

# Magyar Ispell dokumentáció

Németh László

2003. július 22.

# Tartalomjegyzék

<b>1. Bevezetés</b>	<b>3</b>
1.1. A dokumentáció felépítése . . . . .	3
1.2. Elérhetőség . . . . .	3
1.3. Köszönetnyilvánítás . . . . .	3
1.4. A dokumentáció . . . . .	4
1.5. Felhasználási engedély . . . . .	4
<b>2. Telepítés és használat</b>	<b>6</b>
2.1. OpenOffice.org . . . . .	6
2.1.1. Telepítés . . . . .	6
2.1.2. Használat . . . . .	6
2.1.3. A szöveg nyelvének kiválasztása . . . . .	7
2.1.4. Nagybetűs és számokat tartalmazó szavak ellenőrzése . . . . .	7
2.1.5. Alapértelmezett nyelv beállítása . . . . .	7
2.1.6. Automatikus elválasztás beállítása . . . . .	7
2.1.7. Helyesírási állományok frissítése . . . . .	8
2.1.8. Az OpenOffice.org nem magyar változatai . . . . .	8
2.2. A Hunspell telepítése és használata . . . . .	9
2.2.1. Linux és egyéb Unix-típusú rendszerek alatt . . . . .	9
2.2.2. Windows alatt . . . . .	11
2.2.3. Hunspell Mac OS X alatt . . . . .	12
2.3. Emacs . . . . .	12
2.3.1. Az Emacs telepítése Linuxra . . . . .	12
2.3.2. Az Emacs telepítése Windowsra . . . . .	13
2.3.3. A Xemacs telepítése Windowsra . . . . .	13
2.3.4. A Hunspell telepítése az Emacs alá . . . . .	13
2.3.5. A Hunspell telepítése konzolon indított Emacshoz . . . . .	13
2.3.6. Az Emacs használata . . . . .	14
2.4. Az Ispell telepítése és használata . . . . .	14
2.4.1. Telepítés Linuxra . . . . .	15
2.4.2. Telepítés Linuxra, forráskódból . . . . .	15
2.4.3. Windows . . . . .	15
2.4.4. A magyar.hash állomány telepítése és használata . . . . .	16
2.4.5. Az Ispell változat megfeleltetése . . . . .	17
2.4.6. Az Ispell használata . . . . .	17

<b>3. Fejlesztői dokumentáció</b>	<b>19</b>
3.1. Helyesírás-ellenőrzők ismertetése	19
3.1.1. Aspell	19
3.1.2. Ispell	19
3.1.3. Pspell	20
3.1.4. Myspell	20
3.1.5. Mozilla Spell Checker	21
3.1.6. Magyar Myspell	21
3.1.7. OpenOffice.org	22
3.2. Az Ispell felépítése és működése	23
3.2.1. Ragozási szabályok	23
3.2.2. Ragozási osztályok	24
3.2.3. Szótárállomány	25
3.2.4. Affixum tömörítés	25
3.3. Magyar Ispell	25
3.3.1. Történet	26
3.3.2. Alapvető problémák	26
3.3.3. Affixum keretrendszer	28
3.3.4. Szótári keretrendszer	33
3.4. Új szótármodulok létrehozása	34
3.4.1. Szófaji bontás	34
3.4.2. Kivételek a szófaji kategóriákon belül	35
3.4.3. Állományok táblázata	36
3.4.4. Y-ra végződő szavak	37
3.4.5. Kivételek könyvtár	39
3.4.6. További információk	39
3.4.7. Segédprogramok	39
3.5. Magyar Myspell	40
3.6. Javítási cseretáblázat	40
<b>4. Programozói dokumentáció</b>	<b>43</b>
4.1. Hunspell csőfelület	43
4.2. Magyar Myspell felület	44
4.2.1. Konstruktor	44
4.2.2. Metódusok	44
<b>5. Hunspell kézikönyvoldalak</b>	<b>48</b>
5.1. Hunspell(1)	48
5.2. Hunstem(1)	57
5.3. Makedb(1)	58
5.4. Lookdb(1)	59
5.5. Hunspell(4)	60

<b>6. A Magyar Ispell tesztelése</b>	<b>67</b>
6.1. A szókincs tesztelése . . . . .	67
6.2. Tévesztések ellenőrzése . . . . .	68
6.3. A tesztek értékelése . . . . .	68
6.3.1. Szókincs . . . . .	68
6.3.2. Tévesztések . . . . .	68
<b>7. Szabályok</b>	<b>70</b>
Általános tudnivalók . . . . .	71
Általános tudnivalók . . . . .	71
A betűk . . . . .	71
A kiejtés szerinti írásmód . . . . .	72
A szóelemző írásmód . . . . .	75
A hagyományos írásmód . . . . .	77
Az egyszerűsítő írásmód . . . . .	77
A különírás és az egybeírás . . . . .	77
<b>Irodalomjegyzék</b>	<b>79</b>

# 1. fejezet

## Bevezetés

### 1.1. A dokumentáció felépítése

A Magyar Ispell dokumentáció első része a felhasználói ismereteket foglalja össze, második része a fejlesztők számára tartalmaz útmutatást. A függelék első felében a Hunspell helyesírás-ellenőrző és -javító program, és az általa használt szótárállományok formátumának kézikönyvlapjait találjuk. Itt kapott helyet egy részletes összehasonlítás is a Hunspell és a magyar helyesírási szabályzat (illetve a jelenleg legelterjedtebb magyar helyesírás-ellenőrző program, a MS Office XP-ben megtalálható Helyes-e?) között, amit végül egy a Hunspell működésének helyességét vizsgáló leírás követ.

### 1.2. Elérhetőség

A Magyar Ispell levelezési listája: [magyarispell@yahogroups.com](mailto:magyarispell@yahogroups.com)

A lista archívuma: <http://www.yahogroups.com/group/magyarispell>

A Magyar Ispell honlapja: <http://magyarispell.sourceforge.net>

A Szószablya honlapja: <http://www.szoszablya.hu>

### 1.3. Köszönetnyilvánítás

A dokumentáció korábbi változatának elkészítésében, javításában, kiegészítésében Mayer Gyula segített.

A fejlesztésben sokan részt vettek. Szántó Tamás, Bíró Árpád, Hilzinger Marcel, Koblinger Egmont, Godó Ferenc, Trón Viktor, Pásztor György, Noll János, Bencsáth Boldizsár, Fransoa Holop, Goldmann Eleonóra, Halácsy Péter, Kornai András és sokak segítségét köszönöm!

Anyagi jellegű támogatást kaptam Mátó Pétertől, a TypoTeX könyvkiadótól (Mayer Gyula segítségével), a SuSE Linuxtól (Koblinger Egmont és Hilzinger Marcel segítségével), az UHU-

Linuxtól. Jelenleg az állami támogatással működő Szószablya munkacsoport tagjaként a Magyar Ispell fejlesztésén is dolgozom.

A tesztprogramok és a dokumentáció 2002. évi változatának elkészítéséhez, valamint a tesztelés kivitelezéséhez a TypoT<sub>E</sub>X Kft. 300 000 (+áfa) forinttal járult hozzá, a Széchenyi Terv keretében az Informatikai Kormánybiztosság SZ-IS-10/3 pályázati számon elnyert támogatásából. 2002 őszétől az összegből vásárolt számítógéppel készülnek a Magyar Ispell új változatai.

Az IMEDIA Kft. (<http://www.imedia.hu>) különösen nagy segítséget nyújtott a szókincs ellenőrzésében és bővítésében, azzal, hogy nagy és ellenőrzött szókincset nyújtott át a fejlesztés számára.

Az UHU-Linux Kft. (<http://www.uhulinux.hu>) szoftverfejlesztési pályázatának támogatásával készült el a 0.98-as Magyar Ispell és ennek honlapja, valamint a Magyar MySpell használatát lehetővé tevő AbiWord-programkiegészítés (folt). Noll János és Koblinger Egmont munkája révén az UHU-Linux 1.0-s változatába is bekerült a legfrissebb Magyar Ispell szótárát és helyesírás-ellenőrzőt tartalmazó OpenOffice.org irodai csomag, illetve AbiWord szövegszerkesztő.

A Szószablya (<http://www.szoszablya.hu>) révén sor kerül a Magyar Ispell javítására és továbbfejlesztése is, mind szókincsét, mind a magyar nyelv kezelését tekintve, egy nyitott forráskódú magyar morfológiai elemző, a Hunmorph elkészítése céljából. A Szószablya keretén belül elvégzett vizsgálatok szerint a Magyar Ispell 0.99.3-as változata a magyar weboldalakon található (helyes) szóalakok körülbelül 96%-át – beleértve ebbe az összes gyakori szót – felismeri.

## 1.4. A dokumentáció

A dokumentáció forrása L<sup>A</sup>T<sub>E</sub>X nyelven készült. A mesterfájlban a `\usepackage{times}` utasítás abból a célból szerepel, hogy a pdftex program segítségével jó minőségű és kis méretű PDF változatot lehessen előállítani.

A dokumentáció HTML változata a

```
latex2html -split 4 magyarispell
```

parancs segítségével szinte hiba nélkül előállítható.

## 1.5. Felhasználási engedély

A Magyar Ispell szótár jelen változatát Németh László készítette a fejlesztés résztvevőinek és támogatóinak segítségével, és a GNU GPL (General Public License) kettes, vagy későbbi változata alapján szabadon felhasználható.

A Hunspell program, és a Magyar Myspell függvénykönyvtár a GNU LGPL (GNU Library GPL) kettes, vagy későbbi változata alapján szabadon felhasználható.

Ez a dokumentum a Free Software Foundation által kiadott GNU Free Documentation License 1.2-es vagy újabb változatában foglalt feltételek szerint másolható, terjeszthető és módosítható; rögzített fejezet, első borítólapszöveg és hátsó borítólapszöveg nincsen (<http://www.gnu.org/licenses/fdl.html>).

Németh László (nemethl@gyorsposta.hu)

## 2. fejezet

# Telepítés és használat

A Magyar Ispell fejlesztés eredményei Linux, Windows és Mac OS X operációs rendszereken is használhatóak olyan programokkal, mint az OpenOffice.org, AbiWord, Hunspell, Emacs, vagy a magyar nyelv támogatása szempontjából elavult Ispell.

### 2.1. OpenOffice.org

Az OpenOffice.org a legjobb nyitott forráskódú irodai csomag, ezért a legtöbb felhasználó az OpenOffice.org révén kerül kapcsolatba a Magyar Ispell fejlesztés eredményeivel.

#### 2.1.1. Telepítés

Az OpenOffice.org hivatalos magyar változata az FSF.hu Alapítvány honlapjáról érhető el: <http://office.fsf.hu>. Itt tölthető le Windows, és Linux platformra is, illetve már letöltött formában számítástechnikai lapok mellékletén is megtaláljuk.

A program magyar nyelvű telepítővel rendelkezik. Kövessük a telepítő, vagy a honlap (illetve az újság) útmutatását!

#### 2.1.2. Használat

Az OpenOffice.org magyar változataiban alapértelmezett nyelv a magyar, és rendszerint a beírás közbeni ellenőrzés is be van kapcsolva. Ez utóbbi lehetőséget a bal oldali eszköztáron található piros hullámvonallal aláhúzott ABC ikonnal tudjuk ki- és újra bekapcsolni.

Az OpenOffice.org kis hibája, hogy a szövegszerkesztőben az első betű beírása után töltődik be a magyar helyesírási szótár. Ez lassabb gépeken másodperceket is igénybe vesz. Mivel az ellenőrzés párhuzamosan folyik az írással, nem kell megvárnunk a szótár betöltődését, nyugodtan elkezdhetünk gépelni. (Igaz, a gépelt szöveg egy részét a szótár betöltődésének befejeződéséig nem fogjuk látni.)



Ha beírás közben hibás szót írtunk be, a szó piros hullámvonallal kerül aláhúzásra. Ha a jobb egérgombbal kattintunk a szón, a megjelenő helyi menü tartalmazza a javaslatokat a hibás szóra.

Az ellenőrzést külön párbeszédablak segítségével is elvégezhetjük. Ez az ablak az F7-es gomb lenyomásával hívható elő a leggyorsabban. A később tárgyalt beállításokhoz itt egy helyen férhetünk hozzá, így célszerű lehet megismerkedni a használatával.

### **2.1.3. A szöveg nyelvének kiválasztása**

Előfordul, hogy olyan dokumentummal találkozunk, ahol a szöveg, vagy annak egy része hibásként jelenik meg, bár helyes magyar szavakat tartalmaz. Ennek oka, hogy a szöveg – bár magyar – más nyelvűnek lett beállítva.

Jelöljük ki a szöveget, és válasszuk ki a Formátum→Karakter párbeszédablakban a Betűtípus lapot! Válasszuk ki a Nyelv listából a magyart, és kattintsunk a OK gombon!

Ha egy szó, vagy bekezdés nyelvét szeretnénk átállítani, az OpenOffice.org rendelkezik egy kényelmes lehetőséggel: kattintsunk egy hibásnak jelzett szón a jobb gombbal. Az OpenOffice.org a rendelkezésre álló helyesírási szótárak alapján megpróbálja megállapítani, hogy milyen nyelvű a kérdéses szó. Ha ez sikerül, a felnyíló helyi menü alján két új menüpontot is láthatunk. A példánknál maradva, „A szó: magyar” menüponttal a kiválasztott szó, a „bekezdés: magyar” menüponttal a teljes bekezdés nyelvét állíthatjuk egy kattintással magyarra.

### **2.1.4. Nagybetűs és számokat tartalmazó szavak ellenőrzése**

Az Eszközök→Beállítások párbeszédablakban a Nyelvi beállítások→Írástámogatás lapon egyéb beállításokat is megtehetünk. A nagybetűs és számokat tartalmazó szavak ellenőrzését a Beállítások dobozban lehet kérni. Ezt igényesebb dokumentumok ellenőrzése esetén érdemes megtenni.

(A korábbi változatok a nagybetűs szavakat helyességtől függetlenül hibásnak jelezték, ha kértük az ilyen szavak ellenőrzését. Ez a probléma az OpenOffice.org 1.02 magyar változatával megszűnt.)

### **2.1.5. Alapértelmezett nyelv beállítása**

Az Eszközök→Beállítások párbeszédablakban a Nyelvi beállítások→Nyelvek lapon állíthatjuk be az OpenOffice.org dokumentumok alapértelmezett nyelvét. Jelöljük be a Csak az aktuális dokumentumban jelölőnégyzetet, ha csak az éppen szerkesztett dokumentum alapértelmezett nyelvét szeretnénk megváltoztatni!

### **2.1.6. Automatikus elválasztás beállítása**

A helyesírás része a helyes elválasztás. Ennek automatikus változatát az Eszközök→Beállítások párbeszédablakban a Nyelvi beállítások→Írástámogatás lapon vá-

laszthatjuk ki. Kattintsunk a Szerkesztés gombon, a megjelenő párbeszédablakban válasszuk ki a magyar nyelvet, majd az ALTLinux LibHnj Hyphenator jelölőnégyzetet válasszuk ki.

Az OpenOffice.org magyar változatai között akadnak olyanok is, ahol az előző beállítás nem működik (könnyen ellenőrizhető: kattintsunk újra a Szerkesztés gombon, ha az előző jelölőnégyzet újra nincs kiválasztva, a beállítás nem lehetséges). Ilyenkor ha mégis elválasztást szeretnénk, jelöljük ki az egész dokumentumot, és a Formátum→Bekezdés párbeszédablakban válasszuk ki a Szövegfolyam lapot, és jelöljük be az Elválasztás szakaszban az Automatikusan jelölőnégyzetet!

### 2.1.7. Helyesírási állományok frissítése

A helyesírási állományokat (hu\_HU.dic, hu\_HU.aff, hyph\_hu.dic) a Magyar Ispell honlapról (<http://magyarispell.sf.net>) tölthetjük le, és az OpenOffice.org share/dict/ooo (régbben user/wordbook) könyvtárába másoljuk át kibontás után. Az OpenOffice.org újraindítás után már az új állományokat fogja használni.

### 2.1.8. Az OpenOffice.org nem magyar változatai

#### Különbségek

Az OpenOffice.org nem magyar változatai nem tartalmazzák az OpenOffice.org Myspell nevű helyesírás-ellenőrzőjének magyar bővítését (amelyen tulajdonképpen a Hunspell program is alapszik), emiatt sok hibát követnek el a helyesírás ellenőrzésében, különösen az összetett szavak kezelése terén. Például nincs korlátozva az összetett szavak hossza; nem ismeri fel az olyan tipikus hibákat, mint a *színtű*, *karvaj*, *rovarirtószer*; sok helyes szót sem fogad el, mint például *kéthavi*, *Sydneyvel*; nem kezeli a kötőjeles szavakat, mint a *MÁV-ot*, *Voltaire-rel* stb.

Könnyen meggyőződhetünk róla, hogy az OpenOffice.org nem magyar változatával van dolgunk: írjuk be a szövegszerkesztőbe, hogy *helyesírásellenőrző*. A nem magyar változatok helyesnek fogadják el a szót, ha egyébként a magyar helyesírási szótár lett kiválasztva az Eszközök→Beállítások párbeszédablakban a Nyelvi beállítások→Írástámogatás lapon. A magyar változatok hibásnak jelzik a szót, és a *helyesírás-ellenőrző* alakot javasolják helyette.

#### Helyesírási állományok telepítése

Ha mindezek ellenére a nem magyar változat használata mellett döntünk, szükségünk lehet a helyesírási állományok (hu\_HU.dic, hu\_HU.aff, hyph\_hu.dic) kézi telepítésére, ha ezeket nem tartalmazza az OpenOffice változat. Az említett állományokat a Magyar Ispell honlapról (<http://magyarispell.sf.net>) tölthetjük le, és az OpenOffice.org megfelelő könyvtárában kell elhelyeznünk.

A könyvtár helye az újabb változatokban a share/dict/ooo. Az itt található *dictionary.lst* állományba kell még bejegyeznünk a helyesírási állományokat a

DICT hu HU hu\_HU

HYPH hu HU hyph\_hu

sorokkal. Az OpenOffice.org indítása után az Eszközök→Beállítások párbeszédablakban a Nyelvi beállítások→Írástámogatás lapon válasszuk ki a Szerkesztés gombot, és a megjelenő lapon a magyar nyelvhez kapcsoljuk be az OpenOffice MySpell SpellCheckert és az ALTLinux LibHnj Hyphenatort, és az alapértelmezett nyelvet is módosítjuk a korábban ismertetett módon, a Nyelvi beállítások→Nyelvek lapon.

## 2.2. A Hunspell telepítése és használata

### 2.2.1. Linux és egyéb Unix-típusú rendszerek alatt

A Hunspell már része egyes Linux terjesztéseknek, például az UHU-Linuxnak (az 1.0-s változatban még Mispell néven). A legfrissebb változatokat forráskódból fordíthatjuk le.

#### Fordítás

Töltsük le a <http://magyarispell.sf.net> oldalról a legfrissebb Hunspell csomagot.

Telepítsük a következő GNU, vagy ezzel egyenértékű fejlesztői csomagokat: glibc-devel, make, libncurses5-devel, readline-devel, gettext-devel.

Csomagoljuk ki, és fordítsuk le a Hunspell forrást:

```
tar xvzf hunspell.tar.gz
cd hunspell
make
make install
```

Amennyiben a telepítési útvonalak nem megfelelőek, a Makefile állományban pontosíthatunk ezeken.

Ellenőrizzük a program működését a

```
hunspell latin12.html
```

vagy a

```
hunspell oomagyarispellteszt.txt
```

parancs kiadásával.

## Fordítás menüs felület nélkül

A hunspell.cxx állományban kapcsoljuk ki a CURSES makrót, megjegyzésbe téve, vagy törölve a

```
#define CURSES
```

sort. Ez esetben a menüs felület nem fog működni, csak a parancssori. Fordítsuk és telepítsük a programot az előző szakaszban ismertetett módon.

## A Hunspell használata

A program és szótárállománya formátumának leírását a két feltelepített kézikönyvoldal (Hunspell(1), Hunspell(4)) tartalmazza.

A telepítés során kiadott `make install` parancs az Ispell parancshoz is a Hunspell programot állítja be. Az eredeti Ispell `ispell.orig` néven indítható a továbbiakban.

A személyes szótárát a `$HOME/.hunspell_default`, vagy ha nem az alapértelmezett magyar szótárakat használjuk, akkor a `$HOME/.hunspell_[nyelv]` fájlba írja a program.

A saját szótár később kézzel is átszerkeszthető, érdekes tulajdonsága, hogy ragozható szavakat is felvehetünk ide.

A Hunspell az Ispellhez hasonlóan együttműködik az Emacs, LyX, KWord programokkal. A KWordhöz való beállításban Szántó Tamás leírása segít. A leírás a <http://magyarispell.sf.net> oldalon található.

Ha a Hunspell programot állománynév paraméterekkel hívjuk meg, sorban megnyitja az állományokat, és a hibás szavakat egy menüs felület segítségével felkínálja javításra.

A kiterjesztés alapján felismeri a szöveges,  $\text{T}_{\text{E}}\text{X}$ , Nroff/Troff és HTML állományokat, de kapcsolókkal is befolyásolhatjuk ezt: `-t` ( $\text{T}_{\text{E}}\text{X}$ ), `-H` (HTML), `-n` (Nroff/Troff).

A felület további ismertetését a függelékben található Hunspell kézikönyvoldal tartalmazza. Itt csak a használható billentyűkombinációk leírását idézzük. Zárójelben megadjuk a billentyűparancsok angol megfelelőit is.

- *Szóköz*: Tovább lépés javítás nélkül a következő hibás szóra.
- *Számok*: A megfelelő javaslat kiválasztása.
- *c* (*r*): Csere. A hibás szó cseréje a megadott szóra, vagy szavakra. Az Escape billentyű kétszeri lenyomásával megszakíthatjuk a a cserét.
- *j* (*a*): A szó elfogadása a program futásának befejeződéséig.
- *f* (*i*): A szó felvétele a saját szótárba.
- *k* (*u*): A szó kisbetűsként való felvétele a saját szótárba.

- *t* (-): Ragozott tőszó felvétele. A ragozás a másodikként megadott, mintául szolgáló szó alapján történik. Ha ez a szó nincs felvéve a szótárba, a tőszófelvételt előlről kell kezdeni.

Példa: gyakran hiányoznak az idegen és a mozaikszavak. Vegyük fel a hiányzó OTP szót a tv köznévi mozaikszó toldalékaival! Lenyomjuk a t billentyűt, beírjuk, hogy OTP, Enter, beírjuk, hogy tv, és Enter. A tv szó ugyanis már szerepel a szótárban, és ragozása mondja meg a programnak, hogy az OTP szót milyen toldalékokkal szeretnénk látni, pl. OTP-hez, OTP-met stb. a tv-hez, tv-met mintájára. A -val/-vel, -vá/-vé toldalékokat a kötőjellel kapcsolt toldalékú szavak esetében külön kell felvenni, de még így is sokkal kényelmesebb a speciális tőfelvétel, mint minden (akár több száz) toldalékos alakot külön-külön felvenni.

Egy kis segítség a minta kiválasztásához: első menetben a hangrend számít. Pl. mozaikszavaknál a mély hangrendű új szavak felvételénél használhatjuk a MÁV szót példaként, ajakréses magas hangrendnél pedig az említett tv-t. Ajakkerekítéses magas hangrend esetében a HÖK szót használhatjuk (amíg át nem lesz javítva a helyes hök-re). Ha a hiányzó szó nem mozaikszó, hanem például tulajdonnév, a Pál, Péter, Örs szavakat adhatjuk meg példának a hangrendtől függően. Ha meg egyszerű köznévről van szó, akkor használjuk az ok, ék, ük szavakat. Igéknél áll, vél, ül, ikés igéknél ázik, vérzik, nőzik lehet a példa (remélhetőleg a következő változatokban egyszerűsödik ez a tőfelvétel).

Az Escape billentyű kétszeri lenyomásával megszakíthatjuk ezt a műveletet.

Az itt felvett szavak automatikusan a saját szótárba kerülnek. Ha ezt innen törölni szeretnénk, a programból kilépve a saját szótár kézi átszerkesztésével tehetjük meg.

- *m* (q): Kilépés a változtatások mentése nélkül. A program változtatás megléte esetén megerősítést kér.
- *v* (x): Az állomány ellenőrzésének megszakítása a változtatások mentésével. Ha van következő állomány, akkor a program annak ellenőrzését kezdi el.
- ? (?): Segítség. Hasonló rövid leírás megjelenítése a billentyűparancsokról.

### 2.2.2. Windows alatt

Töltsük le a <http://magyarispell.sf.net> oldalról a Windowsra készített csomagot, és csomagoljuk ki a C: meghajtó gyökerében a WinZip programmal.

A kicsomagolás után a futtatható program (Hunspell) és a szótárállományok a C: Myspell könyvtárba kerülnek.

Nyissunk meg egy Parancssor ablakot, majd adjuk ki a C:

Myspell

Hunspell parancsot. A beírt szavakat, illetve mondatokat a Hunspell feldolgozza, és a hibás szavakat &, vagy # jellel kezdve írja ki. (A kimeneti formátum részletes ismertetését a függelékben közölt kézikönyvlapok tartalmazzák.)

Ha egy meglévő állományban szeretnénk a hibás szavakat megkeresni, a *-l* kapcsolót, és esetleg az állomány típusát adjuk meg, az állományt pedig a standard bemeneten keresztül adjuk a Hunspell programnak.

```
hunspell -l -H <próba.html
```

A példában a próba.html állomány hibás szavait irattuk ki a képernyőre. Ha a kimenetet el szeretnénk menteni, irányítsuk fájlba:

```
hunspell -l -H <próba.html >hibás_szavak.txt
```

A Windows alá letölthető futtatható állomány nem tartalmazza a Linux alatt alapértelmezés-ként működő menüs felületet, amely lehetővé teszi a szöveges, HTML,  $\LaTeX$  és Groff formátumú állományok gyors javítását.<sup>1</sup>

### 2.2.3. Hunspell Mac OS X alatt

Fazekas Mihály (boga@mac.com) a Hunspell 0.94-es változatát egy Cocoa felülettel bővítette, ezzel lehetővé téve a helyesírás-ellenőrző rendszerszintű használatát Mac OS X alatt. Letölthető csomagok, és részletes telepítési útmutató található a <http://myspellx.sf.com> címen.

A Hunspell program lefordítható Mac OS X alá is. Telepítsük a GNU fejlesztőeszközöket (glibc-devel, ncurses-devel, readline-devel, gettext-devel), vagy a hunspell.cxx állományban kapcsoljuk ki a CURSES makrót, megjegyzésbe téve, vagy törölve a

```
#define CURSES
```

sort. Ez utóbbi esetben a menüs felület nem fog működni, csak a parancssori.

## 2.3. Emacs

Az Emacs, illetve a Xemacs program mind Windows, mind Linux alatt képes a Hunspell programot használni.

Ismeretlen okból csak a Xemacs kezeli jól a kötőjeleket tartalmazó szavakat, így ennek használata javasolt az Emacs helyett.

### 2.3.1. Az Emacs telepítése Linuxra

Linux esetében a terjesztések tartalmazzák mindkét programot. Az alapsomag az emacs. Ha az emacs-X11 csomagot is telepítjük, az X grafikus felületen egy bővebb változat indul el, ami menüsorral és eszköztárral is rendelkezik. Ez utóbbival jelentős részben megegyezik a xemacs csomagban található Xemacs.

<sup>1</sup>A Cygwin környezet lehetővé teszi a menüs rendszer fordítását. A menüs rendszer Ncurses, Readline, és Gettext függvénykönyvtárakat használ, amelyeket az Visual C++ – szemben a Cygwinnel – alpból nem tartalmaz. A Cygwin címe: <http://www.cygwin.com>.

### 2.3.2. Az Emacs telepítése Windowsra

Windowsra több helyen is található bináris csomagot, például az `ftp://ftp.gnu.org/gnu/windows/emacs/latest/` címen. Általános esetben az `emacs-21.2-bin-i386.tar.gz` csomagra van szükség, melyet az `ftp://ftp.gnu.org/gnu/windows/emacs/utilities/i386/` címen található programokkal vagy a WinZIP programmal is kibonthatunk.

Egy másik lelőhely a CTAN archívum, például: `ftp://ftp.dante.de/tex-archive/systems/win32/fptex/0.5/support/emacs-21.1.1-win32.zip`.

### 2.3.3. A Xemacs telepítése Windowsra

A `http://www.xemacs.org` címről tölthető le a Xemacs windowsos változata.

### 2.3.4. A Hunspell telepítése az Emacs alá

Az Emacs és Xemacs programok a háttérben indított Hunspell-lel együttműködve képesek a helyesírás-ellenőrzésre.

A Hunspell-nek a keresési útvonalon kell elhelyezkednie, méghozzá `ispell` néven. A Linuxra történő telepítés során ez megtörik. Windows alatt készítsünk a C:

Myspell

Hunspell programról egy másolatot `ispell` néven valamelyik rendszerkönyvtárba (például C:

Windows, C:

WinNT).

Windows, és az Emacs X-en futó változatai (Emacs-X11, Xemacs) esetében további beállításra nincs szükség. Az `M-x ispell` parancs kiadása után az Emacs elindítja a Hunspell programot, és ellenőrzi a segítségével az éppen szerkesztett szöveget.

A Xemacs esetében a `M-x ispell-buffer` parancsot adjuk ki elsőnek, majd a következő ellenőrzések során már használható `M-x ispell` is.

### 2.3.5. A Hunspell telepítése konzolon indított Emacshez

Az Emacs (tehát nem Xemacs) konzolon való használathoz szükségünk lehet az `ispell.el` módosított változatára, bár inkább javasolt helyette a Xemacs konzolon történő indítása.

Az `ispell.el` állomány a `http://magyarispell.sf.net/ispell.el` címen található meg. A le-töltött és kibontott állományt másoljuk abba a könyvtárba, ahol a lefordított `ispell.elc` található (például `/usr/share/emacs-21.1/lisp/textmodes/` könyvtár, vagy régebben `/usr/share/emacs/20.7/lisp/`).

Ezután töröljük az ebben a könyvtárban található `ispell.elc` állományt!

A telepített `ispell.el` magyar nevű szótárállományokat keres, ezért győződjünk meg arról, hogy a `default.aff`, `default.dic` állományok ilyen néven is elérhetők (Windows alatt a C:

Myspell, Linux alatt a /usr/share/myspell könyvtárakban). Ha nem, akkor készítsünk másolatot (vagy kötést) ezekről magyar.aff és magyar.dic néven ugyanabban a könyvtárban.

Adjuk ki az M-x ispell-change-dictionary parancsot, és válasszuk ki a magyar szó beírásával a magyar Ispell szótármodult.

Ha magyar az alapbeállításunk, természetesen az Emacs indító konfigurációjába (vagyis a ~/.emacs fájlba) is beleírhatjuk:

```
(ispell-change-dictionary "magyar")  
(ispell)  
(ispell-kill-ispell)
```

A folyamatos ellenőrzés alatt használható billentyűparancsok megegyeznek a Hunspell angol billentyűparancsaival.

A saját szótárba felvett szavak a ~/.hunspell\_default, a módosított ispell.el használata esetén a ~/.hunspell\_magyar állományba kerülnek.

Mínt hogy a Spell-menü tartalmának alapbeállítása az emacs programba bele van fordítva, indításkor az nem tartalmazza a magyar lehetőséget, viszont az ispell legalább egyszeri indítása után már igen, tehát ekkor már a menü is használható a parancssori utasítások helyett: a 20. verzió esetében Edit→Spell, a 21. verzió esetében pedig Tools→Spell Checking.

### 2.3.6. Az Emacs használata

A teljes szöveg ellenőrzésére szolgál az M-x ispell parancs.

Az itt használható billentyűparancsok megegyeznek a Hunspell angol billentyűparancsaival.

Egy szót az M-x ispell-word parancssal ellenőriztethetünk. Ehhez létezik gyorsbillentyű is: M-\$.

Az Emacs *flyspell* üzemmódja lehetőséget nyújt a beírás közbeni helyesírás-ellenőrzésre, de a magyar beállításokkal nem működik tökéletesen együtt. Az üzemmódba az M-x flyspell-mode parancssal léphetünk be.

Ezután a hibás szavak grafikus felületen piros színűvé és aláhúzottá válnak. Ha a 3. egérgombbal (vagy a két szélsővel egyszerre) kattintunk a szón, felnyílik egy menü, amellyel a javasolt cseréket fogadhatjuk el, illetve továbbléphetünk, felvehetjük a hibás szót az elfogadott szavak közé az ellenőrzés idejére, illetve „végérvényesen”.

## 2.4. Az Ispell telepítése és használata

Az Ispell a Hunspell képességeit tekintve elavult, és magyar nyelvű szövegek helyesírás-ellenőrzésére csak végszükség esetén, a Hunspell hiányában javasolt. A különbségeket a függelékben található Hunspell kézikönyvoldal foglalja össze.



### 2.4.1. Telepítés Linuxra

Az Ispell része a legnagyobb Linux terjesztéseknek, és egy átlagos Linux telepítésnél az Ispell telepítésére is sor kerül automatikusan.

Linux alatt az Ispell szótárállományai a `/usr/lib/ispell` állományban kerülnek elhelyezésre. A magyar Ispell szótár, vagyis a `(magyar.hash)` állomány is ide kerül.

### 2.4.2. Telepítés Linuxra, forráskódból

Az `ispell-honlapról` (<http://fmg-www.cs.ucla.edu/fmg-members/geoff/tars/>) letölthető forráscsomag a benne lévő README fájlban leírtak szerint (`make all ; su ; make install`) lefordítható. A `local.h` fájl linux esetében javasolható módosításai a

```
#define NO8BIT /* Remove this if you use ISO character sets */
```

sor törlése és a

```
#define USG /* Define on System V or if term.c won't compile */  
#define LIBDIR "/usr/lib/ispell"
```

sorok eredetijeinek átírása a közölteké.

Némely rendszerek esetében szükség lehet a `/usr/lib/` könyvtárban a `# ln -s libtermcap.so.2 libtermcap.so` kötés létrehozására.

Ezután már csak `make` és `- rootként - make install`, mely parancs a `/usr/lib/ispell/`-be teszi az angol adatfájlokat és `/usr/local/bin/`-be a programokat.

### 2.4.3. Windows

A Windowsra történő telepítéshez Geoff Kuenning Ispell honlapján (<http://fmg-www.cs.ucla.edu/fmg-members/geoff/ispell-winnt.html>) több lehetőséget is találunk.

Windows alatt két fő módja van az Ispell telepítésének és használatának. Az elsőben a teljes Cygwin környezetet telepítjük a Windows NT, 2000, illetve XP operációs rendszerünkre. Ez gyakorlatilag teljes Unix környezetté egészíti ki a Windowst, és utána telepítünk egy Cygwin környezet alá fordított Ispell változatot.

A másik módszer azok számára, akik a szokásos Windows környezetet előnyben részesítik, egyszerűbb: ez a bináris csomag formájában letölthető WIN32-re fordított Ispell változat Cygwin nélküli használata.

Mindkét módszer esetében az [ftp://ftp.franken.de/pub/win32/develop/gnuwin32/cygwin/%porters/Humblett\\_Pierre\\_A/V1.1/](ftp://ftp.franken.de/pub/win32/develop/gnuwin32/cygwin/%porters/Humblett_Pierre_A/V1.1/) címen található verziót javasoljuk. A telepítést az `ispell-3.2.06-cygwin-1.3-bin.README` nevű fájl ismerteti. Az egyszerűbb esetben a csomagból kizárólag a `usr/local/bin/ispell.exe` fájlra van szükség, melyet a PATH-ban található valamelyik

könyvtárba kell elhelyezni. A szótárfájlok (lásd ??) a DICTDIR környezeti változó által meghatározható helyre teendők. (A .tar.gz végződésű fájl kicsomagolásához a Cygwin környezetben kívül pl. a djarnt.exe program, vagy akár a népszerű WinZIP újabb változatai használhatók.)

Ha tehát a C:\AUTOEXEC.BAT fájlban a

```
SET PATH=C:\WINDOWS;C:\WINDOWS\COMMAND;C:\BIN
SET DICTDIR=C:\ISPELL
```

sorok találhatók, akkor az ispell.exe fájlt a c:\bin, a magyar.hash és a magyar.aff fájlt a c:\ispell alkönyvtárba kell helyezni.

Szintén a PATH-ban kell elhelyezni a cygwin1.dll nevű fájlt.

Egy másik bináris csomag formájában letölthető WIN32-re fordított Ispell változat található a <ftp://ftp.tue.nl/pub/tex/GB85/ispell-dutch96/ispellw32.zip> címen.

A letöltés után csomagoljuk ki a ZIP állományt a C:\ meghajtó gyökerében, és adjuk a keresési útvonalunkhoz a C:\ispell\bin könyvtárat. Ehhez nyissuk meg a Start menü→Beállítások→Rendszer párbeszédablakot, és a Speciális lapon a középső részben (Környezet) keressük meg a keresési útvonalakat rögzítő környezeti változót (PATH), és írjuk a végére pontosvesszővel elválasztva a C:\ispell\bin útvonalat.

Sajnos ez a dán Ispell modulhoz készült csomag még kis számú ragozási osztállyal lett lefordítva, ezért ha a legfrissebb Magyar Ispell szótármodult szeretnénk használni vele, újra kell fordítanunk az Ispellt. Ehhez az EMX+gcc csomagokra van szükség.

Egy harmadik win32-változat a CTAN:fp<sub>tex</sub> (<ftp://ftp.dante.de/tex-archive/systems/win32/>) disztribúcióban található.

#### 2.4.4. A magyar.hash állomány telepítése és használata

A magyar.hash állomány, vagyis a magyar Ispell szótár része több Linux terjesztésnek is (különböző neveken: ispell-hu, ihungarian stb.).

A <http://magyarispell.sf.net/> oldalról töltsük le a legfrissebb változatát a Magyar Ispell csomagnak (magyarispell-XXX-tar.gz), majd adjuk ki a következő parancsokat:

```
tar xvzf magyarispell-XXX.tar.gz
cd magyarispell-XXX
make ispell
cp tmp/magyar4ispell.hash /usr/lib/ispell/magyar.hash
```

Windows esetében a magyar.hash állomány a 2.4.3 pontban leírt helyre kell tenni.

Ha a magyart alapértelmezett nyelvként szeretnénk látni, akkor gondoskodjunk arról, hogy a szótárállományokat tartalmazó könyvtárban a default.hash a magyar.hash-ra mutasson.

```
rm /usr/lib/ispell/default.hash
ln -s /usr/lib/ispell/magyar.hash /usr/lib/ispell/default.hash
```

## 2.4.5. Az Ispell változat megfelelősége

Ahogy már említésre került, az Ispell számos fordítás során beállított paramétertől függően végzi tevékenységét. Különösen régebbi Linux terjesztésekben, vagy Windows alatt meg kell győződnünk arról, hogy az Ispell képes helyesen lefordítani és kezelni a Magyar Ispell forrásállományait.

Adjuk ki Linux/Unix alatt a következő parancsot:

```
ispell -vv | grep BIT
```

Ha a bináris Ispell megfelel a céljainknak, akkor a következő két sort kell látnunk:

```
MASKBITS = 64  
!NO8BIT (8BIT)
```

Az első sor jelzi, hogy elegendő számú ragozási osztályt kezel az Ispell (a régebbi Ispell változatoknál a 32 volt az alapértelmezett).

A második sor jelzi, hogy a szótárban található szavak tartalmazhatnak ékezetes betűket is.

Windows alatt a következő parancsot adjuk ki a parancssorban:

```
ispell -vv | more
```

A listában keressük meg a fenti két sort. Ha a MASKBITS értéke 32, vagy a NO8BIT értéke igaz, akkor a Magyar Ispell használatához szükség van az Ispell újrafordítására.

## 2.4.6. Az Ispell használata

### Szöveges állományok

A következő parancssal kérhetjük szövegfájlok ellenőrzését:

```
ispell -d magyar fájl(ok)
```

Ebben az esetben megjelenik egy kényelmes és gyors menü, amely sorban felmutatja a hibásnak vélt szavakat.

Ha van javítási javaslat, akkor a javaslat előtt látható számjegyek leütésével fogadjuk el a javaslatot.

A legfontosabb billentyűparancsok a következők:

- szóköz: Átugrik a szón, azt változatlanul hagyva.

- számok: Választás a javasolt cserék közül.
- *r*: Csere az általunk megadott szóra.
- *a*: Elfogadjuk a szót az ellenőrzés befejeződéséig.
- *i*: Felvesszük a saját szótárunkba. (Ez a `~/ispell_magyar`, vagy ha létezik az aktuális könyvtárban az `.ispell_magyar` állomány, akkor az. Később – ha elhagytuk az Ispell programot – ez akár kézzel is módosítható.)
- *u*: Felvesszük a saját szótárunkba, de kisbetűsre alakítjuk előtte.
- *x*: A következő fájlra ugrunk. A félig ellenőrzött állománynak a fennmaradó része nem lesz kijavítva.
- *l*: A rendszerszótárban végezhetünk kereséseket. A `*` jel segítségével mintázatot is megadhatunk, hasonlóan a fájlrendszerben megadott útvonalakhoz: pl. az `L*` mintázat minden `L` betűvel kezdődő szót kilistáz a `/usr/dict/words` állományból. A rendszerszótár angol nyelvű.
- *q*: Kilépés mentés nélkül, ha a „q” után „y”-t nyomunk megerősítő kérdésre. A vizsgált állomány javításai elvesznek.

A Hunspell program hasonló felülettel rendelkezik Linux alatt, de több újdonsággal is szolgál a magyar nyelv számára. Leírást a Hunspell(1) kézikönyv oldalon találunk.

## **T<sub>E</sub>X állományok**

A következő paranccsal kérhetjük T<sub>E</sub>X, illetve L<sup>A</sup>T<sub>E</sub>X állományok ellenőrzését:

```
ispell -t -d magyar fájl(ok)
```

Amennyiben az állomány(ok) kiterjesztése `.tex`, a `-t` kapcsolót nem szükséges megadni.

A `-t` kapcsoló hatására az Ispell nem törődik a fordított perjellel kezdődő, ékezetes betűket nem tartalmazó szavakkal (vagyis a TeX parancsokkal). Bizonyos parancsokat felismer, és az utána következő egy, vagy két argumentummal sem foglalkozik. A matematikai környezetben lévő szavakat nem ellenőrzi az Ispell.

## **HTML állományok**

A következő paranccsal kérhetjük HTML állományok ellenőrzését:

```
ispell -H magyar fájl(ok)
```

## 3. fejezet

# Fejlesztői dokumentáció

### 3.1. Helyesírás-ellenőrzők ismertetése

Számos helyesírás-ellenőrzésre szolgáló program, illetve függvénykönyvtár áll a szabadszoftver-fejlesztők rendelkezésére.

#### 3.1.1. Aspell

Angol nyelvterületen a legkedveltebb helyesírás-ellenőrző szabad szoftver az *Aspell*, Kevin Atkinson munkája. Az *Aspell* az angol nyelv fonetikai információinak ismeretében egy speciális algoritmussal képes nagymértékű alaki eltérések esetén is helyes javaslatot tenni.

Sajnos a programból hiányzik a nagy mennyiségű ragozási szabály hatékonyan kezelésének képessége, ezért nem megfelelő a magyar és sok egyéb nyelv számára sem.

#### 3.1.2. Ispell

Az *Ispell* egyenes ági leszármazottja R. E. Gorin 1971-ben, PDP-10-re assemblyben írt *Spell* programjának. A *Spell* Pace Williamson írta át C nyelvre 1983-ban. Geoff Kuenning 1987 és 1988 között nagyobb változtatásokat eszközölt a kódon, amelynek számunkra legjelentősebb lépése az affixum tömörítési technika bevezetése volt, az affixum tömörítéssel ugyanis lehetővé vált nagyszámú ragozási szabály futásidejű hatékony kezelése is. Gyakorlatilag ennek köszönhető, hogy az *Ispell*hez több tucat nyelv helyesírási szótármodulja készülhetett el, köztük olyan agglutináló nyelveké, mint a finn és a magyar.

A magyar nyelv vonatkozásában az is kiemelhető, hogy az *Ispell* szabályozott szóösszetétel-képzési lehetőséggel rendelkezik, vagyis a szóösszetétel-képzést korlátozhatjuk a szavak egy bizonyos részhalmazára.

Az *Ispell* egyéb kiemelhető tulajdonságai:

- De facto szabvány a szabad szoftverek világában. A Free Software Foundation GNU fej-

lesztésének alapértelmezett helyesírás-ellenőrzője, és a legtöbb Linux terjesztésben is megtalálható.

- A csőben használható felületének köszönhetően számos programba „beágyazható” (pl. Emacs, Kword, LyX).

### 3.1.3. Pspell

A Pspell függvénykönyvtár Kevin Atkinson kezdeményezése, ami az Aspell és az Ispell előnyeit hivatott a forráskód egybefogásával biztosítani.

### 3.1.4. Myspell

A Myspell az OpenOffice.org irodai programcsomaghoz készült BSD licencű helyesírás-ellenőrző függvénykönyvtár.

2001-ben kezdte fejleszteni Kevin Hendricks azzal a szándékkal, hogy a GPL licenc alatt is elérhetővé tett, de a licencelt szoftverkomponensektől (mint a nyelvi modulok) is megszabadított StarOffice – új nevén OpenOffice.org – helyesírás-ellenőrzővel rendelkezessék.

A Myspell egy kicsi, C++-ban megvalósított program, amely az Ispell tömörítési algoritmusán alapszik és kis változtatással képes az Ispell szótármodulok használatára is.

A Myspellnek az Ispell-lel szemben több határozott előnye van:

- Kicsi és áttekinthető: mivel csak a lényegre tartalmazza, kiváló kiindulási alapot jelent egy professzionális magyar helyesírás-ellenőrző program elkészítéséhez.
- A C++ nyelv megkönnyíti a keresztplatformos C++-alapú fejlesztésekhez való integrációját.
- Nagy mértékben hordozható, mivel eleve Unix/Linuxra és Windowsra fejlesztik egyidejűleg.
- Szintén a hordozhatósággal kapcsolatos, hogy a Myspell futásidőben állítja elő két egyszerű szöveges állományból – a ragozási szabálygyűjteményből (affixum állomány), és a szótárállományból – a működése során használt speciális adatszerkezeteket. Ezzel szemben az Ispell bináris adatállományokat használ, sőt az Ispell program fordítási idejű beállításaitól függ, hogy futásidőben képes-e feldolgozni ezt a bináris adatállományt, vagy sem.
- Egy Unix/Linux, és Windows operációs rendszerek alá készült professzionális irodai programcsomag, az OpenOffice.org része, amely szabad szoftverként minden felhasználó számára elérhető ingyenesen, jogtisztán és korlátozások nélkül.
- Az említett vonzó tulajdonságok miatt több közreműködő is folyamatosan fejleszti a Myspell-t. Így került bele a Myspell kódjába egy jelentős sebességoptimalizáció, egy a javaslattevés minőségét nagy mértékben növelő megoldás (javítási cseretáblázat), valamint a hiányzó szóösszetétel-képzési lehetőség és több kisebb hiba javítása.

- Az említett szóösszetétel-képzési kiegészítés javítja az Ispell egy bosszantó (bár kevésbé feltűnő) hibáját: az Ispell egy szó ragozott alakját akkor is elfogadja egy szóösszetételben, ha amúgy a tőszót nem (pl. kutyaokos hibás, de kutyaokosnak elfogadott szóösszetétel).

### 3.1.5. Mozilla Spell Checker

A Myspell forráskódján alapuló ellenőrző. David Einstein munkája. A Mozilla Spell Checker jelentős sebességoptimalizációja visszakerült a Myspellbe, és a tervek szerint a két ellenőrző egyesülni fog a közeljövőben.

### 3.1.6. Magyar Myspell

A Myspell forráskódján alapuló GPL licencű ellenőrző, amit 2002-ben kezdtem fejleszteni.

Először kisebb hibajelentések illetve foltok készültek a Myspellhez, de később rákényszerültem az idő és a támogatni kívánt nyelvek sajátosságai ismeretének hiányában egy külön ág létrehozására.

A Magyar Myspell fejlesztés eredményeként került az eredeti Myspell ágba, és az így az OpenOffice.org-ba is a szóösszetétel-képzési lehetőség, valamint a javaslattevés minőségét jelentős mértékben emelő javítási cseretáblázat-kezelés.

A külön Magyar Myspell ág a következő fontosabb tulajdonságokkal rendelkezik:

- A magyar helyesírás szabályai 138. számú kötőjellel írási szabályához (3-6-os szabály) igazított szóösszetétel-ellenőrzés.
- Javaslattevés a szótárban nem szereplő szóösszetételek esetén is.
- A szókincs futásidejű bővítésének lehetősége.
- Az újonnan felvett szavak továbbragozása egy mintaként megadott tőszó alapján.
- Javított javaslattevés nagybetűt tartalmazó szavakra.
- A mássalhangzó-triplázások letiltása (sakkör).
- Cső felület a beágyazáshoz.
- Ékezet nélküli szavak esetén helyes javaslattevés.
- Hibásan írt szavak helyes szóösszetételként való elfogadásának megakadályozása. Az egyes tőszavak szóösszetételben való megjelenésének (és ezzel teljes szóbokrok) letiltása helyett a javítási cseretáblázat (esetleg a teljes javaslattevő modul) felhasználásával megvizsgálható, hogy nem tipikus hibáról van-e szó ritka szóösszetétel helyett. (Ilyen hibás összetételek, amelyek korábban csak az egyik tőszó teljes letiltásával vagy sehogy nem voltak lekezelve: szervíz, vizitorma, birság, elitélt, vasárú, tejles, szívéjes, karvaj, stb.)

Számos egyéb tulajdonság megjelenése a közeljövőben várható:

- Kötőjeles és nagyköötőjeles alakok ellenőrzése a kötőjellel kapcsolt toldalékok, ikerszók, a mozgószabály alapján egybeírt szóösszetételek, földrajzi nevek, számnevek és nagyköötőjelet tartalmazó összetételek esetében.
- Szóismétlések kiszűrése (esetleg opcionálisan).
- Kötelező egybeírásra való figyelmeztetés (pl. legalábbis, véghezvisz).

### 3.1.7. OpenOffice.org

Nem kifejezett helyesírás-ellenőrző, de a Magyar Myspell fejlesztésének egyik kiemelt célpontja.

Az OpenOffice.org alapját képező StarOffice a német StarDivision cég kereskedelmi terméke volt a 90-es években. A Sun Microsystems a céget és termékét felvásárolta üzleti megfontolásokból. A cég 2000-ben GPL licenc mellett szabad szoftverré tette a StarOffice nagy (a StarDivision által fejlesztett) részét. A licencelt nyelvi eszközöktől, és adatbázis-technológiától megfosztott program az OpenOffice.org nevet kapta (az OpenOffice védett név volt, ezért a furcsa, a nyitott fejlesztés honlapjának címét hordozó elnevezés). A Sun jelenleg kereskedelmi támogatással vállalati ügyfelek számára az OpenOffice.org-on alapuló StarOffice-t is biztosítja.

Az OpenOffice.org jellemzői:

- Professzionális irodai programcsomag
- Kiváló Microsoft Office export, import szűrők
- Nagy méretű, sok képet tartalmazó dokumentumok gyors kezelése.
- GPL-es: ingyenes, módosítható, terjeszthető (illetve kettős licenclésű).
- XML alapú saját fájlformátum (Java jar formátumú állományokban).
- UNICODE támogatás, még ázsiai nyelvekhez is (mindkettő Sun fejlesztés)
- A GPL licenc alá nem helyezhető licencelt részek fokozatos kiváltása (*lingucomponent* kísérleti fejlesztés: Myspell helyesírás-ellenőrző, ALTLinux elválasztó modul, szinonimaszótár, nyelvhelyességi ellenőrzés).
- Honosított: magyar helyesírás-ellenőrzés a Magyar Myspell helyesírás-ellenőrző, és -javító programmal, és a Magyar Ispell szótármodullal. Magyar elválasztás Mayer Gyula és Miklós Dezső huhyph.tex T<sub>E</sub>X elválasztási moduljával. Mintegy 20 000 programfelirat és -üzenet magyar fordítása.

A számos új funkció miatt is az OpenOffice.org sokáig fejlesztői stádiumban volt. Mivel azonban a program meglepően üzembiztos és jól használható, 2002-ben (marketing okokból is) kibocsátásra került az 1.0-s elnevezésű változat.



A fejlesztői változat jelenleg még az eredeti számozást használja (641D), de ez a számozás várhatóan a közeljövőben megszűnik.

A Magyar OpenOffice.org változatai a <http://office.fsf.hu> oldalon jelennek meg, a fejlesztői változatok a <http://office.fsf.hu/work> könyvtárban találhatóak. A Magyar OpenOffice.org windowsos és linuxos változatai a OOo változatszám mellett egy csomagszámot is kapnak (egymástól függetlenül). Tehát pl. a legfrissebb linuxos Magyar OpenOffice.org teljes „neve”: Magyar OpenOffice.org 1.0, 43-as számú linuxos csomag.

## 3.2. Az Ispell felépítése és működése

Az Ispell adatállományai .hash kiterjesztéssel rendelkeznek, és Unix/Linux alatt a /usr/lib/ispell/ vagy /usr/local/lib/ispell/ könyvtárban kerülnek elhelyezésre.

Egy ilyen hash állomány (pl. a magyar.hash) két fő részből áll: tartalmazza a (ragozási) osztályokba sorolt ragozási szabályokat, és tartalmazza a tényleges hasítótáblát is. A hasítótábla tárolja a tőszavakat és minden egyes tőszónál még az azon alkalmazni kívánt ragozási osztály vagy osztályok kapcsolóit.

Egy Ispell hash-állomány a buildhash paranccsal állítható elő, amely két állományt vár bemenetként: egy ragozási szabályokat (és egyéb beállításokat) tartalmazó affixum állományt és egy szótárállományt, amely a tőszavakat és kapcsolóikat tartalmazza.

### 3.2.1. Ragozási szabályok

Az Ispell affixum állomány formátumát egy kézikönyvoldal (man 4 ispell) ismerteti részletesen. Itt most csak a ragozási szabályokat, és az affixum tömörítés módszerét vizsgáljuk meg közelebbről.

Egy ragozási szabály nem más, mint egy feltételes kifejezés, amely leírja, hogy milyen feltételek között milyen módon helyezhetünk toldalékot egy alapszóhoz.

Az Ispell ragozási szabályainak leírása emlékeztet a szabványos regex reguláris kifejezésekre, és a következő fő jellemvonásokkal rendelkezik:

- A feltétel a tőszó utolsó (prefixum esetén első) legfőljebb nyolc karakterére vonatkozik.
- A feltételben erre a legfőljebb nyolc karakterre vonatkozóan szerepelhetnek karaktertartományok is.
- A toldalékolásnál a toldalék hozzáillesztése előtt lehetőség van a tőszó végének leválasztására. Itt viszont már nem használhatunk karaktertartományokat (pontosan ismerni kell a leválasztandó karaktert vagy karaktersorozatot).

Példák, amelyek bemutatják a ragozási szabályok szintaxisát. (A kettős kereszt után sor végéig tartó megjegyzés következik, amely már nem része a ragozási szabálynak.)

Z > Z # ráz -> rázz  
 Z > ZÁL # ráz -> rázzál  
 E D Z > ENEK # edz -> edzenek, pedz -> pedzenek  
 [ÓUÚ] > VAL # ló -> lóval, apu -> apuval, bú -> búval  
 [^AE] > NAK # okos -> okosnak  
 . > KÉNT # kutya -> kutyaként, okos -> okosként  
 A > -A,ÁNAK # macska -> macskának  
 [^SD] Z | K > -IK,Z # barátkozik -> barátkozz

Az első két példa jelentése, hogy Z betűre végződő tőszó Z, és ZÁL toldalékkal rendelkező alakja is helyes (ez a kisbetűs alakokra is vonatkozik).

A harmadik példa szerint az EDZ betűsorozatra végződő szavak ENEK toldalékot is kaphatnak.

A negyedik példában karaktertartományt adunk meg az utolsó karakterre vonatkozóan: minden Ó, U, és Ú-ra végződő szó VAL toldalékot kaphat.

Az ötödik példa komplementer karaktertartományt ad meg: ha a szó nem A vagy E betűre végződik, megkaphatja a NAK toldalékot.

A hatodik példa a tetszőleges karakter jelét, a pontot mutatja be: bármely szó megkaphatja a KÉNT ragot.

A hetedik példa a tőszó végének levágását mutatja be: az A-ra végződő szavak ÁNAK toldalékot kaphatnak a szóvégi A levágásával.

Az utolsó egy kombinált példa: A nem SZIK, vagy DZIK végű, de ZIK-re végződő szavak Z toldalékot kaphatnak a szó végi IK levágásával. Így pl. a barátkozik→barátkozz helyes átalakítás, de a vakarószik→vakarózz, illetve mászik→mászz már nem.

### 3.2.2. Ragozási osztályok

Az előző példák meglehetősen keverve tartalmaztak különböző szófajú szavakra vonatkozó (és részben hibás) szabályokat. Az Ispell ragozási osztályok kialakításának lehetőségével biztosítja, hogy a tőszavakat a kategóriájuknak megfelelő ragokkal láthassuk el. A ragozási osztályok egy betűs (vagy egy karakteres) névvel rendelkeznek, amit a továbbiakban kapcsolóknak nevezünk.

Példa két ragozási osztály definiálására:

flag A:

[^L] L > LAL # pedál -> pedállal  
 [^L] L > LÁ # pedál -> pedállá  
 L L > AL # szakáll -> szakállal  
 L L > Á # szakáll -> szakállá

flag B:

L > NAK # szaval -> szavalnak  
L L > ANAK # javall -> javallanak

Az első osztály a mély hangrendű főnevek ragozásának egy részletét mutatja, az L betűre végződő szavak -val toldalékolását, míg a másik osztály a mély hangrendű alanyi igeragozás egy részletét.

### 3.2.3. Szótárállomány

Az Ispell szótárállományban (.dict fájl) egyszerűen felsoroljuk a tőszavakat, a megfelelő kapcsolókkal ellátva:

```
macska/A/E/G  
pedál/A/F  
akar/B/C  
szaval/B  
legalábbis
```

Látható, hogy egy tőszó tetszőleges számú ragozási osztályba sorolható, de akár ragozási osztályt sem kötelező megadni.

### 3.2.4. Affixum tömörítés

Minden ragozási szabályhoz egy 256 byte-ból álló tömb kerül a hash állományba. A tömb szerepe, hogy gyorsan eldönthető legyen egy ragozott szóról, hogy az adott ragozási szabály alkalmazásával állt-e elő, vagy sem. A tömb az ASCII karaktertábla elemeivel indexelhető, és visszaad egy byte-ot, ami a már említett 8 karakterpozícióra vonatkozó előfordulási információkat tartalmazza: pl. ha a tömb['A'] elem értéke binárisan 11111010, akkor a tőszóra nem alkalmazható a ragozási szabály, ha a tőszó „A” betűre végződik, vagy a kettővel előtti pozíción is „A” betű található.

A Myspell forráskódja részletesen dokumentálva tartalmazza az affixum tömörítési algoritmust. Amiért itt említésre került, az az, hogy az affixum tömörítés elnevezés nem az affixum állomány kisebb méretre való tömörítésére utal, hanem arra, hogy a lehetséges gyors adatszerkezetek közül talán a legtömörebbet sikerült Geoff Kuenningnek megtalálnia, gyakorlatilag a ragozási szabályokban szereplő karaktertartományok hatékony tárolásával és kezelésével.

## 3.3. Magyar Ispell

A Magyar Ispell egy keretrendszer, amely magyar affixum és szótárállományt állít elő az Ispell számára. Az i2myspell program segítségével ezekből az állományokból hozható létre a megfelelő Myspell affixum és szótárállomány is.

### 3.3.1. Történet

A Magyar Ispell kialakítását 1998-ban kezdtem. 1999-re a tárgyias igeragozás, néhány képzős toldalékolás, sok határozószó és névutó kivételével a helyesírási-modul hibákkal ugyan, de alapjaiban elkészült. A félkész helyesírás-ellenőrzővel 1999-ben a Közművelődésügyi Minisztérium és a Matáv által közösen meghirdetett „A Magyarországi digitális kultúráért” pályázaton is részt vettem, de sikertelenül.

1999 őszén a helyesírás-ellenőrzővel való munkát befejeztem, az eredmények 2000 nyaráig, azonosítóm megszűntéig elérhetőek voltak a <http://sol.cc.u-szeged.hu/~nemeth/> címen.

2001 őszén Szántó Tamás, a KDE grafikus felület magyar honosítója felajánlotta a segítségét az ellenőrző fejlesztésének folytatásához. Tamás folyamatosan tesztelte az ellenőrzőt, számos (ragozási) típushibát megtalálva. A tesztelések eredménye volt a szókincs jelentős bővítése is.

Tamás munkáját Bíró Árpád folytatta, elsősorban további típus- és egészen különleges rejtett hibák feltárásával, a matematikai és egyéb nyelvi modulok összeállításával és a forrásállományok ellenőrzésével.

Godó Ferenc hatalmas szójelentővel az ellenőrző alkalmassá vált a mindennapos munkára.

A Magyar Ispell fejlesztése 2002 januárjáig „titokban” folyt, ekkor jelent meg az első nyilvános változat.

A nyilvánosságra hozatalt követően csatlakozott hozzánk Trón Viktor, aki további hibák, illetve hiányosságok feltárása mellett internetes forrásokból összegyűjtötte, és a Magyar Ispell számára feldolgozta a mai és történelmi magyar helységneveket, illetve – aminek külön örülök – elsőként vállalkozott az alul dokumentált Magyar Ispell keretrendszer működésének megismerésére, és ennek eredményeképp a szótőgenerálás egyes részeinek javítására.

A Magyar Ispellből Koblinger Egmont – aki számos hasznos észrevétellel, és szójelentővel is hozzájárult a Magyar Ispellhez – készített folyamatosan (SuSE) RPM csomagokat, illetve Pásztor György Debian karbantartó DEB csomagokat.

Talán a Magyar Ispell hatására a megalakulófélben lévő FSF.hu Alapítvány az OpenOffice.org honosítása mellett döntött, és szervezésükben egy hétvége alatt a programcsomag tízezres nagyságrendű feliratait, és üzeneteit kerültek lefordításra. Jelenleg a magyar OpenOffice.org linuxos változatának Noll János, windowsos változatának Bencsáth Boldizsár a karbantartója (<http://office.fsf.hu/>).

### 3.3.2. Alapvető problémák

Az agglutináló nyelvekhez nem magától értetődő az Ispell szótármodulok elkészítése. A legjelentősebb megoldandó problémák a következők voltak a Magyar Ispell kidolgozása során:

1. Az Ispellben nincs többszörös toldalékolási lehetőség. Míg a magyar nyelv gazdag a képzőkben, jelekben, és ragokban, addig az Ispell csak a tágabb értelemben vett ragok kezelésére képes. A lehetséges szó végi hasonulások, képzők, jelek és ragok kombinációi miatt igen nagy számú ragozási szabály megadása szükséges, amit „kézzel” helyesen és belátható idő alatt megadni lehetetlen.

2. A szótárállományra ugyanez vonatkozik: közvetlen elkészítése nagy nehézségekbe ütközik, és az eredmény nehezen módosítható a későbbiekben.
3. Az egyes képzők olyan számú képző+jel+rag kombinációt állítanak elő, és ennek megfelelő számú ragozási szabályt igényelnek, amelyet már nem lehet elfogadható méretű állományban tárolni.  
A magyar Ispell hash állomány jelentős részét (több, mint 60%), a jelenlegi 10000 ragozási szabályt tároló affixum tömörítést használó adatszerkezet (tehát nem a tényleges szótári hasítótábla) teszi ki, mintegy 5-7 MB-on. Ha az összes lehetséges képző kombinációt ragozási szabályokkal próbálnánk tárolni, nagy valószínűséggel nagyságrendi növekedésre lehetne számolni a fájl méretben és a helyesírás-ellenőrzés időigényében is.
4. Az Ispellben nincs lehetőség nem szótári tövek felvételére. A morfofonológiai alternációk egy része, mint a szó végi „a”, „e” változása (macska→macskát) jól kezelhető a ragozási szabályokkal, de a hangkivetős főnevek (eper→epret), a nyitótövek (kéz→kezek), a hangugratós igék (bérel→bérlem)<sup>1</sup> már nem, ráadásul ezek igen változatos formát mutatnak a magyar nyelvben.
5. Nemcsak a ragozási szabályok számát sem növelhetjük az évig, hanem a ragozási osztályok száma sem haladhatta meg pusztán gyakorlati megfontolásból a 26-ot (a legtöbb Linux terjesztés ugyanis pár éve ilyen korlátozásokkal fordított bináris Ispell tartalmazott).

Megoldás szerencsére mindegyik pontra akadt:

1. Az affixum állomány, és a szótárállomány is automatikusan van előállítva a Magyar Ispell keretrendszerben. Az affixum állományokban az m4 makrófeldolgozó segítségével gyakorlatilag többszörös toldalékolással vannak leírva a ragozási szabályok. A magyar.aff névre hallgató Ispell affixum állomány a végleges ragozási szabályokkal fordítási időben jön létre.
2. A szótárállomány forrása jelenleg az alapmodulból és külön kezelhető szaknyelvi modulokból áll. A szavak osztályozása egyszerű szólistákkal („egymező” adattáblákkal) történik, ami rendkívül leegyszerűsíti a szókincs bővítést a Magyar Ispellben. A szólistákból héj- és awk programok segítségével áll elő az Ispell buildhash programja számára szükséges magyar.dict szótárállomány.
3. A Magyar Ispell kombinált módszert használ a képzett szóalakok kezelésére: a ragozási szabályok mellett a főnevekhez képest viszonylag kis elemszámú szófajok – mint a melléknevek és az igék – esetében automatikusan állítja elő a képzett alakokat, és ezek külön kerülnek a szótárba más szófajú szavakként. Ezzel a módszerrel a szükséges ragozási szabályok számát jelentős mértékben csökkenteni lehetett, a szótárállomány növekedésének rovására. A szótárba külön felvett szóalakok számának növekedése mintegy kétszeres, de a hash állomány speciális felépítéséből fakadóan a végleges Ispell adatállomány mérete kevesebb, mint ötven százalékkal nő csak. (A ragozási szabályok számának nagyságrendi növelése ennél sokkal nagyobb problémát jelentene).

<sup>1</sup>Lásd pl. [5, 832 skk.].

4. A nem szótári tövek kezelésére kényszerűségből szokatlan megoldást sikerült találni: a nem szótári tőhöz kapcsolódó alakok levezetése nem valódi tőből, hanem az egyik ragozott alakból történik. A módszer rendkívül eredményesnek bizonyult: alkalmas volt a magánhangzó-harmónia szerint elkülönülő kisebb, zárt főnévcsoportok ragozásának kezelésére is, valamint ezt és a különböző hangrendű alakok kezelését is egy ragozási osztályon belül el lehetett intézni (mivel a tőnek választott többes számú alak kötőhangzójából egyértelműen meghatározható a szó hangrendje)!
5. A ragozási osztályok számának csökkentése érdekében több látszólag nem optimális, de mint később kiderült, a fejlesztés szempontjából szerencsés választásra került sor: a ragozási osztályok egy része össze lett vonva, pl. az ikes és iktelen igék ragozása, a főnevek és melléknevek ragozása részben ilyen volt. A hangrend esetében sem a ragozási szabályok számával takarékoskodó négy osztályos megoldás mellett döntöttem (mély hangrend, közös magas hangrend, csak ajakkerekítéses, csak ajakréses magas hangrend), hanem a szabályokat jelentős részben duplázó, de eggyel kevesebb ragozási osztályt kívánó megoldás mellett (mély hangrend, és a két magas hangrend).

Azóta szerencsére a 26 osztályos határt is sikerült átlépni, mindenféle probléma nélkül, mivel a Linux terjesztések is átálltak a több ragozási osztállyal fordított Ispell binárisokra.

### 3.3.3. Affixum keretrendszer

Az affixum állományok az áttekinthetőség miatt több részre vannak feldarabolva. A legfontosabb, amely a ragozási osztályok rövid összefoglalását is tartalmazza, az aff.fej névre hallgató állomány.

Az aff.alanyi és az aff.targyas állományok az alanyi illetve a tárgyas igeragozást tartalmazzák. Ezek az állományok hangrendi összevonást nem tartalmaznak, az m4 makrók, amelyek itt szerepelnek, csak a ragozási szabályokban szereplő mintákat általánosítják. Pl. az aff.alanyi állomány első (valóban is használt) makrójának definíciója a következő:

```
define(dupik,[BDGJKLMNRTZY] [LDGRZ] | K)
```

Egy példa a makró használatára, ugyanebből az állományból:

```
dupik > -IK,ASZ
```

Ez a sor az aff.alanyi állomány m4 makrófeldolgozón való átszűrésének eredményeképp a következőre alakul:

```
[BDGJKLMNRTZY] [LDGRZ] | K > -IK,ASZ
```

A makrók használata itt csökkenti a tévedéseket, egyszerűsíti a minták módosítását a későbbiekben.

Az aff.fonev állomány tartalmazza a főnevekre, és – az állomány megtévesztő neve ellenére – a melléknevekre vonatkozó ragozási szabályokat illetve ezek m4 forrását.

Az m4 makrózás használata ebben az állományban lényegesen megkönnyíti a számtalan rag előállítását. Az m4 makrók paraméteresek lehetnek, egymásba ágyazhatók, egyszerű feltételvizsgálatokat végezhetünk. Segítségével a mély és a két magas hangrend, valamint a tekintélyes számú képző+jel(ek)+rag kombináció rövid makrókkal legeneráltatható. A *pelda* állomány tartalma a bal oldalon, az m4-gyel feldolgozott állomány kimenete a jobb oldalon látható:

define(rag,	Futtatás:
\$1BA	\$ m4 pelda
\$1BAN	kocsiMBA
\$1RA)	kocsiMBAN
	kocsiMRA
define(jel,	kocsiDBA
rag(\$1M)	kocsiDBAN
rag(\$1D)	kocsiDRA
rag(\$1JÁ))	kocsiJÁBA
	kocsiJÁBAN
jel(kocsi)	kocsiJÁRA

A *define(makró\_név, makró\_törzs)* szintaxis az m4 központi utasítása, a paraméterek \$szám azonosítóval érhetőek el. Részletes leírása: info m4.

További affixum állományok: aff.fonev.morfo, amely a hangkivetős és a nyitótövek ragozását tartalmazza (a többes számú alakból levezetve) két ragozási osztályt meghatározva. Az aff.ige\_kiv a hangugratós igék ragozási osztályát definiálja.

A következő táblázat összefoglalja a Magyar Ispell jelenlegi ragozási osztályait. A táblázat oszlopai: a ragozási osztályok betűjelei (a kapcsolók); a szófaj, amit érint a ragozási osztály; a hangrend (o = mély, e = ajakrészes magas, ö = ajakkerekítéses magas); és végül a ragozási osztály leírása példákkal.

Jel	Szófaj	H.	Leírás
A	névszó	o	A magánhangzóra végződő névszók mély hangrendű birtokos személyjeles toldalékolása (ollónkról, macskáitokéinak), többes számú toldalékai (harcsákat), -ék képzős alakjai (kutyáékról), -cska képzős alakjai (birtokos személyjelek nélkül), és -nként, -stul ragok. A két utolsó ragot a mássalhangzóra végződő főnevek is felvehetik (nadrágostul). A tárgyragot kezelő ragozási szabályok nem itt szerepelnek, hanem a K (illetve a hangrendtől függően az L, M) osztályokban.

Jel	Szófaj	H.	Leírás
B	névszó	e	Mint az előző, csak a hangrend más.
C	névszó	ö	Mint az előző, csak a hangrend más.
D	melléknév	–	A leg-, legesleg- prefixum. Nem confixumként kezelve még (legkutya).
E	ige	e	ajakréses magas hangrendű <i>alanyi igeragozás</i> : ikes és iktelen jelen és múlt idő, feltételes és felszólító mód; határozói igenév (-ve/-vén), főnévi igenév (-ni) és személyragozása (-nem, -ned, -nie, ...).
F	melléknév	o	Fokjeles alakok (-bb) és ragozásuk, „a” kötőhangzós toldalékolás (-at, -an, -ak, és -ak ragozása).
G	melléknév	e	Mint az előző, csak a hangrend más. Itt van egy kis redundancia az L osztállyal is.
H	melléknév	ö	Gyakorlatilag megegyezik az előzővel, ezért sor kerül majd a két osztály összevonására.
I	névszó	oeö	Nem szótári tövel rendelkező, valamint a zárt -a kötőhangzót kapó névszavak többes számú alakjaiból levezetett toldalékolás. Nyelvtanilag több típus is tartozik ide, példákkal: bokor/bokrot, varjú/varjak, út/utak, fű/füvek, mű/művek, jó/javak, falu/falvak, valamint az -a kötőhangzós zárt osztály tagjai: vas/vasak. A nem szótári tövek a többes számú, a birtokos személyjeles, és tárgyragos toldalékokat kapják meg az I osztályban. Lásd fonev__morfo.2 és fonev__morfo2.2 állományokat.
J	névszó	oeö	Az előző osztályt kiegészítő -i melléknévképző. Ezt csak a mássalhangzó végű hangkivetős nem szótári tövek kapják meg mind a mai magyar helyesírás szerint (hatalmi, de úti, tűzi!). Ez különbözteti meg a morfo1 és morfo2 csoportokat (bár a kategorizáláson ez nem látszik, mivel az -a kötőhangzós alakoknak ez alapján a fonev__morfo.2 állományban is helyet kellene kapni). Valószínű, hogy szerencsésebb volna a -i képzős alakokat külön szóként felvenni a szótárba, mivel nem fedti le a teljes ragozási sort (paradigmát) ez az osztály (hiányoznak a birtokos személyjeles toldalékok, például hatalminkról). Az osztályba bekerült a -hat, -het végződésű igék E/3. múlt idejű alakját képező két szabály is. A későbbi fejlesztések miatt ez várhatóan kikerül innen.
K	névszó	o	Tárgyrag, és a mássalhangzó végű névszavak toldalékolása az A osztályhoz hasonlóan. Az A osztálytól való szétválasztására az I osztályba tartozó kivételek miatt volt szükség (nincs például szerelemünket).
L	névszó	e	Mint az előző, csak más hangrendű.
M	névszó	ö	Mint az előző, csak más hangrendű.
N	névszó	ö	-höz, -ön ragok. A névszók egy ajakkerekítéses magas hangrendű tőosztálya ajakréses hangrendű toldalékokat vonz (könyvet, zöldet), de a -höz, és az -ön esetében nem (könyvhöz, zöldhöz). A -hez, és -en toldalékú hibás alakok kivételként vannak letiltva a „w” kapcsolóval a főnevek esetében (könyvhez), míg mellékneveknél csak a -hez (zöldhez, bölcshez), mivel itt az -en határozói rag miatt helyesek a zölden, bölcsen. stb. alakok.



Jel	Szófaj	H.	Leírás
O	ige	o	Mint az E kapcsoló, csak ez a mély hangrendű <i>alanyi igeragozás</i> : A -k végű nem ikes igék ragozása külön van lekezelve (lásd kivételek/igekotos/rak).
P	ige	ö	Mint az előző, csak ajakkerekítéses magas hangrendű.
Q	névszó	o	-ja (mély hangrendű I/3 személyű birtokos személyjel és ragozott változatai).
R	névszó	eö	-je (magas hangrendű I/3 személyű birtokos személyjel és ragozott változatai.)
S	névszó	o	-a (mély hangrendű I/3 személyű birtokos személyjel és ragozott változatai).
T	névszó	eö	-e (magas hangrendű I/3 személyű birtokos személyjel és ragozott változatai).
U	névszó	o	Esetragok birtokos személyjel nélkül. A birtokjeles (-é), és birtoktöbbsítő jeles (-i) alakok is megtalálhatók itt (macskáéit), de önmagában a tárgyrag nem.
V	névszó	e	Mint az előző, csak más hangrendű.
W	névszó	ö	Mint az előző, csak más hangrendű.
X	ige	-	Igekötők: abba-, agyon-, alá-, ..., végbe-, végig-, vissza-. Egyszerűsítés, hogy a legtöbb ige minden igekötőt felvehet. Ha egy igénél nem szeretnénk igekötős alakokat, akkor helyezzük el az igét az ige_koto.1 állományban is. Általában két esetben van erre szükség: ha az ige már tartalmaz valamilyen nem tipikus igekötőt ( <i>kétségbeesik</i> ); vagy igenévszóról van szó, és nem akarjuk, hogy a névszóragok megjelenjenek az igekötővel együtt ( <i>megingeiket</i> ). Ez utóbbi esetben külön vannak felvéve az igekötős alakok. Lásd ige_koto.1.
Y	névszó	-	Szóösszetételben való szereplést engedélyező kapcsoló (COMPOUNDFLAG). Lásd a Hunspell(4) kézikönyv oldalt! A legtöbb köznévfő megkapja. Kivételek a fonev_kulon.1 állományban.
Z	névszó	-	Idegen kiejtésű, vagy mozaikszavak toldalékai esetében a kötőhangzó hiányát jelzi. A kotojeles_hangzo.1 állományban vannak megadva azok a szavak, amelyek felveszik a Z kapcsolót.
a	számnév	-	A tízesek prefixumként: tizen-, huszon-, ... kilencven-.
b	számnév	-	A százasként prefixumként: száz-, kétszáz-, ... kilencszáz-.
c	névszó	-	-szerű, -féle, -fajta, -beli, -nemű képzőszerű utótagok. A 6-3-as szabály helyes kezeléséhez szükséges osztály. (Ezek az utótagok nem számítanak bele a szótagszámba, hasonlóan az igazi toldalékokhoz.)
d	ige	-	morf. alt. ragozás
e	ige	e	Ajakkerekítéses magas hangrendű <i>tárgyas igeragozás</i> : ikes és iktelen jelen és múlt idő, feltételes és felszólító mód.
f	névszó	-	Idegen toldalékolású, y-ra végződő főnevek jele.

Jel	Szófaj	H.	Leírás
g	–	–	Toldalékolt alakok jele. A Hunstem tövező számára szükséges információt adja meg. A g kapcsolót közvetlenül követő 1 és 9 közé eső szám jelentése, hogy a tő a számnak megfelelő számú szóvégi karakter levágásával állítható elő. Ha ilyen szám nincs, akkor a háttérben a különleges __ szó szótári bejegyzéssel (annak kapcsolóiként) van tárolva a szó töve. L. ragozatlan.2 állomány leírását.
h	számnév	–	Ezeregyszáz-, ezerkétszáz-, ... ezerkilencszáz- prefixumok a kétezer alatti számok egybeírásához.
i	névszó	o	-nyi. A morfofonetikus csoport miatt van szükség a szétválasztásra, mivel a többi melléknévképzővel ellentétben ez nem járul a nem szótári tövekhez (bokryni, füvnyi).
j	névszó	eö	Mint az előző, csak magas hangrendű.
k	főnév	o	Melléknévképzők: -(o)s, -(j)ú, -i, -tlan/-talan, nem teljes ragozási sorral.
l	főnév	e	Mint az előző, csak más hangrenddel.
m	főnév	ö	Mint az előző, csak más hangrenddel.
n	főnév	oeö	Mint az előző, csak a morfo csoportra vonatkozóan.
o	ige	o	Mint az „e” osztály, csak ajakréses magas hangrendű <i>tárgyas igeragozás</i> .
p	ige	ö	Mint az előző osztály, csak más hangrendű.
q	–	–	-féle toldalék a tulajdonnevekhez.
r	–	–	
s	főnév	o	-ai (I/3 birtokos személyrag + birtoktöbbsítő jel). L. S kapcsoló
t	főnév	eö	-ei (I/3 birtokos személyrag + birtoktöbbsítő jel). L. T kapcsoló
u	–	–	A „w” kapcsoló hatását módosító kapcsoló: ha meg van adva a „w” mellett, akkor a tiltás csak a szótőre vonatkozik, a tiltott szótő toldalékolt változataira nem. A mangrove szót így lehetett egyszerűen és tártakárékosan lekezelni, mivel a felvitt szabályok nem kezelik a -e végződésű főneveket mély hangrendű toldalékokkal. A „mangrove” szó mellett szerepel a „mangrové” is a szótárban „w”, „u” (és egyéb) kapcsolókkal, így a „mangrové” alak helytelen, de a „mangrovét”, „mangrovéinak” stb. alakok helyesek. A „mangrove” említett kapcsolóit a szotar/kivetelek/ragozatlan/mangrove állományban találjuk meg.
v	–	–	szerepelhet összetételekben első szóként (COMPOUNDFIRST). Jelenleg ezzel a kapcsolóval van megoldva az egyszerű tőszámnevek, és a melléknévképzős főnevek egybeírásának engedélyezése (illetve a Magyar Myspell gondoskodik arról, hogy tényleg csak a melléknévképzős alakokat lehessen egybeírni a tőszámnevekkel), pl. harmincnapi, de! két hónapos. Lásd A magyar helyesírás szabályai 119. pontját, és a Hunspell(4) kézikönyv oldalt!

Jel	Szófaj	H.	Leírás
w	–	–	Tiltott szóalakok kapcsolója (FORBIDDENWORD). A Magyar Ispell/Myspell egyszerűsítő felépítésének számos hibája küszöbölhető ki a hibásan képzett szavak tiltásával. A szótárbővítő az idegen kiejtésű szótagra végződik, de a tövel egybeírt toldalékú szavak felvitelénél találkozik ezzel (pl. Madáchhal/w), valamint a rossz szóösszetételek tiltásánál (ifjúpár/w). Adjuk meg a #MYSPELL TILTOTT megjegyzést a szótárállományban, ha ragozott szavak tiltását szeretnénk. (Megtehetnénk ezt úgy is, hogy ha a szót ismételt felvesszük a ragozatlan.2 állományba, a „w” kapcsolóval. A speciális megjegyzés használatának két előnye van: Először is, a tiltott szó nem kerül bele az Ispell szótárba, másodszer az alapszó, és a tiltás egy helyen található, és így könnyebben nyomon követhető, és szerkeszthető.) Lásd a Hunspell(4) kézikönyv oldalt!
x	–	–	szerepelhet összetételekben utolsó szóként (COMPOUNDLAST). Ezt a külön felvett (I osztály által kezelt) többes számú alakok kapják meg.
y	névszó	–	összetett szó (COMPOUNDWORD). A 6-3 szabály helyes alkalmazásához szükség van arra az információra, hogy a tőként felvett szó vajon összetett-e, vagy sem. Lásd a Hunspell(4) kézikönyv oldalt!
z	számnév	–	-magammal, -magaddal, -magával, -magunkkal, -magatokkal, -magukkal toldalékként, mint a sorszámnevek visszaható névmás „ragja”. Lásd szotar/kivetelek/ragozatlan/szamnev2.
0	–	–	A tövezésnél nem megjelenítendő alak jele. A Magyar Ispell számos olyan „tövet” tartalmaz, amelyek pusztán gyakorlati szempontból léteznek. Ilyenek az improduktív névszóragozásnál a többes számú alakok (madarak, szerelmek), illetve a kötőjellel toldalékolt szavaknál a szó és kötőjel alakok (tv-). Ezeket az alakokat a 0 kapcsolóval ellátva a tövezésnél nem jelennek meg.

### 3.3.4. Szótári keretrendszer

A szófajilag elkülönített szóállományok lehetővé teszik a szófajra jellemző ragozási osztályok kapcsolóinak és az alapszavaknak a viszonylag egyszerű párosítását. Héj- és awk program gondoskodnak erről bizonyos kivétellisták figyelembevételével.

Például egy szófajon belül három fő ragozási mintacsoport különíthető el a hangrend alapján. Az alapszavak hangrendje kibővített reguláris kifejezésekkel (awk programokkal) fordítási időben kerül felismerésre, pl. egy erre vonatkozó programrészlet:

```
# mély hangrendű igék
```

```
/[úúóóáá][bcd fghjklmnp rstvxyz]*ik$/ { print $0 "/X/A" }
```

```
/[úúóóáá][bcd fghjklmnp rstvxyz]*$/ && ! /ik$/ { print $0 "/X/A" }
```

A példa csak az utolsó szótag hangrendjének vizsgálatát végzi el, illetve ikes igék esetén az

-ik előtti szótagét, ez sok esetben nem ad jó megoldást. A Magyar Ispell keretrendszerben kivétellisták segítségével tehető pontossá az automatikus hangrend-besorolás. Ilyen kivétellisták az `_alap` könyvtárban található `fonev_mely.1`, `ige_mely.1`, és `melleknev_mely.1` állományok, amelyeket a feldolgozó `awk` programok egy-egy asszociatív tárhoz töltenek be (más kivétellistákhoz hasonlóan), ami a feldolgozás sebességét jelentősen gyorsítja.

### 3.4. Új szótármodulok létrehozása

A Magyar Ispell forrásának telepítése után (ezt a következő fejezet ismerteti) lépünk be a `magyarispell-0.86` könyvtárba.

A szótármodulok külön könyvtárakban helyezkednek el. Az alapkönyvtárak neve aláhúzás-jellel (`_`) kezdődik. A legfontosabb szótármodul az `_alap`.

Egy új szótármodul létrehozása nagyon egyszerű: hozzunk létre egy új alkönyvtárat a `magyarispell-0.86` könyvtárban belül. A könyvtár neve aláhúzásjellel kezdődjék.

Ha szeretnénk, hogy a szótár fordítása során az új szótári modul is bekerüljön a végeredménybe, hozzunk létre egy szimbolikus kötést (`symlink`) a könyvtárra:

```
ln -s _pelda Példa
```

Látható, hogy a többi alkönyvtárhoz hasonlóan a szimbolikus kötés neve aláhúzás nélküli, és nagybetűvel kezdődik. Ezek a tulajdonságok határozzák meg, hogy mely könyvtárak tartalma számít fordításra kijelölt modulnak.

Egy modult egyszerűen a szimbolikus kötés törlésével tudunk kivenni ebből a csoportból:

```
rm Példa
```

#### 3.4.1. Szófaji bontás

A szótármodul szavakat fog tartalmazni, amelyet szófaji alapon el kell tudnunk különíteni. A köznevek a `fonev.1` állományban kerülnek felsorolásra (egy szó egy sorba kerül). A melléknevek a `melleknev.1` állományban, illetve az alanyi és tárgyias igék az `ige_alanyi.1` és az `ige_targy.1` állományokban.

A tulajdonnev.2 állomány két mezőben tárolja a tulajdonnevekre vonatkozó információt, de ha az első mező üres, a második speciálisan (ragozatlan szóként) kerül értelmezésre.

Ezért ha egy tulajdonnév nem kap toldalékot a magyar nyelvben (rendszerint többemű tulajdonnevek nem utolsó tagja, vagy kapcsolóval ellátott, illetve letiltott toldalékolt alakok), akkor a sorban egy tabulátor után (tehát a második mezőbe) írjuk a szavakat, például:

```
Santa  
Vörösmarty/f
```

Gauss| #MYSPELL TILTOTT

Gaussá #MYSPELL TILTOTT

Gauss-szal/g5

Gauss-szá/g4

Ha ragozzuk, akkor rögtön sor elejére.

Einstein

Ha a ragozott alakok között melléknévképzős (kisbetűs) is akad, akkor az ugyanebbe a sorba, egy tabulátorral elválasztva írjuk:

Einstein einsteini

Ha sehová nem illik a szó, felvehetjük a ragozatlan.2 állományba. Ha azért valamilyen mértékben ragozott, akkor puskázzunk. A legenerált magyar.dict állományt szűrjük meg arra vonatkozóan, hogy egy hasonló ragozású, hangrendű, stb. szó milyen ragozási csoportjelzőkkel került a szótárba:

```
grep ^milyen/ magyar.dict  
milyen/B/V/L/R
```

A szavunkhoz ugyanezt rendelhetjük hozzá a ragozatlan.2 állományban:

```
ilyen/B/V/L/R
```

### 3.4.2. Kivételek a szófaji kategóriákon belül

A legtipikusabb kivételek a hangrendi kivételek. Ha egy szó utolsó szótagja magas hangrendű, de a szó mégis mély hangrendű, akkor a szót vegyük fel a szófajhoz tartozó \_mely.1 végződésű névvel rendelkező állományba. Az ilyen szavak utolsó szótagjában többnyire „é”, vagy „i” magánhangzó szerepel, és gyakran vegyes hangrendűek.

Konvenció, hogy tulajdonnev\_mely.1 állomány nem létezik, az ilyen tulajdonneveket a fonev\_mely.1 állományba helyezzük el.

A főnevek esetében megtaláljuk még a fonev\_magas.1 állományt is, ami elsősorban idegen tulajdonneveket, illetve egy-két át nem írt idegen közsót tartalmaz (pl. gentleman, Cadillac), amelyek utolsó szótagja mély hangrendű, mégis a kiejtés miatt magas hangrendű toldalékokat kapnak (gentlemannel, Cadillackel).

A névszók (elsősorban a főnevek) esetében a hangrendi kivételek mellett a I./1-es birtokos személyjeleket (-ja/-je/-a/-e) érintő kivételek jelentősek. Három állománnyal kell megismernünk velük kapcsolatban: a fonev\_ae.1 azokat a főneveket sorolja fel, amelyek *b, d, f, k, l, n, p, r* és *t* betűre való végződés ellenére -a vagy -e birtokos személyjelet kapnak (például váll→válla).

A `fonev_jaje.1` azokat a főneveket sorolja fel, amelyek *c, g, h, j, m, s, x, y, v* és *z* betűre való végződés ellenére *-ja* vagy *-je* birtokos személyjelet kapnak (például *hang*→*hangja*). Amennyiben mind a *j-s*, mind a *j* nélküli változatát megkapja egy szó, akkor a `fonev_jajae.1` állományban kell feltüntetni (például *beszéd*→*beszéde*, *beszédje*).

### 3.4.3. Állományok táblázata

A következő táblázat összefoglalja a jelenlegi állományokat:

Állomány	Leírás
<code>fonev.1</code>	köznevek
<code>fonev_ae.1</code>	<i>b, d, f, k, l, n, p, r</i> és <i>t</i> végű; <i>-a, -e</i> birtokos személyragot kapó főnevek ( <i>váll</i> → <i>váll</i> )
<code>fonev_ing.1</code>	mély és magas hangrendű ragot kapó főnevek ( <i>fotel</i> → <i>fotelban</i> , <i>fotelben</i> )
<code>fonev_jaje.1</code>	<i>c, g, h, j, m, s, x, y, v,</i> és <i>z</i> végű, <i>-ja, -je</i> birtokos személyragot kapó főnevek ( <i>hang</i> → <i>hangja</i> )
<code>fonev_jajae.1</code>	<i>-a/-e,</i> és <i>-ja/-je</i> birtokos személyraggal is előforduló főnevek ( <i>beszéd</i> → <i>beszéde</i> , <i>beszédje</i> )
<code>fonev_kulon.1</code>	főnevek, amelyek összetett szóban nem szerepelhetnek ( <i>január</i> )
<code>fonev_magas.1</code>	főnevek, amelyek utolsó szótagja mély hangrendű, de magas hangrendű ragot kapnak ( <i>gentleman</i> → <i>gentlemannel</i> )
<code>fonev_mely.1</code>	főnevek, amelyek utolsó szótagja magas hangrendű, de mély hangrendű ragot kapnak ( <i>hég</i> → <i>hégj</i> )
<code>fonev_morfo1.2</code>	hangkivető főnevek, többesszámú alakjaikkal ( <i>szerlem</i> → <i>szerlemen</i> , <i>kehely</i> → <i>kelyhen</i> )
<code>fonev_morfo2.2</code>	nyitótüvek, és egyéb kivételek, többesszámú alakjaikkal
<code>fonev_oe.1</code>	<i>-et, -ek, -ön, -höz</i> ragot kapó, ajakkerekítéses magas hangrendű főnevek ( <i>könyvet</i> , <i>könyvön</i> )
<code>fonev_osszetett.1</code>	Összetett szavak (kerékpár).
<code>ige_alanyi.1</code>	nem tárgyias igék (történik, nincs történi, történed, mint a Helyes-e?-ben)
<code>ige_ikes_kiv.1</code>	dupla mássalhangzó+ikes igék, <i>-ó, -ás</i> képzőkkel, l. <code>iges.awk</code> , <code>igesgen</code>
<code>ige_koto.1</code>	igekötőt nem kapó igék (vagy külön kezelve az igekötős alakok)
<code>ige_mely.1</code>	igék, amelyek utolsó szótagja magas hangrendű, de mély hangrendű ragot kapnak ( <i>sír</i> → <i>sírok</i> )
<code>ige_morfo.1</code>	hangkivetős igék ( <i>bicsakol</i> → <i>bicsakló</i> )
<code>ige_targy.1</code>	tárgyas igék ( <i>üt</i> → <i>üti, ütöm</i> )
<code>ige_tat_kiv.1</code>	<i>-tat, -tet</i> gyakorítóképzős alakok nincsenek (pl. <i>mond</i> , <i>mond</i> <i>at</i> , de nincs <i>mond</i> <i>at</i> , <i>mond</i> <i>at</i> <i>at</i> )

Állomány	Leírás
melleknev.1	melléknevek
melleknev_a.1	-o kötőhangzót (főnévi ragozást) nem kap, (nincs állottot)
melleknev_ae.1	b, d, f, k, l, n, p, r és t végű; -a, -e birtokos személyragot kapó melléknevek
melleknev_ing.1	mély és magas hangrendű ragot kapó melléknevek
melleknev_jaje.1	c, g, h, j, m, s, x, y, v, és z végű, -ja, -je birtokos személyragot kapó melléknevek
melleknev_jajae.1	-a/-e, és -ja/-je birtokos személyraggal is előforduló melléknevek
melleknev_mely.1	melléknevek, amelyek utolsó szótagja magas hangrendű, de mély hangrendű ragot kapnak
melleknev_oe.1	-et, -ek, -ön, -höz ragot kapó, ajakkerekítéses magas hangrendű melléknevek (zöldet, zöldön)
ragozatlan.2	Nem ragozott, vagy kivételesen ragozott szavak megadására szolgáló állomány. A kivételes ragozási adatokat kapcsolókkal adhatjuk meg a szó után közvetlenül, egy perjellel elválasztva a szótól. A szavakat külön sorba kell megadni. Egy sorban a szó töve is megadható a tövezőprogram számára, tabulátorjellel elválasztva a szótól.
tulajdonnev.2	Tulajdonnevek: ha az első mező üres, akkor a szó ragozatlan, ha nem, akkor a második mező melléknév.
tulajdonnev-geo.2	földrajzi nevek
tulajdonnev-keresztnev.2	keresztnevek
tulajdonnev-szemelynev.2	személynevek
kotojeles_mely.7	Mély ragozású, idegen ejtésű, vagy mozaikszavak kötőjellel elválasztott toldalékkal (felépítést l. az állományban).
kotojeles_magas.7	Magas ajakréses ragozású, idegen ejtésű, vagy mozaikszavak kötőjellel elválasztott toldalékkal
kotojeles_magas2.7	Magas ajakkerekítéses ragozású, idegen ejtésű, vagy mozaikszavak kötőjellel elválasztott toldalékkal
kotojeles_hangzo.1	Az itt felsorolt szavak kapják meg a Z kapcsolót, ami azt jelzi, hogy az itt megadott szavak esetében nincs kötőhangzó a toldalék és a szó között (például URH-nk, és nem URH-unk, mivel az URH kiejtve „uerhá”).

A tulajdonnevek hangrendjéről: a -i végű melléknévképzős alakoknál a hangrend automatikusan mély, ha a tőszó utolsó magánhangzója mély. Ha ingadozó, (i, í, mint Párizs), akkor külön felveendő a melleknev\_mely.1 állományba a képzett alak (párizsi).

### 3.4.4. Y-ra végződő szavak

Az y végű szavak (pl. Sydneyvel, Uruguayjal, Kölcseyn, sprayen) ragozása vissza lett vezetve az i, és j végű szavakéra.

Az „f” kapcsoló segítségével a következőképp kódolhatóak az y végű szavak:

A tövet nem ragozzuk, csak „f” kapcsolót kap (közneveknél még Y-t, hogy összetett szavakban szerepelhessen).

Felvesszünk egy olyan tövet, amiben kiejtés szerint az y vagy i-re, vagy j-re van átírva (Sydnei, Uruguaj, Kölcsei, spraj), és ezt szerepeltetjük a megfelelő állományokban, de tiltva!

Köznevek (angol eredetű idegen szavak) esetében:

--- részlet az \_alap/fonev állományból ---  
boj #MYSPELL TILTOTT (boy)  
brandi #MYSPELL TILTOTT (brandy)  
cowboj #MYSPELL TILTOTT (cowboy)  
displaj #MYSPELL TILTOTT (display)  
ladi #MYSPELL TILTOTT (lady)  
gázspraj #MYSPELL TILTOTT (gázspray)  
orrspraj #MYSPELL TILTOTT (orrspray)  
spraj #MYSPELL TILTOTT (spray)  
penni #MYSPELL TILTOTT (penny)  
whiski #MYSPELL TILTOTT (whisky)  
złoti #MYSPELL TILTOTT (złoty)

-----  
--- részlet a kivetelek/ragozatlan/y állományból ---  
boy/fY  
brandy/fY  
cowboy/fY  
display/fY  
lady/f  
gázspray/fY  
orrspray/fY  
spray/fY  
penny/fY  
whisky/fY  
złoty/fY

-----  
A köznevek esetében a kivetelek/ragozatlan/y állomány szolgál az f(és az Y) kapcsolós alakok megadására. A tulajdonnevek (régies írásmódú keresztnemek, illetve földrajzi nevek) esetében a tulajdonnev\* állományok szolgálnak erre a célra. Ezekben van a szótó nem ragozva, „f” kapcsolóval felvéve. Felvesszük az átírt, nem valódi tövet, megadjuk a melléknévképzős alakot is, és külön sorban tiltjuk le a nem valódi tövet.

--- részlet a tulajdonnév állományból ---  
Vörösmarty/f  
Vörösmarti #MYSPELL TILTOTT



Vörösmarti vörösmartys  
Vörösmarti vörösmartyas  
-----

Itt gyakran – ebben a példában is – a szavak hangrendje is más. Ezért a *Vörösmarti* szót felvesszük a `fonev_mely.1` állományba is, a melléknévképzős alakokat pedig a `melleknev_mely.1` állományba.

hasonló problémát okoz az idegen szavak esetében a mély hangrendűnek látszó, de magas hangrendű alak: pl. `spray`, `display`. Ezen a nem valódi tövek `fonev_magas.1` állományban való elhelyezése segít.

### 3.4.5. Kivételek könyvtár

A szotar/kivetelek könyvtár tartalmazza azokat a kivételeket, amelyek rendszerint valamilyen zárt toldalékolást mutatnak, de jóval szűkebb csoportot alkotnak a Magyar Ispellben `morfo1/morfo2` néven jelölt névszóosztályoknál. Két alkönyvtárat tartalmaz a könyvtár, az igeekotes és a ragozatlan könyvtárat. A két könyvtár sok állományában tövezési információk is meg vannak adva második mezőként, hasonlóan a `ragozatlan.2` állományhoz. Ezeket az állományokat a `.2` kiterjesztés jelöli.

A kivetelek könyvtár gyökerében található `ige_kepzos.11` állomány az igei kivételek képzős alakjait adja meg (mivel a két alkönyvtárban lévő ige kimaradnak a képzős alakok automatikus előállításából).

Az ige állományban felsorolt képzős alakok útmutatást adnak arra vonatkozóan is, hogy milyen képzős alakok kerülnek külön tőként felvételre a szótárba.

### 3.4.6. További információk

További információkat a Gyakran ismételt kérdések között találunk (GYIK állomány a Magyar Ispell forrásában), valamint a Magyar Ispell levelezőlistán (<http://www.yahogroups.com/group/magyarispell>).

A levelezőlistára a `magyarispell-subscribe@yahogroups.com` e-mail címre írt levéllel jelentkezhetünk. A listára szánt leveleinket a `magyarispell@yahogroups.com` címre címezzük.

### 3.4.7. Segédprogramok

A `bin/` alkönyvtárban található `break` program segít szavakat kigyűjteni szöveges, vagy HTML állományokból.

### 3.5. Magyar Myspell

A Magyar Myspell legfrissebb változata jelenleg a <http://magyarispell.sf.net> oldalról tölthető le. Korábbi változatokat a <http://www.szofi.hu/gnu/magyarispell> oldalon találunk.

A forrásból a `make` paranccsal fordíthatjuk le az `example` nevű példaprogramot, és a leendő – Ispell felváltani képes – programot, a `Hunspell-t`.

### 3.6. Javítási cseretáblázat

A következő táblázat összefoglalja a Magyar Ispell 0.99.3-as változatában megtalálható javítási cseretáblázatot.

A táblázat harmadik oszlopa részben példákat mutat a cseretáblázat segítségével tett javaslatokra, valamint a cseretáblázat segítségével beazonosított hibás szóösszetételekre. Ez utóbbi tulajdonság a Magyar Myspell 0,4-es változatában jelent meg.

A legtipikusabb hibák (*i/í, ü/ű* stb.) elől találhatók, hogy a javaslatoknál is ezek kerüljenek legelőre.

Mit	Mire	Példák cserére, illetve javított hibákra
í	i	szer+víz→szerviz, sí+ma→sima
i	í	kiváncsi→kíváncsi, elit+élt→elítélt
ó	o	szimbólikus→szimbolikus, mikró+számítógép→mikroszámítógép
o	ó	mikro→mikró
o	ő	cselekvo→cselekvő (lapolvasási, kódolási hibák javításához)
ú	u	kultúrált→kulturált, vas+árú→vasáru
u	ú	kultura→kultúra
u	ű	keseru→keserű (lapolvasási, kódolási hibák javításához)
ű	ü	menű→menü
ü	ű	betü→betű, keserü→keserű
j	ly	juk→lyuk, est+éj, kar+vaj, kord+éj, rost+éj, szeg+éj, szem+éj, szem+éjes, szem+éji, szent+éj, szesz+éj, szesz+éjes, tar+táj, tök+éj, ünnep+éj, ünnep+éjes,
ly	j	csevely→csevej, muszály→muszáj, boly+torján, kard+bolyt, tojássár+gálya, váll+bolyt, boly+ár, boly+tár, súly+tó
a	ä	Handel→Händel
S	Š	Skoda→Škoda
s	š	skodás→škodás
r	ř	Dvorák→Dvořák
ra	řa	Dvorak→Dvořák

Mit	Mire	Példák cserére, illetve javított hibákra
l	ł	zloty→złoty
L	Ł	Lódz →Łódz
Lo	Łó	Lodz →Łódz
oliere	oli&grave;re	Molier&grave;→Moli&grave;re
Angström	&Aring;ngström	Angström→Ångström
angström	&aring;ngström	angström→ångström
&Otilde;	Ō	Ō→Ő
&otilde;	ő	ő→Ő
&Ucirc;	Ū	Ū→Ű
&ucirc;	ű	ű→Ű
-	&ndash;	Duna-Tisza köze→Duna–Tisza köze
jj	lly	gajj→gally
jj	llj	ájj→állj
lly	jj	csevellyel→csevejjel
ggy	gyj	haggyon→hagyjon, naggya→nagyja
gyj	ggy	higyjen→higgyen
ggy	dj	maraggy→maradj
gy	dj	horgya→hordja, kargyuk→kardjuk
nny	nyj	annya→anyja
nny	nj	fonnyátok→fonjátok
tty	tyj	attya→atyja
tty	tj	láttyák→látják, bottya→botja
cc	tsz	szerecc→szeretsz
cc	dsz	maracc→maradsz
cc	gysz	eccer→egyszer
cs	ts	kölcség→költség
cs	ds	boloncság→bolondság
ccs	ts	baráccság→barátság
ccs	ds	vaccság→vadság
ccs	gys	naccság→nagyság, őnaccsága→őnagysága
g	kd	csug→csukd, lög→lökd
öss	ős	elősször→először, erőssen→erősen
öll	ől	szöllő→szőlő
ütt	űt	hüttő→hűtő, füttő→fűtő
íjj	íj	dijja→díja, ijja→íja
x	ksz	boxer→bokszer, vax→vaksz
ksz	x	fakszol→faxol, Beatriksz→Beatrix
xx	xsz	kódexxel→kódexszel
xx	kssz	vaxxal→vaksszal
xsz	kssz	vaxszal→vaksszal
x	xsz	kódexel→kódexszel
x	kssz	vaxal→vaksszal
sz	c	licensz→licenc

Mit	Mire	Példák cserére, illetve javított hibákra
sz	ce	licenszt→licencet
ssz	cc	licensszel→licenccel
jl	lj	tejles→teljes, éjlen→éljen
gg	g	heggeszt→hegeszt
file	fájl	file→fájl
file-	fájl	file-lal→fájllal
floppy	flopi	floppy→flopi
bájt	byte	bájt→byte
bájt	byte-	bájtos→byte-os
chh	chcs	Madáchhal→Madáchcsal
ll	l	vállfaj→váfaj
ny	nny	menyország→mennyország
nny	ny	mennyhal→menyhal
o	a	házok→házak
ivess	íves	szivessen→szívesen
-cel	-szel	AIDS→AIDS-szel
-cé	-szé	AIDS-cé→AIDS-szé
őrö	ölrő	előről→előlről
őr	ölr	előről→előlről
ölr	őr	előre→előre
egyenlőre	egyelőre	egyenlőre→egyelőre <sup>2</sup>
kábé	kb.	kábé→kb.
ss	ss-s	Barossal→Baross-sal
ss	ss-sz	Gaussal→Gauss-szal
ssz	ss-sz	Gausszal→Gauss-szal
ll	ll-l	Böllel→Böll-lel
nn	nn-n	Mannal→Mann-nal
mm	mm-m	Grimmel→Grimm-mel
tt	tt-t	Beckettel→Beckett-tel
Jellasics	Jelačić	Jellasics→Jelačić

<sup>2</sup>Bár helyes az *egyenlőre* is, nagyságrendileg ritkább, mint az *egyelőre*.

## 4. fejezet

# Programozói dokumentáció

A Hunspell program, illetve a Magyar Myspell függvénykönyvtár képességeit saját programjainkhoz is hozzáadhatjuk.

### 4.1. Hunspell csőfelület

A -a kapcsoló megadásával a Hunspell program az Ispell -a kapcsolójának hatását utánozza, egy kevés kivételtől eltekintve. Ez az üzemmód ad lehetőséget arra, hogy alkalmazásaink a Hunspell-t használják a helyesírás-ellenőrzésre. Ilyenkor a Hunspell-t háttérben elindítják az alkalmazások, annak ellenőrizendő szöveget adnak át, majd az ellenőrzés végeredményét visszakapva azt feldolgozzák. Számos program él ezzel a lehetőséggel: például a KWord, LyX, Emacs (Windows alatt is).

A Hunspell indulásnál kiír a standard kimenetére egy bemutatkozó sort, majd a standard bemenetet soronként dolgozza fel. Minden egyes sor után a kimeneten a kézikönyvoldalon ismertett információkat jeleníti meg, amit a saját programunk (amennyiben gondoskodtunk a kimenet átirányításáról) feldolgozhat.

A Hunspell a bemenetet szövegsorként értelmezi, hacsak a bejövő sor első karaktere nem egy parancskarakter.

A Hunspell által kezelt parancskaraktereket, és leírásukat a következő táblázat ismerteti:

- ! Nem értelmezi a sort.
- +  $\TeX$ üzemmód bekapcsolása.
- Visszalépés az előző üzemmódba.
- @ A jelet követő szó szótárba kerül a futás idejére.
- \* A jelet követő szó szótárba kerül.
- # A szótár elmentésre kerül a saját szótárállományba.
- ^ A jelet követő sor feldolgozása.

Az & jel (kisbetűsként való felvétel) még nincs a Hunspellbe építve. Egyéb kezdőkértékek esetén a sorok szövegsorként kerülnek feldolgozásra.

## 4.2. Magyar Myspell felület

A Magyar Myspell felület (API) lehetőséget nyújt a függvénykönyvtár közvetlen használata, bár pár dolgot, mint például a saját szótár elmentését és betöltését, magunknak kell megoldani.

### 4.2.1. Konstruktor

A felhasználói programban a MySpell osztály metódusaiként férhetünk hozzá a helyesírás-ellenőrzőhöz.

```
MySpell(const char * affixum, const char * szotar);
```

A MySpell osztály konstruktora két karakterláncot vár paraméterként: az affixumállomány, és a szótárállomány útvonalát.

### 4.2.2. Metódusok

```
int spell(const char * alak);
```

A megadott szó ellenőrzése. Ha a visszatérési érték 0, a szó hibás, egyébként helyes.

```
int suggest(char*** javaslatok, const char * alak);
```

Javítási javaslatok kérése a megadott szóra. A függvény visszatérési értéke a javaslatok száma, illetve a javaslatokat tároló karakterlánc tömb címe. Ez utóbbi tömb a függvényhívás alatt kerül lefoglalásra, címe pedig a javaslatok változóban kerül elhelyezésre. A tömb felszabadításáról nekünk kell gondoskodni a későbbiekben, ügyelve arra, hogy a tömb címe NULL is lehet, ha a helyfoglalás memóriaproblémák miatt sikertelen volt. (Több szálát kezelő programban nem használható.)

```
int suggest_stems(char*** javaslatok, const char * alak);
```

Helyes tövek előállítása az előző függvényhez hasonló szintaxissal. Ha nincs tő, akkor a szót nem ismeri az ellenőrző.

```
int suggest_pos_stems(char*** javaslatok, const char * alak);
```

Valószínű tövek előállítása az előző függvényhez hasonló szintaxissal. A Magyar Ispell szótárban szereplő névszóragokat próbálja leválasztani a megadott ismeretlen szóról a program.

```
char * get_dic_encoding();
```

Visszaadja a betöltött affixumállomány SET paraméterét, vagyis a beállított karakterkódolást. (L. man 4 hunspell)

```
char * get_wordchars();
```

Visszaadja a beállított karakterkészlet betűkarakterének karakterláncát, kiegészítve az affixumállomány WORDCHARS paramétereként megadott karaktereivel. Ez az információ a szavakra bontáshoz hasznos.

```
struct cs_info * get_csconv();
```

Visszaadja a betöltött karaktertáblát, elsősorban a későbbi betűméret-átalakításokhoz. (L. csutil.cxx, csutil.hxx a forrásban.)

```
int put_word(const char * alak);
```

A megadott szó futásidejű felvételére kerül sor. Ha az erre biztosított hely elfogy, a visszatérési érték 0, egyébként 1. (Az így maximálisan felvehető szavak számát az atypes.hxx állomány MAXUSERWORD makrója adja meg.)

```
int put_word_suffix(const char * alak, const char * kapcsolok);
```

A megadott szó futásidejű felvételére kerül sor, a második paraméterként megadott kapcsolókkal (ezek lehetnek prefixumok is, a megtévesztő név ellenére).

```
Például put_word_suffix("leszörpözik", "Pp");
```

```
int put_word_pattern(const char * alak, const char * alak2);
```

A megadott szó futásidejű felvételére kerül sor a második paraméterként megadott szó kapcsolóival.

```
Például put_word_pattern("alfönök", "fönök");
```

```
char * get_prevroot();
```

Visszaadja az előzőleg ellenőrzött, és helyesnek talált szó szótári tövét. (Több szálát kezelő programban nem használható.)

```
int get_prevcompound();
```

A visszatérési érték 0, ha az előzőleg ellenőrzött szó nem összetett szó volt, különben 1. (Több szálát kezelő programban nem használható.)

```
int get_forbidden_compound();
```

Az előzőleg ellenőrzött szóról adja meg, hogy tiltott volt-e, vagy sem (rendelkezett tiltókapcsolóval). Ha nem volt tiltott, a visszatérési érték 0, ha igen, 1, és ha még összetett is volt, akkor 2. Belsőleg annak eldöntésére szolgál, hogy a javaslatokban szóköz, vagy kötőjel szerepeljen. A Hunspell pedig annak jelzésére használja, hogy egyértelmű szótári tiltásról van szó, megerősítve ezzel a felhasználót helyes működéséről.

```
char * get_possible_root();
```

Visszaadja az előzőleg ellenőrzött szóra illeszkedő első ragozási szabály alapján előálló tövet. Ez a függvény a későbbiekben bővítve lesz. (Több szálát kezelő programban nem használható.)

## Példaprogram

A függvénykönyvtár example.cxx programja rövid összefoglalását adja a fontosabb függvényeknek:

```
#include <cstring>
#include <cstdlib>
```

```
#include <stdio>

#include "myspell.hxx"

#ifdef WINDOWS
using namespace std;
#endif

int main(int argc, char **argv)
{
    FILE *wtclst;
    int i;
    int dp;
    char buf[101];
    MySpell *pMS;
    char **wlst;

    /* first parse the command line options */

    for (i = 1; i < 3; i++)
        if (!argv[i]) {
            fprintf(stderr, "correct syntax is:\nexample affix_file");
            fprintf(stderr, "dictionary_file file_of_words_to_check\n");
            exit(1);
        }

    /* open the words to check list */

    wtclst = fopen(argv[3], "r");
    if (!wtclst) {
        fprintf(stderr, "Error - could not open file to check\n");
        exit(1);
    }

    pMS = new MySpell(argv[1], argv[2]);
    while (fgets(buf, 100, wtclst)) {
        *(buf + strlen(buf) - 1) = '\0';
        dp = pMS->spell(buf);
        if (dp) {
            fprintf(stdout, "\"%s\" is okay\n\n", buf);
            int ns = pMS->suggest_stems(&wlst, buf);
            if (ns > 0) {
                fprintf(stdout, "  stems:\n");
                for (int i = 0; i < ns; i++) {
                    fprintf(stdout, "    ...\"%s\"\n", wlst[i]);
                    free(wlst[i]);
                }
            }
        }
    }
}
```



```
    }
    fprintf(stdout, "\\n");
}
if (wlst) free(wlst);
} else {
    fprintf(stdout, "\\\"%s\\\" is incorrect!\\n", buf);
    fprintf(stdout, " suggestions:\\n");
    int ns = pMS->suggest(&wlst, buf);
    for (int i = 0; i < ns; i++) {
        fprintf(stdout, " ...\\\"%s\\\"\\n", wlst[i]);
        free(wlst[i]);
    }
    fprintf(stdout, "\\n");
    if (wlst) free(wlst);

    ns = pMS->suggest_pos_stems(&wlst, buf);
    if (ns > 0) {
        fprintf(stdout, " possible stems:\\n");
        for (int i = 0; i < ns; i++) {
            fprintf(stdout, " ...\\\"%s\\\"\\n", wlst[i]);
            free(wlst[i]);
        }
        fprintf(stdout, "\\n");
    }
    if (wlst) free(wlst);
}
}

delete pMS;
return 0;
}
```

## 5. fejezet

# Hunspell kézikönyvoldalak

### 5.1. Hunspell(1)

hunspell(1)

hunspell(1)

NÉV

hunspell - helyesírás-javító és szótövező program

ÁTTEKINTÉS

hunspell [-1 -a -G -h -H -l -L -n -s -t] [-d szótár] [-p saját\_szótár] [állomány(ok)]

LEÍRÁS

A hunspell lspell-szerű interaktív és csőfelület a (Magyar) MySpell függvénykönyvtárhoz. Felismeri és javítja az elütéseket és típushibákat sima szöveges, (La)TeX, HTML, és groff (kézikönyv oldal) állományokban. A -s kapcsolóval szótövező üzemmódban fut (Hunstem).

Ha nem adunk meg állománynév paramétert, a Hunspell az lspell csőfelületéhez hasonlóan működik: a bemenet sorait szavakra bontja, és \*, +, vagy - jellel kezdődő kimeneti sorral jelzi, ha egy szó jó, és # vagy & jellel, ha rossz. A csőfelület parancsait viszont az lspellhez hasonlóan csak a -a kapcsoló megadásával értelmezi.

Állománynév paraméter esetén elindul egy az lspelléhez hasonló interaktív felület.

A program a következő előnyöket nyújtja a hasonló célokat

szolgáló Ispell-lel szemben:

- Helyes javaslatok a több karakter változásával járó tipikus hibákra (j/ly, íj/ijj, nyj/nny és még több tucatnyi tévesztés megvizsgálásával). Az Ispell csak két karakter felcserélése; vagy egy karakter hiánya, cseréje, illetve felesleges jelenléte esetén ad helyes javaslatot.
- Az összes helyes összetett szónak látszó, de tipikus hibával előálló alak tiltása (karvaj, szintű, súlytó stb.)
- Javaslatok valószínűségi sorrendben jelennek meg (tipikus hibák, i/i, o/ó, u/ú tévesztések stb.).
- Összetett szavak esetén 6-3-as szabály alkalmazása (helyesírás-ellenőrző, gépkocsi-összeszerelés).
- Mozgószabály alkalmazása (kiviteli engedély, kiviteliengedély-kérés).
- A magyar nyelv egyéb összetételi szabályainak alkalmazása (például hatlövetű, tizenkét lövetű, kéthavi, két hónapos, másodmagával).
- Leg-, legesleg- és -bb confixum helyes kezelése (nincs pl. legédes, csak legédesebb).
- Ragozható tőszófelvétel. Megkönnyíti egy új szó ragozott változatainak ellenőrzését, feleslegessé téve ezek külön-külön történő felvételét a saját szótárba.
- Javaslatok mássalhangzó-triplázások (sakkkör, balettáncos, dzsesszszak) esetén. (Ispell helyesnek fogadja el ezeket.)
- Javaslatok szótagduplázások (oktatás, igenenevet) esetén.
- Javaslatok ékezet nélküli szövegre (tukorfurogep->tükörfúrógép).
- Kötőjeles szavak kezelése (unos-untalan).

- Toldalékolt számok (1-jén, 11-én, 5.-et) kezelése.
- °, %, és § jelek (%-kal) toldalékolt alakjainak kezelése.
- Mozaikszavak (MÁV-osokat) kezelése.
- Y-ra végződő szavak (boyjal, pennyvel) kezelése.
- Idegen ejtésű szótagra végződő szavak (Anonymusszal, Voltaire-nek) kezelése.
- Két szóból álló földrajzi nevek (San Franciscó-i) kezelése.
- Hibás alakok tiltása (Babitszal, tanit, alkotóművész stb.)
- Tiltások jelzése a felhasználói felületen.
- Egyéb (például az Ispell ragozott utószó esetében bármilyen szóösszetételt elfogad: macskatekerem, kutyakavarom, ezt a Hunspell nem teszi meg).
- Magyar nyelvű (illetve locale (7) függő) felület.
- Hordozható szöveges szótárállományok, szemben az Ispell platform, és bináris Ispell fordítástól függően (nem) kezelhető adatállományaival.

## KAPCSOLÓK

- 1 Speciális formátum: a bemeneti állomány minden sorából csak az első szót vizsgálja meg.
- a Ispellhez csőfelület. A !, +, -, @, #, vagy ^ jellel kezdődő sorok a csőfelület parancsai, jelentésüket az Ispell kézikönyvoldal részletezi. A -a kapcsoló hatására a kimenetben a jó szavakat csak csillaggal jelzi a Hunspell. Ez nagyobb kompatibilitást biztosít az Ispellt használó programok némelyikével (például a LyX-szel). Ekkor nem kell megadnunk a -a kapcsolót, mivel az Ispellt használó programok automatikusan ezzel a kapcsolóval indítják el az Ispell nevében futó Hunspell is.

-d szótár

Meghatározza a helyesírási szótárt. Az alapértelmezett szótár a `/usr/share/hunspell/default`. A megadott útvonal lehet abszolút és relatív is. Relatív útvonal esetén először az aktuális, majd a `/usr/share/hunspell` könyvtárban keresi a szótárt a program. L. PÉLDÁK.

-G A -l, és -L kapcsolókat kiegészítő kapcsoló. A standard bemenetről érkező hibás szavak, illetve sorok megjelenítése helyett a jó szavak és a hibátlan sorok megjelenítését kérhetjük vele.

-h, --help

A beépített leírás megjelenítése.

-H HTML bemeneti állományformátum beállítása. Ha az állomány nevének `.html`, `.htm`, vagy `.xml` a kiterjesztése, ez automatikusan bekövetkezik.

-l Csak a hibás szavakat írja ki a standard bemenetről érkező állományból. Használható a -H, -t, -n, -1 kapcsolókkal együtt is. A -G kapcsoló megfordítja a Hunspell működését: csak a jó szavak kerülnek kiírásra.

-L Csak a hibás szavakat tartalmazó sorokat írja ki a standard bemenetről érkező állományból. Használható a -H, -t, -n, -1 kapcsolókkal együtt is. A -G kapcsoló megfordítja a működését: csak azok a sorok kerülnek kiírásra, amelyek egy hibás szót sem tartalmaznak.

-n Nroff/troff bemeneti állományformátum beállítása, elsősorban a kézikönyvlapok ellenőrzésére.

-p saját\_szótár

A felhasználó saját szótárának beállítására szolgál.

Az alapértelmezett saját szótár a `$HOME/.hunspell_default` állomány. Ha szótárt is megadunk a -d kapcsolóval, vagy a `DICTIONARY` környezeti változóval, a saját szótár a `$HOME/.hunspell_szótárnév` lesz. Ha a saját szótár

nem létezik, létrehozza.

Ha az aktuális könyvtárban is van egy ugyanilyen nevű állomány, akkor annak tartalma is beolvasásra kerül, mint saját szótár, és az újonnan felvett szavak is ide kerülnek.

A `-p`, vagy a `WORDLIST` környezeti változó megadásával új saját szótárat jelölünk ki. A saját szótárat az aktuális könyvtárban, és a saját könyvtár gyökerében (`$HOME`) keresi a Hunspell. Ha nem található, akkor az újonnan felvett szavak mentésénél létrehozza a saját könyvtár gyökerében.

- s Tokenizálás és szótövek előállítás. Amennyiben állományokat adunk meg paraméterként, a fájlnevek is megjelennek a kimenetben, lehetőséget nyújtva az állományok indexelésére. Ez utóbbiban a `makedb`, és a `lookdb` parancsok segítenek.

A tövező üzemmód kimenete a következő: egy bemeneti szóhoz tartozik egy normalizált szóalak, majd további sorokban a lehetséges tövek, végül egy üres sor. Az ellenőrző által nem ismert szavak egy kettőskereszttel és egy szóközzel bevezetve kerülnek a kimenetbe.

- t LaTeX bemeneti állományformátum beállítása. Ha az állomány nevének `.tex` a kiterjesztése, ez automatikusan bekövetkezik.
- v Verziószám.
- vv Verziószám (Ispell kompatibilitás miatt).

## INTERAKTÍV FELÜLET

A felület legfelső sorában a hibás szóalakot, és az éppen vizsgált állomány nevét látjuk. Amennyiben egy szóalak szótári tiltás miatt nem kerül elfogadásra, a bal felső sarokban a **TILTOTT! (FORBIDDEN!)** üzenet is megjelenik. Ez az üzenet arra utal, hogy nem hiányos szókincs, hanem szándékos tiltás miatt nem került a szó elfogadásra. Az első sor alatt az állomány hibás szót tartalmazó sorát, és szöveggörnyezetét láthatjuk, majd ezt követik a javaslatok, ha vannak.

A használható billentyűkombinációk:

Szóköz Tovább lépés javítás nélkül a következő hibás szóra.

Számok A megfelelő javaslat kiválasztása.

- c Cseré. A hibás szó cseréje a megadott szóra, vagy szavakra. Az Escape billentyű kétszeri lenyomásával megszakíthatjuk a cserét.
- j A szó elfogadása a program futásának befejeződéséig.
- f A szó felvétele a saját szótárba.
- k A szó kisbetűsként való felvétele a saját szótárba.
- t Ragozott tőszó felvétele. A ragozás a másodikként megadott, mintául szolgáló szó alapján történik. Ha ez a szó nincs felvéve a szótárba, a tőszófelvételt előlről kell kezdeni.

Példa: gyakran hiányoznak az idegen és a mozaikszavak. Vegyük fel a hiányzó OTP szót a tv köznévi mozaikszó toldalékaival! Lenyomjuk a t billentyűt, beírjuk, hogy OTP, Enter, beírjuk, hogy tv, és Enter. A tv szó ugyanis már szerepel a szótárban, és ragozása mondja meg a programnak, hogy az OTP szót milyen toldalékokkal szeretnénk látni, pl. OTP-hez, OTP-met stb. a tv-hez, tv-met mintájára. A -val/-vel, -vá/-vé toldalékokat a kötőjellel kapcsolt toldalékú szavak esetében külön kell felvenni, de még így is sokkal kényelmesebb a speciális tőfelvétel, mint minden (akár több száz) toldalékos alakot külön-külön felvenni.

Egy kis segítség a minta kiválasztásához: első menetben a hangrend számít. Pl. mozaikszavaknál a mély hangrendű új szavak felvételénél használhatjuk a MÁV szót példaként, ajakréses magas hangrendnél pedig az említett tv-t. Ajakkerekítéses magas hangrend esetében a HÖK szót használhatjuk (amíg át nem lesz javítva a helyes hők-re). Ha a hiányzó szó nem mozaikszó, hanem például tulajdonnév, a Pál, Péter, Örs szavakat adhatjuk meg példának a hangrendtől függően. Ha meg egyszerű köznévről van

szó, akkor használjuk az ok, ék, ük szavakat. Igéknél áll, vél, ül, ikes igéknél ázik, vérzik, nőzik lehet a példa (remélhetőleg a következő változatokban egyszerűsödik ez a töfelvétel).

Az Escape billentyű kétszeri lenyomásával megszakíthatjuk ezt a műveletet.

Az itt felvett szavak automatikusan a saját szótárba kerülnek. Ha ezt innen törölni szeretnénk, a programból kilépve a saját szótár kézi átszerkesztésével tehetjük meg.

- m Kilépés a változtatások mentése nélkül. A program változtatás megléte esetén megerősítést kér.
- v Az állomány ellenőrzésének megszakítása a változtatások mentésével. Ha van következő állomány, akkor a program annak ellenőrzését kezdi el.
- ? Segítség. Hasonló rövid leírás megjelenítése a billentyűparancsokról.

## CSŐFELÜLET

A MIsPELL soronként dolgozza fel az állományokat, a helyes szavakat \* (tőszó), - (összetett szó), vagy + (ragozott szó) karakterrel jelöli, a helyteleneket # (nincs javaslat), vagy & (van javaslat) karakterrel. A + jelet a kimenetben követi még a szótó. A # jel után a hibás szó, és kezdőpozíciója van feltüntetve. A & jelet követi a hibás szó; a javaslatok száma; a hibás szó kezdőpozíciója; majd kettőspont után a javaslatok, vesszővel elválasztva:

```
macska
*
macskabajusz
-
macskák
+ macska
mcsk
# mcsk 0
macka
& macka 7 0: macska, maca, racka, packa, vacka,
```



marka, mackó

## PÉLDÁK

hunspell -d en\_US english.html

Az amerikai angol helyesírási szótár kiválasztása (aktuális, vagy a /usr/share/hunspell könyvtárból).

hunspell -d /opt/OpenOffice.org1.0.1/share/dict/ooo/de\_DE

A német helyesírási szótár kiválasztása a megadott abszolút útvonalról, a standard bemenet feldolgozásával.

hunspell \*.html

Az aktuális könyvtárban található összes HTML állomány interaktív ellenőrzése.

hunspell -l szöveg.html

A megadott állomány hibás szavainak kiírása.

hunspell -r \*.html

Az aktuális könyvtárban található összes HTML állomány szavainak, és a szavak szótöveinek kiírása.

## KÖRNYEZET

### DICTIONARY

A szótár helyének megadására szolgál. Szerepe megegyezik a -d kapcsolóval.

### WORDLIST

A saját szótár helyének megadására szolgál. Szerepe megegyezik a -p kapcsolóval.

## ÁLLOMÁNYOK

/usr/share/default.aff Az alapértelmezett ragozási táblázat. Lásd hunspell(4).

/usr/share/default.dic Az alapértelmezett szótár az előző ragozási táblázathoz. Lásd hunspell(4).

\$HOME/.hunspell\_default Az alapértelmezett saját szótár. Az állományt létrehozza a Hunspell, ha nem létezik. Lásd ispell(1).

## LÁSD MÉG

hunspell (4), hunstem (1), makedb (1), lookdb (1), ispell (1), ispell (4), Magyar Ispell dokumentáció.

## SZERZŐ

A Hunspell az International Ispell lényegi részének C++-os újramegvalósításának, a MySpell függvénykönyvtárnak a bővítése. Nemcsak felület a könyvtárhoz, hanem egy jelentős bővítés is a magyar nyelv vonatkozásában (Magyar MySpell).

A mintaként szolgáló International Ispell szerzője Geoff Kuenning, sok más elődje munkájára támaszkodva (l. ispell(1)).

A MySpellt Kevin Hendricks készítette a kiváló és ingyenes OpenOffice.org (eredetileg StarOffice) irodai programcsomaghoz 2001-2002-ben.

A Magyar MySpellt, illetve a Hunspell felületet Németh László <nemethl@gyorsposta.hu> készítette 2002-2003-ban.

2003 márciusától a fejlesztést a Szószablya projekt viszi tovább egy nyitott forráskódú magyar morfológiai elemző elkészítésének irányába.

A Hunspell a Magyar Ispell projekt eredményét, a magyar Ispell/MySpell helyesírási szótármodult használja a magyar nyelvű szövegek ellenőrzésekor. A Magyar Ispell levelezőlista a magyarispell@yahoogroups.com címen található. Várunk minden olyan észrevételt, illetve segítséget, ami jobbat tehet a Magyar MySpell helyesírási-ellenőrzőt, és a Magyar Ispell szótármodult.

Az International Ispell-nek, és így a MySpellnek is magyar vonatkozása is van. A programokban szereplő affixumtömörítés a Dömölki-algoritmuson alapszik, Dömölki Bálint egyik 60-as években publikált algoritmusán, amit a magyar M3-as, az első hazai elektronikus számítógép megépítését követően, nyelvészeti jellegű kutatások kapcsán dolgozott ki.

## HIBÁK

A saját szótár mérete jelenleg 1000 szóban van korlátozva. A forrásban a MAXUSERWORD makró beállításával állítható be a kívánt érték.

Az UTF-8, vagy UTF-16 karakterkódolású állományok ellenőrzését még nem támogatja.

A sorok maximális mérete nincs korlátozva, de a 8 kB-os határokon tévesen hibát jelezhet a program. A szóköz lenyomásával ugorjuk át a téves hibajelzést!

Az e-mail címek, URL-ek, és egyéb útvonalak ellenőrzését elkerülhetjük, ha ezeket a javasolt módon; HTML esetében a `<var>útvonal</var>`, LaTeX esetében `\url{útvonal}` formában adjuk meg. (A címek elválasztást is lehetővé tevő `\url` LaTeX parancs az `url` csomagban található, tehát használatához a `\usepackage{url}` parancs kiadására is szükség van a preambulumban.)

2003. június 13.

hunspell(1)

## 5.2. Hunstem(1)

hunstem(1)

hunstem(1)

### NÉV

hunstem - szótövező program

### ÁTTEKINTÉS

hunstem [Hunspell kapcsolók] [állomány(ok)]

### LEÍRÁS

A Hunstem a Hunspell program egyik felülete, amely szótövezésre és állományok indexelésére szolgál. A hunstem parancs megfelel a hunspell -s hívásnak.

A Hunstem egy lehetséges felhasználására a makedb, és a lookdb parancsok mutatnak példát.

### LÁSD MÉG

hunspell (1), makedb (1), lookdb (1), Magyar Ispell dokumentáció.

### 5.3. Makedb(1)

makedb(1)

makedb(1)

#### NÉV

makedb - webhelyek indexállományának létrehozása

#### ÁTTEKINTÉS

makedb [-d könyvtár [-r]] [-o állomány]

#### LEÍRÁS

A makedb parancs az aktuális könyvtár és alkönyvtárainak txt, htm vagy html kiterjesztésű állományait a Hunstem segítségével indexeli, és annak kimenetét a Lookdb által kereshető formára hozza, majd hunstem.index néven elmenti az aktuális könyvtárba.

#### KAPCSOLÓK

-d könyvtár

Az indexelni kívánt könyvtárrendszer gyökere.  
Alapesetben ez az aktuális könyvtár.

-o állomány

A program kimeneti állománya. Alapesetben ez a ./hunstem.index.

-r A kimenet relatív útvonalakat fog tartalmazni a -d kapcsoló használata esetén is.

#### PÉLDÁK

```
makedb -d /var/www -o /var/www/hunstem.index -r
```

A parancs egyenértékű a következővel:

```
cd /var/www; makedb
```

#### LÁSD MÉG

lookdb (1), hunspell (1). A makedb rövid és jól dokumentált héjprogram, ezért érdemes a forrását is megtekinteni.

2003. június 13.

makedb(1)

## 5.4. Lookdb(1)

lookdb(1)

lookdb(1)

### NÉV

lookdb - keresés a makedb paranccsal előállított indexállományban

### ÁTTEKINTÉS

lookdb hunstem\_ adatbázisfájl kulcsszó(k)

### LEÍRÁS

A lookdb parancs gyors keresést hajt végre a makedb paranccsal létrehozott indexállományban. A program kimenete azon állományok listája, amelyek minden megadott kulcsszót tartalmaznak.

Mivel a makedb a hunstem-et futtatja, aminek hatására az indexelt állományok szavainak töve is bekerül a makedb adatbázisba, a lookdb kimenete azokat az állományokat is tartalmazni fogja, amelyek a megadott kulcsszavak toldalékolt alakjait tartalmazzák csak. Például a bokor kulcsszóra a bokrot szót tartalmazó állományok is kiírásra kerülnek. Fordítva, vagyis nem szótő kulcsszót, hanem ragozott alakot használva ez már nem lesz igaz.

Ha nem pontos, hanem szó eleji egyezés alapján akarunk keresni, akkor helyezzünk csillagot a kulcsszó végére. Ennek nagy jelentősége van, mivel az ismeretlen szavak (többnyire tulajdonnevekről van szó) toldalékolt alakjait nem képes a Hunstem tövezni, ezért itt erősen javasolt a csillag használata).

### PÉLDÁK

lookdb hunstem.index macska egér 'Verseghy\*'

A keresés eredménye azon állományok listája, amelyek tartalmazzák a macska ÉS az egér szót (vagy

ezek toldalékolt alakjait) ÉS egy „Verseghy”-vel kezdődő szót.

#### LÁSD MÉG

makedb (1), hunspell (1), look (1). A Lookdb rövid és jól dokumentált héjprogram, ezért érdemes a forrását is megtekinteni.

2003. június 13.

lookdb(1)

## 5.5. Hunspell(4)

hunspell(4)

hunspell(4)

#### NÉV

hunspell - a hunspell állományainak formátuma

#### LEÍRÁS

A Hunspell(1) két állományt igényel egy adott nyelven történő helyesírás-ellenőrzéshez. Az első állomány egy szótárállomány, ami az adott nyelv szavait, és ezek kapcsolóit (attribútumait) tartalmazza. A második állomány egy ragozási (affixum) állomány, ami meghatározza az említett kapcsolók jelentését. A hunspell, ellentétben az ispell-lel, közvetlenül használja fel ezeket az állományokat, tehát nem képez belőlük ún. hash állományt (l. ispell(4)).

#### SZÓTÁRÁLLOMÁNY

Egy hunspell szótárállomány első sora a szótárállományban lévő szavak számát tartalmazza. A többi sor soronként egy szót tartalmaz. Minden szó után nem kötelezően egy perjel, és egy, vagy több kapcsoló következhet. A kapcsolók egybetűsek, és meghatározzák, hogy milyen toldalékokat kaphat a tőszó. Például a

1  
hunspellezik/EeX

szótár egy szót tartalmaz (hunspell), ami a magyar ragozási állomány alapján alanyi és tárgyas igeragozást is vonz (E és e kapcsolók), valamint igekötő prefixumokat is kaphat. A példa szótárállomány és a magyar ragozási állomány alapján a meghunspelleztetek helyes szóként kerülne felismerésre.

## RAGOZÁSI ÁLLOMÁNY

A ragozási állomány rendszerint kisebb méretű, és különféle definíciók mellett a ragozási szabályokat adja meg.

```
SET ISO8859-2
TRY íóúÍÓÚtaeslzánorhgkiédmyőpvöbucfjüűxwqTAESLZÁNO
RHGKIÉDMYŐPVÖBUCFJÜŰXWQ-.
COMPOUNDMIN 2
COMPOUNDFLAG Y
COMPOUNDWORD 2 y
COMPOUNDSYLLABLE 6 aáééííoóóőóúúüü
SYLLABLENUM klmc
COMPOUNDFIRST v
COMPOUNDLAST x
FORBIDDENWORD w
ONLYROOT u
ACCENT áéíóőóúúüü aeiooouuu
CHECKNUM
WORDCHARS -.§%°0123456789

REP 2
REP j ly
REP ly j

PFX X Y 3
PFX X 0 meg .
PFX X 0 rá .
PFX X 0 össze .

SFX e Y 3
SFX e ik tük e[dz]ik
SFX e ik téték e[dz]ik
SFX e ik ték e[dz]ik
```

A kulcsszavak jelentése:

SET A karakterkészlet beállítása (ISO-8859-X).

TRY A javaslattevésnél az 1 karakteres próbálkozásoknál az itt felsorolt karakterekre cserél, illetve bővít.

#### COMPOUNDMIN

A legkisebb méretű szó, ami még összetett szóban szerepelhet. Szerepe csak a COMPOUNDFLAG megadása esetén van.

#### COMPOUNDFLAG

Itt adható meg a szóösszetételben való szereplést engedélyező kapcsoló. Ha hiányzik ez a definíció a ragozási állományból, vagy nincsen használva az itt megadott kapcsoló a szótárállományban, nincsen összetettség-kezelés az adott nyelvben.

#### COMPOUNDWORD

Egy, vagy két paramétert vár. Az első paraméter meghatározza, hogy hány tagszó esetében nem kell még alkalmazni a szótagszámra vonatkozó korlátozásokat az összetett szavaknál. A nem kötelező második paraméterrel azt a kapcsolót adhatjuk meg, ami a szótárba közvetlenül felvett összetett szavakat jelöli majd. (Magyar MySpell)

#### COMPOUNDSYLLABLE

Két paramétert vár. Az első megadja az összetett szavak tövének maximális szótagszámát. A második felsorolja a magánhangzókat, amely a szótagszám megállapításához szükséges. Ha hiányzik ez a definíció, akkor nincsen korlátozás az összetett szavak tövének szótagszámára vonatkozóan. (Magyar MySpell)

#### SYLLABLENUM

Egy paramétert vár. A paraméterben megadhatóak azok a ragozási osztályok, amelyek ragozás esetén eggyel növelik az összetett szavak tövének kiszámított szótagszámot. A magyarban ezek a melléknévképzők: kerékpárjavításainkról, de kerékpár-javítási. (Magyar MySpell)

#### COMPOUNDFIRST

Az a kapcsoló, amely jelöli azokat a szavakat, amelyek az összetett szavak elején fordulhatnak



elő. (Magyar MySpell)

#### COMPOUNDLAST

Az a kapcsoló, amely jelöli azokat a szavakat, amelyek az összetett szavak végén fordulhatnak elő. Használható az előzővel együtt is. (Magyar MySpell)

#### FORBIDDENWORD

A tiltott szavakat jelölő kapcsoló. A tiltott szótövek toldalékolt alakjai is letiltásra kerülnek. (Magyar MySpell)

#### ONLYROOT

Az előző kapcsoló működését befolyásoló kapcsoló. Ha egy szó a FORBIDDENWORD és az ONLYROOT kapcsolóval is rendelkezik, akkor csak a tőalak lesz letiltva, a tő toldalékolt alakjai nem. (Magyar MySpell)

ACCENT Az ékezetesítéshez szükséges betűket két paraméterrel adhatjuk meg itt. Az első paraméter az ékezetes karaktereket, a második az azoknak megfelelő ékezet nélküli párokat adja meg. (Magyar MySpell)

#### CHECKNUM

A számokat tartalmazó szavak ellenőrzésének bekapcsolása. (Magyar MySpell)

#### WORDCHARS

Paramétere azok a karakterek, amelyek a beállított kódolás (pl. a magyar esetében az ISO-8859-2) betűi mellett még szavakban előfordulhatnak. (Magyar MySpell)

REP A javítási cseretáblázat megadására szolgáló definíciók. Az első REP után adjuk meg a többi REP sor számát. Ezek a sorok már két paramétert várnak: először megadjuk a felismerendő mintát (szövegrészletet), utána pedig azt a karaktorsorozatot, amire a felismert mintát cserélni szeretnénk. A helyesírás-ellenőrző a cseretáblázat révén tud helyes javaslatot tenni többkarakteres eltérés (pl. j/ly) esetén. A Magyar Ispellben meglévő keresztellenőrzés miatt egy karakteres cserék (i/i) is szerepelnek a REP

táblázatban. Ennek köszönhető, hogy az amúgy helyes nem szótári szóösszetételek nem kerülnek elfogadásra, ha a cseretáblázat egyik cseréjével előállíthatók egy szótári szóból (például: a „szintű” szó hibás összetett szó lesz, mert az i->í cserével előállítható a szintű szóból).

PFX Prefixum megadására szolgáló definíciók. A prefixumok egykarakteres névvel jelölt osztályokba vannak sorolva. A példában az X osztály definíciója látható. Az egyes osztályokra vonatkozó PFX fejléc tartalmazza a következőket:

(1) az osztály betűjelét (a példában X)

(2) a keresztbe való kombinálás engedélyezése, illetve tiltása (Y/N). Engedélyezés esetén az adott prefixum akkor is alkalmazható, ha a tőszó toldalékolt alakban fordul elő. Ehhez az is szükséges, hogy a toldalék (affixum) ragozási osztályában is legyen engedélyezve a kombinálás.

(3) az osztályba tartozó bejegyzések számát

A PFX fejléct követő sorokban a PFX paraméterek jelentése a következő:

(1) az osztály betűjele

(2) a ragozásnál a szó elejéről leválasztandó karaktersorozat. Ha nem kell leválasztani, akkor 0.

(3) a prefixum (például igekötő, vagy a leg-, legesleg- a mellékneveknél).

(4) a minta, ami ha illeszkedik, alkalmazható a szabály. Például ha itt kutya szerepelne, akkor csak a kutya kezdetű szavak elé lehetne tenni az ebben a sorban megadott prefixumot. A minta megadásánál a pont (.) speciális jelentéssel bír: minden karakterre illeszkedik. Megadhatunk csak bizonyos karakterekre illeszkedő mintát is szögletes zárójelek segítségével. Például az [abcd] illeszkedik az a, b, c, illetve d karakterekre is. Vigyázat! Ellentétben az Ispell-lel, a kötőjel

szerpe a szögletes zárójelek között nem speciális, vagyis kötőjelet jelent, és nem karaktertartományt. Ha a nyitó szögletes zárójel után kalap (^) karakter található, a karaktertartomány komplementerére illeszkedik a minta, hasonlóan a reguláris kifejezésekhez (l. regex(7)).

SFX A PFX-szel megegyező paramétereket kap, csak nem a szó eleji, hanem a szó végi toldalékok leírására szolgál. A példában szereplő definíciók jelentése: az -edik, vagy -ezik végű szavakra alkalmazható a három szabály. A rag illesztésénél levesszük a szó végi -ik karaktorsorozatot, és szabálytól függően illesztjük a -tük, -tétek, -ték toldalékokat, végeredményben -edtük, -edtétek, -edték, illetve -eztük, -eztétek, -ezték végű szavakat kapva.

A ragozási állomány ezeken kívül még megjegyzéseket is tartalmazhat, ezért az adott szótár változatát, szerzőjét, felhasználási engedélyeit itt keressük.

## SAJÁT SZÓTÁR

A saját szótár formátuma közel megegyezik a szótárállomány formátumával. A különbségek a következők:

(1) Nem kell megadni az első sorban az állományban lévő szavak méretét. (Ne is adjuk meg!)

(2) A sorban a kapcsolók helyett megadhatunk két perjellel elválasztva egy mintaszót is, aminek a kapcsolóit örökli a későbbiekben a saját szótárba felvett szó. Például a Patyomkin//Ádám sor jelentése, hogy a Patyomkin szót az Ádáméval megegyező ragozással is el kell fogadni.

(3) Az állomány kézzel módosítható, és ezt időnként javasolt is megtenni, ha esetleg hibásan vettünk fel szavakat.

## LÁSD MÉG

hunspell (1), ispell (1), ispell (4), Magyar Ispell dokumentáció.

## TÖRTÉNET

A MySpell szótár-, illetve ragozási állományának formátumát Kevin Hendricks készítette olyan céllal, hogy lehetővé tegye ezen állományok gyors, futás idejű

feldolgozását.

A COMPOUNDFLAG és a Magyar MySpell formátumbővítést, valamint a REP többkarakteres cseretáblázatot Németh László készítette elsősorban a jobb magyar nyelvű helyesírás-ellenőrzés céljából. A minimális összetettszó-kezelést lehetővé tevő COMPOUNDMIN, COMPOUNDFLAG, és a REP többkarakteres cseretáblázat része a hivatalos MySpell függvénykönyvtárnak is.

A Magyar MySpell szótárállományának formátuma megegyezik a MySpell függvénykönyvtárával, a ragozási táblázat pedig a fentiekben felsorolt bővítéseket tartalmazza. A bővített formátumú affixum állomány gond nélkül együttműködik az eredeti MySpell függvénykönyvtárral is, de a bővítések nem fognak működni. Sőt, mivel a hivatalos MySpell egy megengedőbb összetettszó-kezelést valósít meg, mint a magyar változat, az összetett szavakon belüli toldalékolt alakok is elfogadásra kerülnek (ez különösen sok hibás összetétel elfogadását eredményezi a magyar nyelv esetében).

Ebből fakad, hogy a hivatalos Magyar OpenOffice.org (<http://office.fsf.hu>) - mivel a Magyar MySpell függvénykönyvtárat tartalmazza - jobban támogatja a magyar nyelvet, mint más változatok. Egyszerűen meggyőződhetünk arról, hogy melyik változattal van dolgunk: írjuk be, hogy helyesírásellenőrző az OOo szövegszerkesztőjébe. Ha hibásnak jelzi, és helyesírás-ellenőrzőt javasol helyette, a Magyar MySpell-lel van dolgunk. Ha helyesnek fogadja el, akkor pedig az eredeti MySpell-lel.

2003. március 25.

hunspell(4)

## 6. fejezet

# A Magyar Ispell tesztelése

A tesztrendszer a <http://www.szofi.hu/gnu/magyarispell/magyarispell.zip> állományban található.

A Magyar Ispell működésének tesztelésére egy független szókincstár, a Huhyph 4.0 elválasztási szóadattár szolgált.

A tesztrendszer két tesztelést végez el: az első a Huhyph szókincstár felismerését teszteli, a második a HuHyph-ból kinyert véletlen mintából különféle algoritmusokkal előállított hibás alakok helyesként való elfogadását teszteli.

### 6.1. A szókincs tesztelése

Csomagoljuk ki a magyarispell-teszt.zip állományt, és a könyvtárba lépve adjuk ki a `make test1` parancsot.

A szóadat szavai kötőjellel jelölve tartalmazzák az elválasztási pontokat, valamint fordított perjellel bevezetve, és kapcsos zárójelpárral lezárva a kettőzött többjegyű mássalhangzókat. Ezeket a jeleket az 1. teszt során töröljük, és a Magyar MySpell függvénykönyvtárat használó `badwords` programmal kiszűrjük az állományból a hibás szavakat.

Jelenleg (2002. július) a Magyar Ispell szótármodullal a `huwords.hyph` állomány 98,5%-a helyesnek bizonyul.

A maradék részben (ami mintegy 900 szót jelent a `huwords.hyph` kb. 65 000 szavából) még feldolgozás alatt álló, részben pedig idegen szavakat tartalmaz. Ha ezek elfogadását szeretnénk, az első teszt futtatásánál kapott `GEN.nemismert` állományt vegyük fel a Magyar Ispellben új szó-tári modulként, ragozatlan-ra átkeresztelve az állományt.

## 6.2. Tévesztések ellenőrzése

A tesztelés megkezdése előtt a `teszt.sh` állomány `JAVA_PATH` változójának adjuk meg a java virtuális gép helyét, mivel a téves alakokat előállító program (MI`Test`) Java nyelvű.

Adjuk ki a `make teszt2` parancsot a tévesztések elfogadásának tesztelésére.

A tesztprogram a `huwords.hyph` szavaiból egy véletlen, 1000 szóból álló mintát vesz. Ebből a mintából a MI`Test` program téves alakokat képez az alapvető Ispell javítási algoritmusokkal (betűkihagyás, betűbetoldás, betűcsere, szomszédos betűk felcserélése).

A teszt során három állomány keletkezik: A `GEN.osszes` állomány tartalmazza az összes előállított téves alakot. Ez mintegy 750-szerese az eredetileg 1000 szavas véletlen mintának.

A `GEN.osszes` állományból a `badwords` program kiszűri a helytelennek felismert szavakat, ez kerül a `GEN.hibas` állományba.

Végül ezeket a szavakat kivonjuk a `GEN.osszes` állományból, így kapjuk meg a `GEN.helyes` állományt.

A `GEN.helyes` tartalmazza tehát a `huwords.hyph`-ból nyert véletlen minta „elrontásából” keletkezett, mégis helyesnek felismert szóalakokat.

## 6.3. A tesztek értékelése

### 6.3.1. Szókincs

A `huwords.hyph` első feldolgozása során az eredeti állomány mintegy 5%-a, 3700 szó nem került felismerésre.

A listában az idegen, régies, ill. ritka szavak mellett a hiányzó tövek ragozott alakjai is előfordultak, így mintegy 2000 db. hiányzó töről beszélhetünk, amely szám – tekintve a szavak többnyire nem hétköznapi jellegét – jónak mondható!

A hiányzó szavak listája nagyrészt feldolgozásra került Bíró Árpád munkája révén, így sikerült az 5%-ot jelentősen lecsökkenteni.

### 6.3.2. Tévesztések

A `GEN.helyes` szólista a többszörös tesztelések során átlagosan 16-szor bizonyult nagyobb-nak, mint a véletlen minta.

Íme az „álomszerű” szó téves alakjaiból előállt helyes alakok:

álomászerű  
álomészerű  
álomízerű  
álomőszerű

álomőzerű  
álomsierű  
álomsóerű  
álomszedrű  
álomszegű  
álomszelű  
álomszemű  
álomszenű  
álomszer  
álomszerb  
álomszerbű  
álomszere  
álomszeré  
álomszerfű  
álomszeri  
álomszermű  
álomszert  
álomszertű  
alomszerű  
álomszerű  
álomszérű  
álomszerűr  
álomszerűt  
álomszerv  
álomszervű  
álomszexű  
álomszírű  
álomszóerű  
álomszörű  
álomszúerű  
álomszűrű

Látható, hogy a szavak nagy része szokatlan szóösszetételként került elfogadásra.

Az ilyen szóösszetételek a Magyar MySpell 0.4-es változatában részben már nem kerülnek elfogadásra. Szerencsére erre a legtipikusabb hibák esetében (pl. i/í, o/ó, u/ú, ü/ű, j/l, j/ly, stb.) kerül sor, így a Magyar Ispell/MySpell már (ellentétben pl. a Helyes-e?-vel) nem fogadja el az ilyen típusú, tipikus és súlyos tévesztéseket (pl. szervíz, szeméj, súlytó, stb.).

## 7. fejezet

# Szabályok

A következő táblázat a Magyar Ispell és a Microsoft Office XP-ben megtalálható helyesírás-ellenőrző magyar helyesírásnak való megfelelését részletezi.

A vizsgálat során nem minden esetben a Microsoft Office XP, hanem a Morphologic cég Helyes-e? programjának Linux operációs rendszerre 2002-ben megjelent változata, az Mspell 0.1 program szolgáltatta a szükséges információkat, de a tapasztalatok alapján ez megegyezik a Microsoft Office XP-ben található Helyes-e? változattal. A Magyar Ispell képességei a Magyar Ispell szótármodul 0.99.3-as és a Mispell/Magyar MySpell 0.9.4-es változata alapján lettek összefoglalva. Az ellenőrző és a szótármodul ezen változatát a Magyar OpenOffice.org (<http://office.fsf.hu>) közeljövőben megjelenő változatai tartalmazni fogják.

A táblázat a példaként megadott szóalakokat *kurzív* betűváltozattal jelöli. Az ilyen szóalakok, amennyiben a magyar helyesírás szerint hibásak, csillaggal bevezetve jelennek meg (például \**hejesírás*). Amennyiben helyesek, de a helyesírás-ellenőrző mégis hibásnak jelzi ezeket, a Microsoft Office és az OpenOffice.org szövegszerkesztőkben ismerős módon, vagyis hullámos aláhúzással vannak jelölve (például *helyesírás*).

A leírások A magyar helyesírás szabályai 11. kiadása pontjainak megfelelően számozottak. Egy-két kiegészítő megjegyzés külön sorban foglal helyett.



## Általános tudnivalók

1. <sup>1</sup> Nyitott forráskódú program, amely a közérdeknek megfelelően áttekinthető: folyamatosan dokumentált, ellenőrzött, javított és szabadon hozzáférhető mindenki számára.	Zárt, dokumentáció nélkül megjelenő helyesírás-ellenőrző, amely legalább az MS Office 97 óta bármiféle javítás nélkül jelenik meg az MS Office különféle változataiban. Linuxos változata <i>Mspell</i> néven otthoni felhasználásra ingyenes.
A betűk	
3. A magyar ábécé minden betűjét önálló alakban helyesnek fogadja el.	A magyar ábécé többjegyű betűit hibásnak jelzi: <u>dzs</u> <u>betű</u> . Az ékezetes betűk toldalékolt alakjait hibásnak jelzi: <u>hosszú ú-val</u> .
6. A stilisztikai célból megengedett magánhangzó-halmazt nem fogadja el helyesnek: <u>Één?</u> , <u>Neeem</u> . Kivéve <u>góóól</u> , <u>negyven ó-ig bezárólag</u> .	Hasonlóan, de a <u>góóól</u> alakot hibásnak jelzi.
7.b. A mássalhangzók hosszúságát betűket-tőzéssel jelöli. A többjegyű mássalhangzókat csak az összetett szavak tagjainak határán kettőzi teljesen, egyéb esetekben egyszerűsítve <u>kettőzi őket</u> .	Hiba a <u>-szerű</u> utótag összetett szóként való kezelése: <u>*ésszerű</u> , <u>*vadásszerű</u> , a helyes alakok (kivéve <u>ésszerű</u> ) hibásnak jelzése: <u>mésszerű</u> , <u>viasszerű</u> . L. még 62.
8. Nagybetűs formák esetében a megengedetttek vannak elfogadva: <u>csibe</u> , <u>Csibe</u> , <u>CSIBE</u> , de <u>*CSibe</u> , <u>*cSiBe</u> . Gyakorlati megfontolásból csak nagybetűvel (tulajdonnévként) kerül elfogadásra a <u>kis</u> legtöbb toldalékolt alakja (például <u>kissel</u> , de <u>kissé</u> jó).	Hasonlóan jól működik. (L. még 164.)
9. A stilisztikai célból megengedett mássalhangzó-halmazt nem fogadja el helyesnek: <u>Nnem!</u> , <u>Terringettét!</u>	Hasonlóan.

<sup>1</sup>... A helyesírás valamely nyelv írásának közmegállapodáson alapuló és közérdekből szabályozott eljárás módja, illetőleg az ezt tükröző, rögzítő és irányító szabályrendszer. . .

Magyar Ispell	MS Office XP helyesírás-ellenőrzője
<p>13. A program kezeli az UNICODE karaktereket. Alapértelmezett az ISO 8859–2 karakterkészlet:<sup>2</sup> <i>Dvořák, Händel, Škoda, zloty...</i> stb. Az UNICODE karakterek ISO 8859–1 esetén HTML karakterentitással, egyébként pedig decimális karakterkóddal adhatók meg: <i>Moli&amp;egrave;re</i> (Molière), <i>Duna&amp;dash;Tisza köze</i> (Duna–Tisza köze).</p> <p>Az idegen ékezetek esetében is javítva vannak az ékezet nélküli változatok: <i>Skoda</i> → <i>Škoda</i>, <i>Dvorák</i>, sőt <i>Dvorak</i> → <i>Dvořák</i>.</p> <p>A kiejtés szerinti írásmód</p>	<p>Csak a magyar ábécét, és a <i>q, w, x, y</i> betűket kezeli a szótár: <i>Škoda, Molière</i> stb. A <i>*Skoda</i>, vagy a <i>*zloty</i> alakok elfogadása hiba az UNICODE alapú, vagy ISO-8859–2-t támogató szövegszerkesztők – mint az OpenOffice.org, illetve a Microsoft Office – esetében.<sup>3</sup></p>
<p>17–19. Ragaszkodik a helyesírás által rögzített köznyelvi formához. A leglényegesebb hibák (tipikus tévesztések; a hasonló, de nem rögzített helyesírású alakok előfordulása; hibás, de a szótárba mégis felvett szóösszetételek) alapján ellenőrizve lett a szókincs. A teljes szókincs lektorálására azonban még nem került sor.</p>	<p>Számos, a helyesírásban nem rögzített beszélt nyelvi ingadozást helyesnek fogad el. Ilyen, később nem említett példák: <i>*lyány</i> (<i>lány</i>), <i>*siserehad</i> (<i>siserahad</i>), <i>*bédekker</i> (<i>bedekker</i>), <i>*ekhó</i> (<i>echó</i>), <i>*trikolor</i> (<i>trikolór</i>), stb. Valahol csak a rossz alakot engedi: <i>*puzon</i> (<i>puzón</i>). L. még 28.</p> <p>Egyéb probléma, hogy nem létező szavakat, illetve nem szótári töveket tartalmaz a szótár: <i>*ál, *fesz; *sepr, *sodr; *süv *veé</i> stb.</p> <p>Nem fogad el számos mai magyar szót, különösen az új, szaknyelvi és földrajzi elnevezéseket: <i>karalábé, klónozás, ombudsman, szafari, helló</i> vagy <i>heló</i> (helyette <i>*hello</i>) stb.</p> <p>A szótár pontatlan, illetve elnagyolt besorolásokat tartalmaz. Például hangrendi tévedések: <i>*tettról</i> (<i>tettről</i>), <i>*vésznak</i> (<i>vésznek</i>) stb. Elnagyolások: <i>*fiája, *híjája, *aljájáról</i> (többszörös birtokos személyragok) stb.; <i>*történed, *halja, *évődi</i> stb. (a nem tárgyias igék kezelésének hiánya).</p>

<sup>2</sup>A közép-európai nyelvek közül az albán, bosnyák, cseh, horvát, lengyel, német, magyar, román, szerb, szlovák, szlovén nyelvek ábécéjének betűit tartalmazza, továbbá az angol, finn és ír ábécé betűit is lefedi. A Windows operációs rendszer ún. közép-európai kódtáblája (Windows-1250) egy kibővített ISO 8859–2 kiosztásnak felel meg.

<sup>3</sup>Az OpenOffice.org teljes UNICODE támogatással rendelkezik, lehetőséget nyújt még az ázsiai írások kezelésére is. A Helyes-e? és az OpenOffice.org egybeépítésére a kereskedelmi forgalomban kapható Magyar Office mutat példát.

Magyar Ispell	MS Office XP helyesírás-ellenőrzője
Szaknyelvi modulokat (matematika, informatika stb.) tartalmaz több száz szakszóval. A történelmi és a mai magyar helységneveket is tartalmazza, ami a közigazgatási, valamint földrajz és történelem tárgyú munkák ellenőrzése céljából nélkülözhetetlen. A szókincs a felhasználók által is bővíthető.	Nincsenek szaknyelvi modulok: <i>elempár</i> , <i>alterek</i> , <i>variancia</i> , <i>bijektív</i> , <i>attraktor</i> , <i>flopi</i> , <i>mikrocsip</i> , <i>élkiemelés</i> , <i>interpreter</i> , <i>alhálózat</i> . A városi rangú települések neve közül is sokat nem ismer: <i>Gyomaendrőd</i> , <i>Szécsény</i> , <i>Vásárosnamény</i> , <i>Zalaszentgrót</i> stb. Nincs lehetőség a szókincs bővítésére, csak egyes szóalakok felvételére.
20–21. Az <i>i</i> , <i>u</i> , <i>ü</i> és ezek hosszú változatát mindig ugyanolyan formában tartalmazó szavak írásában pontos. <sup>4</sup>	Számos hibás alakot is elfogad: <i>*egyivású</i> ( <i>egyívású</i> ), <i>*elitélt</i> ( <i>elítélt</i> ), <i>*hivó</i> ( <i>hívó</i> ), <i>*izület</i> ( <i>ízület</i> ); <i>*neutrínó</i> ( <i>neutrínó</i> ), <i>*szivató</i> ( <i>szívató</i> ), <i>*telített</i> ( <i>telített</i> ), <i>*tikfa</i> ( <i>tíkfa</i> ); <i>*kompatibilis</i> ( <i>kompatibilis</i> ), <i>*színtű</i> ( <i>szintű</i> ); <i>*labdarúgócsapat</i> ( <i>labdarúgócsapat</i> ); <i>*mélyfúrás</i> ( <i>mélyfúrás</i> ); <i>*ramazuri</i> ( <i>ramazúri</i> ); <i>*harangbugás</i> ( <i>harangbúgás</i> ); <i>*lámpabúra</i> ( <i>lámpabura</i> ); <i>*küzdőtér</i> ( <i>küzdőtér</i> ) stb.
Gyakorisági megfontolásból sok alakpár ritkán használt tagja le van tiltva: <i>csábit</i> , <i>öblit</i> ; <i>visszairt</i> stb. <sup>5</sup>	Egy-két alak van letiltva: <i>tanít</i> , de például a <i>csábit</i> , <i>öblit</i> , <i>visszairt</i> nem.
22. Az <i>i</i> , illetve <i>í</i> végű szavak esetében nincs ismert hiba.	Sok ilyen főnév toldalékolása hibás, illetve hiányos: <i>kocsijai</i> , <i>*kocsiija</i> , <i>*kocsiijai</i> . Hasonlóan <i>zsenijei</i> , <i>mozijai</i> , <i>masnijai</i> stb. A <i>hí</i> toldalékolása is hibás: <i>híttak</i> , <i>*hittak</i> stb.
23–24. Az <i>u</i> , <i>ú</i> , <i>ü</i> , <i>ű</i> végű szavak esetében nincs ismert hiba.	Elfogadásra kerülnek a <i>*hiu</i> , <i>*savanyu</i> , továbbá <i>*átlagáru</i> , <i>*vasárú</i> stb. alakok. A <i>hú</i> , <i>hű</i> indulatszavak mellett helyesli a <i>hu</i> , és <i>hü</i> alakokat.
25. Kezeli a <i>csepp</i> és <i>csend</i> alakváltozatai közötti különbséget.	A nem használt alakváltozatokat is elfogadja: például <i>*csöppkő</i> , <i>*csöndőr</i> .

<sup>4</sup>Az alapszókincre vonatkozik a pontosság. A magyar helységneveket tartalmazó modul sok kettős alakot tartalmaz még.

<sup>5</sup>L. a szotar/kivetelek/ragozatlan/i állományt a Magyar Ispell forrásában.

Magyar Ispell	MS Office XP helyesírás-ellenőrzője
<p>27. Az <i>í, ú, ű</i> szabálytalan és szabályos váltakozását mutató szavak írásában pontos. Gyakorlati megfontolásból viszont nem fogadja el a ritka <u>tűntet</u> és <u>szűntet</u> alakokat (a <u>tűnt</u>, és <u>szűnt</u> tárgyragos változatát).</p>	<p>a) A szabálytalan váltakozást mutató szavak esetében téves alakokat is elfogad: <i>*hivat (hívat), *hivató (hívató); *irogat (írogat), *irogató (írogató), *irogatás (írogatás); *bízgat (bizgat), *bízgató (biztató), *bízgatás (biztat); *szűntető (szüintető), *szűntetés (szüintetés); *tűntető (tüintető), *tűntetés (tüintetés).</i></p> <p>b) A szabályos váltakozást mutató szavak esetében téves a <i>fűz</i> toldalékolása: <i>*fűzek (füzek), *fűzes (füzes)</i> stb. Súlyos hiba mutatkozik még az ilyen szavak I/3. birtokos személyragos plusz birtokjeles toldalékolásában: <i>*vízée (vizée, vízé), *tűzée (tüzée, tűzé), *kézée (kezée, kézé)</i> stb. Hasonlóan a <i>-nyi</i> képzős alakok esetében: <i>*viznyi (víznyi), *tűznyi (tűznyi), *keznyi (kéznyi)</i> stb.</p>
<p>28. A közkeletű idegen szavak magánhangzó-váltakozásában nincs ismert hiba.</p>	<p>A <i>kultúra</i> szó származtatott alakjait helytelenül kezeli: <i>*kultúrál (kulturál), *kultúrált (kulturált), *kultúrátlan (kulturátlan), *kultúrálódik (kulturálódik)</i>. Számos egyéb esetben is elfogadja még a nem rögzített írásmódú alakokat: <i>*analfabétizmus (analfabetizmus), *polemikus (polemikus), *aszkrétikus (asketikus), *szférikus (szferikus)</i> stb.</p>
<p>29. A szó végi <i>a</i>, és <i>e</i> a megfelelő toldalékolás során válik hosszúvá.</p>	<p>A <i>-szerű</i> toldalék előtt is megnyúlik a magánhangzó minden esetben: <i>*kecskészerű (kecskeszerű), *allegroszerű (allegroszerű), *mangrovészerű (mangroveszerű)</i>.</p>
<p>30. Nem követ el hibát a kettős alakok megengedésében.</p>	<p>Nem létező kettős alakokat is megenged a <i>-nyi</i> toldalék esetében: <i>*borjnyi (borjúnyi); *tetvnyi (tetűnyi); *bokrnyi (bokornyai)</i> stb.</p> <p>Az előzőhöz hasonló probléma, hogy birtokjel is járulhat a nem szótári tőhöz: <i>*bokré (bokré, bokráé), *terhé (teheré, terhéé)</i> stb.</p>
<p>32. A <i>metsz, látszik, tetszik</i> toldalékolt alakjai is helyesek.</p>	<p>A <u>mesd</u>; <u>lássék</u>, <u>tessem</u>, <u>tessek</u> stb. alakokat hibásnak jelzi.</p>

Magyar Ispell	MS Office XP helyesírás-ellenőrzője
41. A műveltető képző kezelése helyes. A régies <i>-atík, -etik, -tatík, -tetík</i> szenvedő ige-képző kezelése csak a mai köznyelvben még megtalálható alakokra szorítkozik: <i>megadatík</i> , de <i>oltatík</i> .	Az <i>-at, -et</i> és a <i>-tat, -tet</i> szabadon járul a tő-höz: <i>*bonttat (bontat); *kértet (kéret); *nyit-tat (nyittat), *oktatatás<sup>6</sup> (oktattatás), *történ-tetnek, *történessél (történik+<i>-et</i>)</i> stb. Általánosan kezeli a szenvedő igeképzőt, de a műveltető képzőhöz hasonlóan hibásan: <i>oldta-tík</i> , de <i>*olttatík, oltatík</i> .
42. A <i>-val, -vel</i> kezelése jó.	Szintén, egy hibával: <i>*mohval (mohhal)</i> .
45. A toldalékolásbeli kettősséget jól kezeli, kivéve gyakorlati okból az elavuló formákat: <i>szeretniök, látniök</i> .	Hasonlóan.
46. Az elhomályosult képzésű szóelemek írá-sában nincs ismert hiba.	Nem ismeri a <i>lélegez</i> ige egyes alakváltozatait: <i>belégzem, kilégzel</i> stb. Több <i>-szt</i> végű igét sem kezel: <i>aggaszt, apaszt, áraszt, ébreszt, fagyaszt, függeszt, horgaszt, kepeszt, lehiggaszt, sorvaszt, szállaszt</i> .
A szóelemző írásmód	
Több tucat –, a beszélt nyelvi ingado-zások, és hasonulások miatt elkövetett, több karaktert is érintő – tipikus tévesztést javít: <i>elősször → először, szöllő → szőlő, hüttő → hűtő, dijjas → díjas, csug → csukd, baráccság → barátság, licensz → licenc</i> stb. <sup>7</sup>	A <i>ly/lj, ts/cs, ggy/dj, ggy/gyj, nny/nj</i> több karak-tert érintő tévesztéseket javítja csak.
62. Az azonos hosszú és rövid mássalhangzók találkozását jól kezeli.	A <i>-szerű</i> és a <i>-féle</i> képzőszerű utótag eseté-ben hibás alakokat is elfogad: <i>*dzsessz-szerű (dzsessz-szerű), *puffféle (puff-féle)</i> . A toldalékolásnál helyenként a nem egysze-rűsített, hibás alakokat is elfogadja: <i>*frissség, *fessség; *feddd; *izzzak, *izzzam...</i> <sup>8</sup> Egyéb példa még a <i>*veddd</i> . A <i>tovább</i> igekötő eseté-ben csak a hibás alakot fogadja el: <i>*tovább-botorkál (tovább-botorkál), *továbbburjánzik (tovább-burjánzik)</i> stb.

<sup>6</sup>Érdekes módon ez egy gyakori tévedés is egyben, aminek háttérében az írás során megismételt szótag áll. Az ilyen hibákat felismeri és javítja a Magyar Ispell.

<sup>7</sup>L. cseretáblázat leírását.

<sup>8</sup>... *\*izzzalak, \*izzz, \*izzzál, \*izzzad, \*izzzon, \*izzza, \*izzzunk, \*izzzuk, \*izzzatok, \*izzzátok, \*izzzanak, \*izzzák*.

Magyar Ispell	MS Office XP helyesírás-ellenőrzője
63. A <i>d</i> végű igék ragozása jó.	A <i>d</i> végű igék módjára, de hibásan ragozza az <i>r</i> , és <i>j</i> végű igéket is: * <i>írdd</i> ( <i>írd, írjad</i> ), * <i>várdd</i> ( <i>várd, várjad</i> ); * <i>fújdd</i> ( <i>fújd, fújjad</i> ), * <i>vájd</i> ( <i>vájd, vájjad</i> ).
64. Helyesen kezeli az <i>-l</i> végű igéket.	Nem ismeri a <i>múlik</i> ige kötőhangzót tartalmazó alakjait: <i>múlanak, múlanának</i> . Elfogadja viszont az <i>-at</i> , és <i>-tat</i> képzős alakjait: * <i>múlat</i> , * <i>múltass</i> ( <i>mulat, mulaszt</i> ). Nem ismeri a <i>telik</i> ige hosszú <i>l</i> -lel írt alakjait: <i>tellene, tellenek</i> stb.
65. Helyesen kezeli az egyalakú <i>-ll</i> végű igéket.	Elfogadja a következő alakot: * <i>álnak</i> ( <i>állnak</i> ) stb. <sup>9</sup>
67. A hosszú magánhangzóra végződő igék ragozásában közel megegyezik a Helyes-e?-vel. A <i>nőszik</i> igéből képzett <i>nőssz</i> alak a jóval gyakoribb <i>nősz</i> ellenőrzése miatt nincs elfogadva.	Hiányzó tő: <i>nyí, nyíjatok</i> . Hibás alak: * <i>hittak</i> ( <i>híttak</i> ). Nem fogadja el a régies <i>hí</i> , és <i>fú</i> egyes toldalékolt alakjait: ( <i>fúvok, húnak</i> stb).
69. A <i>jön</i> igét helyesen kezeli.	Hiányzó alak: <i>jőnek</i> .
70. Az <i>egy</i> számnevet helyesen kezeli.	Hibás alak: * <i>egyűvé</i> ( <i>együvé</i> ).
72. A <i>kis</i> fokozott alakjaiban a hosszan ejtett <i>s</i> -t írásban nem tükrözi.	Hasonlóan, de elfogad egy-két zavaró alakot hibás szóösszetétel-képzés révén: * <i>kisseb</i> ( <i>kisebbség</i> ), * <i>kissebség</i> ( <i>kisebbség</i> ).
72. Az <i>-s</i> végű melléknevekben a hosszan ejtett <i>s</i> -t helyesen kezeli.	Hasonlóan, de elfogad egy zavaró alakot hibás szóösszetétel-képzés révén: * <i>erősseb</i> ( <i>erősebb</i> ).
Helyes javaslatot tesz a <i>belölle, elölle</i> hibás alakokra: <i>belőle, előle</i> .	Nem tesz javaslatot
74. Az <i>új</i> és <i>ujj</i> meg van különböztetve.	Az <i>ujj</i> <i>-ság</i> képzőt is kaphat: * <i>ujjság</i> ( <i>újság</i> ). Az <i>új</i> a főnevekhez hasonlóan bármely szóösszetételben szerepelhet: * <i>újperc</i> ( <i>ujjperc</i> ).
75. A <i>h</i> végű szavak írása pontos.	Szintén, kivéve * <i>mohval</i> ( <i>mohhal</i> ).
78. A rövid magánhangzó + <i>t</i> végű igék, illetve a <i>lát</i> és <i>bocsát</i> igék kezelése helyes.	A hibás * <i>lássd</i> ( <i>lásd</i> ), * <i>bocsássd</i> ( <i>bocsásd</i> ), * <i>hallgassd</i> ( <i>hallgasd</i> ), * <i>kössd</i> ( <i>kösd</i> ), * <i>nyissd</i> ( <i>nyisd</i> ) stb. alakokat elfogadja.
79. Az <i>st</i> és <i>szt</i> végű igék kezelése helyes.	Az <i>st</i> végű igénél ( <i>fest</i> ) elfogadja a * <i>fessd</i> ( <i>fesd</i> ) alakot.

<sup>9</sup>Ennek hátterében a hibás \**ál* (*ál-*) tő felvétele *áll*.

Magyar Ispell	MS Office XP helyesírás-ellenőrzője
80–81. Az <i>s</i> , <i>sz</i> , <i>z</i> és <i>dz</i> végű igék ragozása helyes.	Az <i>s</i> , <i>sz</i> és <i>dz</i> végű igék kezelése hibás: 1. * <i>olvassd</i> ( <i>olvasd</i> ), * <i>késsd</i> ( <i>késd</i> ) stb.; 2. * <i>mászza</i> ( <i>mássza</i> ), * <i>játszák</i> ( <i>játsszák</i> ) stb.; 3. <i>tessem</i> , <i>tessél</i> , <i>lássék</i> , <i>mesd</i> . 4. <i>edzd</i> , <i>pedzd</i> .
82. A kivételes ragozású <i>sz</i> végű igék kezelése helyes.	A <i>tesz</i> , <i>visz</i> , <i>eszik</i> kezelése hibás: * <i>tesszél</i> ( <i>tegyél</i> ), * <i>visszéték</i> ( <i>vigyétek</i> ), * <i>esszünk</i> ( <i>együnk</i> ) stb.
83. A <i>-val</i> , <i>-vel</i> toldalékolás jó.	Hasonlóan, a * <i>mohval</i> , és a betűk toldalékolása (l. 3.) kivételével.
A hagyományos írásmód	
89. Az <i>ly</i> írásában pontos.	Csak a hibás alakot fogadja a <i>csevej</i> esetében: * <i>csevely</i> . Számos <i>ly</i> -t tartalmazó szó esetében a hibás alakot is elfogadja: * <i>estéji</i> , * <i>karvaj</i> , * <i>szeméji</i> , * <i>ünnepéjes</i> ; * <i>bolyár</i> , * <i>bolytár</i> , * <i>súlytó</i> stb.
A keretrendszer lehetőséget ad a zárt <i>ë</i> hangokat jelölő helyesírási szótár létrehozására a magyar nyelv zárt <i>ë</i> szótára alapján.	
91. Az egyszerűsítő írásmód	
94. Jól kezeli az egyszerűsítő írásmódot. Az ezekben elkövetett hibákra javaslatot tesz (például Gaussal→Gauss-szal.)	Akadnak típushibák, például <i>vadásszerű</i> , * <i>vadászszerű</i> , vagy <i>puff-féle</i> és * <i>puffféle</i> . A helyes kötőjeles alakot nem javasolja.
A különírás és az egybeírás	
95. Fokozott hangsúly az egybe- és különírási szabályok betartásán. A helyesírási szótár szintjén kezel nagyon sok szóalakot. A kötőjeles szavak írásánál viszont még nagyon megengedő.	Alapvető hiányosságok (például nagyon sok esetben melléknevek és főnevek egybeírásának engedélyezése). A kötőjeles szavak írásánál is nagyon megengedő.
96–99. Különírandó szóismétlések írásánál feltünteti a helyes kötőjeles alakot is.	Nem javasolja a kötőjeles alakot.
106. Az alanyos kapcsolatok esetében a jelentésváltozással bírókat engedi egybeírni.	Minden alanyos kapcsolatot elfogad (* <i>macskaásított</i> , * <i>egérfutott</i> ).
107–108. A minőségjelzős kapcsolatok esetében a jelentésváltozással bírókat engedi csak egybeírni.	Sok minőségjelző esetében minden kapcsolatot elfogad (* <i>nagyépület</i> , * <i>butakutya</i> , * <i>szépvérs</i> ).

Magyar Ispell	MS Office XP helyesírás-ellenőrzője
111. Az önálló alakban nem szereplő minőségjelzős alakokat nem fogadja el, csak külön felvett összetett szóban, vagy kötőjellel ( <i>al-</i> , <i>ál-</i> , <i>bel-</i> , <i>fesz-</i> ).	Több ilyet elfogad: <i>*ál</i> , <i>*bel</i> , <i>*fesz</i> , <i>*pót</i> .
112–114. Az <i>-ó</i> , <i>-ő</i> képzős minőségjelzős összetett alakokat, ha a szó valamely tagja összetett, szótári tiltások alapján a leggyakoribb esetekben nem engedi egybeírni. Az egybeírt alakokra helyes javaslatot tesz. A lista nem lehet teljes, de a leggyakoribb hibás egybeírásokat tartalmazni fogja a Szószablya fejlesztésnek köszönhetően.	Nagyon sok különírandó alakot helyesen kezel, de így is sok gyakorit enged egybeírni: <i>*ellenőrzőbizottság</i> , <i>*előadó körút</i> .
115. Az anyagnevek felismerése csak a gyakori esetekre van megoldva.	Hasonlóan.
116. A hónapok, illetve napok nevét nem engedi egybeírni.	A hónapok, illetve napok nevét engedi egybeírni. <i>*januárhónapban</i> , <i>*hétfőreggel</i> .



# Irodalomjegyzék

- [1] BAKOS Ferenc, szerk.: Idegen szavak és kifejezések szótára. Akadémiai Kiadó, Budapest, 1984.
- [2] DEME László–FÁBIÁN Pál–TÓTH Etelka: Magyar helyesírási szótár. Akadémiai Kiadó, Budapest, 1999.
- [3] JAKAB László: Tanulmányok az igeragozás köréből. KLTE Nytud. Int. 73. Debrecen, 1999.
- [4] KIEFER Ferenc, szerk.: Strukturális magyar nyelvtan 2. Fonológia. Akadémiai Kiadó, Budapest, 1994.
- [5] KIEFER Ferenc, szerk.: Strukturális magyar nyelvtan 3. Morfológia. Akadémiai Kiadó, Budapest, 2000.
- [6] MAGYAR TUDOMÁNYOS AKADÉMIA, szerk.: A magyar helyesírás szabályai. 11. kiadás. Budapest, 1984, 1994<sup>11</sup>.
- [7] NÉMETH Anikó: A magyar nyelvtan. Merényi Könyvkiadó, Budapest, 1997.
- [8] PAPP Ferenc, szerk.: A magyar nyelv szóvégmutato szótára. Akadémiai Kiadó, Budapest, 1969.
- [9] TÖBBEK. ispell manpages (man ispell, man 4 ispell).