A magyar kvázianalitikus igei toldalékok ingadozó kötőhangzójának kvantitatív elemzése*

1. Bevezetés. A magyar igeragozás egyik régóta ismert jelensége a mássalhangzó-kapcsolatra végződő igetövek és bizonyos toldalékok határán megjelenő kötőhangzó ingadozása. A *mondasz* és a *mondsz,* a *mondanak* és a *mondnak,* a *mondotok* és a *mondtok,* a *mondalak* és a *mondlak* mind élő alakok, bár elfogadottságuk és relatív gyakoriságuk tőtől és toldaléktól függően eltérő. A szóban forgó ingadozást A mai magyar nyelv rendszere (TOMPA szerk. 1961: 531) is megemlíti, elsősorban regiszterhez (beszélt/írott nyelv) és lektushoz (társalgási/irodalmi nyelv) kötve. PAPP FERENC és JÁNOSKA SÁNDOR (1967: 146) döntően az ikes ragozáshoz kötik a kötőhangzó ingadozását.

Az újabb fonológiai-morfológiai szakirodalom a jelenséget egy, az analitikus és szintetikus toldalékok osztályai között található – ún. kvázianalitikus – toldalékosztály segítségével ragadja meg (TRÓN–REBRUS 2001; REBRUS–TÖRKENCZY 2008; REBRUS 2024). A kvázianalitikus toldalékkal ellátott alakok két mássalhangzóra végződő, nem váltakozó igetövek (stabil CC tövek) után tipikusan kötőhangzósak (CC-V minta: *tart-asz*), míg egyetlen mássalhangzóra végződő igetövek (VC tövek) után tipikusan kötőhangzó nélkül jelennek meg (VC-C minta: *ápol-sz*), a két kategorikus minta között azonban széles átmeneti sáv húzódik. Ebben a sávban helyezkednek el a fenti toldalékok is: az alanyi ragozású, jelen idejű, egyes szám második személyű (PRS.NDF.2SG) *-sz (mondsz),* többes szám harmadik személyű (PRS.NDF.3PL) *-nak*/*-nek (mondnak),* többes szám második személyű (PRS.NDF.2PL) *-tok/-tek/-tök (mondtok)* és egyes számú *-lak/-lek* (PRS.2<1SG, *mondlak*). (A múlt idejű paradigma NDF.3SG-n kívül eső személyragot tartalmazó alakjai is mutatnak ilyen váltakozást, mint az *ajánlottak/ajánltak, ajánlotta/ajánlta,* de ezek a fenti példáknál jóval ritkábbak, és dolgozatunkban nem vizsgáljuk őket.) Ezeket az köti össze, hogy a tővégi CC tövekhez (*mond-, tart-, fest-, hajt-* stb.) kötőhangzóval és kötőhangzó nélkül egyaránt járulhatnak, az ingadozás mértéke és iránya azonban tövenként és toldalékonként eltérő.

A mai magyar nyelvben számos helyen jelennek meg hasonló, szabad váltakozást mutató alakok, mint amilyen a tárgyeset kötőhangzójának ingadozása (RÁCZ–REBRUS–TÓTH 2025) vagy az igetövekben található magánhangzó ingadozása és az ide kapcsolódó igei defektivitás (LUKÁCS–REBRUS–TÖRKENCZY 2009; RÁCZ–REBRUS–TÖRKENCZY 2021). Ezekben közös, hogy a nyelv fonotaktikai szabályszerűségei és a fonotaktikailag hasonló alakok viselkedése együtt befolyásolják az ingadozó alakok gyakoriságát (l. pl. SIPTÁR–TÖRKENCZY 2000). Sok helyen, mint a *mondsz/mondasz* pár esetén is, alapvető tényező a fonotaktika, vagyis az, hogy

* A szerzők szeretnék megköszönni bírálóik és kísérleti résztvevőik munkáját, és köszönettel tartoznak Seán Robertsnak.

a nyelvhasználók mennyire fogadják el a kötőhangzó nélküli alakban megjelenő mássalhangzó-kapcsolatot (például a *mondnak* alakban előforduló *ndn* kapcsolatot).

A nemzetközi hangtani szakirodalom az utóbbi két és fél évtizedben rámutatott, hogy a nyelvekben megfigyelhető fonotaktikai szabályszerűségek nem kategorikusak, hanem a nyelv szókészletének valószínűség-alapú leírásai és a nyelvhasználók képesek a szókészletre építő, finomhangolt ítéleteket hozni arról, mennyire találnak bizonyos hangsorokat (fonotaktikai konfigurációkat) elfogadhatónak. Ezek az ítéletek összefüggenek azzal, hogy a célszavak és azok részei mennyire gyakoriak a mentális lexikonban (Coleman–Pierrehumbert 1997). A mentális lexikon közvetlenül nem hozzáférhető a kutató számára, de korpusznyelvészeti eszközökkel közelíteni lehet azt, hogy a beszélt nyelvben milyen szekvenciák milyen gyakorisággal fordulnak elő, és melyek hiányoznak teljes egészében (Rácz–Papp–Hay 2016). Pierrehumbert, Hay és Beckman (1998) kifejezetten a szótaghatáron átnyúló orrhang + zörejhang (nazális-obstruens) kapcsolatokra mutatták meg, hogy a fonotaktikai ítéletek eloszlását a kapcsolatok korpuszban mutatott típusgyakorisága jobban jósolja, mint ezek előfordulási gyakorisága.

A korpusz és a fonotaktikai ítéletek közötti kapcsolat természetét két további megfigyelés árnyalja. Hayes és Wilson (2008) megmutatták, hogy pusztán pozitív bizonyítékok alapján is meg lehet tanulni egy olyan fonotaktikai nyelvtant, amely kategorikus (igen/nem) és fokozatos eloszlásokat egyaránt elfogad. Ez a tanulhatóság számunkra azzal a következménnyel jár, hogy a korpuszbeli eloszlás és a nyelvhasználók ítéletei közötti kapcsolatnak nemcsak tapasztalati (empirikus), hanem műveleti (komputációs) alapja is van. Bailey és Hahn (2001) azt találták, hogy a szóhasonlóság (lexikális szomszédság) és a fonotaktikai valószínűség (a fonotaktikai szekvenciák korpuszban talált arányai) egymástól függetlenül járulnak hozzá az ítéletekhez, és a két komponens közül a lexikális hasonlóság súlya nagyobb.

A két kérdés, amelyre cikkünkben választ keresünk, a következő. Egyrészt megjósolja-e a kvázianalitikus igei toldalékok előtti kötőhangzó megjelenésének korpuszbeli aránya azt, hogy a beszélők egy kitalált ige két alakváltozata közül melyiket választják? Másrészt ha igen, akkor a megfeleltetés mennyire szoros, és melyik szinten figyelhető meg (toldalék, tővégi mássalhangzó-kapcsolat vagy ezek együttesen)? Ezeket a kérdéseket érdemes röviden kontextusba helyezni. A fonotaktikai ítéleteket az említett munkák tipikusan közvetlen kérdésekkel vizsgálták („Elfogadható-e az ön számára ez az alak?”), mi viszont egy szóalak két változata közötti választást vizsgáljuk *(mondasz/mondsz)*. A beszélő tehát konkrét döntést hoz egy ismerős alaktani szerkezeten belül, nem pedig elvont elfogadhatósági ítéletet ad.

Mivel a feladat két alak közötti döntés *(mondasz/mondsz),* nem pedig egy alak elfogadhatóságának megítélése, a korpuszban sem alakokat vagy típusokat számolunk, hanem a hasonló létező alakok eloszlását vizsgáljuk. Ez a megoldás a Bailey és Hahn (2001) által megkülönböztetett két tényezőt – a fonotaktikai valószínűséget és a lexikális szomszédságot – nem választja külön: a korpuszbeli ingadozó igék egyszerre szolgáltatják a tővégi mintázatok eloszlásának jellemzését és a kísérletben szereplő kitalált célszavak lexikális szomszédait.

2. A kvázianalitikus kötőhangzó ingadozása a webkorpuszban. REBRUS és TÖRKENCZY (2005) terminológiája szerint a kötőhangzó ingadozása a szám/személyjelölő kvázianalitikus toldalékok osztályára jellemző. Mivel tudomásunk szerint a miénk a jelenség első átfogó kvantitatív elemzése, vizsgálatunkat négy, általuk említett, ingadozó toldalékra korlátoztuk: a *-sz (mond[a]sz), -nak/-nek (mond[a]nak), -tok/-tek/-tök (mond[o]tok)* és *-lak/-lek (mond[a]lak)* toldalékra.

A Magyar Webkorpusz 2 alapján összeállított gyakorisági listából (NEMESKEY 2020; RÁCZ 2025) kiválasztottuk a fenti négy toldalékkal ingadozást mutató CC végű igéket, és minden tő + toldalék kombinációra kiszámítottuk a kötőhangzós és a kötőhangzó nélküli alak gyakoriságának arányát. Így tizenöt olyan tővégi mássalhangzócsoportot azonosítottunk, amely a fenti négy toldalékkal előfordulva ingadozást mutat. Ezek az 1. táblázatban láthatóak. Az adatelemzést az R statisztikai környezetben végeztük, az ábrákat a ggplot2 programcsomag segítségével hoztuk létre (WICKHAM 2016; R Core Team 2025).

1. táblázat
Ingadozó tövek a webkorpuszban

Tővégi mássalhangzócsoport	Típusgyakoriság	Példák
-ng	125	*inganak/ingnak, ingasz/ingsz, rajonganak/rajongnak*
-nt	77	*jelentesz/jelentsz, jelentetek/jelenttek, jelentelek/jelentlek*
-szt	46	*választanak/választnak, választasz/választsz, választotok/választtok*
-jt	38	*nyújtanak/nyújtnak, nyújtasz/nyújtsz, gyűjtenek/gyűjtnek*
-rt	29	*tartanak/tartnak, tartasz/tartsz, tartotok/tarttok*
-lt	25	*töltenek/töltnek, töltesz/töltsz, töltelek/töltlek*
-ld	16	*küldenek/küldnek, küldesz/küldsz, küldelek/küldlek*
-nd	13	*mondanak/mondnak, mondasz/mondsz, mondotok/mondtok*
-rd	8	*hordanak/hordnak, hordasz/hordsz, hordotok/hordtok*
-zd	7	*kezdenek/kezdnek, kezdesz/kezdsz, kezdelek/kezdlek*
-mt	3	*teremtenek/teremtnek, teremtesz/teremtsz, teremtetek/teremttek*

Tővégi mássalhangzócsoport	Típusgyakoriság	Példák
-nl	3	*ajánlanak/ajánlnak, ajánlasz/ajánlsz, ajánlotok/ajánltok*
-dz (=[dz:])	2	*edzenek/edznek, edzetek/edztek*
-st	2	*festenek/festnek, festelek/festnek*
-tsz (=[ts:])	1	*metszenek/metsznek*

Az ingadozó mássalhangzócsoportok számára kiszámoltuk a Gini együtthatót, amelynek az értéke 0,65. Ez azt mutatja, hogy a mássalhangzócsoportok halmazainak mérete aránytalan, ami a fenti táblázatban is látszik: a legnagyobb, *-ng* csoportban 125 tövet találunk, a legkisebb, *-tsz* csoportban mindössze egyet. Néhány mássalhangzócsoport tehát igen gyakori, míg jelentős részük csak elvétve fordul elő az adatokban. Az elemzésünkhöz négy csoportot választottunk ki, amelyeknek a típus- és tokengyakorisága is megfelelő, tehát elég sok és elég gyakori ingadozó igei alakban találjuk meg őket. Ezek az *-ng, -nt, -szt, -jt.*

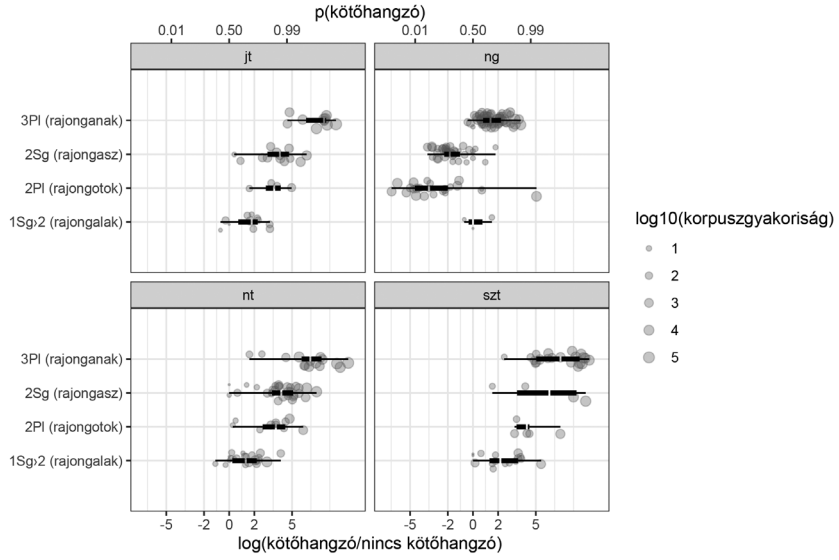
Minden tő + toldalék kombinációhoz megszámoltuk a kötőhangzós (kh.) és kötőhangzó nélküli (kh. n.) alakok előfordulásának számát, majd vettük ezek arányának (odds) a logaritmusát (log odds). A 2. táblázat erre mutat három példát, az 1. ábrán pedig a talált ingadozási eloszlások láthatóak.

2. táblázat
Arányok a webkorpuszban (példák)

Tő	Kh.	Kh. n.	Kötőhangzós gyakoriság	Kötőhangzó nélküli gyakoriság	Odds	Log odds
rajong	*rajongasz*	*rajongsz*	202	3314	0.06	-2.80
dühöng	*dühöngesz*	*dühöngsz*	27	303	0.09	-2.42
csüng	*csüngesz*	*csüngsz*	12	110	0.11	-2.22

1. ábra

A négy tővégi mássalhangzócsoport (a négy panel) a négyféle toldalékkal (függőleges tengely) a webkorpuszban[1]



Látható, hogy a kötőhangzó nélküli alakok aránya a nyelv fonetikai és fonotaktikai viszonyait tükrözi (Siptár–Törkenczy 2000; Rebrus–Szigetvári–Törkenczy 2024; Rácz–Rebrus–Tóth 2025). A könnyen veláris nazálissá redukálódó *ng* csoport vagy a degeminálódó *nt* + *t* szekvenciák esetén (a hasonuló *nd* + *t* szekvenciákat nem ideértve) arányaiban a leggyakoribb a kötőhangzó elmaradása, ezzel szemben a C*t* + nazális csoportokat tartalmazó alakokat a nyelvhasználók jellemzően kerülik.

Fontos megjegyezni, hogy a gyakorisági arányok adott esetben rendkívül kis elemszámokra épülnek. A korpuszban a 2Pl *tekintetek/tekinttek* alakok aránya 5 és 3 megfigyelés. Ehhez mérve a 3Pl *csüngenek/csüngnek* alakok aránya 1121 és 773 megfigyelés. A két aránypár hasonló (1.7 és 1.5), de az utóbbit két nagyságrenddel több megfigyelés támogatja meg, mint az előbbit.

Ezzel együtt aligha meglepő, hogy a nyelvben megfigyelhető fonotaktikai szabályszerűségek tetten érhetőek a kvázianalitikus toldalékok kötőhangzójának ingadozásában is. Dolgozatunk elsődleges kérdése azonban nem ez, hanem az, hogy a nyelvhasználók döntései tükrözik-e a korpuszbeli arányokat. Azért, hogy választ kapjunk erre a kérdésre, a korpuszban látott ingadozó tövek mintájára álszavakat alkottunk, majd ezeket megmutattuk egy online kísérlet résztvevőinek, arra kérve őket, hogy válasszanak a kötőhangzós és a kötőhangzó nélküli alak közül.

[1] A vízszintes tengely az alakváltozatok gyakoriságát adja meg, a kötőhangzós és a kötőhangzó nélküli alak arányának logaritmusa (log odds) szerint. A zárójelbe tett alakok példaként szerepelnek az adott alakváltozatokra. A pontok az egyes igealakok, a pontok mérete pedig az adott alak teljes gyakoriságát mutatja. A dobozdiagram az eloszlás mediánját és interkvartilis tartományát adja meg.

3. A kvázianalitikus kötőhangzó ingadozása kísérleti válaszokban

3.1. Ingerek. Nyolcvan álszót mint igei tövet hoztunk létre, a webkorpuszból kinyert ingadozó alakokat referenciahalmazként használva. A négy kiválasztott tővégi mássalhangzócsoport *(-ng, -nt, -jt, -szt)* egyenként hússzor szerepelt. Minden álszó a fent vizsgált négy toldalékkal jelent meg. Az álszavakat a létező szavak szótagszerkezeti elemeinek véletlenszerű összeválogatásával állítottuk elő, eltávolítva a létező szavakhoz túlságosan hasonló alakokat. A 3. táblázat példákat hoz az ingeranyagra.

3. táblázat

A kísérletben használt álszavak tővégi mássalhangzócsoport és toldalék szerint

Toldalék	Tővégi mássalhangzó-csoport	Példák
1Sg›2	*-jt*	*szüköjtelek/szüköjtlek, nyénnyejtelek/nyénnyejtlek, odojtalak/odojtlak*
1Sg›2	*-ng*	*jököngelek/jökönglek, ejjengelek/ejjenglek, kerzengelek/kerzenglek*
1Sg›2	*-nt*	*pagyontalak/pagyontlak, togyontalak/togyontlak, fedrentelek/fedrentlek*
1Sg›2	*-st*	*ornyosztalak/ornyosztlak, sziptesztelek/sziptesztlek, pollosztalak/pollosztlak*
2Pl	*-jt*	*nyuttojtotok/nyuttojttok, sajlojtotok/sajlojttok, zsömhöjtötök/zsömhöjttök*
2Pl	*-ng*	*pégengetek/pégengtek, ejjengetek/ejjengtek, szüvöngötök/szüvöngtök*
2Pl	*-nt*	*nésszentetek/nésszenttek, lahontotok/lahonttok, fűjöntötök/fűjönttök*
2Pl	*-st*	*dimbesztetek/dimbeszttek, sittyesztetek/sittyeszttek, gínesztetek/gíneszttek*
2Sg	*-jt*	*zsömhöjtesz/zsömhöjtsz, nyénnyejtesz/nyénnyejtsz, nyuttojtasz/nyuttojtsz*
2Sg	*-ng*	*csumlongasz/csumlongsz, csumlongasz/csumlongsz, pömpöngesz/pömpöngsz*
2Sg	*-nt*	*pétlentesz/pétlentsz, hárkontasz/hárkontsz, ofontasz/ofontsz*
2Sg	*-st*	*börkösztesz/börkösztsz, hürösztesz/hürösztsz, kurtosztasz/kurtosztsz*
3Pl	*-jt*	*szüköjtenek/szüköjtnek, sökköjtenek/sökköjtnek, rupojtanak/rupojtnak*
3Pl	*-ng*	*büfröngenek/büfröngnek, ejjengenek/ejjengnek, mógyonganak/mógyongnak*

Toldalék	Tővégi mássalhangzó-csoport	Példák
3Pl	*-nt*	*nésszentenek/nésszentnek, togyontanak/togyontnak, togyontanak/togyontnak*
3Pl	*-st*	*kernyesztenek/kernyesztnek, pollosztanak/pollosztnak, csélesztenek/csélesztnek*

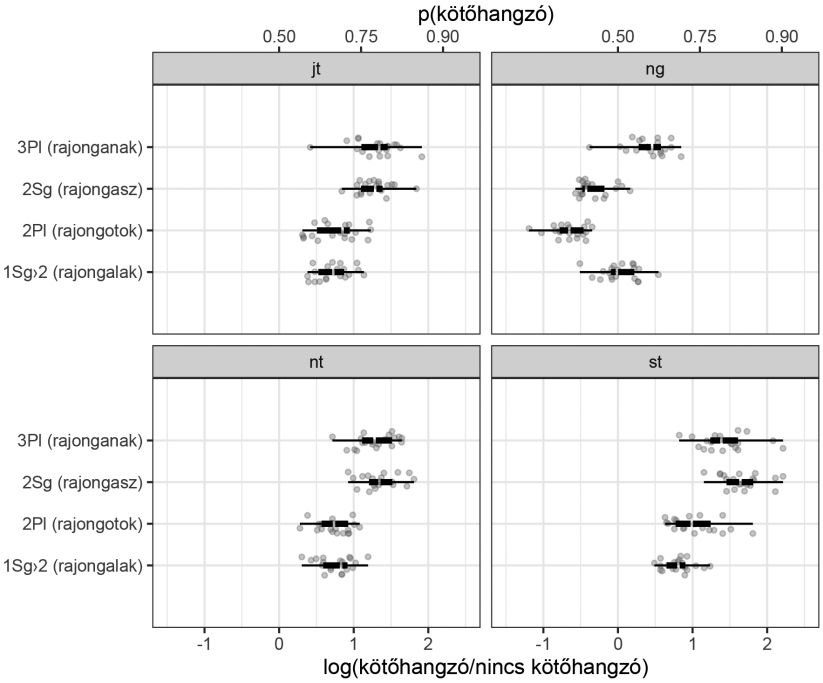
3.2. A kísérlet menete. A kísérletben a Budapesti Műszaki és Gazdaságtudományi Egyetem 76 diákja vett részt (33 férfi, 43 nő), az átlagéletkoruk 22 év. Az adatfelvétel 2025 októbere és decembere között zajlott. A résztvevők a kísérletet otthon, a számítógépükön végezték el. A kísérlet átlagosan húsz percig tartott. A kísérletet javascript nyelven, a jspsych programcsomag segítségével írtuk, és a Pavlovia kísérleti platformon futtattuk (De Leeuw 2015; Peirce et al. 2019). A kísérleti tervet az Egyesített Pszichológiai Kutatási Etikai Bizottság hagyta jóvá (ref. szám: 2021-119).

A kísérlet során a résztvevőknek írott mondatokat kellett befejezniük. Minden esetben két lehetőség között választhattak. Úgy tájékoztattuk őket, hogy mondatokba ágyazott szavakat fognak látni, amelyek közül nem mindegyik lesz ismerős számukra, és mindegyikhez két toldalékolt alak között választhatnak (pl. „Mari szívesen csumlong. Te is szívesen…?" „csumlongasz"/„csumlongsz"). A feladat az volt, hogy azt válasszák a két lehetőség közül, amelyik természetesebbnek tűnik a számukra. Minden résztvevő 320 kísérleti próbát látott: ez a szám a 80 álszói tő és a 4 toldalék kombinációjából áll össze. Megelőzendő azt, hogy a résztvevők figyelme elkalandozzon a kísérlet során, nyolc ellenőrző kérdés is szerepelt, amelyekre egyértelmű a válasz (pl. „Milyen nap jön a szerda után?" „csütörtök"/„hétfő"). Minden résztvevő véletlenszerű sorrendben látta az összes ingert, egymás után.

3.3. Eredmények. Az ingerek szerkezetét és a résztvevők kizárási kritériumait előregisztráltuk (https://aspredicted.org/r3ha7x.pdf). Az itt bemutatott, a kísérleti adatokra illesztett modell eltér az előregisztrációtól. A résztvevők közül eltávolítottuk azokat, akik a nyolc ellenőrző kérdés közül több mint kettőt elrontottak. A válaszok közül eltávolítottuk azokat, amelyeknek a reakcióideje szokatlanul messze esett a többi választól: konkrétan azokat, amelyek reakcióideje a válaszidők mediánjától a tipikus eltérés (az ún. medián abszolút eltérés) háromszorosánál is jobban különbözött. Így 1/76 résztvevőt és 917/24320 választ (3,8%) távolítottunk el az adatokból. A lenti elemzés a megmaradt adatokra vonatkozik.

A 2. ábra megmutatja, milyen arányban választották a résztvevők a kötőhangzós, illetve a kötőhangzó nélküli alakot a különféle tővégi mássalhangzók és toldalékok esetén.

2. ábra
A négy tővégi mássalhangzócsoport (a négy panel) a négyféle toldalékkal (függőleges tengely) a kísérletben[2]



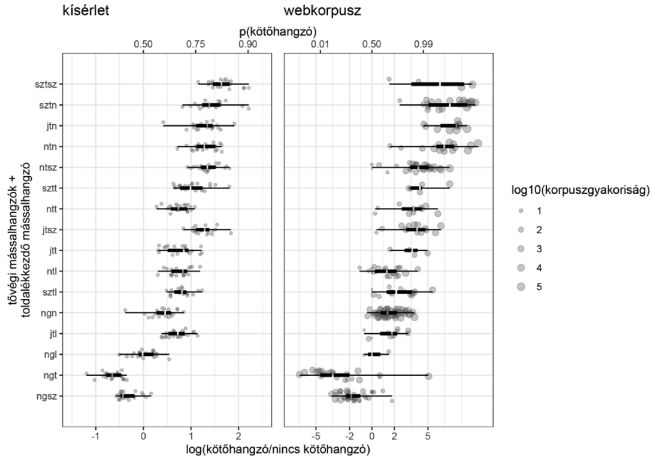
Az 1. ábrával való hasonlóság látványos. Például – ahogyan a webkorpuszban is láttuk – a résztvevők is az *-ng* végű tövek esetén voltak leginkább hajlamosak kiválasztani a kötőhangzó nélküli toldalékos alakot. Valószínűsíthető, hogy a résztvevők a döntéseikben ugyanazokat az implicit fonotaktikai szabályszerűségeket tartották szem előtt, amelyeket a webkorpuszból is vissza tudunk fejteni.

Ez a viszony egyértelműbbé válik a 3. ábrán. Ez az ábra nem a tővégi mássalhangzókat és toldalékokat mutatja meg, hanem azokat a mássalhangzócsoportokat, amelyeket a toldalékolt alakban kapnánk akkor, ha abban nem jelenne meg a kötőhangzó (példa: *mondsz* → *-ndsz* = [nʦ]). Ezeket az ábrán aszerint állítottuk sorrendbe, hogy a korpuszban a kötőhangzó milyen gyakran marad el. Ez látható a bal oldali panelon. A jobb oldali panel megmutatja ugyanezeket a szekvenciákat a kísérletben látható alakok alapján, és azt jeleníti meg, hogy a résztvevők milyen szekvenciák esetében milyen mértékben hagyták el a kötőhangzót (példa: *fűjöntsz* → *-ntsz* = [nʦ]). Látható, hogy a kísérletben kapott arányok nagymértékben együtt állnak a korpuszban látott eloszlásokkal. Ez azt mutatja, hogy a résztvevők aprólékosan le tudják képezni a beszélt nyelvben megtalálható olyan fonotaktikai eloszlásokat, amelyeket a korpusz is megragad.

[2] A vízszintes tengely az alakváltozatokra adott válaszok arányát adja meg. A pontok az egyes igealakok, a dobozdiagram az eloszlás mediánját és interkvartilis tartományát adja meg.

3. ábra

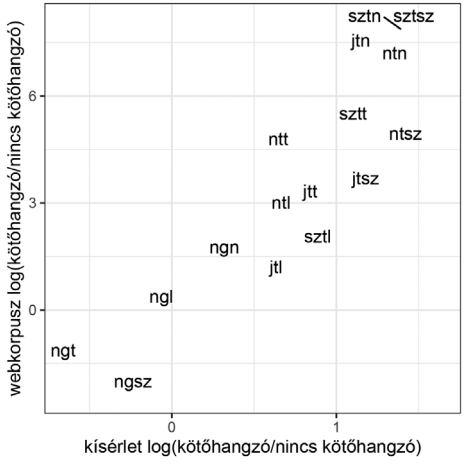
A négy tővégi mássalhangzócsoport és a négyféle toldalék által létrehozott mássalhangzócsoport (függőleges tengely) a webkorpuszban és a kísérletben[3]



Jobban megérthetjük a webkorpuszban látott eloszlások és a kísérleti válaszok közötti viszonyt akkor, ha közvetlenül összehasonlítjuk egymással a fenti eloszlások mediánjait. Ezt mutatja a 4. ábra.

4. ábra

A tővégi mássalhangzó-szekvencia + toldalékkezdő mássalhangzó kombinációk előfordulása kötőhangzóval és anélkül a webkorpuszban (függőleges tengely) és a kísérleti válaszokban (vízszintes tengely)



[3] A vízszintes tengely az alakváltozatokra adott válaszok arányát adja meg. A pontok az egyes igealakok, a dobozdiagram az eloszlás mediánját és interkvartilis tartományát adja meg. A jobboldali panel az 1. ábrához hasonlóan a megfigyelések méretével jelzi az igealakok gyakoriságát.

A webkorpuszban és a kísérletben számolt arányok közötti viszonyra számolt Spearman $\rho = 0{,}94$ ($p < 0{,}01$), ami jól rámutat arra, hogyan idézi fel a webkorpuszban látott arányokat a kísérlet.

4. Összefoglaló. Cikkünkben a magyar igeragozás kvázianalitikus toldalékai *(-sz, -nak/-nek, -tok/-tek/-tök, -lak/-lek)* előtti kötőhangzó ingadozását vizsgáltuk. Először a Magyar Webkorpusz 2 anyagából kinyertük a négy toldalékkal ingadozó CC végű igéket, és kiszámoltuk a kötőhangzós és kötőhangzó nélküli alak arányát minden tő-toldalék kombinációra. Ezután egy 76 fő részvételével zajló online kísérletben derítettük ki, hogy a résztvevők hogyan választanak a létező alakokról mintázott kitalált igék alakváltozatai között. A négy kiválasztott tővégi mássalhangzócsoporton *(-ng, -nt, -jt, -szt)* nézve a korpuszbeli és a kísérleti arányok között a tő-toldalék metszeteken számolt Spearman-rangkorreláció $\rho = 0{,}94$ ($p < 0{,}01$) lett. A beszélők tehát korábban nem látott, számukra ismeretlen igealakoknál is rekonstruálják azt az eloszlást, amely a webkorpuszban megfigyelhető.

Eredményünk jól egybevág a valószínűségi fonotaktika irodalmában található megállapításokkal (Coleman–Pierrehumbert 1997; Bailey–Hahn 2001; Hayes–Wilson 2008): a nyelvhasználók ítéletei fokozatosak, és a fokozatosság a megfigyelhető nyelvhasználat eloszlásával áll együtt. A jelen vizsgálat ezt egy olyan jelenségcsoporton mutatja meg, amelyet a magyar fonológiai-morfológiai irodalom régóta leír (Tompa szerk. 1961: 531; Papp–Jánoska 1967: 146; Trón–Rebrus 2001; Rebrus 2024: 67), de eddig részletes kvantitatív vizsgálatnak nem vetettek alá. A fonotaktikai motiváció jól megmutatkozik: az *-ng* + *-sz* kapcsolat, amelyben az *-ng* könnyen veláris nazálissá redukálódik, és a *-nt* + *-t* kapcsolat, amely degeminálódik, mindkét adatforrásban kevésbé részesíti előnyben a kötőhangzós alakot, mint a stabilabb *-jt* vagy *-szt* végződésű tövek. A nyelvhasználók viselkedése mutatja ezeket a hatásokat.

Annak a kérdésnek az eldöntése, hogy a nyelvhasználók valójában a szavak eloszlásait tanulták meg, vagy egy közvetítő reprezentációt használnak (fonotaktikai nyelvtan, lexikális szomszédság, illetve a kettő keveréke), a jelen dolgozat keretein kívül esik. A kísérletben látott és a korpuszban megfigyelt hangkapcsolatok közvetlen összevetése több szempontból is nehézségbe ütközik:

1. A CC tővég és a kötőhangzó nélküli toldalék tipikusan hármas mássalhangzó-kapcsolatokat (CCC) állít elő, amelyek jelöltsége a C1 és a C3 minőségétől függően több dimenzióban változik, azaz nem adható meg lineáris jelöltségi sorrend a létrejövő kapcsolatokra.

2. Egyes esetekben (pl. *Ct* + *-tOk, Ct* + *-sz*) hasonulás és/vagy degemináció játszódik le, ami nem hármas, hanem kettes kapcsolatokat (CC, pl. [jt], [nt], [st], ill. [jʦ], [nʦ], [sʦ]) eredményez a felszínen, melyek tipikusan (de nem kizárólagosan) jelöletlenebbek a CCC-knél.

3. A kapcsolatok jelöltsége azok fonológiai környezetétől is függ. A négy választott toldalékból háromban magánhangzó is van, itt a létrejövő CC(C) magánhangzóközi helyzetbe kerül *(-nAk, -lAk, -tOk),* de az *-sz* toldalék nem ilyen: kötőhangzó nélküli változata szóvégi kapcsolatot eredményez.

Ezek az egymásnak gyakran ellentmondó tényezők megnehezítik az adatoknak a fonotaktikai jelöltség segítségével való értelmezését. Általánosan annyi mondható, hogy a leggyakrabban létrejövő kapcsolatok ([ŋks], [ŋkt], [nt], [jt], [st], [nʦ], [ŋgl], [ntl]) jelöletlenebbek, a magyarban morfémán belül is előfordulnak, míg a legritkábbak ([sʦ], [ntn], [jtn], [ntn], [stn]) tipikusan jelöltek, és nem szerepelnek morfémán belül (l. Törkenczy 2017: 296). A 4. ábrában bemutatott arányok azonban már nem teljesen követik a jelöltséget: a legjelöletlenebb kettes kapcsolatok ritkábbak, mint az *-ng* végű tövek alakjaival létrejött CCC-k ([ŋks], [ŋkt], [ŋgl], [ŋgn]), ami nem magyarázható a felszíni fonotaktikával. A jelenség a CC-ket előállító degeminációnak a tő visszaállíthatóságát megnehezítő jellegével lehet kapcsolatos (l. erről Papp 1968: 203); ebben a témában további vizsgálat szükséges.

Az eredményeinket jól megragadó modellek mindenesetre biztosan probabilisztikus természetűek lesznek, nem kategorikusak: a megfelelés fokozatossága ezt kikényszeríti.

Külön említést érdemel, hogy a kísérleti tartomány lényegesen szűkebb, mint a korpuszé. A 4. ábrán a kísérleti log-arányok a -0,5 és +1,5 között szóródnak, míg a korpuszban -5 és +6 között. Az értékek rangsora együtt mozog, de a kísérleti hatás amplitúdója kisebb. Ez összhangban van Bailey és Hahn (2001) megfigyelésével, amely szerint a fonotaktikai valószínűség és a lexikális szomszédság egymástól függetlenül járulnak hozzá az ítéletekhez: a kitalált igék a fonotaktikai szerkezetet teljes egészében, a teljes lexikális szomszédság szerepét viszont csak részben tükrözik. Az eredmény így nem azt mutatja, hogy a beszélők viselkedését kizárólag a fonotaktikai eloszlások befolyásolják, hanem azt, hogy a használatban megfigyelhető arányoknak még kitalált szavakra is van mérhető hatásuk.

5. A vizsgálat korlátai és további irányai. A négy kiválasztott tővégi mássalhangzócsoport *(-ng, -nt, -jt, -szt)* az 1. táblázatban felsorolt tizenötnek csak egy szelete. Kiválasztásukat a típus- és tokengyakoriság indokolta, de ez azt is jelenti, hogy a ritkább csoportokról *(-ld, -nd, -rd, -zd, -mt, -nl, -dz, -tsz)* a jelen adatok alapján nem tudunk nyilatkozni: a jelenség címerállatának számító *mondasz/mondsz* váltakozás is ide tartozik. Például a *-tsz* különösen érdekes lenne, mert ott a kötelező degemináció miatt a felszíni alakok szerkezete eltér a többi csoporttól (Rebrus 2024: 112). Az alacsony tokengyakoriságok a vizsgált szegmensben is befolyásolhatják a típusok eloszlásait, ahogyan azt az 1. ábrán is láthattuk. Ez az észrevétel kapcsolódik a korrelációs elemzés lenti kritikájához is.

Ahogyan a 4. pontban is említettük, a kísérleti válaszok tartománya a korpuszhoz képest tömörített (l. a 4. ábrát). Ebből az következik, hogy a hatás iránya megmarad, de a nagysága zsugorodik. Ennek két lehetséges oka van: a kitalált szavak gyengébb lexikális támogatása (Bailey–Hahn 2001) és a kétválasztásos paradigma 0/1 jellege, amely a végpontokat ritkán éri el. A két ok szétválasztása további vizsgálatot kíván, például egy folytonos elfogadhatósági skálán mért válaszokkal.

A megfeleltetést a tő-toldalék metszetek mediánjain mérjük (n = 16), nem alakok vagy akár egyes válaszok szintjén. Ez a Spearman-korrelációt értelmessé teszi, de a metszeteken belüli szóródást, amely az 1. és 2. ábrán látható, nem foglalja bele a hozzá illesztett mérőszámba. Egy általánosított lineáris kevert modell

– válasz ~ korpusz-log-arány + (1|résztvevő) + (1|tő) – több információt tudna kinyerni az adatokból. Az előregisztráció eredetileg ezt írta elő, de az eljárás egyszerűsége kedvéért és a metszetek alacsony száma miatt a leíró elemzés mellett döntöttünk.

A résztvevők egyetlen, viszonylag homogén csoportból kerültek ki (BME-diákok, medián: 22 év). A kötőhangzós alakok eloszlása szociolingvisztikai változó, és életkor, regiszter, valamint dialektus szerint változhat. A jelen eredmények egy fiatal városi populációra vonatkoznak, és aligha általánosíthatóak a magyarul beszélők egészére.

Egy bírálónk hívta fel a figyelmet arra, hogy érdekes volna a jövőben egy másik, itt nem vizsgált tővégi mintázatot is vizsgálat alá vonni: az *-ít* képzős igetöveket *(tanítalak/tanítlak),* valamint a *hűt, fűt* típusú töveket, amelyeknél a kötőhangzó hasonlóan ingadozik. Ezekben az esetekben a tővégen nem mássalhangzó-kapcsolatot, hanem egyetlen mássalhangzót találunk, így a fonotaktikai kérdésfeltevés is másképp merül fel, mint a jelen dolgozatban vizsgált CC tövek esetén. Ha az *-ít* tövek korpuszos vagy kísérleti eloszlása hasonló a CC tövekéhez, az inkább a lexikális, illetve mintaalapú meghatározottság mellett szólna.

A kutatáshoz tartozó adatok és programkód a https://doi.org/10.5281/zenodo.22681904 linken érhetőek el.

Kulcsszók: alaktan, ingadozás, magyar, kísérlet, pszicholingvisztika.

Hivatkozott irodalom

Bailey, Todd M. – Hahn, Ulrike 2001. Determinants of wordlikeness. Phonotactics or lexical neighborhoods? *Journal of Memory and Language* 44: 568–591. ttps://doi.org/10.1006/jmla.2000.2756

Coleman, John – Pierrehumbert, Janet 1997. Stochastic phonological grammars and acceptability. In: Coleman, John ed., *Computational phonology. Third meeting of the ACL special interest group in computational phonology*. ACL, Madrid. 49–56.

Hayes, Bruce – Wilson, Colin 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39: 379–440. https://doi.org/10.1162/ling.2008.39.3.379

De Leeuw, Joshua R. 2015. jsPsych. A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods* 47: 1–12. https://doi.org/10.3758/s13428-014-0458-y

Lukács Ágnes – Rebrus Péter – Törkenczy Miklós 2009. A defektív igék kísérleti vizsgálata. In: Maleczki Márta – Németh T. Enikő szerk., *A mai magyar nyelv leírásának újabb módszerei VII. Modellek, elméletek és elvek érvényessége nyelvi adatok tükrében. Szeged, 2007. október 25–26.* Szegedi Tudományegyetem Általános Nyelvészeti Tanszék, Szeged. 141–152.

Nemeskey, Dávid Márk 2020. *Natural Language Processing methods for Language Modeling.* PhD Thesis, Eötvös Loránd University, Budapest.

Papp Ferenc 1968. Adatok tőszókincsünk etimológiai rétegeinek végződés szerinti eloszlásához. *Nyelvtudományi Közlemények* 70: 200–209.

Papp Ferenc – Jánoska Sándor 1967. A Magyar Szóvégmutató Szótár és az alaktani ingadozások vizsgálata. *Magyar Nyelv* 63: 138–148.

Peirce, Jonathan – Gray, Jeremy R. – Simpson, Sol – MacAskill, Michael – Höchenberger, Richard – Sogo, Hiroyuki – Kastman, Erik – Lindeløv, Jonas Kristoffer 2019. PsychoPy2. Experiments in behavior made easy. *Behavior Research Methods* 51: 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pierrehumbert, Janet – Hay, Jennifer – Beckman, Mary 1998. Speech perception, wellformedness & lexical frequency. In: Local, John – Ogden, Richard – Temple, Rosalind eds., *Papers in Laboratory Phonology* 6. Cambridge University Press, Cambridge. 58–74.

R Core Team 2025. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing. https://www.R-project.org/ (2026. 09. 12.)

Rácz, Péter 2025. *Word frequency list from the Hungarian Webcorpus.* Verzió 1.0. Zenodo, released. https://doi.org/10.5281/zenodo.17508385 (2026. 09. 12.)

Rácz, Péter – Papp, Viktória – Hay, Jennifer 2016. Frequency and corpora. In: Hippisley, Andrew – Stump, Gregory eds., *The Cambridge Handbook of Morphology*. Cambridge Handbooks of Linguistics. Cambridge University Press, Cambridge. 685–704. https://doi.org/10.1017/9781139814720.024

Rácz Péter – Rebrus Péter – Tóth Szilárd 2025. A magyar tárgyeset ingadozó kötőhangzójának kvantitatív elemzése. *Nyelvtudományi Közlemények* 121: 131–143. https://doi.org/10.15776/NyK.2025.121.7

Rácz, Péter – Rebrus, Péter – Törkenczy, Miklós 2021. Attractors of variation in Hungarian inflectional morphology. *Corpus Linguistics and Linguistic Theory* 17: 287–317. https://doi.org/10.1515/cllt-2018-0014

Rebrus Péter 2024. *A fonológiai mintázatok paradigmatikus alapjai.* Akadémiai doktori értekezés. Kézirat. Eötvös Loránd Tudományegyetem, Budapest. https://real-d.mtak.hu/1493 (2026. 09. 12.)

Rebrus, Péter – Szigetvári, Péter – Törkenczy, Miklós 2024. No lowering, only paradigms. A paradigm-based account of linking vowels in Hungarian. *Acta Linguistica Academica* 71: 137–170. https://doi.org/10.1556/2062.2023.00674

Rebrus, Péter – Törkenczy, Miklós 2005. Uniformity and Contrast in the Hungarian Verbal Paradigm. In: Downing, Laura J. eds., *Paradigms in Phonological Theory*. Oxford University Press, Oxford. 263–295. https://doi.org/10.1093/acprof:oso/9780199267712.003.0010

Rebrus Péter – Törkenczy Miklós 2008. Morfofonológia és a lexikon. In: Kiefer Ferenc szerk., *Strukturális magyar nyelvtan 4. A szótár szerkezete.* Akadémiai Kiadó, Budapest. 683–786.

Siptár, Péter – Törkenczy, Miklós 2000. *The phonology of Hungarian.* Oxford University Press, Oxford. https://doi.org/10.1093/oso/9780198238416.001.0001

Tompa József szerk. 1961. *A mai magyar nyelv rendszere.* Akadémiai Kiadó, Budapest.

Törkenczy Miklós 2017. A szótag. In: Kiefer Ferenc szerk., *Strukturális magyar nyelvtan 2. Fonológia*. Akadémiai Kiadó, Budapest. 273–392.

Trón, Viktor – Rebrus, Péter 2001. Morphophonology and the hierarchical lexicon. *Acta Linguistica Hungarica* 48: 101–136. https://doi.org/10.1023/A:1015643405324

WICKHAM, HADLEY 2016. *ggplot2. Elegant Graphics for Data Analysis*. Use R! Springer, New York. https://doi.org/10.1007/978-3-319-24277-4

A quantitative analysis of the variable linking vowel of Hungarian quasi-analytic verbal suffixes

This paper presents a quantitative study of linking-vowel alternation before four Hungarian quasi-analytic verbal suffixes: 2SG.INDEF *-sz* (*rajongasz/rajongsz* 'you adore'), 3PL.INDEF *-nak/-nek* (*rajonganak/rajongnak* 'they adore'), 2PL.INDEF *-tok/-tek/-tök,* (*rajongotok/rajongtok* 'you-pl. adore') and 1SG>2 *-lak/-lek* (*rajongalak/rajonglak* 'I adore you'). These suffixes attach to stems that end in a consonant cluster either with or without a linking vowel, with considerable variation in the rate of each variant across stems and suffixes. Using frequency data from Hungarian Webcorpus 2, we identified the stem-final consonant clusters showing the most robust variation *(-ng, -nt, -jt, -szt)* and calculated the corpus log-odds of the vowel-final to the vowelless variant for each stem-suffix combination. We then tested whether these corpus-based proportions predict speakers' choices for novel, nonce verb stems in a two-alternative forced-choice web experiment (76 participants). Corpus and experimental proportions correlated strongly (Spearman's $\rho = 0.94$, $p < 0.01$) at the level of stem-cluster–by–suffix cells, though the experimental effect was considerably compressed relative to the corpus range. The results contribute to the literature on probabilistic phonotactics by discussing a case of morphologically conditioned variation, showing that speakers generalise corpus-attested distributional patterns to unfamiliar word forms even within a single, well-known inflectional paradigm.

Keywords: morphology, variation, Hungarian, experiment, psycholinguistics.

RÁCZ PÉTER
Budapesti Műszaki és Gazdaságtudományi Egyetem

REBRUS PÉTER
ELTE Nyelvtudományi Kutatóközpont