

---

## Tanulmányok

---

# Új klasszifikációs módszerek alkalmazása a kétnyelvűség és az etnikai identitás kutatásában\*

### Vargha András

egyetemi tanár,  
a Károli Gáspár Református  
Egyetem Pszichológiai Intéze-  
tének intézetvezető-helyettese,  
az Eötvös Loránd Tudomány-  
egyetem professor emeritusa  
E-mail: [vargha.andras@kre.hu](mailto:vargha.andras@kre.hu)

### Borbély Anna,

az MTA Nyelvtudományi  
Intézet tudományos  
főmunkatársa  
E-mail:  
[borbely.anna@nytud.mta.hu](mailto:borbely.anna@nytud.mta.hu)

Jelen tanulmány fő célja, hogy bemutassa a modern személyorientált klasszifikációs módszerek (a klaszteranalízis) alkalmazásának nyelvészeti hasznosságát. Ehhez a szerzők hat közösség, a magyarországi beások, romák, németek, románok, szerbek és szlovákok kétnyelvűségének kutatásából használnak fel egy arra alkalmas adatbázist. Feladatuk azt tűzik ki, hogy azonosítsanak olyan homogén típusokat, amelyek a nyelvcsere folyamatában a saját nyelvükön még megszólalni tudó magyarországi nemzetiségi kétnyelvű felnőtt személyeket jellemzik. Az elemzéseket a következő négy változó segítségével végzik: kisebbségi etnikai identitás; kisebbségi nyelvvel kapcsolatos attitűd; kisebbségi nyelvhasználat a családtagokkal; kisebbségi nyelvtudás szintje. A tanulmányban feltárt tíz klaszter a nyelvcsere nem tisztán lineáris folyamatában tíz létező beszélőtípust azonosít. Az eredmények jelzik, hogy a nyelvcsere több elkülöníthető összetevője van, amelyeket célzottan befolyásolva esély lehet a fenntartható kétnyelvűség megteremtésére.

#### TÁRGYSZÓ:

Kétnyelvűség.  
Etnikai identitás.  
Klaszteranalízis.

DOI: [10.20311/stat2017.08-09.hu0805](https://doi.org/10.20311/stat2017.08-09.hu0805)

---

\* Az NKFP 5/126/2001. pályázat támogatásával készült kutatás konzorciumvezetője *Bartha Csilla*, téma-vezetője *Borbély Anna* volt. Az adatok statisztikai feldolgozását a jelen cikk első szerzője végezte. A tanulmány a „NyelvEsély” MTA-NYTI (2016–2020) pályázat időszakában készült.

A 2011. évi magyarországi népszámlálás adatai szerint Magyarország népességének 6,5 százaléka vallotta magát valamelyik hivatalosan elismert nemzetiséghez tartozónak ([http://www.ksh.hu/docs/hun/xftp/idoszaki/nepsz2011/nepsz\\_orosz\\_2011.pdf](http://www.ksh.hu/docs/hun/xftp/idoszaki/nepsz2011/nepsz_orosz_2011.pdf)), vagyis jelenleg a népesség 6,5 százaléka tartja magát nem magyar etnikumúnak, saját nyelvvel és kultúrával. Ezek a honfitársaink egyben a magyarországi kétnyelvűek táborát is erősítik. Miközben a kétnyelvűség nyilvánvalóan nagy hasznára van a magyar társadalomnak mind kulturális, mind politikai, mind gazdasági-kereskedelmi szempontból, számos tudományos kutatás jelzi, hogy nemzeti kisebbségeink gyorsuló ütemben vesznek el saját nyelvüket és etnikai identitásukat, asszimilálódnak a többségi nemzethez. A témával foglalkozó szakirodalom azt a jelenséget, amikor egy beszélő közösség saját nyelvének használatáról áttér egy másik, leggyakrabban a többségi nyelv használatára, nyelvcsere néven nevezi (lásd például *Gal* [1979], *Borbély* [2001]).

Jelen tanulmány hat magyarországi honos kisebbségi (beás, roma, német, román, szerb és szlovák) közösségben vizsgálja a kétnyelvűséget és a kisebbségi identitást. Ezek a közösségek az asszimiláció és a nyelvcsere folyamatát élik át, melyen belül bizonyos egyének még kétnyelvűek, míg mások már magyar egynyelvűek. Tanulmányunk fókuszában a fenntartható kétnyelvűség áll. Ez „a kétnyelvűségnek az a típusa, amely arra utal, hogy a kétnyelvűség nem egyszerűen stabil, tehát egy hosszantartó, statikus, két nyelvet érintő jelenség, hanem sokkal inkább (az A vagy a B nyelv dominanciájában) változatos, folyamatosan és dinamikusan változó nyelvi jelenség, amely (az egyének és közösségek életében) csak bizonyos feltételek mellett valósulhat meg” (*Borbély* [2014] 92. old.). Empirikus vizsgálatok alapján, mintázatfeltáró statisztikai módszerek segítségével megpróbáljuk azonosítani a nyelvcsere és az etnikai identitásra jellemző típusokat.

## 1. Vizsgálat

A következőkben bemutatjuk vizsgálatunk módszerét, az elemzésbe bevont változókat és személyeket, továbbá kitérünk a statisztikai elemzéseinkre is.

## 1.1. Vizsgálati módszer

Az NKFP 5/126/2001. pályázat támogatásával végzett vizsgálat (MaBiLing kutatás) hat magyarországi kétnyelvű (honos nemzetiségi) közösség egy-egy településén<sup>1</sup>, összesen 421 adatközlővel<sup>2</sup> készült. A szociolingvisztikai összehasonlító kutatás terepmunkája 2003 ősze és 2004 tavasza között zajlott. Ennek keretében egy nyelvhasználati és egy attitűdkérdőív segítségével gyűjtöttek adatokat (vö. *Bartha* [2003]) – a szlovák kutatócsoport egy tagjától eltekintve – e közösségekhez tartozó olyan személyek, akik anyanyelvi szinten beszélték a vizsgált nemzetiség nyelvét. Foglalkozásukat tekintve nyelvészek, egyetemi oktatók, középiskolai és általános iskolai tanárok, illetve egyetemi hallgatók voltak. Jelen tanulmányban nem részletezett kérdőívben rákérdeztünk többek között a kisebbségi és a többségi nyelvhasználat egyes összetevőire különböző nyelvhasználati szintereken (család, vallás, munkahely stb.), a nyelvtudásra, az identitásra, valamint a nyelvi attitűdre is.

## 1.2. Változók

Kutatásunkban a fenntartható kétnyelvűséggel és az etnikai identitással összefüggésben a következő négy nyelvészociológiai változót vontuk be a mintázatfeltáró elemzésekbe:

*Ident*: kisebbségi identitás (a nemzetiségre és az anyanyelvre vonatkozó szubjektív vélemény alapján);

*Attitűd*: kisebbségi nyelvvel összefüggő esztétikai és emocionális attitűd (a saját kisebbségi nyelv szépségére és szerethetőségére vonatkozó válaszok alapján);

*Család*: a családtagokkal való kisebbségi nyelvhasználat gyakorisága (a szülő, testvér, házastárs, gyermek és unoka vonatkozásában kapott adatok átlagolásával);

*KisNyv*: a kisebbségi nyelvtudás szintje önértékelés alapján (a területi/kontaktus-nyelvváltozat és az esetlegesen létező irodalmi nyelv standard tudásának átlagolásával).

<sup>1</sup> A vizsgált települések a következők voltak: Mánfa és Bogyiszló (beások), Tarján (németek), Mezőtúr (romák), Kétegyháza (románok), Pomáz (szerbek), illetve Tótkomlós (szlovákok). Ezúton is köszönjük az adatgyűjtést a következő kutatóknak és terepmunkásoknak: *Orsós Anna* és *Egregyi Borbála* (beások), *Erb Maria* (németek), *Kiss Andrea* (romák), *Abrudán Mária* (románok), *Szimics Milosné* (szerbek), *Hornokné Uhrin Erzsébet*, *Tóth Sándor*, *Tuska Tünde*, valamint *Zsilák Mária* (szlovákok).

<sup>2</sup> Közösségenként 70-71 adatközlőt kérdeztünk meg. A hat település mintáját egységesen (és a közösségek körülményeinek figyelembevételével) az életkor, a nem és az iskolázottság lehetőség szerint minden szintjének arányos képviselőivel alakítottuk ki.

E változóegyüttes tartalmi koherenciáját és pszichometriai megbízhatóságának elfogadható szintjét a Cronbach-féle  $\alpha$  0,68-as értéke jelezte (vö. *Cohen* [1977], [1992]).

### 1.3. Személyek

A megkérdezettek a vizsgálatba beleegyező olyan felnőtt (a vizsgálat idején legalább 20 éves) személyek voltak, akik a korábban jelzett hat kisebbséghez tartoztak, s önként vállalták az interjút. Nemi, életkori és iskolázottsági alapstatisztikáikat az 1. táblázat foglalja össze. A hat minta a nemi megoszlás és az életkor tekintetében meglehetősen hasonlóknak tekinthető, az iskolázottság szempontjából viszont erősen különbözőnek. A beás és a roma mintában ugyanis csak elhanyagolható arányt képviseltek a felsőfokú végzettségűek (szemben a többi mintával, ahol 14-15 százalékot), viszont voltak az általános iskola nyolc osztályánál alacsonyabb végzettségűek is (a roma mintában számottevően [40 százalékban]).

1. táblázat

*A nem, az életkor és az iskolázottság szintjének alapstatisztikái a hat kisebbségi mintában*

Kisebbség	Elemszám	Férfi-nő arány (%)	Korátlag (év)	Korterjedelem (év)	A négy iskolázottsági szint aránya (%)
Beás	70	50-50	46,2	21–81	1,7-94,8-3,7-0,0
Roma	70	49-51	46,5	20–84	40,0-45,7-12,9-1,4
Német	70	50-50	51,0	23–80	0,0-47,1-38,6-14,3
Román	71	51-49	52,1	28–80	0,0-42,3-42,3-15,5
Szerb	70	50-50	49,9	21–81	0,0-42,9-42,9-14,3
Szlovák	70	49-51	51,3	21–76	0,0-44,3-41,4-14,3

*Megjegyzés.* Itt és a továbbiakban a táblázatok és az ábrák adatai a 2003–2004-ben zajlott felmérés eredményei. A jelen táblázat utolsó oszlopában felsorolt százalékos adatok jelentése: első szám – az iskolázottság szintje nyolc általános iskolai osztálynál alacsonyabb; második szám – általános iskola nyolc osztály; harmadik szám – középiskola; negyedik szám – főiskola vagy egyetem.

### 1.4. Statisztikai elemzések

Vizsgálatunk egyik fő célja, hogy feltárjuk a nyelvcsere és az identitás asszimilációs folyamatának különböző típusait, mely segítséget nyújthat ahhoz, hogy ismereteket szerezzünk a fenntartható kétnyelvűségről és a kisebbségi identitás megőrzésének lehetőségéről. Ezeket a típusokat a 421 személy adatai alapján, a már ismertett négy nyelvészociológiai változó segítségével határooltuk le klaszteranalízissel. A sta-

tisztikai elemzéseket ROPstat programcsomaggal (Vargha [2008], [2016]; Vargha–Torma–Bergman [2015]) végeztük.

Tanulmányunk másik fő célkitűzése, hogy bemutassuk a klaszteranalízis korrekt végrehajtásának és értelmezésének módját, rávilágítva azokra a kényes pontokra, ahol a klaszterek számáról, a klasszifikációs struktúra jóságáról és szignifikanciájáról dönthetünk.

## 2. Eredmények

A következőkben kitérünk a klaszterstruktúra létrehozásának, belső és külső validálásának, valamint szakmai értelmezésének részleteire.

### 2.1. A klaszterstruktúra létrehozása

A négy nyelvészociológiai változó (*Ident*, *Attitűd*, *Család*, *KisNyt*) kapcsolatának erőssége a Spearman-féle rangkorreláció alapján 0,13–0,47 volt. A leggyengébb rangkorrelációt az *Ident* és a *KisNyt* ( $r = 0,13$ ), illetve az *Attitűd* és a *KisNyt* ( $r = 0,24$ ), a legerősebbet ( $r = 0,47$ ) a *Család* és a *KisNyt* között kaptuk. Az egyes változók másik három által magyarázott varianciaaránya rendre 24, 24, 40, illetve 24 százalék volt. Mindez arra utal, hogy miközben a négy változó – a belső konzisztenciát kifejező Cronbach-féle  $\alpha$  alapján – tartalmilag koherens együttest képez, vagyis egyazon jelenségkörre fed le, egymással csak gyenge-közepes szintű kapcsolatban van. Ez azt jelenti, hogy mindegyik hordoz számottevő információt, amit a többi változó nem fed le. Ennél fogva a változóegyüttes optimális inputja a klasszifikációs elemzésünknek.

Először reziduálanalízissel ellenőriztük, hogy nincsenek-e kiugró megfigyelési adatok, amelyek torzítanák a klaszterstruktúrát. Mivel outliereket nem találtunk, a 421 személyre HKA-t (hierarchikus klaszteranalízist) végeztünk Ward-féle hierarchikus összevonási módszerrel, az input változókat z-standardizálással közös léptékre hozva, majd a megfigyelési egységek átlagos négyzetes euklideszi távolságát meghatározva (vö. *Bergman–Magnusson–El-Khoury* [2003] 4. fejezet). A ROPstat programmal a következő klaszteradekvációs mutatókat mértük (vö. *Vargha–Bergman* [2015], *Vargha–Bergman–Takács* [2016]):

- EESS%: magyarázott varianciaarány;
- PB: pontbiseriális index, ami azt fejezi ki, hogy az egyazon klaszterbe tartozó esetek mennyivel vannak közelebb egymáshoz, mint a különböző(k)be tartozók;

- XBmod: módosított Xie-Beni index. Ez azt mutatja, hogy mennyivel kisebb az átlagos távolság a saját klaszter középpontjától, mint az egymáshoz legközelebbi két klaszter távolsága;
- SC: Silhouette-együttható, ami arra utal, hogy az esetek mennyivel vannak közelebb saját klasztercentrumukhoz, mint a legközelebbi idegen centrumhoz;
- HC-átlag: a klaszterhomogenitási együtthatók klasztermérettel súlyozott átlaga.

Az adekvációs mutatókat 5–12 klaszterszám esetén a 2. táblázat foglalja össze. A legmegbízhatóbb struktúrának a tízklaszteres megoldás tűnik, mert ennél a klaszterek homogenitását mérő két legfontosabb mutató közül az EESS% a kívánatos 65 százaléknál felel meg (70,47%), a HC-átlag pedig 1 alatt (0,604) úgy, hogy egyetlen klaszter HC-értéke<sup>3</sup> sem haladja meg az 1-et. A klaszterek szeparációját mérő, másodlagos fontosságú mutatók közül a PB eléri a kívánatos 0,300-es szintet, továbbá XBmod és SC sincs nagyon messze az elvárt 0,500-es szinttől.

2. táblázat

*A hierarchikus klaszteranalízis 5–12 klaszterre vonatkozó adekvációs mutatói*

Klaszterek száma	EESS%	PB	XBmod	SC	HC-átlag	HCmin-HCmax
12	74,08	0,288	0,320	0,471	0,533	0,18-0,88
11	72,40	0,288	0,276	0,469	0,565	0,18-1,00
<b>10</b>	<b>70,47</b>	<b>0,317</b>	<b>0,327</b>	<b>0,449</b>	<b>0,604</b>	<b>0,40-1,00</b>
9	68,54	0,319	0,283	0,444	0,641	0,42-1,00
8	65,40	0,346	0,211	0,446	0,703	0,42-1,00
<b>7</b>	62,11	0,389	0,263	0,478	0,768	0,50-1,00
6	58,60	0,382	0,195	0,478	0,838	0,50-1,33
5	54,72	0,377	0,119	0,462	0,915	0,67-1,33

*Megjegyzés.* A táblázatban félkövérrel emeltük ki a legjobb hierarchikus struktúrát.

A HKA tízklaszteres megoldásán próbáltunk relokációval (KKA-val [k-osztópontú klaszteranalízissel]) javítani, ami sikeresnek bizonyult, mert az EESS% értéke nagyobb lett, a klaszterek HC-átlaggal mért heterogenitása pedig lecsökkent. A 3. táblázatban látható, hogy a KKA-klaszterstruktúrához tartozó adekvációs mutatók értéke jónak mondható: EESS% (74,48%) meghaladja a 65 százalékat, PB pedig

<sup>3</sup> A HC homogenitási együttható a klaszterbeli egyedek páronkénti távolságainak átlaga, így minél kisebb egy klaszter HC-értéke, annál homogénebb, mert annál közelebb vannak az egyedei egymáshoz.

a 0,300-es szintet, XBmod megközelíti a 0,500-et, SC 0,600 körüli, és HC-átlag (0,523) is 1 alatt van.

3. táblázat

*A hierarchikus és a nem hierarchikus (k-osztópontú) klaszteranalízis tíz klaszterre vonatkozó adekvációs mutatói*

Klaszteranalízis típusa	EESS%	PB	XBmod	SC	HC-átlag	HCmin-HCmax
Hierarchikus	70,47	0,317	0,327	0,449	0,604	0,40-1,00
K-osztópontú	74,48	0,325	0,438	0,583	0,523	0,26-0,86

*Megjegyzés.* A HC-mutatók kivételével mindig a mutatók magasabb szintje jelez jobb struktúrát.

## 2.2. A belső validitásra vonatkozó eredmények

Egy klaszterstruktúra validálásán olyan statisztikai vizsgálatokat értünk, amelyek hitelesítik az adott struktúrát. Ez történhet hipotézisvizsgálattal, illetve bizonyos adekvációs mutatók kiszámításával. A belső validitás során olyan elemzéseket végzünk, amelyekben csupán a klaszteranalízisbe bevont változók és a kapott struktúra összetevői szerepelnek. Mivel a klaszteradekvációs mutatók (EESS%, HC-átlag stb.) értékszintje erősen függ az alkalmazott változók és a klaszterek számától, egy klaszterstruktúra jóságát célszerű annak alapján megítélni, hogy adekvációs mutatói szignifikánsan és számottevő mértékben jobbak-e, mint egy véletlen változók alapján, ugyanakkora mintán, ugyanannyi input változóval és klaszterrel végzett hasonló elemzés adekvációs mutatói (Vargha–Bergman–Takács [2016]). Jelen esetben huszonöt véletlen adatgenerálást végeztünk 1. az input változók értékeinek random permutálásával; 2. többdimenziós folytonos egyenletes eloszlásból; és 3. többdimenziós normális eloszlásból, mindhárom esetben független random input változókat alkalmazva. Az adekvációs mutatók a tizklaszteres valódi adatokon alapuló struktúra esetében többnyire (az EESS%, az SC és a HC-átlag esetében mindig) ( $p < 0,001$  szinten) erősen szignifikánsan jobbak voltak, mint a véletlen adatokon alapuló struktúráknál.

A szignifikancia mellett Vargha–Bergman–Takács [2016] a MORI-t (measure of relative improvement – relatív javulási mutató) javasolta annak mérésére, hogy a valódi adatokon nyugvó klaszteradekvációs mutatók milyen mértékben jobbak, mint a véletlen mintákon alapulók. A MORI javasolt értelmezési kritériumai a következők: az alacsony szintű (minimálisan elfogadható) belső validitás küszöbértéke 0,15, a közepes szintűé 0,30, a magas szintűé 0,50.

4. táblázat

*A klaszterstruktúra belső validitásának vizsgálata véletlen kontrollváltozók segítségével*

Valódi minta/MORI	Adekvációs mutató, illetve javulásának/romlásának mértéke				
	EESS%	PB	XBmod	SC	HC-átlag
Valódi minta	74,48	0,33	0,44	0,58	0,52
MORI – kontroll: input változók értékeinek random permutációja	0,20	-0,05	0,12	0,07	0,20
MORI – kontroll: független, random, folytonos egyenletes eloszlású változók	0,20	-0,09	-0,29	0,07	0,20
MORI – kontroll: független, random, normális eloszlású változók	0,34	0,04	0,04	0,16	0,34

*Megjegyzés.* A táblázat huszonöt véletlen adatgeneráláson alapul.

A 4. táblázat a valódi minta adekvációs mutatóit és a MORI értékeit mutatja be az öt adekvációs mutató esetén. A korábban leírt értelmezési konvenció alapján megállapíthatjuk, hogy a legfontosabb mutatók, az EESS% és a HC-átlag tekintetében közepes, PB, XBmod és SC esetében pedig gyenge a valódi adatokon alapuló struktúra fölénye a véletlen változókat használóhoz képest. Ez azt jelenti, hogy a kapott klaszterstruktúra a homogenitás szempontjából közepes belső validitású, következésképpen szakmai értelmezésre alkalmas klaszterekhez jutottunk (bár elképzelhető, hogy az nem minden klaszter esetében tehető meg). A PB, az XBmod és az SC alacsony MORI-értékei arra utalnak, hogy a szeparáció nem tökéletes, vagyis a tíz klaszter között lehetnek olyanok, amelyek több – de nem minden – változó tekintetében hasonlítanak egymásra. Ez persze ennyire sok klaszter esetén természetes.

A ROPstat legújabb, 2017. májusi frissítése lehetővé tette, hogy a 4. táblázatban feltüntetett három random kontroll eloszláson kívül egy új, az előzőknél szigorúbb kritérium szerint is tesztelhesük a klaszterstruktúra jóságát. Az ún. „random korreláló normális kontroll” opció olyan többdimenziós normális eloszláson alapul, ahol az összetevők közötti korrelációk megegyeznek az input változók közöttiekkel. Az ezáltal kapott MORI-értékek minden adekvációs mutató esetében szignifikánsan nagyobbak voltak 0-nál, közülük a legfontosabb EESS%-nál és HC-átlagnál pedig elérték a 0,20-as szintet is.

Annak további megerősítése céljából, hogy ez a tízklaszteres megoldás kiemelkedik-e a többi lehetséges struktúra közül, megvizsgáltuk a legjobb nyolc- és kilencklaszteres struktúrák belső validitását is. Ezek MORI-értékei az EESS% és a HC-átlag esetében mind kisebbek voltak (0,34 helyett 0,29, illetve 0,32 legmagasabb értékkel), mint a tízklaszteresé.



## 2.3. A klaszterek értelmezése

A kapott tízklaszteres megoldás értelmezéséhez elkészítettük az input változók standardizált klaszterátlagainak, az ún. klasztercentroidoknak a grafikonját, a jobb áttekinthetőség érdekében külön ábrázolva a magyar (KL1–KL5; lásd az 5. táblázatot és az 1. ábrát), illetve a kisebbségi nyelvhasználat és identitás dominanciájára utaló klasztereket (KL6–KL10; lásd a 2. ábrát). Megjegyezzük, hogy ezeket a többdimenziós skálázás módszerével is megfelelően lehet ábrázolni (vö. *Takács* [2013]).

5. táblázat

*A klaszterek elnevezése, rövid címkeje, gyakorisága, megoszlása, valamint homogenitásának szintje*

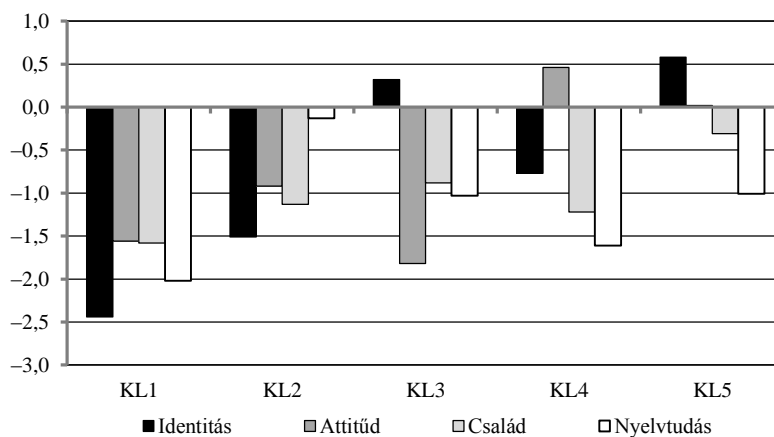
Klaszter	Elnevezés és rövid címke	Gyakoriság	Megoszlás (%)	HC
KL1	Befejeződött nyelvcsere, teljes magyar asszimiláció: K1M++	15	3,6	0,62
KL2	Magyar asszimiláció némi kisebbségi nyelvtudással: K2Nyt+	46	10,9	0,86
KL3	Magyar asszimiláció némi etnikai identitással: K3Id+	24	5,7	0,68
KL4	Magyar asszimiláció némi pozitív kisebbségi nyelvi attitűddel: K4Att+	20	4,8	0,74
KL5	Átmeneti állapot csekély etnikai identitással, gyenge kisebbségi nyelvtudással: K5Nyt–	44	10,5	0,42
KL6	Megőrzött kisebbségi nyelvhasználat, de elvesztett etnikai identitás: K6Id–	25	5,9	0,68
KL7	Megőrzött kisebbségi nyelv és etnikai identitás, de negatívba forduló attitűd a kisebbségi nyelvvel szemben: K7Att–	61	14,5	0,59
KL8	Megőrzött kisebbségi nyelv és etnikai identitás, de megszűnő nyelvhasználat a családban: K8Cs–	67	15,9	0,43
KL9	Megőrzött kisebbségi nyelvhasználat és etnikai identitás romló nyelvtudással: K9Nyt–	50	11,9	0,43
KL10	Megőrzött kisebbségi nyelv és etnikai identitás: K10Kis+	69	16,4	0,26
	<i>Összesen</i>	<i>421</i>	<i>100,0</i>	<i>–</i>

*Megjegyzés.* Itt és a továbbiakban esetenként kerekítési hiba okozza az összeg eltérését a 100,0 százaléktól.

Az 1. ábrán látható klaszterprofilok olyan mintázatot tükröznek, amelyek szerint a nyelvcsere már jócskán előrehaladott állapotban van. Közülük a KL1 klaszter mintázata minden nyelvészociológiai összetevő esetében magyar dominanciájú, a KL2–KL4-nél viszont előfordul egy-egy olyan jellemző (KL2 – kisebbségi nyelvtudás, KL3 – identitás, KL4 – saját kisebbség nyelvével összefüggő attitűd), mely a magyar

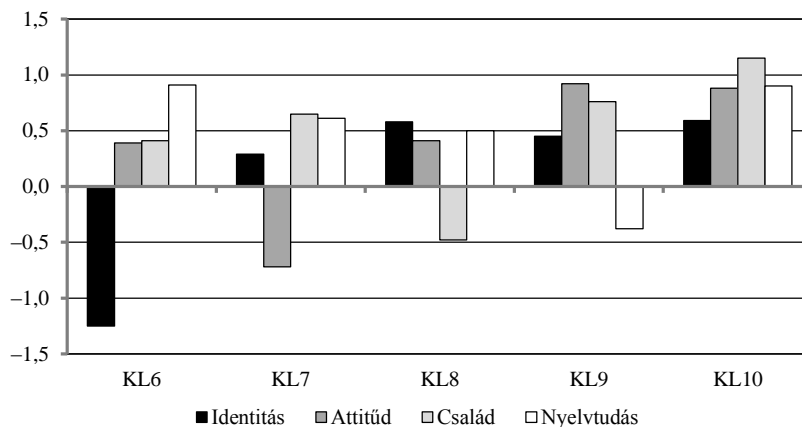
nyelvhasználat még nem teljesen uralkodó szerepére, illetve a kisebbségi nyelv enyhe hangsúlyára utal a megkérdezettek életében. A KL5-be tartozók bár döntően a magyar nyelvet használják, még őrzik kisebbségi identitásukat, amit a saját kisebbségi nyelvükkel összefüggő attitűdjük is jelez.

1. ábra. A nyelvhasználat és az identitás szempontjából magyar dominanciájú klaszterek standardizált átlagainak mintázata a négy input változó tekintetében



Megjegyzés. Itt és a továbbiakban a klaszterek megnevezéséért lásd az 5. táblázatot. Az alacsonyabb értékek magasabb fokú magyar asszimilációra utalnak.

2. ábra. A nyelvhasználat és az identitás szempontjából kisebbségi dominanciájú klaszterek standardizált átlagainak mintázata a négy input változó tekintetében



A 2. ábrán olyan klaszterek láthatók, amelyek esetében a kisebbségi nyelvhasználat és az etnikai identitás négy nyelvészeti összetevője közül legalább

három kisebbségi dominanciára utal. Közülük KL10 mutatja a saját kisebbséghez való tartozás legerősebb (egyöntetű) kifejezését, de KL6, KL7, KL8 és KL9 is csak egyetlen tényező (rendre az etnikai identitás, a kisebbségi nyelvvel szembeni attitűd, a családi nyelvhasználat, illetve a kisebbségi nyelv tudása) tekintetében „húz” a többségi magyar asszimiláció felé.

A klaszterstruktúra stabilitásának tesztelésére, a ROPstat centroid modulja segítségével összevetettük a tízklaszteres megoldást a hasonló módon kapott legjobb nyolc- és kilencklaszteressel. Az elemzés kimutatta, hogy a tízklaszteres struktúra hat klasztere (az első és az utolsó három) szinte teljes mértékben megjelent a nyolcklaszteresben is, nyolc klasztere pedig (az első és az utolsó négy) a kilencklaszteresben. Az utóbbi esetén még a legkevésbé összeillő klaszterek (a KL8 és párjának) távolsága is csak 0,014 volt, vagyis ezek centruma is csaknem egy pontba esett. Az egyetlen különbséget a kilenc- és a tízklaszteres megoldás között az jelentette, hogy KL4 és KL5 a kilencklaszteres megoldásban összevonódott. Ez azonban az identitásszint ellentétes irányultsága miatt (lásd az 1. ábrát) szakmailag elfogadhatatlan, ami azt is jelzi, hogy tíznél kevesebb klaszterrel a vizsgált kisebbségi minta nem írható le megfelelően, ismételten megerősítve a kapott tízklaszteres megoldás helyességét.

Az előbbi elemzésekből az a következtetés is levonható, hogy a nyelv- és az identitáscsere folyamatában a két szélsőséges pólushoz közeli állapotok a legstabilabbak, az azok közöttiek pedig a legkevésbé tartósak. A sokféle átmeneti típus feltárása és azonosítása jelzi a folyamat összetettségét, és különféle sikeres beavatkozásokra adhat lehetőséget a fenntartható kétnyelvűségi állapot megteremtése érdekében.

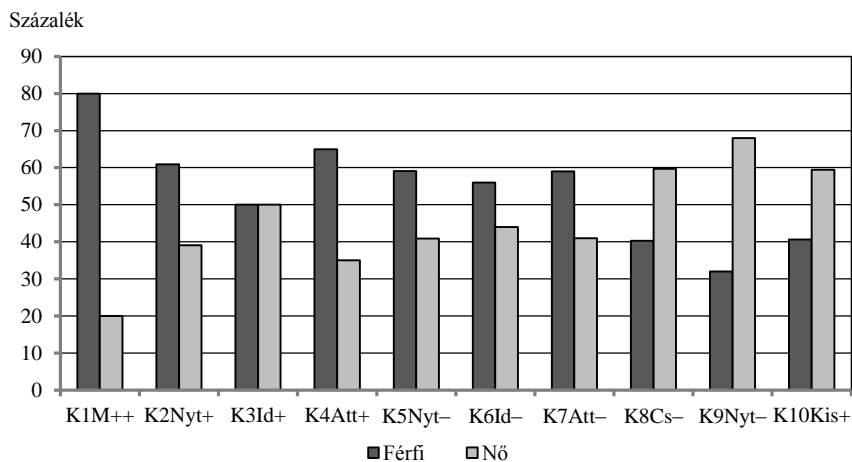
A klaszterek megoszlására vonatkozó alapadatokat tartalmazó 5. táblázat alapján megállapítható, hogy a domináns magyar nyelvhasználat és identitás klasztereibe a megkérdezettek mintegy 35, a kisebbségi klaszterekbe pedig 65 százaléka tartozik, vagyis tetemes részük még megőrizte etnikai identitását és kisebbségi nyelvhasználatát, fenntartva kétnyelvűségét.

## 2.4. A külső validitásra vonatkozó eredmények

A külső validitás során a klaszteranalízisbe be nem vont változók segítségével próbáljuk kimutatni egy klaszterstruktúra jóságát, szakmai értelmezhetőségét. Esetünkben ezek a következők: klaszterekre jellemző nemi arányok, továbbá ezek kapcsolata az életkorral, az iskolázottsággal és a hat kisebbségi csoporttal.

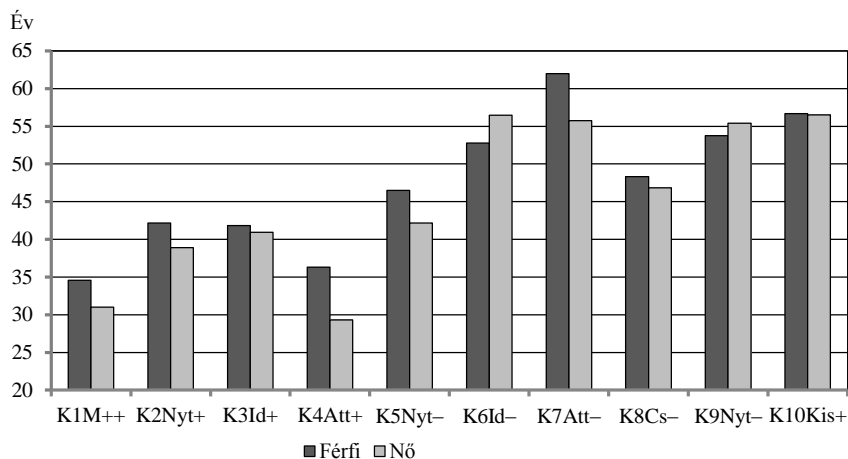
Az egyes klaszterekbe tartozók nemi összetételét a 3. ábra szemlélteti. Eszerint a KL1–KL5-ben egy kivétellel (KL3) a férfiak voltak többen (K1M++ esetében 80 százalékos arányban), a kisebbségi nyelvet és identitást legmarkánsabban megőrzők klasztereiben (KL8–KL10) viszont a nők. A kapcsolat erősen szignifikáns ( $\chi^2(9) = 24,604$ ,  $p = 0,0034$ , Cramér-féle  $V = 0,242$ ).

3. ábra. Nemi összetétel a klaszterekben



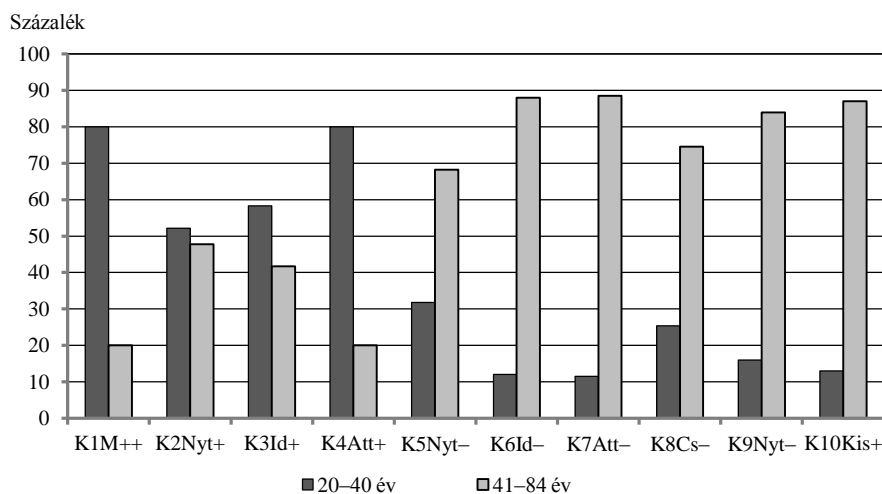
Az életkori átlagokat klaszterenként, nemenkénti bontásban a 4. ábra szemlélteti. Eszerint a magyar asszimilációval leginkább jellemezhető négy klaszter (KL1–KL4) korátlagai 30–45 év közé esnek, míg a kisebbségi nyelvüket és etnikai identitásukat továbbra is őrző személyek öt klaszterében (KL6–KL10) az átlagok 45 évnél magasabbak. Az utóbbiak közül az életkorátlag érthetően a KL8 klaszterben a legalacsonyabb, tehát a megkérdezettek azon csoportjában, ahol a családon belül már nem használják a kisebbségi nyelvet. Az életkori hatás erősen szignifikáns (varianciaanalízisben  $F(9; 401) = 14,352, p < 0,0001$ ). Figyelemre méltó, hogy az életkori átlagok mintázata férfiakkal és nőknél szinte teljesen megegyezik.

4. ábra. Életkori átlagok a klaszterekben

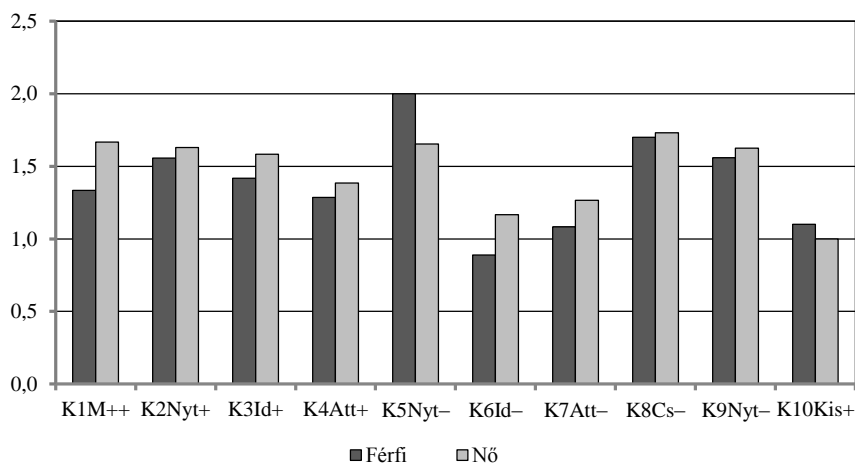


Az életkori hatást két korcsoport (a 20–40, illetve a 41–84 évesek) megoszlásán keresztül vizsgáltuk a tíz klaszterben. Az 5. ábra világosan jelzi az idősebbek erős dominanciáját a kisebbségi nyelvet megőrzők klasztereiben (KL6–KL10) és az átmeneti állapotot jellemző KL5-ben, valamint a fiatalabbakét a magyar asszimilációval érintett legtöbb klaszterben.

5. ábra. Életkori összetétel a klaszterekben



6. ábra. Iskolázottsági átlagok a klaszterekben nemi bontásban



Megjegyzés. 0 – az általános iskola nyolc osztályánál kevesebb; 1 – általános iskola nyolc osztály; 2 – középiskola; 3 – főiskola/egyetem.

A klaszterekre jellemző iskolázottsági átlagokat nemenkénti bontásban a 6. ábrán szemléltetjük. Az iskolázottsági hatás szignifikáns (robosztus varianciaanalízisben  $F(9; 62,3) = 5,926$ ,  $p = 0,0178$ ). A klasztereket páronként összehasonlítva, KL5 (K5Nyt–) iskolázottsági fölénye szignifikáns ( $p < 0,01$  szinten) a tőle jobbra elhelyezkedő KL6 (K6Id–), KL7 (K7Att–) és KL10 (K10Kis+) klaszterekkel szemben, valamint KL10 (K10Kis+) iskolázottsági hátránya az összes többi egyesített átlagával szemben (Scheffé-féle kontraszt módszerrel  $F(9; 389) = 1,951$ ,  $p = 0,0439$ ; vö. Scheffé [1959]). Az iskolázottsági átlagok mintázata – hasonlóan az életkori átlagokhoz – itt is szinte teljesen megegyezik a férfiaknál és a nőknél.

A kisebbségi csoportok százalékos megoszlási arányait az egyes klaszterekben a 6. táblázat tartalmazza. Mindegyik nemzetiség esetén dominál egy-egy klaszter. A szlovákok teszik ki a KL1 (K1M++) klaszter 40 százalékát, a németek a KL2 (K2Nyt+) 45,7 százalékát, a románok a KL3 (K3Id+) 50 százalékát. Ugyanakkor a romák alkotják a kisebbségi nyelvhasználatban kiemelkedő klaszterek közé tartozó KL10 (K10Kis+) 42 százalékát, a szerbek pedig a KL9 (K9Nyt–) 34 százalékát. Figyelmet érdemel a beások 60 százalékos részesedése is a KL6 (K6Id–) klaszterben.

6. táblázat

*A kisebbségi csoportok százalékos megoszlása az egyes klaszterekben*

Klaszter	Beás	Roma	Német	Román	Szerb	Szlovák	Összesen
K1M++	6,7	13,3	13,3	20,0	6,7	<b>40,0</b>	100,0
K2Nyt+	13,0	8,7	<b>45,7</b>	8,7	4,3	19,6	100,0
K3Id+	16,7	4,2	8,3	<b>50,0</b>	0,0	20,8	100,0
K4Att+	30,0	5,0	15,0	15,0	25,0	10,0	100,0
K5Nyt–	9,1	4,5	4,5	25	27,3	29,5	100,0
K6Id–	<b>60,0</b>	8,0	16,0	4,0	0,0	12,0	100,0
K7Att–	19,7	26,2	19,7	14,8	8,2	11,5	100,0
K8Cs–	13,4	11,9	23,9	16,4	14,9	19,4	100,0
K9Nyt–	8,0	10,0	10,0	28,0	<b>34,0</b>	10,0	100,0
K10Kis+	13,0	<b>42,0</b>	4,3	4,3	26,1	10,1	100,0
Összesen	16,6	16,6	16,6	16,9	16,6	16,6	100,0

*Megjegyzés.* A táblázatban félkövérrel jelöltük a 30 százalékot meghaladó arányokat.

A 7. táblázatban azt mutatjuk be, hogy a tíz klaszter miként oszlik meg egy-egy kisebbségi csoporton belül. A táblázat szerint a nyelvcseré előrehaladott állapotát képviselő KL1–KL5 a teljes mintában 35,5 százalékat képvisel. Ennél azonban érzékelhetően nagyobb az arányuk a szlovák (50,1%), a román (46,4%) és a német (43,0%) kisebbségek esetén. A románoknál viszont ugyanez az adat csak 14,3 százalék, vagyis e

nemzetiségi csoport nagy többségére (85,7 százalékára) még a romani nyelv és a roma identitás megőrzése jellemző. Az átlaghoz viszonyítva nagyobb mértékben őrzik kisebbségi nyelvüket és etnikai identitásukat a szerbek (71,4%) és a beások (70%) is.

7. táblázat

*A klaszterek százalékos megoszlása az egyes kisebbségi csoportokban*

Klaszter	Beás	Roma	Német	Román	Szerb	Szlovák	Összesen
K1M++	1,4	2,9	2,9	4,2	1,4	8,6	3,6
K2Nyt+	8,6	5,7	30,0	5,6	2,9	12,9	10,9
K3Id+	5,7	1,4	2,9	16,9	0,0	7,1	5,7
K4Att+	8,6	1,4	4,3	4,2	7,1	2,9	4,8
K5Nyt–	5,7	2,9	2,9	15,5	17,1	18,6	10,5
<i>Összesen</i>	<i>30,0</i>	<i>14,3</i>	<i>43,0</i>	<i>46,4</i>	<i>28,5</i>	<i>50,1</i>	<i>35,5</i>
K6Id–	21,4	2,9	5,7	1,4	0,0	4,3	5,9
K7Att–	17,1	22,9	17,1	12,7	7,1	10,0	14,5
K8Cs–	12,9	11,4	22,9	15,5	14,3	18,6	15,9
K9Nyt–	5,7	7,1	7,1	19,7	24,3	7,1	11,9
K10Kis+	12,9	41,4	4,3	4,2	25,7	10,0	16,4
<i>Összesen</i>	<i>70,0</i>	<i>85,7</i>	<i>57,1</i>	<i>53,5</i>	<i>71,4</i>	<i>50,0</i>	<i>64,6</i>

### 3. Összefoglalás

Jelen tanulmány fő célja az volt, hogy bemutassa a modern személyorientált klasszifikációs módszerek alkalmazásának nyelvészeti hasznosságát. Ehhez hat közösség, a magyarországi beások, romák, németek, románok, szerbek és szlovákok kétnyelvűségének kutatásából használtunk fel egy arra alkalmas adatbázist (*Bartha* [2003], *Borbély* [2014]). Feladatul azt tűztük ki, hogy azonosítsunk olyan homogén típusokat, amelyek a nyelvcseré folyamatában a saját nyelvükön még megszólalni tudó magyarországi nemzetiségi kétnyelvű felnőtt személyeket jellemzik.

A klaszteranalízis régóta kedvelt módszere az objektumok (személyek, tárgyak, változók stb.) homogén alcsoportokba (klaszterekbe) sorolásának (*Hartigan* [1975]). Az általunk alkalmazott klasszifikációs módszer újszerűsége, hogy speciális klaszteradekvációs mutatók (EESS%, PB stb.) segítségével képes volt olyan, mind a matematikai kritériumok, mind a szakmai szempontok alapján elfogadható, értelmes klaszterstruktúrát találni, amely a statisztikai validitásvizsgálatok (*t*-próbák, MORI) alapján megfelelőnek bizonyult.

Az általunk feltárt tíz klaszter a nyelvcsere nem tisztán lineáris folyamatában tíz létező beszélőtípust azonosít (lásd az 1. és a 2. ábrát, illetve az 5. táblázatot). Közülük a teljes minta közel kétharmadát kitevő KL6–KL10 klaszterek még dominánsan kisebbségiak (lásd a 2. ábrát), de az öt közül négyben a megkérdezettek már lépés(ek)e)t tettek a nyelvcsere (nyelvvesztés), illetve az asszimiláció irányába: gyengült/megszűnt/a többségi magyar kultúra felé elmozdult etnikai identitásuk (KL6), kisebbségi nyelvvel szembeni attitűdjük (KL7), kisebbségi nyelvtudásuk (KL9), vagy a magyar nyelvet használják családjukban (KL8). Mindemellett ezek még továbbra is bizakodásra adnak okot szemben KL1–KL4-gyel, amelyek már a nyelvcsere és az asszimiláció végső fázisát képviselik, vagy ahhoz közel állnak (lásd az 1. ábrát és az 5. táblázatot). Figyelmet érdemel, hogy csak egyetlen klaszter (KL5) képvisel átmeneti típust, de ez is inkább az asszimiláció felé „húz” (lásd az 1. ábrát). Mindez arra is utalhat, hogy a nyelvcsere folyamatának középső fázisa instabil, egyedi és gyors, mely nem hoz létre markáns átmeneti típust.

Eredményeink megerősítik, hogy a fokozatosan, de megállíthatatlanul végbemenő nyelvcsere a magyarországi nemzetiségek esetében nem lineáris folyamat. Több elkülöníthető összetevője van (családi nyelvhasználat, nyelvtudás, nyelvi attitűd, identitás stb.), amelyeket ha célzottan és hatékonyan befolyásolunk/alakítunk, esély lehet a fenntartható kétnyelvűség megteremtésére. Ezáltal – a domináns magyar nyelvhasználat kialakulásával párhuzamosan – bizonyos nyelvi komponensekben megőrizhető lenne a kétnyelvűség, valamint a saját közösséghez és a kisebbségi nyelvhez való kötődés, melynek emocionális vetülete (saját nyelv iránti pozitív attitűd, etnikai identitás), nyelvikompetencia-összetevője (nyelvtudás), valamint nyelvi viselkedési komponense (nyelvhasználat) is van.

Az elemzésekbe bevont változók között csak gyenge-közepes a korreláció, ami lehetőséget teremt arra, hogy e kisebbségi komponenseket külön-külön is befolyásoljuk. Például az etnikai identitást a különféle kisebbségi kulturális, tudományos, ismeretterjesztő, szociális, sportprogramok stb. erősíthetik; a kisebbségi nyelvtudás javítására minőségi oktatás, illetve élethosszig tartó nyelvtanfolyamok szervezhetőek; a kisebbségi nyelvhez való pozitív attitűdöt országosan ismert kisebbségi személyek, szakemberek nyílt állásfoglalásával, illetve a nyelvtudás társadalmi hasznosságára hivatkozó meggyőző érvekkel lehet pozitív irányban alakítani. Az intervencióban a helyi kisebbségi önkormányzatok kulcsszerepet tölthetnek be, de a folyamatot az is segítheti és erősítheti, ha a helyi és az országos média a magyarországi nemzetiségi közösségekről és a környező országokról, illetve népeikről a jelenleginél több pozitív megnyilvánulást, véleményt, valamint jó együttműködésre való törekvést közvetít.

Az elemzésben felhasznált adatok a 2003/2004-es állapotokat tükrözik. Az életkori összefüggéseket bemutató eredmények (lásd a 4. és az 5. ábrát) arra engednek következtetni, hogy a kisebbségi nyelvükhöz leginkább ragaszkodó idős emberek eltávozásával az elmúlt 10–12 évben romlott a helyzet. Azt azonban, hogy a nyelv-



csere folyamata nem feltétlenül fokozatosan egy irányba mutatóan zajlik, vagyis hogy a nemzetiségi közösségeknek vannak olyan kisebb csoportjai, amelyekben a nyelvcsera folyamata lassabban megy végbe és az identitás, illetve a saját nyelvvel szembeni pozitív attitűd hosszabban megmarad, egy magyarországi románok közösségében végzett, nemrég publikált longitudinális kutatás eredményei jelzik (Vargha–Borbély [2016]).

## Irodalom

- BARTHA CS. [2003]: A nyelvi másság dimenziói: a kisebbségi nyelvek megőrzésének lehetőségei – egy országos szociolingvisztikai-kétnyelvűségi vizsgálatról. In: *Hajdú M. – Keszler B. (szerk.): Köszöntő Kiss Jenő 60. születésnapjára*. ELTE Magyar Nyelvtudományi és Finnugor Intézet, Magyar Nyelvtudományi Társaság. Budapest. 304–311. old.
- BERGMAN, L. R. – MAGNUSSON, D. – EL-KHOURI, B. M. [2003]: *Studying Individual Development in an Interindividual Context. A Person-Oriented Approach*. Lawrence-Erlbaum Associates. Mahwa, London.
- BORBÉLY A. [2001]: *Nyelvcseré. Szociolingvisztikai kutatások a magyarországi románok közösségében*. MTA Nyelvtudományi Intézet. Budapest.
- BORBÉLY A. [2014]: *Kétnyelvűség – Variabilitás és változás magyarországi közösségekben*. L'Harmattan Könyvkiadó. Budapest.
- COHEN, J. [1977]: *Statistical Power Analysis for the Behavioral Sciences (Revised Edition)*. Academic Press. New York.
- COHEN, J. [1992]: A power primer. *Psychological Bulletin*. Vol. 112. No. 1. pp. 155–159. <http://dx.doi.org/10.1037/0033-2909.112.1.155>
- GAL, S. [1979]: *Language Shift: Social Determinants of Linguistic Change in Bilingual Austria*. Academic Press. New York.
- HARTIGAN, J. A. [1975]: *Clustering Algorithms*. John Wiley and Sons. New York.
- SCHEFFÉ, H. [1959]: *The Analysis of Variance*. John Wiley and Sons. New York.
- TAKÁCS SZ. [2013]: Többdimenziós skálázás. *Psychologia Hungarica Caroliensis*. I. évf. 1. sz. 140–151. old.
- VARGHA A. [2008]: Új statisztikai módszerekkel új lehetőségek: a ROPstat a pszichológiai kutatások szolgálatában. *Pszichológia*. 28. évf. 1. sz. 81–103. old.
- VARGHA A. [2016]: A ROPstat statisztikai programcsomag. *Statisztikai Szemle*. 28. évf. 11–12. sz. 1165–1192. old. <http://dx.doi.org/10.20311/stat2016.11-12.hu1165>
- VARGHA, A. – BERGMAN, L. R. [2015]: *Finding Typical Patterns in Person-Oriented Research within a Cluster-Analytic Framework Using ROPstat*. Conference on Person-Oriented Research. 8–9 May. Vienna.
- VARGHA, A. – BERGMAN, L. R. – TAKÁCS, SZ. [2016]: Performing cluster analysis within a person-oriented context: Some methods for evaluating the quality of cluster solutions. *Journal for Person-Oriented Research*. Vol. 2. Nos. 1–2. pp. 78–86. <http://dx.doi.org/10.17505/jpor.2016.08>

VARGHA A. – BORBÉLY A. [2016]: Modern mintázatható módszerek alkalmazása a kétnyelvűség kutatásában. In: *Kissné Viszket M. – Puskás-Vajda Zs. – Rácz J. – Tóth V. (szerk.): A pszichológiai tanácsadás perspektívái. Tisztelegő kötet Ritoók Magda 80. születésnapjára.* L'Harmattan. Budapest. 173–186. old.

VARGHA, A. – TORMA, B. – BERGMAN, L. R. [2015]: ROPstat: A general statistical package useful for conducting person-oriented analyses. *Journal for Person-Oriented Research.* Vol. 1. Nos. 1–2. pp. 87–98. <http://dx.doi.org/10.175 05/jpor.2015.09>

## Summary

The paper demonstrates the usefulness of some new and promising modern analytic techniques of cluster analysis in linguistics by means of data from a study of language shift and national minority assimilation of six (Boyash, Roma, German, Romanian, Serb, and Slovak) communities living in Hungary. Using four variables of major importance (minority identity, attitude towards minority language, language choice in family, and minority language proficiency), it explores statistically valid and linguistically meaningful homogeneous types, which characterize bilingual adult persons of these national minorities. The cluster analyses identify ten main types of language shift. The results show that the long way from a monolingual minority to a monolingual Hungarian state is not a linear process. This may give a chance to a successful intervention to slow down the process of assimilation and language shift.