

# Az OpenOffice.org magyar nyelvi támogatása

## Védjegyek és Jogi nyilatkozat

A Novell, és termékei a Novell, Inc. bejegyzett védjegyei az Egyesült Államokban és más országokban. A bejegyzett védjegyek teljes listája a Novell weboldalán található: <http://www.novell.com/company/legal/trademarks/tmlist.html>.

A Linux Linus Torvalds bejegyzett védjegye. Az egyéb védjegyek a birtokos cégek tulajdonát képezik.

A jelen dokumentáció Németh László írásából és engedélyével készült, OpenOffice.org Novell Edition programmal.

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

Novell Professional Services Hungary  
1124 Budapest, Csörsz u. 45.  
Tel.: +36 1 4894600 Fax.: +36 1 4894601

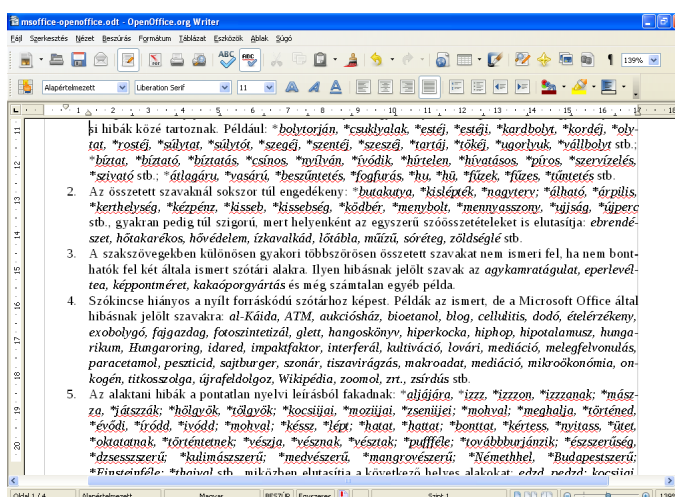
## Bevezető

Ahogy több független<sup>1</sup> vizsgálat korábban megmutatta, a magyar szavak helyesírás-ellenőrzésében a Microsoft Office elmarad az ingyenes, nyílt forráskódú OpenOffice.org-tól<sup>2,3</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>4</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>5</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 2007 helyesírás-ellenőrzőjének, a Helyes-e?-nek ismert hibáiból azokat mutatja be, amelyeket a fejlesztő MorphoLogic Kft. a saját szoftverének *webforditas.hu* oldalon kipróbálható újabb változatában sem javított még:

1. Elfogad számos ly-j, i-í, u-ú és ü-ű tévesztést, amelyek hagyományosan a legsúlyosabb magyar helyesírási hibák közé tartoznak. Például: \*bolytorján, \*csuklyalak, \*estéj, \*estéji, \*kardbolyt, \*kordéj, \*rostéj, \*súlytat, \*súlytát, \*szegéj, \*szentéj, \*szeszéj, \*tartáj, \*tökéj, \*ugorlyuk, \*vállbolyt stb.; \*bíztat, \*bíztató, \*bíztatás, \*csínos, \*nyílván, \*ívódiok, \*hírtelen, \*hívatásos, \*píros, \*szervizelés, \*szivató stb.; \*átlagáru, \*vasáru, \*beszúntetés, \*fogfúrás, \*hu, \*hü, \*fúzek, \*fúzes, \*túntetés stb.



1. ábra: Helyesírás-ellenőrzés az OpenOffice.org-ban

2. Az összetett szavaknál sokszor túl engedékeny: \*butakutya, \*kislépték, \*nagyterv; \*álható, \*árpilis, \*kerthelység, \*kézpénz, \*kisseb, \*kissebség, \*kődbér, \*menybolt, \*mennyasszony, \*ujjság, \*újjperc 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. Szókincese hiányos a nyílt forráskódú szótárhoz képest. Példák az ismert, de a Microsoft Office által hibásnak jelölt szavakra: *al-Káida, aukciósház, bioetanol, blog, cellulitis, dodó, ételérzékeny, exobolygó, fajgazdag, fotoszintetizál, glett, hangoskönyv, hiperkocka, hip-hop, hipotalamusz, hungarikum, Hungaroring, idared, impaktfaktor, interferál, kultiváció, lovári, mediáció, melegfelvonulás, paracetamol, peszticid, sajtburger, szonár, tiszavirágzás, makroadat, mediáció, mikroökönómia, onkogén, titkosszolga, újrafeldolgoz, Wikipédia, zoomol, zrt., zsírdús* stb.

1 Az egyik, Mártonfi Attila, az MTA Nyelvtudományi Intézet munkatársa, aki az Osiris Kiadónál megjelent Helyesírás című kötet társszerzője segítségével készült

2 <http://index.hu/tech/szoftver/ispell/>

3 <http://www.origo.hu/techbazis/szamitogep/20080923-megvizsgaltuk-a-helyesirasellenorzoket-microsoft-office-vs-openoffice.html>

4 <http://magyarispell.sourceforge.net/>

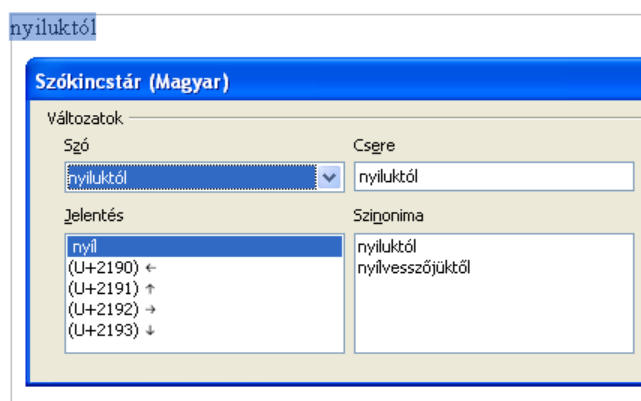
5 Magyar Ispell dokumentáció 7. fejezet

5. Az alaktani hibák a pontatlan nyelvi leírásból fakadnak: \*aljáára, \*izzz, \*izzzon, \*izzzanak; \*mászza, \*játszzák; \*hölgyök, \*tölgyök; \*kocsijai, \*mozijai, \*mohval; \*meghalja, \*történed, \*évődi, \*íródd, \*ivódd; \*mohval; \*késsz, \*lépt; \*hatat, \*hattat; \*bonttat, \*kértess, \*nyitass, \*ütet, \*oktatatnak, \*történtetnek; \*vészja, \*vésznak, \*vésztak; \*puffféle; \*továbbburjánzik; \*észszerűség, \*dzsesszszerű; \*kulimászszerű; \*medvészerű, \*mangrovészerű; \*Némethhel, \*Budapestszerű; \*Einsteinféle; \*thaival stb., miközben elutasítja a következő helyes alakokat: edzd, pedzd, kocsijai, mozijai, Béláék, unokámék, kulimászszerű; New York-iaak; linuxos, windowsos; thajjal stb.
6. Egyes szavak hibás alakban is szerepelnek a szótárban: \*aszkrétikus (aszketikus), \*gargalizál (gargarizál), \*trikolor (trikolor), \*szférikus (szferikus) stb., de van, ami csak hibásan: \*adrez (adrezz), \*bájtos (byte-os), \*entuziazmus (enthuziazmus), \*libling (liebling), \*Lódz (Łódź), \*mancsaft (mannschaft), \*szanszevéria (szanszeviéria), \*szkunksz (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, fész, pót, süv, tel, veé.
7. A jóval 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 (α-ról, β-sugárzás, ε-nal, μ-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 javaslattevő képességével nemcsak a hasonló szavakra ad javaslatot (pl. előről → előlről, Csaucseszku → Ceauşescu, arvizturo → árvíztűrő), hanem az eltérőbbekre is (file-t → fájl, 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. *olajáremelkedés*), helyes összetételeket jelezhet hibásnak, mint például régebben a *ligatúrahaszniát*, *notebooküzletág* szavakat (ez utóbbi kettő a Microsoft Office-ban még hibásan van kezelve).

## Szinonimaszótár

Az OpenOffice.org szinonimaszótára amellet, hogy 20 ezer szavas szókincse összemérhető a Microsoft Office-éval, számos újdonságot is tartalmaz. Ilyen a tövezés és todalékolás, 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 todalékolva. 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.



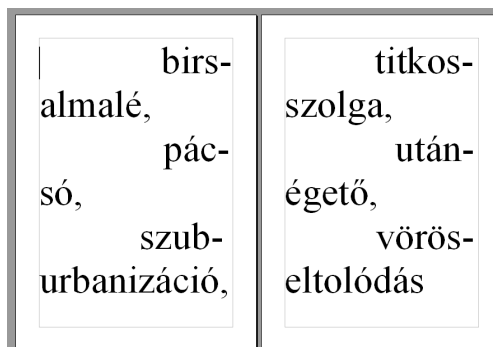
2. ábra: Ragozott szinonimák az OpenOffice.org-ban

## Elválasztás

Az elválasztás lehetővé teszi a szép, egyenletes szedést sorkizárás esetén, amellett, hogy takarékos nyomtatást eredményez. A Microsoft Office 2007 nem választja el azokat a szavakat, amelyeket nem ismer fel a helyesírás-ellenőrző. Mivel ez gyakran előfordul a programban, 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](http://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.



3. ábra: Helyes elválasztás az OpenOffice.org-ban

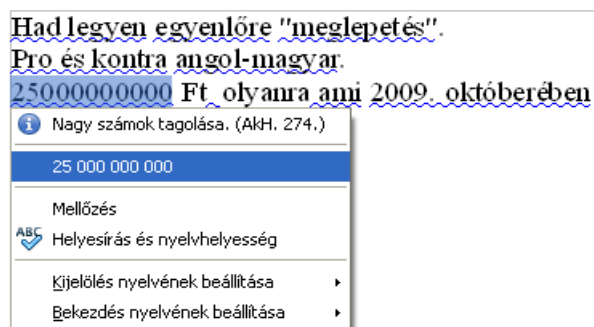
További különbség, hogy az OpenOffice.org elválasztóprogramja támogatja az Unicode karakterkódolást. Az OpenOffice.org magyar elválasztási szótárának unicode-os kiegészítése helyesen választja el a nyílt forráskódú helyesírási szótárban szereplő idegen ékezetes szavakat.

A frissített szótárak az OpenOffice.org Novell Edition 3.2 verziójában jelennek meg, a korábbi verziókhöz az OpenOffice.org oldaláról érhető el:

[http://extensions.services.openoffice.org/project/hu\\_dicts](http://extensions.services.openoffice.org/project/hu_dicts)

## Mondatszintű helyesírás-ellenőrzés

A szóellenőrzőn lép túl a mondatellenőrző (nyelvhelyesség-ellenőrző), aminek egyik leghasznosabb tulajdonsága, hogy a mondat szinten jelentkező elütések, hiányzó vagy felesleges szóközök és írásjelek, írásjel- és szótévesztések, a súlyosabb egybe- és különírási hibák 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 eszközbe se jusson kikapcsolni. A következő változatokban pedig 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.



4. ábra: Mondatellenőrzés az OpenOffice.org-ban

A frissített szótárak az OpenOffice.org Novell Edition 3.2.1 verziójában jelennek meg, a korábbi verziókhöz az OpenOffice.org oldaláról érhető el:

<http://extensions.services.openoffice.org/project/lightproof>

## Számok gépi fordítása

Az OpenOffice.org-hoz kapcsolódó nyelvtechnológiai újdonságok jó példája a magyar fejlesztésű NUMBERTEXT OpenOffice.org-kiegészítés. A számok automatikus számnévvé alakítása megkönnyíti a szerződések, számlák, banki dokumentumok elkészítését. Míg a Microsoft Office csak thai nyelven képes a számokat számnévre és thai pénznemre, *bátra* és *szatanga* átalakítani az azóta már nemzetközi szabványként is elkönyvelt BAHTTEXT függvényrel, 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>6</sup>

	A	B	C	D	E	F	G
1	2009	kétezer-kilenc					
2	1232424	egymillió-kétszázharminckétezer-négyszázhuszonnégy					
3	-5535,34	minusz 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							

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

a megfelelő OASIS munkacsoporthoz is benyújtásra került, hogy a nemzetközi ISO/OASIS OpenDocument dokumentumformátum-szabvány részévé váljon.<sup>6</sup>

A frissített szótárak az OpenOffice.org Novell Edition 3.2.1 verziójában jelennek meg, a korábbi verziókhöz az OpenOffice.org oldaláról érhető el:

<http://extensions.services.openoffice.org/project/numbertext>

## 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 sűgőval rendelkezik. Pár különbséget érdemes megemlíteni: a rövid hónapnevet a dátumformátumokban (ami a Microsoft Office 2003-ban hibásan, két 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 helyesírási szabályzatban is szereplő kötőjeles (pl. 2009-09-18) dátumformátumot használja alapértelmezésként a táblázatkezelőjében.

	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				

6. ábra: Az OpenOffice.org helyes dátumformátumai

<sup>6</sup> <http://numbertext.org>

## Más nyelvek

Gyakran fordulnak elő idegen szavak, idézetek a dokumentumokban. A Microsoft Office magyar változata angol és német nyelvtámogatást is tartalmaz, más nyelvi eszközöket külön kell megvásárolni. A magyar OpenOffice.org FSF.hu kiadása az angolon és a németen kívül még francia helyesírási, elválasztási és szinonimaszótárral érkezik, és az OpenOffice.org honlapjáról még további száz nyelvhez érhető el 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 helyesírási-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>7</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> <http://hu.wikipedia.org/wiki/Hunspell>