

# A Microsoft Office és az OpenOffice.org magyar nyelve\*

Németh László ([nemeth@openoffice.org](mailto:nemeth@openoffice.org))

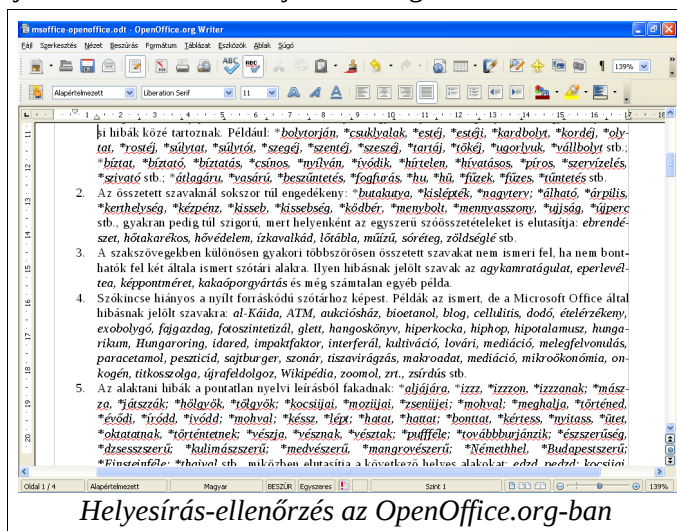
2009. október 6.

Ahogy több független (az egyik esetben Mártonfi Attila, az MTA Nyelvtudományi Intézet munkatársa, a mérvadó, Osiris Kiadónál megjelent Helyesírás című kötet társszerzője segítségével készült) vizsgálat korábban megmutatta, a magyar szavak helyesírás-ellenőrzésében a Microsoft Office elmarad az ingyenes, nyílt forráskódú OpenOffice.org-tól.<sup>1,2</sup> Az elmúlt évek fejlesztéseinek köszönhetően a többi nyílt forráskódú nyelvi eszköz (elválasztóprogram, nyelvhelyesség-ellenőrző, szinonimaszótár) is felzárkózott és számos pontban felül is múlja zárt vetélytársát.

## Helyesírás-ellenőrző

A leggyakrabban használt nyelvi segédeszköz a szavak szintjén működő helyesírás-ellenőrző. A nyílt forráskódú magyar helyesírási szótár<sup>3</sup> dokumentációja külön fejezetben sorolja fel a Microsoft Office és a nyílt forráskódú alternatívája közötti különbségeket, sok súlyos hibát feltárva a Microsoft Office magyar helyesírás-ellenőrzőjében.<sup>4</sup> 2003-ban készült az összehasonlítás, az azóta javított hibák (mint a csak hibásan elfogadott \*csevely cseréje a helyes csevej alakra, vagy a \*kultúrált és \*analfabétizmus tiltása) ellenére a Microsoft Office elmarad a nyílt forráskódú megoldástól. A következő felsorolás a Microsoft Office helyesírás-ellenőrzőjének, a Helyes-e?-nek ismert hibáiból azokat mutatja be, amelyeket a gyártó MorphoLogic Kft. a saját szoftverének Webforditas.hu oldalon kipróbálható újabb változatában sem javított még:

1. Elfogad számos ly-j; i-í és ü-ű tévesztést, amelyek hagyományosan a legsúlyosabb magyar helyesírási hibák közé tartoznak. Például: \*bolytorján, \*csuklyalak, \*estéj, \*estéji, \*kardbolyt, \*kordéj, \*olytat, \*rostéj, \*súlytat, \*súlytót, \*szegéj, \*szentéj, \*szeszéj, \*tartáj, \*tökéj, \*ugorlyuk, \*vállbolyt stb.; \*bíztat, \*bíztató, \*bíztatás, \*csínos, \*nyilván, \*ívódik, \*hirtelen, \*hívatásos, \*píros, \*szervizelés, \*szivató stb.; \*átlagáru, \*vasáru, \*beszünntetés, \*fogfurás, \*hu, \*hü, \*fúzek, \*fúzes, \*tűntetés stb.



2. Az összetett szavaknál sokszor túl engedékeny: \*butakutya, \*kislépték, \*nagyterv; \*álható, \*árpilis, \*kerthelység, \*kézpenz, \*kisseb, \*kissebség, \*ködbér, \*menybolt, \*mennyasszony, \*ujjság, \*újperc stb., gyakran pedig túl szigorú, mert helyenként az egyszerű szóösszetételeket is elutasítja: *ebrendészet, hőtakarékos, hővédelem, ízkavalkád, lőtábla, műízű, sóréteg, zöldséglé* stb.

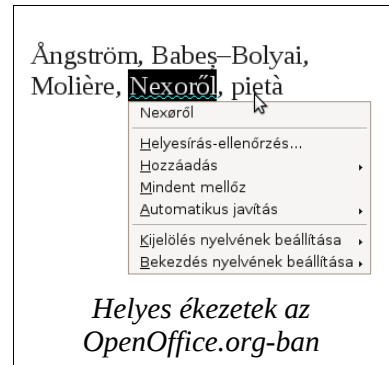
3. A szakszövegekben különösen gyakori többszörösen összetett szavakat nem ismeri fel, ha nem bonthatók fel két általa ismert szótári alakra. Ilyen hibásnak jelölt szavak az *agykamratárgulat, eperlevéltea, képpontméret, kakaóporgyártás* és még számtalan egyéb példa.

4. Szókincse hiányos a nyílt forráskódú szótárhoz képest. Példák az ismert, de a Microsoft Office által hibásnak jelölt szavakra: *al-Káida, ATM, aukciósház, bioetanol, blog, cellulitis, dodó, ételérzékeny, exobolygó, fajgazdag, fotoszintetizál, glett, hangoskönyv, hiperkocka, hiphop, hipotalamusz, hungarikum, Hungaroring, idared, impaktfaktor, interferál, kultiváció, lovári, mediáció, melegfelvonulás, paracetamol, peszticid, sajtburger, szonár, tiszavirágzás, makroadat, mediáció, mikroökonómia, onkogén, titkoszolga, újrafeldolgoz, Wikipédia, zoomol, zrt., zsírdús* stb.

5. Az alaktani hibák a pontatlan nyelvi leírásból fakadnak: \*aljájára, \*izzz, \*izzzon, \*izzzanak; \*mászza, \*játszzák; \*hölgyök, \*tölgök; \*kocsiijai, \*moziijai, \*zseniijei; \*mohval; \*meghalja, \*történéd, \*évódi, \*íródd, \*ívódd; \*mohval; \*késsz, \*lépt; \*hatat, \*hattat; \*bonttat, \*kértess, \*nyitass, \*ütet, \*oktatatnak, \*történtetnek; \*vészja, \*vésznak, \*vésztak; \*pufféle; \*továbbbujjánzik; \*ézszerúség, \*dzsesszszerű; \*kulimásszerű; \*medvészerű, \*mangrovészerű; \*Némethhel, \*Budapestszerű; \*Einsteinféle; \*thaival stb., miközben elutasítja a következő helyes alakokat: *edzd, pedzd; kocsijai, moziyai, zseniyei; Béláké, unokámék; kulimásszerű; New York-iat; linuxos, windowsos; thajjal* stb.

\*Ez a dokumentum szabadon másolható, terjeszthető a Creative Commons Attribution-NoDerivs licenc alapján.

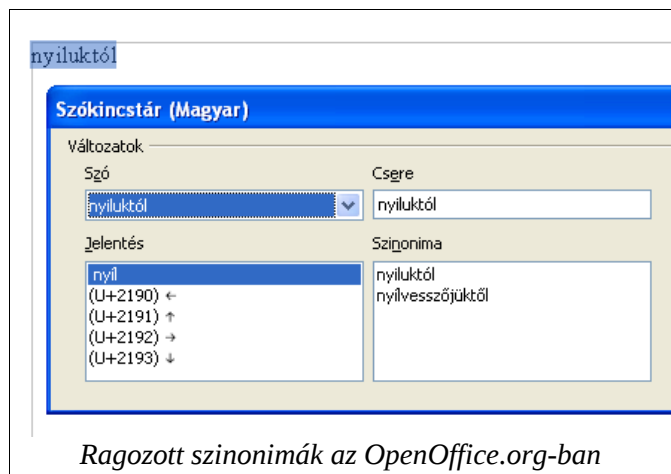
6. Egyes szavak hibás alakban is szerepelnek a szótárban: \**aszkrétikus* (*aszketikus*), \**gargalizál* (*gargarizál*), \**trikolor* (*trikolór*), \**szférikus* (*szferikus*) stb., de van, ami csak hibásan: \**adresz* (*adrezz*), \**bájtos* (*byte-os*), \**entuziazmus* (*entuziazmus*), \**libling* (*liebling*), \**Lódz* (*Lódz*), \**mancsaft* (*mannschaft*), \**szanszevéria* (*szanszeviéria*), \**szkunsz* (*szkunk*), \**Skoda* (*Škoda*), \**zloty* (*złoty*) stb. Példák még az önálló alakban nem létező, elavult vagy értelmetlen szavakra, amiket a Microsoft Office helyesnek ítél: *ál*, *bel*, *fesz*, *pót*, *süv*, *tel*, *veé*.
7. A jól szélesebb lehetőségeket nyújtó Unicode karakterkódolás kezelése teljes mértékben hiányzik a Microsoft Office helyesírás-ellenőrzőjéből, ezért sem ismeri a következő szavakat: *Ångström*, *Babeş-Bolyai*, *Camões*, *Chişinău*, *Ceauşescu*, *crô-magnoni*, *curaçao*, *déjà vu*, görög ábécé ( $\alpha$ -ról,  $\beta$ -sugárzás,  $\varepsilon$ -nal,  $\mu$ -ról stb.), *El Niño*, *Molière*, *Nexø*, *œuvre*, *pietà*, *São Paulo*, *voilà* stb. (A helyesírási szabályzat előírja az idegen ékezetek és latin betűk használatát, ha erre lehetőség van, mint a mai szövegszerkesztőkben.)
8. Hibára adott javaslatai rendkívül hiányosak, rosszak vagy hiányoznak. Ezzel szemben az OpenOffice.org kifinomult javaslattévő képességével nemcsak a hasonló szavakra ad javaslatot (pl. *előről* → *előlről*, *Csautseszku* → *Ceauşescu*, *arvitzturo* → *árvíztűrő*), hanem az eltérőbbekre is (*file-t* → *fájlt*, *sármos* → *charme-os*, *csivava* → *chihuahua*).



Az OpenOffice.org magyar helyesírási szótárából 16 új kiadás jelent meg a legutóbbi, 2003-as összehasonlítás óta. A folyamatosan megújuló szókincs jól tükrözi a magyar köznyelv változásait. Például a *zrt.*, *nyrt.* rövidítések már a megjelenésüket követően bekerültek a nyílt szótár következő kiadásába. A Microsoft Office-nál felsorolt hiányosságok nem vagy csak sokkal kisebb mértékben jellemzik. Ilyen hibák például, hogy néhol kötőjel nélkül is elfogadja a többszörösen összetett, hat szótagnál hosszabb szavakat (pl. *olajár-emelkedés*), helyes összetételeket jelezhet hibásnak, mint például régebben a *ligatúrahaszálat*, *notebooküzletág* szavakat (ez utóbbi kettő a Microsoft Office-ban még hibásan van kezelve).

## Szinonimaszótár

Az OpenOffice.org szinonimaszótára mellett, hogy 20 ezer szavas szókincse összemérhető a Microsoft Office-éval, számos újdonságot is tartalmaz. Ilyen a tövezés és toldalékolás, az Unicode karakterkészlet támogatása, és több száz Unicode karakter felvétele is. A képen látható példában a „nyiluktól” szóra kattintás és a szinonimaszótár előhívása után megjelenik a *nyíl* tőszó, és a *nyílvesszőjüktől* szinonima, ami a keresett szóalaknak megfelelően van toldalékolva. Láthatóak még az Unicode nyíl karakterek is, amelyek így nemcsak a speciális karakterek beillesztése párbeszédablak, hanem a szinonimaszótáron keresztül is elérhetővé váltak. Nemcsak a nyilak, hanem a görög ábécétől kezdve matematikai és sok egyéb szimbólum is kereshető név alapján az OpenOffice.org szókincstárban.

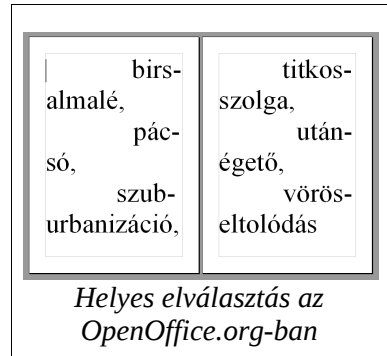


## Elválasztás

Az elválasztás lehetővé teszi a szép, egyenletes szedést sorkizárás esetén, amellyel, hogy takarékos nyomtatást eredményez. A Microsoft Office XP, valamint a MorphoLogic parancssori helyesírás-ellenőrző és elválasztóprogramjának 2005-ös kiadása nem választja el azokat a szavakat, amelyeket nem ismer fel a helyesírás-ellenőrző. Mivel ez gyakori a Microsoft Office-nál, ráadásul éppen a hosszú összetett szavaknál (lásd a szóellenőrzésről szóló szakaszt), a program nem alkalmas a teljesen automatikus szedésre: ellenőrizni kell a szöveget és kézzel javítani a szedési hibákat, csak így lehet eltüntetni a szövegből az elfogadhatatlanul nagy szóközöket.

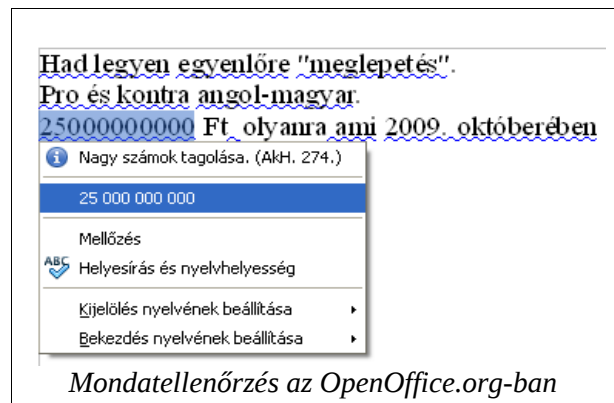
A Webforditas.hu oldalon kipróbálható MorphoLogic elválasztóprogram már megpróbálkozik az ismeretlen szavak elválasztásával is, de az OpenOffice.org-énál lényegesen egyszerűbb és így pontatlanabb megoldással. Például a következő hibákat véti a mellékelt szavak elválasztásánál: \*bir-salmalé, \*pá-csó, \*szu-burbanizáció, \*titkosz-szolga, \*utá-négető, \*vörö-seltolódás. Ebből a szuburbanizáció a Helyes-e? helyesírás-ellenőrző által felismert szó, tehát ez sem garancia a helyes elválasztásra.

További különbség, hogy az OpenOffice.org elválasztóprogramja támogatja az Unicode karakterkódolást. Az OpenOffice.org magyar elválasztási szótárának unicode-os kiegészítése helyesen választja el a nyílt forráskódú helyesírási szótárban szereplő idegen ékezetes szavakat.



## Mondatszintű helyesírás-ellenőrzés

A szóellenőrzőn lép túl a mondatellenőrző (nyelvhelyesség-ellenőrző), aminek egyik leghasznosabb tulajdonsága, hogy a mondat szinten jelentkező elütések, hiányzó vagy felesleges szóközpontok és írásjelek, valamint tipikus írásjel- és szótévesztések nem kerülnek el a felhasználó figyelmét. 2009-ben, pár hónappal azután, hogy az OpenOffice.org irodai csomagban megjelent a nyelvhelyesség-ellenőrzés támogatása, elkészült a magyar nyelv helyesség-ellenőrző első változata is, több száz szabállyal az említett hibák kiszűrésére. A Microsoft Office nyelv helyesség-ellenőrzőjével ellentétben alapértelmezett beállításként minél kevesebb téves hibajelzést ad, hogy a felhasználóknak eszükbe se jusson kikapcsolni. Bár még nem tartalmaz minden lehetőséget, amit a Microsoft Office nyelv helyesség-ellenőrzője, a következő változatokban helyet kap egy igazi újdonság: a kényelmes, de sok hibát okozó összetett szó-felismerő algoritmus több szintes kikapcsolási lehetősége, amivel a korrektorok az elütéseket vagy az optikai karakterfelismerő programok hibáit sokkal könnyebben fogják tudni javítani.



## Számok gépi fordítása

Az OpenOffice.org-hoz kapcsolódó nyelvtechnológiai újdonságok jó példája a magyar fejlesztésű NUMBERTEXT OpenOffice.org-kiegészítés. A számok automatikus számnévé alakítása megkönnyíti a szerződések, számlák, banki dokumentumok elkészítését. Míg a Microsoft Office csak thai nyelven képes a számokat számnévre és thai pénznemre, bátra és szatangra átalakítani az azóta már nemzetközi szabványként is elkönyvelt BAHT-TEXT függvénnyel, addig a NUMBERTEXT és MONEYTEXT függvények ezt tetszőleges (jelenleg mintegy 30) nyelvre és nyelvenként több pénznemre megteszik, így magyarra és forintra is. A fejlesztés során kidolgozott új programnyelv- és függvényspecifikáció a megfelelő OASIS munkacsoporthoz is benyújtásra került, hogy a nemzetközi ISO/OASIS OpenDocument dokumentumformátum-szabvány részévé váljon.<sup>5</sup>

	A	B	C	D	E	F	G
1	2009	kétezer-kilenc					
2	1232424	egymillió-kétszázharminckétezer-négyszázhuszonnégy					
3	-5535,34	mínusz ötezer-ötszázharmincöt egész harmincnégy század					
4	120,2	százhusz forint húsz fillér					
5							
6	2009	two thousand and nine					
7	1232424	one million, two hundred and thirty-two thousand, four hundred and twenty-four					
8	-5535,34	negative five thousand five hundred and thirty-five point three four					
9	120,2	one hundred and twenty euro and twenty cents					
10							
11	2009	две тысячи девять					
12	1232424	один миллион двести тридцать две тысячи четыреста двадцать четыре					
13	-5535,34	минус пять тысяч пятьсот тридцать пять целых и тридцать четыре сотых					
14	120,2	сто двадцать целых и две десятый рубля					
15							

Számok magyar, angol és orosz szöveggé alakítva az OpenOffice.org Calc táblázatkezelőjében

## Honosítás

Bár a nyelvtechnológia fontos része az irodai csomagoknak, a használhatóság mögött azért jóval többnek kell állnia. Egy nem hazai fejlesztésű programnál különösen fontos szerepe van a honosításnak, ami a program felületének, beépített dokumentációjának és egyéb magyar nyelvre átültetett funkcióinak összessége. A nyelvtechnológián felül kiemelhető, hogy az OpenOffice.org a Microsoft Office-hoz hasonlóan kiváló magyar nyelvű felülettel és sűgóval rendelkezik. Pár különbséget érdemes megemlíteni: a rövid hónapnevet a dátumformátumokban (ami a Microsoft Office 2003-ban hibásan, két

	A	B	C	D
1	2009-01-01	2009-07	2009. I. 1.	VII. 1.
2	2009-02-07	2009-08	2009. II. 7.	VIII. 7.
3	2009-03-12	2009-09	2009. III. 12.	IX. 12.
4	2009-04-19	2009-10	2009. IV. 19.	X. 19.
5	2009-05-23	2009-11	2009. V. 23.	XI. 23.
6	2009-06-30	2009-12	2009. VI. 30.	XII. 31.
7				

*Az OpenOffice.org helyes dátumformátumai*

ponttal jelenik meg, pl. „szept.”) az OpenOffice.org-ban a jóval gyakrabban (pl. jogi és történelmi anyagokban) használt római számozás helyettesíti: 2009. XII. 25-e. Ennél nagyobb különbség, hogy a Microsoft Excel magyar helyesírás által nem ismert, szóköz és záró pont nélkül írt dátumformátuma (pl. 2009.09.18) helyett az OpenOffice.org az MSZ ISO 8601 szabvány, egyben a magyar helyesírás szabályzatban is szereplő kötőjeles (pl. 2009-09-18) dátumformátumot használja alapértelmezésként a táblázatkezelőjében.

## Más nyelvek

Gyakran fordulnak elő idegen szavak, idézetek a dokumentumokban. A Microsoft Office magyar változata nem tartalmaz angol, német stb. helyesírás-ellenőrzőt, elválasztót és egyéb nyelvtechnológiát, ezeket külön kell megvásárolni. A magyar OpenOffice.org több világnyelv helyesírási, elválasztási és szinonimaszótárával érkezik, és az OpenOffice.org honlapjáról még további száz nyelvhez érhető el javarészt ingyenes szótárak. Az OpenOffice.org-ban újnak számító nyelvhelyesség-ellenőrzéshez is több ingyenes megoldás közül választhatunk, de megvásárolható például a német nyelvhez készült jó nevű Duden Korrektor is.

## Elterjedtség

Az OpenOffice.org magyar nyelvtechnológiai fejlesztései 2008-ban elnyerték az OpenOffice.org közösségi innovációs díját. A legnagyobb elismerés azonban, hogy a magyar fejlesztések helyet kaptak az OpenOffice.org-ban és a leginnovatívabb informatikai termékekben: a böngészők közül a Mozilla Firefox, az Opera és a Google Chrome döntött a Hunspell helyesírás-ellenőrző használata mellett, a fordítómemóriáknál a piacvezető SDL Trados és a feltörekvő magyar MemoQ, az operációs rendszereknél az Apple Mac OS X (a legfrissebb Snow Leopard változattól), a különböző Linux terjesztések és az OpenSolaris.<sup>6</sup> A Hunspell tette lehetővé nemcsak a magyar, hanem az arab, baszk, kopt, koreai, nepáli stb. nyílt forráskódú helyesírási szótárak elkészítését és más egybeíró (nagy számú összetett szót használó), Unicode karakterkódolást igénylő vagy bonyolult alaktannal rendelkező nyelvek kezelését.

## Összefoglalás

Az irodai csomagok nyelvi eszközei nemcsak az ismert hibák, hanem a nyelv változásai miatt is rendszeres frissítést igényelnek, ahol a szabad programok gyártófüggetlensége határozott előny. Mártonfi Attila megállapítása ma is érvényes: a jelek szerint a közeljövőben érdemi fejlődésre nem lehet számítani a Microsoft Office magyar nyelvi támogatásában. A szabad magyar nyelvi eszközöket ezzel szemben folyamatos fejlesztés és innováció jellemzi, az új változatok pedig késedelem nélkül helyet kapnak az OpenOffice.org-ban, Mozilla Firefoxban és más szabad programokban. A fejlesztések legfőbb támogatója a hazai és nemzetközi szabad szoftveres közösség, jelenleg a magyar FSF.hu Alapítványon keresztül.<sup>7</sup>

1. [Szabad a helyesírás](#), Index, 2005-05-17.

2. [Szövegszerkesztők helyesírásversenye](#), Origo, 2008-09-23.

3. [Szabad magyar szótár](#).

4. [Magyar Ispell dokumentáció](#), 7. fejezet, 2003-05-12.

5. [NUMBERTEXT.org](#).

6. [Hunspell szócikk](#), Wikipédia.

7. [FSF.hu Alapítvány a szabad szoftverek magyarországi népszerűsítéséért és honosításáért](#).