

Közzététel: 2021. június 15.

A tanulmány címe:

**Egy ritkábban használt survey-kísérleti elrendezés,
az alanyokon belüli ismételt mérés egy nyelvészeti kutatás példáján**

Szerző:

NÉMETH RENÁTA,
az Eötvös Loránd Tudományegyetem intézetvezetője,
az Empirikus Tanulmányok Intézetének társvezetője
E-mail: nemeth.renata@tat.k.elte.hu

DOI: <https://doi.org/10.20311/stat2021.6.hu0495>

Az alábbi feltételek érvényesek minden, a Központi Statisztikai Hivatal (a továbbiakban: KSH) Statisztikai Szemle c. folyóiratában (a továbbiakban: Folyóirat) megjelenő tanulmányra. Felhasználó a tanulmány vagy annak részei felhasználásával egyidejűleg tudomásul veszi a jelen dokumentumban foglalt felhasználási feltételeket, és azokat magára nézve kötelezőnek fogadja el. Tudomásul veszi, hogy a jelen feltételek megszegéséből eredő valamennyi kárért felelősséggel tartozik.

1. A jogszabályi tartalom kivételével a tanulmányok a szerzői jogról szóló 1999. évi LXXVI. törvény (Sztj.) szerint szerzői műnek minősülnek. A szerzői jog jogosultja a KSH.
2. A KSH földrajzi és időbeli korlátozás nélküli, nem kizárólagos, nem átadható, térítésmentes felhasználási jogot biztosít a Felhasználó részére a tanulmány vonatkozásában.
3. A felhasználási jog keretében a Felhasználó jogosult a tanulmány:
 - a) oktatási és kutatási célú felhasználására (nyilvánosságra hozatalára és továbbítására a 4. pontban foglalt kivétellel) a Folyóirat és a szerző(k) feltüntetésével;
 - b) tartalmáról összefoglaló készítésére az írott és az elektronikus médiában a Folyóirat és a szerző(k) feltüntetésével;
 - c) részletének idézésére – az átvevő mű jellege és célja által indokolt terjedelemben és az eredetihez híven – a forrás, valamint az ott megjelölt szerző(k) megnevezésével.
4. A Felhasználó nem jogosult a tanulmány továbbértékesítésére, haszonszerzési célú felhasználására. Ez a korlátozás nem érinti a tanulmány felhasználásával előállított, de az Sztj. szerint önálló szerzői műnek minősülő mű ilyen célú felhasználását.
5. A tanulmány átdolgozása, újra publikálása tilos.
6. A 3. a)–c.) pontban foglaltak alapján a Folyóiratot és a szerző(ke)t az alábbiak szerint kell feltüntetni:

„*Forrás: Statisztikai Szemle c. folyóirat 99. évfolyam 6. számában megjelent, Németh Renáta által írt, 'Egy ritkábban használt survey-kísérleti elrendezés, az alanyokon belüli ismételt mérés egy nyelvészeti kutatás példáján'* című tanulmány (link csatolása)”

7. A Folyóiratban megjelenő tanulmányok kutatói véleményeket tükröznek, amelyek nem esnek szükségképpen egybe a KSH vagy a szerzők által képviselt intézmények hivatalos álláspontjával.

Németh Renáta

Egy ritkábban használt survey-kísérleti elrendezés, az alanyokon belüli ismételt mérés egy nyelvészeti kutatás példáján*

**A less commonly used survey experimental design:
the within-subject design on the example
of a linguistic research project**

NÉMETH RENÁTA,
az Eötvös Loránd Tudományegyetem intézetvezetője,
az Empirikus Tanulmányok Intézetének társvezetője
E-mail: nemeth.renata@tatk.elte.hu

A társadalomtudományi felvételekben az utóbbi évtizedben tért hódított az erősebb oksági bizonyítékot szolgáltató kísérleti megközelítés. Népszerűségét még inkább növeli az online platformok elérhetőségéből fakadó könnyű kivitelezhetőség. Ugyanakkor az alanyok közötti (*between-subjects*) elrendezés sokkal elterjedtebb, mint az alanyokon belüli (*within-subjects*), annak ellenére, hogy az utóbbi, bár nem mindig megvalósítható, olcsóbb és statisztikailag hatékonyabb. A tanulmány először röviden összeveti a két megközelítést, és összefoglalja a társadalomtudományi alkalmazhatóságukat, majd, mivel az alanyokon belüli elrendezés vizsgálata statisztikailag nem triviális, bemutatja az ahhoz illeszkedő többszintű keresztezett randomhatásmodell használatát egy hangszimbolikai kutatás példáján. Az elemzés során a szerző kitér az alanyokon belüli elrendezés potenciális gyengeségeinek, a külső, illetve a konstrukciós érvényesség korlátjainak vizsgálati lehetőségére is.

TÁRGYSZÓ: alanyokon belüli elrendezés, többszintű keresztezett randomhatásmodell, survey-kísérlet

In social science surveys, the experimental approach has become widespread in the last decade. It provides stronger causal evidence than observational research, and its popularity is further enhanced by the ease of implementation due to the availability of online platforms. The between-subjects design is much more common than the within-subjects design, while the latter, although

* A szerző köszönetét fejezi ki Kovács Lászlónak, az Eötvös Loránd Tudományegyetem Társadalomtudományi Kar (Savaria Egyetemi Központ) docensének, hogy a nyelvészeti esettanulmányban elemzett adatbázist a rendelkezésére bocsátotta.

not always feasible, is cheaper and statistically more efficient. This study briefly compares the two approaches and summarizes their application possibilities in social sciences. As the analysis of a within-subjects design is statistically not trivial, the application of the appropriate multilevel crossed random-effects model is presented taking a sound symbolism research project as an example. The author also explains how to examine potential weaknesses of within-subject designs regarding external validity and construction validity.

KEYWORD: within-subject design, multilevel random-effects model, survey experiment

Az utóbbi évtizedben a társadalomtudományi kutatásokban nagymértékben megnövekedett az erősebb oksági bizonyítékot szolgáltató kísérleti vizsgálatok, azon belül is a *survey-kísérletek* népszerűsége (egy gondosan kivitelezett hazai szociológiai példát nyújt erre *Janky* [2019] tanulmánya). A Google Scholar tudományos-publikáció-kereső oldalon a „survey experiment” (survey-kísérlet) keresőszóra kiadott találatok számát tekintve drámai emelkedést tapasztalhatunk, ha azt 2000 és 2020 között, ötéves szakaszonként vizsgáljuk: 810, 2 060, 5 180, 14 900. A kísérleti megközelítés gyors politikatudományi terjedése kapcsán *Mutz* és *Kim* ([2020] 1. old.) megjegyzik: „Fiatalabb kollégáink számára meglepő lehet, hogy nem is olyan régen a politikatudományban a kísérleteket teljesen alkalmatlan módszernek tartották, egyedül a való világban végzett terepkísérletek jelentettek kivételt.”

A társadalomtudományokban tradicionálisan használt megfigyeléses vizsgálat hátránya a randomizált kísérlettel szemben, hogy az *okszági kapcsolat* kimutatása szempontjából kisebb bizonyító erővel rendelkezik (*Németh* [2015], *Freedman–Pisani–Purves* [2005] 1–2. fejezet). Ennek a „korrelációból nem következik okság” jelenségnek a közismert oka, hogy felvétel esetén a „kezelés” kiosztását nem tudjuk kézben tartani, így fellép az összemosódás (*confounding*) lehetősége. Példaként: ha a médiafogyasztás bizonyos attitűdökre gyakorolt hatását szeretnénk mérni, a megfigyelt korrelációk mögött nem feltétlenül áll oksági kapcsolat, hiszen számtalan olyan tényező van, amely mindkét változót befolyásolja (politikai preferencia, nem, iskolázottság stb.). A kísérleti megközelítés előnye ezzel szemben, hogy a kutató olyan absztrakt környezetben vizsgálhatja alanyait, ahol maga végzi a „kezelés”, a független változó különböző értékeinek hozzárendelését a vizsgálati alanyokhoz. Ilyen kísérleti elrendezést alkalmazott például *Wozniak, Drakulich* és *Calfano* [2020] a médiában megjelent rendőr-civil interakciókról készült fotók rendőrséggel kapcsolatos közvéleményre gyakorolt hatásának mérésére. Az alanyoknak három, a barátságostól az ellenségesig értékelhető interakciót ábrázoló fotó közül egyet mutattak, míg a kontrollcsoport nem kapott képet. Ezután a résztvevőket a rendőrségről alkotott véleményükről kérdezték.

Ideális esetben tehát a különböző kezeléseknél való kitettség kísérleti mérése korrekt oksági következtetést tesz lehetővé, az absztrakt körülmények ugyanakkor episztemológiai korlátokkal is járnak: a kísérlet a megfigyeléses vizsgálattal szisztematikus módon, különböző érvényességi kritériumok alapján vethető össze (Cutler [2017]). A *belső érvényesség* arra vonatkozik, hogy biztosak lehetünk-e benne, hogy a vizsgálat valóban a kérdéses kapcsolatot méri, s csakugyan oksági-e a kimutatott kapcsolat. Az előbbieknél alapján a kísérletek e dimenzió szerint jól teljesítenek. Nem így a *külső érvényesség* dimenziójában: sokszor kérdéses, hogy mennyire általánosíthatók kisebb, gyakran kényelmi mintán kapott eredmények a teljes populációra. Általában a kísérletet végzők úgy tekintik, hogy a külső érvényesség korlátozottsága megfizethető ár a belső érvényességért cserébe. Végül a *konstrukciós érvényesség*, amely az alkalmazott mérőeszköz és a mérni kívánt mögöttes fogalom megfeleltethetőségére vonatkozik. De vajon jó mérőeszköze-e az idegenellenességnek a European Social Survey (Európai Társadalmak Összehasonlító Vizsgálata) vonatkozó kérdésblokkja? Vajon tényleg a médiában közölt fotók hatását méri-e *Wozniak, Drakulich és Calfano* [2020] korábban idézett vizsgálata? A konstrukciós érvényességi kérdések a kutatási kérdések kísérleti elrendezésre való lefordítását, azaz az operacionalizációt érintik, s egyaránt fellépnek a megfigyeléses és a kísérleti vizsgálatoknál.

1. Survey-kísérletek, online survey-kísérletek

A survey-kísérletek, azaz a felvételi (survey) körülmények között végzett randomizált kísérletek bevezetőmben említett növekvő népszerűsége éppen arra vezethető vissza, hogy reprezentatív mintára támaszkodva, a kísérletek külső érvényességét tudják növelni a belső érvényesség megtartása mellett. Ezenkívül az egyre olcsóbbá váló online felvételek (lásd például a kollaboratív adatgyűjtési projekteket, amikor megosztott platformokon egyszerre több kísérlet végezhető; a témát többek között *Mutz és Kim* [2020] is tárgyalják) a költségek oldalán sem jelentenek már nagy hátrányt.

Az online survey-kísérlet relevanciája még inkább nő olyan területeken (például a hagyományosan kényelmi mintával dolgozó pszichológiában vagy nyelvtudományban), ahol a reprezentativitásnak elvi okokból kisebb szerepe van. A mögöttes elméleti feltevés szerint az alapvető pszichológiai működésünk, illetve nyelvhasználatunk a legtöbb kutatási területen minden emberre nézve azonos, ezért a mintavétel módja tetszőleges lehet. Ez a feltevés az empirikus adatok birtokában tesztelhető is, ahogy ezt nyelvészeti esettanulmányomban a következőkben bemutatom.

Ha a kényelmi mintával is megelégedik a kutató (akár megalapozott, akár megalapozatlan módon), az online survey-kísérletre bármikor elérhető és olcsó

megoldást nyújtanak az ún. crowdsourcing¹ platformok, amelyen például az Amazon által működtetett Mechanical Turk, a Figure Eight (korábban CrowdFlower), a Lighttag vagy a kínai Weichaiishi. Az ezeken kérdezni kívánó kutatók egyszerűen felhívást tesznek közzé a platformok bedolgozói között, amelyben a juttatáson kívül a válaszadók elvárt képességeit (minimális képzettségét, anyanyelvét, a korábbi munkáikkal kapcsolatos elégedettségi arányt stb.) is megadják. E platformok az empirikus tudományok, így a kísérleti pszichológia figyelmét szintén felkeltették, noha összetettebb feladatok esetén bizonyítható az online eredmények jelentős eltérése a valódi kísérleti körülmények között szerzettektől (*Crump–McDonnell–Gureckis* [2013]). *Mutz* és *Kim* [2020] szerint a politikatudományi survey-kísérletek nagyobb része ma már kényelmi mintán működik. Ennek indokoltsága lehet, hogy a kényelmi minta a random kiosztott kezelésekből mért oksági hatás meghatározására (egy specifikus alpopulációra vonatkozóan) megfelelő, kimondva, hogy a hatás heterogenitásának, általánosíthatóságának vizsgálata további tanulmányokat igényel. A crowdsourcing platformok politikatudományi alkalmassága témájában az egyik legtöbbet hivatkozott írás (több mint 3 600 Google Scholar hivatkozással) *Berinsky, Huber* és *Lenz* [2012] munkája, akik csak annyit állítanak, hogy ezek a minták heterogénebbek, mint a gyakran használt egyetemi hallgatói vagy egyéb kényelmi minták, és kevésbé reprezentatívak, mint az internetes panelek vagy az országos véletlen minták.

Érdeemes megemlíteni, hogy a világháló nem csak survey, hanem terepkísérletek végzésére is alkalmas az internetes szolgáltatások kérdéskörét tekintve. *Simonovits, Shvets* és *Taylor* [2018] például egy hazai autómegosztó platformon tesztelték az etnikai diszkrimináció meglétét, különböző, csak etnikai háttérük és nemük szerint eltérő profilokat hozva létre.

Az előbbieken felsorolt módszertani korlátok mellett is nyilvánvaló a survey-kísérletek további terjedése, mivel azok olcsón, gyorsan végezhetőek, belső érvényességük biztosított, és – ha valószínűségi mintára épülnek – a külső érvényességük is megfelelő.

2. Alanyokon belüli versus alanyok közötti kísérleti elrendezés

A kutatók alapvetően kétfajta módon oszthatják ki a kezeléseket a résztvevők között. Alanyokon belüli (*within-subject* vagy ismételt mérések [*repeated measurements*]), a biostatisztikai irodalomban még: párhuzamos csoportos) elrende-

¹ A crowdsourcing munkaszervezeti forma sajátossága, hogy a hagyományos esetben saját dolgozók által végzett feladatokat főleg online formában független személyek nagy csoportjának szervezik ki (erre utal az angol „crowd”, azaz tömeg kifejezés) úgy, hogy a csoport tagjai csak egy kis részletet végeznek el a teljes feladatból.

zés esetén minden résztvevő csak egyfajta kezelésben részesül. Ilyen volt *Wozniak, Drakulich* és *Calfano* [2020] már idézett kutatása, amelyben a három képből minden kísérleti csoport csak egyet kapott. Emellett idetartozna egy, például a női, illetve férfi vezetőikkel kapcsolatos attitűdök feltárását célzó vignettás vizsgálat is (a vignettát bővebben lásd később), amelyben a minta random módon kettéosztott fele számára csak egy hipotetikus női, míg a másik felének egy hipotetikus férfi vezetőről szóló rövid történetet vázolnak fel, és azok főszereplőjére vonatkozóan tesznek fel kérdéseket. Alanyok közötti (*between-subjects*) elrendezés esetében ezzel szemben minden résztvevő több, gyakran minden kezelést megkap egymás után.² Tehát utóbbi példánknál maradva, a résztvevőket mind a női, mind a férfi vezetőről szóló történettel kapcsolatban megkérdeznék. Nyilvánvaló, hogy a két elrendezés közötti választás gyakran praktikus szempontból dől el: *Wozniak, Drakulich* és *Calfano* [2020] kutatása nem megvalósítható alanyok közötti elrendezést használva, mert nem életszerű, hogy az alanyok megváltoztatják a rendőrséggel kapcsolatos megítélésüket a vizsgálat két időpontja között.

A kétféle elrendezés kombinált formája a kevert elrendezés (*mixed design*), amely egyaránt tartalmaz alanyok közötti és alanyokon belüli elrendezést is. Például, ha egy „kezelés” hatását akarjuk mérni (tegyük fel, hogy valamilyen újságcikk elolvasásának hatását valamely attitűdre), akkor a kezelés előtti és utáni attitűdöt tanulmányozzuk (alanyokon belül), mely elrendezést még tovább bővíthetjük úgy, hogy kétféle kezelést vizsgálunk (például ugyanazt az újságcikket két különböző médium neve alatt, ha a médium hatását is mérni kívánjuk) az alanyok random kettéfelé osztásával (alanyok közötti elrendezés).

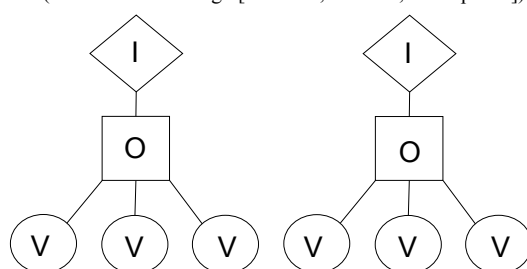
A két elrendezés közötti különbség nem csak megvalósíthatósági alapon nyugszik (*Charness–Gneezy–Kuhn* [2012]). A legfőbb különbség az oksági kapcsolat mérésére vonatkozik. Egy alanyokon belüli elrendezésnél minden résztvevő több kezelést is kap, és az oksági következtetés az ezekre adott egyéni válaszok eltérésére épül. Ezzel szemben alanyok közötti elrendezésnél az oksági következtetés a különböző kezelést kapott csoportok összevetéséből adódik – mivel a csoportokhoz történő hozzárendelés random, a csoportok várhatóan csak a kezelés szempontjából különböznek, így az oksági következtetés érvényes.

Egy további különbség statisztikai alapon fogalmazható meg: mivel az alanyokon belüli elrendezés esetén ugyanahhoz az alanyhoz több mérés is tartozik, ezeket nem tekinthetjük egymástól függetlennek, hiszen adott alany várhatóan hajlamos egymáshoz hasonló válaszokat adni (például azért, mert saját implicit skálát használ). Ezért itt a független, azonos eloszlású mintaelemeket feltevő modellek nem megfelelők, helyettük a korrelált mintaelemeket kezelni képes többszintű (vagy hierarchikus) modellek használhatók (jó összefoglalót ad a többszintű modellek társadalomtud-

² Faktoriális elrendezésnek nevezzük, ha az alanyok minden kezelést megkapnak.

mányi alkalmazásáról *Snijder és Bosker* munkája [2012]). Az utóbbiak elnevezése abból ered, hogy a válaszok közötti eltérések több szinten létrejövő hatásokból erednek. Például egy oktatásszociológiai vizsgálatban a tanulók teljesítménye az egyéni tényezőknél túl függhet az osztálytól és az iskolától is, vagyis az adatok három szinten szerveződnek, a szintek egymásba ágyazottak (*nested* vagy *hierarchical* elrendezés).

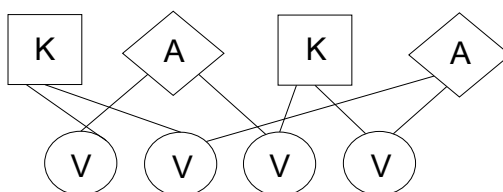
1. ábra. Egymásba ágyazott elrendezés
(Nested effects design [I: school; O: class; V: response])



Megjegyzés. I: iskola; O: osztály; V: válasz.

Az egyéneken belüli elrendezés esetén ez nem teljesül: ott a kezelés és az egyén mint szint keresztezi egymást (*crossed effect design*): a kezelések több egyénhez is hozzá vannak rendelve, és az egyének többféle kezelést kapnak.

2. ábra. Keresztezett adatstruktúra
(Crossed effects design [K: item; A: subject; V: response])



Megjegyzés. K: kezelés; A: alany; V: válasz.

Az elrendezés azonosítását követően olyan statisztikai modellre van szükségünk, amely képes megjeleníteni a vizsgált jelenség kialakulásának három szintjét és a keresztezett elrendezést, továbbá részlegesen magyarázni a jelenséget az egyes szintek jellemzőinek bevonásával. Erre a célra a többszintű regressziós modellek alkalmasak, mégpedig az elrendezésnek megfelelően a kereszthatást leíró változat (*crossed random-effects regression*; hasonló problémára lásd például *Németh-Luksander* [2013]).

Egy másik fontos különbség az alanyok közötti, illetve az alanyokon belüli elrendezés között a mintanagyságra, így a *költségekre* vonatkozik. A két kezelést összevető, alanyok közötti elrendezés ugyanannyi válasz előállításához nyilván kétszer annyi résztvevőt igényel, mint az alanyokon belüli. Az utóbbi *statisztikai hatékonyságra* vonatkozó előnye abban fogható meg, hogy csökkenti az egyéni különbözőségekből adódó mérési hibát, hiszen, ha ugyanazon alanyon belül értékelünk különböző kezeléseket, akkor minden résztvevő saját magának szolgál viszonyítási pontként, így kisebb random zajt várhatunk, mintha két (nem csak a kezelésben különböző) csoportot vetünk össze. A két csoport ugyanis csak várható értékben azonos, az aktuális random kettéosztás kisebb-nagyobb különbségeket mutathat közöttük.

Az alanyokon belüli elrendezés ugyanakkor definíció szerint magában hordoz egy olyan mérési problémát, amely sok esetben ellehetetleníti a használatát. Ez a probléma *kontextushatásként* ismert az irodalomban (*Greenwald* [1976]); lényege, hogyha egymás után tesszük ki különböző kezeléseknél alanyainkat, akkor e kezelések hathatnak egymásra, vagyis a kontextus befolyásolja az adott kezelés eredményét. A kontextushiba egyik alelete az *áthúzódo (carryover) hatás*, amikor egy kezelés (például gyógyszer) hatása még a következő kezelés idejében is kimutatható. De idetartozik a *gyakorlási (practice-) hatás* is, amikor az alany tanul a korábbi kezelésből, vagy a *terjedelem- (range-) hatás*, amikor korábbi kezelésre adott válasza az új kérdésnél viszonyítási pontként szolgál számára. Fontos megjegyezni, hogy a terjedelemtartás nem feltétlenül negatív előjelű. Egy olyan közgazdasági kísérletben, ahol a fizetési hajlandóság (*willingness to pay*) vizsgálata a cél, a viszonyítási pont megléte növelheti a belső érvényességet. Például, ha az a kérdés, hogy alanyaink mennyit fizetnének egy szendvicset a sarki boltban és mennyit a repülőtéren, az alanyokon belüli elrendezés esetén a kérdezt a második választ az első által beállított skálához szabja. Végül meg kell említeni a *kifáradást (fatigue)* is: ha ugyanannak az alanynak teszünk fel egymás után több hasonló kérdést, van rá esély, hogy a későbbi kérdésekre már kevésbé átgondolt, sematikus választ ad. E faktorok szisztematikus torzító hatása megelőzhető, ha a kezelések sorrendje nem azonos minden alanyra, hanem randomizált.

Összefoglalva: bár az alanyokon belüli elrendezés nem mindig megvalósítható adott kutatási kérdés mellett (lásd a rendőrséggel kapcsolatos attitűd mérését), ha megvalósítható, érdemes megfontolni használatát, hiszen olcsóbb és statisztikailag hatékonyabb. Ennek ellenére a társadalomkutatási gyakorlatban ritkán használják, talán azért, mert a klasszikus randomizált kontrollált kísérlet az orvostudományból jön, és a kezelt, illetve a kontrollcsoport összevetését célozza. A survey-kísérlet tipikus alkalmazási lehetőségeit szemlélő *Gaines, Kuklinski és Quirk* [2007] például csak alanyok közötti elrendezést említenek. Az alanyokon belüli elrendezés méltatlan mellőzöttségéről és megalapozatlanul vélt hátrányairól a neves amerikai statisztikus, *Andrew Gelman* is gyakran ír ismeretterjesztő írásaiban (például *Gelman* [2017]).

Véleménye szerint a kontextushatástól való félelem olyan esetben is visszatartja annak használatától a kutatókat, amikor az megelőzhető lenne, vagy amikor jelenléte ellenére az előnyök és hátrányok mérlege az elrendezés felé billenne.

Léteznek ugyanakkor az alanyokon belüli elrendezést hagyományosan gyakran használó területek is (ilyen többek között a kísérleti pszichológia), ahol tipikusan különböző, ugyanolyan szempontból értékelendő ingereknek teszik ki egymás után az alanyokat. *Deffenbacher, Leu és Brown [1981]* egy kutatásban a résztvevők arc-memóriáját tesztelték, mindegyiküknek 50 különböző arcot mutatva, majd egy hét múlva visszaemlékezésre kérve őket.

A pszichológián kívül a szociológia is ismer egy másik, alanyokon belüli elrendezést, a vignetta-módszert. A vignetták olyan hipotetikus szituációk, amelyekről az ismertetésük után kérdéseket tesznek fel a kísérlet alanyai számára, a társadalmi normákat, értékeket szándékozva ezzel felderíteni. Egy hazai példa alapján *Koltai [2011]* a nyugdíjrendszerrel kapcsolatos igazságossági attitűdöket vizsgálta a módszer segítségével.

Mindig a konkrét kutatási kérdéstől függ, hogy az alanyok közötti vagy az alanyokon belüli elrendezés az előnyösebb; ennek eldöntésekor praktikus és statisztikai megfontolásokat egyaránt figyelembe kell venni. Jelen írás célja a két elrendezés közötti explicit választás támogatása és az alanyokon belüli elrendezés statisztikailag nemtriviális elemzésének bemutatása.

3. Az alanyokon belüli elrendezés elemzése: egy nyelvészeti esettanulmány

A következőkben egy nyelvészeti esettanulmányon keresztül részletesen ismertetem az alanyokon belüli elrendezés elemzéséhez illeszkedő többszintű keresztezett hatásmodell alkalmazását, kitérve a hatótényezők kapcsolatának, valamint a külső és a konstrukciós érvényesség vizsgálatára is.

A nyelvészet legkülönbözőbb területein alkalmaznak kísérleti módszereket, mind klasszikus laboratóriumi körülmények között, mind az interneten, például a közösségi média kényelmi mintájára alapozva (egy hazai friss összefoglaló *Kenesei–Bánréti [2017]* tanulmánya). Jól példázza a webes platform nyelvészeti lehetőségeit *Skirgard, Roberts és Yencken [2017]* munkája, amely „The Great Language Game” (A nagy nyelvi játék) adatain alapul. E webes játék résztvevői ismeretlen emberekkel készített hangfelvételeket meghallgatva megtippelhatték, hogy azok milyen nyelven beszélnek. Nyolcvan országból egymillió ember próbálta ki a játékot, ezzel a valaha volt legnagyobb nyelvészeti kísérleti adatbázist hozva létre. A szerzők ezen az adat-

bázison vizsgálták a különböző nyelvekre vonatkozó találati valószínűségeket, és meghatározták az egyes nyelvpárok összetévesztését befolyásoló tényezőket.

3.1. Adatok és módszertan

A következő esettanulmányban *Elsen, Németh és Kovács* [2021] munkájának adataira épülő, leegyszerűsített másodelemzéseket mutatok be illusztrációs célt szolgáló modellezéssel, de részletezett módszertannal, az eredeti német-magyar összehasonlító vizsgálatnak csak a magyar adatbázisát és a szűkített változókészletét használva – a tényleges tartalmi eredmények iránt érdeklődőknek az eredeti dolgozatot, illetve további, készülében levő cikkeinket ajánlom.

Kutatótársaimmal a hangszimbolika jelenségét vizsgáltuk – e jelenség lényege, hogy a beszédhangokhoz a beszélők hajlamosak valamilyen szemantikai természetű minősítést asszociálni, például a jelölt tárgy méretére, alakjára, anyagára vonatkozóan. Ezeknek az asszociációknak a jelentős része univerzális, azaz nem függ a beszélő anyanyelvétől. Célunk annak vizsgálata volt, hogy a beszédhangok tulajdonságai (zöngéesség stb.) milyen univerzális összefüggést mutatnak azzal, hogy mekkora tárgyat jelölhet a szó. A kérdőívben egy képzelten idegen nyelvhez tartozó (a magyarban nem létező), hat betűből álló, felváltva mással- és magánhangzókat tartalmazó szavakat soroltunk fel (például vivivi, gegege); az alanyok feladata annak eldöntése volt, hogy azok egy 1–4 skálán mekkora tárgyat jelölnek. Hét magánhangzót és 4 mássalhangzót vontunk vizsgálatba, utóbbiakat zöngés-zöngétlen formában is, így összesen $7 \cdot 4 \cdot 2 = 56$ szót, azaz „kezelést” definiáltunk. A kényelmi mintát 121 szombathelyi egyetemista alkotta, egy önkéntes offline kérdőívben mindegyikük mind az 56 szót megkapta, azaz az alanyokon belüli elrendezést, mégpedig annak faktoriális változatát alkalmaztuk. A kényelmi minta indokolhatóságát korábban már említettük (a nyelvhasználat univerzalitására hivatkozva), de a reprezentativitás szükségességére vonatkozó feltevéseket a következőkben tesztelni is fogom.

Az alanyokon belüli elrendezés hátrányát jelenti a nyelvészeti kutatásokban is, ha a kutatási kérdés kontextushatást valószínűsít. *Arunachalam* [2013] példájában a kontextushatás egyik típusát, az áthúzódo hatást említi: ez a mondattani előfeszítés (*priming*), amikor egy nem sokkal korábban hallott mondatszerkezet valószínűbbé teszi ugyanannak a szerkezetnek a használatát. Nálunk ilyen áthúzódo hatás praktikusán nem volt várható, egyedül a kifáradási hatást valószínűsítettük. Annak érdekében, hogy ez ne okozzon szisztematikus torzítást, randomizáltuk a kérdéssorrendet. Az elrendezés előnye viszont, mint általában, hogy a kisebb minta is sok információt adott ($121 \cdot 56$ mérés), illetve statisztikailag hatékony mérést definiáltunk, hiszen a kísérleti alanyok hatását pontosabban tudtuk mérni azáltal, hogy minden kezelésre ismertük mindenki válaszát.

Fő kutatási kérdésünk a beszédhangok fonetikai jegyeinek hatása az elképzelt tárgy vélt méretére. Jelen illusztrációként szolgáló elemzésben csak négy jegyet vizsgálunk: magánhangzókra vonatkozóan a képzés helyét (elöl vagy hátul képzett) és a nyelvvállást (felső, középső vagy alsó); mássalhangzókra vonatkozóan a zöngésséget, illetve a képzés módját (zár- vagy réshang). Mivel e négy jegy hatását egyszerre és egymásra korrigálva szeretném mérni, többszörös lineáris regressziót alkalmazok. Ahogy korábban már részleteztem, az ugyanazon alanyra adott 56-féle kezelés várhatóan korrelál, ezért többszintű, keresztezett random hatást leíró lineáris regresszió (*crossed random effect linear regression*) alkalmazandó (Snijders–Bosker [2012]; nyelvészeti alkalmazásáról lásd Baayen [2008]). A Stata *xtmixed* parancsát használtam (a Stata többszintű modulját, benne az *xtmixed*-et részletesen Rabe, Hesketh és Skrondal [2012] kötete mutatja be). Mivel minden alany/kezeléscsoportban ugyanannyi megfigyelés van, és a minta kiegyensúlyozott, korlátozott (*restricted*) maximum likelihood becslést végeztem.

A többszintű modelleket használó nyelvészeti kísérletek leggyakrabban Baayen [2008] munkáját hivatkozzák, amely a statisztika nyelvészeti alkalmazásait foglalja össze. Azonban sem ez az átfogó munka, sem az általunk ismert, mérettel kapcsolatos hangszimbolikai tanulmányok (például Shinohara–Kawahara [2010], Knoeferle et al. [2017]) nem értelmezik azokat a modellparamétereket (az osztályon belüli korrelációt vagy a varianciacsökkenés arányát), amelyeknek egyébként nagy interpretációs ereje lenne. Ezért döntöttem úgy, hogy egy nyelvészeti esettanulmányon illusztrálom e modellek használatát.

A többszintű modell feltevése szerint a kezelésnek és az alanyra is van hatása a válaszra. Formálisan:

$$Y_{ak} = \gamma_{00} + \rho_{0k} + \mu_{a0} + \varepsilon_{ak}. \quad /1/$$

Az egyenletben γ_{00} az átlagosan várható válasz. Ezen mind a kezelés, mind az alany változtat, mégpedig a k . kezelés ρ_{0k} -val (az eltérés normális eloszlású 0 várható értékkel és V_K varianciával), az a . alany pedig μ_{a0} -lal (az eltérés szintén normális eloszlású 0 várható értékkel és V_A varianciával, illetve független a kezelés okozta eltéréstől). A kezelés és az alany nem határozza meg teljesen a választ, a mérésben van némi további randomitás (reziduálismérés-szintű hatás, ε_{ak}), amely független az előbbi két hatástól, és szintén normális eloszlást követ 0 várható értékkel, illetve V_{Rez} varianciával. A függetlenség feltevéséből következően a válaszok teljes varianciájára: $V_{Teljes} = V_K + V_A + V_{Rez}$. A modell teljes statisztikai tárgyalását lásd Snijders és Bosker [2012] összefoglaló munkájában.

A kezelések, illetve alanyok hatásának volumenét az osztályok közötti korrelációs együtthatók (intra-class correlation coefficients; ICC_K és ICC_A) segítségével számszerűsíthetjük, melyek a vonatkozó varianciakomponensek teljes varianciához

viszonyított arányát mutatják. Ha például az ICC_A értéke 0,08, azt mondhatjuk, hogy a teljes variancia 8 százaléka az alanyok közötti különbségeknek tulajdonítható. Az 1 értékű ICC_A arra utal, hogy minden kezelésre azonos választ ad az alany (vagyis a válaszok közötti különbségek 100 százaléka az alanyok szintjén jelentkezik), a 0-val egyenlő ICC_A esetén pedig a válaszokban nincs jelen alanszintű jellemző. A 0-tól jelentősen eltérő ICC_A úgy interpretálandó, hogy az alanyok maguk is fontos tényezők a válaszok közötti különbségek megértésében.

3.2. A fonetikai jellemzők hatása, a hatások közötti kapcsolat

Első lépésben egy üres modellt hoztam létre, amely még nem tartalmaz magyarázó változót, célja inkább az elrendezés előbbieken említett fontos paramétereinek becslése. (Lásd az 1. táblázatot.) Az ICC_K értéke közepesnek mondható, és lényegesen nagyobb, mint az ICC_A -é. Vagyis a kezelések szerepe számottevően meghaladja az alanyokét, ami a kutatási kérdést tekintve megnyugtató. Fontos viszont megemlíteni, hogy a teljes variancia jó részét a reziduális, tehát a kezelés és az alany által meg nem magyarázott random hatás adja. Az 1. táblázat annak a likelihoodhányados-próbának az eredményét is mutatja, amely az adott modellt veti össze a kezelések és az alanyok okozta korrelációt figyelmen kívül hagyó egyszintű modellel – a próba szignifikanciája arra utal, hogy érdemesebb a többszintű modellt alkalmazni (ehhez hasonlóan az is tesztelhető lenne, hogy az alanyok szintjét érdemes-e bevonnani a kezelések szintje mellé, hiszen alacsony volt az ICC_A értéke; ám ettől jelen tanulmányban eltekintek).

1. táblázat

A nullmodell paraméterbecslései
(Parameter estimates for the null model)

Paraméter	Becslés
ICC_K	0,10
ICC_A	0,02

LR-teszt vs. lineáris regresszió: $\chi^2(2) = 561,06$
(p -érték < 0,0001)

Megjegyzés. LR (likelihood rate): likelihood-hányados.

A kezelésszintű hatás hipotézisünk szerint a kezelésként definiált fiktív szavakat alkotó beszédhangok fonetikai jellemzőivel magyarázható, ezért ezeket a jellem-

zöket (rögzített hatású) magyarázó változóként vontuk be a modellbe (1. modell). Hasonlóan az üres modellhez, a magyarázó változók bevonása után megmaradt illesztett (adjusted) variancia (VT_{Teljes}^1) is három részre partícionálható: V_K^1 , V_A^1 és V_{Rez}^1 , ahol a felső indexbeli 1-es az 1. modellre utal. Ezekből illesztett ICC^1 -mutatók számíthatók. A fonetikai jellemzőkre való korrigálással a kezelésszintű hatás egy része várakozásaink szerint magyarázható, ennek sikerét a varianciacsökkenés arányával (proportional change in variance) számszerűsíthetjük:

$$PCV_K^1 = \frac{v_K^0 - v_K^1}{v_K^0}.$$

Egy 0,3 értékű PCV_K^1 például azt jelenti, hogy a válaszokban megfigyelhető kezelésszintű különbségek 30 százaléka az újonnan bevont fonetikai jellemzőknek tulajdonítható.

Az 1. modell illesztésével kapott eredmények szerint a képzés helyének, módjának és a zöngésségnek van statisztikailag bizonyítható (5 százalékos szinten szignifikáns) hatása. (Lásd a 2. táblázatot.) A szignifikancián felül a hatás nagyságára is érdemes kitérni; eszerint a képzés helyének van a legnagyobb szerepe: 0,36 egységgel változtatja a választ az 1–4 skálán.

2. táblázat

Az 1. modell becslései
(Parameter estimates for Model 1)

Megnevezés	Becslési eredmény
Rögzített hatás	
Képzés helye (referencia: elöl): hátul	+0,36*** (0,000)
Nyelvállás (referencia: alsó): középső	+0,10 (0,300)
Nyelvállás (referencia: alsó): felső	0,18 (0,064)
Zöngésség (referencia: zöngétlen): zöngés	+0,18*** (0,001)
Képzés módja (referencia: réshang): zárhang	-0,17** (0,010)
Random hatás	
ICC_K	0,04
ICC_A	0,02
PCV_K	0,63
Log-korlátozott likelihood	-8,944

*** $p < 0,001$, ** $p < 0,01$.

Megjegyzés. Zárójelben a p -értékek találhatók. A rögzített hatásokhoz tartozó paraméterek arra utalnak, hogy hány egységgel változtatta meg a fiktív tárgy méretének értékelését az adott fonetikai jellemző.

Az 1. modellt a nullmodellel összevetve megállapítható, hogy az ICC_K jelentősen lecsökkent, vagyis a négy fonetikai jellemzőre való korrigálás után a fennmaradó varianciából már csak 4 százalék tulajdonítható a (még figyelembe nem vett, tehát nem a négy jellemzővel magyarázható) kezelésszintű hatásnak. Ezzel párhuzamosan a kezelésszintű (reziduális) variancia 63 százalékkal csökkent a nullmodellhez képest (PCV_K^1), vagyis a kezelésszintű eltérések kétharmada ennek a négy jellemzőnek tudható be. Azaz e jellemzők valóban fontosak a mérettel kapcsolatos hangszimbolika megértésében, az annak létrejöttében szerepet játszó nyelvi jellemzők nagyobb részét csupán ezek segítségével sikerült megragadnunk.

A modell bővítésével a négy fonetikai jellemző hatásának egymással való kapcsolata is vizsgálható, mellyel választ adhatunk arra a kérdésre, hogy vajon ezek egymástól függetlenül hatnak-e, vagy módosítják egymás hatását. Így például, ha hátul képzett magánhangzó és zöngés mássalhangzó is jelen van egy szóban, teljesül-e, hogy együttes hatásuk erősebb, mint ami a két jellemző önálló hatásának összege alapján várható lenne. Statisztikailag ez a kérdés a jellemzők mint regressziós magyarázó változók közötti interakciók tesztelésével válaszolható meg. Több modellt teszteltem, mindegyik jellemzőre külön megvizsgálva, hogy az módosítja-e az összes többi hatását. Az eredmény nagymértékben robusztus: egyetlen esetben sem kaptam statisztikai bizonyítékot az interakció jelenlétére. Vagyis a fonetikai jellemzők egymástól függetlenül határozzák meg a választ.

3.3. A külső érvényesség vizsgálata

Ahogy már említettem, olyan kényelmi mintákkal dolgozó tudományok (implicit) feltevése szerint, mint az empirikus pszichológia vagy a nyelvészet, a legtöbb kutatási kérdés esetén a vizsgálati alanyok egymással felcserélhetők, tehát azok jellemzői nem befolyásolják lényegesen az eredményeket. Mivel kutatásunkban a vizsgálati alanyok idegennyelv-tudását, életkorát és nemét is rögzítettük, lehetőség adódott a feltevés tesztelésére e tényezőkre vonatkozóan. Az idegennyelv-tudás esetében a nyelvismeret szintjét, a beszélt nyelvek számát és típusát (germán, újlatin vagy szláv nyelvcsaládhoz tartozását) tanulmányoztuk.

Első lépésben a tényezőket mint az alanyokat jellemző magyarázó változókat vontuk be a modellbe. Egyedül az életkor hatása bizonyult statisztikailag szignifikánsnak, de nagysága elhanyagolható: a legalább 24 évesek az 1–4 skálán átlagosan 0,078 egységgel nagyobb értéket adtak válaszként, mint a fiatalabbak (minden más magyarázó változót rögzítve, azaz a négy fonetikai jellemző szerint megegyező kezelésre reagálva). Ugyanakkor az eredmény érvényességét korlátozza, hogy mivel egyetemisták töltötték ki a kérdőívet, az idősebb korcsoportokról nem áll rendelkezésre

zésre adat. Elképzelhető, hogy nagyobb életkori terjedelem nagyobb életkorhatás kimutatását tette volna lehetővé.

Második lépésben azt vizsgáltuk, hogy az alanyok tulajdonságai módosítják-e a fonetikai jegyek hatását. Igaz-e például az, hogy a több idegen nyelvet beszélők (több egymást erősítő benyomásra támaszkodva) nagyobbban vélik a zöngés, illetve zöngétlen mássalhangzók alkotta szavakkal jelölt tárgyak mérete közötti különbséget? Ha ilyen jellegű interakció kimutatható, az érvet szolgáltat a kényelmi minták használatával szemben, hiszen a zöngésesség kimért hatása függni fog a minta nyelvtudás szerinti összetételétől.

A kérdés ismét interakciók bevonásával formalizálható statisztikailag. Csak másodrendű interakciókat vizsgáltam, azaz olyan típusú kérdést nem, amely három magyarázó változóra terjedne ki (például van-e különbség a nők és férfiak között abban, hogy mekkora hatást gyakorol az idegennyelv-ismeret a zöngés-zöngétlen különbségre). Az alanyok minden tulajdonságához külön modellt rendeltem, kiegészítve az 1. modellt az adott tulajdonság és a fonetikai jegyek közötti interakciókkal. Mivel egyszerre több paraméter szignifikanciáját vizsgáltam, Bonferroni-korrekciót alkalmaztam az elsőfajú hiba korrigálására (lásd *multiple comparison* [többszörös összehasonlítási] probléma). Az életkor és a nem esetében statisztikailag bizonyítható volt a hatásmódosítás jelenléte. A nem hatásmérete is jelentős, a nők a hangképzés helye szerint nagyobb különbséget tesznek: körükben kétszer nagyobb a hátul, illetve az elől képzett magánhangzók közötti különbség, mint a férfiaknál. Mindez a kényelmi minták bizonyos fokú torzítására figyelmeztet: minél több nő van egy mintában, annál nagyobb hatást fog mutatni a hangképzés helye. Ugyanakkor egyetlen vizsgált alanszintű tulajdonság sem változtatott a fonetikai jegyek hatásának előjelén vagy nagyságrendjén, vagyis nem találtunk bizonyítékot arra, hogy a kényelmi minták hamis tendenciákat mutatnának.

3.4. A konstrukciós érvényesség egy vizsgálati lehetősége

Kísérletünk konstrukciós érvényességével, azaz az alkalmazott mérési mód és a mérni kívánt hangszimbolikai jelenség megfeleltethetőségével kapcsolatban feltehető kérdésként az, hogy a használt fiktív szavak (gegege, tititi stb.) mögött biztosan nem sejlik-e fel a kérdezettek számára olyan szemantikai asszociáció, amely befolyásolná a jelölt tárgy nagyságára vonatkozó értékelésüket. Bár a három ismétlődő szótag egyike sem alkot magyar szót, két szótagú változatuk több esetben is a babanyelv eleme (baba, popo, bebe, didi, baba, bibi, kaka, pipi). Felmerül tehát, hogy a babanyelvi szavak három szótagos változatát nem értékelték-e másképp a kérdezettek – kézenfekvő módon feltehetjük, hogyha fellépett hatás, akkor az csökkentette a jelölt tárgy vélt méretét a babákhoz kapcsolódó asszociáció révén.

E kérdés vizsgálatához létrehoztunk egy kezelésszintű változót, amely azt jelezte, hogy a babanyelvhez tartozik-e a szó két szótagú változata. Az 1. modellt ezen új magyarázó változóval bővítve újraillesztettük. Eredményünk szerint a változó hatása szignifikáns, mégpedig negatív előjelű (ahogy feltettük), és nagysága a leg-erősebb hatású fonetikai jegyhez (a képzés helyéhez) mérhető. Ugyanakkor bevonása nem változtatott a fonetikai jegyek hatásának előjelen vagy nagyságrendjén. Mindez arra utal, hogy bár nem sikerült szemantikai asszociációktól teljesen mentes kezelést definiálni, a kutatás érvényességét ez nem gyengítette.

4. Összefoglalás

Tanulmányom a társadalomtudományokban az utóbbi évtizedben egyre inkább tért hódított kísérleti megközelítés egyik formájával, az alanyokon belüli elrendezés-sel foglalkozott. Ez az elrendezés, bár nem mindig megvalósítható, olcsóbb és statisztikailag hatékonyabb, mint gyakrabban használt párja, az alanyok közötti elrendezés. Dolgozatomban előbb röviden összevettem a két megközelítést a társadalomtudományi alkalmazási lehetőségeiket tartva szem előtt, majd, mivel az alanyokon belüli elrendezés vizsgálata a megfigyelések korreláltsága miatt statisztikailag nem triviális, egy hangszimbolikai kutatás példáján bemutattam az annak elemzéséhez illeszkedő többszintű keresztezett randomhatásmodell használatát. A statisztikai modellek paramétereit részletesen interpretáltam a kutatási kérdésnek megfelelően, és a fonetikai jegyek mint magyarázó változók hatásának mérésén túl e hatások függetlenségének vizsgálati lehetőségéről is áttekintést adtam. Végül azt ismertettem, hogy miként lehet (legalább részben) vizsgálni az alanyokon belüli elrendezés két potenciális módszertani gyengeségét, a külső, illetve a konstrukciós érvényesség korlátjait.

Irodalom

- ARUNACHALAM, S. [2013]: Experimental methods for linguists. *Language and Linguistics Compass*. Vol. 7. No. 4. pp. 221–232. <https://doi.org/10.1111/lnc3.12021>
- BAAYEN, R. H. [2008]: *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge University Press. Cambridge.
- BERINSKY, A. J. – HUBER, G. A. – LENZ, G. S. [2012]: Evaluating online labor markets for experimental research: Amazon.com’s Mechanical Turk. *Political Analysis*. Vol. 20. No. 3. pp. 351–368. <https://doi.org/10.1093/pan/mpr057>

- CHARNESS, G. – GNEEZY, U. – KUHN, M. A. [2012]: Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*. Vol. 81. No. 1. pp. 1–8. <https://doi.org/10.1016/j.jebo.2011.08.009>
- CRUMP, M. J. C. – MCDONNELL, J. V. – GURECKIS, T. M. [2013]: Evaluating Amazon’s Mechanical Turk as a tool for experimental behavioral research. *PLoS ONE*. Vol. 8. No. 3. e57410. <https://doi.org/10.1371/journal.pone.0057410>
- CUTTLE, C. [2017]: *Research Methods in Psychology*. University of Washington. Seattle.
- DEFFENBACHER, K. A. – LEU, J. R. – BROWN, E. L. [1981]: Memory for faces: Testing method, encoding strategy, and confidence. *The American Journal of Psychology*. Vol. 94. No. 1. pp. 13–26.
- ELSEN, H. – NÉMETH, R. – KOVÁCS, L. [2021]: The sound of size revisited – New insights from a German-Hungarian comparative study on sound symbolism. *Language Sciences*. Vol. 85. May. Paper No. 101360. pp. 1–13. <https://doi.org/10.1016/j.langsci.2021.101360>
- FREEDMAN, D. A. – PISANI, R. – PURVES, R. [2005]: *Statisztika*. Typotex. Budapest.
- GAINES, B. – KUKLINSKI, J. – QUIRK, P. [2007]: The logic of the survey experiment reexamined. *Political Analysis*. Vol. 15. No. 1. pp. 1–20. <https://doi.org/10.1093/pan/mpi008>
- GELMAN, A. [2017]: Poisoning the well with a within-person design? What’s the risk? *Statistical Modeling, Causal Inference, and Social Science* page. 25 November. <https://statmodeling.stat.columbia.edu/2017/11/25/poisoning-well-within-person-design-whats-risk>
- GREENWALD, A. G. [1976]: Within-subjects designs: To use or not to use? *Psychological Bulletin*. Vol. 83. No. 2. pp. 314–320. <https://doi.org/10.1037/0033-2909.83.2.314>
- JANKY, B. [2019]: Changing connotations and the evolution of the effect of wording: Labeling asylum-seekers in a political campaign. *International Journal of Public Opinion Research*. Vol. 31. No. 4. pp. 714–737. <https://doi.org/10.1093/ijpor/edy035>
- KENESEI I. (főszerk.) – BÁNRÉTI Z. (szerk.) [2017]: *Általános Nyelvészeti Tanulmányok XXIX.: Kísérletes nyelvészet*. Akadémiai Kiadó. Budapest.
- KNOEFERLE, K. – LI, J. – MAGGIONI, E. – SPENCE, C. [2017]: What drives sound symbolism? Different acoustic cues underlie sound-size and sound-shape mappings. *Scientific Reports*. Vol. 7. No. 5562. <https://doi.org/10.1038/s41598-017-05965-y>
- KOLTAI J. [2011]: A nyugdíjrendszerrel kapcsolatos igazságossági attitűdök – a vignettás módszer – elemzési lehetőségei. In: *Örkény A. – Székelyi M. (szerk.): Az igazságosság labirintusai-ban: Társadalmi méltányosság és generációs igazságosság a 21. század Magyarországn.* Sík Kiadó. Budapest. 147–170. old.
- MUTZ, D. – KIM, E. [2020]: The progress and pitfalls of using survey experiments in political science. *Oxford Research Encyclopedia of Politics*. <https://oxfordre.com/politics/view/10.1093/acrefore/9780190228637.001.0001/acrefore-9780190228637-e-929>
- NÉMETH R. [2015]: Oksági következtetés az empirikus szociológiai kutatásban. *Szociológiai Szemle*. 25. évf. 2. sz. 2–30. old.
- NÉMETH R. – LUKSANDER A. [2013]: A kérdezőbiztosok pártpreferenciájának hatása a politikai közvélemény-kutatások eredményeire: bizonyítékok és magyarázatok. *Szociológiai Szemle*. 23. évf. 2. sz. 52–71. old.

- RABE-HESKETH, S. – SKRONDAL, A. [2012]: *Multilevel and Longitudinal Modeling Using Stata*. Stata Press. College Station.
- SHINOHARA, K. – KAWAHARA, S. [2010]: A cross-linguistic study of sound symbolism: The image of size. *Berkeley Linguistics Society*. Vol. 36. No. 1. pp. 396–410. <https://doi.org/10.3765/bls.v36i1.3926>
- SIMONOVITS, B. – SHVETS, I. – TAYLOR, H. C. [2018]: Discrimination in the sharing economy: Evidence from a Hungarian field experiment. *Corvinus Journal of Sociology and Social Policy*. Vol. 9. No. 1. pp. 55–79. <https://doi.org/10.14267/CJSSP.2018.1.03>
- SKIRGARD, H. – ROBERTS, S. G. – YENCKEN, L. [2017]: Why are some languages confused for others? Investigating data from the Great Language Game. *PLoS ONE*. Vol. 12. No. 4. <https://doi.org/10.1371/journal.pone.0165934>
- SNIJDER, T. A. B. – BOSKER, R. J. [2012]: *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Sage. London.
- WOZNIAK, K. H. – DRAKULICH, K. M. – CALFANO, B. R. [2020]: Do photos of police-civilian interactions influence public opinion about the police? A multimethod test of media effects. *Journal of Experimental Criminology*. 25 January. <https://doi.org/10.1007/s11292-020-09415-0>