

MNSZ – MAZSOLA – FORDÍTÁS

Sass Bálint

sass.balint@nytud.mta.hu

MTA Nyelvtudományi Intézet – PPKE ITK

Fordítók és Tolmácsok Őszi Konferenciája
Budapest, 2012. szeptember 28.

KORPUSZOK A FORDÍTÁS SZOLGÁLATÁBAN

- forrásnyelvi korpusz:
a forrásnyelvi szöveg értelmezésében segít
- célnyelvi korpusz:
a célnyelvi szöveg megfogalmazásában segít
- párhuzamos korpusz:
mindkettő; a két nyelv viszonya, hasonlóságai, különbségei

Magyar nyelvű korpuszok –
a magyar célnyelvi szöveg megfogalmazásának támogatása.

1 MNSZ

2 MAZSOLA

MAGYAR NEMZETI SZÖVEGTÁR


- a magyar nyelv reprezentatív korpusza
- 187,6 millió szó
- *stílusrétegek*: sajtó, szépirodalom, tudományos, hivatalos, személyes
- *regionális változatok*: magyarországi, szlovákiai, kárpátaljai, erdélyi, vajdasági
- szófajlilag elemzett + egyértelműsített

<http://corpus.nyttud.hu/mnsz>

Szavak, szókapcsolatok használatát térképezhetjük fel általa.

MNSZ: „ember”

1. ember szófaj: MSD-kód:



Egy elemű véletlen minta szó kontextusban.

A szóalakon kívül az MSD-kód is jelenjen meg. Attribútumok felbukkanó ablakban.

Rendezés: Bibliográfiai adatok:

Nyelvváltozat:

Alkorpusz:

Alkorpuszok szerinti megoszlás. Szerző:

[keresés indítása](#)

MNSZ: „ember”

Nyelvváltozat: magyarországi, szlovákiai, kárpátaljai, erdélyi, vajdasági

Alkorpusz: sajtó, szépirodalom, tudományos, hivatalos, személyes

Lekérdezés: [lemma = "ember"] ;

Lekérdezett korpusz mérete: 181328691 szó

Találatok száma: 262868 db 1449,68 db / millió szó

Keresési idő: 0,7s

- | | | |
|-----|--|--|
| 1. | magam. Nem egy szimpatikus ember <i>N.NOM</i> | a Simicska, az tutti |
| 2. | meg) leginkább a náthás ember <i>N.NOM</i> | á betűihez lehet hasonlítani. |
| 3. | már magában is vagy százezer embert <i>N.ACC</i> | akadályozott meg abban, hogy |
| 4. | az, hogy nem szabad emberek <i>N.PL.NOM</i> | akaratának ellenszegülni, hanem az |
| 5. | a döntés, ezek az emberek <i>N.PL.NOM</i> | csak végrehajtották a kötelességüket " |
| 6. | el, és mintegy hatvanmillió ember <i>N.NOM</i> | használja. A rekordot nem |
| 7. | hisz nem azért vesz az ember <i>N.NOM</i> | ingatlant drága pénzért, hogy |
| 8. | Ivan Karamazov mondaná. Az embereket <i>N.PL.ACC</i> | korántsem az érdekli, hogyan |
| 9. | , még nem szűntem meg ember <i>N.NOM</i> | lenni: gondolkoztam. Viselkedésem |
| 10. | különböző munkatáborokban, ebből 56 ember <i>N.NOM</i> | maradt úgy oda, hogy |
| 11. | s hogy az itt élő emberek <i>N.PL.NOM</i> | minden gondolata, babonáik, |
| 12. | Lottó a matematikához nem értő emberek <i>N.PL.NOM</i> | önkéntes adója. Nem én |
| 13. | igaz. De bizonyára minden ember <i>N.NOM</i> | számára megvan a horog, |
| 14. | hogy a zsidó állam 400 embert <i>N.ACC</i> | , a szélsőséges Hamász és |
| 15. | ilyet? Azon önmagukba fordult emberek <i>N.PL.NOM</i> | , akiket a vezetőik, |
| 16. | hogy Krisztus a földön nem ember <i>N.NOM</i> | , hanem tökéletes istenség volt |
| 17. | Holist, mielőtt legorombítod az embereket <i>N.PL.ACC</i> | , hogy nem olvasnak, |
| 18. | kisded tisztikarral, középén az ember <i>N.NOM</i> | - vagy állatvonta, a |
| 19. | buta, sot bunko muveletlen emberek <i>N.PL.NOM</i> | . Konyorgom 250 millió emberrel |
| 20. | végződött azzal, hogy az ember <i>N.NOM</i> | ... Tradoni nem " akarta |

MNSZ: „fut után”

1. fut MSD-kód:

2. után MSD-kód:

Egy elemű véletlen minta kontextusban.

A szóalakon kívül az MSD-kód is jelenjen meg. Attribútumok felbukkanó ablakban.

Rendezés: Bibliográfiai adatok:

Nyelvváltozat:

Alkorpusz:

Alkorpuszok szerinti megoszlás. Szerző:

keresés indítása 

MNSZ: „fut után”

Nyelvváltozat: magyarországi, szlovákiai, kárpátaljai, erdélyi, vajdasági

Alkorpusz: sajtó, szépirodalom, tudományos, hivatalos, személyes

Lekérdezés: [lemma = "fut"] [] { 0, 5 } [word = "után"] ;

Lekérdezett korpusz mérete: 181328691 szó

Találatok száma: 295 db 1,63 db / millió szó

Keresési idő: 12,89s

- | | | |
|-----|---|--|
| 1. | home-videózás céljából, és most fut <i>V.e3</i> | a befektetett energiája után <i>NU</i> . Tán így. Vigasztaljon |
| 2. | kell kezdeni - van lábuk fussanak <i>V.Pt3</i> | és busz után <i>NU</i> ! Aztán jöhetnek az öt |
| 3. | részezen dalolt a gazda, futottam <i>V.Me1</i> | a gazda után <i>NU</i> . Ugrottam, buk fenceztem is |
| 4. | azért feltápázkodott, s újra futott <i>V.Me3</i> | a királyfi után <i>NU</i> . Mikor még egyszer körbefutották |
| 5. | árak újra estek, mindenki futott <i>V.Me3</i> | a pénze után <i>NU</i> , ami már rég nem |
| 6. | önkormányzatnak, amely március óta fut <i>V.e3</i> | a pénze után <i>NU</i> . Az alpolgármester cáfolta a |
| 7. | tagjukból. A MIÉP meg fut <i>V.e3</i> | a pénze után <i>NU</i> . Jelenleg a Rökk Szilárd |
| 8. | - hitből vagy kényszerűségből - fut <i>V.e3</i> | a pénze után <i>NU</i> . Kellemetlen a dolog Kiszelynek |
| 9. | ? Hányan vannak, akik futnak <i>V.t3</i> | a pénzük és az álmaik után <i>NU</i> ? Erre a bíróságokon, |
| 10. | Tudósítónktól 2001._augusztus_9., csütörtök Futottunk <i>V.Mt.</i> | a pénzünk után <i>NU</i> - mondja az IT-Com ügyvezetője |
| 11. | őket Segesvárra. Éva hosszan futott <i>V.Me3</i> | a vonat után <i>NU</i> . Egyik kezével integetett, |
| 12. | , a síri éjbe, fuss <i>V.Pe2</i> | az eltűnt cél után <i>NU</i> ! (Kálnoky László fordítása |
| 13. | következménye is volt, hogy futottunk <i>V.Mt1</i> | az eredmény után <i>NU</i> . Sőt zsinórban három norvég |
| 14. | Nyírfa-ügy két főszereplője magas karriert futott <i>V.Me3</i> | be a rendszerváltás után <i>NU</i> . Nikolits István a Népszava |
| 15. | Jöjjön velünk. - Ne fuss <i>V.Pe2</i> | déliab után <i>NU</i> . - Semmi okod kételkedni |
| 16. | emberek ilyenkor. Hát így futunk <i>V.t1</i> | egymás után <i>NU</i> . Kiabálok Kukuj után. |
| 17. | ilyen elszántan, ilyen szemérmetlenül futunk <i>V.t1</i> | életünk után <i>NU</i> . Velünk szemközt a már |
| 18. | számlát kell ellenőriztetnem, s futottam <i>V.Me1</i> | Judit után <i>NU</i> . Éppen letette a kagylót |
| 19. | fiúra. - Én nem futok <i>V.e1</i> | senki után <i>NU</i> , tőlem jöhet akárki, |
| 20. | szemközti oldalon megjelent Pista, futott <i>V.Me3</i> | , egy szánkót húzott maga után <i>NU</i> . Átintett a fiúnak. |

MNSZ – JÖVŐ

Jelenleg készül az MNSZ2.

- nagyobb
- jobb elemzés
- részletesebb elemzés:
 - összetett szavak
 - tagmondatjelölés
 - főnévi csoportok megjelölése
 - tulajdonnevek megjelölése

1 MNSZ

2 MAZSOLA

MAZSOLA

- az igék és bővítmények viszonyainak feltérképezésére szolgáló korpuszlekérdező
- az MNSZ teljes anyagát tartalmazza
- tagmondatokra bontott
- igék, bővítmények megjelölve

<http://corpus.nytud.hu/mazsola>

Milyen igék milyen névszókkal állnak együtt?

Milyen névszók milyen igékkel?

Hogyan *szokás* megfogalmazni bizonyos dolgokat?

MAZSOLA: *áttekint -t*



Korpusz:

Igető:

Nem: Eset/névtű:

Nem: Vonzató:

Nem: Eset/névtű:

Nem: Vonzató:

Nem: String:

Teljes mondatfedés:

Mehet

Eloszlás:



MAZSOLA: *áttekint -t*



Korpusz: Magyar Nemzeti Szövegtár



Igető: áttekint

Nem: Eset/névtű:

Nem: Vonzató:

Nem: Eset/névtű:

Nem: Vonzató:

Nem: String:

Teljes mondatlefedés:

Mehet

Eloszlás:



MAZSOLA: *áttekint -t*



Korpusz: Magyar Nemzeti Szövegtár

Iget: **áttekint**

Nem: Eset/névutó:

Nem: Vonzató:

Nem: Eset/névutó:

Nem: Vonzató:

Nem: String:

Teljes mondatlefedés:

Mehet

Eloszlás:



MAZSOLA: *áttekint -t*



Korpusz: Magyar Nemzeti Szövegtár

Igető: *áttekint*

Nem: Eset/névtő:t

Nem: Vonzató:

Nem: Eset/névtő:

Nem: Vonzató:

Nem: String:

Teljes mondatlefedés:

Mehet

Eloszlás:



MAZSOLA: *áttekint -t*



Korpusz: Magyar Nemzeti Szövegtár

Igető: *áttekint*

Nem: Eset/névt.ő:t

Nem: Vonzató:

Nem: Eset/névt.ő:

Nem: Vonzató:

Nem: String:

Teljes mondatlefedés:

Mehet

Eloszlás:



MAZSOLA: *áttekint* -t



Korpusz:

Igétő:

Nem: Eset/névtű: Nem: Vonzattű:

Nem: Eset/névtű: Nem: Vonzattű:

Nem: String:

Teljes mondatlefedés:

Mehet

Eloszlás:

◊

◊

◊

3238 találat. [helyzet](#) [226] [kérdés](#) [166] [NULL](#) [367] [lehetőség](#) [80] [feladat](#) [55] [működés](#) [42] [alakulás](#) [33] [tapasztalat](#) [37] [ügy](#) [40] [tevékenység](#) [35] [rendszer](#) [33] [történet](#) [28] [dokumentum](#) [26] [kérdéskör](#) [20] [az](#) [86] [eredmény](#) [32] [anyag](#) [27] [probléma](#) [25] [viszony](#) [22] [javaslat](#) [22] [folyamat](#) [21] [gazdálkodás](#) [16] [ez](#) [37] [munka](#) [22] [adat](#) [20] [program](#) [18] [hatás](#) [18] [terv](#) [17] [szerződés](#) [17] [esemény](#) [16] [végrehajtás](#) [14] [teendő](#) [12] [kapcsolat](#) [18] [egész](#) [17] [állapot](#) [15] [jogszabály](#) [13] [gyakorlat](#) [13] [típus](#) [12] [rész](#) [16] [pont](#) [15] [terület](#) [14] [kör](#) [13] [szabály](#) [12] [időszak](#) [12] [állás](#) [12] [intézkedés](#) [11] [fejlesztés](#) [11] [problémakör](#) [7] [év](#) [15] [törvény](#) [13] [összefüggés](#) [10] [jelentés](#) [10] [szabályozás](#) [9] [témakör](#) [8] [érvényesülés](#) [7] [amely](#) [14] [feltétele](#) [10] [váltás](#) [9] [téma](#) [9] [módszer](#) [9] [együttműködés](#) [9] [változat](#) [8] [tervezet](#) [8] [irodalom](#) [8] [fejleszt](#) [8] [esztendő](#) [8] [előterjesztés](#) [8] [teljesítés](#) [7] [menet](#) [7] [jegykönyv](#) [7] [hatékony](#) [7] [fejlemény](#) [7] [előkészítés](#) [7] [ui](#) [9] [szerep](#) [8] [költészet](#) [8] [gond](#) [8] [lépés](#) [7] [indítvány](#) [7] [terep](#) [6] [struktúra](#) [6] [kínálat](#) [6] [doku](#) [7] [pályá](#) [6] [megvalósítás](#) [6] [megállapítás](#) [6] [szervezmény](#) [6] [koncept](#) [6] [irat](#) [6] [fejlesztés](#) [6] [eszköz](#) [6] [szel](#) [8] [létszám](#) [6] [munkaközi](#) [6]

adat

a Fidesz elnökségének ülésén áttekintették a legújabb közvélemény-kutatási adatokat.
 A korai ómagyar kor névrendszerében - ha pusztán statisztikai adatokat tekintünk is át -
 A szakemberek órák alatt áttekinthetik a T. Ház számítógépes rendszerében lévő adatokat -
 Az IMF-delegáció a Román Nemzeti Bank vezetőivel áttekintette a pénzügypolitika legutóbbi évének adatait, majd Radu Berceanu ipari és kereskedelmi miniszterrel megbeszélte a gazdasági helyzetet.
 d) Ha a megyék adatait tekintjük át,
 Ehhez át kell tekinteni néhány olyan adatot,
 Ezt követően egy pszichiátriai centrumban létrehozott multidiszciplináris team pszichiátriai áttekintik a körtörténeti adatokat, standardizáltan.
 Ha a részletes adatokat áttekintjük, az is nyilvánvaló,
 Ha az európai adatokat áttekintjük,
 Ha korcsoportonként tekintjük át adatainkat,
 hogy érdemes hosszabb távot vizsgálni és valamennyi intézet adatait áttekinteni és akkor inkább látjuk,
 hogy Tokióban az illetékesek áttekintették a légkörben, a talajban és a vízben mért sugárzási adatokat.
 miközben a roppant mennyiségű adatot áttekinttem.
 Most felidéztem az előző évi statisztikai adatokat a két 663. rendszerről.

MAZSOLA A FORDÍTÓI MUNKÁBAN

A magyar célnyelvi szöveg megfogalmazásában segít.

- meet the requirements → *megfelel* a követelményeknek
- make a contract → szerződést *köt*

„Modern magyar kollokációs szótár hiányában nagyon jó szolgálatot tesz az adatbázis.”

MAZSOLA: magasra ??? lécet



Korpusz:

Igető:

Nem: Eset/névtű: Nem: Vonzató:

Nem: Eset/névtű: Nem: Vonzató:

Nem: Eset/névtű: Nem: Vonzató:

Nem: Szó:

Teljes mondatlefedés:

Eloszlás:

-
-
-
-

A konkrét igeötvet nem tartalmazó lekérdezések lefuttatása akár 2-3 percig is tarthat.

44 találat. [tesz](#) [21] [emel](#) [10] [helyez](#) [8]

állít

Elvégre a lécet éppen a mieink állították olyan magasra, majd menet közben aztán eldőlt, érdemes -e változtatni és magasabbra állítani a lécet.

áttesz

hogy a múltkorinál még magasabbra tegyük a lécet - és át

emel

A szerző szerint azonban az új tagok teljesítménye magasabbra emelte a lécet:

és nem emeli egyre magasabbra a lécet -

hogy a tiltakozók most magasabbra emelték a lécet, a tavalyinál

- Magasra emelte a lécet a Városháza,

mert Antall József egy nagyon nehéz időszakban rendkívül magasra emelte a lécet.

mert Bécs a többi EU-tagállamtól eltérően túlzottan magasra akarta emelni a lécet a jövődöbéli tagok előtt az atomerőművek biztonsága terén.

mint ellenzék, magasra emelte a lécet a működő kormányokkal szemben.

Olyan magasra kell emelni a lécet,

Sem a szerepe, sem a körítés nem emelte olyan magasra a lécet,

Szász János fantasztikus atmoszférájú filmje talán még magasabbra ' emeli a lécet '.

helyez

ha a Nyugat szándékosan magasabbra helyezné a lécet Szlovákia előtt,

hogy nemcsak magasra helyezi a lécet,

Ilyen magasra azért nem kellett volna helyezni a lécet!

Már nagyon magasra helyezte a lécet.

milyen magasra helyezik ott Brüsszelben a lécet.

Müller Péter és Kapás Dezső azonban már 1965-ben is az eldődknél magasabbra helyezték önmaguk előtt a lécet:

nagyon magasra helyezte a lécet.

sőt a lécet talán magasabbra is helyezi, az utóbbi néhány GENESIS album populáris hangvételű opuszainál...

lök

És ha már, akkor lökjük a legmagasabbra a lécet,

tesz

FŐNÉV ÉS JELZŐ

„Vajon használta-e valaki is valaha ezt a jelzőt ezzel a főnévvel, merül fel bennem olykor fordítás közben.”

– tipikus kérdés a célnyelvi megfogalmazás kapcsán

Adott főnévhez mely *jelzők* illenek a legjobban?

- *kaparós sorsjegy* (9) ↔ „vakarós” (0)
- *közlekedési lámpa* (72) ↔ „forgalmi” (18)
- *??? helyzetbe sodor*

a Mazsola kimenetének feldolgozásával a leggyakoribbak:
lehetetlen, nehéz, kritikus, kilátástalan, kellemetlen. . .

Szükségük van-e a fordítóknak
egy modern magyar kollokációs szótárra?

→ Magyar igei szerkezetek

MAZSOLA – JÖVŐ

- 1 A jelző-főnév kapcsolatot közvetlenül kezelő eszköz.
- 2 Hasonlóí, igéket és bővítményeket kezelő korpuszlekérdező *párhuzamos* korpuszra.
Kísérlet: holland–francia
pl.: *houden + van = aimer + tárgy*
- 3 → új magyar kollokációs szótár

HOZZÁFÉRÉS

Az MNSZ és a Mazsola
közös regisztráció után szabadon hozzáférhető.

```
http://corpus.nytud.hu/mnsz  
http://corpus.nytud.hu/mazsola
```

ideiglenes felhasználói név: vendeg

ideiglenes jelszó: mazsola

HOZZÁFÉRÉS

Az MNSZ és a Mazsola
közös regisztráció után szabadon hozzáférhető.

```
http://corpus.nytud.hu/mnsz  
http://corpus.nytud.hu/mazsola
```

ideiglenes felhasználói név: vendeg

ideiglenes jelszó: mazsola

Köszönöm a figyelmet!

sass.balint@nytud.mta.hu