

VÁRADI TAMÁS – LUDÁNYI ZSÓFIA – KOVÁCS RÉKA

# Géppel segített helyesírás

## A helyesírás.mta.hu portál készítéséről

### **Computer aided spelling advice. Designing the helyesírás.mta.hu site**

The present article describes a newly launched portal ([helyesírás.mta.hu](http://helyesiras.mta.hu)) that employs innovative language technology to provide spelling advice. The article discusses the changing social context in the age of digital communication where, thanks to the Internet and social media, masses of people now have the means to express themselves publicly but lack the necessary literacy skills, most prominently spelling. The focus of the article is on the design principles of the portal, how it compares with the traditional way of acquiring spelling competence, what are the functionalities of the spelling advisory service and how it can be used.

### **Bevezetés**

Jelen cikkben bemutatjuk az MTA Nyelvtudományi Intézete által kifejlesztett helyesírási tanácsadó portált, amely nyelvtechnológiai módszerek alkalmazásával nyújt online tanácsadást idő- és helybeli kötöttségek nélkül. Az MTA Nyelvtudományi Intézet megalakulása óta foglalkozik nyelvhelyességi tanácsadással, amelynek zöme a helyesírás kérdéseire irányul. Az utóbbi években azzal a helyzettel szembesültünk, hogy a most már e-mailen és telefonon nyújtott szolgáltatásunk kapacitása nem volt képes megfelelően szolgálni az igényeket. A társadalmi elvárásokat feltehetően növelte, hogy az internet és különösen a társas média megjelenésével a nyilvános írásbeliség rendkívüli módon kiterjedt. A blogok, fórumok tömegeknek kínálnak nyilvános megszólalási lehetőséget, amihez a helyesírás ismerete elengedhetetlen. Szerencsére éppen az a technológia, amely az igényt kifejlesztette, képes megoldást is nyújtani. Ez a felismerés vezetett oda, hogy – amennyire csak lehetséges – a helyesírási tanácsadási tevékenységünket kiegészítsük automatikus online szolgáltatással. Így jött létre a [http://helyesírás.mta.hu](http://helyesiras.mta.hu) címen található portál.

E cikkben a portál tervezésével és működésével kapcsolatos elvi, nyelvészeti és nyelvtechnológiai kérdésekre helyezük a hangsúlyt, egyben ismertetjük a webhely szolgáltatásait és kezelését is. Először a helyesírással kapcsolatban felmerülő általános kérdésekkel foglalkozunk, különös tekintettel a nyelvészeti vonatkozásokra. Ezután következik a portál kialakításában alkalmazott nyelvtechnológia bemutatása, majd a webhely szerkezetét, az ott nyújtott szolgáltatásokat és használatukat írjuk le. Végül a fejlesztési terveinket ismertetjük.

## **A helyesírás céljáról**

A helyesírás alapvető célja az olvasás megkönnyítése, valamint az információ-visszakeresés elősegítése.

Ha egy szöveg olvasása során ugyanaz a lexéma többféle írásképpel fordul elő, az megnehezíti a szöveg befogadását, míg az egy szóalak – egy íráskép segíti az olvasást: a kölcsönösen egyértelmű megfeleltethetőség mindenképpen könnyíti a megértést (Mártonfi 2007: 135). Arról nem is beszélve, ha egy szövegen belül két (vagy több) párhuzamos alak fordul elő (tehát egy lexémához többféle íráskép tartozik), a szöveg olvasója hajlamos az eltérő alakoknak eltérő értelmezést tulajdonítani: azt gondolhatja tehát, ha egyszer így írjuk, egyszer úgy, annak bizonyára valamiféle oka van, mást jelenthet a két szó. Ez az érzés zavart kelt, elbizonytalanítja az olvasót, mivel szándékoltságot feltételez, holott erről nincs szó.

## **A helyesírás a digitális kommunikáció korában**

A helyesírás egészen a közelmúltig az írástudó elit, a művelt nagyközönség ügye volt. Az írásos nyilvánosság a társadalom szűk körére korlátozódott, akik valamilyen értelemben hivatásszerűen foglalkoztak írásos közlemények közreadásával. A digitális kommunikáció, leginkább az internet jóvoltából ez a helyzet mára teljesen átalakult, és továbbra is rohamosan változik. Különösen az ún. társas média megjelenése, a blogok, fórumok exponenciális növekedése azt eredményezte, hogy ma már olyan tömegek jutnak írásos nyilvánossághoz, akiknek azelőtt semmi lehetőségük nem volt erre, ugyanakkor nem gyakorlott íráshasználók, ennél fogva a helyesírással hadilábon állnak. A mobilkommunikációs eszközök, elsősorban az okostelefonok elterjedtével egyre többen kommunikálnak rendszeresen a digitális térben. Nem önmagában az írás kedvéért, hanem hogy tudakozódjanak valamilyen szakmai kérdésben, megosszák élményeiket, tapasztalataikat, szórakozzanak, kapcsolatot teremtsenek stb. Általános tapasztalat, hogy a helyesírást elvétő fórumozók igen könnyen szembesülnek durva helyreigazítással, társadalmi stigmával. Ez rögtön földre is viszi a beszélgetés fonalát: egyrészt érzelmileg rosszul érinti az egyik felet, másrészt senkinek sem hasznos. Ezért van tehát szükség a helyesírás széles körű „házhöz vitelére”.

## **A helyesírási portál társadalmi céljai**

A helyesírási tanácsadó portál annak a társadalmi igénynek igyekszik megfelelni, amelyet a rohamos mértékben terjedő digitális kommunikáció felvet. Célközönségünk nem korlátozódik a művelt nagyközönségre, elvileg minden olyan felhasználónak segítséget kívánunk nyújtani, akihez a digitális technológia segítségével eljutunk. Az okostelefonok ezt a kört olyannyira kibővítik, hogy néhány év távlatában már gyakorlatilag a teljes lakossággal számolhatunk. Nem csupán a technológia iránt fogékony, számítógép-használó emberekre kell gondolnunk tehát. Iskolázottság tekintetében sem fogalmazhatunk meg semmilyen küszöböt. Ha korlátra gondolunk, az inkább az igény felől jelentkezik. Az online elektronikus szolgáltatások, a társadalmi e-részvétel várható növekedése azonban egyre szélesebb rétegeket motivál majd a digitális kommunikációra, ahol a helyesírás alapkövetelmény.

Távlatilag olyan szolgáltatás kiépítése a célunk, amely a legszélesebb tömegek számára tenyérben hordozható segítséget nyújt, minimális tudás és iskolázottság feltételezésével. Mi az, amire építhetünk a felhasználók „tudását” illetően? Az alap természetesen a magyar anyanyelvi kompetenciájuk, vagyis az, hogy ugyan a nyelvészeti szakterminológiában nem jártasak, de határozott és helyes intuíciónak van nyelvi, főleg szemantikai különbségekről. A beszéd hangzó oldaláról pedig azt a változatot feltételezhetjük, amelyet kötetlen beszédhelyzetben használnak. Abból kell kiindulnunk, hogy a helyesírásban teljesen járatlan beszélő ösztönös késztetése, hogy az általa ejtett alakot próbálja írásban rögzíteni. Az eredményt akár egyfajta naiv fonematikus lejegyzésnek is nevezhetnénk. Ezt tekinthetjük a rendszer bemenő adatának, és a feladatunk az, hogy az így leírt alaknak megtaláljuk a helyesírási szabályzat szerinti helyes formáját. Ez tágabban értelmezett feladatot jelent, mint amit a helyesírási szabályzat maga elé kitűz. A szabályzat ugyanis eleve feltételezi a köznyelvi sztenderd kiejtést.

Az általunk előfeltételezett ejtismód, amelyet az inputon feltételezünk, több okból eltérhet a beszélt nyelvi sztenderdtől. A teljesen általános mássalhangzó-hasonulás következtében a köznyelvi beszélt sztenderdben is rengeteg olyan alak van, amelynek a kiejtése és írása nem egyezik. Mivel a hasonulás összetételekben és toldalékolt alakokban is produktívan előfordul, olyan sok szóalakot érint, amelyet szótárba foglalni is lehetetlen. További forrásai az eltérésnek a sztenderd beszélők kötetlen beszédében is gyakorta előforduló olyan lazítási jelenségek, mint például a *mér* alak használata a *miért* helyett, amelyeknek a szabály szerinti alakját tehát szintén meg kell adnunk. Ha ugyanis eleve feltételezzük a beszélt köznyelvi norma használatát, akkor már nincs is többé helyesírási probléma. További eltérést jelent az, hogy a beszélők igen nagy része nem a sztenderd kiejtést követi, hanem valamilyen dialektus szerint beszél.

## Helyesírás és a nyelvészet

A helyesírásnak mint diszciplínának a státusza igen sajátos. Mártonfi (2011: 68) joggal állapítja meg, hogy „a közvélekedés a nyelvtant és a helyesírást igen közeli, sőt átfedő területeknek tekinti, a mai tudományos konszenzus ellenében”. Számunkra elsősorban az a releváns kérdés, hogy a helyesírás mennyire támaszkodik nyelvészeti fogalmakra, még élesebben fogalmazva: mennyire a nyelvtani fogalmak, kategóriák, a köztük lévő különbségek határozzák meg az egyes kifejezések helyesírását?

Ha például azt akarjuk megtudni, hogy egybe kell-e írni a *félreállít* és a *végére jár* szavakat, akkor azt találjuk, hogy egybe- vagy különírásuk azon múlik, hogy a *félre* alak igekeető, míg a *végére* alak „raggal jelölt határozós kapcsolat” tagja [AkH. 125. a)]. Itt látszólag egyszerű kérdésről van szó, amely azonban intuíción alapján nehezen eldönthető, és végső soron a szófaji kategorizálás határozza meg a helyes alakot. Ez a szófaji besorolás az igekeetők esetében ráadásul nem mindig egyértelmű a nyelvészek számára sem, és a nyelvhasználat változásával idővel módosul. Vannak azonban a helyesírási szabályzatban ennél elvontabb nyelvtani kategóriák (például alanyos alárendelés, összetétel vs. szókapcsolat, jelentéstömörítő összetételek), amelyeknek pusztán ismeretét sem igen lehet feltételezni nyelvészeti előtanulmányok nélkül, nemhogy biztos alkalmazásukat. Gyakran a helyesírás szóalaktani vagy mondattani elemzésre is ha-

gyatkozik. Például sokszor nehézséget okoz a jelöletlen birtokos jelzős alárendelések, jelentéssűrítő összetételek elkülönítése a különírt főnévi jelzős szerkezetektől. A grammatikai szerkezet azonban gyakran szemantikai viszonyokat tükröz, így például a túl hosszú elemeket a fő tagok határán kötőjellel javasolja szétválasztani [AkH. 128. c), 138.], az azonban, hogy ez hova esik, feltételezi a szóösszetétel belső szerkezetének az ismeretét (például egyaránt értelmes a *vitamin-hiánybetegség* és a *vitaminhiány-betegség* többszörös összetétel). Hasonló a helyzet azokkal a szó szerkezetekkel, ahol a szerkezet alaptagját bővítő szavak, tipikusan jelzők, jelzős szerkezetek, hatóköre szabja meg, hogyan tagoljuk a kifejezést. (Például a *bontott + téglá + fal* szavak írását az határozza meg, hogy a *bontott* jelző a *téglára* vonatkozik-e, vagy a *téglafalra*.)

Megállapíthatjuk tehát, hogy a magyar helyesírási szabályzat kiterjedt módon támaszkodik a leíró nyelvészet fogalmi készletére, kategóriáira. Ez felveti a következő problémát. A helyesírás és az ezt szabályozó *A magyar helyesírás szabályai* mindenkinek szól. Ez esetben tehát nem célszerű magasabb szintű tudást feltételezni. Például a kiejtés szerinti írás elvének magyarázatában (AkH. 18.) a szerzők azt feltételezik, hogy az olvasó ismeri a *képző, jel* és *rag* terminusokat, tisztában van a fogalmukkal, annak ellenére, hogy ezeket külön nem definiálják. A szabályzat nyelvészeti kifejezések ismeretének feltételezésével készült. Bekezdésről bekezdésre találunk azonban olyan szakszavakat, amelyeket nem feltétlenül ért mindenki. Az alárendelő és mellérendelő szókapcsolatokról és összetételekről szóló szabálypontok használatához az olvasónak tisztában kell lennie a morfológia, a szintagmatan és a mondatan alapfogalmaival. Ilyen például a „mondatrészi viszonyú szókapcsolat” (AkH. 104.) kifejezés, amely szintén nem segíti a laikus felhasználót a szabálypont alkalmazásában, hiszen ismernie kell a *mondatrész* és a *mondatrészi szerep* terminusokat.

A felhasználók szempontjából a nyelvészeti terminusok használata különösen akkor jelent gondot, ha egy helyesírási dilemma megoldása olyan nyelvtani kategóriára vonatkozik, amely intuitíve nehezen megragadható. A nyelvészeti szakkifejezések alkalmazása helyett célszerűbb lenne közvetlenül azokra a különbségekre hivatkozni, amelyekre a nyelvészeti kategóriák igen absztrakt módon utalnak. Például a *fenn marad* és a *fennmarad* különbségét leírhatjuk úgy is, hogy a *fenn* alak az előbbiben határozószó, az utóbbiban igekötő. Ez azonban sok ember számára nem jelent érdemleges segítséget. Ugyanakkor sokkal nagyobb eséllyel hivatkozhatunk az anyanyelvi beszélők intuíciójára a jelentéskülönbség (‘valahol a magasban marad’ vs. ‘továbbra is tart’) megragadásában. Az érthetőség iránti igényt mutatják a portálon és e-mailben érkezett kérdések is, melyekben a felhasználók jelzik, hogy nem tudják értelmezni a szabálypontokat. Megjegyezzük, hogy annak ellenére tapasztaljuk ezt, hogy a helyesírás része a kerettantervnek.

## Hogyan lehet elsajátítani a helyesírást?

A helyesírási kérdésekben tájékozódni próbálók számára két hagyományos forrás áll rendelkezésre: az akadémiai helyesírási szabályzat és a helyesírási szótárak. Az előbbi elektronikus közzététele még kereshető és hipertextes formában sem igazán ad választ arra a kérdésre, melyik szabály vonatkozik az adott kifejezésre, és azt hogyan is kell érteni. Hathatósabb megoldást jelentenek a helyesírási szótárak. Ezek általában olyan szavak és szóalakok gyűjteményei, amelyek a helyesírási szabályok alól kivételt

jelentenek, vagy amelyek szabályosak ugyan, de a szerkesztők megítélése szerint a használok számára várhatóan gondot okoznak. A kivételes alakok tulajdonképpen a szabályzat mellékletének is tekinthetők, amelyek kiegészítik a szabályok által lefedett jelenségeket. A szabályok szerinti alakok szerepeltetése a szótárban első gondolatra egyfajta kényelmi szolgáltatásnak tűnik, hiszen látszólag nem nyújtanak új információt, csak példákat adnak arra, amit a szabályok megfogalmaznak. Valójában azonban ennél hasznosabb szerepet játszanak. A példák ugyanis segítenek a szabály megértésében és elsajátításában.

A szótárak a helyes alakok megadásával nyújtanak segítséget. Ha az olvasó megtalálja az alakot, máris láthatja, hogyan kell azt helyesen írni.<sup>1</sup> Ha azonban az olvasó nem találja a szótárban a kérdéses alakot, az számára nem jelent egyértelmű konklúziót. Gondolhatja azt is, hogy a keresett alak helytelen, de azt is, hogy a szótár hiányos. Márpedig elvileg is leszögezhetjük, hogy minden szótár óhatatlanul hiányos, már csak amiatt is, hogy a nyelvhasználat folyamatosan változik, így egyre újabb és újabb szavak, főleg kifejezések merülnek fel, amelyek helyesírásiilag szabályosak ugyan, de a szótár példatárában nem szerepelnek.

Azt látjuk tehát, hogy a nyelvhasználók a helyesírási szabályzat és a szótár kettős eszközére vannak utalva a helyesírás megtanulásában. A szótárak példaanyaga nemcsak a kérdéses kifejezések helyes alakjának megtalálásában, hanem az elvont szabályok elsajátításában is nagy haszonnal járnak. Ezenkívül vélhetően a szótár gazdag példatárára épülő analógiák segítségével próbálják megtalálni a helyes írott alakot.

## A helyesírás.mta.hu innovatív jegyei

Mielőtt részletesen rátérünk a portál kialakításában alkalmazott nyelvtechnológiai megoldásokra, röviden összefoglaljuk, miben jelent újdonságot a helyesírás.mta.hu portál. A legfontosabb annak a hangsúlyozása, hogy a portál *elvi*, és véleményünk szerint az egyetlen szakszerű megoldást kínálja a problémára. A kérdéses kifejezés helyes alakját nem listákkal (azaz szótárakkal) való összevetéssel próbálja megállapítani, hanem algoritmikus úton, a szóalak vagy kifejezés elemzésével, annak alapján a releváns helyesírási szabály megtalálásával és alkalmazásával. Nemcsak a helyes alakot állapítja meg, hanem a magyarázatot is megadja, ami egyelőre többnyire a vonatkozó helyesírási szabályzat idézését jelenti.

Az elemző algoritmus elkészítéséhez előzetesen szükség volt a helyesírási szabályzat megfelelő szabályainak az elemzésére is. Ebben az értelemben tehát mondhatjuk, hogy a gépi szabályrendszer leképezi a helyesírási szabályok elsajátításának azt a módját, amelyet az emberek esetében feltételeztünk. A helyesírási tanácsadó rendszerünk a szabályrendszerben megfogalmazott tudás, valamint a keresett kifejezés elemzésénél alkalmazott nyelvtechnológia segítségével nemcsak egy véges számú kifejezés helyes alakját tudja megállapítani, hanem képes a nyelvhasználat változását is követni és megállapítani a felmerülő új szavak, kifejezések helyes alakját.

<sup>1</sup> Megjegyezzük azonban, hogy a helyesírási szótárak nem a helytelen alakokból vezetik az olvasót a helyes alakokra, azaz azt nem találjuk bennük, hogy például „tuggya = tudja”. Ennélfogva már a helyes alak megtalálása sem teljesen triviális feladat.

## Hogyan lehet a szabályzat megértését gépesíteni?

Ahhoz, hogy egy interaktív helyesírási tanácsadó portál létrejöjjön, első lépésként részletesen meg kell vizsgálni a helyesírási szabályokat abból a szempontból, hogy lehetséges-e egyáltalán azok automatizálása, és ha igen, milyen szemantikai és szintaktikai jellemzőket kell ismernünk (Selyem 2009). A szabályokat számítógépes feldolgozhatóságuk alapján már korábban is tanulmányozták (például Kis 1999). Selyem (2009) alapján egy jól működő online tanácsadó portál elkészítéséhez az alábbi kategóriákat szükséges felismerni:

- szófajok (részletesen, például befejezett melléknévi igenév, sorszámnev);
- összetételség;
- összetételi tagnak számító idegen elő- és utótagok (*aero-*, *deka-*; *-centrikus*, *-fóbia* stb.);
- idegen szótság;
- családnevek;
- színnévi alaptagok (*piros*, *kék*, *zöld*, *antracit*, *okker* stb.);
- összetételi előtagok (*al-*, *bel-*, *köz-*, *kül-*, *szak-* stb.);
- folyamatos melléknévi igenévi jelzős szerkezetekben egyszerű szóként viselkedő összetételek (*műszer*, *rendszer*, *kórház*, *vendéglátó*, *előadó*);
- sportot jelentő folyamatos melléknévi igenevek (*labdarúgó*, *birkózó*, *súlyemelő*);
- főnévi jelzős szerkezetek tipikus elő- és utótagjai (*ajándék*, *férfi*, *barát*, *néni*, *úr* stb.);
- foglalkozásnevek (*színész*, *lakatos*, *katona*, *tüzér*, *tanár*);
- beosztások, fokozatok, rangok, tisztségek megjelölései (*százados*, *alhadnagy*, *bíboros*);
- egybeírandó, „fajtajelölő” főnévi jelzős szerkezetek (*diáklány*, *kölyökkutya*, *lakatosmester* stb.);
- anyagnevek (*acél*, *bőr*, *gyapjú*, *kő*, *műanyag*, *vas* stb.);
- ételnevek (vagy tipikus összetételi tagjaik) (*leves*, *pörkölt*, *ragu*, *filé*, *öntet* stb.);
- események, rendezvények, ünnepek, nevezetes napok neve (vagy az ilyen kifejezések tipikus szavai), (*napja*, *hete*, *ünnepe*, *játékok*, *kongresszus*, *szimpózium*, *konferencia*, *verseny* stb.);
- névelőzések (*doktor*, *idősb*, *professzor*, *vitéz* stb.);
- tulajdonnevek közzói jelentéssel nem rendelkező tagjai (*Huron*, *Kab*, *Kaszpi* stb.);
- embercsoportok gyűjtőnévi (jellegű) szavai (*család*, *kör*, *kórus*, *duó* stb.);
- földrajzi köznevek (*sziget*, *öböl*, *barlang*, *mellék*, *szántás*, *zsombék*, *zátony* stb.);
- utcanévi utótagok (*utca*, *körút*, *park*, *sor*, *sétány* stb.);
- földrajzi jellegű jelzők (*magas*, *kis*, *hosszú*, *felső*, *keleti* stb.).

Amit lényeges kiemelni: a portál megvalósításához mind morfológiai-szintaktikai adatokra (például szófaj, szótagszám, összetételi tagok száma), mind szemantikai információkra szükség van (például az adott szó egy foglalkozás vagy egy szín neve). Ezek az erőforrások a rendszerhez csatolva állnak rendelkezésre (morfológiai elemző program, illetve lexikális adatbázis formájában). A helyes alak azonban néha csak az adott kontextus ismeretében dönthető el, amely tudást csupán a felhasználóktól kaphatunk meg. Ezt a jelenleg még hiányzó „erőforrást” a jövőben úgy tervezzük pótolni,

hogy a felhasználót az alternatívák megvilágítására szolgáló példamondatok alkalmazásával bevonjuk a megoldás megtalálásába.

## **A helyesírás.mta.hu nyelvtechnológiájának és felhasználásának ismertetése**

Miután megvizsgáltuk, melyek a magyar helyesírásnak azon területei, amelyek gépesítése vállalható, eléggé hamar nyilvánvalóvá vált, hogy az online tanácsadó portál eredeti elképzelését – miszerint a felhasználók egyetlen beviteli mezőbe írhatták volna a kérdéses kifejezéseket (Pintér–Mártonfi–Oravecz 2009) – nem lehet megvalósítani. A nyelvtechnológiai eszközökkel támogatható helyesírási területek, szabályok más-más megvalósítást, eszközöket igényelnek (gondoljunk csak arra, hogy például a dátumok helyesírásához egészen más ismeretekre van szükség, mint a különírás-egybeírásnál). A végleges elképzelés szerint tehát a portál hétféle eszközt kínál: külön- és egybeírás, helyesírás-ajánló (a szó szintjén), elválasztás, tulajdonnevek írása, számnevek helyesírása, keltezés, betűrendbe sorolás, amelyek közül ki kell választania a felhasználónak, milyen típusú kérdésben szeretne automatikus választ kapni.

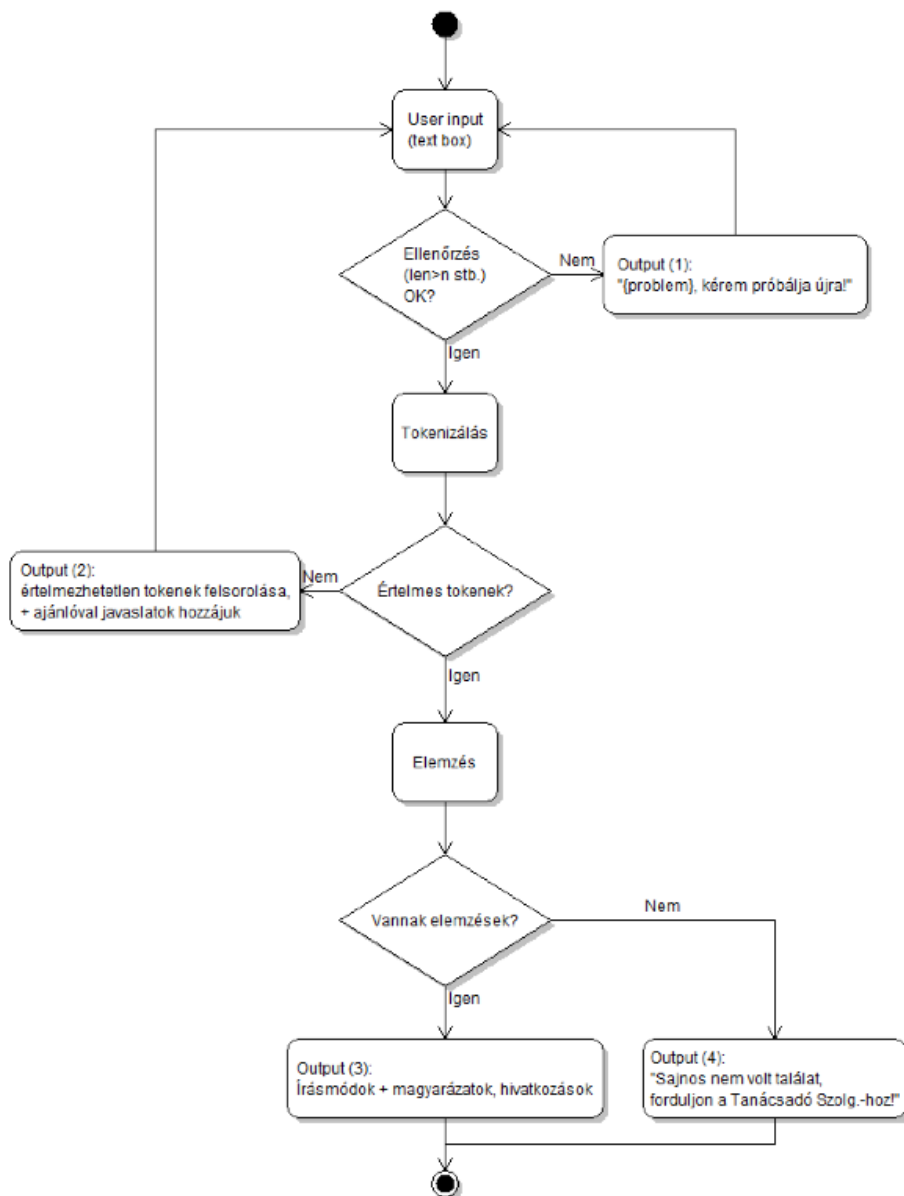
A választás ténye feltételez bizonyos ismereteket, és az interakció a felhasználó és a tanácsadó portál között már itt elkezdődik. Az eszközök (*Külön vagy egybe?*, *Helyes-e így?*, *Névkereső*, *Elválasztás*, *Számok*, *Dátumok*, *Ábécébe rendezés*), amelyek a fenti, majd továbbélve a bal oldali menüsorban is megjelennek, a kezdőoldalon ikonokkal középre rendezve serkentik a használót a döntésre. Az eszközökön kívül az oldalon megtalálható *A magyar helyesírás szabályai*, útmutató a használatához, blog, *Archívum* a régi kérdésekből és válaszokból, valamint lehetőség a visszajelzések elküldésére. A *Hogyan működik?* pont alatt részletesen olvashatunk az eszközök alkalmazásáról.

Az említett hétféle eszköz közül némelyik teljesen saját fejlesztés (*Dátumok*, *Számok*), míg mások már létező nyelvtechnológiai eszközöket fejlesztenek tovább, egészítenek ki (*Elválasztás*, *Helyes-e így?*), vagy használnak fel alapul (*Külön vagy egybe?*). Az alábbiakban az egyes eszközök működését, azoknak nyelvtechnológiai háttérét ismertetjük röviden (Miháltz et al. 2012 alapján). A háttértechnológia mellett röviden foglalkozunk az eszközök használatával, különös figyelmet fordítva az alkalmazás során eddig tapasztalt problémákra.

### **Külön vagy egybe?**

A különírás-egybeírás szabályainak elsajátításához szükséges, hogy különbséget tudjunk tenni a szószerkezet és a szóösszetétel között. Hogy egy kifejezés szerkezet vagy összetétel-e, nem mindig egyértelmű, olykor mind a kettő lehetséges (például *gyorsvonat* 'vonatfajta' – *gyors vonat* 'gyorsan közlekedő vonat'). Ennek a megítélésnek az automatizálása nem lehetséges, ilyenkor van szükség az emberi értelemre. Ez a terület tehát a jelenlegi, rendelkezésünkre álló technológiával nem automatizálható teljes körűen (Pintér–Mártonfi–Oravecz 2009). Bizonyos szabályok azonban leírhatók formális nyelven a szemantikai kategóriák (lásd fentebb) és a morfológiai elemző segítségével. Ennyiben a portál *Külön vagy egybe?* modulja több, mint az egyéb helyesírás-ellenőrzők, tudniillik nemcsak szótárt, hanem formális szabályokat (formális nyelvtant) is használ a tanácsadáshoz.

A rendszer általánosan az 1. ábrán látható módon épül fel.



**1. ábra. A rendszer felépítése (Ludányi–Miháltz–Hussami 2013: 119)**

A bemenet (*user input*) maximum 70 karakteres lehet. Ezt a rendszer először általánosságban ellenőrzi (nem tartalmaz-e nem megfelelő karaktereket stb.). Ezután következik a tokenizálás, amelynek során a kifejezést atomi szintű, azaz tovább már nem



bontható egységekre bontja. A morfémákat a HUMor morfológiai elemző látja el szófaji, morfológiai, szótagszámra és összetételei tagok számára vonatkozó információkkal (Ludányi–Miháltz–Hussami 2013: 118). Ezeken az információkon túl a szemantikai kategóriákat is hozzárendeljük. Körülbelül 90 kategória áll rendelkezésre, például a színnevek, foglalkozások és rangok, számnevek, keresztnevek, népek és nyelvek nevei, rövidítések, közsői betűszók, önálló szóként nem használatos előtagok, a helyesírási szabályzatban az egyes szabályokban hivatkozott további kategóriák és különösen az egyes kivételek listája (125 ezer [Miháltz–Ludányi 2013]). Miután a tokenek morfológiai és szemantikai információkkal történő ellátása megtörtént, megkezdődhet az elemzés. Az elemző modul bemenetét az előző részben említett szegmentált, morfológiai és szemantikai jegyekkel ellátott tokenek képezik.

A rendszer az elemzést formális nyelven megfogalmazott szabályok alapján végzi. A szabályok modellezhetők, tehát számos, azonos módon szabályozott, valós esetre felírható egy modell, amely azután alkalmazható a többi, ugyanezen a módon szabályozott valós esetre. Az eszköz alapja egy környezetfüggetlen, jegystruktúrás formális nyelvtan (a formális nyelvostályokról lásd bővebben például Prószték–Kis 1999: 114–118).

Az elemzések során a rendszer szintaktikai fákat állít elő, majd ezeket optimalizálja. Az optimalizálás a szabályt alkotó szükséges és elégséges feltételek megfeleltetésével történik. A szabálminták azt az információt tartalmazzák, hogy egy bizonyos esetben milyen feltétel mellett milyen írásmód érvényesül.

Természetesen ezek nem egy lépésben történnek. Vannak úgynevezett újíró szabályok. Összesen 372 formális szabály áll rendelkezésre, amelyek az alábbi területeket ölelik fel:

- jelölt és jelöletlen alárendelői összetételek/szintagmák;
- a szótagszámlálás szabálya (6:3-as szabály);
- mozgósabályok;
- rövidítéseket és mozaikszókat tartalmazó összetételek;
- néhány speciálisabb szabály, például a színnévi összetételek, anyagnévi összetételek (Miháltz et al. 2012: 141).

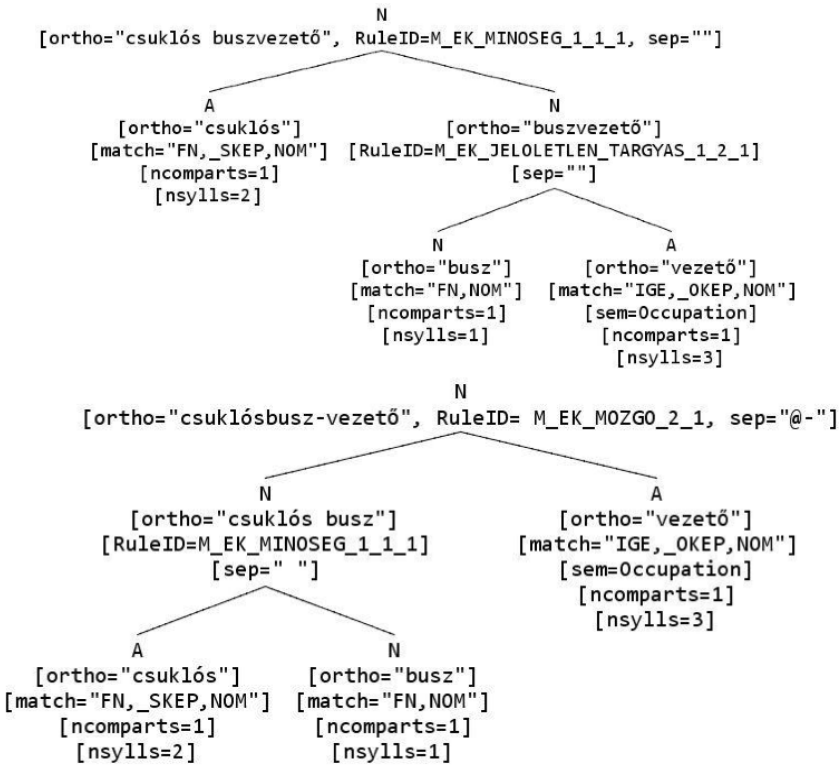
A rendszer a szervesetlen, jelentéssűrítő, mellérendelő és morfológiai típusú összetételeket korlátozott mértékben képes kezelni.

Lássuk egy példán, hogyan is elemez a program két, eltérő szabályból fakadó javaslatot.

A második mozgósabály:

Az AkH. 139. b) szabálya (az ún. második mozgósabály) a következőképpen szól: ha egy különírt szókapcsolat (például *hajlított bútor*) olyan utótagot kap (például *gyár*), amely az egészhez járul, az egyébként különírandó előrészt az új alakulatban egybeírjuk, és ehhez az utótagot kötőjellel kapcsoljuk: *hajlítottbútor-gyár*.

Vegyük a *csuklás + busz + vezető* bemenetet. Ennek kétféle elemzési fája a 2. ábrán látható módon ábrázolható.



**2. ábra. A csuklós + busz + vezető bemenet kétféle javaslatának elemzése (Miháltz–Ludányi 2013)**

Mi történik tehát az első, és mi a második esetben? Az első javaslat értelmezése: 'buszvezető, aki csuklani szokott'. Ennek értelmében a *busz* + *vezető* tokenek mint jelöletlen tárgyias szerkezet összetétellé válnak, amelyekhez azután a *csuklós* mint minőségjelző kapcsolódik. A második esetben érvényesül a mozgószabály. A jelentés itt: 'csuklós busznak (két részből álló buszfajta) a vezetője'. A minőségjelzős szerkezetért felelős szabály összefűzi az első két tagot (token) egy különálló szerkezetté: *csuklós busz*. A második mozgószabály bemenetei így a *csuklós busz* + *vezető* lesznek, a különírt *csuklós busz* szerkezet egészéhez kapcsolódik utótagként a *vezető*. Ezek után generálódnak aztán a javaslatok melletti magyarázatok.

A *Külön vagy egybe?* eszközt abban az esetben kell használnunk, ha nem tudjuk eldönteni egy kifejezésről, hogy az szerkezet, szóösszetétel-e, kell-e kötőjelet alkalmaznunk, és ha igen, akkor hol. Ez önmagában nem zárja ki, hogy a betűvel leírt számokkal és a tulajdonnévvel alkotott szerkezetekkel kapcsolatos kérdéseket is itt tegyük fel. Ezekre azonban két másik eszköz áll rendelkezésre: a *Számok* és a *Névke-reső*. *Helyes használat: Külön vagy egybe?* címmel posztot írtunk a *Helyes* blogba (1), ahol a beérkezett inputok és e-mailek alapján gyűjtöttük össze, milyen esetekben használják a felhasználók tévesen ezt az eszközt. Az első kettő, a már említett betűvel

leírt számok és a tulajdonnévvel alkotott szerkezetek. Ezenfelül a toldalékok, leginkább az idegen szavakhoz kapcsolódó végződések helyes írásmódja, amelynek szisztematikus kezelésére egyik eszköz sem alkalmas a számos, különböző nyelvekből átvett szavak kiejtése miatt (vö. AkH. 215–217).

## 2 javasolt alak: (a sorrendnek nincs jelentősége)



### I. "bontott téglafal"

Magyarázat

Részletek ▼

Két szabályt alkalmaztunk egymás után az alábbi sorrendben:

1. A "tégla" főnevet és a "fal" főnevet egybeírjuk az alábbi szabály alapján:

Az anyagnévi jelzőt ("tégla"), ha egyszerű szó, egybeírjuk a nem összetett főnevekkel. [AkH-115, OH-117]

2. A "bontott" melléknévet és a "téglafal" főnevet különírjuk az alábbi szabály alapján:

A befejezett melléknévi igenévi jelzős szerkezetek tipikusan (állandó vagy alkalmi) szókapcsolatok, tehát különírandók. [OH-114]



### II. "bontott-tégla fal"

1. lehetséges magyarázat

Részletek ▼

Három szabályt alkalmaztunk egymás után az alábbi sorrendben:

1. A "bontott" melléknévet és a "tégla" főnevet különírjuk az alábbi szabály alapján:

A befejezett melléknévi igenévi jelzős szerkezetek tipikusan (állandó vagy alkalmi) szókapcsolatok, tehát különírandók. [OH-114]

2. A "bontott tégla" főnévi szerkezetet és a "fal" főnevet különírjuk és az első szerkezetet egybeírjuk (összerántjuk) az alábbi szabály alapján:

Ha az anyagnévi jelző szerepét különírt szó szerkezet tölti be ("bontott tégla"), az eredetileg különírt szerkezetet egybeírjuk, az alkalmi összetétellé váltó jelző és a jelzett szó azonban különírandó. [OH-134]

3. A "bontotttégla fal" főnévi szerkezetet a szóösszetétellel határon összetörődött három azonos mássalhangzót kötőjellel elválasztva írjuk az alábbi szabály alapján:

Az olyan összetett szavakat, illetőleg tulajdonnévi szóalakokat, amelyekben a tagok, illetőleg a szóelemek határán három azonos, mássalhangzót jelölő betű kerül egymás mellé, kötőjellel tagoljuk. Ha mozgószabály miatt kerül egymás mellé 3 azonos msh., akkor is kötőjelet alkalmazunk. [AkH-262a, OH-27, OH-49, OH-87]

2. lehetséges magyarázat

Részletek ►

### 3. ábra. A bontott + tégla + fal bemenetre kapott javaslatok, magyarázatok

Ami a portál előnye, az egyszersmind a hátránya, hiszen azáltal, hogy nem korlátoztuk a javaslatokat azokra, amelyek a szabályzat szótári részében találhatóak, megnövekedett a potenciálisan jó válaszok száma. A javaslatok közül pedig a felhasználónak kell döntenie arról, melyik az az alak, amely az adott kontextusnak a leginkább megfelel. Ehhez a szabálypontokra való hivatkozások szolgálnak támpontokként.

A megoldások sorrendje nem mérvadó. A visszajelzésekből kiderült, hogy a felhasználók szükségét érzik a sorrendiség bevezetésének. Ennek oka a nyelvészeti kifejezések nagyszámú használata lehet. Tekintsünk meg egy bemenetet, amely példaként szolgál a kétféle, jelentésalapú értelmezésre, és a javaslatok lehetséges magyarázatára (3. ábra).

Az első javaslat tehát a *bontott téglafal*. Ennek részletes magyarázata további két szabályra van osztva, amelyek úgy generálódnak, hogy a rendszer felismeri a *téglát* mint anyagnévi jelzőt, valamint a *fal* egyszerű – tehát nem összetett – főnevet, és ezen információk tudatában adja hozzá a kapcsolódó szabálypontot.

A második szabály a *bontott* jelzővel ellátott szóalakra vonatkozó tudnivalókat adja meg. E javaslat értelmezéséhez ismerni kell az *anyagnévi jelző* és a *befejezett melléknévi igenévi jelzős szerkezet* terminusokat, hiszen a kontextusban való használat eldöntéséhez ezek a releváns információk. A második javaslatnál ez kiegészül az állandó vagy alkalmi összetétellel, a *szókapcsolatokkal* és a *jelzett szóval*.

A nehézkes megértés problémáját úgy orvosolnánk, hogy a magyarázatokat idővel érthetőbbé tesszük. Addig azonban a különböző interaktív platformokon segítjük a terminusok megértését.

A *Külön vagy egybe?* használatakor ügyelnünk kell arra, hogy a kérdéses szavakat szökökkel elválasztva írjuk a keresőmezőbe.

## Helyes-e így?

A *Helyes-e így?* modul azt vizsgálja, létezik-e a beírt szó (a modul egyszavas bemeneteket kezel). Ha nem talál egyezést, helytelennek ítéli az alakot, outputként javaslatokat tesz<sup>2</sup>. Az eszközt abban az esetben kell használni, ha egy szó kiejtése és az ortográfia közötti kapcsolat, különbség nem triviális (Miháltz–Ludányi 2013).

A használat során tapasztalt problémák közül az első a nagybetű használata vagy éppen a tulajdonnevekkel kapcsolatos kérdések. Bizonyos esetekben itt is megkapjuk a megoldást, azonban a *Névkereső* eszközünkben található a tulajdonnév-adatbázis, így azokra ott biztosabb válasz érkezik.

A másik tendencia a *Külön vagy egybe?* eszköz általi javaslatok kipróbálása a *Helyes-e így?*-ben a megfelelő alak kiválasztása céljából. Ez azért okoz problémát, mert a *Helyes-e így?* eszköz mögötti algoritmus más, mint a *Külön vagy egybe?* mögötti: nem foglalkozik a különírás-egybeírás szabályaival, csupán ellenőrzi a beírt szóalakot (ha a két tagot különírva adjuk meg, külön-külön ellenőrzi azokat). Ebből következik, hogy ha egy szókapcsolat az egybeírás-különírás szabályainak nem is felel meg, de az

<sup>2</sup> Az egyik használt motor a nyílt forráskódú Hunspell 1.3.2-es verziója (2), a másik a MorphoLogic HUMor elnevezésű eszköze (3). Helyesnek azokat fogadja el a modul, amelyeket a HUMor morfológiai elemző képes összerakni a tárolt morfémák, a hozzájuk tartozó információk és a nyelvi szabályok alapján. A javaslatok generálására eleinte a Hunspell-t és a HUMor spell checkert is használtuk, azóta azonban eltávolítottuk a HUMort a túlgenerálások miatt.

összetételi tagok helyesen vannak leírva, az eszköz helyesnek ítéli meg azt, következésképpen a két modul eltérő javaslatokat ad. A két eszközt a jövőben kívánjuk összehangolni, ez a fejlesztés azonban több időt igényel.

A helytelen szó helyett javasolt alakokat program generálja karakterhasonlóság alapján, emiatt születnek néha jelentésükben különböző vagy értelmetlennek tűnő megoldások (\**kőcsodás*, \**helyősíres*). A kiszűrésükre irányuló fejlesztések folyamatosan zajlanak.

## Névkereső

A tulajdonnevek helyesírására számos szabály vonatkozik (AkH. 154–201.), amelyek gépesítését adatbázisokkal oldottuk meg (jelenleg több mint kétszáz ezer adattal rendelkezünk). A *Névkereső* eszköz agyaga több forrásból származik: a hunmorph, a huntag és a Hunspell szabadon letölthető erőforrásai, a Magyar Posta nyilvános listái, a publikusan hozzáférhető telefonkönyvi adatok stb. (Miháltz et al. 2012: 139). Jelenleg több mint százezer magyar vonatkozású földrajzi név, több mint százezer vezetéknev és körülbelül hatezer keresztnév és becenév található az adatbázisban. Az intézmény- és márkanevekkel alapvető probléma, hogy a bejegyzett nevek közül sok nem normatív helyesírással szerepel a cégbíróság nyilvántartásában (Váradi 2013: 326).

A szabad keresés helyett a *Névkereső* prediktíven, folyamatosan bővülő listában jeleníti meg a találatokat a gépelés elkezdésével. A tulajdonnevek alatt egyértelműsítő besorolások (kategóriák és egyéb jegyek) szerepelnek – vezetéknev/keresztnév, férfi/női vagy becenév, illetve településnév; közterület neve, magyarországi vagy nem (Miháltz et al. 2012: 139). Ezenkívül földrajzi neveknél bizonyosan az *-i* képzővel ellátott alakok is megtalálhatók.

Sajnos egyelőre az olyan esetekben, amikor kétséges, hogy bizonyos alakot, amely nem számít tulajdonnévnek – a képzőktől függetlenül –, kis- vagy nagybetűvel kell-e írni, a rendszer nem ad választ. Ilyenek például az egyezmények, a fontosabb napok vagy ünnepek (*tordai országgyűlés*, *anyák napja*) (Kovács 2013).

## Elválasztás

A portál eszköze egy létező alkalmazás továbbfejlesztett változata: az OpenOffice/LibreOffice és az arra létrehozott huhyphn. Azoknak a hibáknak, hiányosságoknak a kiküszöbölésére, hogy a többféleképpen elválasztható szavakat is felismerje és kezelje a rendszer, a HUMor morfológiai elemzőt alkalmaztuk – rendszerint az összetett és igekötős szavak esetében. Az elemző szóegyértelműsítést végez, felismeri a szóhatárokat, és jelöli az összetétel határát is | jellel. Ha a kifejezés potenciálisan összetétel is lehet, és emiatt több megoldás lehetséges, mindet megjeleníti (*me-gint*, *meg-int*) (Miháltz et al. 2012: 139–140).

## Számok

A számjegyek betűvé alakítása viszonylag könnyen automatizálható, mivel a számjegyek és a betűvel írott számok kölcsönösen megfeleltethetők egymással, továbbá a hosszú számjegyek írásmódjának szabályai közérthetőek, egyértelműek. A teljesen saját fejlesztésű számjegy-szöveg átalakító a tőszámneveken kívül a sorszámneveket, valamint a hagyományos és a tizedes törteket is képes betűvel írt kifejezéssé alakítani.

A törtek és a *kettő* – *két* tőszámszavak esetében több helyes írási lehetőség: a törteknél az egybe- vagy különírás változik jelentéstől, kontextustól függően, a *kettő* – *két* esetében pedig a szóalak, szintén a jelentéstől függően (jelzői értelemben a *két* használata az elfogadott). Ezekre megjegyzésekkel hívjuk fel a felhasználók figyelmét.

A *Számok* eszköz használatakor pontosan be kell tartani bizonyos formai előírásokat: számokat kell beírni a megadott központozással.

## Dátumok

A dátumok helyesírása gyakran okoz problémát (például mikor kell az évszám vagy a nap után pontot tenni). A dátumírás szabályai viszonylag könnyen gépesíthetők, ezt használja ki ez a saját erőből fejlesztett webalkalmazás. A modulban a hónapok és napok listáját használjuk fel. Ezt a hangrend szerint csoportosítottuk, s ez alapján javaslatok generálhatók. Bizonyos jelenségeket azonban nem kezel, amelyekre a *Hogyan működik?* oldalon (4) adott leírásban hívjuk fel a figyelmet, például a szökőévet vagy a Gergely-naptár bevezetésekor kimaradt napokat. Emellett csak potenciálisan létező dátumokra ad megoldást, tehát például a 2012-10-42-re nem.

A *Dátumok* eszközről ugyanaz mondható el, mint a *Számok* eszközről: a formai előírások betartásával gyakorlatilag nem ad hibás javaslatot. A dátum beírása után az összes lehetséges módon központozott forma közül a felhasználónak csupán ki kell választania, amit keres.

## Ábécébe rendezés

Annak érdekében, hogy a szabályzatnak az ábécébe rendezésre vonatkozó pontjait működésbe hozzuk, először négy előfeldolgozási lépést hajtottunk végre; ezt követően használhatóvá vált a klasszikus rendezési algoritmus, azaz a két karaktorsor (sztring) összehasonlítása. Az első különböző karakterpár adja a két sor összehasonlításának eredményét, és ha az egyik karaktorsor egy az egyben a másik elejét alkotja, akkor a hosszabbat tekintjük a másodiknak.

A portál más nyelvek ábécéje szerinti rendezésére nem alkalmas. Nem kezeli továbbá a bibliográfiai tételek betűrendjét, illetve a számokat is tartalmazó tételeket.

## A fejlesztési tervekről

A portál az e-mailek, az oldallátogatások és a visszatérő felhasználók alapján sikeresnek mondható. Természetesen a munka nem állt le, és a javítások mellett vannak rövid, közép- és hosszú távú terveink. A rövid távú tervek között szerepel a folyamatos tesztelés, valamint a visszajelzett hibák kiszűrése, illetve a hiányosságokat pótlandó a mostani Archivum bővítése. A kategóriák finomításán kívül a válaszokat megjegyzésekkel látjuk el, és az azóta érkezett e-maileket is feldolgozzuk.

Az egyik legnagyobb, több időt igénylő probléma az, miként korlátozzuk a rendszer túlgenerálását; a szabályok sok olyan összetételt is előállítanak, amelyek szabályosan vannak képezve, azonban nem vagy csak nehezen értelmezhetők (Váradi 2013: 330).

Fontos feladatnak tekintjük továbbá, hogy a magyarázatokat, amelyek most többnyire szó szerinti idézések az AkH.-ból és az *Osiris Helyesírásból* (Laczkó–Mártonfi 2004), közérthetőbbé tegyük. Azt tervezzük, hogy a könnyedebb stílus kialakításán kívül a nehezen érthetőnek ítélt terminusokhoz rövid magyarázatot adunk.

## Összefoglalás

A helyesírás.mta.hu nemzetközi összehasonlításban is ritka, úttörő vállalkozás, amely nyelvtechnológiai eszközök felhasználásával nyújt helyesírási tanácsokat. A helyesírási szabályzat nagy részének szabályalapú rendszerbe foglalása az egyetlen olyan szakszerű megoldás, amely a nyelvhasználat változásait is képes követni, ennél fogva időtálló is.

## ÁBRÁK

1. ábra: Ludányi Zsófia – Miháltz Márton – Hussami Péter (2013): Különírás-egybeírás – automatikusan. In: Várad Tamás (szerk.) *AlkNyelvDok7. Doktoranduszok tanulmányai az alkalmazott nyelvészet köréből*. Budapest: MTA Nyelvtudományi Intézet, 119.
2. ábra: Miháltz Márton – Ludányi Zsófia (2013): *helyesírás.mta.hu: automatizált helyesírási tanácsadás nyelvtechnológiai módszerekkel*. Magyar Tudomány Ünnepe 2013. Budapest, 2013. november 11. (Prezentáció, kézirat.) [http://www.nytud.hu/archiv/matu2013/mihaltz\\_ludanyi.pdf](http://www.nytud.hu/archiv/matu2013/mihaltz_ludanyi.pdf) (Letöltve: 2013. 11. 20.)
3. ábra: <http://helyesiras.mta.hu/helyesiras/default/kulegy?q=bontott+t%C3%A9gla+fal> (Letöltve: 2013. 12. 03.)

## FORRÁS

- (1) <http://helyesiras.mta.hu/helyesiras/blog/show/helyes-hasznalat-kulonvagyegybe>
- (2) <http://hunspell.sourceforge.net/>
- (3) <http://www.morphologic.hu/>
- (4) <http://helyesiras.mta.hu/helyesiras/default/howitworks>  
(Valamennyi letöltve: 2014. 03. 10.)

## IRODALOM

- AkH. = Magyar Tudományos Akadémia (2000): *A magyar helyesírás szabályai*. 11. kiadás. 12. (példaanyagában átdolgozott) lenyomat. Budapest: Akadémiai Kiadó.
- Kis Ádám (1999): Az akadémiai helyesírási szabályzat és a számítógép. *Magyar Nyelvőr* 123/2, 149–168.
- Kovács Réka (2013): *Helyesiras.mta.hu*. helyesiras.mta.hu, az intelligens helyesírási tanácsadó portál. *Anyanyelv-pedagógia* 3. (Elektronikus dokumentum.) <http://anyanyelv-pedagogia.hu/cikkek.php?id=474> (Letöltve: 2013. 12. 03.)
- Laczkó Krisztina – Mártonfi Attila (2004): *Helyesírás*. Budapest: Osiris Kiadó.
- Ludányi Zsófia – Miháltz Márton – Hussami Péter (2013): Különírás-egybeírás – automatikusan. In: Várad Tamás (szerk.): *AlkNyelvDok7. Doktoranduszok tanulmányai az alkalmazott nyelvészet köréből*. Budapest: MTA Nyelvtudományi Intézet, 116–130.
- Mártonfi Attila (2007): Időszerű-e A magyar helyesírás szabályai 12. kiadásának az előkészítése? In: Domonkosi Ágnes – Lanstyák István – Posgay Ildikó (szerk.): *Műhelytanulmányok a nyelvvelésről*. Budapest: Gamma Nyelvi Iroda – Tinta Könyvkiadó, 134–138.
- (2011): A helyesírás mint szabályrendszer. *Café Babel* 65: 63–71.

- Miháltz Márton – Hussami Péter – Ludányi Zsófia – Mittelholcz Iván – Nagy Ágoston – Oravecz Csaba – Pintér Tibor – Takács Dávid (2012): Helyesírás.hu – Nyelvtechnológiai megoldások automatikus helyesírási tanácsadó rendszerben. In: Tanács Attila – Vincze Veronika (szerk.): *MSZNY 2013. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged: JATEPress, 135–148.
- Miháltz Márton – Ludányi Zsófia (2013): *helyesírás.mta.hu: automatizált helyesírási tanácsadás nyelvtechnológiai módszerekkel*. Magyar Tudomány Ünnepe 2013. Budapest, 2013. november 11. (Prezentáció, kézirat.) [http://www.nytud.hu/archiv/matu2013/mihaltz\\_ludanyi.pdf](http://www.nytud.hu/archiv/matu2013/mihaltz_ludanyi.pdf) (Letöltve: 2013. 11. 20.)
- 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*. Szeged: JATEPress, 172–182.
- Prószéky Gábor – Kis Balázs (1999): *Számítógéppel emberi nyelven*. Budapest: Szak Kiadó.
- Selyem Zoltán (2009): *Szemantikai kategóriák és rákérdezés. A helyesírási kérdések eldöntéséhez lényeges szemantikai jegyek listája, valamint a felhasználó által beírt kifejezések félinteraktív elemzése*. (Kézirat.)
- Váradi Tamás (2013): Gépesített helyesírási tanácsadás. *Argumentum* 9, 321–331. [http://argumentum.unideb.hu/2013-anyagok/kulonszam/20\\_varadit.pdf](http://argumentum.unideb.hu/2013-anyagok/kulonszam/20_varadit.pdf) (Letöltve: 2014. 03. 10.)

## TINTA KÖNYVKIADÓ

MARGALITS EDE – KOVÁCS ERZSÉBET  
MAGYAR–LATIN KÖZMONDÁSSZÓTÁR

2000 magyar közmondás, szólás klasszikus és újkori latin megfelelője  
226 oldal, 1490 Ft

A TINTA Könyvkiadó jóvoltából most először kerül egy kötetben az érdeklődő nagyközönség elé Margalits Ede *A magyar közmondások és szólások rostája* című, páratlanul izgalmas gyűjteménye. A szerző munkáját több részletben adta közre a Nyelvtudományi Közlemények folyóiratban 1899-től.

A *Magyar–latin közmondásszótárban* 2000 magyar közmondásszerű bölcsesség található latin megfelelőikkel, pontos bibliográfiai adatok feltüntetésével. A latin anyag forrásai többek között Cicero, Erasmus, Szent Jeromos, Horatius, Ovidius, Plinius, Seneca és Vergilius művei. A Bibliából 175 közmondás származik.

Az utóbbi évtizedek filológiai kutatásainak eredményeit felhasználva Kovács Erzsébet modern szövegkiadások és számítógépes szövegtörzsek felhasználásával egészítette ki Margalits Ede eredeti gyűjteményét és pontosította annak bibliográfiai hivatkozásait. A szótár célja, hogy felfedezzük, mennyi mindennel gazdagította a latin nyelv a magyar művelődést, és hogy milyen szervesen építette magába az antik kultúrát a közmondásokon keresztül a magyar nyelv.

Megvásárolható a kiadóban:

TINTA KÖNYVKIADÓ

1116 Budapest, Kondorosi út 17., Tel.: (1) 371-0501; fax: (1) 371-0502

E-mail: [info@tintakiado.hu](mailto:info@tintakiado.hu); honlap: [www.tintakiado.hu](http://www.tintakiado.hu)