

**Kovács Réka**

**Helyesiras.mta.hu**

**Helyesírási tanácsadó portál**

**(Váradi Tamás, MTA Nyelvtudományi Intézet Nyelvtechnológiai és Alkalmazott Nyelvészeti Osztály)**

MTA NYELVTUDOMÁNYI INTÉZET. BUDAPEST. 2013.

### **helyesiras.mta.hu, az intelligens helyesírási tanácsadó portál**

A Magyar Tudományos Akadémia Nyelvtudományi Intézete 2009 óta dolgozik a helyesiras.mta.hu portál elméleti hátterének a kidolgozásán és a megvalósításon. Az eredményt, vagyis a portált 2013. április 30-án az MTA Székházában rendezett megnyitón tárták a nagyközönség elé. A fogadtatás vegyes, de többnyire pozitív és bizakodó volt. Abban viszont mindenki egyhangúlag egyetértett, hogy hiánypótló szolgáltatást hoztak létre a fejlesztők. Váradi Tamás, a Nyelvtudományi Intézet Nyelvtechnológiai és Alkalmazott Nyelvészeti Osztályának vezetője a portál előnyeinek kívül arra is felhívta a figyelmet, hogy a magyar helyesírás kezelése nyelvtechnológiai eszközökkel nem egyszerű feladat, és a rendszer folyamatos fejlesztés alatt áll (jelenlegi verzió: 1.0.7 [2013-07-02]). A fejlesztők név szerint: Váradi Tamás, Miháلتz Márton, Ludányi Zsófia, Takács Dávid, Oravecz Csaba, Hussami Péter, Nagy Ágoston, Mittelholcz Iván, Pintér Tibor, Sass Bálint, Mártonfi Attila, Kardos Tamás, Selyem Zoltán (1).

#### **A portálról általában**

A helyesiras.mta.hu egy olyan nyelvtechnológiai eszközökre támaszkodó tanácsadó portál, amely az érvényben lévő akadémiai szabályzat, *A magyar helyesírás szabályai* 11. kiadása alapján szolgál helyesírási javaslatokkal azoknak, akik helyesen szeretnének írni (2).

A portál egyszerűsége és közérthetősége törekszik, és a nyelvészekről a tanárokon át a laikusokig mindenkinek készült. Dizájnya és struktúrája is ezt hivatott reprezentálni. A közérthetőségen még szükséges javítani, de – a fejlesztők reményei szerint – rövidesen egy olyan oldalt sikerül létrehozniuk, ahol a hét részterületet lefedve a számítógép lehetőségeihez mérten érthető és alkalmazható választ kap a felhasználó. A portál a javaslatok után magyarázattal és a szabálypontokra való hivatkozással is szolgál. A grafikat a Farm Stúdió tervezte (3). A weblapra lépve a következő nyitóképp látható:



1. ábra  
Nyitólap (4)

Az eszközökön kívül a *Helyesírási szabályzat*, az *Archívum*, a *Magunkról* és a *Kapcsolat* menüpont található. A *Helyesírási szabályzat* pont alatt az AkH. tizenegyedik kiadása, tizenkettedik (példaanyagában átdolgozott) lenyomata van feldolgozva utalóként szolgáló ugrópontokkal. Azoknak, akik nem szívesen keressélnék a papíralapú szabályzatban, ez a megoldás tökéletes, hiszen amellett, hogy a weblapon lehetőség van arra, hogy szavakra keressünk, a fejlesztők a portál bal oldalán pontokba szedték az AkH. főbb fejezeteit, amelyek szintén ugrópontként szolgálnak.

Az *Archívum* menüpont alatt találja meg a látogató azokat a szavakat és kifejezéseket – különböző kategóriák és szempontok szerinti keresési lehetőséggel –, amelyeket a Nyelvtudományi Intézet nyelvi tanácsadó szolgálata az évek során kérdésként telefonon, illetve levélben kapott, és meg is válaszolt. Bár ennek az anyagnak a feldolgozottsága még nem teljes, a nyelvhasználók példáinak realitása és a kategóriák létrehozása miatt ez az adatbázis nagyon hasznos része az oldalnak. Kereshet az érdeklődő nyelvtani kategóriák szerint, például: betűk, intézménynevek, elválasztás; szakmák szerint, ilyen a föld(rajz), a pol(itika), a kém(ia) stb.; valamint egyéb jelzések alapján: mozg 'mozgószabály', össz 'összetétel', őő 'folyamatos melléknévi igenévvel alkotott összetétel', emellett pedig a szabálypontokra mutató hiperlinkek is segítik a felhasználót. Az *Archívum* frissítésekor a fejlesztők azt tervezik, hogy bizonyos eseteknél megjegyzéssel is ellátják a kérdéses szavakat.

A *Magunkról* pontnál alapos leírást lehet olvasni arról, hogy kik készítették, és hogyan működik a portál; hogyan használjuk, mit nyújt a rendszer, mit nem tud a rendszer, és mik a portál előnyei. Emellett fellelhető az eszközök használatának a részletes leírása, a felhasznált alkalmazások, fogalommagyarázat és a közreműködők publikációi is (5). Bár a fejlesztők törekedtek arra, hogy minél több információval lássák el a felhasználót, és előre választ adjanak a felmerülő kérdéseire, a *Hogyan működik* pontot kevesen veszik figyelembe (április 30-tól június 27-ig az 590 376 oldallátogatásból 1944-szer nézték meg a *Hogyan működik* oldalt – a Google Analytics statisztikái alapján [6]).

A *Kapcsolat* menüpont három különböző alternatívával szolgál, amelyeken keresztül a látogató felteheti a kérdéseit, jelezheti a megjegyzéseit: a Nyelvművelő és Nyelvi Tanácsadó Kutatócsoport e-mail elérhetősége, a visszajelző űrlap és a hivatalos Facebook-oldal található itt.

## Az eszközök

A választható eszközök vagy modulok a *Külön vagy egybe?*, a *Helyes-e így?*, a *Névkereső*, az *Elválasztás*, a *Számok*, a *Dátumok* és az *Ábécébe rendezés*. Az eszközökre irányítva a kurzort rövid leírást olvashatunk arról, hogy milyen problémákkal kapcsolatban használható az adott modul, a *Számok*nál például a „Számjegyekkel megadott tőszámnevek, sorszámnevek, tizedes törtek, törtek átalakítása betűvel írt alakba” használati útmutató található.

A javaslatok megadása zöld háttérben kis pipával történik, a helytelen alakoké pedig piros háttérben felkiáltójellel. Főként a *Külön vagy egybe?* modulban előfordulhat, hogy több megoldással, javaslattal áll elő a portál, ezek alatt szerepelnek a magyarázatok is. Ezek a szövegek néha nem egészen érthetők, és a javaslatok sorrendjének sincs jelentősége, ez ugyancsak zavaró lehet. A fejlesztők törekednek arra, hogy a többértelműséget teljes mértékben megszüntessék, és a felhasználók egyértelműen választ kapjanak a kérdéseikre.

## Külön vagy egybe?

Sorrendben az első a *Külön vagy egybe?* modul. „A magyar helyesírás egyik legnehezebb kérdésköre” (Laczkó–Mártonfi 2004: 91) a szavak külön- és egybeírása, így nem véletlen, hogy általában ezzel a területtel van a legtöbb baj. A *magyar helyesírás szabályai* meglehetősen nagy részét a külön- és az egybeírás, valamint a kötőjelek alkalmazására vonatkozó szabálypontok teszik ki (AkH. 95–142.). A mozgó- és egyéb összetételekkel kapcsolatos szabályok használata sokszor nem egyértelmű. Itt azokra az esetekre lehet gondolni, ahol a jelentésárnyalatok a perdöntők (*bontott-tégla fal – bontott téglafal*) (Laczkó–Mártonfi 2004; Pintér et al. 2009; Ludányi et al. 2013a), vagy éppen az idegen elemekkel alkotott összetételekre (*franchise-rendszer*). Emellett számba kell vennünk az emberi tényezőt is, sokan nem ismerik részletesen a külön- és az egybeírás szabályait, például az *-ól/-ő* képzős melléknévi igenevekkel alkotott összetételekre vagy a jelentéssűrítő összetételekre vonatkozó szabályokat, és nem szívesen fordulnak segítségért a helyesírási szótárakhoz, különösen a papíralapú segédeszközökhöz. Míg a szótagok és a tagok felismerése, szótövesítése, a morfémák azonosítása a mostani nyelvtechnológiai segédeszközöknek nem okoz gondot, az idegen szavakkal alkotott összetételekből (*outsourcingpiac*) és a többletjelentésekből (*mozgássérült-segítő kutya – mozgássérült segítőkutya*) fakadó problémákat nehezen kezeli a rendszer. Az alkalmazott morfológiai elemzők a *HUMor* és a *Hunspell*, egyéb alkalmazások a *Helyes-e?* helyesírás-ellenőrző és -ajánló, valamint a *Magyar elválasztási szótár* a *PyHyphen* szabad forrású alkalmazáshoz (7). Amikor a fejlesztők úgy döntöttek, hogy a többi helyesírási tanácsadó portáltól eltérően az egyszerű beépített adatbázis és a szótári keresés helyett formálisan megfogalmazva automatikusan, nyelvtechnológiai módszerekkel készítik el az oldalt, számolniuk kellett ehhez hasonló problémákkal. Ezeknek a megoldására több módszer létezik, ám sajnos tökéletes sohasem lesz. Természetesen a problémás esetek nagyobb részére maguk a felhasználók világítanak rá, a portál *Visszajelzés* szolgáltatása révén. Ha a kérdéses szó vagy kifejezés speciális, egyedi vagy nem fogalmazható meg formálisan, azaz nem írható rá szabály, de ennek ellenére fontosnak tartható, a rendszer kivételista-adatbázisába kerül. Amennyiben általánosabb dologra derül fény, például az állatnevek *vízi-*, *fekete-*, *fehér-*, *házi-* stb. nemi, nemzetségi vagy faji jelzős fajnevek külön- és egybeírásai szabályaira (*házilégy – házi egér*), akkor a fejlesztők vagy szabály alapján oldják meg az újonnan felmerült problémát, vagy végső esetben valamilyen korpuszból történik a bővítés.

A *Külön vagy egybe?* modul alkalmazásához az oldalon instrukciókat is kap a felhasználó. Az egyik ilyen fontos utasítás, hogy ha egy kifejezés vagy szóösszetétel külön- és egybeírásával, valamint a kötőjelhasználattal kapcsolatos kérdésünk van, akkor ezt az eszközt válasszuk. Sajnos, a *Helyes-e így?* nagyon csábítóan hangzik, bármilyen típusú kételyünk is akad, itt azonban külön- és egybeírással kapcsolatos problémánkra

nem feltétlenül kapunk releváns választ. A két leginkább használt alkalmazás a *Helyes-e így?* (177 932) és a *Külön vagy egybe?* (173 075) (április 30-tól június 27-ig az 590 376 oldallátogatásból – a Google Analytics statisztikai alapján). A másik instrukció a tagok bevitelére vonatkozik. A kérdéses alakokat tagonként, azaz külön-külön kell beírni. Ha nem talál egyezést az adatbázissal, akkor a rendszer szabály alapján adja meg a megoldást, ám ekkor maximum három tagot képes inputként kezelni.

### Helyes-e így?

A *Helyes-e így?* eszközmegnevezés jogosan jelentheti a felhasználónak azt, hogy itt minden kétes alakkal kapcsolatban választ kap. Sajnos ez nem feltétlenül történik így. Az oldal fejlesztőinek hosszabb távú tervei között szerepel, hogy összefűzzék a modulokat, és ha nem is valamennyit – hiszen akkor nem volna értelme a külön eszközöknek –, de a leggyakrabban előforduló szavakat (például a tulajdonneveket a *Helyes-e így?*-ben) máshol is meg lehessen találni. „Szóalak helyességének vizsgálata, helytelen alakhoz helyes alakok javaslata” – olvasható a menüpontra kattintva. Ám itt éppen a javaslatok köre az, ami kifogásolható. Amikor ugyanis a beviteli alak helytelen, a rendszer hasonlósági mintaillesztéssel, automatikusan ad ki hozzá javaslatokat, amelyekből aztán olyan alakok is előkerülhetnek, amelyek grammatikailag helyesek, ám értelmük nincsen (*négykézlábazó* – *légykézlábazó*). Ez még abban az esetben is zavaró lehet, ha megtaláltuk a helyes megoldást a javaslatok között. A fejlesztők tisztában vannak a jelenséggel, megoldására pedig több lehetőség is felmerült. Az egyik az, hogy a megjelenő javasolt alakok mennyiségét csökkentik, hogy így szűrjék ki az értelmetlen megoldásokat. A másik az, hogy a javaslatokat az MNSz2 korpuszból adják. Az utóbbinak azonban az a hátránya, hogy nem tartalmaz minden létező alakot. Mindezek mellett – új szolgáltatásként – felmerült, hogy a Nyelvtudományi Intézet birtokában lévő archívum ellenőrzött szavai, kifejezései közötti hasonlóság alapján (az alaki hasonlóságon túl) is adjon találatokat a rendszer, amelyet aztán a felhasználó analógiával alkalmazhat a későbbiekben. Ezek azonban egyelőre a közép- és a hosszú távú fejlesztések között szerepelnek.

### Névkereső

A *Névkereső* a tulajdonnevek és ezek toldalékolásának helyesírásában segít. Annak ellenére, hogy a márkanévek, az intézménynevek, az állatnevek, a címek, a csillagnevek, a kiténtetés- és a díjnevek is tulajdonnévnek számítanak (AkH. 154–200.), a portál egyelőre csak a földrajzi neveket és a személyneveket ismeri. Emellett hiányosságként elmondható, hogy az olyan esetekben, amikor a felhasználó nem biztos benne, hogy egy bizonyos alakot kis- vagy nagybetűvel kell írni, ám az nem számít tulajdonnévnek, a rendszer ugyancsak nem ad választ. Erre példaként szolgálnak az egyezmények, a fontosabb napok vagy ünnepek (*karácsony*, *tordai országgyűlés*, *anyák napja*). A tulajdonneveknél begépelve a szót egy legördülő listában kapunk folyamatosan szűkülő találatokat, a felsorolás a megoldásokon túl a névtani kategóriákat is megadja (földrajzi név, településnév, személynév, vezetéknev, keresztnév, természetföldrajzi név, családnév, férfinév, női név, becenév). A *Névkereső* modul folyamatosan bővül mind az adatmennyiséget, mind a kategóriák pontosítását tekintve. Emellett rendelkezésre áll az *Archívum* listája is.

### Elválasztás

Az elválasztó eszköz a huhyphn és a HUMor morfológiai elemző alapján működik (8). Amellett, hogy a rendszer a szótagokat jelöli, a szóösszetételek határát is jelzi virgulával. Az elválasztás viszonylag könnyen automatizálható területe a helyesírásnak, ám éppen a formalizálás miatt találkozni olyan jellegű hibákkal, mint a *Püthagorasz* – *Püt-ha-go-rasz* javaslat. Hiányossága – ahogy erre a fejlesztők is utalnak –, hogy a tulajdonnevek esetében csak a leggyakoribb régies írásmódú családnéveket képes elválasztani.

## Számok

A számok leírása, ahogy az elválasztás, a dátumok és az ábécébe rendezés, egyszerűbben automatizálható (Ludányi et al. 2013b). A rendszer arra is lehetőséget ad, hogy törteket, tizedesvesszővel ellátott számokat vagy előjeleket használjunk. Erre vonatkozó instrukciót az eszköz neve alatt találunk. Ha több válasz lehetséges, a magyarázatban meglehetősen pontos leírást kap a látogató az alakok használatáról az alábbiak szerint:

Szám:

? 2/3 Mehet

'2/3' betűkkel írva:

- ✓ két harmad [291] (ha a tört nem egy főnév jelzőjeként használatos, pl. „két harmad nagyobb, mint egy harmad”)
- ✓ kétharmad [291] (ha a tört egy főnév jelzője, pl. „kétharmad csésze liszt, háromnegyed óra”)
- ✓ kettő harmad [291] (ha a tört nem egy főnév jelzőjeként használatos, pl. „két harmad nagyobb, mint egy harmad”)
- ✓ kettőharmad [291] (ha a tört egy főnév jelzője, pl. „kétharmad csésze liszt, háromnegyed óra”; Nem része a sztenderd nyelvváltozatnak.)

2. ábra  
A Számok eszköz használata (9)

## Dátumok

Ahogy azt a *Hogyan működik?*-ben írják a fejlesztők „a *Dátumok* elnevezésű alkalmazás a felhasználó által *évszám-hónap-nap* (éééé-hh-nn, azaz például 2013-04-15) formában beírt vagy egy sablonból kiválasztott dátumot a magyar helyesírás szabályai szerinti lehetséges alakokra alakítja át”. (10) A lista a lehetséges alakokkal gyakorlatilag valóban lefed mindent kezdve a római számoktól, a *tól-ig* formákon át a birtokos jelzővel ellátott hónapok helyesírásáig. A rendszer nem tudja kezelni a szökőévet és a (Gergely-naptár bevezetésekor) kimaradt napokat, de ennek nem igazán érezni a hiányát.

## Ábécébe rendezés

Ez a modul szintén egy nagyon jól működő részét képezi az oldalnak. Tekintve, hogy az oldal a magyar helyesírást támogatja, más ábécék szerinti rendezésre nem képes (11). Bemenetként a szavakat külön sorba kell írni, amelyre a kitöltési mező fölötti rövid instrukció hívja fel a figyelmet.

Az eszközök között található még egy *További információk a helyesírási szabályzatban* és egy *További információk az archívumban* menüpont, amelyek alatt a kérdéses nyelvtani kategóriára vonatkozó szabálypontok és az archívum szintén oda tartozó kategóriáira mutató hiperlinkek szerepelnek. Ezek az oldalak közötti könnyű átjárhatóságot biztosítják, és a magyarázatok körét bővítik.

## Összegzés

A helyesírási portál remek segédeszköze lehet a magyartanároknak, a magyar mint idegen nyelv tanárainak, az általános és a középiskolásoknak, valamint a magyart mint idegen nyelvet tanulóknak egyaránt. Mivel a felhasználók nem csupán a helyes alakot tanulják meg, hanem rögtön a hozzá tartozó magyarázatot is láthatják, a portál – szemben a papíralapú helyesírási szabályzattal – sokkal inkább lekötheti a figyelmet, és lehetőséget adhat szórakoztató, interaktív, egyéni vagy közös helyesírás-tanulásra is.

## Irodalom

- AkH. = *A magyar helyesírás szabályai*. 11. kiadás. 12. (példaanyagában átdolgozott) lenyomat. Akadémiai Kiadó. Budapest.
- Laczkó Krisztina – Mártonfi Attila 2004. *Helyesírás*. Osiris Kiadó. Budapest.
- Pintér Tibor – Mártonfi Attila – Oravecz Csaba 2009. Online helyesírási szótár és megvalósítási nehézségei. In: Tanács Attila – Szauter Dóra – Vincze Veronika (szerk.) *MSZNY 2009. Magyar Számítógépes Nyelvészeti Konferencia*. JATEPress. Szeged. 172–182.
- Ludányi Zsófia – Miháltz Márton – Hussami Péter 2013a. *Különírás-egybeírás – automatikusan*. VI. Alkalmazott Nyelvészeti Doktoranduszkonferencia. Megjelenés alatt.
- Ludányi Zsófia – Miháltz Márton – Pintér Tibor – Takács Dávid 2013b. *helyesiras.mta.hu – Az intelligens helyesíróportál*. XXIII. MANYE Kongresszus konferenciakötete. Megjelenés alatt.

- (1) MTA Nyelvtudományi Intézet. helyesiras.mta.hu portál. <http://helyesiras.mta.hu/helyesiras/default/aboutus> (2013. június 28.)
- (2) MTA Nyelvtudományi Intézet. helyesiras.mta.hu portál. <http://helyesiras.mta.hu/helyesiras/default/akh> (2013. június 28.)
- (3) Farm Stúdió. <http://farm.co.hu/> (2013. június 28.)
- (4) MTA Nyelvtudományi Intézet. helyesiras.mta.hu portál. <http://helyesiras.mta.hu/helyesiras/default/index> (2013. június 28.)
- (5) MTA Nyelvtudományi Intézet. helyesiras.mta.hu portál. <http://helyesiras.mta.hu/helyesiras/default/howitworks> (2013. június 28.)
- (6) Google Analytics. <http://www.google.com/analytics/> (2013. június 28.)
- (7) MTA Nyelvtudományi Intézet. helyesiras.mta.hu portál. <http://helyesiras.mta.hu/helyesiras/default/howitworks#S11> (2013. június 28.)
- (8) MTA Nyelvtudományi Intézet. helyesiras.mta.hu portál. <http://helyesiras.mta.hu/helyesiras/default/howitworks#S4> (2013. június 28.)
- (9) MTA Nyelvtudományi Intézet. helyesiras.mta.hu portál. <http://www.helyesiras.mta.hu/helyesiras/default/numerals#> (2013. június 28.)
- (10) MTA Nyelvtudományi Intézet. helyesiras.mta.hu portál. <http://www.helyesiras.mta.hu/helyesiras/default/howitworks#S6> (2013. június 28.)
- (11) MTA Nyelvtudományi Intézet. helyesiras.mta.hu portál. <http://www.helyesiras.mta.hu/helyesiras/default/howitworks#S7> (2013. június 28.)

**Kovács, Réka: Helyesiras.mta.hu, the intelligent online spelling consultation portal**

### Az írás szerzőjéről

*Kovács Réka*

MTA, Budapest, Magyarország

[kovacs.reka\[kukac\]nytud.mta.hu](mailto:kovacs.reka[kukac]nytud.mta.hu)

<https://orcid.org/0009-0003-9109-6633>