

## **Az -ó/-ő képzős igenévi jelzős alakulatok helyesírásáról – korpusznyelvészeti alapon**

1. Az írásgyakorlat és a helyesírási szabályok által meghatározott norma közti eltérések korpusznyelvészeti elemzése számos tanulsággal szolgálhat mind a helyesírással foglalkozó alapkutatás, mind a nyelvi tervezés számára. E tágabb kutatási területen belül a nagyméretű korpuszokba foglalt szövegek kvantitatív elemzése különösen alkalmas lehet arra, hogy az írásbeli magyar nyelvhasználat minél nagyobb részére nézve helytálló, pontos választ kapjunk számos lényeges kérdésre. A helyesírás tervezése szempontjából ilyen kérdés például, hogy mely helyesírási szabályok által lefedett jelenségek terén áll fenn jelentős mértékű ingadozás a szabályozás által helyesnek, illetve helytelennek minősített írásváltozatok között, tehát mely pontokon különösen bizonytalanok a nyelvhasználók abban, hogy melyik írásváltozatot kell alkalmazniuk. Ugyanígy hasznos lehet azt is megvizsgálni, hogy mely pontokon távolodik el leginkább az írásgyakorlat a normától, más szóval mely szabályokat sértik meg a nyelvhasználók a leggyakrabban.

Ugyan ebben az értelemben a korpuszokon végzett kvantitatív helyesírási vizsgálatok egyrészt roppant tanulságosak volnának, másrészt a kutatható anyag is viszonylag könnyen és tömegesen elérhető a magyar nyelvű weben, meglepő módon – legalábbis tudomásunk szerint – nincs példa ilyen jellegű publikációkra a magyar nyelvészeti szakirodalomban. A helyesírási norma betartásának kérdéseit általában anyanyelv-oktatási és szövegírói tapasztalatok kapcsán, anekdotikus módon szokták felvetni, azaz egyéni példákra vagy jobb esetben kisebb adatgyűjtésekre, jellegzetesen iskolai felmérések eredményeire támaszkodva (a számos idevágó írás közül vö. BENYHE 2004; SZÖKÉNÉ–SZÖKE 1987).

Erre tekintettel jelen tanulmányban arra vállalkozunk, hogy egy jelentős méretű korpusz alapján elvégzett elemzés formájában demonstráljuk a kvantitatív korpuszvizsgálatokon alapuló helyesírás-kutatás hasznosságát. Az itt ismertetett eset-tanulmány egy konkrét szabály követésével, illetve az attól való eltérésekkel foglalkozik: arra összpontosítunk, hogy miként ingadozik az előtagként *-ó/-ő* képzős igenévet (pl. *élő zene, törlesztőrészlet*), illetve ilyen fejű összetételt (pl. *rovarölő szer, vendéglátóegység*) jelzői funkcióban tartalmazó alakulatok írásgyakorlata.

2. A vizsgált kérdéskör választását az indokolta, hogy egyfelől úgy véltük, a helyesírási szabályzat vonatkozó pontja korántsem periferikus, azaz aránylag gyakori jelenséget ölel fel. Az utóbbi sejtésünket – ahogyan azt az adatok kvantitatív elemzéséből alább látni fogjuk – a korpusz meg is erősítette. Másrészt előzetes megfigyeléseink alapján a szabály alkalmazásában egyéb helyesírási kérdésekhez képest különösen nagy fokú bizonytalanságra számítottunk még a helyesírásukra fokozottan odafigyelő professzionális szerzők esetében is. Úgy gondoltuk, ez a bizonytalanság nem pusztán abból fakad, hogy a szövegezők hiányosan ismerik

a vonatkozó szabályokat, hanem maguknak a szabályoknak a bonyolultságából, nehezen értelmezhető voltából, és ebből következő rossz alkalmazhatóságából ered.

Kiinduló feltételezéseinket a téma szakirodalma is alátámasztani látszik. A külön- és egybeírás ingatag voltára már maga az AkH. 1984 95. pontja (42) és az OH. (90) is utal, de fontos problémaként jellemzik ezt a helyesírással, illetve anyanyelv-pedagógiával foglalkozó szakemberek és a leíró nyelvészek is (FÁBIÁN 1984: 394; MISAD 2004: 188; MÁRTONFI 2011: 69; NÁDASDY 2004: 79; ANTALNÉ 2009: 129 stb.). E tekintetben olyannyira nagy az egyetértés, hogy LACZKÓ és MÁRTONFI a különírás és egybeírás témájával foglalkozó vitaanyag nyitómondatában egyenesen így fogalmaz: „Köztudott tény, hogy a magyar helyesírás legneuralgikusabb pontja a különírás és az egybeírás problémaköre” (LACZKÓ–MÁRTONFI 2004: 141). Ezen belül is kiemelt helyet foglalnak el az -ó/-ő képzős igenévi jelzőt tartalmazó összetételek és szószerkezetek. Helyesírásukat az OH. szerzői a külön- és egybeírás „talán egyik legnehezebb kérdéskörének” vélik (113), amire egyébként már az OH. megjelenése előtt külön is felhívták a figyelmet: „Az -ó/-ő képzős melléknévi igenévi előtaggal rendelkező szerkezetek helyesírási szabályainak megfogalmazása feltétlenül átgondolást igényel a bizottság részéről, s ennek alapján a szótári rész egyes példáinak rögzítését is újra kellene gondolni” (LACZKÓ–MÁRTONFI 2004: 150).

Úgy tűnik tehát, hogy itt a helyesírási szabályzat egy olyan pontjának a megfogalmazásával, értelmezhetőségével kapcsolatban merülnek fel aggályok, amely pont egyébként különös fontosságú, ugyanis az általa tárgyalt összetétel-, illetve szószerkezet típusnak mind az előfordulási gyakorisága, mind pedig az írásmódját illető bizonytalanság mértéke jelentős. Tanulmányunkban a rendelkezésünkre álló empirikus adatok alapján ez utóbbi tényezőket – a kifejezéstípus gyakoriságát és az ingadozás mértékét – próbáljuk meg minél pontosabban számszerűsíteni, hogy tisztábban láthassunk abban a kérdésben, objektíven nézve milyen súlyú problémával állunk szemben.

**3. Az -ó/-ő képzős igenévi jelzős alakulatok helyesírását az AkH. 1984 alábbi, 112. pontja szabályozza (AkH. 1984: 46–7).**

„**112.** Az -ó, -ő képzős igenevet, ha csak alkalmi minőségjelző, általában különírjuk jelzett szavától: *dolgozó nő, izzó fém, közvetítő javaslat, parancsoló hang, sajtó térd, felszabadító mozgalom, kiváló termék*, (az újoncokat) *kiképző tiszt* stb.; különösen olyankor, ha a kapcsolatnak valamelyik vagy mindkét tagja összetett szó: *mutató névmás, szélező körfűrész, növényvédő szer, rakétaindító állvány, földrengésjelző műszer, ellenőrző bizottság, összekötő híd, előadó körút* stb.

Ha viszont az igenévi jelzős kapcsolatok tagjai jelentés tekintetében összeforrtak (az igenév nem folyamatot, hanem valamire való képességet, rendeltetést fejez ki; illetőleg a jelzett szó az igenévben kifejezett cselekvésnek valamilyen határozója), a két egyszerű szót egybeírjuk: *cséplőgép, izzólámpa, desztillálótorony, védőnő, ivóvíz, költőpénz, eladó lány* (üzletben), *kiképzőtiszt* (beosztás), *felvonóhid; hálószoba, kutatóintézet, pihenőnap, evezőlámpa, átütőpapír, belépőjegy*; stb. (Vö. **138.**)

A kialakult szokást megtartva több olyan (részben jelentésváltozást is mutató) kapcsolatot is egybeírunk, amelynek valamelyik tagja összetett szó: *forgószínpad, javítóműhely, mozgókórház, mérőműszer, előhívótank, vendéglátóipar* stb.”.

Az OH. ezt részletezi azzal a kiegészítéssel, hogy ha a jelző és/vagy az alaptag önmagában is összetétel, a jelentésváltozás ellenére többnyire különírást kell alkalmazni (113–114), majd pedig a következőket jegyzi meg (OH. 114):

„[a] vonatkozó szabály azonban nem szigorú, csupán tendencia. Elsősorban azért, mert több olyan jelzős szókapcsolatot egybe kell írni, amelyben az egyik tag önmagában is összetétel. Jellemzően egybeírandók azok az összetett alaptagot tartalmazó szerkezetek, amelyekben az alaptagnak mind az előtagja, mind az utótagja egy szótagos szóalak, s összetétel voltuk a mai nyelvi tudás számára részben elhomályosult: *javítóműhely, forgószínpad, mozgókórház, mérőműszer*. Ezen túlmenően léteznek olyan melléknévi igenévi jelzős kapcsolatok, amelyekben az egybeírás pusztán az írásszokáson alapul: *vendéglátóipar, előhívótank*. A folyamatos melléknévi igenévi előtagú szerkezetek formai megítélésükor a két vagy több szótagos igekötők összetételi tagnak számítanak, tehát például az *ellenálló képesség, összekötő híd 2 + 1* szerkezetű összetétel. Kivétel azonban e szabály alól az *előadóművész* egybeírása a formailag szabályos különírás helyett”.

**4.** Vizsgálatunk során a P.A.G.-Studium Bt. által összeállított „TöSzTeKK” (Többnyelvű Szövegtechnológiai Kutatási Korpusz) elnevezésű korpusz három magyar nyelvű részkorpuszát használtuk fel: az első 2000 óta keletkezett online újságcikkeket tartalmaz (mintegy 250 millió szövegszó terjedelemben), a második zömében fiatal amatőr szerzők által írt prózai szövegeket (2005 óta; kb. 150 millió szövegszó), végül a harmadik blogbejegyzéseket (2005 óta; kb. 5000 blogról kb. 300 millió szövegszó). Ezeket a részkorpuszokat rendre SAJ-jal, AP-vel, illetve B-vel rövidítjük az alábbiakban.

Abból indulhatunk ki, hogy a három vizsgált részkorpusz a viszonylag igényesen megírt, többnyire szabálykövetésre törekvő szövegektől (online újságcikkek) a normát gyakran figyelmen kívül hagyó stílusig (blogok) a helyesírás szempontjából nézve eltérő színvonalú írások igen széles spektrumát tartalmazza. A teljes vizsgált szövegmennyiség (700 millió szövegszó) viszonylag nagynak mondható (összehasonlításképp a Magyar Nemzeti Szövegtár mérete kb. 200 millió szövegszó), és a korpuszunk téma, illetve műfaj szempontjából is sokoldalú. Ugyan a szövegek szerzőinek számáról pontos adatokkal csak az AP részkorpusz kapcsán rendelkezünk (ebben kb. 5000 szerzőtől találunk szövegeket), becsléseink szerint számuk nagyjából 10 000-re tehető az egész korpuszban, és így a magyar lakosságnak több nagyságrenddel nagyobb részét fedi le, mint az 1. pontban említett felmérések. E szempontok alapján úgy véljük, a kutatás empirikus háttere viszonylag magas fokon reprezentatív a normakövetésre törekvő írásbeli magyar nyelv-

használat egészére nézve. A vizsgálati adatok halmazába ugyanakkor szándékosan nem vettünk bele olyan környezeteket, amelyekben magától értetődő, várható, illetve szubkulturális szabályként jelenik meg a helyesírási normák (akár szándékos) megsértése, például kisiskolások, fiatalok kommunikációja a közösségi hálózatokon, blogbejegyzésekhez fűzött kommentek stb.

Az esettanulmány empirikus adathalmazát az említett részkorpuszokban előforduló szóalakok gyakorisági listái, táblázatai jelentették. Ezen belül kétféle gyakorisági listával dolgoztunk: az egyik (a továbbiakban: 1g-táblázat) a számítógépes nyelvészeti zsargonban 1-gramoknak nevezett szóalakokat, a másik (2g-táblázat) a 2-gramoknak vagy bigramoknak nevezett alakpárokat – tehát valamely szövegben egymás mellett előforduló szóalakokat – foglalta magában. Mindkét típusú gyakorisági lista a konkrét alakokhoz természetesen megadta az alak, illetve pár megfelelő részkorpuszon belül számolt gyakoriságát. E táblázatokból csak bizonyos sorok relevánsak a tanulmány kérdésfelvetése szempontjából: a 2-gramok listájából csak azok, amelyek tokenpárjai egybeírva megtalálhatóak az 1-gramok listájában; ezen belül is csak azok a sorok, ahol a 2-gram előtagja folyamatos melléknévi igenév, illetve ilyen alaptagú szó; továbbá értelemszerűen e 2-gramok egybeírt változatai az 1-gramok listájából. Hogy ezt konkrét példákkal szemléltessük: az AP részkorpusz 2g-táblázatából a *röhögő görcs* tokenpár (113 előfordulás) sora érdekes számkra, ugyanis az 1g-táblázatban szerepel a *röhögőgörcs* alak (150 előfordulás), továbbá a *röhögő* előtag folyamatos melléknévi igenév. Ugyanígy a B részkorpusz 2g-táblázatából releváns a *vendéglátó egységekben* (14 előfordulás) sora, ugyanis az 1g-táblázatban szerepel ennek egybeírt változata, a *vendéglátóegységekben* (9 előfordulás), és előtagja folyamatos melléknévi igenévi fejű összetétel.

Felhívjuk rá az olvasó figyelmét, hogy a vizsgálat során sem az elő-, sem az utótagokat nem lemmatizáltuk, vagyis a toldalékolt alakokat nem vezettük vissza szótári alakjukra. Így például a *vendéglátó egység* és a *vendéglátó egységekben* szóalak előfordulásait nem vontuk össze a táblázatban egyetlen sorba, és az utóbbi alak (valamint egyéb toldalékolt alakok) előfordulásainak számát nem vontuk össze az előbbi előfordulásainak számával. A lemmatizálástól pusztán az egyszerűség kedvéért tekintettünk el, a korpusz ilyen irányú előfeldolgozása ugyanis jelentős mennyiségű további munkát igényelt volna, miközben nem világos, hogy bármi hasznot hozott volna. A lemmatizálás hiánya ugyanis csak olyankor okozhat problémát, ha például egy különírt változat megtalálható a korpuszban (pl. *röhögő görcsödet*), annak egybeírt párja (*röhögőgörcsödet*) viszont nem (vagy épp fordítva). Ha lemmatizáltunk volna, ilyenkor a pár nélküli, adott esetben a különírt változatot figyelembe vehetnénk a megfelelő „lemma” (*röhögő görcs*) gyakoriságának meghatározásakor, míg lemmatizálás nélkül a pár nélküli alakok „elvesznek”, ezeket figyelmen kívül hagyjuk. Vegyük azonban észre, hogy ilyen csak két szituációban fordulhat elő jelentős valószínűséggel: Az egyik lehetőség, hogy összességében sokkal ritkább a konkrét ragozott alaktól eltekintve is a különírt vagy az egybeírt változat a másikkal; ilyen esetekben viszont, amint lentebb látni fogjuk, eleve figyelmen kívül hagynánk az adott párt a jelen vizsgálat keretében. A másik

lehetőség, hogy a változatok egyébként nagyjából hasonló gyakran fordulnak elő, így ugyanezt várjuk az egyes alakjaik esetében is. Ha viszont egy ilyen alakulat egy bizonyos toldalékolt alakjából például az egybeírt változat nulla gyakorisággal fordul elő, akkor annak különírt megfelelője is nullához közeli gyakoriságú, tehát többnyire hapax lesz, de legalábbis nagyon ritka. Feltételezzük, hogy az ilyen nagyon kis gyakoriságú elemek elvesztése sem a kvantitatív adatainkat nem torzítja jelentősen, sem a kvalitatív jellegű észrevételeink érvényességére nincs hatással.

Visszatérve a gyakorisági listákra: ahhoz, hogy a 2g-táblázatokból kiszűrjük a számunkra ebben az értelemben hasznos sorokat, a 2-gram előtagjának és utótagjának formai jegyei alapján selejteztük ki a nem releváns párokat. Mindenekelőtt csak azokat a 2-gramokat tartottuk meg, amelyek csak betűkből állnak, az előtag *ó/ő*-re végződik és legalább 3 karakter hosszú, az utótag legalább 2 karakter hosszú, stb. További lépésekben fokozatosan eltávolítottuk az adathalmazokból az olyan 2-gramokat, amelyek az említett durvább szűrőkön nem akadtak fenn, különösen az eleve nem melléknévi igenévi (pl. *nettó, olcsó, rádió, idő, erő*), illetve az összetételek előtagjaként gyakorlatilag soha nem melléknévi igenévi funkcióban megjelenő előtagúakat (pl. *való, adó, késő, jövő, szóló*), valamint az irreleváns utótagúakat (pl. *ként, ért, képp, képpen, képes*).

Az így kapott adathalmaz ugyanakkor a gondos szűrés ellenére is tartalmazott nagy mennyiségű olyan külön- és egybeírt alakváltozatot, amely tanulmányunk kérdésfelvetése szempontjából lényegtelen. Ilyenek például az elgépelések (pl. *következődolgoz*) és a normának megfelelő változathoz képest nagyon ritkán előforduló helyesírási hibák (pl. *forogató könyvét*), amelyek nem járulnak hozzá annak megválaszolásához, hogy mely szókapcsolatok esetében okoz nehézséget a 112. szabály alkalmazása, tehát hol tapasztalunk jelentős ingadozást a külön- és egybeírt változat között. Ebből a megfontolásból tovább szűkítettük a részletesebb vizsgálataink során használt adathalmazt: Csak azokat a tokenpárokat vettük figyelembe, amelyek leírása erőteljesen ingadozik. Az erőteljes ingadozást (önkéntesen, de az így kapott eredmények ésszerűségét ellenőrizve) úgy definiáltuk, hogy ez akkor áll fenn, ha a különírt és az egybeírt változat előfordulási gyakorisága között az arány szigorúan kisebb, mint 1:4 bármelyik irányban.

5. A hozzávetőlegesen 700 millió szövegszó méretű korpuszban mintegy 23 ezer olyan pártípust (tokenhalmazt) találtunk, amelynek a külön- és egybeírt változata a fenti értelemben erőteljesen ingadozik. Ezeket összesen több mint 200 ezer token, azaz konkrét előfordulás (szövegszó, illetve -szópár) képviseli a részkorpuszokban, ami azt jelenti, hogy durván minden 3500. szövegszónál okoz nehézséget a 112. szabály alkalmazása a nyelvhasználók számára, tehát tíz gépelt oldalanként (kb. 350 szós oldalakat alapul véve) egy ilyen szót találunk. A részletet az alábbi táblázat foglalja össze:

**1. táblázat**

A 112. szabály alkalmazásának gyakorisága részkorpuszokra lebontva

Részkorpusz	Típusok, ~	Tokenek, ~	Token/típus, ~	Gyakoriság, 1/hány oldal?
SAJ	9 200	96 000	10	7
B	14 100	99 000	7	9
AP	2 650	21 000	8	20
Összesítve	22 900	216 000	9	9

Ezen a ponton megjegyzendő, hogy az itt és az alábbi táblázatokban közölt számok nem teljesen pontosak, valójában felülről közelítik a jelenség tényleges gyakoriságát. Egyrészt mivel nem tokenizáltunk, egy kis mennyiségű releváns adat nem került be az adathalmazunkba, másrészt az adathalmaz méreténél fogva aránytalanul nagy mennyiségű idő- és energiaráfordítást igényelt volna minden egyes bekerült tokenet vagy akár csak a típusokat tételesen ellenőrizni abban a tekintetben, hogy valóban relevánsak-e a 112. szabály szempontjából. Az adott körülmények között ugyanakkor igyekeztünk körültekintően eljárni és a fentiekben vázolt módon, illetve ezenfelül a fennmaradó anyag szűrőpróbaszerű ellenőrzésével kiszűrni minden irreleváns adatot. Ennek eredményeként az 1. táblázatban szereplő számokat alább, a 9. szakaszban egy részletesebben átvizsgált minta alapján pontosítjuk, illetve az általunk szintén közzétett részletes adathalmazok alapján az olvasó megbizonyosodhat arról, hogy a fennmaradó irreleváns adatok mind a típusok, mind a tokenek számát tekintve viszonylag elhanyagolhatóak, a számadataink arányain, nagyságrendjén számottevően nem változtatnak.

**6.** A szűrt adathalmazunk alapján elsőként azt a kérdést vizsgáltuk, hogy mely előtagok kapcsán fordul elő a legnagyobb mértékű ingadozás a külön- és az egybeírt változat között a három részkorpuszban összesítve. Ez az információ hasznos lehet helyesírási szótárak készítése szempontjából.

A leggyakrabban ingadozó tíz előtagot az alábbi táblázatban közöljük. Minden előtaghoz zárójelben megadjuk azt a két utótagot, amellyel az adott előtag a leggyakrabban megtalálható a teljes korpuszban, valamint a 2. oszlopban az adott előtaggal előforduló erőteljesen ingadozó utótagok számát. Miután az előtag és az utótag alkotta pár határoz meg egy ingadozó típust, az ebben az oszlopban megadott számok azt jelentik, hogy hány ingadozó típus esik az 1. táblázatban látott kb. 22 900-ból az adott előtagra. Megjegyzendő, hogy az utótagokat itt sem lemmatizáltuk, tehát egyazon tő két ragozott vagy képzett alakja két típusnak számít. A 3. oszlopban közölt számok a különírt és egybeírt tokenek számának összegét adják meg. A táblázat sokkal részletesebb változatát, amelyben megtalálható a 10 előfordulásnál gyakoribb összes típus, valamint minden típus különírt változatának összes konkrét előfordulása konkordancia formájában, a <http://mnytud.arts.klte.hu/tkny/112-szabaly/elotagok> címen közöljük.

**2. táblázat**

Azon előtagok összesített táblázata, melyekkel a legtöbb utótag ingadozik erőteljesen

Előtag	Utótagok (típus)	Összesített gyakoriság (token)
<i>záró (napon, fordulóban)</i>	710	18 700
<i>kezdő (lökést, játékos)</i>	552	16 600
<i>nyitó (fordulóban, körben)</i>	550	15 100
<i>élő (zene, állat)</i>	473	27 000
<i>védő (hatása, falakat)</i>	466	14 000
<i>futó (játékos, csapat)</i>	437	9 500
<i>támadó (fal, futballt)</i>	418	9 900
<i>vezető (bíró, nő)</i>	387	35 500
<i>büntető (feljelentést, intézkedéseket)</i>	347	30 500
<i>kereső (tevékenységet, programok)</i>	307	7 200

7. Vizsgálatunk következő kérdésfelvetése, hogy mely erőteljesen ingadozó típusok a leggyakoribbak a korpuszunkban. Ez a kimutatás szintén helyesírási szótárak készítése szempontjából lehet hasznos, viszont a látszólagos hasonlóság ellenére teljesen más adatokról ad felvilágosítást, mint a 2. táblázat: egyfelől a kiugróan gyakori típusok előtagja nem okvetlenül ingadozik sok egyéb utótaggal, másfelől az összesítve legnagyobb mértékben ingadozó előtagokkal alkotott típusok egyenként nem feltétlenül gyakoriak.

Mivel a részkorpuszok igen eltérő témájúak, és így a bennük szereplő szerkezetek és összetételek is jelentősen különböznek, eredményeinket az alábbiakban részkorpuszok szerint bontva közöljük. A táblázatok az adott részkorpusz tíz leggyakoribb típusát tartalmazzák. A részkorpuszonként leggyakoribb 100 típust tartalmazó listát a <http://mnytud.arts.klte.hu/tkny/112-szabaly/parok> címen tesszük közzé.

**3. A) táblázat**

A SAJ részkorpusz leggyakoribb ingadozó párjai

Ingadozó pár	Különírva (darab)	Egybeírva (darab)
<i>vizsgáló + bizottság</i>	918	2 405
<i>kutató + intézet</i>	566	1 590
<i>felügyelő + bizottság</i>	464	722
<i>szálló + por</i>	897	277
<i>felügyelő + bizottsági</i>	278	732
<i>felügyelő + bizottságának</i>	291	694

Ingadozó pár	Különírva (darab)	Egybeírva (darab)
<i>nyomozó + hatóság</i>	461	467
<i>működő + tőke</i>	222	602
<i>büntető + feljelentést</i>	211	584
<i>szórakoztató + elektronikai</i>	301	429

### 3. B) táblázat

Az AP részkorpusz leggyakoribb ingadozó párjai

Ingadozó pár	Különírva (darab)	Egybeírva (darab)
<i>ülő + helyzetbe</i>	790	325
<i>felfedező + útra</i>	133	429
<i>kábító + átkot</i>	325	98
<i>kábító + átok</i>	198	66
<i>röhögő + görcs</i>	113	150
<i>röhögő + görcsöt</i>	110	107
<i>tisztító + bűbájt</i>	97	91
<i>támadó + állásba</i>	63	124
<i>lovagló + ülésben</i>	116	59
<i>kopogó + szellem</i>	37	132

### 3. C) táblázat

A B részkorpusz leggyakoribb ingadozó párjai

Ingadozó pár	Különírva (darab)	Egybeírva (darab)
<i>köszönő + viszonyban</i>	374	419
<i>képviselő + testület</i>	213	520
<i>tanító + néni</i>	386	181
<i>élő + zene</i>	209	252
<i>vizsgáló + bizottság</i>	98	338
<i>alkotó + eleme</i>	88	264
<i>törlesztő + részlet</i>	105	243
<i>felügyelő + bizottsági</i>	100	246
<i>kieső + helyen</i>	219	112
<i>sétáló + utca</i>	66	261

8. Végül azt a kérdést vetettük fel, hogy az ingadozó írású párokat milyen arányban írják a szabályok értelmében helyesen a nyelvhasználók, tehát milyen



nagy a hibázási arány. Az adatok e tekintetben történő feldolgozása során az irányadó AkH. 1984 mellett ennek az új, változatlan kiadásához virtuális mellékletként csatolt Internetes helyesírási szótárt, a DEME–FÁBIÁN–TÓTH szerkesztette akadémiai szótárat (1999), az akadémiai szabályozáson alapuló, azt részletező-kiegészítő OH. javaslatait és tekintélyes szótárát, valamint az Akadémiai Kiadó által közzétett, BÁRÁNY szerkesztette (2009) kötetet használtuk fel.

A korpuszban talált *-ó/-ő* képzős igenévi jelzős alakulatok lejegyzésének helyes vagy helytelen voltát az AkH. 1984 112. pontjának bizonytalanságai miatt leginkább a szótárak példaanyagára hagyatkozva állapítottuk meg. Azt, hogy a vonatkozó korpuszadatokat külön vagy egybe kell-e írni, akkor tartottuk eldönthetőnek, ha azok a szótárakban előfordultak (*lovagló ülés, fényképezőgép*) vagy azonos irányba mutató analóg példákat találtunk rájuk (*azonosító + lapka* külön, ugyanis: *azonosító állítmány, azonosító értelmező, azonosító kártya, azonosító mondat, azonosító szám; röhögő + görcs* egybe, ugyanis: *írógörcs, nevetőgörcs, rángógörcs, sírógörcs*). Az írásmódot kétféle esetben minősítettük eldönthetetlennek: egyrészt akkor, amikor a jelzős alakulat nem szerepelt a szótárakban, és ezenfelül – bár ez ritkán fordult elő – még analóg példái sem voltak (*barnító + szalon*); másrészt akkor, ha a talált analóg szavak és szó szerkezetek helyesírása eltérő irányba mutatott (*közvetítő + rendszer: közvetítő kereskedelem, büntető rendszer ↔ közvetítőállomás, hűtőrendszer*). Eldönthetetlen helyesírásúnak tekintettük továbbá azt a néhány *-ó/-ő* képzős igenévi jelzős alakulatot is, amelyeket a különböző szótárak példaanyagában eltérő írásmóddal találtunk meg, pl. *irányító központ* (DEME–FÁBIÁN–TÓTH 1999) ~ *irányítóközpont* (OH.).

9. Annak vizsgálata kapcsán, hogy a nyelvhasználók milyen gyakran írják a fenti értelemben helyesen az ingadozó írású párokat, a következő problémával szembesülünk. Ahhoz, hogy teljes képet kapjunk a szabálykövetés mértékéről, végig kellene nézni több mint 20 000 szót (lásd 1. táblázat) a helyesírási szótárakban, adott esetben nyomon követve a szótárak közötti eltéréseket, továbbá kezelve azt a problémát is, hogy egy összetétel nem található meg egyikben sem. Ezenfelül a konkrét kontextusokat is ellenőrizni kellene több mint 200 000 tokenél konkordanciák alapján, hogy meggyőződjünk arról, szókapcsolatszerű vagy összetételyszerű szintaktikai funkcióban használja-e a szerző az adott párt. Mindez gyakorlatilag kivitelezhetetlen volt számunkra. Ezért azt a megoldást választottuk, hogy a fenti lépéseket nem a korpusz egészén hajtjuk végre, hanem egy a teljes adathalmazból vett mintán, majd ez alapján becsüljük meg a keresett arányokat.

A mintavételhez az egyes részkorpuszok gyakorisági listáit felosztottuk centilisközökre úgy, hogy ezekben azonos számú token essen (tehát a bontás tokenszám, nem típusok darabszáma alapján történt). Minden centilisközből kiválasztottunk véletlenszerűen 5-5 típust, vagy ha a centilisköz legfeljebb 5 típus tartalmazott, az összeset. Az így kapott minta mérete tokenek tekintetében a teljes adathalmaz egyharmada, típusok tekintetében 5%-a (durván 70 ezer token, 1300 típus). A mintában szereplő párokat és tokenszámukat részkorpuszonként az alábbi címen tesszük közzé: <http://mnytud.arts.klte.hu/tkny/112-szabaly/mintavetel>.

A minta feldolgozása során célzottan kiszűrtük konkordanciák segítségével az olyan adatokat, amelyeket az 5. szakaszban körvonalazott durvább szűrési módszerek nem azonosítottak, viszont mégsem tartoznak a 112. szabály hatókörébe.

Erről van szó egyrészt akkor, ha egy főnévként is lexikalizálódott előtaggal és egy bizonyos utótaggal alkotott kapcsolatok leírását nem a 112. szabály határozza meg, hanem a főnévi előtagokkal történő külön- és egybeírást szabályozó különböző pontok [114. a), illetve b), 128. c), 129.)] valamelyike, például: *böngésző + kliens, varázsló + dinasztia; tűzoltó + daru; tanító + néni, képviselő + asszony, takarító + néni*. Ezeket típusként szűrtük (tehát nem tokenenként), ugyanis bár elméletileg előfordulhatnak ezeknek a kapcsolatoknak a 112. szabály alá eső, azaz melléknévi igenévi előtagú használatai (pl. *az iskolánkban tanító néni; az újságot böngésző kliens*), ezek azonban tanulmányunk kérdésfelvetése szempontjából érdektelenek. Gyakoriságuk elenyésző vagy nem létező, továbbá helyesírásuk nyilvánvalóan nem okoz nehézséget (pl. *az iskolánkban tanítónéni [nagyon kedves]* típusú leírások gyakorlatilag nem adathozhatóak).

Másrészt szintén figyelmen kívül hagyunk olyan tokeneket (ritkábban egész típusokat), ahol az előtag nem jelzője az utótagnak; például az alábbi konkordanciasorban az *álló + helyi* típus előfordulása ebben az értelemben nem releváns a témánk szempontjából: *által hordott vackoktól elég messze álló helyi batikművészet él és virágzik*.

A körütekintőbb szűrési eljárásnak köszönhetően a mintánk viszonylag nagy hányada, mintegy 20 százaléka bizonyult a tanulmányunk kérdésfelvetése szempontjából irrelevánsnak. A kb. 70 ezer tokenből releváns mintegy 60 ezer (~ 83%), illetve a kb. 1300 típusból 1000 (~ 78%). Ez – a mintából az adathalmaz egészére extrapolálva – arra utal, hogy az 1. táblázatban közölt számokat némileg felülbecsülik a vizsgált helyesírási probléma gyakoriságát. A pontosított becslés alapján durván minden 4000. szónál, tehát átlagosan bő 11 gépelt oldalanként okoz problémát a 112. szabály alkalmazása a nyelvhasználók számára.

**10.** A szabálykövetés arányára vonatkozó kérdésfelvetésünkre a következő eredményeket kaptuk részkorpuszonként, illetve összesítve:

#### 4. táblázat

Szabálykövetés és szabálytól való eltérés részkorpuszonként

Részkorpusz	Helyes (token)	Hibás (token)	Eldönthetetlen (token)	Összesen
SAJ	18 000 (54%)	9 650 (29%)	5 450 (17%)	33 100
B	7 900 (47%)	4 800 (29%)	3 950 (24%)	16 650
AP	5 500 (55%)	3 200 (32%)	1 300 (13%)	10 000
Összesen	31 400 (53%)	17 650 (30%)	10 700 (17%)	59 750

Amint a táblázatból leolvasható, a hibás írású változatok aránya átlagosan 30%. Előzetes sejtésünkkel ellentétben (vö. 4. szakasz) a részkorpuszok írásgya-

korlata között ebből a szempontból nem tapasztaltunk számottevő különbséget, sőt a hibák aránya mondhatni teljesen homogén. A blogok helyesírása a 112. szabály alá eső jelentősen ingadozó leírású párok tekintetében nem rosszabb, mint a sajtószövegeké. Mindössze az amatőr prózai szövegekben látunk minimálisan több hibát.

Érdekes azt is megvizsgálni, hogy milyen gyakoriak egymáshoz viszonyítva a mintánkon belül az olyan típusok, amelyeket 1. többnyire helyesen írnak, 2. amelyeket többnyire hibásan írnak, illetve 3. amelyeket közel egyforma gyakran írnak le helyesen és hibásan. Természetesen ehhez csak azokat az alakulatokat vettük figyelembe, amelyeknek biztosan eldönthető volt a helyes írásmódja. „Többnyire helyes”-nek (a határt önkényesen meghúzva) akkor tekintettünk egy típust, ha előfordulásainak legalább 60%-a lett helyesen leírva. Ehhez analóg módon határoztuk meg a „többnyire hibás” kategóriát is, míg az „egyforma gyakran hibás vagy helyes” kategóriába azokat a típusokat számítottuk be, amelyek egyik előbbibe sem lettek besorolva. Az 5. táblázat összegzi ennek eredményeit. Részkorpuszok szerint nincs értelme tagolni ezeket az adatokat, ugyanis majdnem pontosan ezeket az arányokat találjuk mindháromban.

### 5. táblázat

Többnyire helyesen, illetve hibásan írt típusok eloszlása

Típus írásmódja	Arány tokenzám alapján
Többnyire helyes	51%
Többnyire hibás	17%
Egyforma gyakran hibás és helyes	32%

A 6. táblázatban szemléltetésképpen néhány konkrét példát közlünk a szabálykövetésről szóló vizsgálatunk anyagából. Az első oszlopban első helyen (tehát helyesként) adtuk meg és félkövér szedéssel kiemeltük egyrészt azokat az alakokat, amelyeket a 112. szabály és (konkrétábban) a szótárak alapján csak külön írhatunk helyesen, tehát ahol az előtag ezek szerint csak alkalmi minőségjelzőnek ítéltető. Másrészt azokat az egybeírt alakokat jelöltük ugyanígy, amelyek szintén a szótárak szerint és a 112. szabály értelmében egybeírhatóak, azaz „az igenévi jelzős kapcsolatok tagjai jelentés tekintetében összeforrtak” (lásd 3. szakasz) tekinthetőek. Természetesen az, hogy ezek írhatóak egybe is, nem zárja ki azt, hogy bizonyos kontextusokban alkalmi minőségjelzői funkciót lásson el az előtag, tehát hogy ilyenkor a különírt alak legyen helyes. Egyértelműen erről van szó olyan kontextusokban, ahol a melléknévi előtagot bővítmény, azaz vonzat (pl. *az ellenzéket bíráló bizottság*) vagy szabad határozó (pl. *a feltűnően jelző táblák*) módosítja. Az ilyen okból különírt változatokat elkülönítettük a hibásan különírtaktól és helyesként kezeltük. A második oszlopban a helyes, illetve hibás előfordulások számát ennek megfelelően adjuk meg. Így fordulhat elő, hogy például a *jelző táblák* alak minden esetben helyesként lett besorolva, ugyanis bár a mintában szereplő 100 előfordulásból 25 külön van írva, ezek azonban mind így helyesek, mivel ténylegesen vonzattal állnak a konkordanciák tanúsága szerint.

**6. táblázat**  
Szabálykövetés és szabálytól való eltérés

Ingadozó pár (helyes / hibás)	Előfordulások száma (helyes / hibás)	Helyes előfordulás aránya
<i>beszélő viszonyban</i> – <i>beszélőviszonyban</i>	89 > 25	78%
<i>kormányzó párt</i> – <i>kormányzópárt</i>	225 > 73	76%
<i>habzó szájú</i> – <i>habzósájú</i>	132 > 49	73%
<i>köszönő viszonyban</i> – <i>köszönőviszonyban</i>	22 < 46	32%
<i>növényvédő szer</i> – <i>növényvédőszer</i>	56 < 129	30%
<i>jelzőtáblák</i> – <i>jelző táblák</i>	100 > 0	100%
<i>vizsgálóbizottság</i> – <i>vizsgáló bizottság</i>	3405 > 159	95%
<i>sétálóutca</i> – <i>sétáló utca</i>	261 > 66	80%
<i>alkotóeleme</i> – <i>alkotó eleme</i>	264 > 88	75%
<i>bírálóbizottság</i> – <i>bíráló bizottság</i>	92 < 145	39%

Végül a 7. táblázatban néhány példával illusztráljuk azokat a típusokat, amelyeket a 8. szakaszban kifejtettek értelmében eldönthetetlennek soroltunk be. Hangsúlyozzuk, hogy ezzel nem akarjuk azt állítani, hogy az összes ilyen példa esetében képtelenség volna kiválasztani egy „helyes” írásmódot, és a döntés mellett szóló érveket találni. Mindazonáltal ezek a párok kivétel nélkül nem szerepelnek a szótárakban, és más szócikkekben találunk mind olyan analóg példákat, amelyek a külön-, mind olyanokat, amelyek az egybeírás mellett szólnak. A szótárak általában nem tartalmazzak arra vonatkozóan magyarázatot, hogy a szerkesztő miért döntött egy-egy adott szókapcsolat adott leírása mellett, így a használó ilyen jellegű segítséget sem kap dilemmájának megoldásához. Továbbá a 112. szabály értelmében a döntés meghozatalához a beszélőnek határozott intuícióval kellene rendelkeznie arról, hogy az állandóság mint szemantikai jegy mennyiben része a szókapcsolat jelentésének. Sok esetben ilyen intuícióval nem rendelkezünk, illetve ha igen, akkor különböző beszélők intuíciói különböznek. Ismét sok esetben eleve értelmetlen maga a kérdésfelvetés is, hogy „állandó” jellemzője-e az előtagban megnevezett tulajdonság az utótagnak (pl. ilyen szerintünk a *kezdő + mondat*), így tetszés szerint kreálhatunk mondvacsinált érveket akár az egybe-, akár a különírás mellett. Mindezen érvekre tekintettel úgy véljük, a nyelvhasználók szemszögéből nézve nem durva túlzás gyakorlatilag eldönthetetlen helyesírású alakokról beszélni ebben az összefüggésben.

Táblázatunk 1. és 2. oszlopában az ingadozó leírású párt (a morfológiai szót mint típust) és előfordulásainak számát adjuk meg, 3. oszlopában pedig olyan hozzájuk kapcsolódó különírt szókapcsolatokat, illetve egybeírt összetételeket említünk példaként, amelyek a helyesírású szótárakban megtalálhatóak, de megítélésünk szerint nem segítenek egyértelműen annak eldöntésében, hogy az ingadozó pár helyesen külön- vagy egybeírandó-e.

**7. táblázat**  
Eldönthetetlenként besorolt alakok

Ingadozó pár	Előfordulások száma (külön / egybe)	Analóg példák
<i>kutató cég – kutatócég</i>	275 > 153	<i>kutató tudós – kutatóút, kutatóhely</i>
<i>záró napon – zárónapon</i>	48 < 135	<i>záró rendelkezés – zárónyilatkozat, záróár</i>
<i>folyó mérleg – folyómérleg</i>	121 < 342	<i>folyó év, folyó kiadás, folyó szöveg –folyóírás, folyószám</i>
<i>nyomozó iroda – nyomozóiroda</i>	69 < 200	<i>nyomozó hatóság, nyomozó szerv –nyomozóhivatal</i>
<i>záró értéke – záróértéke</i>	59 < 94	<i>záró rendelkezés – zárónyilatkozat, záróár</i>
<i>támadó részleg – támadórészleg</i>	47 < 60	<i>támadó fél, támadó jelleg – támadóállás, támadósor</i>
<i>ösztönző csomagot – ösztönzőcsomagot</i>	24 > 9	<i>ösztönző hatás – megszorítócsomag</i>
<i>tárgyaló delegációjának – tárgyalódelegációjának</i>	14 < 45	<i>tárgyaló fél, tárgyaló küldöttség –tárgyalóasztal</i>
<i>fertőző osztályára – fertőzőosztályára</i>	18 > 9	<i>uralkodó osztály, fertőző beteg – idegosztály, röntgenosztály</i>

**11.** Az elvégzett vizsgálatunk esettanulmányának tekinthető, amely három, különböző absztrakciós szintű, de egymással szorosan összekapcsolódó kérdés-felvetésre kínál választ.

Az első, legkonkrétabb kérdés, amelyre választ kerestünk, az volt, hogy egy adott helyesírási szabálypont tartalma, megfogalmazása miként javítható az írásgyakorlatra vonatkozó tapasztalatok tükrében. Munkánk során – amikor az adatok értékelése részeként meg kellett ítélnünk nagyszámú alakról, hogy megfelel-e a normának – világossá váltak a 112. szabály alkalmazásával kapcsolatban bizonyos nehézségek, amelyek a megszövegezés módosításával, pontosításával enyhíthetők volnának:

Egyrészt szerencsés volna kerülni az „általában különírjuk”, „különösen olyankor”, „részben jelentésváltozást is mutató” kitételeket, amelyek felmerülő helyesírási kérdések eldöntésében nem nyújtanak segítséget, ugyanis a nyelvhasználaton nincs semmilyen támpontja arra vonatkozóan, hogy az ő konkrét esetében a szabály által ismertetett tendencia érvényesül-e.

Másrészt célszerű volna, ha a szabálypont a kivételek magyarázataként nem a „kialakult írásszokásra” hivatkozna, hiszen ez a felhasználók számára nem informatív. Járható megoldás volna ehelyett tételesen felsorolni az összes kivételes esetet, amely a hagyomány alapján sérti meg az általános szabályt. Amennyiben

egy ilyen felsorolás túl hosszú volna, a szabálypont átvehetné az OH. által alkalmazott megoldást, és egyértelműen utalhatna arra a tényre, hogy azon szókapcsolatok külön- és egybeírása ingadozik, amelyek elő- vagy utótagja maga is összetett szó. Ebben az esetben a szabálynak a szótárhoz kell irányítania az olvasót, a szótárral szemben pedig követelményként kell megfogalmazni, hogy tartalmazza az összes kivételes alakot. Természetesen a legjobb az lenne, ha a szabályzat eleve megszüntetné ezeket a kivételeket, amelyek miatt – a szabálypontban kiemelt példák (*forgószínpad, javítóműhely, mozgókórház, mérőműszer, előhívótank, vendéglátóipar*) alapján – külön kategóriát fenntartani nem érdemes.

Továbbá úgy találtuk, hogy a szabálypont megfogalmazásának módja más tekintetben sem eléggé felhasználóbarát. Mivel a helyesírási szabályzat célközönségét nem elsősorban nyelvész szakemberek alkotják, a leíró és az előíró szemléletnek a 112. szabályban is tetten érhető keveredése kifejezetten zavaró lehet. Ugyan egy helyesírási szabálynak természeténél fogva előírónak kell lennie, az „általában különírjuk”, „különösen olyankor” típusú fordulatok szemmel láthatóan leíró jellegű kommentárként értelmezendők. Más szóval ezek (bár így is érthetőek) nem a nyelvhasználóra kívánják bízni a külön-, illetve egybeírás közötti döntést, hanem – némileg sarkítottan fogalmazva – arról szólnak, hogy a szótár szerint többnyire hogyan kell írni az adott kapcsolatokat. Ez helyesírás-kutatók számára talán érdekes lehet, konkrét helyesírási problémák megoldása szempontjából azonban teljesen haszontalan. A nyelvhasználót nem érdekli, hogy a külön- vagy az egybeírás a gyakoribb általában, hanem csak arra kíváncsi, hogy az általa éppen leírni kívánt szó helyesen hogyan írandó. Szerencsésebb volna tehát a szabályban következetesen preskriptív szemléletet alkalmazni, és olyan stratégiákat bemutatni a nem nyelvész olvasó számára, amelyek önkényes kivételek nélkül, kontextusba ágyazott életszerű példákkal gazdagon illusztrálva egyértelműen megmondják számára, hogy milyen jellegű szókapcsolatokat írjon külön, illetve legybe. Például nem kielégítő minden magyarázat nélkül, egymástól viszonylag messze elhelyezve hozni egyetlen olyan példapárt, hogy „(az újoncokat) *kiképző tiszt*” és „*kiképzőtiszt* (beosztás)”, hanem helyette explicitté kellene tenni, hogy amennyiben a jelző a konkrét kontextusban bővítve van, akkor egészen biztosan a különírás választandó, vö. *újoncokat kiképző tiszt*.

Mindezeken túl problémát jelent az is, hogy a szabálypont olyan szemantikai szempontra hivatkozik – ti. „jelentés tekintetében összeforrtak (az igenév nem folyamatot, hanem valamire való képességet, rendeltetést fejez ki)” –, amelyek érvényesítésekor számolni kell azzal, hogy a nyelvhasználók eltérően ítélnék. Ha ki kívánjuk zárni a nyelvhasználók írásgyakorlata közötti ilyen jellegű váltakozást vagy ingadozást, akkor szubjektív, illetve természetüknél fogva fokozatosságot mutató szemantikai különbségtetelekre nem szabad alapozni. Ha viszont megtartjuk ezt a jelentéstani kritériumot, akkor tudomásul kell venni az intuíciók közötti eltéréseket, és következőképpen meg kell engedni, hogy a nyelvhasználók ezeket a kapcsolatokat belátásuk szerint írják külön- vagy egybe.

A második típusú kérdésfelvetés, amelyre válaszokat kerestünk, arra vonatkozott, hogy a 112. szabályponttal kapcsolatban fentebb jelzett problémák létezése

alátámasztható-e empirikus, kvantitatív adatokkal, azaz nem csupán általunk konstruált álproblémákról van-e szó. Továbbá ezen adatok alapján azt is fel kívántuk táni, hogy mely konkrét szókapcsolatok leírásánál jelentkeznek e problémák leginkább.

A 6. és 7. pontban láthattuk, hogy mely előtagok és mely szópárok leírása okoz a leggyakrabban nehézséget a nyelvhasználóknak. A 9. pontban választ kaptunk továbbá arra, hogy milyen gyakoriak abszolút értelemben azok az esetek, ahol a 112. szabály alkalmazása nyilvánvalóan nehézséget jelent a nyelvhasználóknak: kb. minden 4000. szó ilyen egy átlagos szövegben. A 10. pontban azt is megtudtuk, hogy e nehézségeket milyen gyakran sikerül a helyesírási normával összhangban megoldaniuk a nyelvhasználóknak: kb. 30%-ban hibáznak, az esetek 17%-ában pedig három – helyesírási kérdésekkel is foglalkozó, szótárakkal és más segédletekkel felszerelt – nyelvésznek sem sikerült kiderítenie, hogy mi lenne a helyesírási norma.

Nyilvánvaló, hogy ezek a számok önmagukban kevésbé informatívak. Ahhoz, hogy lássuk, a minden 4000. szó gyakori vagy ritka nyelvi jelenségnek tekinthető-e, illetve hogy a 30%-os hibázási, 17%-os eldönthetetlenségi arány sok-e vagy kevés, más helyesírási szabályok alkalmazásának hasonló rendszerszerű feltárására, majd a kapott eredmények összehasonlítására volna szükség. Úgy gondoljuk viszont, hogy ezek a számok elég magasak ahhoz, hogy ne menjünk el szó nélkül a vizsgált helyesírási kérdés mellett. Az is egyértelmű, hogy egy olyan szabály, amely az esetek 17%-ában maximális körültekintés mellett sem alkalmazható, továbbá amelyet (tokeneket tekintve) a releváns esetek 30%-ában nem követnek a nyelvhasználók (4. táblázat), illetve amely alá eső típusokat 30%-ban (5. táblázat) teljesen véletlenszerűen írnak le, az nem kielégítő.

Ugyanakkor a vizsgálat kiterjesztésével könnyedén feltárható, hogy más, szórványos, egyedi megfigyelések alapján gyakorinak vélt vagy akár a szakirodalomban, nyelvi tanácsadó anyagokban ilyenként említett helyesírási hibákhoz képest milyen súlya van a 112. pont által szabályozott jelenségeknek. Amint az 5. és 10. szakaszban megtudtuk, legalább az utóbbiak közül már az 1:4 arányban ingadozó párok is hozzávetőlegesen 200 ezerszer fordulnak elő korpuszunkban, ezek közel egyharmada, azaz 60-70 ezer alak leírása hibás. Érdemes ezzel összehasonlítani például a nyest.hu-n a félreérthető A tíz leggyakoribb helyesírási hiba címmel megjelent cikkben szereplő hibákat. Ez utóbbi egy magyartanárok körében végzett felmérés alapján a tíz leggyakoribbnak vélt helyesírási hibajelenséget ismertette. Ezek közül négy (*new yorki*, *muszály*, *higyj*, *kultúrális*) valójában mind abszolút, mind relatív értelemben jóval ritkább a 112. szabály alá eső hibahalmaznál: a hibás alakok tokenszáma 500 és 1200 közé esik, a helyes változat pedig legalább 16-szor olyan gyakori, mint a hibás (*higyj*), de a *muszály* esetében ez eléri az 1:60 arányszámot is. Az *időjárásjelentés* legalább relatív értékben gyakorinak mondható (1:1 hibaarány, de csak kb. 400-400 token); csupán az *l-e* és az *össze-vissza* bizonyult abszolút és relatív értelemben is viszonylag gyakorinak (kb. 5000, illetve 7700 hibás alak, 1:4 és 2:1 hibaarány) a 112. szabály alá eső esetekhez képest. Hasonló eredményekkel járt a magyarhelyesiras.hu oldalon gyakori hibákként feltüntetett esetek közül szűrőpróbaszerűen kiválasztott öt hiba kvanti-

tatív elemzése is: Három esetében (*mindíg, savanyu, elenfél*) a hibaarány legalább 1:180 volt, tehát ezeket szinte soha nem rontják el. Egy hiba (*örjöng*) ennél aránylag gyakoribb, de nem igazán meggyőző (kb. 500 token, 1:10 hibaarány). Egyedül a *brilliáns* bizonyult valóban problematikusnak (kb. 1000 token, 1:2 hibaarány).

Ez a rövid kitérő is jelzi, hogy az írásunkban tárgyalt megközelítésmóddal mennyire más alapokra helyezhető a helyesírási kérdések súlyának, gyakoriságának megítélése, mint ha ad hoc módon, szubjektív benyomások alapján járnánk el. Racionálisan vizsgálhatóvá válik például az is, hogy a magyar helyesírás reformjának sarkalatos kérdéseként kell-e kezelni az *ly* megőrzésének vagy eltörlésének a problémáját (ami lépten-nyomon felmerül), vagy vannak-e esetleg olyan kifejezéscsoportok, amelyek leírása ennél objektíve mérhetően jóval nagyobb nehézséget okoz a nyelvhasználóknak.

Végezetül a harmadik, legáltalánosabb kérdésfelvetés, amelyre választ kívánunk adni tanulmányunkban, az volt, hogy miként alkotható meg egy olyan elemzési módszertan, amellyel nagyméretű korpuszokon tanulmányozhatóak a helyesírás és az írásgyakorlat közti eltérések. Úgy véljük, ezek az általunk bemutatott eljárással pontosan és átfogóan felmérhetőek, és így a helyesírás-tervezést a jövőben immár valóban empirikus alapon, nagyméretű korpuszokra támaszkodva hajthatjuk végre, ezáltal pedig megalapozottabbá és hatékonyabbá tehetjük. Ha megkapjuk a választ arra a kérdésre, hogy pontosan mely jelenségek azok, amelyek különösen gyakran okoznak nehézséget a nyelvhasználóknak, akkor a szabálypontok megszövegezése vagy helyesírási tankönyvek, tanácsadó kiadványok, szótárak írásakor ezekre különös tekintettel lehetünk, vagyis e segédeszközöket úgy alakíthatjuk ki, hogy a nyelvhasználók számára hasznos információkkal szolgáljanak.

**Kulcsszók:** helyesírás, különírás, egybeírás, összetétel, folyamatos melléknévi igenév, minőségjelző.

### Hivatkozott irodalom

- AkH. 1984= 1984. *A magyar helyesírás szabályai. 11. kiadás.* Akadémiai Kiadó, Budapest.
- ANTALNÉ SZABÓ ÁGNES 2009. A különírás és az egybeírás rendszeréről. Javaslat egy helyesírási fejezet megújítására. *Magyar Nyelvőr* 133: 129–149.
- BÁRÁNY MARGIT szerk. 2009. *Akadémiai helyesírási szótár.* Akadémiai Kiadó, Budapest.
- BENYHE JÁNOS 2004. Ésszerű szabálykönyv. Több mint szótár. *Heti Válasz* 4/47: 58.
- DEME LÁSZLÓ – FÁBIÁN PÁL – TÓTH ETELKA szerk. 1999. *Magyar helyesírási szótár. A Magyar Tudományos Akadémia szabályai szerint.* Akadémiai Kiadó, Budapest.
- FÁBIÁN PÁL 1984. A magyar helyesírás szabályainak 11. kiadásáról. *Magyar Nyelvőr* 108: 385–401.
- Internetes helyesírási szótár:* <http://www.helyesiras.akkrt.hu>. (2014. 11. 21.)
- LACZKÓ KRISZTINA – MÁRTONFI ATTILA 2004. A különírás és az egybeírás. Vitaanyag. *Magyar Nyelvőr* 128: 141–57.
- [magyarhelyesiras.hu](http://www.magyarhelyesiras.hu) *Leggyakoribb hibák.* [http://www.magyarhelyesiras.hu/gyakori\\_\[-\]hibak.php](http://www.magyarhelyesiras.hu/gyakori_[-]hibak.php) (2014. 11. 21.)



- MÁRTONFI ATTILA 2011. A helyesírás mint szabályrendszer. *Café Babel* 20: 63–71.
- MISAD KATALIN 2004. Laczkó Krisztina–Mártonfi Attila: Helyesírás. Osiris Kiadó, Budapest, 2004, 1539 p. *Fórum Társadalomtudományi Szemle* 6/4: 187–192.
- NÁDASDY ÁDÁM 2004. A November 7 Erőműtől DJ Newlig. Laczkó Krisztina – Mártonfi Attila: Helyesírás. *Magyar Narancs* 16/51–52: 78–79.
- nyest.hu 2012. A tíz leggyakoribb helyesírási hiba. *Nyelv és Tudomány* 2012. szeptember 11. <http://www.nyest.hu/hirek/a-tiz-leggyakoribb-helyesirasi-hiba> (2014. 11. 21.)
- OH. = LACZKÓ KRISZTINA – MÁRTONFI ATTILA 2004. *Helyesírás*. Osiris Kiadó, Budapest.
- SZŐKE PÉTERNÉ – SZŐKE PÉTER 1987. Helyes írás – vagy helyesírás tíz körömmel? *Magyar Nyelvőr* 111: 257–275.

### On the spelling of expressions involving participles in -ó/-ő – a corpus linguistic study

A corpus linguistic study of differences between spelling practice and the norms laid down in the rules of orthography may yield a number of conclusions both for orthography research and for language planning. Within that broad field of study, a quantitative analysis of texts in large corpora is especially well suited to give answers to a number of important questions, answers that are both exact and representative with respect to the whole of written language use in Hungarian. In this paper, the authors demonstrate, in the form of an analysis of a large corpus, the usefulness of quantitative corpus-based orthography research. The case study focuses on the way a specific rule of orthography is observed or ignored: on the spelling practice concerning expressions that involve a participle in -ó/-ő, or a compound headed by such a participle, in an attributive role. With reference to paragraph 112 of the current Rules of Hungarian Orthography (1984), a method is proposed whereby differences between orthographic rules and spelling practice can be explored exactly and comprehensively. The authors think that their procedure makes it possible to carry out orthography planning on an empirical basis by using large corpora, thereby making it more well-founded and more efficient.

**Keywords:** orthography, writing as two words, writing as one word, compounds, present participle, attribute.

PETHŐ GERGELY  
Debreceni Egyetem

VERÉCZE VIKTÓRIA  
Debreceni Egyetem

FEHÉR KRISZTINA  
Debreceni Egyetem