



NYELV- ÉS BESZÉDTECHNOLÓGIAI PLATFORM

„Mazsola”



NYELVTUDOMÁNYI INTÉZET
MAGYAR TUDOMÁNYOS AKADÉMIA

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

Áttekintés

- Nagy méretű korpuszok → a kinyert nyelvi adatok összegzésére van szükség
- Ilyen eszköz: „Mazsola” – a magyar igék bővítményszerkezetének vizsgálatára
 - korpuszalapú
 - kvantitatív
- Alapegység: *tagmondat* = ige + bővítmények
pl.: „Valakinek csak át tudom adni a végső üzenetet.”
→ reprezentáció: *ige*: „átad” + *-t*: „üzenet” + *-nAk*: „valaki”
- Erre ad választ:
Mely **jellemző** szavak fordulnak elő adott ige mellett adott bővítményként?
→ Jellegzetes igei szókapcsolatokat, idiómákat, szólásokat találunk.

Szövegből szótár

szöveg



morfológiai elemzés, egyértelműsítés

korpusz (Magyar Nemzeti Szövegtár)



részleges szintaktikai elemzés

korpusz – gazdagabb elemzéssel



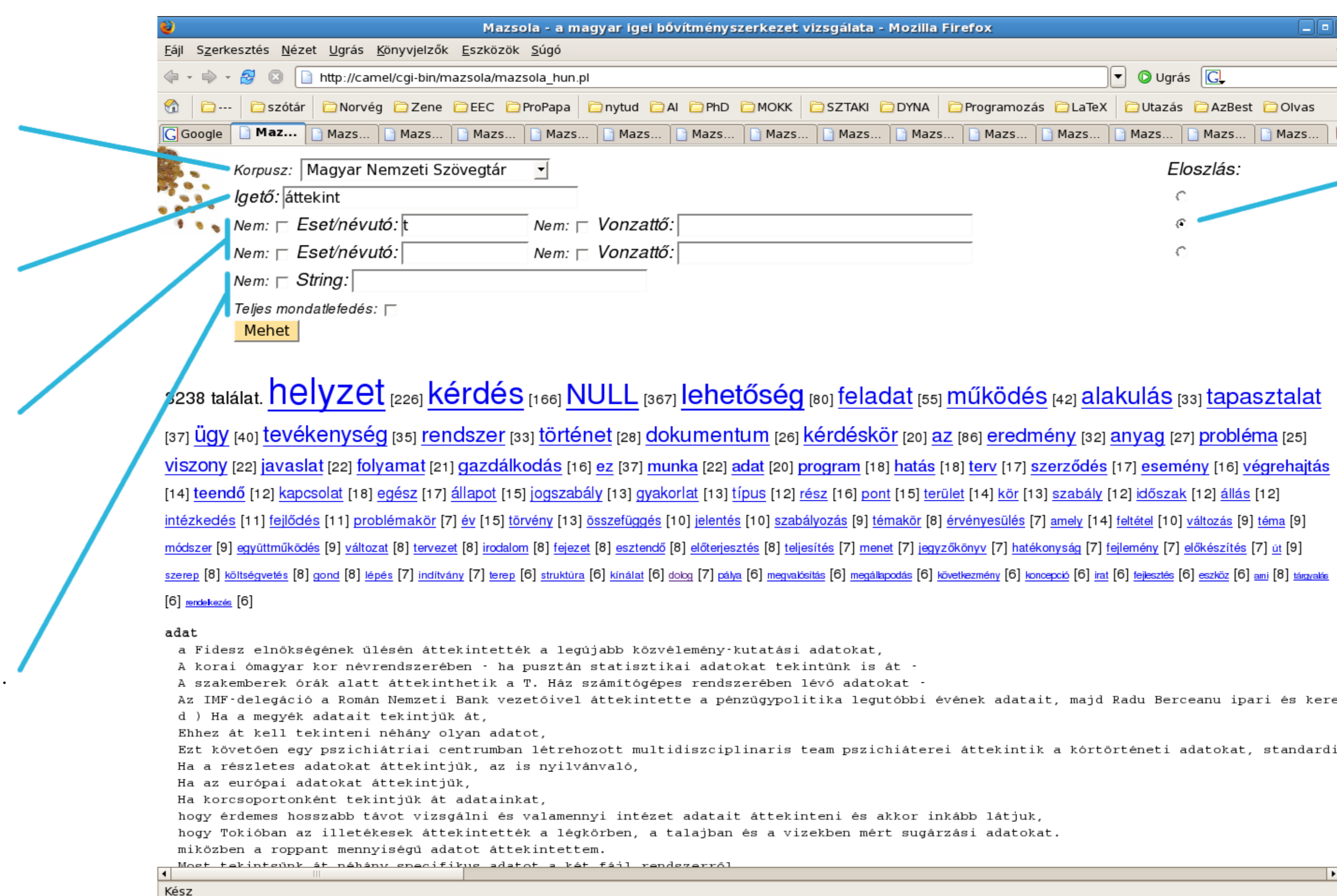
A „Mazsola” felülete és használata

A korpuszt választhatjuk ki.
Lehetőségek: teljes MNSZ;
illetve ennek részkorpuszai:
rövid mondatok, Index fórum,
Magyar Nemzet, szépirodalom.

Az igetövet adhatjuk meg,
pl.: *áttekint*

Egy vagy két bővítményt adhatunk meg eset (*Eset/névtűtő* mező) és/vagy szótót (*Vonzattűtő* mező) segítségével.
pl.: *-t* – a jellemző tárgyakat keressük.

Egyszerű szabadszavas kereséssel tovább szűkíthetjük a lekérdezést.
A teljes mondatlefedés funkcióval olyan példákat kereshetünk, melyekben a megadottakon kívül más bővítmény nem szerepel.



Itt jelölhetjük meg, hogy melyik pozícióban megjelenő jellemző szavak listáját kérjük. pl.: most az áttekint ige jellemző *tárgyaira* vagyunk kíváncsiak.

A megjelölt pozícióban megjelenő szavak listája jellegzetesség (ún. *saliencie*) szerinti sorrendben. Zárójelben a gyakoriság. pl.: *áttekint a helyzetet, áttekinti a kérdést.* (NULL: explicit tárgy nélküli, tárgyasan ragozott igével bíró mondat)

Megfelelő korpuszpéldák a Magyar Nemzeti Szövegtárból, a megjelölt pozícióban megjelenő szavak szerint csoportosítva.

Példák

- szókapcsolatok, szólások
 - *eszik -t* – különféle ételek; de: kása = „Nem eszik olyan forrón a kását.”
 - *üt alany* – üt az óra, Üsse kól!, Csalánba nem üt a mennykő, x forint üti a markát
 - *hány -t* – fitty, szikra, cigánykerék, hányja-veti magát, kereszt, borsót falra

A *hány* igét tartalmazó mondatok 8%-a a *fittyet hány -rA* szerkezetet tartalmazza.
→ A szókapcsolatok, szólások **gyakoriak**.
A „szó szerinti” jelentés ismerete önmagában nem elegendő.

- szemantikai osztályok
 - *veszélyben forog alany* – élet, győzelem, lét, továbbjutás, épség – pozitív konnotációjú fogalmak
 - *fest -t -rA* – színek; de: fal = „Ne fessd az ördögöt a falra!”

Alkalmazás

- a jellemző szerkezetek fordításukkal együtt egy gépi fordító rendszer adatbázisába kerültek: <http://www.webforditas.hu>
pl.: „Fülön csípték a tolvajt. Csípi a szememet a füst.”
- magyar WordNet építése
- szótárak alapanyaga lehet
- nyelvoktatás – elsődlegesen a gyakori szerkezetek

Hozzáférhetőség

<http://corpus.nytud.hu/mazsola>
Felhasználói név: **platform** – jelszó: **mazsola**
Email: mazsola@nytud.hu

szervezetek jellemző szavai



jellemző szerkezetek összegyűjtése

jellemző szerkezetek



manuális javítás, lexikográfiai munka

szótár

Mutatvány a szótárból

beül [2288]
beül -bA [530] pedig legszívesebben beült volna valamelyik padba, ...
beül kocsi-bA [98] Beültünk a kocsiba, ...
beül autó-bA [74] Ül be az autódba, ...
beül [335] Beültünk, ...
beül -rA [106] szépen felöltözve beülnek egy előadásra, ...

elver [744]
elver -t [284] hogy minap elvertelek azért, ...
elver jég -t [36] Már ahol a jég nem verte el a termést!
elver -n por-t [95] vagy hogy egy pár túlbuzgó helyi tanácselnökön verjük el a port.

zúdit [743]
zúdit -rA -t [212] majd hideg vizet zúdítanak rájuk.
zúdit -rA ösztűz-t [32] hogy a szervezett bűnözésre most ösztűzet tudunk zúdítani.
zúdit nyak-bA -t [77] mégis rengeteg bajt zúdítottak a nyakunkba.
zúdit -t [60] Komor felhők szakadatlanul zúdították az esőt.