

Bevezetés a nyelvtechnológiába 7.

Fordítástámogatás: szótárak



A nyelvtechnológia szerepe a lexikográfiában

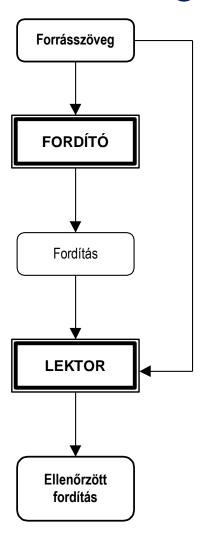
- ☐ A szótárak célja ma
 - emberek számára készülnek
 - gépek számára készülnek
- Miből hozunk létre ma szótárakat?
 - semmiből, technikai támogatással
 - szövegekből, gépi támogatással
 - meglevő szótár(ak)ból, gépi támogatással
- ☐ A géppel támogatott szótárkészítés típusai
 - embernek írt forrásokból, ember számára
 - embernek írt forrásokból, gép számára
 - gépi forrásokból, ember számára
 - gépi forrásokból, a gép számára



A szótár helye a fordítási folyamatban

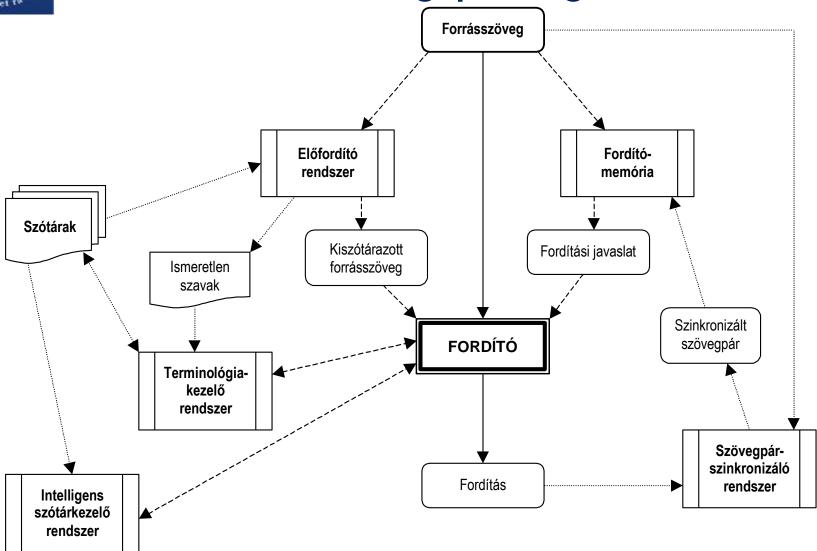


Az emberi fordítás gépi támogatása



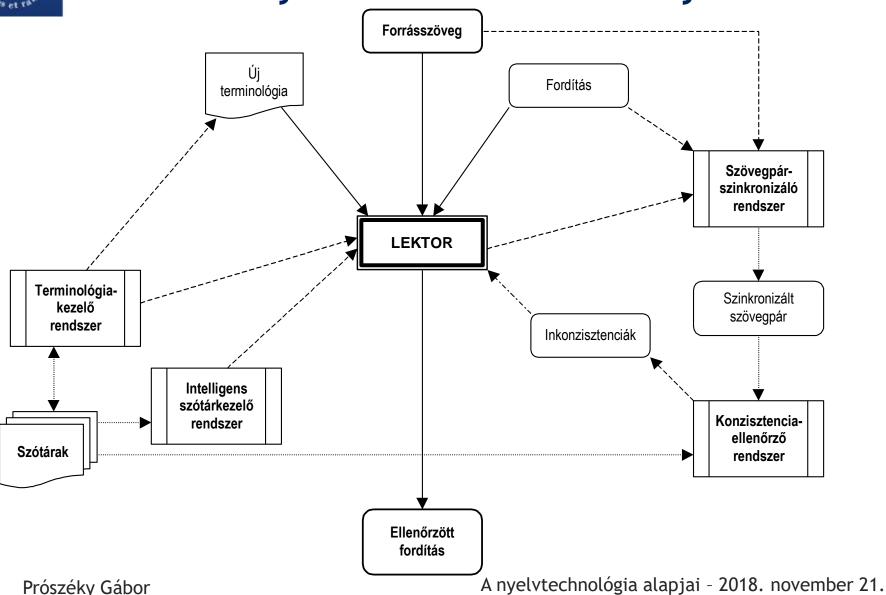


A fordító számítógépes segédeszközei

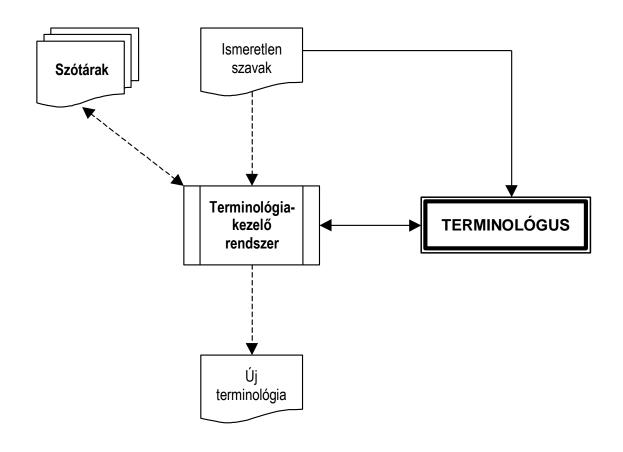




A lektor helye a fordítási munkafolyamatban



A terminológus helye a fordítási folyamatban





Leegyszerűsített szótártipológia

- Nyomtatott szótárakElektronikus szótárak
- SzótárakTerminológiai adatbázisok
- Közvetlen szótárakKözvetett (elektronikus) szótárak
- Kétnyelvű szótárakEgynyelvű szótárakTöbbnyelvű szótárak
- A forrásnyelv és a célnyelvek aszimmetriája papír- és elektronikus alapon



Szótártípusok



A nyelvek száma szerint

- ☐ Ezek voltak először: lexikon, enciklopédia
- Mai egynyelvűek: értelmező, nyelvtörténeti, etimológiai, nyelvjárási, szólás-, szleng-, idegen szavak, helyesírási, gyakorisági, ragozási, a tergo (szóvégmutató) szótárak
- □ Kétnyelvűek
- Többnyelvűek: valójában párba állított kétnyelvűek
- Sőt, az egynyelvűek is két különböző nyelvet használnak: a címszavak és a szócikkek "nyelve, nyelvezete" különbözik



A szótárak általános szerkezete

- ☐ Önálló és utaló szócikkek
- ☐ Szócikkfej (=a szótári baloldal): címszó, homonimák, alak- és írásváltozatok, kiejtés, elválasztás, szófaj, főbb toldalékos alakok, nyelvtani megjegyzés, stílusminősítés
- ☐ Jelentéscsoportok (=a szótári jobboldal) alapjelentés, jelentésárnyalatok, értelmezések és példák, továbbá szólások, közmondások, más szavakkal alkotott összetételek, származékszók

A szótárelemek nyelvtechnológiai felhasználása

🗖 Címszó: kiindulás helyesírási programokhoz
☐ Variánsok és toldalékolt alakok: a morfológiai rendszerhez
☐ Szótagolás: elválasztó programokhoz
☐ Kiejtés: beszédkeltő rendszerekhez
☐ Szófaj: egyértelműsítőkhöz
☐ Témakód: szövegtípus-azonosításhoz
Definíciók: jelentés-egyértelműsítéshez
Példák: a címszó körüli többszavas kifejezések
azonosításához
🗖 "Lásd még" szavak: szinonimák, antonimák



Segéd- és szakszótárak

Szakmai szótárak (enciklopédiák, lexikonok?) ☐ Terminológiai adatbázisok □ Keveredés az alapszótárakkal: általános nyelvi nagyszótárak (tele szakszócikkekkel!) Az egyidejű használat lehetősége miatt megváltozott szerepük az elektronikus világban ☐ Dinamikus szakszótár-kiegészítés alapszótárakhoz

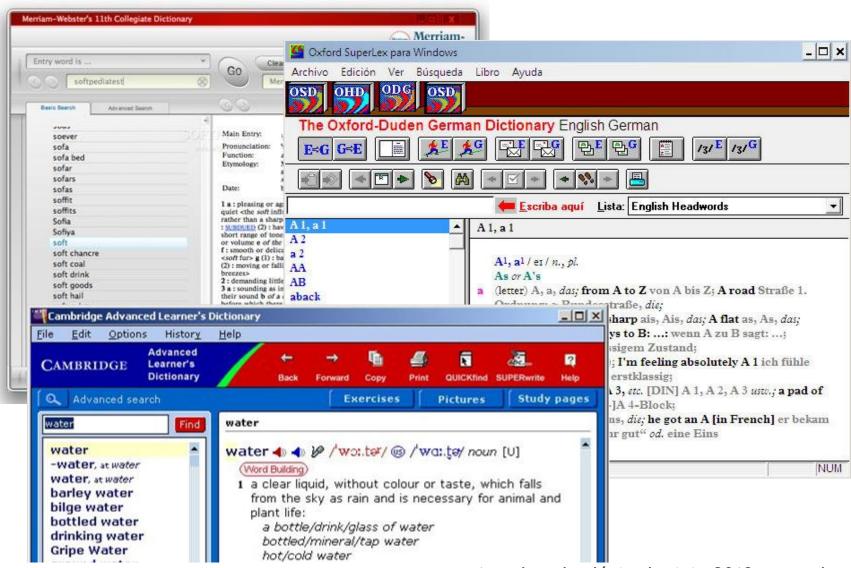


Nyomtatott szótárak vs. elektronikus szótárak

Nyomtatott szótárak	Elektronikus szótárak
Zárt szerkezetű szöveget tartalmaznak.	Zárt vagy nyitott szerkezetű szöveget tar- talmaznak.
Van terjedelmi korlát a címszójegyzék összeállításában és a jelentésstruktúra felépítésében.	Nincs terjedelmi korlát.
A címszavak hierarchikusan, betűrendben követik egymást.	A címszavakat asszociatív útvonalon is elérhetjük.
A szótári rész szócikkekből vagy utaló szócikkekből áll.	A szótári rész szócikkekből áll, az utaló szócikk szerepét a hiperlink veszi át.
A szócikkeket statikus ábrák, képek il- lusztrálják.	A szócikkeket interaktív, multimediális anyagok is illusztrálják.
Manuális keresésre van lehetőség.	Gépi keresésre van lehetőség.
Csak címszó szerint kereshető.	Többszavas kifejezések, toldalékolt ala- kok szerint is kereshető.
Több szótár együttes használata manuáli- san lehetséges.	Több szótár együttes használata kereső- felület segítségével lehetséges.



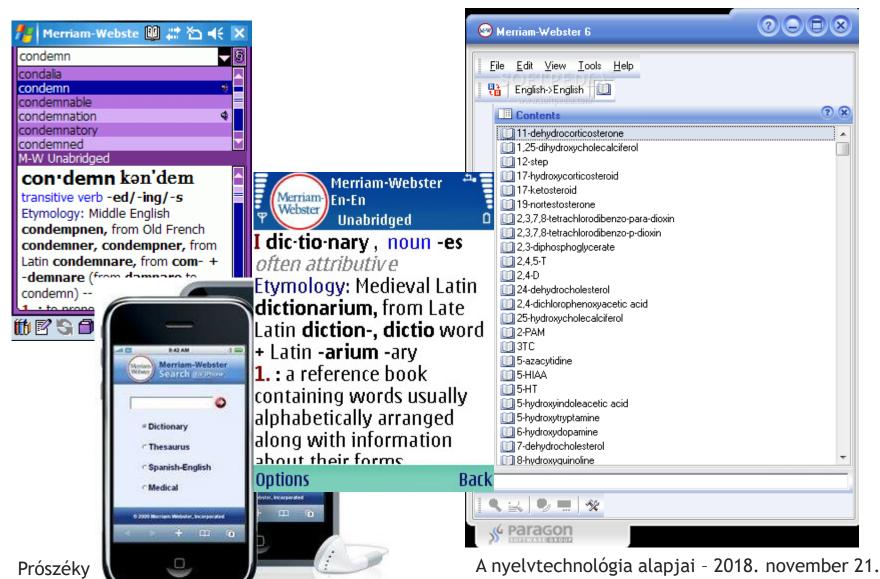
A "hagyományos" elektronikus szótárak



Prószéky Gábor

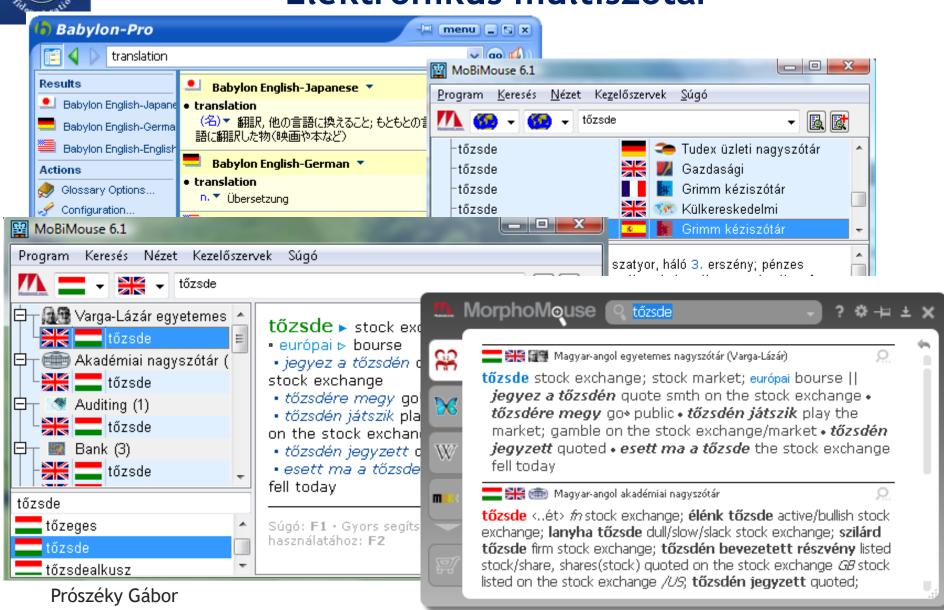


Mindenféle platformon



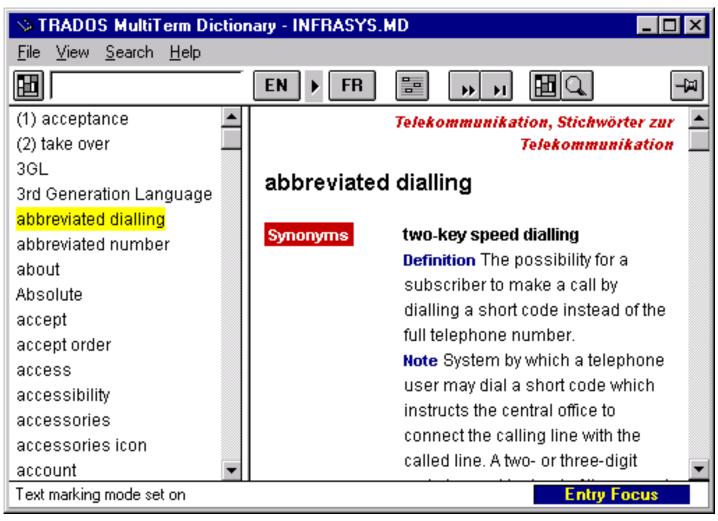


Elektronikus multiszótár





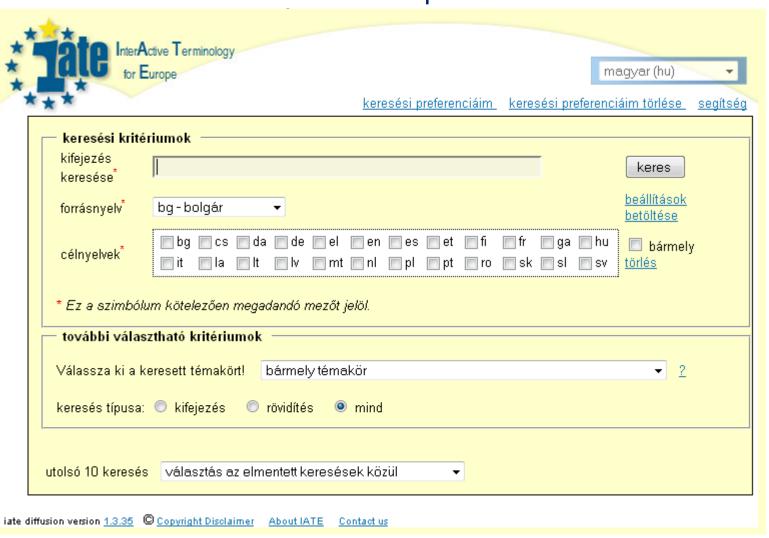
Terminológiakezelők (1)





Terminológiakezelők (2)

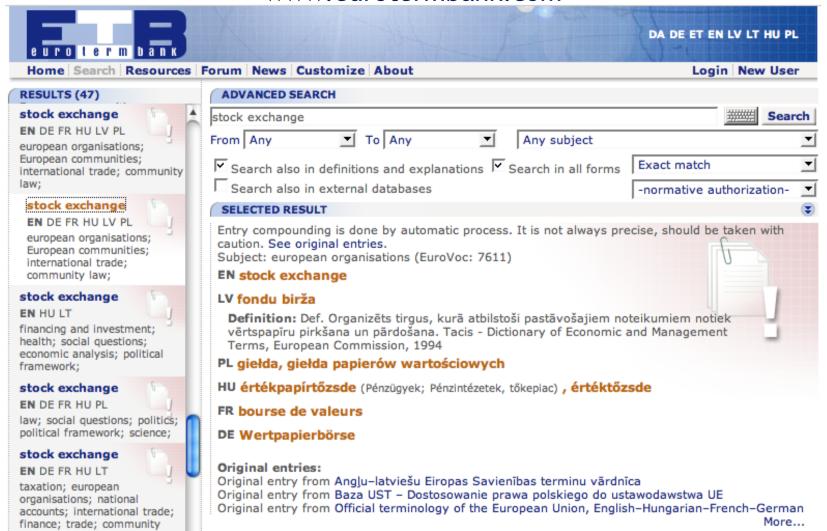
iate.europa.eu





Terminológiakezelők (3)

www.eurotermbank.com



A korszerű internetes szótárszolgáltatás kritériumai

☐ Folyamatosan bővíthető szótárkínálat Sajátszótár-készítési lehetőség Tetszőleges webes tartalom integrált megjelenítése □ A kifejezések intelligens kezelése □ Közösségi jelenlét ☐ Egymás segítésének és a (jogos) kritikának a fóruma A rendszer szemantikus ismereteinek erősítése a felhasználó keresési szokásainak elemzésével Könnyű keresés-indítási lehetőség Saját menthető beállítások a környezet személyre szabásához



Szótárszolgáltatás + közösségépítés





A szótári keresés



Nyelvfüggő szótárproblémák

- □ A forrás- és célnyelv karakterkészleteinek ismerete
- □ A forrás- és a célnyelv ábécérendjeinek ismerete
- □ A fonetikai információ kezelése
- Egységes jelölés: nyelvi keresésnél a szótár grammatikai információival való kompatibilitás

Keresési technikák elektronikus szótárakban

- □ Betű szerint□ Csonkolt keresés
- ☐ Hasonlósági keresés (fuzzy, soundex, spell)
- Nyelvi alapú keresés a bemeneti oldalon
- Nyelvi alapú keresés a találati oldalon
- □ A kifejezések kezelésének problémái: alcímszók, kulcsszó-választás, indexek, egyazon kifejezés több címszó alatt



Többszavas kifejezések keresési módjai

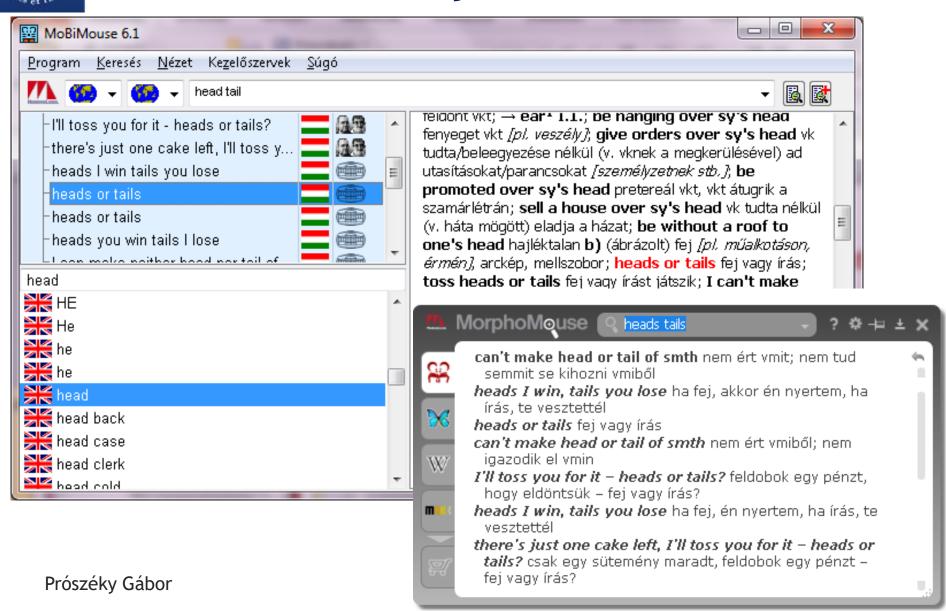
- ☐ Csak címszóként
- ☐ Betű szerint
- □ Teljes szövegű kereséssel
- ☐ Reguláris kifejezésként
- □ Tőindexekkel:

készítéskor vagy elemzési időben (is)

PÁZMÁNY PÉTER KATOLIKUS EGYETEM - KIEMELT PELSŐOKTATÁSI INTÉZMÉNY

INFORMÁCIÓS TECHNOLÓGIAI ÉS BIONIKAI KAR

Többszavas kifejezések keresése



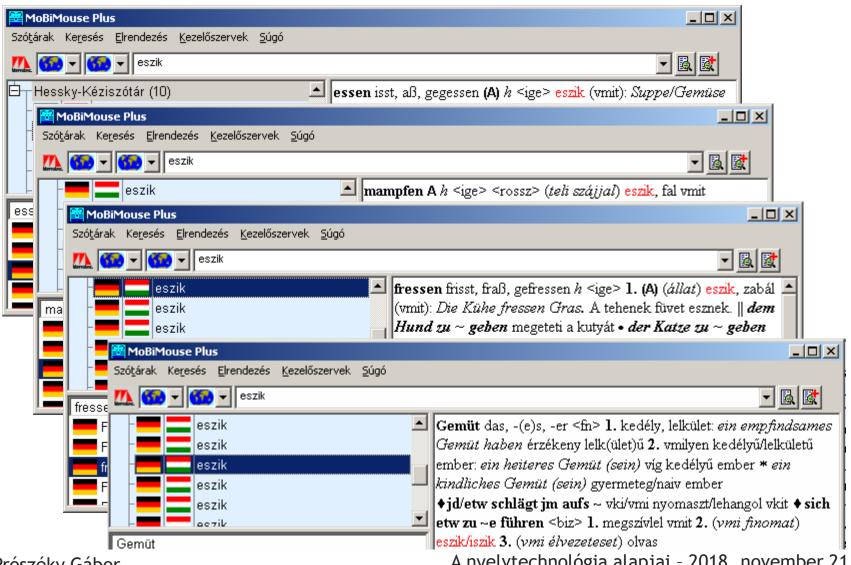


A szótári "jobboldal" szerepe

- ☐ Papírszótárak esetében: csak tipográfiai
- □ Elektronikusan: új lehetőség
- ☐ Ábécé-környezet helyett szinonimák
- ☐ Többféle jelentés kezelése a baloldali címszavak segítségével
- ☐ Új találati ablak
- Elektronikusan érdemes "kifordítani" a szótárakat



A "kifordítható" szótár



Prószéky Gábor



Szótármegjelenítés



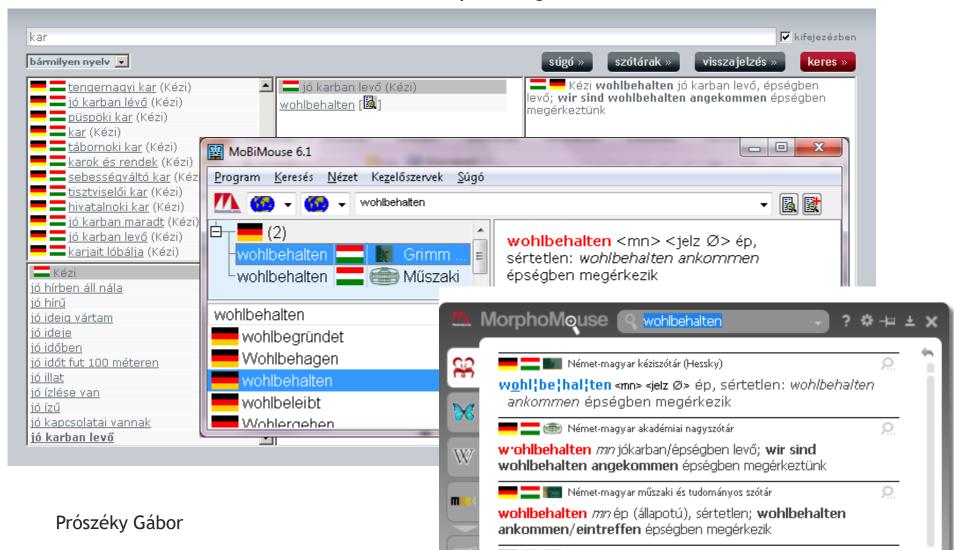
LMF

■ Lexical Markup Framework (LMF): ISO-szabvány (ISO-24613:2008) szótárak leírásához ☐ 1. lépés: létező szótárak struktúráinak konzisztens feltérképezése ■ 2. lépés: az összes feldolgozott szótárat lefedő leírás létrehozása ☐ 3. lépés: 61 szakértő bevonásával az összes szóba jövő szótárszerkezet megvizsgálása Fontos szempont volt a morfológiai támogatás biztosítása ☐ A fejlesztés eredménye (2004-2008): koherens UML-modell



Szótármegjelenítés

Ma már szinte csak XML + XSLT alapú megoldások





Kétféle szótármegjelenítés

(Atkins 2002)

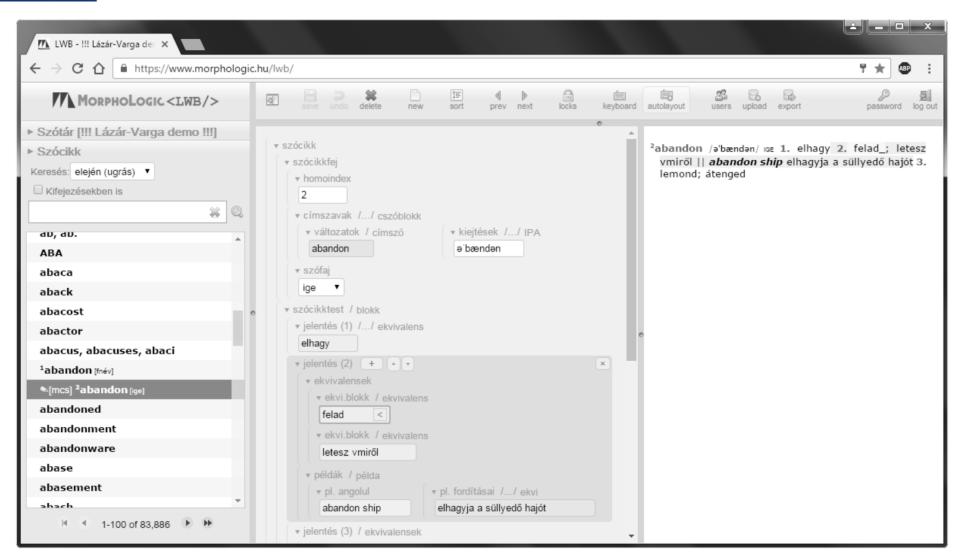




Szótárkészítés



Szótárszerkesztő eszközök (LWB)



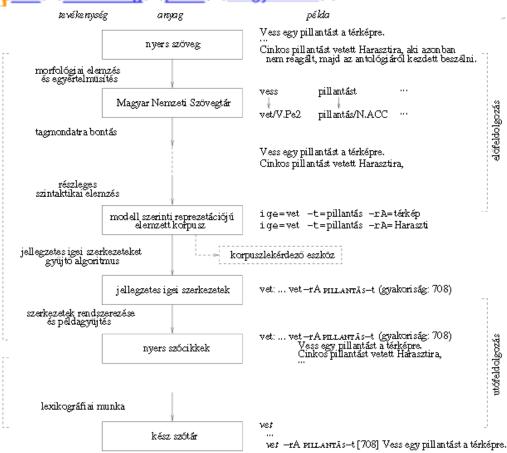
Szótárszerkesztő eszközök (Mazsola)

hang [477] lehetőség [412] válasz [47] otthon [41] tájékoztatás [47] hír [42] ok [43] áldás [43] esély [144] lendület [14] hely [160] pénz [129] magyarázat [87] szám [129] alkalom [161] tanács [86] koncert [71] munka [144] utasítás [01] felvilágosítás [02] hangverseny [45] támogatás [50] kép [65] igaz [62] lőkés [41] százalék [52] segítség [70] engedély (sa) megbízás (sa) forint (sa) információ (sa) felhatalmazás (sa) voks (sa) ízelítő (sa) alap (sa) mód (sa) jel (sa) nyomaték (sz) tér (sa) igaz 🕍 válasz 🖂 tanács 🖂 hang 🔥 tipp 🗞 ami 🖂 lehetőség 🖂 pénz 🖂 magyarázat 📾

hála (67) OK (140) név (100) esély (63) ötlet (64) kegyelem (40) mi (140) hi felmentés (12) több (15) valami (16) fegyver (11) áldás (26) az (13)) maga (16) al parancs (28) megbízás (28) puszi (18) erő (40) hir (35) engedély (33) hit (30) kedve

Mazsola vonzatkereső (Sass 2007)

> Igei szerkezetek gyakorisági szótára (Sass 2011)





PÁZMÁNY PÉTER KATOLIKUS EGYETEM - KIEMELT FELSŐOKTATÁSI INTÉZMÉNY INFORMÁCIÓS TECHNOLÓGIAI ÉS BIONIKAI KAR



Hasonló fogalmak gyűjtése korpuszból, szóbeágyazással

0	PPKE	1	970
1	KRE	0.8159	629
2	BCE	0.8029	1052
3	ÁJK	0.7895	895
4	BMGE	0.7840	276
5	PTE	0.7788	7721
6	JPTE	0.7733	366
7	ÁJTK	0.7728	331
8	BMF	0.7719	762
9	NYME	0.7712	966
10	SZIE	0.7683	2574

0	bionika	1	196
1	sejtbiológia	0.7714	298
2	biofizika	0.7539	501
3	biokémia	0.7464	1993
4	genomika	0.7420	714
5	bioinformatika	0.7371	420
6	rendszerbiológia	0.7292	113
7	mikrobiológia	0.7286	859
8	filogenetika	0.7264	41
9	gyógyszerkémia	0.7132	64
10	nanotechnika	0.7073	166



PÁZMÁNY PÉTER KATOLIKUS EGYETEM - KIEMELT FELSŐOKTATÁSI INTÉZMÉNY INFORMÁCIÓS TECHNOLÓGIAI ÉS BIONIKAI KAR



Szemantikus jegyek automatikus kinyerése korpuszból

"bionika[FN]"	"PPKE[FN]"	"Pázmány[FN]"	"Dzsudzsák[FN]"
bionika	PPKE	Pázmány	Dzsudzsák
atom	educate	@Catholic_Church	team
chemistry	institution	holy	sport
science	science	christian	city_2
method	group_5	book	two_3
knowledge	study	religion	INSTRUMENT_8
HAS_31	student	priest	ball
bionika[FN]	PPKE[FN]	Pázmány[FN]	Dzsudzsák[FN]
CHEMISTRY_22	COLLEGE_12	CHRISTIANITY_5	SPORT_18
BIOLOGY_35	COLLEGE_13	CHRISTIANITY_4	FOOTBALL_5
EDUCATION_10	EDUCATION_9	CHRISTIANITY_16	SPORT_10
MICROBES,_GENETICS,_BIOCHEM_9	COLLEGE_8	CHRISTIANITY_23	THEATRE_12
PHYSICS_1	COLLEGE_11	CHRISTIANITY_7	SPORT_14
HARD_SCIENCE_1	EDUCATION_13	CHRISTIANITY_2	SPORT_7



Kapcsolat más nyelvészeti megoldásokkal



Az "ablakos" kommunikáció nehézségei

- ☐ Kilépni az adott alkalmazásból
- Elindítani
- ☐ Kinyitni vagy felnagyitani
- □ Beírni
- □ Klikkelni
- □ Átmozgatni
- Lekicsinyíteni vagy bezárni
- □ Visszalépni az eredeti alkalmazásba



A gyorsfordítók megjelenése

- ☐ Amikor *információ* kell, pl. szótári, akkor: csak amit kérek, nem többet,
 - de azt gyorsan,
 - kevés aktív művelettel
 - és a lehető legautomatikusabban!
- ☐ Kialakul a "pop-up" viselkedés
- ☐ A kijelölhetőség, ill. az automatikus indíthatóság szerepe

PÁZMÁNY PÉTER KATOLIKUS EGYETEM - KIEMELT FELSŐOKTATÁSI INTÉZMÉNY INFORMÁCIÓS TECHNOLÓGIAI ÉS BIONIKAI KAR

A "rávetítős" megoldás nyelvtechnológiai elemei

- □ Szöveg(rész)-felismerés
- ☐ Nyelvi elemzés: morfológia, lemmák, szókapcsolatok (esetleg környezetelemzés)
- ☐ Szótári keresés: tövesítve vagy csak literálisan
- Megjelenítés: buborékban vagy fix ablakban
- Log-fájlok segítségével az automatikus információgyűjtés lehetősége (nemcsak rávetítős szótáraknál!)
- Megjelennek a felhasználók valódi javaslataival feljavított szótárak



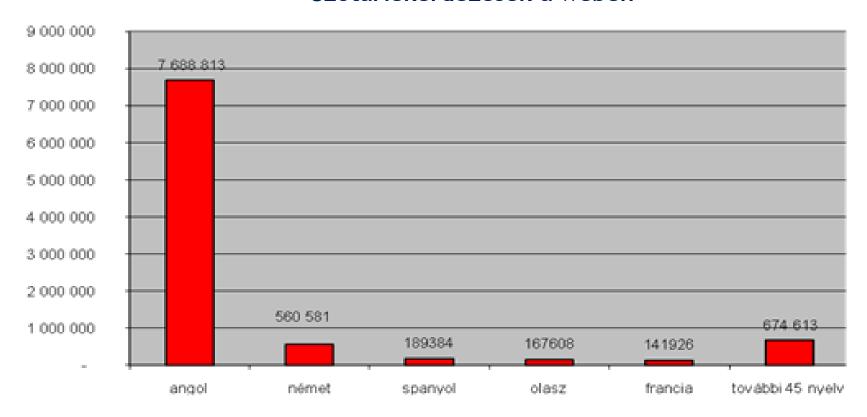
A "rávetítés" a gyakorlatban





A log-fájlok szociolingvisztikailag elemezhetőek

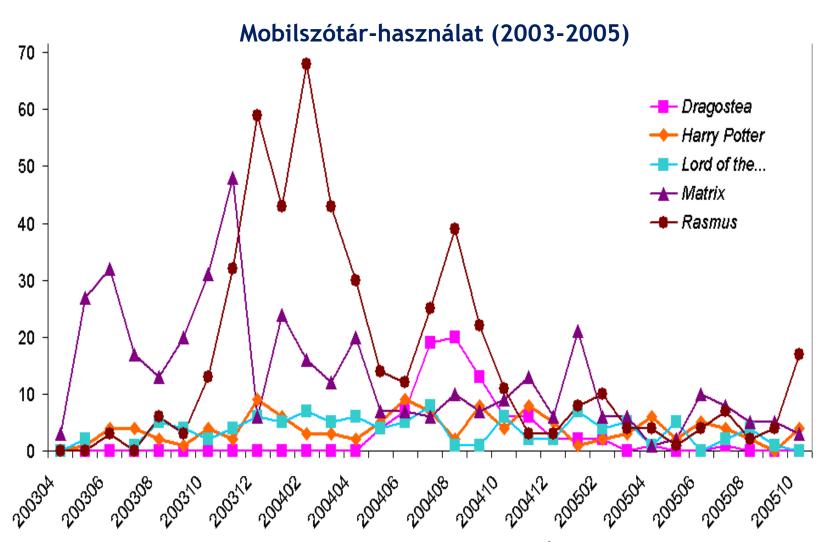
Szótárlekérdezések a weben





PÁZMÁNY PÉTER KATOLIKUS EGYETEM - KIEMELT FELSŐOKTATÁSI INTÉZMÉNY INFORMÁCIÓS TECHNOLÓGIAI ÉS BIONIKAI KAR

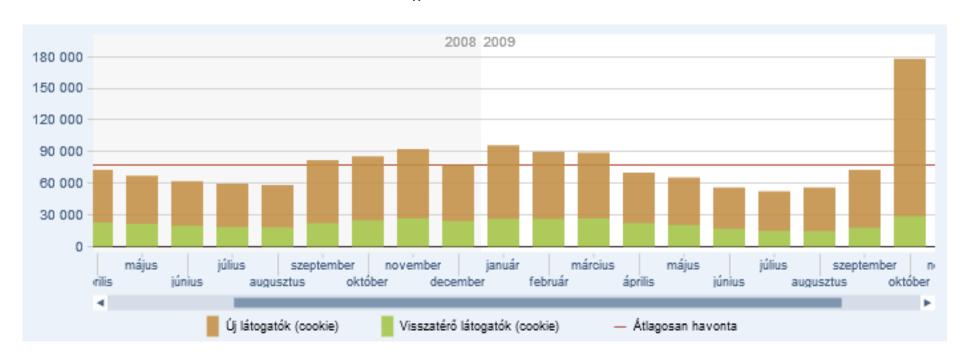
A log-fájlok szociolingvisztikailag elemezhetőek





A log-fájlok szociolingvisztikai elemezhetőek

"A szótárazás tanítása"

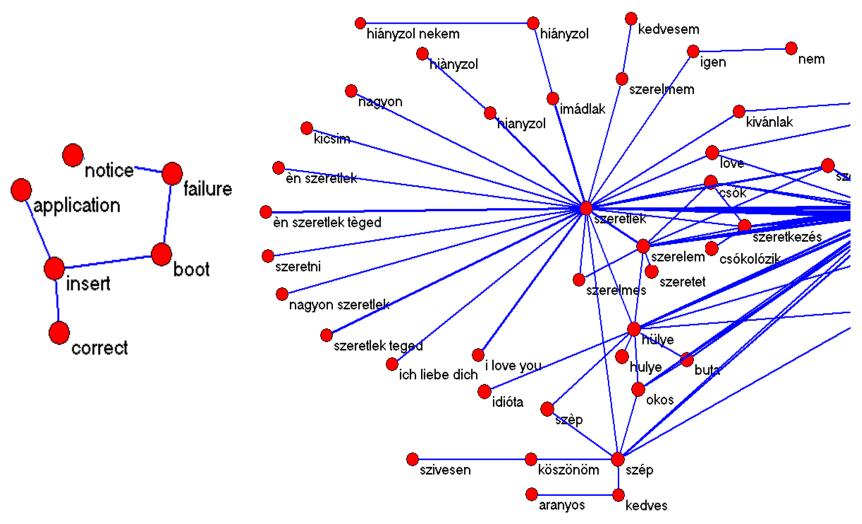




PÁZMÁNY PÉTER KATOLIKUS EGYETEM - KIEMELT PELSŐOKTATÁSI INTÉZMÉNY INFORMÁCIÓS TECHNOLÓGIAI ÉS BIONIKAI KAR

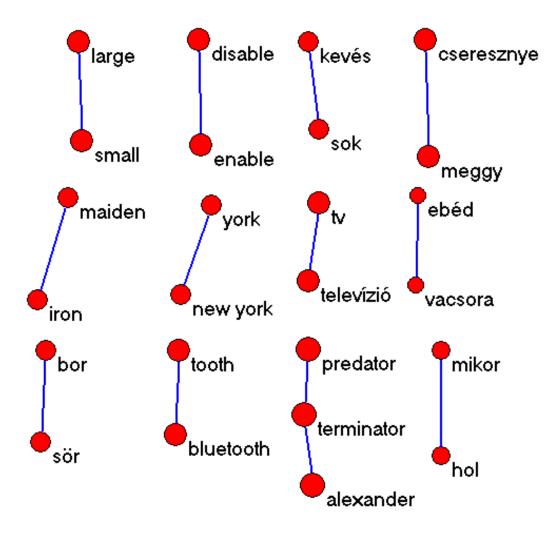
A szótárlekérdezés koincidencia-részgráfjairól

"Aki ezt kérdezte, az még mit kérdezett?"





A szótárlekérdezés koincidencia-részgráfjairól (2)





A hagyományos és az "intelligens" szótárak viszonyáról

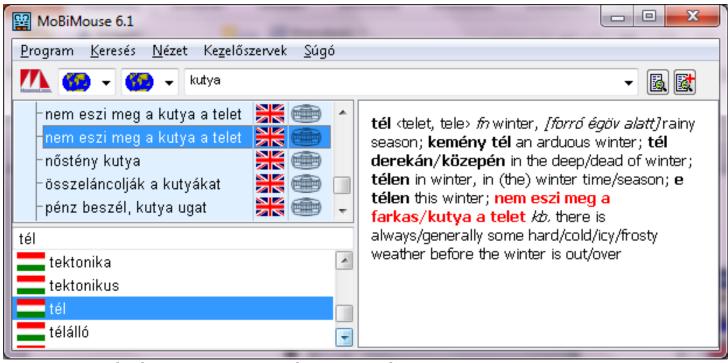


Az "intelligens" szótárak készítésének problémái

- ☐ A legfőbb baj: a szótárforrások XML-változatainak "amatőr" vagy legalábbis nyomtatás-centrikus megoldásai
- □ A második ok: a szótár az embereknek, nem a gépnek készül
- ☐ Egy sor technikai probléma, ami a szótárak "papírszótár" mivoltából ered, ám a gépi változatban át kell ezeket alakítani



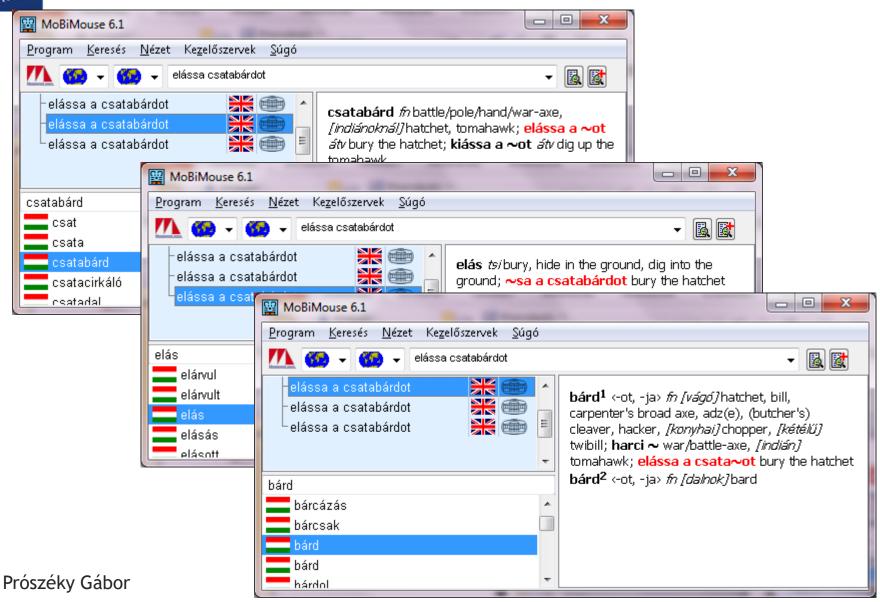
A perjel-probléma



- A szótárak nem a gépnek készülnek (pl. perjelek):
 - nem szavazó/szavazásra nem jogosító részvény
- Tudhatja-e egy program összefogni az egybetartozó kifejezést?
 - {arcára fagy/ráfagy az arcára/lefagy az arcáról} a mosoly

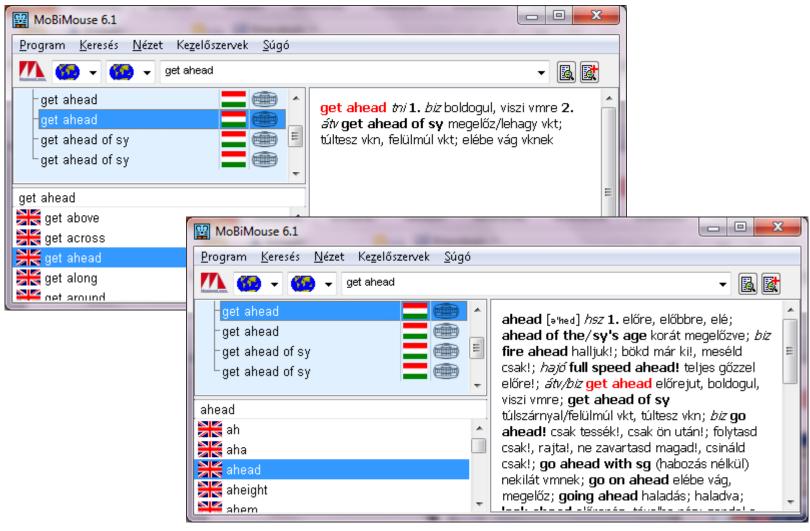


A többszörös előfordulások problémája



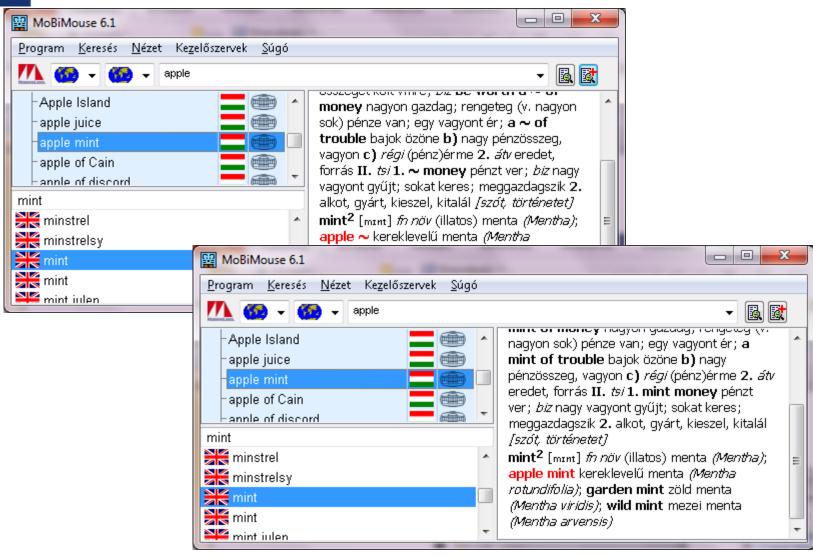


Az ellentmondó előfordulások problémája



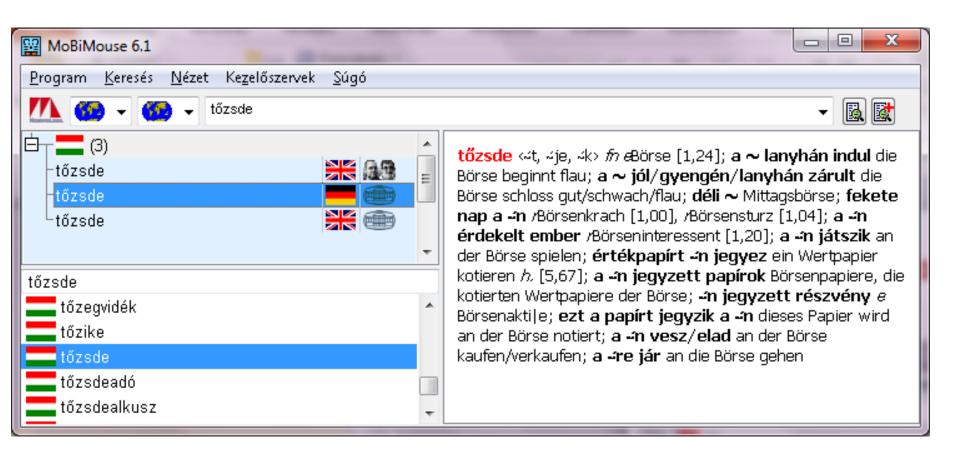


A tilde-probléma



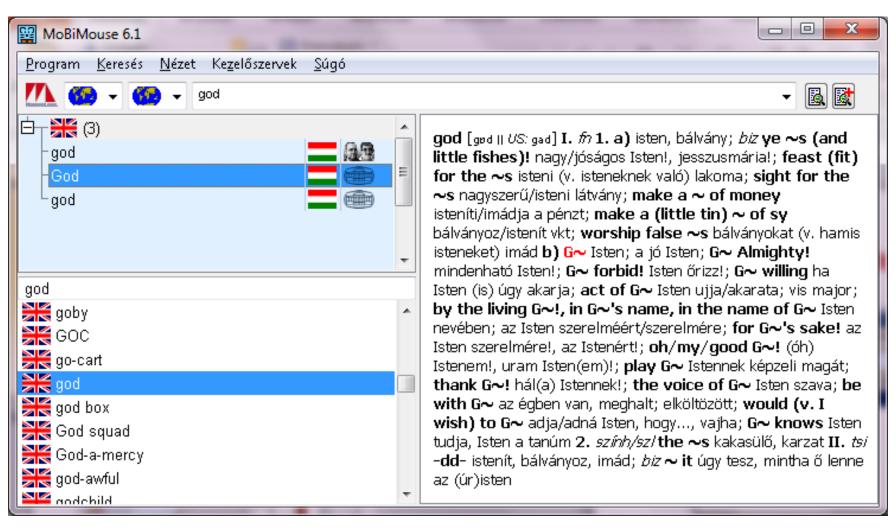


A morfológia-probléma



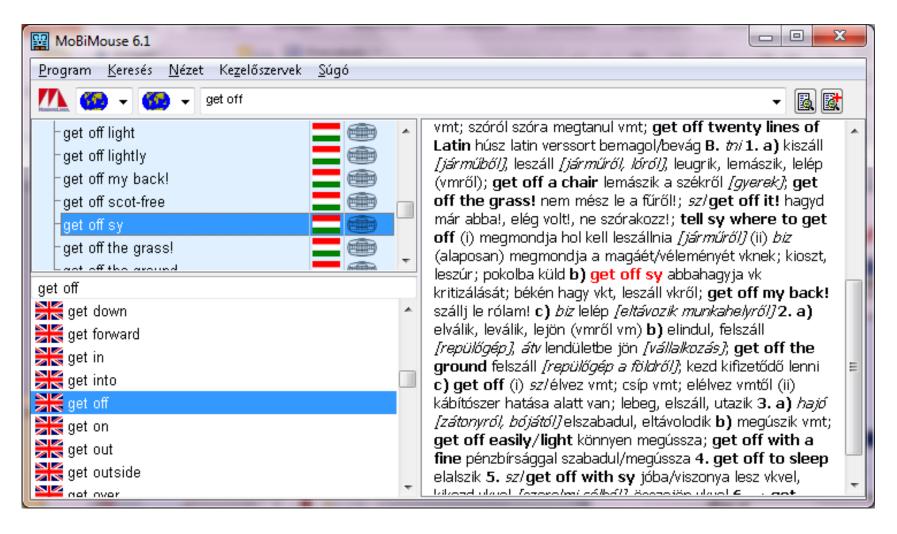


A nagybetű-probléma





A vonzat-probléma



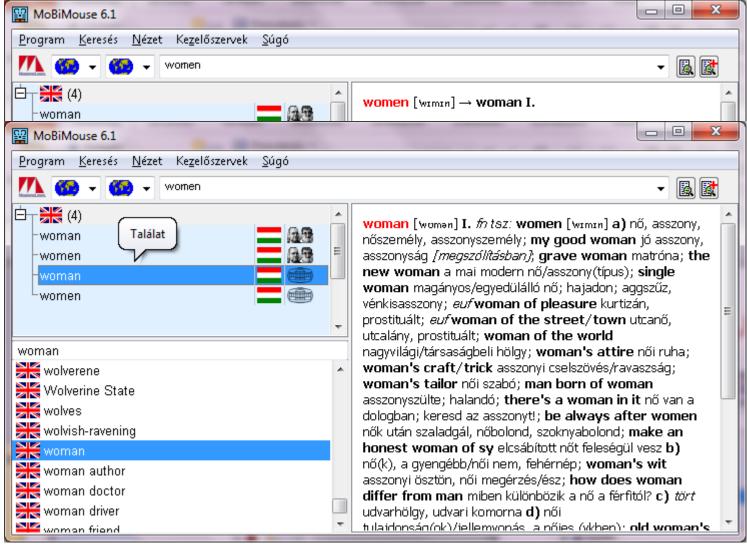


A példa-probléma

/hajó] b) a disease will run its course egy betegségnek megvan a (szabályos) lefolyása; egy betegségnek nem is nagyon lehet megváltoztatni a lefolyását, **things must run their course** a dolgoknak ki kell futniok magukat, szabad folyást kell engedni a dolgoknak **c) run a traffic signal** elnéz (v. nem vesz észre) (közúti) jelzőtáblát **2. a)** (meg)futtat, (meg)szalaszt b) űz, üldöz, hajszol [vadat], versenyt fut (vkvel), vad run a scent nyomon van /kutya]; átv run sy close/hard szorongat/megszorít vkt; nyomában van vknek, komoly versenytársa vknek, megközelít vkt, átv **be hard run** szorongatják, szorongatott/szorult helyzetbe kerül; szutyongatják; *biz* (anyagilag) megszorult, le van égve (anyagilag) 3. futtat [lovat, versenylovat]; biz run a candidate jelöltet futtat 4. a) vezet, irányít, kormányoz, visz [járművet]; run a ship to London elvezeti/elkormányozza a hajót Londonba b) run sy up to town bevisz vkt a városba [autón] 5. üzemben tart, járat, működtet [gépet], járat, fenntart [járművet *bizonyos útvonalon1*; **run a car** kocsit/autót tart **6. a)** vezet ∫*szállót, üzletet1*, irányít ∫*gazdaságot1*, üzemben tart [gyárat], kezel, vezet [ügyeket], igazgat [színházat], ellát, (el)vezet [háztartást]; **run a business** üzletet (v. üzleti vállalkozást) irányít/vezet, *biz* it is he who runs the show ő a fejes/góré; *biz* he is run by his secretary a titkára dirigálja, a titkára intéz helyette mindent b) gazd/biz we are running a cheap line just now that might suit you jelenleg igen jutányosan beszerezhető árutételt tartunk raktáron, ami talán önt is érdekelné 7. a) folyat, ereszt [folyadékot], (ki)önt [formába]; run sy a bath, run a bath for sy fürdőt készít vknek, fürdővizet ereszt/enged vknek; **her eyes run tears** szeme könnyet hullat/ont, szeméből könny patakzik; **the** streams run blood a folyók vértől zavarosak/áradnak b) kiolvaszt, run butter vajat kisüt 8. (meg)húz, megvon [vonalat, határt]; run a parallel too far túl messzire megy az összehasonlításban 9. biz csempész [pl. kábítószert]; run liquor/rum szeszt/alkoholt csempész 10. biz run a temperature láza/hőemelkedése van 11. behúzó öltéssel varr [minden 4–5-ik öltés után húzza ki a tűt [12. a) sp kúgy üti a labdát, hogy leesés után továbbguruljon a lyuk felé> *[golfban]* **b)** *sp* átüt *[labdát kapun, krikettben]* **13.** *biz* **run a rig** kirúg a hámból 14. kat/szl run a man fegyelemsértéssel vádol vkt 15. US run an ad (állandó) hirdetést ad fel (v. tesz közzé) 16. sp versenyez 17. elfuvaroz, elvisz [vkit járművön] B. tni 1. a) fut, szalad, rohan; run home hazaszalad; $| \Delta t v |$ befut a célba; run like a hare, run like blazes/hell, run like the devil fut, mint a nyúl; run to meet sy vk elé(be) fut/szalad; biz run to meet one's troubles keresi a bajt, rules so clear that he who runs may read olyan világos szabályok, hogy azonnal megért(het)i mindenki, **he ran a poor third** gyenge harmadiknak futott be; also ran futottak még; an also ran egy a próbálkozók/résztvevők/futottak közül; set off running nekiiramodik, futni kezd, futásnak ered; *biz c*ut and run elpucol, eliszkol, kereket old b) *biz* kifutóként/küldöncként működik 2. a) fut, megy, halad fjármű], jár fkeréken]; hajó run free hátfélszéllel (v. háromnegyed széllel) megy/halad/vitorlázik b) gördül /vers/ c) let one's ideas run freely gondolatait szabadjára engedi (v. csapongani hagyja) **3. a**) jár, működik, megy, dolgozik, üzemben van [gép, gyár], forog [kerék]; műsz run hot túlhevül, hőn fut [csapágy] b) jár, közlekedik, megy [jármű menetrendszerűen]; a bus runs between the two villages autóbusz jár/közlekedik a két falu között c) how her tongue runs! be nem áll a szája!, hogy pereg a



A "lásd"-probléma





A pontos találatok problémája

