

Mazsola – mindenkinek


Sass Bálint
MTA Nyelvtudományi Intézet

2018. január 18.
MSZNY 2018, Szeged

Példa

http://corpus.nytud.hu/mazsola

Lekérdezés: *iszik* + *-t*



Mazsola - a magyar igei bővítményszerkezet vizsgálata

[English version](#)

Korpusz: Magyar Nemzeti Szövegtár

Igető: iszik

Nem: Eset/névutó: t Nem: Vonzattó:

Nem: Eset/névutó: Nem: Vonzattó:

Nem: Eset/névutó: Nem: Vonzattó:

Nem: Szó:

Teljes mondatlefedés:

Mehet

Eloszlás:

-
-
-
-

7981 találat. [sör](#) [583] [bor](#) [581] [kávé](#) [318] [korty](#) [206] [víz](#) [459] [tea](#) [238] [pálinka](#) [138] [tej](#) [161]
[konyak](#) [94] [sok](#) [311] [lé](#) [96] [valami](#) [204] [alkohol](#) [86] [rum](#) [65] [whisky](#) [54] [NULL](#) [1045] [mi](#) [217] [több](#) [196]

Háttér – adatbázis

Ha a város a tárcáktól pénzt nyer ,
stem@@nyer ABL@@tárca ACC@@pénz NOM@@város

Háttér – adatbázis

Ha a város a tárcáktól pénzt nyer ,
stem@@nyer ABL@@tárca ACC@@pénz NOM@@város

Igető.

Háttér – adatbázis

Ha a város a tárcáktól pénzt nyer ,
stem@@nyer **ABL@@**tárca **ACC@@**pénz **NOM@@**város

Bővítmények.

Háttér – adatbázis

Ha a város a tárcáktól pénzt nyer ,
stem@@nyer **ABL@@**tárca **ACC@@**pénz **NOM@@**város

Lekérdezés: *nyer + -t*

Mi történik? Megnézi, hogy mely szavak jellegzetesek ezen a helyen.

Jellegzetesség: gyakrabban fordul elő ebben a környezetben, mint általában.

A jellegzetes szavak között ott lesz a *pénz*.

Hasonlóan: rá lehet kérdezni az igére is.

Lekérdezés: *pénz-t*

A jellegzetes igék között ott lesz a *nyer*.

Példa

http://corpus.nytud.hu/mazsola



Mazsola - a magyar igei bővítményszerkezet vizsgálata

[English version](#)

Korpusz: Magyar Nemzeti Szövegtár

Igető: iszik

Nem: Eset/névutó: t Nem: Vonzató:

Nem: Eset/névutó: Nem: Vonzató:

Nem: Eset/névutó: Nem: Vonzató:

Nem: Szó:

Teljes mondatlefedés:

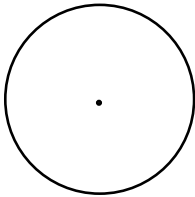
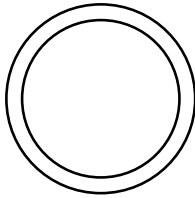
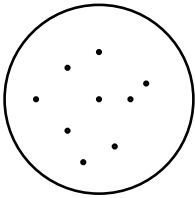
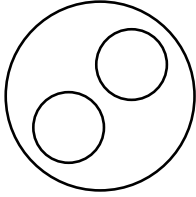
Mehet

Eloszlás:

-
-
-
-

7981 találat. [sör](#) [583] [bor](#) [581] [kávé](#) [318] [korty](#) [206] [víz](#) [459] [tea](#) [238] [pálinka](#) [138] [tej](#) [161]
[konyak](#) [94] [sok](#) [311] [lé](#) [96] [valami](#) [204] [alkohol](#) [86] [rum](#) [65] [whisky](#) [54] [NULL](#) [1045] [mi](#) [217] [több](#) [196]

Eredmény: szemantikailag koherens szóosztályok

méret szám	kicsi	nagy
1		
sok		

Visszafelé: bővítmény → ige

Lekérdezés: *hideg hátán*



Korpusz: Magyar Nemzeti Szövegtár ▼

Igető:

Nem: <input type="checkbox"/>	Eset/névtő: <input type="text" value="alany"/>	Nem: <input type="checkbox"/>	Vonzattő: <input type="text" value="hideg"/>
Nem: <input type="checkbox"/>	Eset/névtő: <input type="text" value="-n"/>	Nem: <input type="checkbox"/>	Vonzattő: <input type="text" value="hát"/>
Nem: <input type="checkbox"/>	Eset/névtő: <input type="text"/>	Nem: <input type="checkbox"/>	Vonzattő: <input type="text"/>

Nem: Szó:

Teljes mondatlefedés:

Mehet

Eloszlás:

-
-
-
-

113 találat. [végigfut](#) [35] [futkos](#) [30] [futkározik](#) [13]

Feladatmegoldás Mazsolával

1/2

kiküszöböli a csorbát (?)

csorbát → szenved, ejt, kiköszörül (sok), kiküszöböl (kevés)

kiköszörül -t → csorba

*kiküszöböl -t (gyakoribb!) → számos különféle tárgy, gyakori jelentés-
elem a valamiféle hiányosság (hiba, hiányosság, probléma, anomália stb.).*

Feladatmegoldás Mazsolával

1/2

kiküszöböli a csorbát (?)

csorbát → szenved, ejt, kiköszörül (sok), kiküszöböl (kevés)

kiköszörül -t → csorba

*kiküszöböl -t (gyakoribb!) → számos különféle tárgy, gyakori jelentés-
elem a valamiféle hiányosság (hiba, hiányosság, probléma, anomália stb.).*

A gyakoribb és nagyon hasonló hangzású kifejezés
magába olvasztotta a ritkább, hasonló jelentésű kifejezést.

Feladatmegoldás Mazsolával

2/2

össze és -vAl

1. jellemzően *-vAl* ragos bővítménnyel:
összefügg, összeköt, összehasonlít, összekever, stb.
2. jellemzően *-vAl* ragos bővítmény nélkül:
összegyűlik, összeállít, összehív, összeszed stb.

Feladatmegoldás Mazsolával

2/2

össze és -vAl

1. jellemzően *-vAl* ragos bővítménnyel:
összefügg, összeköt, összehasonlít, összekever, stb.
2. jellemzően *-vAl* ragos bővítmény nélkül:
összegyűlik, összeállít, összehív, összeszed stb.

két dolgot állítunk relációba \leftrightarrow sok dolgot állítunk relációba
 \rightarrow két csoportra osztva strukturálhatjuk
az *össze* igekötő jelentéseit (az alapigétől függetlenül)
mondhatjuk: a *-vAl* az első esetben **az igekötő vonzata**

Cél

A közzététel célja: a fentiekhez hasonló, magyar igék és bővítmények viszonyainak vizsgálatára alkalmas eszközt bárki készíthessen a saját adataival.

Cél

A közzététel célja: a fentiekhez hasonló, magyar igék és bővítmények viszonyainak vizsgálatára alkalmas eszközt bárki készíthessen a saját adataival.

Vagy akár **más nyelvre**.

Cél

A közzététel célja: a fentiekhez hasonló, magyar igék és bővítmények viszonyainak vizsgálatára alkalmas eszközt bárki készíthessen a saját adataival.

Vagy akár más nyelvre.

Vagy akár **párhuzamos korpuszra**.

Cél

A közzététel célja: a fentiekhez hasonló, magyar igék és bővítmények viszonyainak vizsgálatára alkalmas eszközt bárki készíthessen a saját adataival.

Vagy akár más nyelvre.

Vagy akár párhuzamos korpuszra.

Vagy akár **egyéb strukturákra**.

Cél

A közzététel célja: a fentiekhez hasonló, magyar igék és bővítmények viszonyainak vizsgálatára alkalmas eszközt bárki készíthessen a saját adataival.

Vagy akár más nyelvre.

Vagy akár párhuzamos korpuszra.

Vagy akár egyéb strukturákra.

Minden esetben ugyanúgy igaz lesz: az adott pozíciókban előforduló jellegzetes entitásokat fogja eredményül adni a Mazsola.

Minden azon múlik, hogy az adatbázisban mit hogyan reprezentálunk.

Alkalmasan választott kódolással a fentiek megvalósíthatók.

Párhuzamos korpusz

```
nl_deze te vertalen in wiskundige formules . /  
fr_les traduire en expressions mathématiques .  
stem@@nl_vertalen+fr_traduire  
fr_en@@expression nl_in@@formule
```

Párhuzamos korpusz

```
nl_deze te vertalen in wiskundige formules . /  
fr_les traduire en expressions mathématiques .  
stem@@nl_vertalen+fr_traduire  
fr_en@@expression nl_in@@formule
```

Nyelvek címkével megjelölve.

Párhuzamos korpusz

```
nl_deze te vertalen in wiskundige formules . /  
fr_les traduire en expressions mathématiques .  
stem@@nl_vertalen+fr_traduire  
fr_en@@expression nl_in@@formule
```

Igepár.

Párhuzamos korpusz

```
nl_deze te vertalen in wiskundige formules . /  
fr_les traduire en expressions mathématiques .  
stem@@nl_vertalen+fr_traduire  
fr_en@@expression nl_in@@formule
```

Bővítmények.

Párhuzamos korpusz

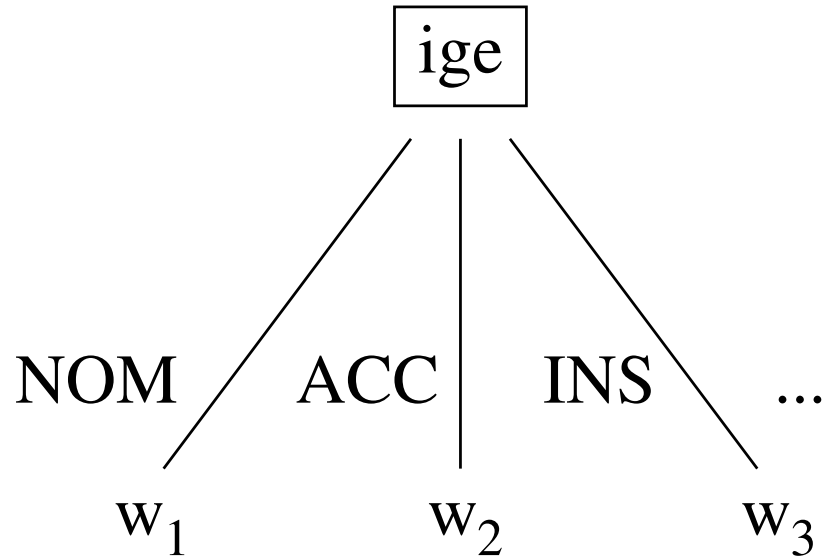
```
nl_deze te vertalen in wiskundige formules . /  
fr_les traduire en expressions mathématiques .  
stem@@nl_vertalen+fr_traduire  
fr_en@@expression nl_in@@formule
```

Lekérdezés:

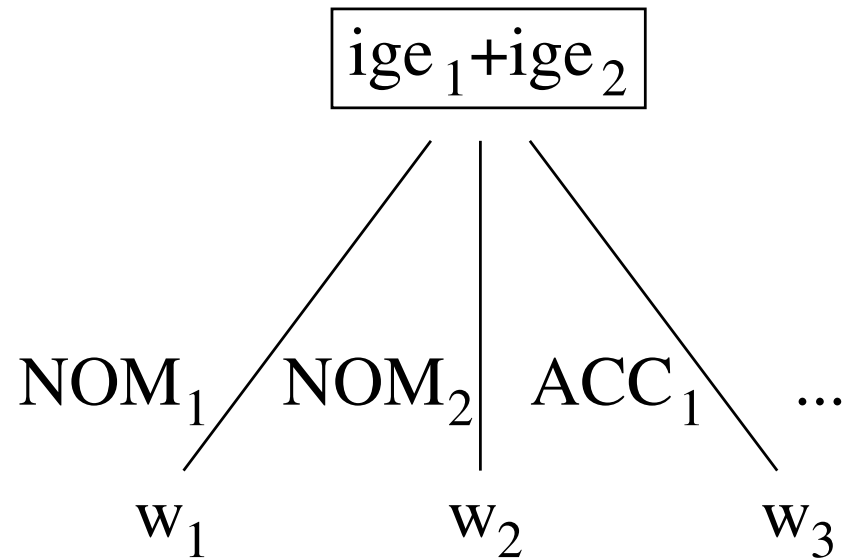
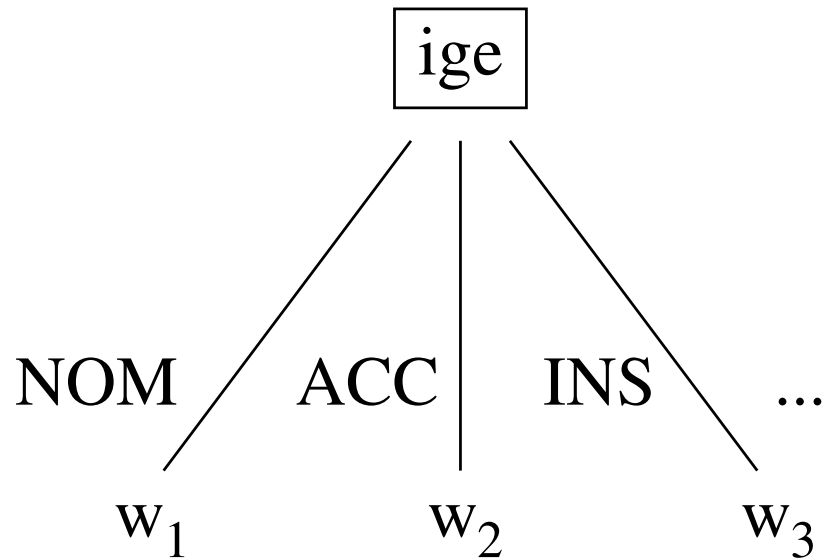
nl_maken+fr_faire nl_ACC=deel fr_ACC=?

→ partie

Párhuzamos korpusz – modell



Párhuzamos korpusz – modell



Főnévre

szabad szemmel

stem@@szem ADJ@@szabad CAS@@INS NUM@@sg

Főnévre

szabad szemmel

stem@@szem ADJ@@szabad CAS@@INS NUM@@sg

A főnévi fej.

Főnévre

szabad szemmel

stem@@szem **ADJ@@**szabad **CAS@@**INS **NUM@@**sg

A főnévi csoport jellemzői.

Ezek a jellemzők eltérő típusúak.

Főnévre

szabad szemmel

stem@@szem ADJ@@szabad CAS@@INS NUM@@sg

Lekérdezés:

fül CAS=-rA NUM=PL ADJ=?

Főnévre

szabad szemmel

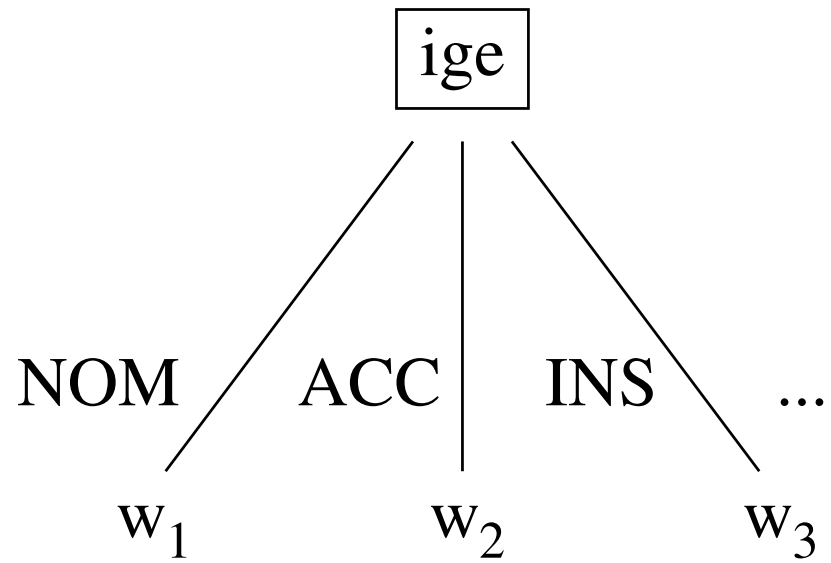
stem@@szem ADJ@@szabad CAS@@INS NUM@@sg

Lekérdezés:

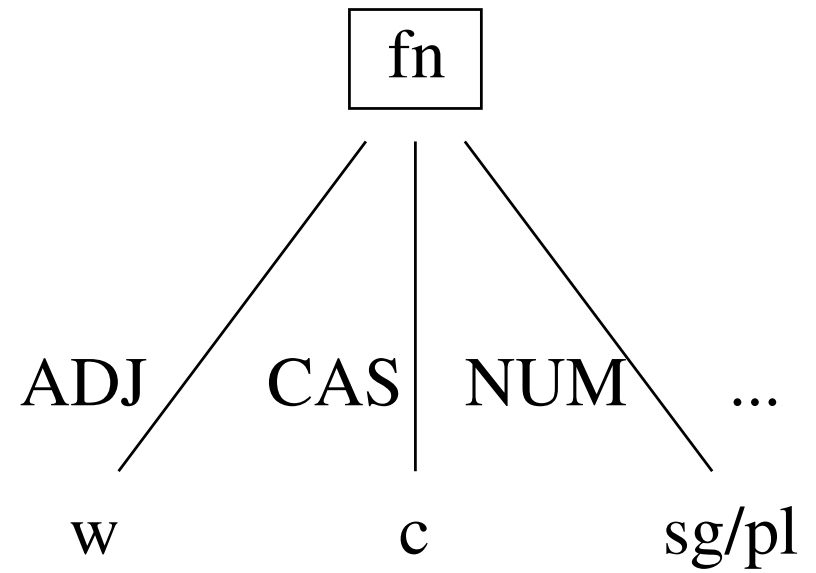
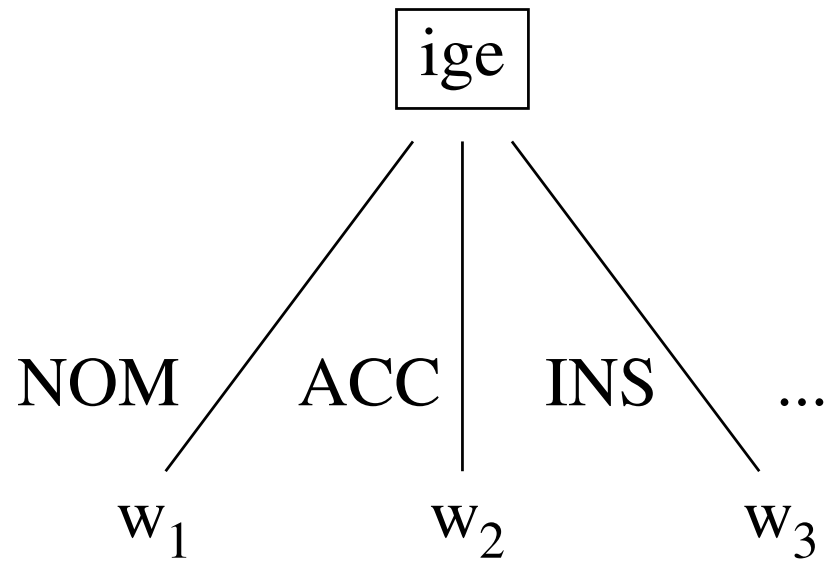
fül CAS=-rA NUM=PL ADJ=?

→ *süket*

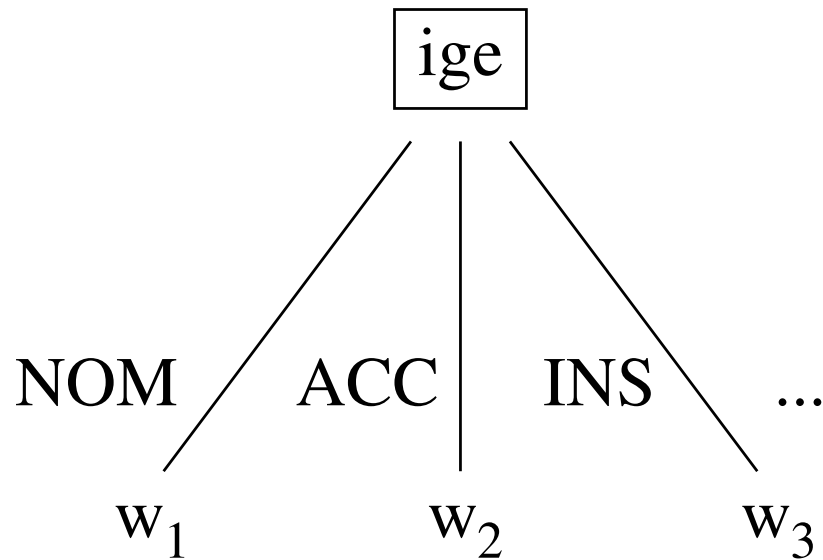
Fónévre – modell



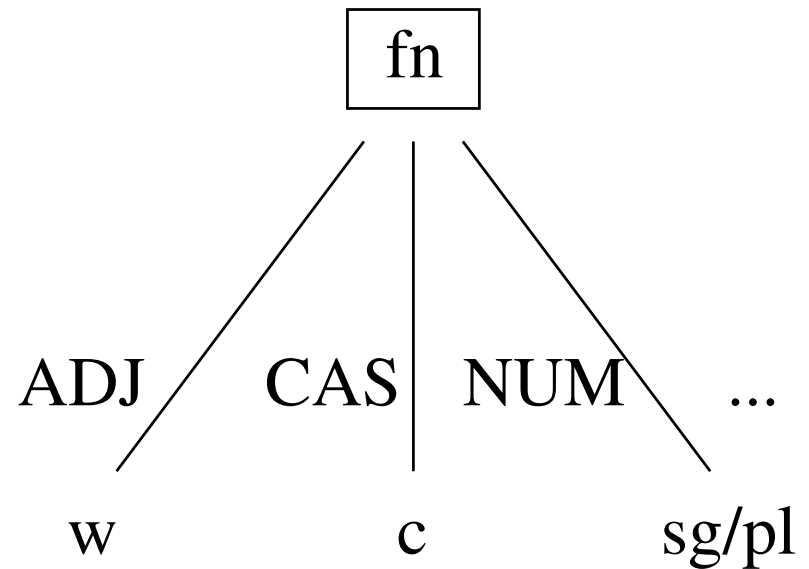
Fónévre – modell



Főnévre – modell



összekever szezont fazonnal



süket fülekre

Megj: a klasszikus igés Mazsola csak a szótöveket tekintette, a morfológiai jegyeket figyelmen kívül hagyta, most tekintetbe vesszük a számot, így adott esetben ki is tud jönni jellegzetesként.

Elérhető

<https://github.com/sassbalint/mazsola>

Itt megtalálható:

- minta-adatbázisok
- installálás leírása
- config-minták – új adatbázishoz

Böngészőből használható.

<http://corpus.nytud.hu/mazsola3>

← itt elérhető egy ideiglenes install mintaadatokkal

Összegzés

- Szabadon elérhető, használható a Mazsola eszköz.
- Más/saját korpusz használatával bárki egyszerűen készíthet hasonló funkcionalitású kutatóeszközt.
- A Mazsola általánossága révén alkalmazható...
nem csak magyar nyelvű,
hanem *más nyelvű* szövegek kezelésére is;
nem csak egynyelvű,
hanem *párhuzamos korpuszra* is;
nem csak igék és bővítmények,
hanem *egyéb struktúrák* vizsgálatára is.