

**Szintaktikai dialektusok statisztikai elemzése felé: Egy új alárendelő  
szerkezet elfogadottságának felmérése\***

Kálmán László (1957-2021) emlékének

**Szeteli Anna**

Pécsi Tudományegyetem, Pécs  
anna.szeteli@gmail.com

**Kenesei István**

MTA Nyelvtudományi Kutatóközpont, Budapest  
és Szegedi Tudományegyetem, Szeged  
kenesei@nytud.hu

*Kivonat*

Tanulmányunk tárgyát olyan *hogy* kötőszós szerkezetek képezik, melyekben a kötőszó nem véges igéjű tagmondatban, hanem főnévi igenév mellett jelenik meg. E kevésbé kutatott használat terjedését Kenesei (2020) szerint erősen motiválja, hogy a főnévi állítmánnyal álló infinitívuszi mellékmondat egy hiátust tölt ki a magyar grammatika rendszerében. Két adatközlői csoportnak a szerkezetre adott grammatikalitási ítéletei alapján megállapítható, hogy egyes anyanyelvi beszélők sokszor csak alacsony szinten, de elfogadják azt. A nyelvészekből és a diákokból álló populációkat klaszteranalízis segítségével elfogadóbb és elutasítóbb csoportokra bontottuk. A tanulmány ennek menetét mutatja be, illetve azt, hogy milyen adatközlői stratégiák rajzolódnak ki a létrejött klaszterek alapján a két populációban. Az adatközlői mintázatok keresésével a cél értékelni a szerkezetek elfogadhatóságát és feltárni a konstrukció terjedésének lehetséges irányait.

*Kulcsszavak:* nem véges igéjű tagmondat, kötőszó, alárendelés, klaszteranalízis, grammatikalitási ítéletek

---

\* A tanulmány megírásának idején a szerzők kutatói tevékenységét az NKFI FK-128518 (Szeteli Anna), illetve az NKFIH 120073 (Kenesei István) projektek támogatták.

A szerzők köszönetet mondanak a cikk első változatát bíráló lektoroknak: számos tanácsukat megfogadtuk, illetve kérdéseikre a cikkben válaszoltunk.

*Abstract*

In this study constructions are investigated containing the Hungarian complementizer *hogy*, unusually appearing in a nonfinite embedded clause. According to Kenesei (2020), the recent spread of this underresearched usage can be motivated by the gap-filling property of such structures attached to a matrix clause with nominal predicate. Based on the grammaticality judgements of two populations, certain native speakers accept the construction, although often only on a low level. The populations of linguists and students were each separated into accepting and refusing subgroups by applying cluster analysis. The paper presents this process and the strategies of the participants, based on the identified clusters. The aim is to explore the acceptability and the directions of the spreading of the construction.

*Keywords:* nonfinite clause, complementizer, embedding, cluster analysis, judgment data

*1. Bevezetés*

Tanulmányunk célja, hogy részletesen bemutassuk a Kenesei (2020) által tárgyalt *hogy* kötőszós szerkezetek első empirikus vizsgálatát. Ezekben a szerkezetekben a kötőszó nem véges igéjű tagmondatban, hanem főnévi igenév mellett jelenik meg, ahogyan azt a második szakaszban részletesebben bemutatjuk.

A harmadik szakaszban tárgyaljuk az adatgyűjtés és -feldolgozás részleteit. Az adatok egy nyelvészek, majd diákok körében elvégzett írásbeli kérdőívezésből származnak. A különböző szerkezetekre adott grammatikalitási ítéletek statisztikai feldolgozásának elméleti megfontolásairól a harmadik szakasz második felében írunk.

A negyedik szakaszban beszámolunk a kérdőívezés eredményeiről először a nyelvészek, majd a diákok csoportját górcső alá véve. A szerkezet terjedésére vonatkozó tanulságok levonása mellett a diszkusszióban egy alternatív klaszterezési eljárással is megvizsgáljuk az adatokat, hogy megállapításaink más módszerekkel is igazolhatók-e. Elemzésünk fő hozadéka, hogy az alkalmazott klaszterező eljárásnak köszönhetően nem csupán arról tehetünk kijelentéseket, hogy a vizsgált szerkezet az adatközlők körében milyen állítmány mellett elfogadottabb. A nyelvtani dialektusok azonosítása kétféle klaszterezési eljárással robosztus eredményhez vezet: megmutatja, hogy az állítmány típusa a jólformáltság elsődleges prediktora, illetve hogy nincs olyan adatközlői részhalmaz, amelyik a főnévi állítmánnyal álló infinitivuszi *hogy* kötőszós szerkezeteket ne fogadná el, más állítmánnyal állókat azonban igen. Ez összhangban áll Kenesei (2020) elméleti megfontolásaival, melyek szerint a szerkezet terjedése a főnévi állítmányos változat hiátuskitöltő szerepe felől értelmezhető.

## 2. A jelenség

Az elmúlt néhány évben lehetett megfigyelni egy új szerkezetű felbukkanását, amelyben a főnévi igenévvel (infinitívusszal) alkotott tagmondat egy főnévi csoport részeként, viszont szokatlanul *hogy* kötőszóval bevezetve fordul elő, ahogy azt az alábbi példákban illusztráljuk, melyek közül az élő beszédből feljegyzettekért Kálmán Lászlónak (KL), illetve a Magyar Nemzeti Szövegtárból (MNSZ) származókért Sass Bálintnak tartozunk köszönettel.<sup>2</sup>

- (1) a. Az a cél, hogy az értelmiségiek alól kihúzni a talajt. (KL)  
 b. Egy életet lehetne arra áldozni, hogy azokat megírni. (KL)  
 c. Nyilvánvalóan kínálja magát az a lehetőség, hogy eltörölni az egészségügyi hozzájárulást. (MNSZ)  
 d. Majd meglátja, egy idő után eszébe se jut az a csacskaság, hogy virulni meg illatozni. (MNSZ)

Amint azt Kenesei (2020) bemutatta, az infinitívuszt tartalmazó *hogy* kötőszós mellékmondatok, azaz IHKM-ek egyrészt az (alkalmasint esetjelölt) *az* utalószóval állnak konstrukcióban, másrészt a véges igéjű HKM-et jellemzésükben egyébként is megengedő lexikális fej mellett jelennek meg, az alábbi sematikus formában.<sup>3</sup>

- (2) a. a cél (van) [DP az [CP hogy V-ni ...]]  
 b. ... [DP az a [N lehetőség [CP hogy V-ni ...]]

Az idézett tanulmányban azt is állítottuk, hogy az IHKM kötőmódú értelmezést tesz szükségessé, továbbá hogy az IHKM rejtett alanya a főmondat valamely referáló kifejezésével koreferens, illetve általános vagy tetszőleges referenciájú is lehet.

Mivel mind az igei (V), mind a melléknévi (A) fejek mellett nemcsak *hogy* kötőszós mellékmondatok (HKM), de infinitívuszi (INF) tagmondatok is szabadon megjelenhetnek, az volt a konklúziónk, hogy az IHKM-ek felbukkanása a főnévi (N) vagy névmási (Pron) fejek mellett egyfajta nyelvtani hiátus kitöltése, hiszen a főnevek vagy a névmások tiltják az infinitívuszi vonzatokat, de a HKM-et ugyanebben a pozícióban megengedik.

Sematikusan ábrázolva a helyzetet és % jellel mutatva a dialektális elfogadottság változó mértékét, az alábbi rendet fedezhetjük fel.

- (3) a. V – HKM / INF / \*IHKM

<sup>2</sup> L. még Oravecz et al. (2014).

<sup>3</sup> A (2a) szerkezet nem tükrözi a felszíni sorrendet, amelyben az *az* utalószó a(z) (akár implicit) kopula elé kell kerüljön.

- b. A – HKM / INF / \*IHKM  
 c. N/Pron – HKM / \*INF / %IHKM

A szerkezet számos grammatikai problémája további elemzésre vár, például a lent idézett (7a), illetve (9a) mondat, amelyről egyik lektorunk találóan állapította meg, hogy itt egyfajta „kvótatív” szerkezetéről lehet szó. Hogy felvetett gondolatát továbbfejtsük, a *kérdés* szó beszédaktusra utal, de szemben a szokásos vonzatával, a HKM-mel, amelyben bármely igemód lehetséges, az IHKM-ben csak a kötőmódú értelmezéssel összeférő szerkezet helyezhető el, aminek a diszjunkció teljes mértékben eleget tesz (vö.: *az a kérdés, hogy menjünk vagy maradjunk*). A *hogy* mondatbevezető más szerkezetekben is tapasztalható ilyen kvótatív vonatkozásával egy folyamatban lévő kutatásunk is foglalkozik. Amint bírálónk megjegyzéséből is következik, a kérdő diszjunkciók amúgy is teljes értékű, bár elliptikus, megnyilatkozásokként szerepelhetnek akár mint főmondatok, akár mellékmondatként, vö. (4)

- (4) a. Most menni vagy maradni? (- kérdezte Mari.)  
 b. A könyvet vagy a lemezt? (- érdeklődött az eladó)

A következő részekben az IHKM-ek elfogadottságát, illetve elterjedtségét vizsgáljuk statisztikai módszerekre támaszkodva.

### 3. A felmérés

A fenti példákból kitűnik, hogy az IHKM-ek változatos állítmányokkal állhatnak, ezen felül a kérdőív összeállításakor további tényezőként adódott például a *hogy* kötőszavas tagmondat kihelyezése topik pozícióba. Jelen szakaszban mutatjuk be a kérdőíveket és azok rögzítésének körülményeit a két adatközlői csoportban, emellett beszámolunk a klaszteranalízishez kapcsolódó elméleti megfontolásokról.

#### 3.1. A kérdőívek felvétele

Előzetes tájékozódáshoz egy 16 nyelvészből álló csoport töltötte ki a következőkben ismertetett kérdőívet, majd ugyanezt a kérdőívet 63 budapesti diáknak is kiosztottuk.<sup>4</sup> A nyelvészek változatos életkorúak voltak, de mindannyian 30 év felettek, míg a diákok életkora 16–20 év között alakult. Itt szögezzük le, hogy a minták az eltérő elemszám és a nyelvészeti képzettségbeli különbség miatt statisztikailag nem vonhatók vagy hasonlíthatók össze, azonban külön-külön jó tájékoztatásul szolgálhatnak arról, hogy mely konstrukciók a legelfogadottabbak, illetve milyen téren érdemes tovább tesztelni. A kérdőív 22 szerkezetet tartalmazott, melyeket a függelékben ismertetett sorrendben értékelték a résztvevők. Ezek

<sup>4</sup> Utóbbi lebonyolításáért Schiller Mariann (ELTE Radnóti Miklós Gyakorló Gimnázium) illeti köszönet.

között egyaránt megtalálható volt igei (5a-c), melléknévi (5d-e) és főnévi állítmánnyal álló IHKM (5f-g), és emellett ezek egy részének topikalizált változatai (x'). Utóbbi azért gondoltuk szükségesnek, hogy leteszteljük: más szórendben elfogadhatóbbak lesznek-e az IHKM-ek.

- (5)
- a. azt szeretném, hogy alaposan ledöngölni a padlót
  - b. az kellene, hogy kifesteni a kamrát
  - c. azon gondolkodom, hogy mindenkinek megmondani az igazat
  - c'. hogy mindenkinek megmondani az igazat, azon gondolkodom
  - d. az fontos volt, hogy megismerkedni az elnökkel
  - e. az a legnehezebb, hogy mindenkinek megmondani az igazat
  - e'. hogy mindenkinek megmondani az igazat, az a legnehezebb
  - f. az a mérce, hogy kitörni a magyar valóságból
  - f'. hogy kitörni a magyar valóságból, az a mérce
  - g. az a cél, hogy túlélni
  - g'. hogy túlélni, az a cél

Bekerültek olyan változatok is, amelyekben az utalószó (6a-b) vagy a kötőszó nem szerepel (6c), ezek topikalizált változatai ((6b') és (6c')) illetve helyet kapott olyan mondat is, amely nem IHKM és elméletileg – illetve a nyelvészek ítéletei alapján is – kifogástalan (6d). Ennek párja (6d').<sup>5</sup>

- (6)
- a. fontos volt, hogy megismerkedni az elnökkel
  - b. a legnehezebb, hogy mindenkinek megmondani az igazat
  - b'. hogy mindenkinek megmondani az igazat, a legnehezebb
  - c. az a legnehezebb, mindenkinek megmondani az igazat
  - c'. mindenkinek megmondani az igazat, az a legnehezebb
  - d. a legnehezebb mindenkinek megmondani az igazat
  - d'. mindenkinek megmondani az igazat a legnehezebb

---

<sup>5</sup> Egyik lektorunk kérdésére, amely az (6c, c', d, d') grammatikai státuszára vonatkozott, itt annyit mondhatunk, hogy az (6c) állítmánya elvben nem enged *hogy*-törlést még a szokásos HKM-ben sem, ráadásul "kettőspontos stílusban" sem értelmezhető a mutatószó miatt. Az (6c') egyszerű topikalizáció és visszautalás eredménye, ami garantálja az elfogadottságát. Az (6d, d') infinitívuszi alany mellett álló melléknévi állítmányt tartalmaz, ami szintén problémamentes szerkezet. (Vö. Kenesei 1992, 673-80)

Emellett bekerült még néhány szerkezet, melyeknek a *hogy* kötőszós alárendelésében diszjunkció (7a-a') vagy kérdőszó szerepel (7b-c).

- (7) a. fölmerül az a kérdés, hogy túrni vagy lázadni  
a'. hogy túrni vagy lázadni, az a kérdés merül föl  
b. fölmerül az a kérdés, hogy meddig túrni  
c. az a legnehezebb, hogy mikor mondani meg

A gyűjtést követően az adatok a 23-as verziójú SPSS szoftver segítségével kerültek elemzésre. Az adatközlők eleve azt az utasítást kapták, hogy a mondatokat egy 1-től 5-ig terjedő Likert-skálán értékeljék (hasonlóan például Surányi és Turi (2017) megoldásához), ahol egy pontot kap a teljesen elfogadhatatlan és ötöt a kifogástalan mondat. Így tehát nem volt szükség az adatok intervallumos változóvá történő transzformálására, a feladatkijelölés eleve rögzítette, hogy az egyes skálapontok egymástól egyforma távolságot jelölnek (Schütze–Sprouse 2013, 8).

### 3.2. Klaszteranalízis

Az adatközlők mindkét csoport esetében szemmel láthatóan is elkülönülnek. Vannak közöttük, akik minden IHKM-konstrukcióval szemben elutasítóak, egy részük néhány szerkezet esetében valamivel megengedőbb, egy másik részük pedig még több szerkezetet elfogad. Tehát a különböző típusú IHKM-ekre a különböző emberek eltérően reagálnak, ami a nyelvészek esetében még némileg lehetővé teszi e csoportok intuitív úton történő azonosítását, azonban a diákoknál (n=63) ez a rendszer kevésbé átlátható.

Ezért a csoportok azonosítására a nyelvészeknél és a diákoknál is hierarchikus klaszterező eljárást alkalmaztunk. Az adatokat korábban más módszerekkel is vizsgáltuk (Kenesei–Szeteli 2021). Az eljárást azt a célunkat szem előtt tartva választottuk, hogy különböző szintaktikai dialektusokat és adatközlői stratégiákat tárjunk fel, melyek korábbi munkánkban csak részben bontakozhattak ki. Úgy ítéltük meg, hogy bár a populációstatisztikai mintázatok feltárására más bevett módszerek is léteznek, mégis az általunk használt eljárással azonosíthatjuk, egyben átláthatóan szemléltethetjük is az egyes adatközlői alcsoportokat.

A hierarchikus klaszterezés során nem kell előre megadni a keresendő klaszterek számát, ami alkalmassá teszi előzetes vizsgálódásra. Ide tartozik a széleskörűen használt és elfogadott (Cooksey 2014, 179; Sajtos–Mitev 2007, 295), varianciamódszernek is nevezett Ward-féle eljárás.

Kenesei (2020) alapján a hipotézisünk az, hogy a főnévi állítmánnyal álló IHKM-eket fogadják el a legszélesebb körben, aki ezeket is rossznak értékeli, azt az eljárás az elutasítók közé fogja sorolni. A létrehozott klaszterek elemzése során kiderülhet, ha mégis létezne olyan

alcsoporthoz, amelyik valamilyen más állítmánnyal álló IHKM-et fogad el, miközben a főnévi állítmánnyal állót elutasítja.

A klaszteranalízis változóiként a kérdőívben megadott 22 konstrukcióhoz tartozó grammatikalitási ítéleteket jelöltük meg. A változók faktoranalízissel történő csökkentéséhez nem állt rendelkezésre az ajánlott – a változók számához képest ötszörös – elemszám. Azonban az ökölszabályt, miszerint nem adunk hozzá olyan változót az elemzéshez, ami egy másik változóval erősen korrelál, szem előtt tartottuk (0.9-es korrelációs együttható, l. pl. Sajtos–Mitev 2007). Minden esetben ellenőriztük a korrelációkat, de nem merült fel ilyen változó a példamondatok között.

A távolságfüggvény varianciamódszer esetén kötelezően a négyzetes euklidészi távolság, használatakor a nagyobb távolságok még nagyobb hangsúlyt kapnak (Sajtos–Mitev 2007, 293). Emellett Cooksey (2014, 179) szerint is ez a legmegfelelőbb távolságfüggvény adatközlők klaszterekbe sorolásához.

#### 4. Eredmények

A két csoporton az előzőekben leírtak értelmében hierarchikus klaszteranalízist végeztünk, ezen belül is a Ward-féle varianciamódszert alkalmaztuk (Ward 1963) négyzetes euklidészi távolságfüggvénnyel.

##### 4.1. A nyelvészek csoportjai

A nyelvészek esetében a fenti beállításokat alkalmazva a hierarchikus klaszterezés két nagy csoportot hoz létre. Az 1. ábra felső klaszterébe hat olyan nyelvész kerül, aki általában bármilyen típusú IHKM-et elutasít. Ezzel szemben az alsó klaszterbe került az a tíz nyelvész, aki szisztematikusan több IHKM-et is elfogad. Hierarchikus klaszterezés esetén fontos kérdés, hogy hol „vágjuk el” a klaszterezési fát. Ebben az esetben IHKM elfogadókat és elutasítókat tartalmazó nagy klasztereken kívül az alsó – IHKM elfogadókat tartalmazó – fa még egyszer kettéválasztható, mind statisztikai, mind nyelvészeti megfontolások alapján. A statisztikai ok az, hogy az *e*, *g*, *b*, *l* tartalmú klaszternek a másik hat IHKM elfogadót tartalmazó klaszterrel való összevonásakor a koeficiens jelentősen megemelkedik (ld. Függelék), ami azt jelenti, hogy két távoli klaszter összevonása történt meg ebben az esetben, tehát ezeket indokolt különálló klaszterekként kezelni. Emellett nyelvészeti szempontok alapján is jól beazonosítható, hogy mely mondatokat értékeli eltérően a négy fős csoport, ami szintén a szétválasztás mellett szóló érv. Tagjai a tesztsorban szereplő két igei állítmányos IHKM közül (8a-b) az egyiket (8b) nem tartják elfogadhatatlannak, azonban nem fogadják el például a melléknévi állítmánnyal álló IHKM-eket (8d-e). A főnévi állítmányosokat (8f-g) viszont már jobbra értékeli. A példák az (5) tesztmondatainak ismétlései, már az elfogadóbb alcsoporthoz, illetve a hat fős elfogadó csoportokban, per jellel elválasztva.

- |   |         |
|---|---------|
| (8) a. azt szeretném, hogy alaposan ledöngölni a padlót     | 1/1     |
| b. az kellene, hogy kifesteni a kamrát                      | 2/2     |
| c. azon gondolkodom, hogy mindenkinek megmondani az igazat  | 1/1     |
| c'. hogy mindenkinek megmondani az igazat, azon gondolkodom | 1.5/2   |
| d. az fontos volt, hogy megismerkedni az elnökkel           | 1/1     |
| e. az a legnehezebb, hogy mindenkinek megmondani az igazat  | 1/2     |
| e'. hogy mindenkinek megmondani az igazat, az a legnehezebb | 2.5/2   |
| f. az a mérce, hogy kitörni a magyar valóságból             | 1.5/3   |
| f'. hogy kitörni a magyar valóságból, az a mérce            | 2.5/2.5 |
| g. az a cél, hogy túlélni                                   | 2/3.5   |
| g'. hogy túlélni, az a cél                                  | 2.5/2.5 |

Bár a melléknévi állítmánnyal álló IHKM-eket a nyelvészek inkább elutasítják, annak topikalizált változatát nem tartják teljesen rossznak, a négyfős csoportra ez különösen jellemző. Kenesei (2020) kimutatta, hogy a diákok hajlamosak elfogadni olyan topikalizált IHKM-eket, melyeknek az egyenes szórendbeli párjukat rosszabbra értékelték. A nyelvészekre ez összességében nem jellemző, de ennek a négyfős klaszternek a tagjai a diákokéhoz hasonló jelenséget produkálnak azzal, hogy ők az egyenes szórendű verziót teljesen elfogadhatatlannak ítélik. Természetesen az elemszám alacsony ahhoz, hogy ezt statisztikailag is igazolni lehessen.

Ugyanezt a négyfős társaságot jellemzi, hogy általában ötösre értékeli a diszjunkciót tartalmazó IHKM-eket (9), mind egyenes szórendű (9a), mind topikalizált változatban (9a').

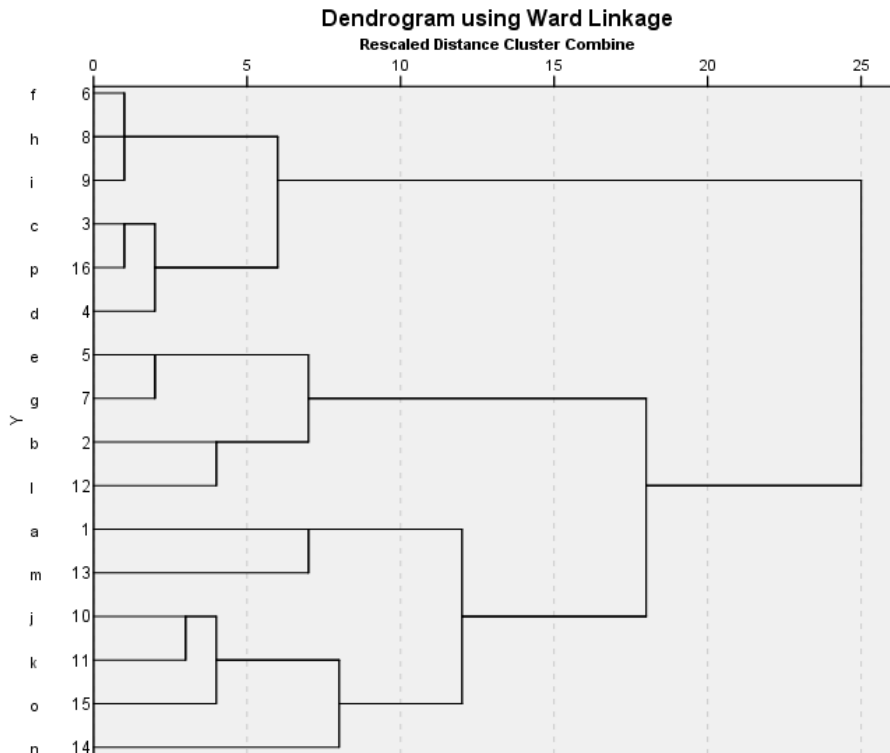
- |  |         |
|--|---------|
| (9) a. fölmerül az a kérdés, hogy tőrni vagy lázadni | 5/5     |
| a'. hogy tőrni vagy lázadni, az a kérdés merül föl   | 4.5/2.5 |
| b. fölmerül az a kérdés, hogy meddig tőrni           | 2/1.5   |
| c. az a legnehezebb, hogy mikor mondani meg          | 1/1.5   |

Ezzel jobbnak tartják őket, még a leginkább IHKM-párti 6 fős csoport egyes tagjainál is. A tesztbe bekerült a diszjunkciót tartalmazó IHKM főmondata kérdőszós beágyazással is (9b), ezt a kisebbik elfogadó csoport nem találta elfogadhatatlannak, azonban a nagyobbik csoportot jobban megosztotta a szerkezet. Egy melléknévi állítmányos főmondatú kérdőszós IHKM-et is teszteltünk (9c), ez ugyanolyan népszerűtlen a nyelvészek körében, mint a többi melléknévi állítmányos IHKM, a kérdőszós beágyazásnak feltehetően nincs nagy hatása a grammatikalitási ítéletekre, ellentétben a diszjunkcióval.

Mivel a nyelvészek életkora rendkívül széles spektrumon mozgott (30-70+), azt is



megvizsgáltuk, hogy van-e összefüggés az életkoruk és a grammatikalitási ítéleteik között valamelyik változóra nézve. A Pearson-féle korrelációs együttható minden esetben legfeljebb gyenge összefüggést mutatott és ezek az összefüggések egyik vizsgált szerkezet esetében se bizonyultak szignifikánsnak ( $|r| < 0.3$ ). Ez azt jelenti, hogy a nyelvészek esetében nem életkori tényezők befolyásolták az IHKM-ek elfogadásának mértékét.



1. ábra: Dendrogram a nyelvészek hierarchikus klaszterezése alapján

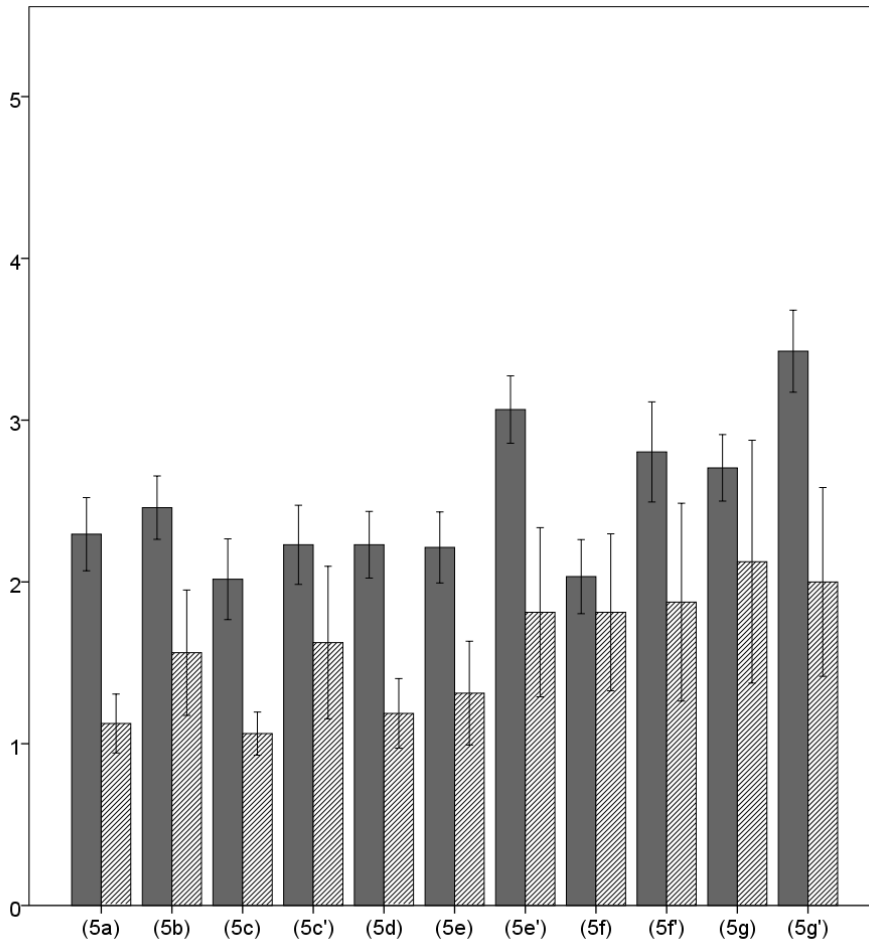
Összefoglalva: a fenti dendrogram két csoportra bontja a nyelvészeket aszerint, hogy IHKM-pártiak-e vagy sem. Emellett az IHKM elfogadó csoport még tovább bontható, amire statisztikai és nyelvészeti érvet is hoztunk. Az ábra alján szereplő hat fős klaszter általánosságban jó pontokat ad a különféle szerkezeteknek, míg a négyfős csoport a melléknévi állítmánnyal és az egyik igei állítmánnyal álló IHKM-et is teljes mértékben elfogadhatatlannak találja, de például a főnévi állítmányt elfogadja, azt a csoport tagjai esetenként kifejezetten jóra értékelik. A négyfős csoport tagjait jellemzi még, hogy több esetben jobb pontokat adnak a topikalizált változatokra, mint a hatfős elfogadó csoport, illetve osztatlan sikert arat körükben a diszjunkciót tartalmazó IHKM. A diszjunkciókkal szemben a kérdőszós IHKM-ek nem javítanak a szerkezet megítélésén, ilyenkor továbbra is a főmondat állítmánya az irányadó abban, hogy milyen megítélése lesz a szerkezetnek.

#### 4.2. A diákok csoportjai

A diákok klaszterezésében csak 61 fő döntéseit vizsgálhattuk, mivel két esetben hiányzott egy-egy grammatikalitási ítélet (az elemzéshez tartozó dendrogram ezúttal a függelékben található). Ebből egy 12 fős csapat bizonyult a legelutasítóbbnak az IHKM-ekre nézve, így az egyik klasztert ők alkotják. A másik, elfogadó klaszter tovább bontva egy 19 és egy 30 fős csoportra ezúttal nem szolgál olyan letisztult képpel, mint a nyelvészeknél. Náluk két világosan elkülöníthető stratégia rajzolódott ki a tekintetben, hogy mely szerkezeteket értékeli jobbnak az egyik és a másik csoport. A diákok esetében azonban nem ez a helyzet, a kisebbik csoport egyetlen markáns közös tulajdonsága, hogy az első főnévi állítmányos (5f) mondatot tagjai elfogadhatatlannak találják, míg a 30 fős csoport 3 körüli elfogadhatóságúnak értékeli azt. A 30 tagú csoport más mondatokat is fél-egy ponttal jobbra értékeli a kisebb csoportnál, de fordított helyzetek nincsenek, mint ahogyan az a nyelvészeknél előfordult.

További változás a nyelvészekhez képest, hogy a diákoknál már nem az IHKM elfogadókon belül válnak el egymástól azok, akik „még nem kedvelik” a melléknévi vagy az igei állítmányos IHKM-eket, hanem az IHKM elutasító és kedvelő csoport között. Ennek oka, hogy a diákok között mást jelent IHKM elutasítónak lenni, mint a nyelvészek között. Ott a klaszteranalízis azokat választotta le, akik kategorikusan minden típusra az elfogadhatatlant jelentő egyes ítéletet adták. A diákoknál azonban már az számít a legkonzervatívabb álláspontnak, ha valaki csak a főnévi állítmányos szerkezetekre ad magasabb pontokat és a többi típust elutasítja.

Megemlítendő, hogy a diákok eleve minden mondatra magasabb pontot adtak, mint a nyelvészek (ld. 2. ábra), illetve azokban az esetekben, ahol pedig a nyelvészek a maximális 5 pontra értékelték a mondat jóformáltságát, ott a diákok óvatosabbak voltak, inkább 4.5-ös átlagokat állítottak be. Azonban ez nem jelenti azt, hogy a diákok a populációjukon belül ne tennének különbséget a különböző típusú állítmányok között. Nemparaméteres tesztek – amelyeknél a konkrét értékek csak rangok formájában vésnek számításba – kimutatják, hogy a Likert-skála csekélyebb mértékű kihasználása mellett is különbséget tesznek a diákok a főnévi, igei és melléknévi állítmányok között (Friedman ANOVA:  $p=0.001$ ).



2. ábra: Nyelvszerek (sraffozott) és diákok (szürke) grammatikalitási ítéleteinek összehasonlítása (95%-os konfidenciaintervallumok)

A 2. ábrán látható, hogy a főnévi állítmányos IHKM-ek ((5f) és (5g)) megítélésének várható értéke nem különül el a diákok és a nyelvészek között. Jelentős különbség azonban, hogy a topicalizált változatok mind ezeknél a főnévi állítmányos IHKM-eknél ((5f') és (5g')), mind a melléknévi állítmányosnál (5e') népszerűbbek a diákok körében, mint az egyenes szórendű alakok, már a konfidenciaintervallumok alapján is. Ezen felül a diákok jóval elfogadóbbak az egyenes szórendű igei és melléknévi állítmánnyal álló IHKM-ekkel szemben is, mint a nyelvészek.

Végül tekintsük át azokat a tesztmondatokat, amelyekben nem szerepel utalószó és/vagy kötőszó.

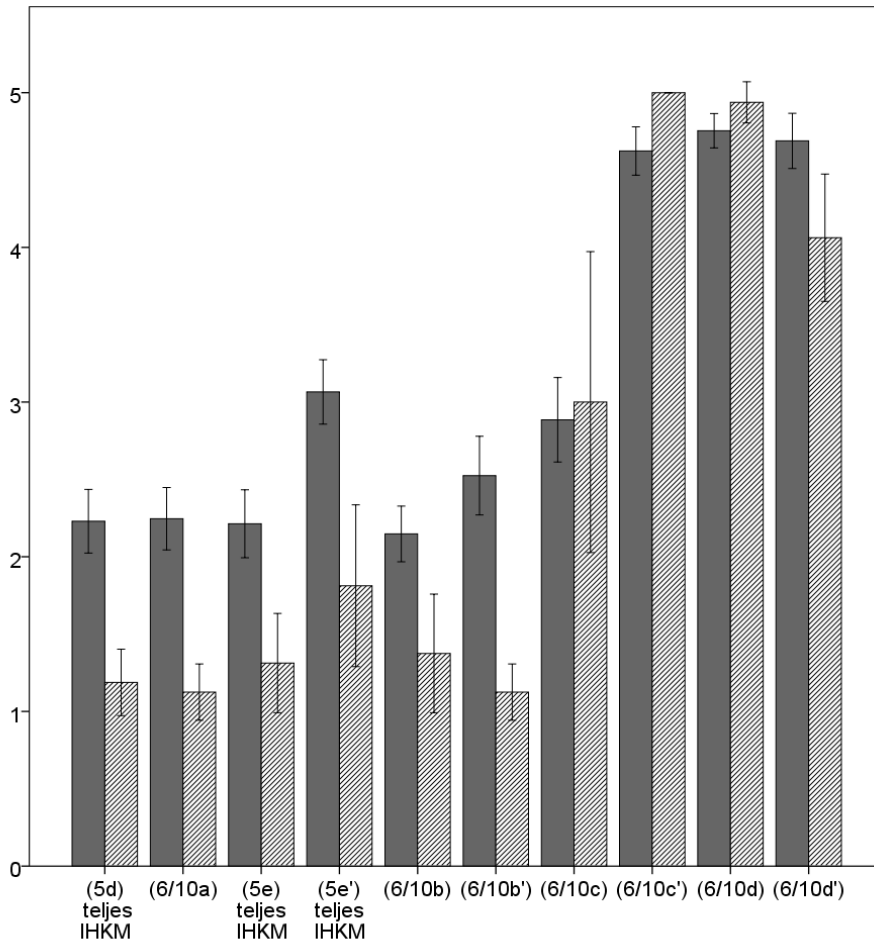
(10) a. fontos volt, hogy megismerkedni az elnökkel

- b. a legnehezebb, hogy mindenkinek megmondani az igazat
- b'. hogy mindenkinek megmondani az igazat, a legnehezebb
- c. az a legnehezebb, mindenkinek megmondani az igazat
- c'. mindenkinek megmondani az igazat, az a legnehezebb
- d. a legnehezebb mindenkinek megmondani az igazat
- d'. mindenkinek megmondani az igazat a legnehezebb

A 3. ábrán (10) mondatai előtt feltüntettük a szerkezetekhez tartozó teljes IHKM-eket is, hogy szemléltessük az utalószó és/vagy a kötőszó elhagyása esetén fellépő különbségeket. Látható, hogy az utalószó elhagyása esetén nem változnak meg a várható értékek, (10a) szerkezet és a hozzá tartozó (5d) teljes IHKM megítélése a nyelvészeknél és a diákok megbízható, nagy mintás csoportjánál (n=63) sem okoz változást. Ugyanez a helyzet (10b) és teljes párja (5e) esetén, azonban itt a topikalizált pároknál ((10b')-(5e')) az utalószó elhagyása mindkét csoportban rosszabb megítélést von maga után. A kötőszó nélküli szerkezetek várható értéke pedig (10c-d') természetesen megugrik az IHKM-ekhez képest.

Összefoglalva: A diákok, akárcsak a nyelvészek, feloszthatók IHKM pártiakra és IHKM elutasítókra, azonban ez az ő esetükben már mérsékeltbb módon valósul meg. Az IHKM-ek elutasítása náluk abban merül ki, hogy a főnévi állítmánnyal álló IHKM-eket nem tartják elfogadhatatlannak, viszont az egyéb állítmánnyal állókat igen. Az elfogadó csoporton belül pedig elkülöníthető ugyan két alcsoport, de a különbségük kimerül abban, hogy a 30 fős még elfogadóbb és nem áll fenn olyan helyzet, mint a nyelvészeknél, ahol az adatközlők eltértek abban, hogy pontosan melyik szerkezeteket fogadják el. A diákok IHKM elfogadó része tehát jóval homogénebbnek bizonyult, mint a nyelvészeké.

Ugyanitt összehasonlítottuk a hiányos IHKM-eket a teljes IHKM-ekkel, melynek eredménye, hogy az utalószó elhagyásának nincs nagy jelentősége a szerkezet megítélésében, sem a diákok, sem a nyelvészek esetében.



3. ábra: Nyelvszerek (sraffozott) és diákok (szürke) grammatikalitási ítéleteinek összehasonlítása hiányos IHKM-ek esetében (95%-os konfidenciaintervallumok)

### 5. *Diszkusszió*

A klaszteranalízis érvényességének ellenőrzésére több módszer is alkalmas (ld. pl. Sajtos–Mitev 2007, 313–316; Riazi 2016, 35). Ebben a szakaszban mutatjuk be a kiegészítésként alkalmazott kétlépes klaszterezés tanulságait, ami a varianciamódszerrel ellentétben egy nem hierarchikus klaszterezési eljárás. Emellett tárgyalunk még néhány általános jelenséget a grammatikalitási ítéletek körül, melyek hatással lehetnek az IHKM-ek megítélésére is.

Ha egy csoporton klaszteranalízist végzünk infinitívuszi HKM-ekre adott

grammatikalitási ítéleteik alapján és a klaszterek számát – a nem hierarchikus eljárások esetében – kétféle rögzítjük, akkor általában egy az IHKM-ek tekintetében elfogadóbb és egy elutasítóbb alcsoport jön létre.

Azonban például különböző populációkon elvégzett kétlépes klaszterezés különböző eredményeket hozhat a tekintetben, hogy az alcsoportok mely szerkezetek megítélései alapján különülnek el egymástól, tehát mely mondatok erősebb prediktorai a csoportosításnak. A nyelvészekből álló csoporton végzett elemzésben ilyen erős prediktornak bizonyultak például a főnévi állítmánnyal álló utalószavas fejtű IHKM-ek megítélései, de még ezen szerkezetek topikalizált megfelelői is. Ellenben az igei vonzatban megjelenő IHKM-ek gyenge prediktorok, hiszen ezeket még sok IHKM-párti nyelvész is elutasítja. A diákokon végzett elemzésben viszont éppen ezek az igei IHKM-ek lépnek elő első számú prediktorokká. A főnévi IHKM-ek jelentősége csökken, a topikalizált változataikat pedig egységesen még jobbra értékeli, így azok a leggyengébb prediktorok.

A főnévi állítmánnyal álló IHKM-ek jelentőségének csökkenése arra utal, hogy a diákok csoportja sokkal egységesebben fogadja el őket, mint a nyelvészeké. Ez összhangban van azzal, hogy a diákok hierarchikus klaszterelemzésének az elfogadó csoportjában már nem nyilvánultak meg olyan éles különbségek, mint a nyelvészeknél. A diákokat már az igei IHKM-ek megítélése osztja meg, ami arra utal, hogy – a főnévi IHKM mintájára, ami a magyar nyelv grammatikájában egy részt tölt ki (Kenesei 2020) – analógiásan további IHKM-ek is egyre elfogadhatóbbá válhatnak majd. Érdemes lehet további felméréseket folytatni idősebb és fiatalabb korosztályokban is e megosztottság tekintetében, különös tekintettel az idősebb nem nyelvészek, illetve nem budapesti adatközlők bevonására, mivel ezáltal nyomom követhetőek lennének a szerkezet terjedésének összefüggései az életkorral. Ennek megállapításához a nyelvészek ítéletei nem szolgáltatnak elégséges adatot, a diákokéi pedig semmilyen módon nem utalnak erre, mivel csoportjuk a korra nézve homogén.

A nyelvészek ítéletei kapcsán az életkorral való korreláció hiánya (vö. 4.1) első ránézésre akár ellentmondásosnak is tűnhet azzal az állítással szemben, hogy a szerkezet terjedőben van, ezért ehelyütt érdemes (újra)tárgyalnunk néhány a populációk és a résztvevők közötti (esetleges) eltérést, illetve ezeknek a lehetséges hatásait a résztvevők grammatikalitási ítéleteire. A korreláció hiányát ebben a kis populációban több tényező is okozhatja<sup>6</sup> együtt vagy külön-külön.

Az egyik megemlíthető különbség, hogy a kísérletben résztvevő diákok Budapesten, illetve annak környékén élők, míg a nyelvészek ebből a szempontból egy heterogén társaság volt.

Egy másik megfontolandó szempont az eredmények értékelésekor, hogy – még a szakmailag felkészült nyelvészek csoportján belül is – különbség merülhet fel annak a

---

<sup>6</sup> Kérdéses, hogy egyáltalán támasszunk-e egy (képzett beszélőkből álló) populáció felé olyan elvárást, hogy egy szerkezet időbeli terjedését grammatikalitási ítéleteiben a résztvevők életkorára kötelezően leképezze (Schütze 2016, 112).

jelenségnek a szempontjából, melyet Nagata (1989) a szubjektív és objektív tudatosság (*subjective/objective self-awareness: SSA/OSA*) paradigmájában vizsgál. Nagata kutatásában ugyan nem mutatja ki, hogy szignifikáns különbséget jelentene a résztvevők ítéleteire nézve, ha a szerkezetet kontextus nélkül (SSA) vagy valamilyen pragmatikai kontextust figyelembe véve (OSA) értékelnék, azonban megállapít tendenciákat (Schütze 2016, 142). A jelenséget az IHKM-ek megítélésére értelmezve több beszélői stratégiára is magyarázatot kaphatunk. Egyrészt a nyelvészeknek a diszjunkciót preferáló hat fős alcsoportja felfogható úgy, mint egy az objektív tudatosságot (OSA) stratégiájául választó csoport: a diszjunkciót tartalmazó IHKM-eknek kvótatív értelmezést adva értékelhetik ezen szerkezeteket pozitívabban. Másrészt ezzel az is magyarázható, hogy például a könnyebben kontextusba helyezhető igei állítmányos IHKM-nek (3g) jelentősen jobb a megítélése a teljes beszélői közösségben, de a fenti hat fős nyelvész csoportban is a négy főshöz képest. Ez azt jelenti, hogy a kontextuskereső (OSA) stratégiához mindkét szerkezet esetében ugyanazok folyamodhattak.

Továbbá ugyan Ross (1979) adatai e tekintetben Schütze (2016, 110) szerint nem túlságosan meggyőzőek, azonban nem kizárt, hogy az életkor és az elfogadóbb grammatikalitási ítéletek között összefüggés áll fenn, ami szintén az ellen hatna, hogy a szerkezet időbeli terjedése az életkoron képeződjön le.

Ross (1979, 149) azt is megállapítja, hogy a naiv anyanyelvi beszélők a konzervatívabbak, vagyis ugyanabban az esetben ők adnak negatívabb ítéletet,<sup>7</sup> illetve kevésbé használják ki a rendelkezésükre álló skálát. A skála kisebb mértékű kihasználása jelen kutatásban is megjelent, viszont a negatívabb értékítéletek csak az egyéb szerkezetekre voltak igazak, az IHKM-ek esetében erősebb volt a diákok körében az elfogadás.

Összességében tehát elmondható, hogy kutatásunk egyes eredményei a szerkezet időbeli és a magyar nyelv grammatikájában történő terjedésére utalnak, méghozzá a főnévi állítmánnyal álló IHKM-ek felől. Más eredményei pedig – mint például a korban heterogén csoporton belüli korreláció hiánya – nem cáfolják meg azt.

## 6. Összefoglalás

Tanulmányunkban bemutattuk a Kenesei (2020) által tárgyalt *hogy* kötőszós szerkezetek első empirikus vizsgálatát. Feldolgoztuk egy nyelvészekből és egy diákokból álló adatközlői csoport grammatikalitási ítéleteit, melyeket infinitívuszt tartalmazó *hogy* kötőszós mellékmondatokra adtak. A két populációt hierarchikus klaszterezési eljárással további csoportokra osztottuk és megvizsgáltuk az elfogadóbb csoportok stratégiáit. A diákok eredményeit a nyelvészek eredményeivel összehasonlítva tekintettük, szem előtt tartva azonban azt, hogy a két csoport az elemszám, az életkori homogenitás és a szaktudás

---

<sup>7</sup> A kiegyensúlyozottság érdekében ide citálandó Spencer (1973) munkája is, ami ezzel ellentmondásos eredményre jut, azonban abban a kutatásban csak két kategóriába kellett osztani a szerkezeteket (grammatikus - agrammatikus), így jelen tanulmány tárgyával Ross kutatása mutat több párhuzamot.

tekintetében is különbözik, így ítéleteik várható értékeit leszámítva statisztikai módszerekkel nem hasonlítottuk össze őket, inkább a csoportokon belül felfedezhető stratégiákat tekintettük át. Elmondható, hogy az adatközlők megítélését leginkább a főmondat állítmánya befolyásolja, de sok esetben pozitív hatással van rá a topikalizáció és a beágyazott tagmondatban szereplő diszjunkció.

Tanulmányunk egyik fő hozadéka, hogy rámutattunk, az adatok alátámasztják Kenesei (2020) hipotézisét, hogy a szerkezet terjedésének középpontjában a főnévi állítmánnyal álló IHKM állhat, mivel mindkettő adatközlői csoportban ezeknek a pozitív megítélése volt egyértelműen megállapítható. A létrehozott klaszterek elemzése során kiderült volna, ha mégis létezik olyan adatközlői alcsoport, amelyik valamilyen más állítmánnyal álló IHKM-et elfogad, miközben a főnévi állítmánnyal állót elutasítja, de ilyen adatközlői stratégiát vizsgálatunk nem tudott kimutatni.

### *Hivatkozások*

- Cooksey, Ray W. 2014. *Illustrating Statistical Procedures: Finding Meaning in Quantitative Data*. Melbourne: Tilde University Press
- Kenesei, István. 1992. Az alárendelt mondatok. In Kiefer Ferenc, (szerk.), *Strukturális magyar nyelvtan, I. Mondattan*. Budapest: Akadémiai. 529-714.
- Kenesei, István 2020. A terjeszkedő igenév. *ÁNYT* 32: 123–136.
- Kenesei, István – Anna Szeteli 2021. Surprise: Nonfinite Clause with Finite Complementizer. In: Judit Gervain – Gergely Csibra – Kristóf Kovács (szerk.) *A Life in Cognition: Studies in Cognitive Science in Honor of Csaba Pléh*. Springer. 93-106.
- Nagata, Hiroshi. 1989. Judgments of sentence grammaticality with differentiation and enrichment strategies. *Perceptual and Motor Skills* 68(2): 463–469.
- Oravecz, Csaba – Tamás Váradi – Bálint Sass 2014. The Hungarian Gigaword Corpus. In Nicoletta Calzolari et al. (szerk.): *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. European Language Resources Association (ELRA). 1719–1723  
[http://www.lrec-conf.org/proceedings/lrec2014/pdf/681\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2014/pdf/681_Paper.pdf)
- Riazi, A. Mehdi. 2016. *The Routledge Encyclopedia of Research Methods in Applied Linguistics*. London: Routledge
- Ross, John Robert 1979. Where's English? In Charles J. Fillmore – Daniel Kempler – William S-Y. Wang (szerk.): *Individual differences in language ability and language behavior*. New York: Academic Press. 127–163.
- Sajtos László – Mitev Ariel 2007. *SPSS Kutatási és adatelemzési kézikönyv*. Budapest: Alinea Kiadó.
- Spencer, Nancy J. 1973. Differences between linguists and nonlinguists in intuitions of grammaticality-acceptability. *Journal of Psycholinguistic Research* 2(2): 83–98.
- Schütze, Carson T. – Jon Sprouse 2013. Judgement Data. In: Robert J. Podesva – Devyani



Sharma (szerk.): *Research Methods in Linguistics*. Cambridge: Cambridge University Press. 27–50.

Schütze, Carson T. 2016. *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Berlin: Language Science Press.

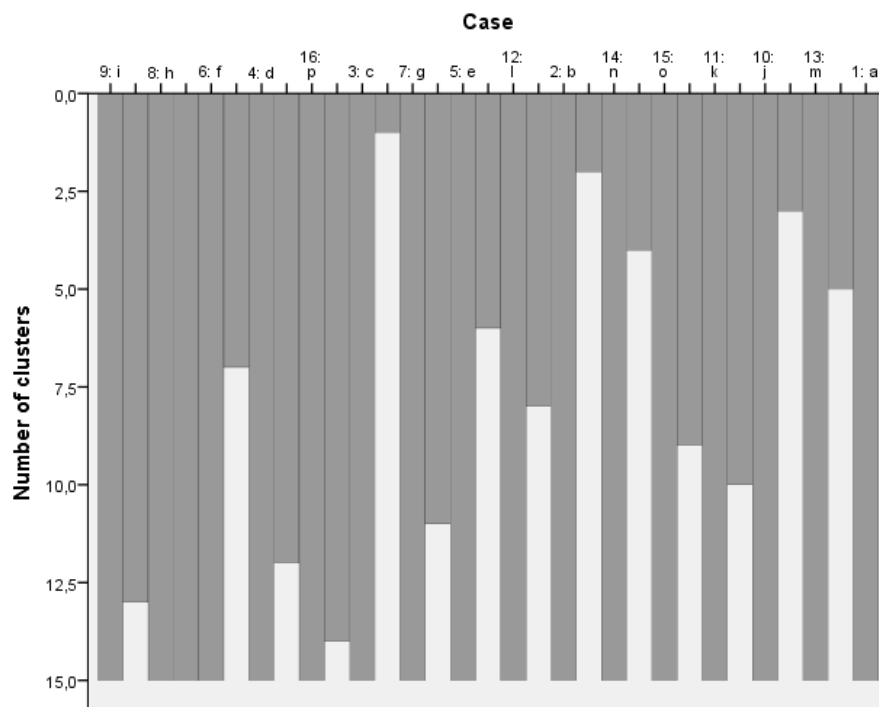
Surányi, Balázs – Gergő Turi 2017. Focus and quantifier scope: An experimental study of Hungarian. In: Harry van der Hulst – Anikó Lipták (szerk.): *Approaches to Hungarian* 15. Amsterdam: John Benjamins. 209–238.

Ward, Joe H. 1963. Hierarchical groupings to optimize an objective function. *Journal of the American Statistical Association* 58: 236–244.

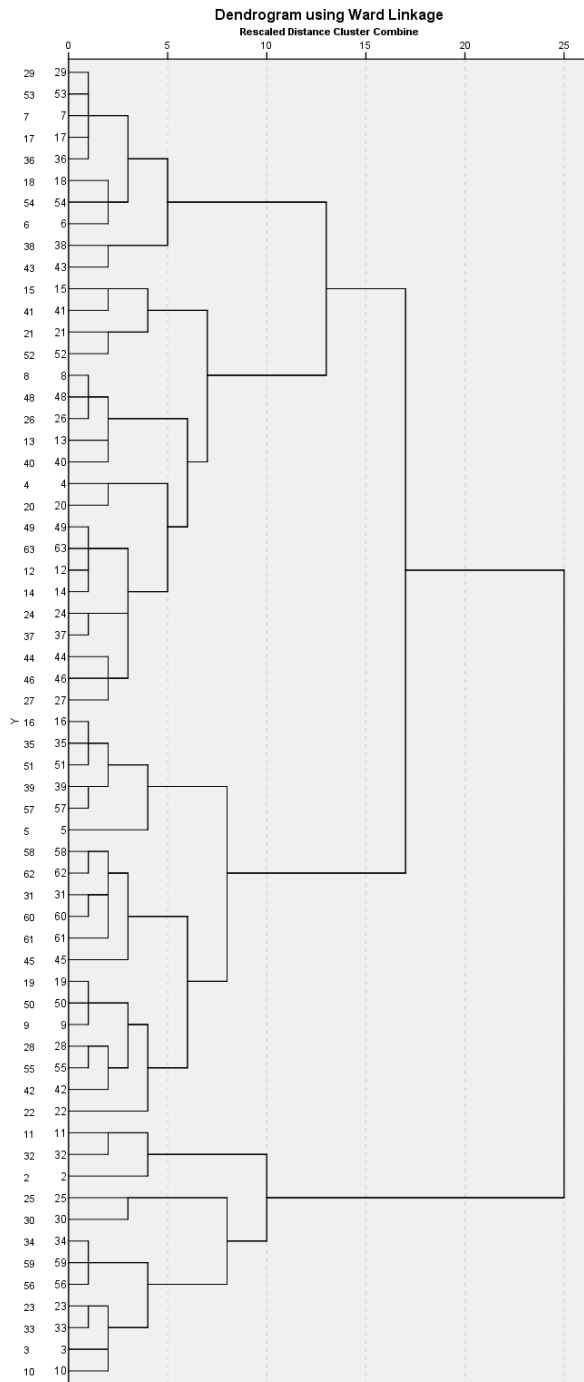
*Függelék*

## Kérdőív:

- 1) az a legnehezebb, hogy mindenkinek megmondani az igazat
- 2) a legnehezebb, hogy mindenkinek megmondani az igazat
- 3) az a legnehezebb, mindenkinek megmondani az igazat
- 4) a legnehezebb mindenkinek megmondani az igazat
- 5) hogy mindenkinek megmondani az igazat, az a legnehezebb
- 6) hogy mindenkinek megmondani az igazat, a legnehezebb
- 7) mindenkinek megmondani az igazat, az a legnehezebb
- 8) mindenkinek megmondani az igazat a legnehezebb
- 9) azon gondolkodom, hogy mindenkinek megmondani az igazat
- 10) hogy mindenkinek megmondani az igazat, azon gondolkodom
- 11) az a mérce, hogy kitörni a magyar valóságból
- 12) hogy kitörni a magyar valóságból, az a mérce
- 13) az a cél, hogy túlélni
- 14) hogy túlélni, az a cél
- 15) fölmerül az a kérdés, hogy tűrni vagy lázadni
- 16) hogy tűrni vagy lázadni, az a kérdés merül föl
- 17) fölmerül az a kérdés, hogy meddig tűrni
- 18) az a legnehezebb, hogy mikor mondani meg
- 19) az fontos volt, hogy megismerkedni az elnökkel
- 20) fontos volt, hogy megismerkedni az elnökkel
- 21) azt szeretném, hogy alaposan ledöngölni a padlót
- 22) az kellene, hogy kifesteni a kamrát



4. ábra: Nyelvészek klaszterelemzése – A koefficiens megemelkedik *b* és *n* klasztereinek összevonásakor, tehát ezek jelentősen különböznek



5. ábra: A diákok klaszterelemzése