



A NEW MONOLINGUAL, CORPUS-BASED EXPLANATORY DICTIONARY PROJECT

Veronika Lipp, László Simon
Research Institute for Linguistics, Budapest

Encyclopedistics 2020, Zagreb

Background

(1) Explanatory Dictionary of the Hungarian Language (1959-1962)

- ▶ It was compiled in the Research Institute for Linguistics, Hungarian Academy of Sciences in 7 volumes
- ▶ 60.000 entries
- ▶ Each sense unit is illustrated by a few examples: citations from the classic Hungarian literature and example sentences created by the lexicographers
- ▶ The citations came from the 6 million dictionary cards collected mostly from the 19th century
- ▶ This is one of the best used dictionaries, but the vocabulary and the examples are old-fashioned



1879

Bore pure pánusé, szajkón egy alma

Annaly. 22. v. 6. 99

alma fn almát, almája

1. Az almafának kb. gyermekököl nagyságú, gömbölyded, húsos, leves bélű, apró magvú ehető termése. → *Borízű ~*; *édes, savanykás ~*; → *kormos ~*; *nemes, piros ~*; *nyári, téli ~*; *éretlen, érett, rothadt ~*; *~ bundában* ←; *~t eltesz, hámoz, ráz, reszel, szed. Kerek, mint az ~. Szh: olyan képet vág v. olyan az arca, mintha savanyú ~ba harapott volna: siralmasan fanyar, kedvetlen kifejezést ölt az arca. □ Asztal utóján már kedvökre lakoznak; S ím, erdei süldőt, büszke vadat hoznak... Bőre piros páncél, szájában egy alma. AR. Az almafa alá ment s leszakított egy édes almát, mert azt nagyon szerette. MÓ. Piros, nagy alma volt, | a kislány ráhajolt. R A D.*

2. (ritk) *Almafa. Egy sor ~t ültetett a gyümölcsösében.*

3. Az almafa összes termése, hozama. *Lefagyott az ~.*

4. (Mit, irod) *Erisz ~ja: vmely viszályt v. nézeteltérést kirobbantó ok.*

5. (rends. állitm-ként) (*argó*) <Nyomatékos tagadás kifejezésére:> semmi; nincs. *~ az egész. [Van vmi eredmény?] ~.*

Sz: beleharap a savanyú ~ba: kénytelen-kelletlen belefog vmely kellemetlen munkába, kényszeredetten vállal vmi rosszat. K: Nem esik messze az ~ a fájától: a gyermek rendsz. olyan (természetű, jellemű), mint a szülője.

(2) Comprehensive Dictionary of Hungarian (2006-)

- ▶ 110.000 entries
- ▶ XML database
- ▶ available in printed and electronic form also
- ▶ <http://nagyszotar.nytud.hu/index.html>

- ▶ 3 sources:
- ▶ Hungarian Historical Corpus (1772-2010)
- ▶ http://clara.nytud.hu/mtsz/run.cgi/first_form
- ▶ 6 million dictionary cards
- ▶ a huge CD-ROM collection with 400 million words

Advanced query in the Folio View 4.2

alma | almába | almában | almábann | almából | almád | almáé | almáér
| almáert | almáért | almáho | almához | almái | almaié | almáim | almain
| almaink | almáit | almáitól | almáival | almája | almájából | almájához | a
lmáján | almájának | almájára | almáját | almájává | almájokkal | almák |
almakat | almákat | almákban | almákba | almákból | almaként | almák
hoz | almákkal | almákon | almákat | almákról | almákul | almám | almámat
| almán | almának | almánál | almánk | almára | almáról | almát | almától
| almával



Találati mutató

- 7019024 – 1772 – Barcsay Ábrahám: Az Éva almájának eredete s értelme. In Szira Béla (szerk.): Barcsay Ábrahám költeményei. [= Költeményei.] Egyetemi Nyomda, Budapest, 1933. Ritmikus, fikciós.
- ...1772 – Barcsay Ábrahám: Az Éva **almájának** eredete s értelme. In Szira...
- 7019024 – 1772 – Barcsay Ábrahám: Az Éva almájának eredete s értelme. In Szira Béla (szerk.): Barcsay Ábrahám költeményei. [= Költeményei.] Egyetemi Nyomda, Budapest, 1933. Ritmikus, fikciós. \ < bibl> < wdate>1772< /wdate> < author>Barcsay Ábrahám< /author> < id type...
- ...Vajjon mi lehetett Évának **almája**?
- ...Hogyha Ádám uram e titkos **almára**...
- 7019026 – 1772 – Barcsay Ábrahám: Alomfejtés. In Szira Béla (szerk.): Barcsay Ábrahám költeményei. [= Költeményei.] Egyetemi Nyomda, Budapest, 1933. Ritmikus, fikciós. \ < id type="corp" id="7019026" /> < p>...
- ...reá is kapott a tiltott **almára**...
- 7044026 – 1772 – Bessenyei György (1): Delfén. In Az Eszterházi vígasságok. [= Az Eszter-házi vígasságok; Delfén.] Bécs, 1772. Ritmikus, fikciós. \ < id type="corp" id="7044026" /> < p>21< /p> < bibl>...
- ...Az arany **almáér**, melyen vetekednek...
- 7249001 – 1772 – Orczy Lőrinc: Futó gondolat a szabadságról. In Két nagyságos elme. [= Két nagyságos elmének költeményei.] Egyetemi Nyomda, Budapest, 1933. Ritmikus, fikciós. \ < id type="corp" id="7249001" />...
- ...Meg vallják, **almáva** falták fel mas43zlagát...
- 7141004 – 1773 – Horányi Elek (ford.) [eredeti szerző: Nádasdy Ferenc]: In Magyarország királyainak koporsóépülete. [= Magyarország királyainak koporsóépülete.] Landerer, Pozsony–Pest, 1773. Ritmikus, fikciós. \ < id type="corp" id="7141004" />...
- ...az Ors43zág **Almája** a' ves43zekeedés **almája** lett...
- ...és az Ors43zág ezen **almá**-játéknak laptázo-helye...
- 7332009 – 1773 – Szilágyi Sámuel (2) (ford.) [eredeti szerző: Wiegand, Johann]: In Mezei gazdaságra oktató könyvecske. [= A Mezei gazdaságra oktató könyvecske.] Landerer, Pozsony–Pest, 1773. Ritmikus, fikciós. \ < id type="corp" id="7332009" />...
- ...Hűsös Gyümölcson éretnek kváltképen az **Alma** a' Körtvélynek s43ok nemei, Birs **alma**. Füge és egy s43zóval minden...
- 7283005 – 1774 – Révai Miklós (ford.) [eredeti szerző: Révai Miklós]: VI. alagya. In Révai Miklós elegyes versei. [= Elegyes versei.] Landerer, Pozsony–Pest, 1774. Ritmikus, fikciós. \ < id type="corp" id="7283005" />...
- ...Vad vagy, az éretlen **almá** le s43zakas43zod erő4vel...
- 7062001 – 1775 – Csapó József: In Új kert. [= Új füves és virágos magyar kert.] Landerer, Pozsony–Pest, 1792. Aritmikus, tényirodalom, szakny. Mezőg. \ < id type="corp" id="7062001" />...
- ...tudja, mitsoda has43znos a' Farkas **alma** leveleit reá rakni, akár zölden...
- 7062001 – 1775 – Csapó József: In Új kert. [= Új füves és virágos magyar kert.] Landerer, Pozsony–Pest, 1792. Aritmikus, tényirodalom, szakny. Mezőg. \ < id type="corp" id="7062001" />...
- ...fes43s43lilbus folitariis.Egyébként az **Aris43tolochiak** (Farkas-**almák**) nemeihez tartozik ez-is...
- 7062001 – 1775 – Csapó József: In Új kert. [= Új füves és virágos magyar kert.] Landerer, Pozsony–Pest, 1792. Aritmikus, tényirodalom, szakny. Mezőg. \ < id type="corp" id="7062001" />...
- ...Szerelem **almája** Deák: Solanum pomiferum. Német: Liebsapfel...
- 7232028 – 1775 – Molnár János: In Pásztorember. [= Pásztor-ember, avagy A pásztorok tanításáról.] Landerer, Pozsony, 1775. Ritmikus, tényirodalom, szakny. Mezőg. \ < id type="corp" id="7232028" />...
- ...s43zerént támad a' borsóban, lentsében **almábar** 's több a' féliben a...
- 7401002 – 1775 k. – Heraclius, s. Exaltatio S. Crucis. In Alszeghy Zsolt–Szlávik Ferenc (szerk.): Csíksomlyói iskoladramák. [= Csíksomlyói iskoladramák.] Landerer, Pozsony–Pest, 1775. Ritmikus, tényirodalom, szakny. Nyelvt. \ < id type="corp" id="7401002" />...
- ...nincs kezébe Heracliusnak se arany **almája** országunknak, se páltája, se koronája...
- 7384015 – 1776 – Weszprémi István (ford.) [eredeti szerző: Wood, Edward]: In Gazdaember könyvecskéje. [= Gazda ember könyvecskéje.] Landerer, Pozsony–Pest, 1776. Ritmikus, tényirodalom, szakny. Mezőg. \ < id type="corp" id="7384015" />...
- ...gondolom azt, hogy az Ádám **almájának** kes43erúsége, 's mérge még az...
- 7021007 – 1779 – Baróti Szabó Dávid (ford.) [eredeti szerző: Vanie4re, Jacques]: In Paraszti majorság. [= Paraszti majorság.] Landerer, Pozsony–Pest, 1779. Ritmikus, tényirodalom, szakny. Mezőg. \ < id type="corp" id="7021007" />...
- ...A' telelő körtvélyt, **almá** a' s43záráz helyekre...
- 7283025 – 1779 – Révai Miklós: In A magyar nyelv tanításának részeli. [= A magyar nyelv tanításának két részeli. Az igaz kimondás helyébe ne léttés43s43enek, p. o. **alma** nem pedig olma...]
- 7467001 – 1780 e. – Comoedia de artibus. In Bernáth Lajos (szerk.): Protestáns iskoladramák. [= Protestáns iskoladramák.] Landerer, Pozsony–Pest, 1780. Ritmikus, tényirodalom, szakny. Nyelvt. \ < id type="corp" id="7467001" />...
- ...Ennek ajándéka gyakran arany **almá**...
- 7283016 – 1780 – Révai Miklós: In Ortográfia. [= A magyar nyelvnek helyes írása, és kimondása felől4i kettő4s tanúság, avagy Ortografia a nemzeti oskolák számára.] Egyetemi Nyomda, Buda, 1815. Aritmikus, tényirodalom, szakny. Nyelvt. \ < bibl> < wdate>1780< /wdate> < autho...
- ...egybekeverni, és eltserélni; p. o. **alma** magamat újabb, vóna, 's a...
- 7283016 – 1780 – Révai Miklós: In Ortográfia. [= A magyar nyelvnek helyes írása, és kimondása felől4i kettő4s tanúság, avagy Ortografia a nemzeti oskolák számára.] Egyetemi Nyomda, Buda, 1815. Aritmikus, tényirodalom, szakny. Nyelvt. \ < bibl> < wdate>1780< /wdate> < autho...
- ...fellyebb-is megjegyzénk; p. o. **alma** nem olma...
- 7397001 – 1780 – In Bévezetés a számvetésre. [= Bé-vezetés a szám-vetésre. A magyar és hozzá tartozandó tartományok nemzeti iskoláinak számára.] Egyetemi Nyomda, Buda, 1780. Aritmikus, tényirodalom, szakny. Mat. \ < bibl> < wdate>1780< /wdate> < pubtitle>Bévezetés a ...
- ...minden dolog Egység. P. O. **Alma** Egység. Ház Egység...
- 7397001 – 1780 – In Bévezetés a számvetésre. [= Bé-vezetés a szám-vetésre. A magyar és hozzá tartozandó tartományok nemzeti iskoláinak számára.] Egyetemi Nyomda, Buda, 1780. Aritmikus, tényirodalom, szakny. Mat. \ < bibl> < wdate>1780< /wdate> < pubtitle>Bévezetés a ...

Bővített keresés

Szó:

- álmatok
- álmatoknak
- álmától
- álmától
- álmatóredék
- álmatorta
- álmatortája
- álmatortát
- álmátskát
- álmagyó
- almával**
- almával
- almavirág
- almavirágai
- almavirágait
- almavirágok
- almavirágos
- almavirágot
- almavirágról
- almavirágszín
- almavirágszínben
- almaviva
- almav

Rekordok találatokkal - 1179

- alma - 552
- almába - 21
- almában - 15
- almábann - 1
- almából - 2
- almából - 22
- almád - 4
- almáé - 3
- almáér - 1
- almáert - 1
- almáért - 7
- almáho - 1
- almához - 7
- almái - 5
- almáié - 2
- almám - 1
- almám - 1
- almáink - 1
- almáit - 5
- almától - 1
- almával - 1
- almája - 26
- almájából - 2

Keresési kifejezés

| almákat | almákban | almákba | almákból | almákért | almákhoz | almákkal | almákon | almákat | almákról | almáktól | almáku | almám | almámat | almán | almanak | almának | almánál | almánk | almára | almáról | almát | almától | almával

Előző

Következő

[]

 Megjelölt ágak

Mindre

Súgó

OK

Mégsem

A Multilingual Evaluation Dataset for Monolingual Word Sense Alignment

alma				
	noun- az almafa termése	exact	noun- Az almafának kb. gyerme	noun- Az almafának kb. gyermekököl nagyságú, gömbölyded, húsos, leves bélű, apró magvú ehető termése
	noun- vizályt kirobbantó ok	exact	noun- vmely vizályt v. nézet	noun- Almafa
	noun- országalma	NONE		noun- Az almafa összes termése, hozama
	noun- vmilyen növénynek, pl. burgonyának, karalábénak a gumója	NONE		noun- vmely vizályt v. nézeteltérést kirobbantó ok
	noun- a rózsafélék családjába tartozó, gömbölyded, húsos termésű, nemesített gyümölcsfa (<i>Malus domestica</i>), ill. e fajba tartozó növény; almafa	exact	noun- Almafa	noun- <Nyomatékos tagadás kifejezésére:> semmi; nincs
	noun- az ezt a fajt is magában foglaló növénynemzetség (<i>Malus</i>)	NONE		
	noun- az almafa v. az almáskert teljes termése, hozama	broader	noun- Az almafa összes termés	
	noun- nincs, ill. semmi	exact	noun- <Nyomatékos tagadás kif	
	noun- ne, nehogy	NONE		

The new monolingual Hungarian explanatory dictionary

- ▶ The Explanatory Dictionary of the Hungarian Language is a good starting point
- ▶ We are working on converting this dictionary's HTML files into XML format
- ▶ We started to build a corpus with 3-5 billion words from texts from 1960s

First of all, we need a BIG corpus

- ▶ <https://adtplus.arcanum.hu/hu/>
- ▶ cca. 40 billion words
- ▶ we „bought” 10 billion words
- ▶ cleaning the dirty OCR
- ▶ mapping each style register

Hungarian Web 2012 (huTenTen12)

- ▶ <https://www.sketchengine.eu/hutenten-hungarian-corpus/>
- ▶ corpus made by SketchEngine
- ▶ 2.5 billion words
- ▶ data was crawled by the SpiderLing web spider in June 2012
- ▶ POS annotated

- ▶ 500 million words of manually downloaded text
- ▶ cca. 6000 files

filename	date	autho	auth	fiction	műfaj	ifjúság	reg1	reg2	reg3	words	qualit
Abraham, Cyril - Az Onedin család 2	1997			f	regény		családrege	történelmi re	filmadaptáció	46300	
Abraham, Cyril - Az Onedin család 3	1997			f	regény		családrege	történelmi re	filmadaptáció	39900	
Bagdy Emőke et al. - Fejben dől el?	2017			nf			pszichológia	önsegítő	ismeretterjesz	22500	
Balogh Béla - A tudatalatti tízparancsolata	2005			nf			ezotéria	életmód	meditáció	17500	javíta
Brown, Dale - A bombázó	1990	Vándor	Dale E	f	regény		kalandregé	techno-thrille	repülő	107391	
Corey, James S. A. - Cibola meghódítása	2015	Galamb	Zoltár	f	regény		sci-fi	horror	tévéorozat-a	147800	
Kárpáti György, Peterdi Pál - Cini és a többi	1985			nf			sport	humoros	irodalomtörté	41000	
Laurie, Hugh - A balek	2019	Ördögh	Hugh	f	regény		kalandregé	humoros	krimi	94611	
Lennox, John C. - Isten és Stephen Hawking	2016	Visky S.	Béla	nf	esszé		vallás	tudományfilo	ismeretterjesz	18600	
Lőrincz L. László - Kegyetlen csillagok	1992			f	regény		kalandregé	történelem	sci-fi	85700	
Lőrincz L. László (Leslie L. Lawrence) - A ló	2008			f	regény		kalandregé	Kelet	Tibet	60000	
Lőrincz L. László (Leslie L. Lawrence) - A ló	2008			f	regény		kalandregé	Kelet	Tibet	61000	
Lőrincz L. László (Leslie L. Lawrence) - Kuk	2015			f	regény		kalandregé	Mexikó	krimi	62400	
Lőrincz L. László (Leslie L. Lawrence) - Kuk	2015			f	regény		kalandregé	Mexikó	krimi	54500	
Moldova György - A félelem kapuja	1992			f	regény		kortárs próza	politikai regé	korrajz	64721	
Nemere István - Cirkáló	1992			f	regény		kalandregé	politika	háborús	44600	
Robertson D. C., Breen B. - LEGO. Kockárc	2014	Németh	Ádám	nf			cégtörténet	marketing	közgazdaság	82000	
Róna-Tas András - Nomádok nyomában	1961			nf			útleírás	kultúrantropc	folklór	84300	
Rólfredrik Emma (Alaska Diamond) - De ha	2015			nf			életmód	vallás	regény	28400	

Summary

- ▶ 1. we have the headword list
- ▶ 2. corpus building
- ▶ 3. new corpus in SketchEngine – analysis
- ▶ 4. based on the analysed data and the frequency we choose the examples for illustrating senses
- ▶ 5. future: we are planning to upload entries to Lexonomy

Thank you for your attention!