

Igényesség, helyesírás, számítógép

Prószéky Gábor
proszeky@morphologic.hu

MorphoLogic

<http://www.morphologic.hu>

Pázmány Péter Katolikus Egyetem
Információs Technológiai Kar

<http://www.itk.ppke.hu>



Hogy kell leírni, amit hallunk?

(azaz: miért is született meg a „spell-checker”?)

**Eye halve a spelling chequer,
It came with my pea sea,
It plainly marques four my revue
Miss steaks eye kin knot sea.**

**Eye strike a key and type a word
And weight four it two say
Weather eye am wrong oar write
It shows me strait a weigh.**

**As soon as a mist ache is maid
It nose bee fore two long
And eye can put the error rite
Its rarely ever wrong.**

**Eye have run this poem threw it
I'm shore your pleased two no
Its letter perfect in its weigh,
My chequer tolled me sew.**

**I have a spelling checker,
It came with my PC,
It plainly marks for my review
Mistakes I cannot see.**

**I strike a key and type a word
And wait for it to say
Whether I am wrong or right,
It shows me straightaway.**

**As soon as a mistake is made
It knows be for too long
And I can put the error right,
It's rarely ever wrong.**

**I have run this poem through it.
I'm shure you're pleased to know,
It's letter perfect in its way,
My checker told me so.**

Helyesírásellenőrzőprogram-történelem

- ❑ Szövegszerkesztők IBM PC-re, kb. 1981-től
- ❑ Az első helyesírásellenőrző-gyártók a 80-as évek végétől
- ❑ Egy teljes merevlemez mint korrektúra-tárhely lehet-e a megoldás?
- ❑ Az elválasztás szólistás megoldásokkal nem megy könnyen, pl. *ba-goly*, *bag-ly+ok*
- ❑ Csak a szavak morfológiai szintű kezelése lehet a megoldás: kétszintes morfológia (1983-től), unifikációs morfológia (1990-től)

Szóellenőrzés számítógépes morfológiai elemző segítségével

Ha pl. az ismeretlen alak a kérdése, akkor

kérdése

kérdés[FN]+e[PSe3]

főnévi

kérd[IGE]+és[IF]+e[PSe3]

főnévi

kérdéses

kérdéses[MN]

melléknévi

kérdés[FN]+es[SKEP]

melléknévi

kérd[IGE]+és[IF]+es[SKEP]

melléknévi

kérdesse

kérd[IGE]+es[MUV]+se[TPe3]

igei

kérdéssé

kérdés[FN]+sé[FAC]

főnévi

kérd[IGE]+és[IF]+sé[FAC]

főnévi

kérdéssel

kérdés[FN]+sel[INS]

főnévi

Helyesírás-ellenőrzés különböző nyelvekre

- ❑ Angol (házi feladat Knuth: „A programozás művészeté” -ben)
- ❑ Az angolnál bonyolultabb, de azért „felsorolhatóan sok” szóalakot használó nyelvek gépi leírása
- ❑ Komplex latinbetűs nyelvek: finn, magyar, észt, török (agglutinálnak és fontosak!)
- ❑ Sémi nyelvek (sok a formai többértelműség!)
- ❑ Ideografikus írások és átírásaik (pl. japán írások, egyszerűsített kínai)

Idővel további kihívások

- ❑ Kétféle helyesírási rendszer (pl. norvég)
- ❑ A reformhelyesírások megjelenése (pl. német, holland, román)
- ❑ Megjelennek „újabb” nyelvek (szerbhorvát > szerb, horvát, majd bosnyák; most az ír; ...)
- ❑ Megjelennek a nem-államnyelvek beszélőinek igényei (katalán, szorb, velszi, fríz, ...)
- ❑ Helyesírás-ellenőrzés és a politika találkozásai: számi, romani (lovári), beás, ...
- ❑ Korábban nem egyértelműen szabályozott nyelvek: a máltai és a baszk esete a gépi helyesírással

Csak lábjegyzetként: két máltai „kódolás”

Xideu il cada, ye gireni tale nichadithicum
Mensab fil gueri uele nisab fo homorcom
Calb mehandihe chakim soltan ui le mule
Bir imgamic rimitne betiragin mucsule,
Fen hayran al garca nenzel fi tirag minzeli
Nitla vu nargia ninzil deyem fil bachar il hali.

Huakit hi mirammiti lili zimen nibni,
Mectatilix mihallimin, me chitali tafal morchi;
Fen timayt insib il gebel sib tafal morchi;
Vackit hi, mirammiti.

Xidew il-qada, ja ġirieni, talli nħadditkom,
Ma nsab fil-weri u la nsab f'għomorkom
Qalb m'għandha ħakem, sultan u la mula
Bir imgħammiq irmietni, b'turġien muħsula,
Fejn ħajran għall-għarqa, ninżel f'taraġ minżeli
Nitla' u nerġa' ninżel dejjem fil-baħar il-għoli.

Waqgħet hi, imrammti, l'ili żmien nibni,
Ma ħtatlix mgħallmin, 'mma qatagħli tafal merħi;
Fejn tmajt insib il-ġebel, sibt tafal merħi;
Waqgħet hi, imrammti.

Jól kodifikált rendszerben nehéz változásokat bevezetni

- ❑ A meglevő szabályok megváltoztatására elvileg csak akkor kerülhet sor, ha „az úzus egyáltalán nem képes elfogadni és alkalmazni hosszú idő után sem, vagy a korábban kodifikált szabály teljességgel ellentmond a nyelv működésének, rendszerszerűségének”

(Laczkó K.: Gondolatok a helyesírási szabályzat 12. kiadásáról
In: Balázs G. (szerk.) *A magyar nyelvi kultúra jelene és jövője I-II.*)

- ❑ Ilyen eset például „egy-egy szóalak írásképeinek a korábbi alakhoz képest eltérő módon való kodifikálása” (L.K)
- ❑ Ki által? (pl. a *kompatibilis* nem „korábbi” alak volt)
- ❑ Észre kell vennünk, hogy **a mai szövegek nagy része olyan típusú írásbeliség keretében jelenik meg, mellyel korábban nem találkoztunk**, és a szabályozásért sokak által felelősnek gondolt szakemberek egy része ma sem találkozik

Ami szokásos, az felsorolható?

- Kérdés, hogy az „úzus diktálta kivételek” (*szúette, jótett, bérbeadás, napkelte*) egyediek, vagy újabb nyomtatott helyesírási szótár megjelenése nélkül is bővíthető a lista?
- „A meggyökeresedett szokást megtartva nem egyszer jelentésváltozás nélküli szókapcsolatokat is egybeírunk...” (HKSz)
- Kinek a szokását? Pl. a *nyitvatartás* alak hibás vagy inkább ide tartozik?

Nézzük meg a „szokásokat”: mit írnak a magyar weben? És ennek következtében: mit olvasnak a magyar weben?

jobbkez: 13 600 - jobb kéz: 38 300
házifeladat: 57 200 - házi feladat: 106 000
csevely: 21 800 - csevej: 64 300
muszály: 147 000 - muszáj: 954 000
szervíz: 1 020 000 - szerviz: 1 880 000
köppen: 2 420 - köpeny: 173 000
bakkancs: 9 890 - bakancs: 124 000
jappán: 42 600 - japán: 2 160 000
egy: 28 100 - egy: 5 390 000
helyesírásellenőrző: 2 900
helyesírás ellenőrző + helyesírás-ellenőrző: 638
helyesírás ellenőrző + helyesírás-ellenőrző: 35 200
meg-e lehetne: 558
nem-e lehetne: 8810

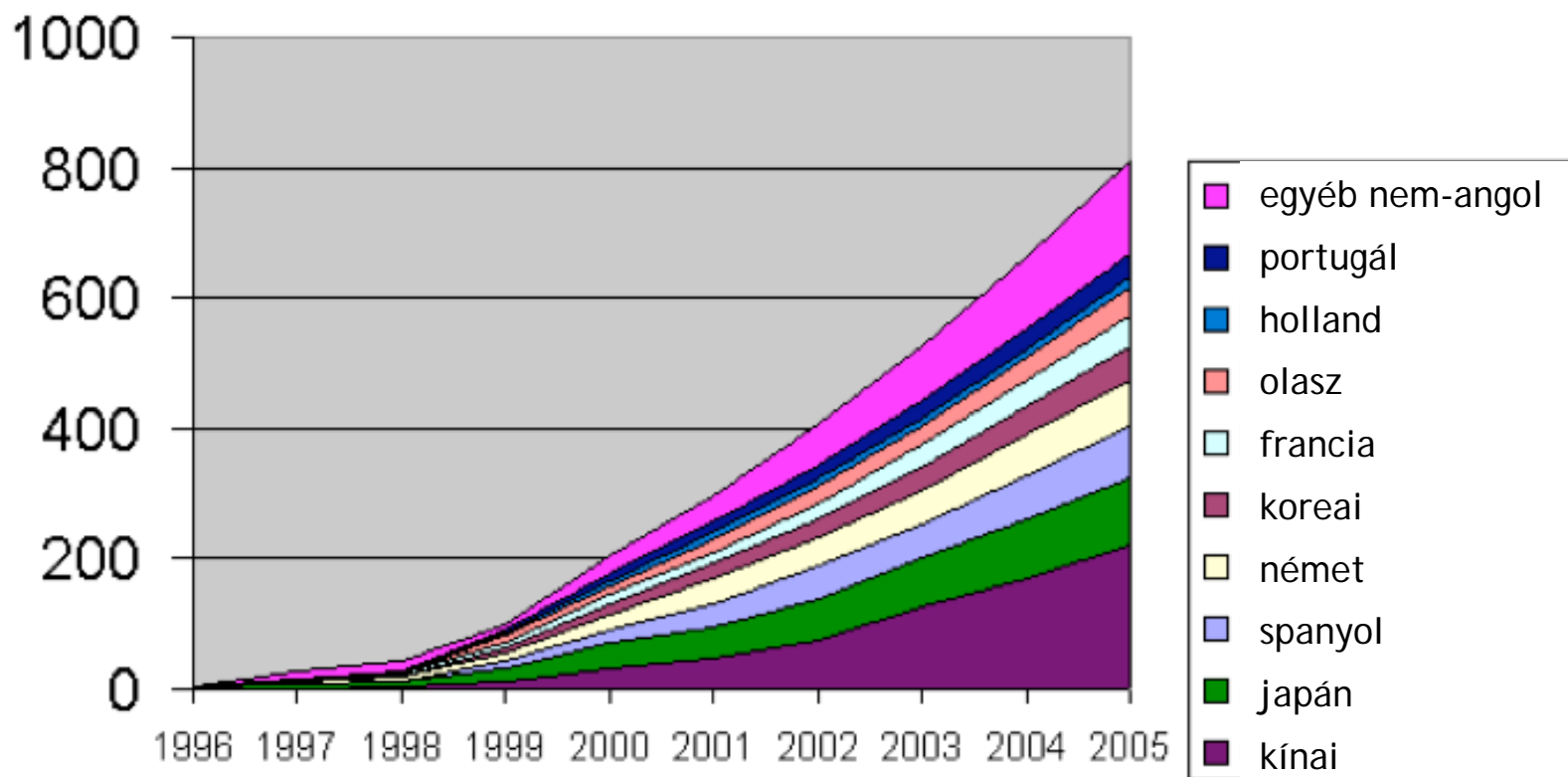
Becslések latin betűs internetes szövegek lehetséges méretéről

	Nyelv	Szó a weben
1	angol	76 598 718 000
2	német	7 035 850 000
3	francia	3 836 874 000
4	spanyol	2 658 631 000
5	olasz	1 845 026 000
6	portugál	1 333 664 000
7	holland	1 063 012 000
8	svéd	1 003 075 000
8	norvég	609 934 000
10	cseh	520 181 000
11	magyar	457 522 000
12	dán	346 945 000
13	finn	326 379 000
14	lengyel	322 283 000
15	szlovák	216 595 000
16	katalán	203 592 000



	Nyelv	Szó a weben
17	török	187 367 000
18	maláj	157 241 000
19	horvát	136 073 000
20	szlovén	119 153 000
21	észti	98 066 000
22	ír	88 283 000
23	román	86 392 000
24	eszperantó	57 154 000
25	latin	55 943 000
26	baszk	55 340 000
27	izlandi	53 941 000
28	lett	39 679 000
29	litván	35 426 000
30	velszi	14 993 000
31	breton	12 705 000
32	albán	10 332 000

A nem angol nyelvű szövegek növekedése az interneten



Gépelési tévesztések vizsgálata (1)

	A teljes web	>>	A magyar web
internet	2 460 000 000	918	2 680 000
internte	67 400	4212	16
interent	681 000	1114	611
intenet	116 000	2762	42
intrenet	193 000	541	357
inetrnet	128 000	538	238
itnernet	66 400	4150	16
ninternet	47 700	-	0
interne.	19 200 000	1466	13 100
intern.t	1 940 000	7132	272
inter.et	19 400 000	27714	700
inte.net	2 480 000	246	10 100
int.rnet	436 000	682	625
in.ernet	522 000	418	1 250
i.ternet	441 000	1400	315
.ninternet	1 150 000	2758	417

Gépelési tévesztések vizsgálata (2)

	A magyar web
nyelv	4 410 000
nyevl	11
nylev	0
neylv	32
ynelv	0
nyel.	(57 000)
nye.v	288
ny.lv	60
n.elv	123
.yelv	219

Mit tud egy helyesírási program a helyesírásunkat meghatározó elvekkel kezdeni?

- Kiejtés szerinti írásmód - semmit
- Szóelemző írásmód - sok mindent (néha túl sokat is!)
- Hagyományos írásmód - amit megadunk „előre”, az mindig úgy lesz!
- Egyszerűsítő írásmód - nem tud dinamikusan reagálni, csak bizonyos korlátok között
- De még azzal sem tud mit kezdeni, hogy „normálisan” a bekezdésvége-jel egyúttal mindig szó, illetve mondat végét is jelöli, hiszen a számítógépes szövegben sokszor törnek szét a mondatok, pl. felsorolójelek használata miatt (megoldás van, csak nem szokták tudni!)

Helyesírásunk alapelemei a gép szemével

- Betűk és betűrendbe sorolás
 - sokkal jobb, mint az ember, de a felhasználók nem is tudják kezelni!
- Szótövek írása és toldalékok kapcsolása
 - ld. fent: a négy alapelv érvényesülése
- Különírás és egybeírás
 - listás megoldás: tökéletes; de analógia: csak felügyelet mellett!!
- Kis és nagy kezdőbetűk, a tulajdonnevek kérdése
 - ha a „köznéviség” nem zavar be, ui. a kisbetűs elvileg nem tulajdonnév (de lehet olyan eredetű!), a nagybetű oka pedig nem feltétlen tulajdonnévi!
- Az idegen szavak írása
 - amit nem tud, azt nem tudja!
(de: megtanítható egy-egy konkrét nyelven – pl. angol – kiejteni)
- Írásjelek
 - környezetelemzéssel valamit lehet tenni, de ez max. tanács
- Rövidítések és mozaikszók
 - minden lehet rövidítés, tehát csak a felsorolhatókkal tud mit kezdeni
- Elválasztás
 - gyakorlatilag is közel 100%-os megoldás (pontosság/fedés)

Az igazán problémás jelenségek és a nyelvhelyességi programok viszonya

- ❑ Összetételek
(pl. *Lebensversicherungsgesellschaftsangestellter*)
- ❑ Szófajismeretet is igénylő produktív szabályok
(pl. a magyar 6-3)
- ❑ A mozgószabály formális kezelése
(pl. *jobbkéz-szabály*)
- ❑ Formaváltoztató szóalakok
(pl. a német vagy a magyar elválasztás)
- ❑ Tipográfiai szokások és helyesírás
(pl. *autói*)
- ❑ Nyelvismeret-függő szegmentálások kezelése
(pl. *post+sc-riptom*, *diā+g-nózis*, *kon+k-rét*, *di+sz-tichon*)
- ❑ ... és ami nem helyesírás, hanem „népnevelés” volna inkább: hogyan kell használni a gépi helyesírás-ellenőrzőket, és hogyan kell érteni a „javaslataikat” ...

Nyelvhelyesség-ellenőrzés a szóhatáron túl

- ❑ Hiba-nyelvtan vs. nyelvten
- ❑ Parciális elemzések
- ❑ Hibaelemzések, a hibák súlyozása
- ❑ A hiba és a nem-hiba határának elmosódása
- ❑ A nyelvi vagy a formai természetű hibák szűrésének preferálása
- ❑ Stílusellenőrzés számítógéppel


Hogy fest ez a gyakorlatban?

Aki másnak vermet és maga esik bele.

Elment mellettem anélkül hogy köszönt volna.


A **hogy** kötőszó elé vesszőt kell írni!
Javaslat: **mellettem,**

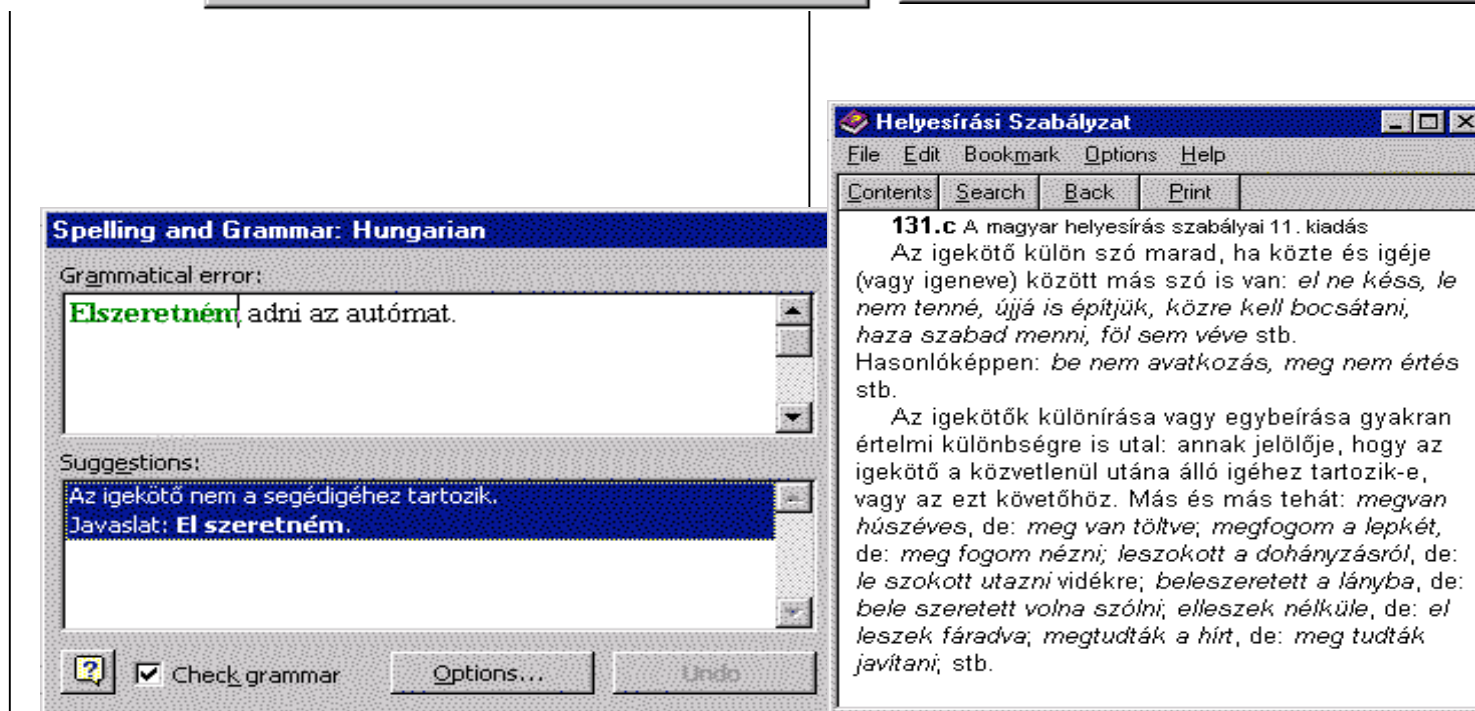
Ignore Sentence

 Grammar...

Ha ez egy összetett mondat, a jelzett helyen vagy környékén egy vessző hiányzik.
Ellenőrizze!
Javaslat: **maga,**

Ignore Sentence

 Grammar...



Spelling and Grammar: Hungarian

Grammatical error:
Elszeretném adni az autómat.

Suggestions:
Az igeikötő nem a segédigéhez tartozik.
Javaslat: **El szeretném.**

Check grammar Options... Undo

Helyesírási Szabályzat

File Edit Bookmark Options Help

Contents Search Back Print

131.c A magyar helyesírás szabályai 11. kiadás
Az igeikötő külön szó marad, ha közte és igéje (vagy igeneve) között más szó is van: *el ne késs, le nem tenné, újjá is építjük, közre kell bocsátani, haza szabad menni, föl sem véve* stb. Hasonlóképpen: *be nem avatkozás, meg nem értés* stb.

Az igeikötők különírása vagy egybeírása gyakran értelmi különbségre is utal: annak jelölője, hogy az igeikötő a közvetlenül utána álló igéhez tartozik-e, vagy az ezt követőhöz. Más és más tehát: *megvan húszéves*, de: *meg van töltve; megfogom a lepkét*, de: *meg fogom nézni; leszokott a dohányzásról*, de: *le szokott utazni vidékre; beleszeretett a lányba*, de: *bele szeretett volna szólni; elleszek nélküle*, de: *el leszek fáradva; megtudták a hírt*, de: *meg tudták javítani*, stb.

A helyesírás-tanulás motivációjáról

„Miért tanulják a diákok a helyesírást? Sokféle választ kaphatunk erre a kérdésre. Talán az iskola szorításának hatására vagy egy rokonszenves magyartanár kedvéért, esetleg szülői ráhatásra, de akadnak olyan diákok is, akiket saját meggyőződésük ösztönöz helyesírás-tanulásra. Érzik, tudják, belátják, hogy a helyesírástudásra szükségük van és szükségük lesz egész életükben. A magyar helyesírás szabályainak ismerete határozottságot, biztonságot ad, segít a magán-, a hivatalos, a köz- és a tudományos életben: önéletrajzok, pályázatok, hirdetések, üzenetek, cikkek, tanulmányok, iskolai dolgozatok, személyes és hivatalos levelek megfogalmazásában, a szövegszerkesztő okos használatában.”

Antalné Szabó Ágnes (2001): A helyesírás-tanulás motivációja (Beszámoló a Simonyi Zsigmond helyesírási verseny 2001. évi Kárpát-medencei döntőjéről). *Nyr* 125(3)

Mit írnak erről a webes fórumokon?

Egy fórumról, 2007-ből:

„Ezentúl a helyesírás hozzátartozik az alpműveltséghez, de szerintem némely hiba hozzátartozik az emberi mivoltunkhoz, arra van a helyesírásellenőrző.”

(sic!)

<http://shamalt.scft.hu/1652/helyesiras-es-ertelmesseg/>

Mit gondolnak az emberek a gépi helyesírás-ellenőrzésről?



#1 - KJ - 2006.11.06 - 16:23:37

Olyan természetes, ha az ember ír, legalapvetőbb, hogy a helyesírásra figyel. Ha nem biztos valamiben megkérdezi, vagy megnézi valahol. Van, úgy hogy véletlenül, csak a tartalomra figyelve becsúszik egy-egy hiba, de itt van lehetőség az utólagos javításra is. Velem is előfordult már. Nekem fizikai fájdalom, ha komoly helyesírási hibákat látok. Aki arra szánja magát, hogy másnak is megmutatja az írásait, annak kötelessége mondatról, mondatra átnéznie művét, jól írta-e!

#9 - ER - 2006.11.07 - 13:32:53

ha mindenki a szövegszerkesztőbe írná be először a szöveget, akkor ott automatikusan ki lenne javítva.

#10 BM - 2006.11.07 - 13:48:43

hát igen, ennyi kellene, ebben meg is egyezhetünk.

#11 BM - 2006.11.07 - 13:49:27

"Szerintem, a helyesírás képessége: az általános műveltség tartozéka."
szerintem meg nem. a telefonkönyv sem az.

#12 - ÖÜ - 2006.11.07 - 13:50:43

Egy francot, az ellenőrző se javít ki mindent, ami hibásan is elfogadott szó, az átmegy. Persze a 90-98 %-át javítja.

#13 - ÖÜ - 2006.11.07 - 13:52:17

De egy cikkben legalábbis esztétikailag nem elhanyagolható szempont

#14 - BM - 2006.11.07 - 14:11:45

senki nem mondta, hogy a helyesírásellenőrző mindent javít. (vagy írta vki?)
a helyesírás egy cikkben fontos szempont, egyetértek.



A web a gépi nyelvi eszközök megítéléséről

PZS, 2005. május 6. 16:31

A kérdésem, hogy hogyan lehet (kell) megcsinálni azt, hogy egy angol office-hoz legyen magyar nyelvű helyesírás ellenőrzés?

HM, 2005. május 7. 8:20

El kell menni a Morphologichoz és megvenni a helyes-e csomagot, és akkor lesz magyar helyesírás ellenőrző az officeban.

BE, 2005. augusztus 24. 7:22

Hú ez nagyon jó! Én már régóta szenvedek de semmi.... Megnézem mit tud ez a Morfolódzsik:)

Drága? :-)

BE, 2005. augusztus 24. 7:25

VÁÁÁÁÁ! Ez 20.000 ft, ennyiért én magamtól bepötyögöm a saját szótárba :-) Hát milyen rablás ez kérem szépen, egy bolti 2 nyelvű szótár az 1/3adába kerül, és itt még papáír sincs, csak egy 200ft-os korong...

SB, 2005. augusztus 25. 13:3

A helyesírás-ellenőrző nem egy szótár! Annál egy picit több. Ismernie kell a szavak toldalékolását, az igék összes alakját és idejét, továbbá a nyelvtani ellenőrzőjének még elemezni is kell tudnia a mondatokat.

Van azért objektív tájékoztatás is...

Helyesírás-ellenőrzés

A beépített helyesírás-ellenőrző sem pótolja teljesen a szigorú „kézi” ellenőrzést, mert

- a magyar nyelvtan és az azt felügyelő szervezetek sajátosságai miatt a nyelvi szabályozás messze nem „hézagtalan”, aminek következtében a program alkotói, és a szöveg felhasználói között mindig van „nézeteltérés”. Továbbá, szintén a „hézagos” szabályzás miatt a ritkábban előforduló helyzeteket egyedi módon oldják meg;
- még egy nyelvtanilag teljesen „hibátlan” program sem birkózna meg a magyar vagy idegen tulajdonnevekkel, termékek típuszámaival, egyéb, a szövegbe ágyazott számokkal;
- az adott szervezet szakterületének speciális kifejezései nem szerepelnek az alapszókincsben, így a szóellenőrző nem ismerheti azokat;
- a helyesírás-ellenőrző csak a szavakat vizsgálja, az írásjelek helyes használatát nem, vagy csak korlátozott érvényességgel.

A fentiek ellenére az automatikus helyesírás-ellenőrző hasznos szolgáltatás, mert a tapasztalatok szerint egy gépelt oldalon átlagosan 5-10 elgépelési hibát hagy az, aki begépelte, és *kétszer átolvasta*. Az emberi agy működés módja miatt ugyanis az, aki begépelte, egyszerűen képtelen felfedezni a saját hibáit – amelyeket viszont a helyesírás-ellenőrző kijelöl.

<http://itl7.elte.hu/~zsolt/Oktatas/SzA/Irogep.pdf>

Mi kell a helyes (és a szabályzattal szinkronban levő) íráshoz?

- Hibafelismerési készség
- Szabályzatismertet
- Analógiaalkalmazási készség
- De leginkább: igényesség (amiből a fentiek könnyebben következhetnek...)

Tanulni kell ezt is használni!

„A számítógép úgy segít a hibák javításában, hogy az írónak nem kell félnie egy külső személy ítéletétől, így kevésbé frusztráló. Bízunk tehát a géppel javítható eseteket a gépre, és az oktatás és a nyelvművelés erőforrásait koncentráljuk a népszerűtlen-gépszerűtlen helyesírási esetekre. Persze, ez csak akkor lehetséges, ha az emberek (és nem utolsósorban a magyar szakos tanárok) megtanulják, hogyan kell a gépet erre a célra használni.”

Kis Á. (1999) Az akadémiai helyesírási szabályzat és a számítógép. *Nyr* 123(2)

A gép nem ember

(és viszont: az ember nem gép!)

“Több mint húsz éve vagyok olvasószerkesztő, könyveket, napi- és üzemi lapokat, üzleti dokumentumokat, pályázatokat korrektúrázok. A számítógépek elterjedésével sokan azt hitték, ezt a szakmát felváltja majd a szövegszerkesztő szoftverek helyesírás-ellenőrző opciója. Tévedtek. Az olykor igénytelenül, sokszor magyartalanul és érthetetlenül megfogalmazott szövegeket javítani, a helyesírás komplex szabályait alkalmazni egyetlen szoftver képes csak: a kreatív, munkájával szemben igényes, gondolkodó ember.”

(Budán László olvasószerkesztő, Magellán PR & Hírügynökség,
<http://www.magellanpr.hu/index.php?inc=inc.fedelzet.php&title=Fed%E9lzet>)

... az igényességet viszont
nem pusztán a nyelvtanórán
kell(ene) tanítani!



Köszönöm a figyelmet!