

# TÉZISEK

Rung András

Magyar főnévi alaktani jelenségek analógiás megközelítésben  
című doktori értekezéséhez

## 1. Célkitűzések

Ha bármely nyelv analógiás nyelvtanát kívánjuk megírni, annak egyik alapfeltétele az, hogy tudjuk, mely fonémák-hangok, alakok (Bybee 2001), összetételi elemek (Krott 2009: 132), konstrukciók (Fillmore és Kay 1987, Goldberg 1995, 2006) hasonlóak az adott nyelvben, és ezek **hasonlósága milyen mértékű, min alapszik**. Ehhez azonban meg kell határoznunk, hogy a hasonlóságot milyen paraméterek mentén mérjük, és hogy a nyelvi működésben milyen tulajdonságokban, viszonyokban számít egyáltalán a hasonlóság. Ezek ismerete nélkül az analógiás vizsgálatok nehezen megfogható spekulációkká válhatnak (Bybee 2010: 62). Vizsgálódásaink szempontjából a **viszonyok hasonlósága** épp olyan fontos, mint maguknak az elemeknek a hasonlósága. Ezeket azonban az elemek hasonlósági mértékének ismerete nélkül nem tudnánk megállapítani. Másrészt az analógiás nyelvtanok feltételezése szerint a hasonló elemek gyakrabban rendelkeznek hasonló viszonyokkal, amelyek felismerése ezen elemek közt valószínűbb és könnyebb is.

Disszertációmban azt vizsgáltam, hogy a **hasonlóság miképp befolyásolja önállóan és a gyakorisággal interakcióban a nyelvi változást, az analógiás kiegyenlítődést és kiterjesztést**, illetve a produkciót. A formai hasonlóságok és az ingadozás elemzése során céлом volt, hogy ezek természetét feltárjam, s leírásommal hozzájáruljak az analógiás megközelítés eszköztárának finomításához, pontosságának, egzaktságának növeléséhez. A formai hasonlóságok vizsgálatának **középpontjában a szavak hasonlósági viszonyainak feltérképezése és összehasonlítási módjaik modellezése állt**. Egyes vonatkozásaiban újszerű megközelítéssel modellt kívántam adni az írásomban nem elemzett morfofonológiai jelenségek elemzéséhez mind a magyar, mind más gazdag morfológiával rendelkező nyelvek esetében. Kutatásomban cél volt megmutatni azt is, hogy a szavak holisztikusan

felfogott szerkezete és a fonémánál nagyobb egységet magukba foglaló szövegek milyen nagy szerepet játszanak abban, hogy egy-egy szó miképp viselkedik, amit használati körülményei és jelentései is befolyásolnak.

## 2. Elméleti háttér

A **szabályalapú nyelvtanok** (akár hagyományosak, akár generatívak) sokszor jó közelítő leírást adnak az alaktani viselkedésről, azonban **több nyelvi jelenségre, folyamatra nem tudnak megnyugtató magyarázatot adni.** Így megválaszolatlanul hagyják azokat a kérdéseket, hogy gyakran miért **fokozatosak az átmenetek** az egyes nyelvi kategóriák közt (Chandler 2002: 57, Lakoff 1987, Taylor 1995), mi a **valószínűség** szerepe a nyelvhasználatban, melyek a nyelvi **ingadozás** okai, illetve mi a **gyakoriság** hatása a nyelvi változásra (Skousen 1989). Ezeket a problémákat a generatív nyelvészet a performancia és a kompetencia szétválasztásával kezeli. A kompetencia alá tartozó reprezentációkat redundanciamentesnek és kategorikusnak (pl. bináris jegyek) veszi, a nyelvi elemekhez egyféle viselkedést rendel, míg számos nehezen magyarázható jelenséget (pl. beszédtevesztések (Frisch 1996: 109)) a performanciának tulajdonít, amelyeknek működését és jellegét azonban homályban hagyja, így a performancia és a kompetencia látszólagos elvi szétválasztása mögött inkább praktikus okok húzódnak meg (Bybee 2001, Bybee 2010, Skousen és mtsai 2002, Blevins és Blevins 2009).

Ezzel szemben ezekre a kérdésekre az általam **választott analógiás megközelítés rugalmasságának köszönhetően képes jobb feleletet adni,** és kezelni tudja azokat a helyzeteket is, amikor a nyelvi adatok látszólag egyediek, nem egyértelműek, rosszul formáltak, vagy a zaj, felejtés, vagy bármilyen más ok hatására azok értelmezésében és produkciójában korlátozottak vagyunk, miközben a szabályos viselkedésre is tud magyarázattal szolgálni (Eddington 2003, Skousen 1989: 54–60). Ezekben a bizonytalan esetekben egy szabályrendszer „működésképtelenné” válik, ha nem tartalmaz olyan szabályt, amely alkalmazható az adott nyelvi elemekre,

vagy akár több is van belőlük. Ilyenkor a szabályalapú megközelítésben kénytelenek vagyunk önkényes módon e nyelvi elemek tömegét rendhagyóként megjelölni, holott ezek egy másik, alkalmazkodóbb keretrendszerben magyarázhatóak lennének.

További problémákat vet fel, hogy a generatív elméletek a **nyelven kívüli hatásokat, például a használati gyakoriságot, teljesen kiküszöbölték a nyelvi leírásból**, holott számos esetben ezeknek jelentős befolyása lehet magának a rendszernek a formálódására is (Ullman 1999, Pinker 1999, Kraska-Szlenk 2007, Rung 2008, Rung 2009). Ezzel szemben a használat alapú nyelvtanok (Halliday 1961, Bybee 2010) a kommunikáció hatékonyságára, módjaira és egyéb szociális, pszichológiai funkcióira helyezik a hangsúlyt, így **jobban közelítik a pszichológiai realitást**, azaz a valós nyelvi működést. Ennek a törekvésnek az eredménye a gyakoriság fogalmának beépítése az elméletbe, amelynek hatásait a nyelvhasználatban pszicholingvisták már évtizedekkel korábban megmutatták (MacDonald 1994, Hare és mtsai 2001). Írásomban határozottan elkötelezem magam amellett is, hogy a **nyelv vizsgálatának nagy mennyiségű adatra** kell támaszkodnia (Sinclair 1991, Jurafsky és mtsai 2001), és a reális, valós folyamatokat leíró modelleknek ki kell állniuk a számítógépes tesztelés próbáját is (Skousen és mtsai 2002).

Disszertációmiban részletesen **ismertetem** azokat a **kutatásokat**, amelyek vagy **elméleti elgondolásaiknak, vagy technikai megoldásaiknak köszönhetően szorosan összefüggnek vizsgálataimmal**. Áttekintem az analógiás nyelvi megközelítés kialakulását, alapelveit és a vele kapcsolatos vitákat. Tárgyalom az analógia szempontjából különösen fontos hasonlóság és gyakoriság szerepét a nyelvben, illetve a paradigmák felépítésével és változásával kapcsolatos tudományos nézeteket. Bemutatom a magyar szerzőktől származó analógiás kutatásokat is, amelyekre adataimban és elképzeléseimben leginkább támaszkodom. A paradigmák szerveződését a nyelvi változás folyamatosan átalakítja, ennek áttekintésére is sor kerül, amit az analógiás modellezés ismertetése követ. Ebben bemutatom az analógiás modellezés szempontjából legjelentősebb AM (Analogical Modeling,

Analógiás modellezés, Skousen 1989) és TiMBL (Tilburg Memory Based Learner, Tilburgi memóriaalapú tanuló algoritmus, Daelemans és van den Bosch 2005) programokat, valamint röviden kitérek más, kevésbé ismert, de a vizsgálataimhoz kapcsolódó modellezési kezdeményezésekre is. Ezen túl röviden bemutatom, hogy miképp látom a hasonlósági hatások és a gyakoriság funkcióját egy analógiás nyelvtenban. Külön kitérek a prototípusok szerepére, hogy egyes, valamilyen szempontból kivételes szavak hasonlósága és gyakorisága miképp határozhatja meg és befolyásolhatja más szavak viselkedését.

### 3. A vizsgálat módszerei és anyaga

Vizsgálatomat a magyar főnevek egy közepes méretű csoportjának alaktani viselkedésével kapcsolatban végeztem el. A **hangkivető főnevek** önmagukban is eléggé érdekesek ahhoz, hogy érdemes legyen teljes körűen megvizsgálnunk őket, de leírásukra nem csak célként tekintettem, hanem arra szolgáló eszközként is, hogy a nyelvi viselkedés szervezésében alapvető jelentőségű tényezőket, a hasonlóságot és a gyakoriságot jobban megismerhessük.

Vizsgálatom a belső felépítés tanulmányozása helyett a hangkivető szavak **egymás közti viszonyainak és a többi főnévtől való különbségüknek** az alaposabb megismerésére irányult. A hangkivető főnevek belső és a többi főnevekhez viszonyított külső viszonyainak jellegéről sokat elárul az egyes hangkivető szavak hangkivetésének mértéke és ennek a mértéknek az egyes toldalékaik közt való megoszlása. Amennyiben azt feltételezzük, hogy a szavak hangkivető volta nem véletlenszerűen alakult ki és maradt fenn, akkor szükségszerűen ezeknek a szavaknak valamilyen szempontból hasonló jegyekkel kell rendelkezniük.

A vizsgálatra **1211 hangkivető főnevet választottam ki a BME MOKK morphdb.hu szótárából**, amely jelenleg a legnagyobb ingyenesen is

hozzáférhető nyelvi adatbázis (130 ezer szó, Trón és mtsai 2006). Az 1211 szó összesen 229 szóból és az azokból létrehozott összetételekből áll. Vizsgálatomat a Rebrus és Törkenczy (2008) által meghatározott  $\sim VC_a(o/e/ö)C_b\#$  mintán túl az utolsó magánhangzóként *-a-t* és *-u-t* tartalmazó szavakra is kiterjesztem, mint pl. *ajak*, *bajusz*, *vacak* stb. (17 szó). Kutatásom során a *Szószablya Korpusz* (Halácsy és mtsai 2003) alapján számított **gyakorisági adatokat** használtam fel.

A **szavak hasonlóságának vizsgálatában azok felszíni szerkezetét vettem alapul** (Kálmán 2008, Bybee 2001, Fűköh és Rung 2005, Rebrus és Törkenczy 2008). Feltételezésem szerint az analógia alkalmazását további szempontok is meghatározzák (használati mód, jelentés, stb.), de ezekből a legfontosabb a hangtani/fonológiai hasonlóság és a gyakoriság (Lukács 2002).

**Vizsgálati anyagomat három eltérő módszerrel közelítettem meg**, amelyeket a továbbiakban röviden be is mutatok. A hangkivető főneveket a végeik alapján megállapítható viszonyaik alapján jellemeztem a korábbi kutatásoknál alaposabban és pontosabban, általános és atomi szinten, aminek során vizsgálataimat statisztikai számításokkal is alátámasztottam. Az elemzett szóanyagot gráfstruktúrákban is tanulmányoztam, illetve ennek két állapotát össze is vettem korpuszadatokra támaszkodva. A vizsgált szóanyag viselkedését ezt követően modelleztem az analógiás modellezés bevett módszereivel. Végezetül azt teszteltem, hogy álszavakat elméletemmel összhangban ragoznak-e magyar anyanyelvű beszélők vagy sem.

A három eltérő jellegű vizsgálathoz **a szavak hasonlóságát saját fejlesztésű algoritmusokkal határoztam meg**, amelyek a kurrens hasonlóság mérésére használt algoritmusoknál (Skousen és mtsai 2002) finomabb összehasonlításokat is lehetővé tesznek. A *komplex jegymérték* és a *komplex tengelymérték* nevű algoritmusok a szavak hasonlóságát azok jobb szélétől véve számítják ki úgy, hogy a megfeleléseknek, hasonlóságoknak egyre kisebb súlyt adnak a szavak bal széle felé haladva. Így mind a két számítógépes algoritmus a *vas* és *sas* szavakat hasonlóbbnak tekinti, mint a *vas* és a *vaj* szavakat. Az algoritmusok a hasonlítást az egyes fonémák jegyei

alapján végzik el, de a *komplex jegymérték* (Rung 2008, Rung 2009) fonémákat hasonlít össze, míg a *komplex tengelymérték* az egyes jegyek tengelyeinek hasonlósága alapján számítja ki két szó hasonlósági értékét. Ezeket az értékeket egy 0-1 terjedő skálán adom meg. 0 értéket kapnak az egyáltalán nem hasonlító szavak, míg 1 a szavak önmagukhoz mért hasonlósági értéke.

#### 4. Eredmények összefoglalása

- A hangkivető főnevek nem egységsen viselkednek. Ebben a viselkedésben jelentős szerepet játszik egymáshoz való hasonlóságuk és gyakoriságuk.
- Az analógiás források kiválasztása formalizálható, precízen is meghatározható.
- A nyelvi változásban az egyes paradigmacellák nem egyformán vesznek részt.
- A magyar főnevek viselkedése modellezhető analógiásan, amelyben a kurrens tanuló algoritmusokkal egyező vagy jobb eredményeket tudtam elérni.
- A szavak végéhez közelebbi fonémák hasonlósága, azonossága esetén két szót hasonlóbbnak érznek a beszélők a toldalékolás szempontjából (szuffixumok esetében), mint ha a hasonlóságok, azonosságok a szavak belsejében vagy a bal szélén figyelhetők meg.
- A szavak viselkedését nem csak a hozzájuk leghasonlóbb szó befolyásolja, hanem további szomszédai is hatással vannak erre.

#### 5. Hangkivető főnevek viselkedésének elemzése

A magyar hangkivető főnevek viselkedését az analógiás alapelvekre támaszkodva **elemeztem**. Első lépésben a végszekvenciáik alapján tártam fel a hangkivető szavak viselkedésének mozgatórugóit.

utolsó fonéma	példánygyakoriság	példánygyakoriság alapján aránya	típusgyakoriság	hányszor gyakoribb a hangkivetőknél	hangkivetés mértéke (típus-alapon)	hangkivetés mértéke (példány-alapon)	hangkivetés az összes alakban
m	1412628	65,39%	558	21,25	99,6%	99,9%	52%
g	378878	17,54%	55	1,97	99,7%	99,9%	52%
k	183660	8,50%	170	2,47	98,1%	99,5%	52%
r	105764	4,90%	186	0,49	97,3%	98,2%	46%
ny	23048	1,07%	36	0,30	99,5%	99,7%	49%
l	22492	1,04%	26	0,23	74,7%	84,1%	33%
n	15644	0,72%	20	0,27	98,1%	99,6%	37%
j	12011	0,56%	21	0,27	96,7%	98,5%	39%
cs	4704	0,22%	13	0,63	99,6%	99,9%	69%
sz	1136	0,05%	5	0,02	32,4%	36,2%	48%
z	471	0,02%	2	0,02	43,7%	33,6%	17%

Hangkivető főnevek csoportjai utolsó fonémáik alapján

A későbbi vizsgálatok során megmutattam, hogy a **hangkivető főnevek esetében azoknál kisebb a hangkivetés mértéke az eredetileg hangkivetést elváró toldalékaiknál, amelyek összes alakjai közt** (a képzetteket is beleértve) **a hangkivetéses alakok aránya alacsony**. Ez az összefüggés azért jelentős, mert az összes alakok viszonylatában mérhető hangkivetéses alakok alacsonyabb aránya nem lehet kizárólagosan a szavak ingadozásának következménye. Azok a szavak, amelyeknél a hangkivetéses alakok aránya összes alakjaikban kisebb, hajlamosabbak arra, hogy hangkivetésükben elbizonytalanodjanak, és részt vegyenek az analógiás kiegyenlítésben. Az összes alak alapján mért hangkivetési mértéknek és az eredetileg hangkivetést elváró toldalékos alakok hangkivetési mértékének ezen összefüggését az egyes tövégek alapján kialakított hasonlósági csoportoknál (utolsó és utolsó két fonéma, utolsó két mássalhangzó) is igazolni tudtam.

A hangkivető főnevek hasonlósági csoportokba való szerveződését gráfstruktúrákban is áttekintettem. Megmutattam, hogy a változásban **nemcsak a bizonyos csoportokhoz való hasonlóságnak, hanem az ezektől való eltérésnek is szerepe van**. A felépítésükben egyedi, magányos szavak



jobban eltávolodtak a hangkivető séma által meghatározott viselkedéstől, mint azok a hangkivető főnevek, amelyek a hangkivetésükben hasonlóan viselkedő szavakkal közösen hasonlósági csoportokba rendeződtek. Ebből azt látjuk, hogy az analógiás alapú regularizálódásban elsődleges szerepe lehet a saját viselkedési csoporttal való kapcsolatok meggyöngyülésének (formai vagy jelentésbeli autonómia), amely hatásában akár jelentősebb lehet, mint az a vonzóerő, amelyet a nem hangkivető főnevek fejtenek ki ezekre a szavakra.

Ezzel szemben a formailag heterogén hangkivető szavak közt **zárt mintát alkotó -alom, -elem végűek következetesen hangkivetők, ami kapcsolatba hozható erős hasonlósági viszonyaikkal és magas gyakoriságukkal** is. Közel azonos viselkedésük nem tulajdonítható az *-alom, -elem* morféma-alkotóknak, hisz több esetben ezek nem jól vagy egyáltalán nem szegmentálhatóak (*cimbalom, malom, alom, halom, gyalom* stb.), és morféma-alapon azt sem tudnánk magyarázni, hogy több nagyon hasonló, de némileg eltérő végű szó (*áalom, ólom*) miért viselkedik közel azonos módon velük. A hasonlóság és a különbözőség tényezőjének fontosságát megmutattam azzal is, hogy a *komplex jegymérték* és a *tengelymérték* által meghatározott legszorosabb kapcsolatok száma szignifikánsan összefügg a hangkivetés mértékével. Gyakorisági hatásokat tudtam kimutatni az összetett szavak esetében is, amelyeknél azt tapasztaltuk, hogy az alapszótól eltérő hangkivetési mértékkel rendelkező összetett szavak az átlagosnál kisebb típus- és példánygyakoriságú összetett szóbokrokban találhatóak.

Két nagyméretű szöveges korpusz összehasonlításával (*Szószablya Korpusz*, 2010-es saját gyűjtés a *Google* kereső segítségével: *Google Gyakorisági Gyűjtés*) a változás jellegét, illetve a különféle toldalékos alakok sajátos viselkedését vizsgáltam. *Google Gyakorisági Gyűjtés*emet 2010 tavaszán készítettem, amely a vizsgálatomban szereplő hangkivető szavak leggyakoribb hangkivetéssel együttjáró toldalékos alakjainak gyakorisági számait tartalmazza. A *Szószablya Gyakorisági Szótár* és a *Google Gyakorisági Gyűjtés* összehasonlítása során a legfontosabb felismerésem az volt, hogy a **változásban a paradigmaticus cellák meglehetősen eltérően vesznek részt.**

Ezzel érintőlegesen már Rebrus és Törkenczy (2008) is szembesített minket. Az egyes paradigmacellák viselkedése nem teljesen autonóm, hisz megfigyelhetők benne általános tendenciák, de a változás nem úgy megy végbe, hogy a szavak egyenletesen vagy akár hirtelen sorolódnak át egy másik paradigmába.

Az egyes paradigmacellák kisebb, mások magasabb hangkivetési mértékkel rendelkeznek, amelyek statisztikailag igazolhatók. Az egyes toldalékoknál a fonotaktikai, rendszerbeli vagy használati okokra visszavezethetően a hangkivetési mértékek és azok változásának tempója eltérnek, és ezek közt nincs egyértelmű összefüggés, hisz legalacsonyabb mértékben a szuperesszívusz esetén figyelhető meg a hangkivetés, de egyedül csak **az E.3 birtokosra utaló alak változása vagy az informális regiszterekben való erősebb jelenléte igazolható szignifikáns módon.** Azonban az egyes toldalékok közt felállítható hierarchia a hangkivetési mérték szerint számos egyedi esetben eltéréseket mutat, és a változásban részmintázatok is megfigyelhetők (pl. tárgyesetükben legkevésbé hangkivető, gyorsan változó szavak)

	Szószablya 2003 hangkivetés mértéke	Google 2010 hangkivetés mértéke	változás mértéke	változás dinamikája
tárgyeset	96,82%	96,36%	0,46%	1,14
szuper-	95,31%	93,11%	2,2%	1,47

esszívusz				
<b>többes szám</b>	98,62%	98,54%	0,08%	1,06
<b>E.1 birtokos</b>	97,83%	97,68%	0,15%	1,07
<b>E.3 birtokos</b>	97,56%	96,66%	0,9%	1,37
<b>T.3 birtokos</b>	98,71%	98,55%	0,16%	1,12
<b>E.3 birtokos több birtokkal</b>	98,99%	98,71%	0,28%	1,28
<b>összes toldalék</b>	97,57%	96,97%	0,6%	1,25

A hangkivető főnevek hangkivetésének mértéke az egyes toldalékoknál a *Szósablya* korpuszban és a *Google Gyakorisági Gyűjtésben*

## 6. Hasonlósági hatások modellezése

Az analógiás megközelítéssel szemben felhozott **kifogások** többnyire arra irányulnak, hogy **az analógiás források kiválasztásának a módja esetleges, nem elég egyértelmű**, ezért **algoritmusaim segítségével teszteltem a szavak hasonlóságáról alkotott elképzeléseim helyességét** olyan feladatokon, amelyekkel a szabályalapú elméletek nehezen boldogulnának. Nyelvi tudásunk része, hogy új, vagy legalábbis a beszélő számára kevésbé ismert szavakat kategorizálunk hasonlósági alapon, és besoroljuk őket egy már ismert paradigmába. Ezt a képességet kívántam a hasonlósági mértékeken alapuló algoritmusaimmal megragadni több modellezési teszt segítségével.

Az **első teszt**em azt vizsgálta, hogy különféle algoritmusok megfelelő analógiás forrást választanak-e a hangkivető főnevek egy csoportjához **eltérő méretű szóminta alapján**. Módszereiben ez egy korábbi, a településnevek lokatíviszaival kapcsolatos vizsgálatomat követi (Rung 2008). **Második** tesztem középpontjában az állt, hogy ha a rendelkezésünkre álló szótári állomány jól reprezentálja egy magyar felnőtt mentális lexikonát, akkor

mennyire jól tudnának az algoritmusok az **összes hangkivető szóhoz a teljes lexikont figyelembe véve analógiás forrást választani**. Harmadik tesztet ugyanezen a szóanyagon hajtottam végre, de ezúttal a *komplex jegymértéket* más gépi tanuló algoritmusokkal vettem össze tízszeres keresztellenőrzés segítségével. **Végezetül** a hasonlósági mértékek alapján a legközelebbi források helyett **prototípusokat választottam ki** az egyes hangkivető főnevekhez egy olyan algoritmus segítségével, amely a vizsgálataim során megszerzett tapasztalatok egy részét összegzi működésében. A kiválasztott prototípusok segítségével a hangkivető szavak hangkivetési mértékében megfigyelhető különbségek okaira kerestem magyarázatot.

A szavak viselkedésének tanulmányozása során megfigyelt jelenségek alapján készített **algoritmusokról** bebizonyosodott, hogy **jól képesek megragadni a hangkivető szavak formai jellegzetességeit**, és több tanuló algoritmusnál is sikeresebben osztályozták azokat (egyedül a maximum entrópia modell hozott hasonlóan jó eredményt). A bináris döntési helyzetekben a *komplex jegymérték* nevű hasonlítási mód teljesített a legjobban, ami azt mutatja, hogy ha kategorikus döntéseket kell hoznunk szavak viselkedéséről, akkor elsősorban a végek hasonlóságára (nem azonosságra) hagyatkozunk. A szavak hangkivetési mértékében tapasztalható különbségeket azonban már a szerkezeti hasonlóságoknak nagyobb súlyt adó, így holisztikusabb *komplex tengelymérték* is jól tudta megragadni. Ez a hasonlítási mód a többinél jobban teljesített, ha csak néhány prototípus segítségével kellett modellezni a szavak hangkivetési mértékét. A szavak viselkedésében szerepet játszó **prototípusok kiválasztásában azonban legnagyobb szerepet a példánygyakoriság játszotta**. Ebben az esetben ismét a tövégek hasonlósága számított jobban.

	F pontszám egyéb főnév	F pontszám hangkivető	Tévesztés száma: egyéb főnév -> hangkivető	Tévesztés száma: hangkivető -> egyéb főnév
Döntési fa (J48)	0,999	0,955	39	58
Maximum entrópia	0,999	0,977	27	23
Komplex jegymérték	0,999	0,979	31	14
TiMBL (Tilburg Memory Based Learner)	0,999	0,955	60	37

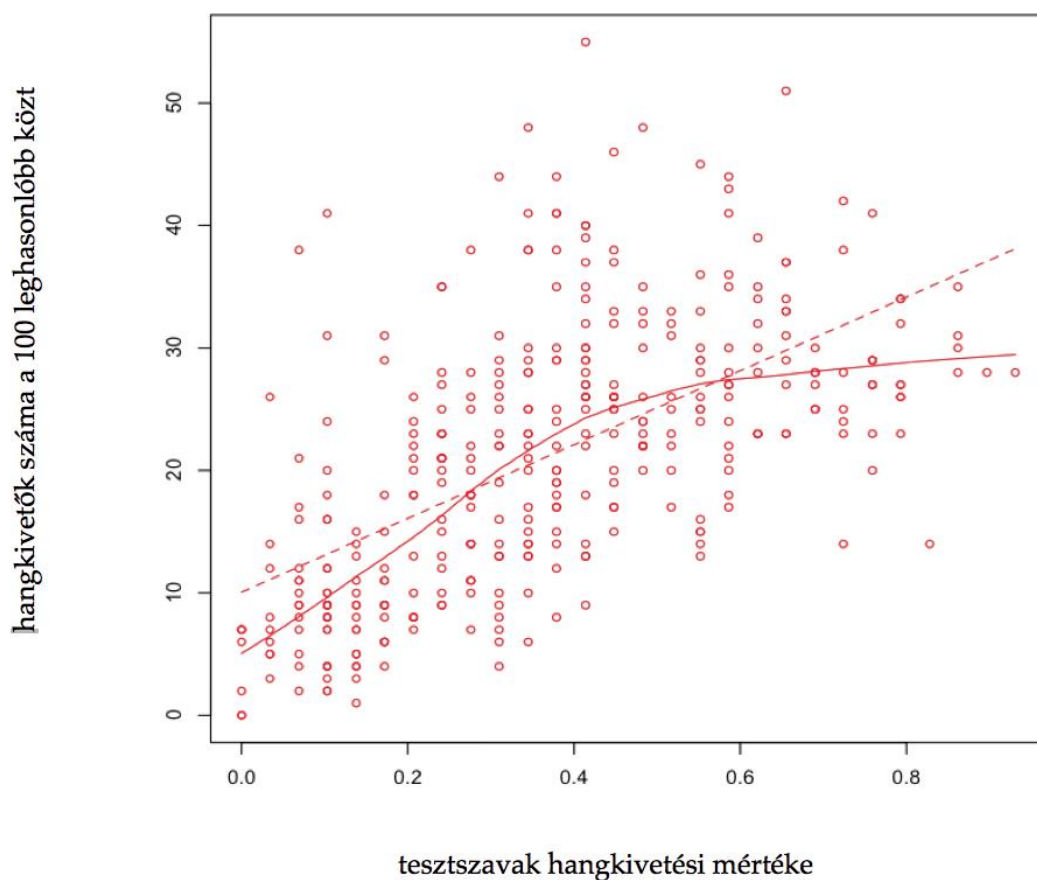
Az egyes algoritmusok eredményei a tízszeres keresztellenőrzésben 1078 hangkivető és 48389 nem hangkivető főnév besorolása alapján

## 7. Az analógiás források kiválasztásában szerepet játszó tényezők mérése CVCVC szerkezetű álszavakkal

Nyelvi tesztben azt vizsgáltam, hogy **eltérő pozíciókban a fonémák mennyiben járulnak hozzá két szó hasonlóságának a megítéléséhez**, abból a feltételezésből kiindulva, hogy minél jobban hasonlít egy álszó egy valódi szóhoz viselkedésében, annál közelebbi lesz ahhoz alaki hasonlóságuk alapján. A vizsgálatnak ezen túl az is célja volt, hogy kimutassam, milyen tényezők befolyásolják együttesen egy új nyelvi elem viselkedését, amelyekből a legjelentősebbnek a szerkezetileg leghasonlóbb szavak analógiás hatása bizonyult. Nyelvi tesztet 91 CVCVC szerkezetű hangkivető főnév alapján végeztem el összesen 116 magyar anyanyelvi beszélővel.

Nyelvi tesztet több olyan feltételezést is igazolt, amelyre már korábban is támaszkodtam elemzéseimben, és amelyeket Lukács (2002) is felvázolt

iránymutató munkájában. Ezek közül a legfontosabb, hogy az **egyes fonémapozíciókban megfigyelhető hasonlóságok és eltérések különböző fontossággal bírnak szavak összevetése esetén**. Ezt abból láthattuk, hogy minél inkább balra változtattam meg egy fonémát egy CVCVC szerkezetű álszón belül, annál hasonlóbb volt viselkedése (hangkivetési mértéke) ahhoz a szóhoz, amelyből létrehoztam. A fonémapozíció hatásától függetlenül megfigyelhettük, hogy az eredeti szó befolyásán túl leginkább az új szó egyedi viszonyrendszerének van szerepe hangkivetési mértékének alakulásában. Azaz a résztvevők nem feltétlenül az eredeti szót használták analógiás forrásnak az álszó alakjainak kiválasztásában, hanem az álszót már önállóan értékelték, és ennek megfelelően határozták meg viselkedését. Ebben legnagyobb szerepe a szerkezetileg is hasonló szavaknak van, amit kiegészít a vége alapján legközelebbinek számító leghasonlóbb szó hatása is.



A komplex tengelymérték alapján meghatározott leghasonlóbb hangkivető szavak számának és a tesztszavak hangkivetési mértékének együttjárása

lecserélt fonéma	hangkivetés mértéke	szignifikáns eltérések	szórás	min. hangkivetés	max. hangkivetés	leginkább hangkivető szavak
1. fonéma	44,2%	> 3. fonéma ** > 5. fonéma ***	21,9	0%	93,1%	lücsök, rucsok, böcök, pürök, dücsök
2. fonéma	41,5%	> 5. fonéma ***	20,1	0%	86,2%	pücsök, vücsök, vocok
3. fonéma	39,2%	> 5. fonéma ***	22,1	0%	79,3%	vöcsök, rücsök, tücsök, tüsszök, surom
5. fonéma	22,5%		13,5	0%	58,6%	sulyog, hurocs, bögl

A fonémapozíció hatása a hangkivetés mértékére

\*\* =  $p < 0,01$

\*\*\* =  $p < 0,001$

## 8. Rung Andrásnak a témához kapcsolódó publikációi

Fűköh Borbálával: Az -esz és az -er végű becézett szóalakokról (2005).  
*Nyelvtudomány* I. 115–130.

Determining word similarity in the Hungarian language (2008). In: Kálmán László (szerk.): *Papers from the Mókus Conference*. Tinta Kiadó. Budapest. 112–118.

- Szóhasonlóság mérése analógiás megközelítésben (2009). In: Tanács Attila–Szauter Dóra–Vincze Veronika (szerk.): VI. Magyar Számítógépes nyelvészeti konferencia. MSZNY 2009. Szegedi Tudományegyetem. Szeged. 104–113.
- Kálmán Lászlóval: Klaszterek helyett prototípusok (2010). In: Tanács Attila–Csendes Dóra (szerk.): *Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2010)*. 325–332.

## 9. Bibliográfia

- Blevins James P.–Blevins, Juliette (szerk.) 2009.** *Analogy in Grammar: Form and Acquisition*. Oxford University Press. Oxford.
- Bybee, Joan L. 2001.** *Phonology and Language Use*. Cambridge University Press. Cambridge.
- Bybee, Joan L. 2010.** *Language, Usage and Cognition*. Cambridge University Press. Cambridge.
- Chandler, Steve 2002.** Skousen's analogical approach as an exemplar-based model of categorization. In: Skousen, Royal–Lonsdale, Deryle–Parkinson, Dilworth B. (szerk.): *Analogical Modeling. An exemplar-based approach to language*. John Benjamins. Amsterdam. 51–105.
- Daelemans, Walter–van den Bosch, Antal 2005.** *Memory-Based Language Processing*. Cambridge University Press. Cambridge.
- Eddington, David 2003.** Issues in modeling language processing analogically. *Lingua*. 114. 849–871.
- Fillmore, Charles J.–Kay, Paul. 1987.** *The goals of Construction Grammar*. Berkeley Cognitive Science Program Working Paper 50. University of California at Berkeley. Berkeley, CA.
- Frisch, Stefan A. 1996.** *Similarity and Frequency in Phonology*. (PhD-disszertáció) <http://www.cas.usf.edu/~frisch/Frisch96.pdf> (2010.07.01.)
- Fűkőh Borbála–Rung András 2005.** Az -esz és az -er végű becézett szóalakokról. *Nyelvtudomány* I. 115–130.
- Goldberg, Adele 1995.** *Constructions. A Construction Grammar approach to argument*



- structure*. University of Chicago Press. Chicago.
- Goldberg, Adele 2006.** *Constructions at Work. The Nature of Generalization in Language*. Oxford University Press. Oxford.
- Halácsy Péter–Kornai András–Németh László–Rung András–Szakadát István–Trón Viktor 2003.** A Szószablya projekt. In: Alexin Zoltán–Csendes Dóra (szerk.): *Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2003)*. 299.
- Halliday, Michael A. K. 1961.** Categories of the theory of grammar. *Word* 17. 241–292.
- Hare, Mary L.–Ford, Michael–Marslen-Wilson, William D. 2001.** Ambiguity and frequency effects in regular verb inflection. In: Bybee, Joan–Hopper, Paul (szerk.): *Frequency and the emergence of linguistic structure*. John Benjamins. Amsterdam. 181–200.
- Jurafsky, Daniel–Bell, Alan–Gregory, Michelle–Raymond, William D. 2001.** Probabilistic Relations between Words: Evidence from Reduction in Lexical Production. In: Bybee, Joan–Hopper, Paul (szerk.): *Frequency and the emergence of linguistic structure*. John Benjamins. Amsterdam. 229–254.
- Kálmán László 2008.** *A mögöttes és ami mögötte van*.  
<http://www.szv.hu/cikkek/a-mogottes-es-ami-mogotte-van> (2010.07.01.)
- Kraska-Szlenk, Iwona 2007.** *Analogy. The Relation between Lexicon and Grammar*. Lincom. München.
- Krott, Andrea 2009.** The role of analogy for compound words. In: Blevins, James–Blevins, Juliette (szerk.): *Analogy in Grammar: Form and Acquisition*. Oxford University Press. Oxford. 118–136.
- Lakoff, George 1987.** *Women, fire, and dangerous things*. University of Chicago Press. Chicago.
- Lukács Ágnes 2002.** *Alaktanilag kivételes tövek vizsgálata a magyarban. A leíró általánosítások mentális realitása*. (szakdolgozat)
- MacDonald, M. C. 1994.** Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes* 9: 157–201.
- Pinker, Stephen 1999.** *Words and Rules*. New York. Basic Books.
- 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.
- Rung, András 2008.** Determining word similarity in the Hungarian language. In:

- Kálmán László (szerk.): *Papers from the Mókus Conference*. Tinta Kiadó. Budapest. 112–118.
- Rung András 2009.** Szóhasznosság mérése analógias megközelítésben. In: Tanács Attila–Szauter Dóra–Vincze Veronika (szerk.): *VI. magyar számítógépes nyelvészeti konferencia. MSZNY 2009*. Szegedi Tudományegyetem. Szeged. 104–113.
- Sinclair, John 1991.** *Corpus, concordance, collocation*. Oxford University Press. Oxford.
- Skousen, Royal 1989.** *Analogical Modeling of Language*. Kluwer Academic Publisher. Dordrecht.
- Skousen, Royal–Lonsdale, Deryle–Parkinson, Dilworth B. (szerk.) 2002.** *Analogical Modeling*. John Benjamin. Amsterdam.
- Trón, Viktor–Halácsy, Péter–Rebrus, Péter–Rung, András–Vajda, Péter–Simon, Eszter 2006.** Morphdb.hu: Hungarian lexical database and morphological grammar. In: *Proceedings of 5th International Conference on Language Resources and Evaluation. ELRA*. 1670–1673.
- Taylor, John R. 1995.** *Linguistic categorization prototypes in linguistic theory* (2. kiadás). Clarendon Press. Oxford.
- Ullman, M. T. 1999.** Acceptability ratings of regular and irregular past tense forms: Evidence for a dual system model of language. from word frequency and phonological neighborhood effects. *Language and cognitive processes* 14: 47–67.