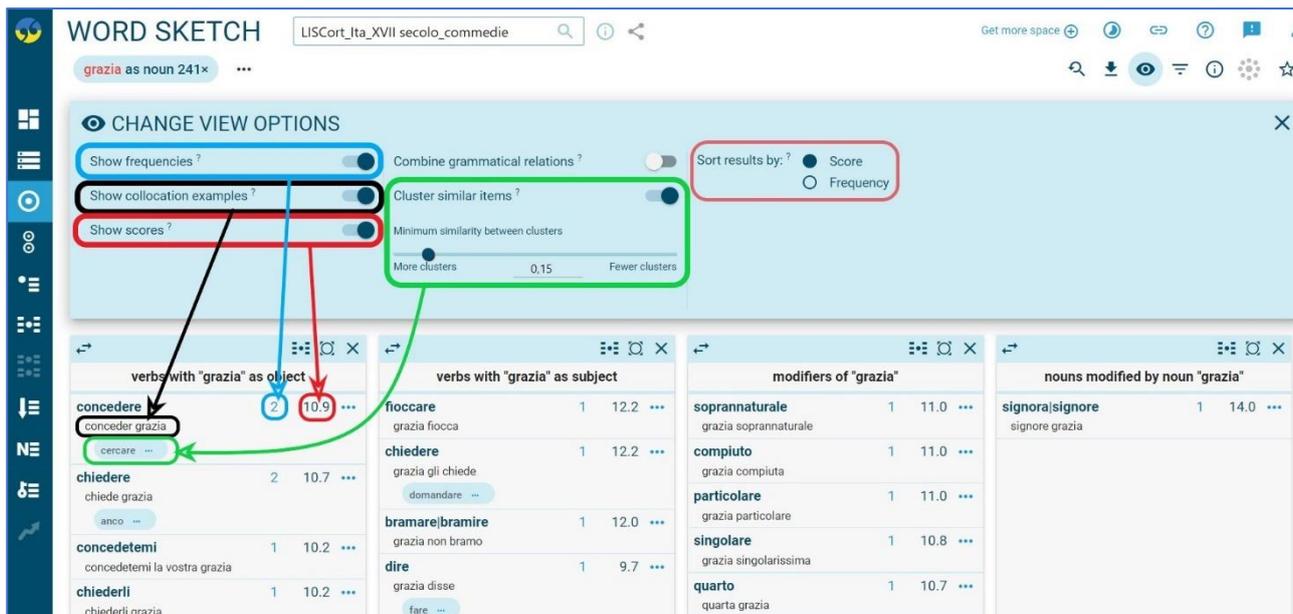
verbs with "grazia" as object	verbs with "grazia" as subject	modifiers of "grazia"	nouns modified by noun "grazia"	prepositional phrases with nouns	pronominal possessors of "grazia"
concedere conceder grazia	fioccare grazia fiocca	soprannaturale grazia soprannaturale	signora signore signore grazia	"grazia" di "grazia" a "grazia" del "grazia" con "grazia" de "grazia" per "grazia" dal	tuo tua grazia vostro vostra grazia nostro grazia nostra suo con sua buona grazia mio in grazia mia
chiedere chiede grazia	chiedere grazia gli chiede	compiuto grazia compiuta			
concedetemi concedetemi la vostra grazia	domandare grazia vi domando	particolare grazia particolare			
chiederli chiederli grazia	bramare bramire grazia non bramo	singolare grazia singolarissima			
mercè mercè la vostra benigna grazia	dire grazia disse	quarto quarta grazia			
spirare spira tutta grazia	fare grazia le fece	amato amata grazia			
fare farà grazia		buono con sua buona grazia			
anco anco una grazia		sol sol grazia			
cercare cerco la grazia		benigno benigna grazia			
andare grazia andate		sommo somma grazia			

Figura 39 - Esempio di ricerca con Word Sketch

Come possiamo vedere dalla Figura 39, le concordanze vengono mostrate suddivise per classe grammaticale, mettendole in ordine dalle più frequenti alle meno frequenti. Inoltre in automatico ci viene fornita, sotto ogni parola mostrata, la rappresentazione più frequente della collocazione, riportando un esempio preso direttamente dal corpus selezionato.

Da **Change view options** (Figura 40, Figura 41) possiamo modificare le informazioni relative alle concordanze come preferiamo. In particolare, possiamo scegliere:

- Show frequencies:** mostra quante volte è stata trovata la collocazione.
Show collocation examples: mostra la rappresentazione più comune della collocazione, ovvero la corrispondenza più comune.
Show scores (= punteggio di tipicità): indica quanto è forte la collocazione. Più alto è il punteggio, più forte è la collocazione. Un punteggio basso significa che le parole nella collocazione si combinano frequentemente anche con molte altre parole.
- Combine grammatical relations:** combina le collocazioni di tutte le relazioni grammaticali in un unico elenco.
Cluster similar items: organizza le collocazioni in gruppi in base alla somiglianza nel significato. Valori più bassi producono più gruppi contenenti meno collocazioni. Valori più alti producono meno gruppi contenenti più collocazioni. Possiamo regolare quanti items vedere dal dispositivo di scorrimento o digitando il valore che vogliamo. Attenzione: lo stesso valore può produrre risultati diversi con parole diverse.
- Sort results by Score:** visualizza le collocazioni più forti.
Sort results by Frequency: mostra le collocazioni più frequenti (ma non necessariamente le più tipiche).
 NB: questa opzione è legata a Show scores, nella prima sezione delle opzioni.



The screenshot shows the Word Sketch interface for the search term "grazia". The "CHANGE VIEW OPTIONS" panel is open, with several options highlighted by colored boxes and arrows:

- Show frequencies?**: Checked (blue box).
- Show collocation examples?**: Checked (blue box).
- Show scores?**: Checked (red box).
- Combine grammatical relations?**: Unchecked (grey box).
- Cluster similar items?**: Checked (green box).
- Sort results by?**: Score (red box).
- Minimum similarity between clusters**: Set to 0.15 (green box).

The search results are displayed in four columns:

- verbs with "grazia" as object**:

concedere	2	10.9
conceder grazia		
cercare		
chiedere	2	10.7
chiede grazia		
ancò		
concedetemi	1	10.2
concedetemi la vostra grazia		
chiederli	1	10.2
chiederli grazia		
- verbs with "grazia" as subject**:

fioccare	1	12.2
grazia fiocca		
chiedere	1	12.2
grazia gli chiede		
domandare		
bramare bramire	1	12.0
grazia non bramo		
dire	1	9.7
grazia disse		
fare		
- modifiers of "grazia"**:

soprannaturale	1	11.0
grazia soprannaturale		
compiuto	1	11.0
grazia compiuto		
particolare	1	11.0
grazia particolare		
singolare	1	10.8
grazia singolarissima		
quarto	1	10.7
quarta grazia		
- nouns modified by noun "grazia"**:

signora signore	1	14.0
signore grazia		

Figura 40 - Word Sketch-Change view options

→ SCORE VS. FREQUENCY

Un **punteggio (score) molto alto** della collocazione significa che c'è poca concorrenza da parte di altre collocazioni. Il nodo (la parola ricercata, la parola chiave) non si combina spesso con altri collocati. In altre parole, i concorrenti non sono frequenti né per questi motivi né per la loro combinazione:

- il numero di collocazioni diverse in competizione è molto basso.
- il numero di collocazioni concorrenti può essere elevato, ma la frequenza di ciascuna di esse è bassa e quindi il totale rimane basso.

Un **punteggio (score) molto basso** significa che c'è un'estrema concorrenza da parte di altri collocati per uno di questi motivi o per la loro combinazione:

- il numero di collocazioni diverse in competizione è molto alto.
- il numero di collocazioni diverse in competizione può essere piccolo, ma la frequenza di ciascuna di esse è estremamente elevata, il che produce una forte concorrenza per la collocazione in questione.

Di conseguenza, è abbastanza comune che le parole più comuni e frequenti non ricevano quasi mai punteggi elevati come collocati, perché sono usate così spesso in combinazione con così tante altre parole che c'è molta competizione. Possono esserci delle eccezioni se la collocazione è così estremamente frequente da battere tutti i suoi concorrenti.

(<https://www.sketchengine.eu/guide/word-sketch-collocations-and-word-combinations/#toggle-id-6>)

Word	Grammatical relation	Score ↓
1 buono con sua buona grazia	modifiers of "grazia"	8.2 ...
2 vostra vostra grazia	pronominal possessors of "grazia"	8.0 ...
3 hà grazia , hà	"grazia" and/or ...	7.9 ...
4 concedere conceder grazia	verbs with "grazia" as object	7.9 ...
5 chiedere verb chiede grazia	verbs with "grazia" as object	7.9 ...
6 signora signore noun Di grazia , signore	"grazia" and/or ...	7.9 ...
7 tuo tua grazia	pronominal possessors of "grazia"	7.7 ...
8 fare verb farà grazia	verbs with "grazia" as object	7.7 ...
9 porre grazia di non porvi	"grazia" di	7.5 ...
10 suo con sua buona grazia	pronominal possessors of "grazia"	7.1 ...
11 stà grazia , stà	"grazia" and/or ...	7.1 ...
12 sopranaturale	modifiers of "grazia"	7.1 ...

Figura 41 - Word Sketch: risultato opzione Combine grammatical relation su Change view options

In cima ad ogni colonna dei risultati, abbiamo una barra di selezione rapida da cui possiamo scegliere di vedere tutte le collocazioni della colonna contestualizzate (☰), di visualizzare solo quella colonna e nascondere le altre (👁), oppure di nascondere solo quella colonna (✕).

Dal **menù** vicino ad ogni collocazione (⋮, Figura 42) da cui possiamo scegliere di:

- usare la funzione Concordance per vedere le occorrenze di quella data collocazione contestualizzate (si noti che SkE individua la collocazione anche se le due parole non compaiono vicine) e quindi i suoi esempi d'uso tratti dal corpus.

CONCORDANCE LIScort_Ita_XVII secolo_commedie

CQL **grazia + concedere** • 2
9.41 per million tokens • 0.00094%

☰ Details Left context KWIC Right context

1	☰	doc#4	ito di conforto, e, ripigliato lo smarrito spirito, pregai Porzia mia di volermi conceder grazia di tollerare con pazienza la mia partita, la lontananza e l'assenza mia per tre anni so
2	☰	doc#4	cizia la quale era tra mio padre e voi, io vi prego affettuosamente a concedermi una grazia , e so che non me la negherete.</s><s>demetrio.</s><s>Come, Giulia, domanda pur

- Usare la funzione Word Sketch per vedere le collocazioni in un sintagma.

WORD SKETCH LIScort_Ita_XVII secolo_commedie

concedere grazia 2x ...

grazia noun > concedere verb ✕

↔ ☰ ✕

prepositional phrases with nouns

"concedere grazia" di	50.0% ...
-----------------------	-----------

- Usare la funzione Thesaurus per vedere tutti i sinonimi della collocazione.

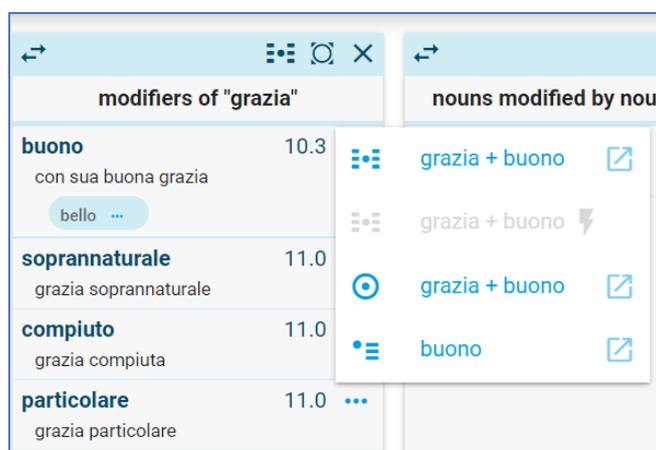


THESAURUS LISCort_Ita_XVII secolo_commedie

concedere as verb 36x

Lempos	Frequency ?
1	insegnare 23 ...
2	parare 28 ...
3	facci 14 ...
4	negare 32 ...
5	rallegrare 24 ...
6	sforzare 17 ...
7	palesare 16 ...
8	fidare 17 ...
9	dispiacere 24 ...
10	domandare 20 ...

Rows per page: 10 1-10 of 131 1 / 14



modifiers of "grazia"	nouns modified by noun
buono 10.3 con sua buona grazia bello ...	grazia + buono grazia + buono ⚡ grazia + buono buono
soprannaturale 11.0 grazia soprannaturale	
compiuto 11.0 grazia compiuta	
particolare 11.0 ... grazia particolare	

Figura 42 - Word Sketch-Tendina del Menù per ogni collocazione

4.3. Thesaurus

Se vogliamo vedere i **sinonimi** di una parola, possiamo usare la funzione Thesaurus.

Digitiamo il lemma (la forma della parola come da dizionario) che ci interessa e clicchiamo su "Go" (Figura 43): otterremo un elenco di tutti i sinonimi di quella data parola all'interno del corpus che abbiamo selezionato (Figura 44).



THESAURUS LISCort_Ita_XVII secolo_commedie

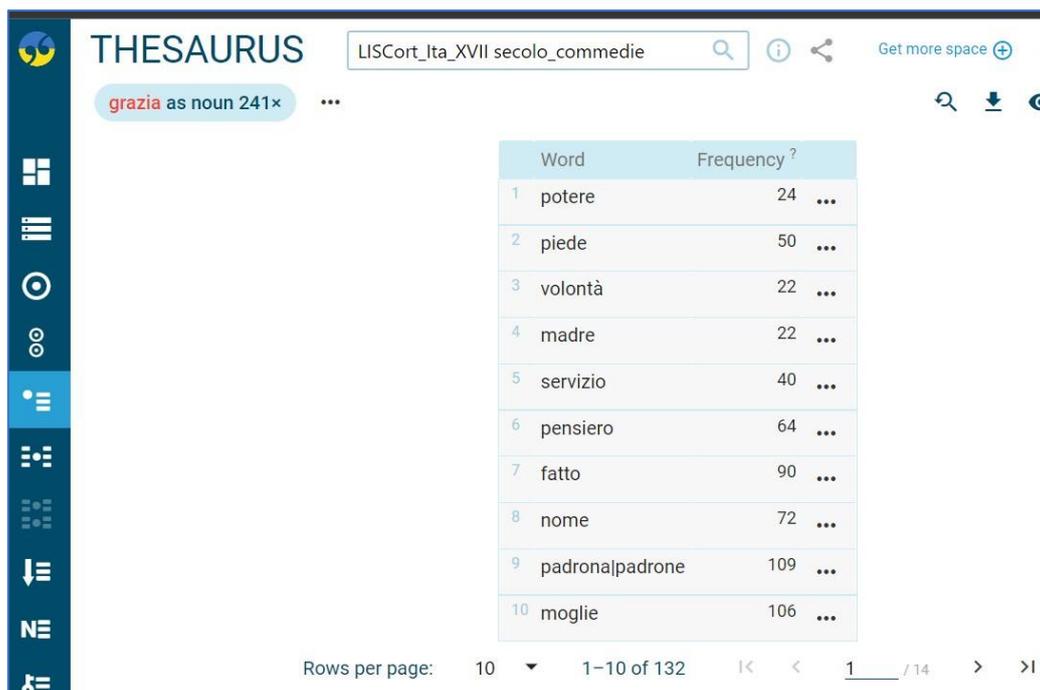
BASIC ADVANCED ABOUT

Search ?

lemma

GO

Figura 43 - Thesaurus



THESAURUS LIScort_Ita_XVII secolo_commedie

grazia as noun 241x

	Word	Frequency ?
1	potere	24 ...
2	piede	50 ...
3	volontà	22 ...
4	madre	22 ...
5	servizio	40 ...
6	pensiero	64 ...
7	fatto	90 ...
8	nome	72 ...
9	padrona padrone	109 ...
10	moglie	106 ...

Rows per page: 10 1-10 of 132 1 / 14

Figura 44 - Esempio di ricerca con Thesaurus

Da **Change view options** (Figura 45) possiamo scegliere:

- Show line numbers: numera l'elenco fornito.
Show counts: mostra la frequenza di ogni parola.
Show similarity scores: mostra il punteggio di somiglianza per ogni sinonimo.
- Single column: passa da una lunga colonna che richiede lo scorrimento verso l'alto/il basso al riempimento dello schermo con diverse colonne brevi
Cluster similar items: posiziona le parole in gruppi in base alla somiglianza nel significato: valori più bassi producono più gruppi contenenti meno collocazioni; valori più alti producono meno gruppi contenenti più collocazioni. Possiamo regolare i raggruppamenti attraverso il dispositivo di scorrimento o digitando il valore desiderato.



CHANGE VIEW OPTIONS

Show line numbers ? Single column ?

Show counts ? Cluster similar items ?

Show similarity score ?

	Lempos	Frequency ?	Similarity ? ↓
1	potere	24	0.448 ...
2	piede	50	0.439 ...
3	volontà	22	0.435 ...
4	madre	22	0.428 ...
5	servizio	40	0.421 ...

Figura 45 - Thesaurus-Change view options

Dal **menù** ... vicino ad ogni sinonimo (Figura 46) possiamo scegliere di:

- usare la funzione Word Sketch Difference per comparare il sinonimo con la parola chiave.

WORD SKETCH DIFFERENCE LIScort_Ita_XVII secolo_commedie

grazia 241x | potere 24x

verbs with "grazia/potere" as object			verbs with "grazia/potere" as subject			modifiers of "grazia/potere"			"grazia/potere" di		
concedere	2	0	fioccare	1	0	soprannaturale	1	0	porre	2	0
chiedere	2	0	chiedere	1	0	compiuto	1	0	promettermi	1	0
concedetemi	1	0	domandare	1	0	particolare	1	0	tollerare	1	0
chiederli	1	0	bramare bramire	1	0	singolare	1	0	ritrovare	1	0
mercè	1	0	dire	1	0	quarto	1	0	lasciare	1	0
spirare	1	0	fare	1	0	amato	1	0	vedere	1	0
anco	1	0	mancare	0	1	buono	4	0	dare	1	0
cercare	1	0				sol	1	0			
andare	1	0				benigno	1	0			
fare	14	1				sommo	1	0			
eccedere	0	1				alcuno	0	1			
registrare	0	1				immenso	0	1			

b. usare la funzione Thesaurus per vedere i sinonimi di quel sinonimo.

THESAURUS LIScort_Ita_XVII secolo_commedie

potere as noun 24x

Lempos	Frequency ?	Similarity ? ↓
1	piede	50 0.999
2	danno	20 0.749
3	contento	55 0.712
4	fratello	82 0.712
5	consorte	34 0.692
6	figlia	19 0.664
7	nome	72 0.657
8	salute	23 0.653
9	madre	22 0.650
10	petto	29 0.629

Rows per page: 10 1-10 of 117

c. usare la funzione Concordance per vedere il sinonimo contestualizzato nel corpus e quindi i suoi esempi d'uso.

CONCORDANCE LIScort_Ita_XVII secolo_commedie

COQ [lempos='potere-n'] • 24
112.93 per million tokens • 0.011%

Left context KWIC Right context

1 doc#0 ./s><s>Vogliamo così perché possiamo, possiamo perché vogliamo, e il volere e il potere è al piacere congiunto; si che vogliamo perché ne diletta e piace.</s></s>alberto.</s>

2 doc#0 o schietto, ma del signor conte Alberto, e noi vi doniamo la contea, la quale ora è in potere del Gran Turco, ma gle la piglieremo bene fra poco.</s></s>alberto.</s></s>Pur tropp

3 doc#0 :>Fermatevi signora, non s'affatichi correndo di tór la vita a chi la vita dar le vuole in potere , se d'ascoltar due parole sole si farà degno chi, per sentenza sua, debbe morire.</s>

4 doc#1 e vostre braccia.</s></s>scappino.</s></s>E le mie braccia vi serviranno a tutto suo potere , e così ogn'altra mia cosa che vi possa dar gusto.</s></s>Orsù, non vi è tempo da p

5 doc#1 n prometto se non quello che voglio fare; ma alle volte il volere è oppresso del non potere .</s></s>lavinia.</s></s>Il mio amore è una pianta, quale non è abbarbicata nel terre

d. usare la funzione Word Sketch per vedere le collocazioni del sinonimo.

WORD SKETCH LISort_Ita_XVII secolo_commedie

potere as noun 24x

verbs with "potere" as object			verbs with "potere" as subject		
eccedere	1	13.0	mancare	1	12.4
eccedere il mio potere			potere ha mancato		
registrare	1	13.0			
registrare il mio potere					
fare	1	6.1			
fatto ogni vostro potere					

Lempos	Frequency ?	Similarity ? ↓	
1	potere	24	0.448
2	piede	50	0.439
3	volontà	22	0.435
4	madre	22	0.428
5	servizio	40	0.421
6	pensiero	64	0.419
7	fatto	90	0.399

-  Word Sketch Difference 
-  Thesaurus 
-  Concordance 
-  Concordance with macro
-  Word Sketch 

Figura 46 - Thesaurus-Tendina del Menù per ogni sinonimo

→ **Curiosità: Come funziona Thesaurus?** Una volta inserita la parola di cui vogliamo ottenere i sinonimi, SkE opera una ricerca Word Sketch per verificarne le collocazioni, dopodiché opererà ricerche Word Sketch su tutte le parole della stessa classe grammaticale che appaiono in rapporto di collocazione con la parola chiave. Infine compara le collocazioni della parola chiave con quelle delle sue collocazioni (ovviamente tutto questo sulla base del corpus che noi abbiamo scelto di usare) e ci mostra, in ordine di punteggio (score), le parole che contengono la più alta proporzione di collocazioni identiche, che sono anche quelle più simili nel significato.

Questo criterio è basato sulla teoria della **semantica distribuzionale**, secondo la quale le parole che occorrono in contesti simili sono anche simili nel loro significato.

ATTENZIONE:

- Siccome Thesaurus funziona così, la lista che ci fornisce è (potenzialmente) infinita; per questo motivo la sua affidabilità e quindi utilità è limitata alla parte più alta di questa lista, dove le parole sono più simili.
- Talvolta può succedere che Thesaurus ci mostri sinonimi che non dovrebbero essere sulla lista: questo può accadere perché le parole più comuni sono usate in così tanti contesti differenti che le loro collocazioni possono essere molto simili anche se il loro significato non lo è. Ciò significa che i risultati per le parole meno comuni tendono ad essere più precisi.