

# Sketch Engine

## Bevezetés

A számítógéppel olvasható és elemezhető korpuszok a nyelvtudomány több területén számos új kutatási területet és lehetőséget nyitottak meg. Közülük például a korpuszalapú fordítástudományi kutatások megjelenése Magyarországon az 1990-es évekre tehető (Klaudy 2005), elterjedésükről azonban a 2000-es évek elejétől beszélhetünk, ami a technológia fejlődésének köszönhető (Robin & Seidl-Péché, 2020). Az ilyen típusú vizsgálatok jelentősége abban rejlik, hogy a szövegmennyiség méretének köszönhetően statisztikailag is megalapozott következtetések vonhatók le a nyelvhasználatra vonatkozóan. Bár a kutatóknak több ingyenes, vagy részben ingyenes eszköz áll rendelkezésükre a korpuszok elemzésének elvégzésére, ezeket gyakran csak bizonyos korlátozásokkal használhatjuk – pl. az annotálás vagy a lemmatizálás hiánya miatt (Robin et al., 2016; Seidl-Péché & Robin, 2019).

Jelen ismertető a Sketch Engine (<https://www.sketchengine.eu/>) nevű eszközt mutatja be, amely lehetővé teszi a nyelvészeti korpuszkutatások során jellemzően használt elemzések elvégzését.

## Sketch Engine

A Sketch Engine korpuszelemző eszköz, melynek segítségével különféle kereséseket és műveletek végezhetünk saját és az adatbázisban rendelkezésre bocsátott szöveggyűjteményeken. A kínált lehetőségek – amelyek között a lemmatizálás automatikus elvégzése is szerepel – hasznosak lehetnek nyelvészek, fordítók, oktatók és diákok számára is, hiszen a korpuszok elemzése során számos olyan információhoz juthatunk, amelyek a szótárakból és glosszáriumokból hiányoznak. A használathoz előfizetés szükséges, melynek díja az előfizetés típusától függ: a felsőoktatásban hallgatók és dolgozó kutatók – az ismertető megírásának pillanatában – egyénileg havi bruttó 10,66 euróért juthatnak licenchez, azonban lehetőség van éves fizetési ütemezésre, illetve csoportos előfizetésre is, melynek köszönhetően az egy főre eső havidíj mérséklődik. Amennyiben nem hallgatóként vagy kutatóként, hanem egyénileg szeretnénk előfizetni, a havi bruttó alapdíj 19,99 euró, azonban ez esetben is lehetőségünk van a fizetési gyakoriságtól függően alacsonyabb összegért hozzájutni a szolgáltatáshoz. Az említett díjak a rendszer használatán túl magukban foglalnak egymillió szónyi tárhelyet arra az esetre, ha saját korpuszokkal is dolgozni szeretnénk (amennyiben ettől nagyobb méretű a korpuszunk, úgy további tárhely vásárlása szükséges és lehetséges). Mielőtt azonban fizetnénk a használatért, a Sketch Engine összes funkcióját ingyen kipróbálhatjuk 30 napon át.

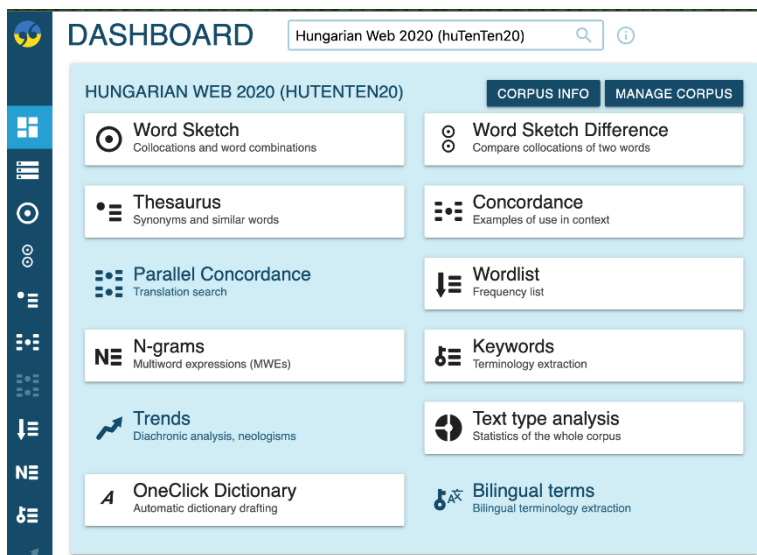
## A Sketch Engine segítségével végezhető elemzések

A regisztrációhoz csupán néhány alapvető adat (név, e-mail-cím, munkahely, a használat célja, számlázási adatok) megadása szükséges, majd néhány percen belül megérkezik postafiókunkba a rendszer használatához szükséges jelszó. Bejelentkezés után a korpusz kiválasztása (*Select Corpus*) nézet jelenik meg. Itt a nyelv megadása után lehetőségünk van az adott nyelven elérhető bármely korpuszt kiválasztani. A rendszerben számos nyelv és az egyes nyelveken gyakran több korpusz is hozzáférhető. Az elérhető nyelvek, az azokhoz tartozó korpuszok és azok jellemzői (például annotált, lemmatizált) a rendszer súgójában található táblázatban tekinthetők meg (*Languages in Sketch Engine*<sup>1</sup>). A magyar nyelvű szövegek tanulmányozását jelenleg 21 különböző korpusz segítségével végezhetjük, melyek között megtalálhatók általános (így internetről gyűjtött) és szűkebb területről származó (például európai uniós) gyűjtemények egyaránt.

A nyelv kiválasztása után a *Dashboard* felületre érkezünk, amely tulajdonképpen egyfajta vezérlőpult – innen érhetők el az egyes műveletek, amelyek a kiválasztott korpuszon elvégezhetők.

### 1. ábra

A Sketch Engine kezdőlapja



### Szóvázat – a szavak nyelvtani és kollokációs viselkedése (Word Sketch)

A Word Sketch funkció a szavak kollokációjának vizsgálatára alkalmas, használatához csupán a keresett lemmát kell megadnunk, majd rövid időn belül a rendszer kilistázza az adott szó környezetével kapcsolatos adatokat. Ennek segítségével meg-

<sup>1</sup> <https://www.sketchengine.eu/corpora-and-languages/>

tudhatjuk például, hogy milyen igékkel, névszókkal, névelőkkel, névutókkal fordul elő leggyakrabban az adott kifejezés, de a megjelenő összesítés egyes listáit ki is bonthatjuk, így még több információhoz juthatunk az éppen vizsgált jelenséggel kapcsolatban. A keresést optimalizálhatjuk úgy is, ha a lemma megadása után az *advanced* fülön pontosítjuk a beállításokat (például szófaj megadása). A 2. ábrán az *egészség* szóra keresve megjelenő összefoglaló látható.

## 2. ábra

A szóvázat (Word Sketch) művelet eredménye az *egészség* szóra keresve

The screenshot shows the Word Sketch interface for the word "egészség" (health) in Hungarian. The search results are organized into four main columns:

- modifiers of "egészség"**: Includes terms like "mentális egészség", "lelki egészség", "gyakorolt egészségre gyakorolt", "ártalmas az egészségre ártalmas", "káros az egészségre káros", "testi-lelki egészség", "testi egészség", "csontok egészségének", "bőr egészségének", "emberi egészségre", "gazda növekednek, a gazda szervezet egészségét is befolyásolhatják", and "fogak egészségének".
- verbs with "egészség" as object**: Includes "veszélyeztet egészségét veszélyezteli", "örvend jó egészségre örvend", "károsít az egészséget", "megőrzi megőrizzik egészségüket", "kíván jó egészséget kívánok", and "vigyáz vigyázzon az egészségére".
- verbs with "egészség" as subject**: Includes "a-z Összes cikk Egészség A-Z témában: Az", "cenzuráz Cenzurázott egészség", "helyreáll az egészség helyreáll", "romlik az egészség romlik", "felértékelődik felértékelődik az egészség", "lefog rd egészség lefog", and "összefügg egészség szorosan összefügg".
- "egészség" and/or ...**: Includes "biztonság egészség és biztonság", "szépség Szépség és egészség", "környezet egészségre és a környezetre", "gazda növekednek, a gazda szervezet egészségét is befolyásolhatják", "jólet egészség és jólet", "megőrzés egészségének megőrzésében", "közérzet egészség és a jó közérzet", "boldogság egészségét és boldogságot", "épség épséget vagy egészséget", "élet egészségének és életének", "érdek egészségének és életének megővése érdekében", and "jólet egészségre és jólete".

## Tezaurusz (Thesaurus)

A Tezaurusz szógyűjtemény a megadott lemmával egy szemantikai mezőbe tartozó kifejezéseket gyűjt össze. Ahhoz, hogy az összeállított lista megbízható legyen, nagy méretű korpuszokra van szükség. A fejlesztői ajánlások szerint a minimális méret kb. százezer szó, a megbízható működéshez azonban milliárdos nagyságrendű szószámot javasolnak – emiatt célszerű lehet e funkció esetében a beépített korpuszokkal dolgozni, ha a saját szöveggyűjteményünk mérete ettől elmarad.

## Párhuzamos konkordancia (Parallel Concordance)

A párhuzamos konkordancia vizsgálatához párhuzamos korpuszokra van szükség. A művelet segítségével arra kaphatunk választ, hogy bizonyos szavakat vagy kifejezéseket hogyan fordítottak le más nyelvre. Az 3. ábrán látható, hogy a német *Parlament* szót tartalmazó szövegrészeket hogyan fordították magyarra.

### 3. ábra

A Parlament szó fordításainak listája a párhuzamos konkordancia segítségével

PARALLEL CONCORDANCE EUR-Lex 22016 parallel - German

single parliament • 246,147  
add 60 per million items • 200%

12008E215 <-> (1) Sieht ein nach Teil V Kapitel 2 des Vertrags über die Europäische Union erlassener Beschluss die Aussetzung, Einschränkung oder vollständige Einstellung der Wirtschafts- und Finanzbeziehungen zu einem oder mehreren Drittländern vor, so erlässt der Rat die erforderlichen Maßnahmen mit qualifizierter Mehrheit auf gemeinsamen Vorschlag des Hohen Vertreters der Union für Außen- und Sicherheitspolitik und der Kommission. <-> <-> Er unterrichtet hierüber das Europäische Parlament. <->

12008E263 <-> Der Gerichtshof der Europäischen Union überwacht die Rechtmäßigkeit der Gesetzgebungsakte sowie der Handlungen des Rates, der Kommission und der Europäischen Zentralbank, soweit es sich nicht um Empfehlungen oder Stellungnahmen handelt, und der Handlungen des Europäischen Parlaments und des Europäischen Rates mit Rechtswirkung gegenüber Dritten. <-> Er überwacht ebenfalls die Rechtmäßigkeit der Handlungen der Einrichtungen oder sonstigen Stellen der Union mit Rechtswirkung gegenüber Dritten. <->

12008E263 <-> Zu diesem Zweck ist der Gerichtshof der Europäischen Union für Klagen zuständig, die ein Mitgliedstaat, das Europäische Parlament, der Rat oder die Kommission wegen Unzuständigkeit, Verletzung wesentlicher Formvorschriften, Verletzung der Verträge oder einer bei seiner Durchführung anzuwendenden Rechtsnorm oder wegen Ermessensmissbrauchs erhebt. <->

12008E048 <-> Das Europäische Parlament und der Rat beschließen gemäß dem ordentlichen Gesetzgebungsverfahren die auf dem Gebiet der sozialen Sicherheit für die Herstellung der Freizügigkeit der Arbeitnehmer notwendigen Maßnahmen; zu diesem Zweck führen sie insbesondere ein System ein, das zu- und abwandernden Arbeitnehmer und Selbstständigen sowie deren anspruchsberechtigten Angehörigen Folgendes sichert: <->

12008E043 <-> (2) Das Europäische Parlament und der Rat legen gemäß dem ordentlichen Gesetzgebungsverfahren und nach Anhörung des Wirtschafts- und Sozialausschusses die gemeinsame Organisation der Agrarmärkte nach Artikel 40 Absatz 1 sowie die anderen Bestimmungen fest, die für die Verwirklichung der Ziele der gemeinsamen Agrar- und Fischereipolitik notwendig sind. <->

EUR-Lex 22016 parallel - Hungarian

(1) Ha az Európai Unióról szóló szerződés V. címének 2. fejezete szerint elfogadott valamely határozat az egy vagy több harmadik országgal fennálló gazdasági és pénzügyi kapcsolatok részbeni vagy teljes megszakítását vagy korlátozását írja elő, a Tanács az Unió külügyi és biztonságpolitikai főképviselője és a Bizottság által előterjesztett közös javaslat alapján, mindkét fél közeggegyet elfogadja a szükséges intézkedéseket. <-> Erről tájékoztatja az Európai Parlamentet. <->

Az Európai Unió Bírósága megvizsgálja a jogalkotási aktusok jogszerűségét, valamint a Tanács, a Bizottság és az Európai Központi Bank jogi aktusait, kivéve az ajánlásokat és a véleményeket, továbbá az Európai Parlament és az Európai Tanács harmadik személyekre joghatással járó aktusait. <-> Az Európai Unió Bírósága megvizsgálja továbbá az Unió szervei vagy hivatalai által elfogadott, harmadik személyekre joghatással járó jogi aktusok jogszerűségét. <->

E célból az Európai Unió Bírósága hatáskörrel rendelkezik az olyan keresetek tekintetében, amelyek valamely tagállam, az Európai Parlament, a Tanács vagy a Bizottság nyújt be hatáskör hiánya, lényeges eljárási szabályok megsértése, a Szerződések vagy az alkalmazásra vonatkozó bármely jogi rendelkezés megsértése vagy hatáskörrel való visszaélés miatt. <->

Az Európai Parlament és a Tanács rendes jogalkotási eljárás keretében a szociális biztonság területén elfogadja azokat az intézkedéseket, amelyek a munkavállalók szabad mozgásának biztosításához szükségesek; e célból olyan eszközrendszert hoz létre, amely a migráns munkavállalók és önként vállalkozók, valamint az azok jogán jogosultak számára biztosítja. <->

(2) A 40. cikk (1) bekezdésében előírt közös piacszervezet, továbbá a közös agrárpolitika és a közös halászati politika célkitűzéseinek megvalósításához szükséges egyéb rendelkezéseket a Gazdasági és Szociális Bizottsággal folytatott konzultációt követően rendes jogalkotási eljárás keretében az Európai Parlament és a Tanács állapítja meg. <->

### N-gramok (N-grams)

Az N-gramok, azaz a többszavas kifejezések funkció segítségével a nyelvben előforduló többszavas kifejezéseket gyakoriságát vizsgáljuk. A keresés elindításakor lehetőségünk van beállítani, hogy hány szóból álló kifejezésekre van szükségünk, de természetesen további keresési feltételeket is megadhatunk (pl. a keresési eredmény tartalmazzon bizonyos szavakat). A 4. ábrán olyan, három vagy négy tagból álló kifejezések láthatók, amelyek tartalmazzák a *néz* igét.

### 4. ábra

Három vagy négy tagból álló kifejezések, amelyek tartalmazzák a *néz* igét

N-GRAMS Hungarian Web 2020 (huTenTen20)

3-4-grams, word (Items: 3,635, total frequency: 106,738)

N-gram	Frequency ?	N-gram	Frequency ?	N-gram	Frequency ?
1 úgy néz ki	11,470 ...	11 Így néz ki	821 ...	21 Hogyan néz ki	515 ...
2 néz ki a	3,875 ...	12 és úgy néz	608 ...	22 nagyon jól néz ki	512 ...
3 jól néz ki	3,317 ...	13 Hogy néz ki	588 ...	23 de úgy néz ki	511 ...
4 Úgy néz ki	2,511 ...	14 jobban néz ki	549 ...	24 néz ki és	496 ...
5 Így néz ki	2,154 ...	15 és úgy néz ki	533 ...	25 hogy néz ki a	478 ...
6 hogy néz ki	1,753 ...	16 nagyon úgy néz	526 ...	26 nem néz ki	477 ...
7 is néz ki	1,308 ...	17 nagyon úgy néz ki	522 ...	27 nem úgy néz	476 ...
8 néz ki egy	1,152 ...	18 de úgy néz	520 ...	28 úgy néz ki a	461 ...
9 hogyan néz ki	1,093 ...	19 néz ki mint	518 ...	29 nem úgy néz ki	456 ...
10 néz ki az	1,079 ...	20 nagyon jól néz	518 ...	30 néz ki ez	415 ...

You are only allowed to access 1,000 items. [Get more](#)

## Trendek (Trends)

A trendek opció a diakrón vizsgálatok során jelenthet segítséget, a használat gyakoriságának változásairól adnak felvilágosítást. A funkció a működéséből adódóan csak időbélyeggel ellátott korpuszokon használható. Lehetőség van egy-egy konkrét szó vagy kifejezés vizsgálatára, alapbeállítás mellett pedig azokat az elemeket listázza a rendszer, amelyek a vizsgált időszakban a legnagyobb – akár pozitív, akár negatív – változáson mentek keresztül.

## Szóvázatok eltérései (Word Sketch Difference)

A Word Sketch Difference, azaz a szóvázatok eltérései funkció akkor segít, ha két lemma kollokációját szeretnénk összehasonlítani, ezáltal jobban meg szeretnénk ismerni a kettő közötti különbségeket. A keresési eredmények nagyjából a Word Sketch lehetőségnél megjelenőkkel azonosak azzal az eltéréssel, hogy az egyes szavak mellett két számérték jelenik meg. Az első számérték az adott találat és az első keresett szó, a második számérték az adott találat és a második keresett szó együttes előfordulásának gyakoriságát mutatja. Az 5. ábrán a *megy* és a *jár* igék eltérő kollokációi láthatók. Az ábrából az is kiderül, hogy a *megy* szó jelentősen gyakrabban fordul elő a *játék* szó mellett (*megy a játék*), míg a *jár* és a *köszönet* együttes előfordulása sokkal jellemzőbb, mint a *megy* és a *köszöneté*.

### 5. ábra

A *megy* és a *jár* ige kollokációinak összehasonlítása

The screenshot shows the 'WORD SKETCH DIFFERENCE' tool interface. At the top, it displays 'Hungarian Web 2020 (huTenTen20)' and search filters for 'megy' (4,454,495x) and 'jár' (2,982,335x). Below are three panels showing concordance data for 'megy/jár' and/or ...', 'objects of "megy/jár"', and 'subjects of "megy/jár"'. Each panel contains a table with words and their respective counts.

"megy/jár" and/or ...			objects of "megy/jár"			subjects of "megy/jár"		
jön	5,323	219	rovás	14,727	0	játék	13,767	383
megy	4,398	283	csőd	14,690	0	változás	3,845	189
marad	1,264	82	nyugdíj	17,123	66	dolog	21,723	1,782
csinál	839	80	férj	31,784	330	min	3,925	373
megnéz	781	162	irány	16,036	624	munka	3,725	455
van	5,052	2,445	változás	16,933	2,114	s	6,989	3,006
lát	738	645	él	5,619	19,291	idő	3,440	9,064
szeret	527	704	iskola	11,937	43,137	pont	2,266	8,759
kel	306	316	út	16,238	63,797	hír	333	5,611
beszél	433	779	hely	7,380	31,015	büntetés	39	3,110
olvas	176	350	siker	718	47,623	kedvezmény	0	3,477
jár	283	2,050	következmény	0	35,545	köszönet	0	4,117

## Konkordancia (Concordance)

A konkordanciavizsgálat segítségével szavak vagy kifejezések környezetét vizsgálhatjuk. A kereséshez meg kell adnunk egy lemmát és a rendszer kilistázza a megadott szó

(vagy kifejezés) konkrét előfordulásait a szövegben a környezettel együtt. Az 6. ábra a *kérdez* lemma előfordulásait, illetve azok forrásait mutatja be.

## 6. ábra

A *kérdez* lemma előfordulásai

CONCORDANCE Hungarian Web 2020 (huTenTen20)

simple **kérdez** • 1,063,333  
168.25 per million tokens • 0.017%

Left context KWIC Right context

- 1 worldshots.hu emzeti Községi Egyletem Hadtudományi és Honvédtisztviselői Kar kibeháborús szakértőjét arról **kérdeztük** , hogy fenyeget-e az internetes háború Magyarországot?
- 2 worldshots.hu konok **kérdez** Mire észbekaptam, már egy rádiómű közepén voltam. **kérdez** Kínél van a pogácsa? **kérdezte** egyik unokadóm **kérdez** Kristóf **kérdez** Kristóf **kérdez** Meg ne ígess magad - sikította keresztül a
- 3 worldshots.hu Ádám van még abból a pálnkából? **kérdez** Kínél van pálnka? **kérdez** Ki meri ábrvni letez a kezét? **kérdezte** Kristóf, miközben csillogó szemtel a fátýrája szegete **kérdez** Csenedesebben **kérdez** Más síroknál oh
- 4 keesz.hu ke és Galgóczy Gábor társelnök is **kérdez** A járvány második hullámában személyes tapasztalatairól **kérdeztünk** meg ismert embereket: Honnan merítjük az erőt, amikor tele vagyunk szorongással és a tartalékaink
- 5 keesz.hu járjátói a világhálón érhető el **kérdez** A Magyar Katolikus Rádío Zénés napköszöntő című műsorában **kérdezte** a megújulás okáról és a tervéről Rochlitz Bernadett főszerkesztőt Szentáheyl Csongor **kérdez** Az is
- 6 gportal.hu hetente összejönnek, hogy együtt megbeszéljék, és lelki erőt nyerjenek belőle **kérdez** A vezetőjük azt **kérdezte** : "Régóta magvan már ez a kiadvány? **kérdez** A kör kibővítése Évente több millió olvasó és ír
- 7 tv2.hu Inél kevesebb gyereket szüő nők lassabban öregednek - állítja egy új kutatás **kérdez** Szűzanyukákat **kérdeztünk** : HÁBORÚBAN A **kérdez** Herczeg-Kis Bálint gimnazista évek óta várja, hogy megmértesse magá
- 8 gov.hu helyettes államtitkár mutatta be **kérdez** A mai Országos Érdekegyeztető Tanácsülésen napirend előtt **kérdezte** a munkavállalói oldal a kormány képviselőit a korongedményes nyugdíj jövőjéről **kérdez** Az esemény
- 9 avorospostakocis. egy végre újíték össze egy itaira, igaz, most hármasban lennének, de ott válnak rád cse- desben, ne **kérdezen** semmit, **kérdez** és nem mondatod el, mi volt a fékezés, ám határozott terv, melyel legnap este hozzá ér
- 10 gportal.hu 1 új vagyok az oldalon **kérdez** Régóta olvasgatom, de csak most regisztráltam. **kérdez** Azt szeretném **kérdezeni** , hogy tudok-e diasorozatokat vagy anekdotákat innen e-mailen küldeni a barátainak **kérdez** Tud va
- 11 c3.hu **kérdez** Egészen elképvne attól, hogy németül kifogásolhatóan tudó ismerőőm, a szövegre rákérve, azt **kérdezte** : "Polen? **kérdez** Már 1967-ben is lengyel volt a pápa? **kérdez** " Tehát a wittgenstein-behaviourista filo
- 12 gportal.hu **kérdez** Ki ne hagyó **kérdez** Sziasztok! **kérdez** Ha valakinek kérdése van, kérlek gyere el a honlapra, és **kérdezd** meg! **kérdez** Jöhet szerelem-, barátságpanasz is **kérdez** Válaszok **kérdez** Barka votam, a www.tini
- 13 andranum.hu an közvetlen, közeli és környéken tapasztalható, mintha csak a szomszéd szobából kiabálnának át." **kérdezd** meg az Urat! **kérdez** - mondják. **kérdez** Erre megkérdések, a válasz pedig egyértelmű, homály, réhék
- 14 magyarizmusip. ulásának kockázatát **kérdez** A kockázatokról és mellékhatásokról olvassa el a betegjelöztetőt vagy **kérdezte** meg kezelőorvosát, gyógyszerészét **kérdez** Magyar Hypertonia Társaság, szerkesztette: Dr. **kérdez** K

## Szólista (Wordlist)

A szólista (Wordlist) funkció segítségével különböző szempontok szerint összeállított gyakorisági listát hozhatunk létre. A keresési szempontok között megadhatunk szófajra vonatkozó kritériumokat, tartalmazandó karaktersorozatot, konkrét lemmákat stb. A 7. ábrán látható lekérdezésben azok a szavak jelennek meg, amelyek a *hoz* karaktersorozattal kezdődnek.

## 7. ábra

A *hoz* karaktersorozattal kezdődő szavak

WORDLIST Hungarian Web 2020 (huTenTen20)

word (7,983 items | 10,145,318 total frequency)

Word	Frequency ? ↓	Word	Frequency ? ↓	Word	Frequency ? ↓
1 hozzá	2,239,655 ...	11 hozzájuk	212,647 ...	21 hozzászólására	79,193 ...
2 hozott	527,890 ...	12 hozzászólások	191,767 ...	22 hoztam	78,239 ...
3 hozta	335,734 ...	13 hozták	149,911 ...	23 hozzávalók	77,155 ...
4 hozzátette	293,191 ...	14 hoztak	144,371 ...	24 hozzájárult	75,603 ...
5 hoz	292,815 ...	15 hozzájárul	130,108 ...	25 hozzátéve	63,570 ...
6 hozza	284,257 ...	16 hozzáférés	115,815 ...	26 hozhat	58,344 ...
7 hozzánk	275,498 ...	17 hozzád	111,915 ...	27 hozzáférést	55,989 ...
8 hozzám	274,547 ...	18 hozzák	104,311 ...	28 hozzáadott	54,625 ...
9 hozni	261,823 ...	19 hoznak	99,418 ...	29 hozzon	52,509 ...
10 hozzászólás	261,476 ...	20 hozzáadjuk	83,329 ...	30 hozzájárulás	52,177 ...

You are only allowed to access 1,000 items. [Get more](#)

## Kulcsszavak (Keywords)

A kulcsszavak funkció tulajdonképpen automatikus terminológiakivonatolás, melynek során az elemző bizonyos algoritmusok segítségével igyekszik kiszűrni a terminusok kritériumainak megfelelő szavakat a korpuszból. A rendszer az egy és a több szóból álló terminusok kivonatolására is képes. A 8. ábrán látható lekérdezési eredmény a német nyelvű EUR-Lex korpuszból automatikusan kigyűjtött terminusokat tartalmazza.

### 8. ábra

A német nyelvű EUR-Lex korpuszból automatikusan kivonatolt terminusok

## KEYWORDS

EUR-Lex 2/2016 parallel – German



SINGLE-WORDS ✓

MULTI-WORD TERMS ✓



reference corpus: German Web 2020 (deTenTen20) (items: 8,640,514)

Term	Term
1 europäische Union ...	14 Kommission der Europäischen Gemeinschaften ...
2 europäische Gemeinschaft ...	15 vorliegender Fall ...
3 europäisches Parlament ...	16 betreffender Mitgliedstaat ...
4 zuständige Behörde ...	17 Gründung der Europäischen Gemeinschaft ...
5 Amtsblatt der Europäischen Union ...	18 folgende Fassung ...
6 anderer Mitgliedstaat ...	19 Entscheidung der Kommission ...
7 vereinigtés Königreich ...	20 Eg des Europäischen Parlaments ...
8 Ewg des Rates ...	21 Vorschlag der Kommission ...
9 nachstehender Grund ...	22 angefochtene Entscheidung ...
10 Erwägung Nachstehender Gründe ...	23 folgende Verordnung ...
11 europäische Kommission ...	24 vorliegende Verordnung ...
12 staatliche Beihilfe ...	25 vereinigter Staat ...
13 Name der Kommission ...	26 hergestellte Ware ...

## Szövegtípus-elemzés (Text type analysis)

A szövegtípus-elemzés a korpusz metaadatairól nyújt felvilágosítást. Segítségével többek között azt is megtudhatjuk, hány dokumentumból áll a korpuszunk, mekkora a terjedelme, milyen forrásból származnak a szövegek, melyik évben jelentek meg, de számos egyéb szűrési és keresési opció is rendelkezésre áll. A megjelenő eredmények száma és részletessége a korpusz annotáltságától függ.

## Összegzés

Az ismertető a Sketch Engine nevű korpuszelemző alkalmazás legfontosabb funkcióit mutatta be. Fontos azonban megjegyezni, hogy a leírások nem mindenre terjednek ki, csupán az egyes funkciókban rejlő alapvető lehetőségekre fókuszálnak. A bemutatott műveletek mindegyike tovább pontosítható – ennek részleteiről a rendszer súgó-jában találhatunk információkat –, így némi gyakorlás után a különböző szűrési és keresési beállítások finomhangolásával precízen összeállíthatjuk az aktuális kutatáshoz vagy bármilyen kereséshez szükséges pontos lekérdezést.

---

## IRODALOM

- Klaudy, K. (2005). Párhuzamos korpuszok felhasználása a fordításkutatásban. In Lanstyák, I. & Vančoné Kremmer, I. (szerk.), *Nyelvészetről – változatosan. Segédkönyvek egyetemisták és a nyelvészet iránt érdeklődők számára* (pp. 153–185.) Gramma Nyelvi Iroda.
- Robin, E., Dankó, Sz., Götz, A., Nagy A., L., Pataky, É., Szegh, H. & Zolczer, P. (2016). Fordítástudomány és korpuszkutatás: bemutatkozik a Pannonia Korpusz. *Fordítástudomány*, 18(2), 5–26.
- Robin, E. & Seidl-Péché, O. (2020). Korpuszalapú fordításkutatás Magyarországon. In Robin, E. & Seidl-Péché, O. (szerk.), *Fókuszban a fordított és a tolmácsolt szöveg: Korpuszalapú fordításkutatás Magyarországon* (pp. 9–24.) ELTE BTK Fordítástudományi Doktori Program, MANYE Fordítástudományi Szakosztály. <https://doi.org/10.36252/Nyelvikozvsegedkonyv1.1>
- Seidl-Péché, O. & Robin, E. (2019). Alkalmas-e a korpuszalapú módszer a terminuskezelési stratégiák kutatására? Dróth, J. (szerk.), *Korpusz és kontrasztivitás a szakfordítás oktatásában és gyakorlatában* (pp. 93–104.) Károli Gáspár Református Egyetem, L'Harmattan Kiadó.

*Szlávik Szilárd*