

## 2.

# A társasnyelvészet módszertani alapjai

A társasnyelvészet elméleti alapvetéséből – azaz hogy a valóban létező nyelvet kell megfigyelnünk – következik, hogy a nyelvész nem elégedhet meg saját nyelvi intuíciójával, amikor a nyelvet leírja, hanem a valódi nyelvhasználatból kell adatokat gyűjtenie. Sőt: William Labov szerint a nyelvészek kifejezetten „béna kacsák” vagy „kibicek” még saját nyelvhasználó közösségükben is, mert tanulmányaik, kutatásaik, a nyelvről gyűjtött sok információ befolyásolja nyelvérzéküket, intuícióikat. (Labov a *lame 'sánta, béna'* kifejezéssel írta ezt le, az afroamerikai angolból kölcsönözve a szót, ahol azokat nevezik így, akik egy-egy közösség periferikus tagjai.)

A nyelvészeti megállapításokhoz tehát olyan adatok tömegére van szükségünk, amelyek valódi nyelvhasználók természetes nyelvhasználatából származnak. Ez azonban önmagában nem elég: ahhoz, hogy megállapításaink megalapozottak legyenek, kutatásunknak számos föltételnek kell megfelelnie. Itt most természetesen csak áttekintést tudunk erről adni – egész könyveket írnak az empirikus nyelvészeti vizsgálatok módszertanáról –, hozzátevé, hogy mindezt igazából a gyakorlatban lehet elsajátítani.

## Az empirikus nyelvészet alapfogalmai

Az empirikus nyelvészeti kutatások alapvető módszertani eljárásainak megismerése előtt tisztáznunk kell néhány alapfogalmat: hogy mi a változó, a változat és a változószabály, valamint hogy milyen gyakorlati, illetve etikai nehézségekkel kell szembenéznünk lényegében minden empirikus nyelvészeti vizsgálat lefolytatásakor.

## Változó, változat, változószabály

A társasnyelvészet egyik alapfogalma a változó: többnyire a változók viselkedését vizsgáljuk, amikor megpróbáljuk megállapítani az egyes nyelvi formák társas jelentését. **Változóknak** azoknak a nyelvi formáknak az együttesét nevezzük, amelyek funkciója azonos, de társas jelentésük eltérő. Az azonos változóhoz tartozó, de eltérő társas jelentéssel rendelkező formák a változó **változatai**. A magyarban változó például az egyes szám első személyű feltételes módú toldalak: két változata van, a sztenderd *-nék* (pl. *olvasnék*), illetve a nemsztenderd *-nák* (pl. *olvasnák*).

A változóknak három típusát szokás megkülönböztetni. Az **indikátor** változatai nem mutatnak stílusrétegződést, azaz nem köthetők a formálisabb, illetve informálisabb beszédhelyzetekhez, használatukat elsősorban a kontextus befolyásolja. Ilyen lehet a magyarban a vegyes hangrendű szavak egy részének toldalékolása (pl. *fotelban/fotelben*, *Ágnesnak/Ágnesnek*), amelynek előfordulását befolyásolhatja (de nem determinálja), hogy előtte közelre vagy távolra mutató névmás áll (pl. az *abban* után a *fotelban* gyakoribb lehet, mint a *fotelben*).

A beszélők az ilyen változók létéről többnyire nem is tudnak. A **marker** kötődhet csoporthoz is, beszédhelyzethez is, azaz változatainak megoszlása lehet jellemző egyes nyelvhasználó közösségekre, illetve formális vagy informális helyzetekre, de a beszélők nem fűznek a változatokhoz nyílt értékítéletet. Marker például a *hol* kérdésre válaszoló *-ba(n)/-be(n)* változó, mert a beszélők gyakrabban használják az *-n*-re végződő változatot a formális, az *-n* nélküli alakot informális beszédhelyzetben. A **sztereotípiá** szintén kötődhet csoporthoz vagy stílushoz, de ennek változatairól a beszélőközösségben nyílt véleményt fogalmaznak meg. Sztereotípiá például a *-t* végű igék határozott ragozásának kijelentő módja: ennek nemsztenderd, „szukszükölő”, illetve „sukszükölő” ragozásához a magyar beszélőközösségben nagyon határozott vélemények („műveletlen”, „bunkó”, „helytelen” stb.) tapadnak.

Egy-egy változó több típusba is tartozhat egyszerre: ugyanaz a változó lehet sztereotípiá a nyelvközösség egy része számára, míg másoknak marker, de lehetnek olyanok is, akiknek a beszédében indikátorként viselkedik. Például a *hol* kérdésre válaszoló *-ban/-ben* változó (nemsztenderd változata a *-ba/-be*) a magyar anyanyelvűek egy részének sztereotípiá: az iskolában megtanulták, hogy aki a nemsztenderd változatot használja, az „műveletlen”, esetleg „nem tud rendesen magyarul” vagy egyszerűen „helytelenül beszél”, és ezt az értékítéletet maguk is megfogalmazzák időnként. Mások figyelnek arra, hogy formális beszédhelyzetekben a sztenderd

változatot használják, de informális beszédhelyzetekben természetesnek tartják a nemsztenderd változatot. És vannak olyanok is, akik számára a beszédhelyzet formalitásához sem kötődik a változó használata, de választásukat befolyásolhatja a nyelvi kontextus, például az, ha a változót tartalmazó szó közvetlen környezetében előfordult a *-ban/-ben*.

A változók előfordulásának valószínűségét írják le a **változószabályok**: azt, hogy egy változó egyes változatai milyen eséllyel jelennek meg meghatározott jegyekkel rendelkező beszélők meghatározott beszédhelyzetben elhangzó szövegeiben. A társasnyelvészet szabályai ugyanis nem kategorikusak, hanem valószínűség-jellegűek: azaz nem arról beszélünk, hogy egy-egy forma egy-egy nyelvváltozatnak része vagy nem része, hanem arról, hogy a változó egyes változatai egy meghatározott korú, nemű, végzettségű, foglalkozású, földrajzi kötődésű stb. beszélő beszédében a nyelvterület meghatározott részén meghatározott beszédhelyzetben milyen valószínűséggel fordulnak elő.

## A megfigyelői paradoxon

A természetes nyelvhasználatot nem könnyű megfigyelni, különösen nem könnyű rögzíteni, mert a beszélők megváltoztatják természetes nyelvi viselkedésüket, ha észreveszik, hogy figyelik beszédüket. A társasnyelvészek azonban éppen azt szeretnék megfigyelni, hogyan viselkednek az emberek akkor, amikor nem figyelik őket – ezt a módszertani problémát nevezik **megfigyelői paradoxonnak**.

Az adatgyűjtéskor használt módszertani fogások jelentős részét éppen azért hozták létre, hogy ezt a nehézséget leküzdjék, többnyire úgy, hogy eltereljék valahogyan az adatközlők figyelmét arról, hogy megfigyelik a beszédüket. Ennek egyik, talán legegyszerűbbnek tűnő módja, hogy másra tereljük az adatközlők figyelmét: például úgy építünk föl egy szintaktikai szerkezetek használatát vizsgáló kérdőívet, mintha az adatközlők szókincsét vizsgálnánk vele; vagy olyan témáról kezdünk beszélgetni, amelyek az adatközlőt közelebről érintik, így az érzelmi közelség miatt sokkal inkább magára a témára s nem a nyelvhasználatra összepontosítanak – erre az Egyesült Államokban a „halálveszedelem”-témát vetették be az első kutatásokban (mert a „volt-e már halálveszedelemben?” kérdés többnyire eléggé fölzaklatja az embereket) –, természetesen a vizsgált közösség kultúrájától függ, hogy mi lehet az a téma, amely érzelmi kötődései miatt alkalmas arra, hogy az adatközlőkből előcsalogathassuk legtermészetesebb beszédmódjukat.

## Etikai követelmények

Fölmerülhet, hogy a megfigyelői paradoxont könnyű lenne áthidalni, ha titkos felvételeket készítenénk – ehhez ma már bőségesen rendelkezésre állnak különféle hang-, vagy akár képrögzítő technikák. Ezt azonban a társasnyelvészet etikai előírásai nem engedik meg: az adatközlőket nem csaphatjuk be, tudtuk nélkül nem gyűjthetünk tőlük adatokat.

Vannak más etikai előírások is – ezek hasonlóak az antropológia és a szociológia empirikus vizsgálataiban alkalmazott szabályokhoz. Vigyáznunk kell például arra, hogy adatközlőink anonimitását megvédjük: nevüket az adatok közzétevésekor nem adjuk meg, sőt, ha valamilyen személyes információból azonosíthatók lennének, akkor ezeket meg szoktuk változtatni vagy nem közöljük. Az adatközlő természetesen az adatfelvétel után is nyilatkozhat úgy, hogy mégsem akar részt venni a vizsgálatban, ilyenkor a tőle gyűjtött adatokat meg kell semmisíteni.

## Mit vizsgálunk?

Természetesen minden empirikus vizsgálat azzal kezdődik, hogy pontosan körülírjuk, mit akarunk egyáltalán vizsgálni, miről akarunk adatokat gyűjteni, s mindezt minek érdekében tesszük, vagy másképp fogalmazva, hogy milyen kérdésekre akarunk majd választ kapni. Ezt követően választhatjuk ki a megfelelő gyűjtési módokat, majd ezek ellenőrzése és a szükséges korrekciók elvégzése után következik az adatgyűjtés, aztán az adatok rögzítése és elemzése.

## A kiinduló hipotézis megfogalmazása

A kutatás a kiinduló hipotézis megfogalmazásával kezdődik: ekkor írjuk körül, mi az, amit vizsgálni akarunk. Ehhez fölhasználhatjuk korábbi kutatások eredményeit és saját megfigyeléseinket is. A kiinduló hipotézis tartalmazza azt az elképzelést, hogy a megvizsgálandó jelenség hogyan működik, a föltételezett eredmények fölvázolását, és annak megokolását is, hogy miért érdemes elvégezni a tervezett kutatást: mert újat hoz, mert módosít egy korábbi ismeretet, pontosítja, megerősíti vagy cáfolja egy másik vizsgálat következtetéseit, mert szükségünk van a válaszra valamilyen probléma megoldásához stb. A kutatás lényegében a kiinduló hipotézis empirikus ellenőrzése.

## Függő és független változók

Az empirikus kutatásokban **függő változónak** nevezzük azokat a változókat, amelyek változatainak megoszlását vizsgáljuk, **független változónak** pedig azokat a jegyeket, amelyeket a vizsgálatban mint a függő változó viselkedését lehetségesen befolyásoló tényezőket veszünk figyelembe. Ha például arra vagyunk kíváncsiak, hogy az ikes ígéket nagyobb eséllyel ragozzák-e ikesen a felsőfokú végzettségűek, mint a középiskolát végzettek és a csak általánost végzettek, akkor a függő változó az ikes/iktelen forma (pl. *eszem/eszek*), a független változó pedig az iskolázottság, ez esetben három lehetőséggel (általánost végzettek, érettségizettek, diplomások). Egy-egy kutatásban általában több függő változó viselkedését vizsgáljuk, több független változót alkalmazva.

Nyilvánvaló, hogy minél több független tényezőt veszünk egy-egy vizsgálatban figyelembe, annál több adatközlőre lesz szükségünk, hogy később megfelelő elemszámmal rendelkezünk az érvényes statisztikai elemzéshez, és annál költségesebb és nehezkesebb lesz a kutatásunk. Ugyanakkor az is nyilvánvaló, hogy minél kevesebb független változót építünk be a kutatásba, annál nagyobb az esélye, hogy tévedjünk, és valós összefüggések helyett valamilyen látszólagos hatást mutassunk ki. Ezért fontos a megfelelő arányok s főként a releváns tényezők megtalálása – ebben segíthetnek a korábbi kutatások és saját előzetes megfigyeléseink.

## Formális és informális nyelvhasználat

Mindannyian tudjuk, hogy másképpen beszélünk, amikor úgy érezzük, hogy valamilyen körülmény miatt „jobban” kell beszélnünk, több figyelmet fordítunk a beszédünkre, és nagyobb eséllyel választjuk egy változó formális változatát, mint amikor oldottan beszélgetünk. A beszédreflexió mértékének nagy szerepe van nyelvi viselkedésünk alakulásában, ezért minden társasnyelvészeti adatgyűjtés fontos eleme a beszédhelyzet formalitásának ellenőrzése. Ha azt akarjuk, hogy adataink összemérhetőek és eredményeink érvényesek legyenek, akkor adatközlőinktől azonos beszédhelyzetben kell adatokat gyűjtenünk.

Célszerű úgy tervezni a kutatást, hogy ugyanazoktól az adatközlőktől többféle beszédhelyzetben gyűjthessünk adatokat – azaz független változóként vezessük be a beszédreflexió mértékét (a formalitást) is. Így nemcsak egymással lesznek összehasonlíthatók a formális és informális

helyzetbeli választások, de ez az eljárás megvilágíthat más nyelvhasználati összefüggéseket is. Lehet például, hogy egy-egy csoport – például a fiatalabb és idősebb férfiak – választásai nem mutatnak lényeges eltéréseket formális helyzetben, mikor nagy a beszédrefigyelés mértéke, de igen, ha az informális beszédhelyzetből származó adataikat vetjük össze.

## A kutatás megtervezése

A kutatást úgy tervezzük meg, hogy a kiinduló hipotézisben megfogalmazott kérdésekre minél pontosabb és minél adekvátabb válaszokat kapjunk. Ehhez választjuk ki a legmegfelelőbb vizsgálati módokat és eszközöket, szem előtt tartva, hogy a lehető legjobban kontrollálható vizsgálatot kell terveznünk.

## Ellenőrizhetőség

Tudnunk kell, kik azok, akiktől az adatokat gyűjtöttük – fontos lehet a nyelvhasználatban az életkor, nem, iskolázottság, csoporthovatartozás, a szülők iskolázottsága, hogy az adatközlő hol élt élete során stb. Ismernünk kell azt a helyzetet, amelyből az adat származik, hogy az adatközlő mennyire figyelhetett beszédére. Ha statisztikai elemzést akarunk végezni, megfelelő számú adatközlőre lesz szükségünk, hogy statisztikailag értelmezhető adatokat kapjunk. A kutatást úgy kell leírni, hogy ha mások hasonló kutatást szeretnék végezni, esetleg megismételni a miénket, akkor pontosan rekonstruálni lehessen, mi hogyan történt, milyen adatközlői mintától való az adataink, milyen módszerekkel gyűjtöttük őket.

## A kutatási eszközök kiválasztása

Attól függően, hogy mit akarunk vizsgálni, tervezhetünk kvantitatív és kvalitatív vizsgálatot vagy akár a kettő ötvözetét. Az egész kutatást meghatározza, hogy sikerült-e a kutatás tárgyához illeszkedő, annak megfelelő vizsgálati eszközöket választanunk, így a tervezési szakasz akár jóval hosszabb is lehet, mint maga az adatfölvétel.

## Kvantitatív vizsgálatok

**Kvantitatív vizsgálatokat** akkor szoktunk tervezni, amikor nagy mennyiségű adatra van szükségünk, mert statisztikailag értelmesen elemezhető **korpuszt** (adatgyűjteményt) szeretnénk kapni. Az ilyen fölmérések többnyire általánosító összefüggések föltárására alkalmasak; jól alkalmazhatók folyamatban lévő nyelvi változások föltérképezésére, az egyes társadalmi csoportok nyelvhasználatának leírására. Ugyanakkor a finomabb motivációkat, az eltérések okait nem mindig tudjuk meg az ilyen kutatásokból.

A kvantitatív vizsgálatok adatait a legtöbbször magnetofonnal, máskor kérdőívvel gyűjtik. Sok adat összegyűjtése a cél, és a kérdőív lekérdezése és kódolása is gyorsabb, mint a kötetlenebb beszélgetéseké, és magáról az adatközlőről is egyszerűen lehet megtudni azokat az információkat, amelyeket független változóként szeretnénk majd használni, ugyanakkor a spontán – vagy afelé közelítő – beszélgetések adatai, a természetesebb kontextus miatt, igen értékesek. A két adatfölvételi mód kombinálható is.

A kérdőív tartalmazhat nyitott és zárt kérdéseket. Az előbbiekre az adatközlők szabadon válaszolnak, a zárt kérdésekre viszont megadott válaszlehetőségek közül kell választaniuk, illetve valamilyen skálának megfelelően kell meghatározniuk saját válaszukat, véleményüket. A nyelvészeti kérdőívekben többnyire úgy állítják össze a kérdéseket, hogy az informális és formális beszéd közötti skáláról (az előbit minél természetesebb beszélgetésből, az utóbbit pl. minimális párok fölolvasásával lehet begyűjteni), illetve a beszéd percepciójáról (azaz egy-egy változat elfogadásáról, pl. „hibajavító” feladattal), illetve produkciójáról (pl. mondatkiegészítéssel) is legyenek adatok.

A kérdőívek összeállításakor vigyázni kell arra, hogy az adatközlő lehetőleg ne találja ki, valójában mi érdekel bennünket (ezzel valamelyest csökkenthető annak esélye, hogy olyan válaszokat adjon, amelyeket az elvárt válasznak vél), ezért olyan „vattakérdéseket” is beépíthetünk a kérdőívbe, amelyek adatait nem dolgozzuk föl, mert irrelevánsak a kutatás szempontjából. Fontos, hogy a kérdések megfogalmazásával, illetve sorrendjével se befolyásoljuk a válaszokat.

## Kvalitatív vizsgálatok

Ha nem az a célunk, hogy statisztikai összefüggéseket keressünk, hanem részletesebben szeretnénk megismerni egy jóval kisebb adatközlői csoport nyelvhasználatát, gondolkodását bizonyos kérdésekről, akkor célszerű

**kvalitatív vizsgálatot** terveznünk. Ebben az esetben a kapott eredmények nem lesznek ugyan általánosíthatók, de több oldalról, alaposabban ismerhetjük meg, amire kíváncsiak vagyunk. Amikor érvelési módokra, egy-egy jelenség mögött rejlő ideológiákra, attitűdökre, véleményekre vagyunk kíváncsiak, akkor ezek megismeréséhez kvalitatív vizsgálatot célszerű terveznünk.

A kvalitatív vizsgálatban készíthetünk **irányított interjúkat** az adatközlőkkel. Ez a lehető legjobban próbál közelíteni a spontán beszélgetéshez, szintén a megfigyelői paradoxon hatásainak kiküszöbölése érdekében, de a terepmunkás meghatározott témákat mindenképpen fölvet, illetve meghatározott kérdéseket mindenképpen megkérdez. Ezek sorrendje azonban kötetlen: az egyes téma- és kérdésmodulokat a terepmunkás szabadon mozgathatja a beszélgetésben, hogy a beszélgetés a lehető legtermészetesebben alakuljon. Arra azonban ilyenkor is figyelni kell, hogy a megbeszéltek témák sorrendje befolyásolhatja az adatközlők véleményét. „Irányításra”, azaz a modulokra, kötelezően fölített kérdésekre azért van szükség, hogy a különböző interjúkból származó adatok később összehasonlíthatók legyenek.

Kvalitatív adatokat gyűjthetünk a **résztevő megfigyelés** módszerével is. Ezt az antropológiából átvett vizsgálati módszert alkalmazva a kutató beépül a vizsgált közösségbe (hónapokra, nem ritkán évekre), és belülről, a közösség tagjaként nyílik módja megfigyelni a közösség nyelvi viselkedését. Ez a módszer nyújtja a kutatónak a legtöbb lehetőséget, hiszen ha befogadja őt a vizsgált közösség, akkor nagyobb eséllyel viselkednek a jelenlétében természetesen, még akkor is, ha tudják, hogy rögzíti a beszélgetést. A közösség belső viszonyait, dinamikáját, tagjainak pozícióját – azaz a kapcsolathálózat jellemzőit – is így ismerheti meg a legnagyobb biztonsággal.

## Ügynökvizsgálat

A megfigyelő paradoxon hatásait jelentősen csökkenthetik az **ügynökvizsgálatok**, ezeket főként **nyelvi attitűdök** fölterképezésére szokták alkalmazni. Az ügynökvizsgálat lényege, hogy a beszélőknek különféle nyelvváltozatokon vagy nyelveken elmondott szövegeket játszanak le, és arra kéri őket, jellemezzék az elhangzott szövegek beszélőit különféle szempontok szerint. Ezek többnyire legalább két nagyobb csoportba sorolhatók, az egyik a barátságosság (azaz hogy mennyire kedves, barátságos, „jófej” a beszélő), a másik a kompetencia (mennyire hozzáértő, értelmes, rátermett stb.).



Az adatközlők nem tudják, hogy két szöveg ugyanattól a beszélőtől származik, de a szövegeket eltérő nyelvváltozatban vagy nyelven mondja el. Ez esetben tehát a beszélő neme, életkora, hangja ugyanaz, így ha az adatközlők az „ügynök” két alakját eltérően értékelik, az nagy valószínűséggel a két eltérő (nyelv)változattal szemben kialakított attitűdjeik miatt van.

## Próbavizsgálat, korrekció

A kiválasztott vizsgálati módszereket kisebb adatközlői mintán szokás tesztelni: érthetőek-e a kérdések, nem befolyásolják-e a válaszadást, a zárt kérdések előre megadott válaszai megfelelőek-e, nem túl sok vagy nem túl kevés-e a skálán megadandó értékek száma, és egyáltalán, hogy alkalmas-e a vizsgálati eszköz arra, hogy a kiinduló hipotézisben megfogalmazott kérdésekre választ kapjunk a segítségükkel. A **próbavizsgálat** (az angol elnevezést követve gyakran *pilot*-nak is nevezik) tanulságainak elemzése után elvégezhetjük a szükséges korrekciókat, mielőtt a költségesebb és időigényesebb valódi adatgyűjtéshez hozzákezdünk.

## Az adatközlői minta

Szintén nagyon gondosan kell megtervezni, hogy kiktől gyűjtjük majd az adatokat – ez természetesen szintén attól függ, mi a vizsgálat célja. Ha a teljes beszélőközösségre érvényes megállapításokat akarunk tenni, akkor a mintának reprezentálnia kell a közösséget, ha más szempontjaink vannak (például egy-egy független változó hatásának vizsgálata vagy egy kisebb beszélőközösség szokásainak megismerése), akkor más típusú mintákkal dolgozunk.

## Rétegzetten reprezentatív minta

Ritkán fordul elő, mert igen költséges kutatási mód, hogy sikerül olyan mintával dolgoznunk, amelynek összetétele meghatározott szempontok szerint pontosan tükrözi a vizsgált nagyobb beszélőközösség arányait. Ezt a mintát hívjuk **rétegzetten** (a megadott szempontok szerint) **reprezentatív**nak, eredményei, éppen a megfelelő arányok miatt, rávetíthetők a teljes beszélőközösségre. Tehát: ha például a magyarországi felnőtt lakosság nyelvhasználatát vizsgáljuk azt illetően, hogy a kiválasztott változók hasz-

nálatát befolyásolják-e egyes szociológiai jellemzők, akkor a mintánkban ezeknek a kiválasztott tényezőknek (független változóknak) megfelelően ugyanolyan arányban kell szerepelniük az adatközlőknek, mint a teljes lakosság körében. A legegyszerűbb példával: ha kíváncsiak vagyunk a nem mint független változó hatására, és a magyar lakosság 52 százaléka nő, akkor a mintában szintén 52 százaléknak kell lennie a nők arányának. Ha az is érdekel, hogyan oszlik meg a vizsgált függő változó használata Budapest-vidék viszonylatban, akkor a minta ötödének budapestinek, négyötödének vidékinek kell lennie, és így tovább, a többi számunkra érdekes független változóval. E föltételeknek természetesen egyszerre kell teljesülniük. Ugyanakkor a mintát véletlenszerűen kell kialakítani, azaz a teljes beszélőközösség minden tagjának azonos esélye kell hogy legyen arra, hogy bekerüljön az adatközlői mintába.

## Kvótaminta

A kvótaminta jól körülhatárolható, lehetőleg kevés, de ellenőrzött szempont szerint elkülönülő, azonos létszámú adatközlőből álló csoportokat tartalmaz. Nem reprezentatív tehát, az eredmények nem vetíthetők ki a beszélőközösség egészére, de kiválasztása gyorsabb és olcsóbb, a kutatás lefolytatása is egyszerűbb ilyen mintával. Ha például arra vagyunk kíváncsiak, van-e eltérés valamilyen nyelvi forma használatát illetően a nők és a férfiak, illetve a diplomások és az érettségivel rendelkezők között, akkor a mintában ugyanannyi nőnek, illetve férfinak, diplomásnak, illetve érettségizettnek kell szerepelnie. Ha az így létrehozott csoportok nyelvhasználatában eltéréseket találunk, akkor azok jó eséllyel függhetnek össze a kiválasztott tényezőkkel. Nagyon fontos azonban, hogy a kvótamintás vizsgálatok elemzésekor megfelelően szűkre szabjuk következtetéseink érvényességét.

## Hólabda-módszerrel kialakított minta

A kvalitatív elemzésekben alkalmazott mintaképzési eljárás: amikor egy ismerősünk ismerőseit, majd azok ismerőseit vonjuk be a kutatásba. Ezzel a módszerrel természetesen nem lehet a teljes beszélőközösségre érvényes megállapításokat tenni, elsősorban akkor alkalmazható, ha kisebb közösségek társas kapcsolathálózatára és ennek (nyelvi) viselkedésbeli hatásaira vagyunk kíváncsiak.

## Anonim elektronikus (internetes) adatgyűjtés

Szakdolgozatokban, de akár konferenciákon, publikációkban is gyakran lehet találkozni olyan adatokkal, amelyeket internetes kérdőívvel gyűjtöttek névtelen adatközlőktől. Kétségtelen, hogy ez az adatgyűjtés leggyorsabb és legkényelmesebb módja, de nagy hátránya, hogy az adatközlőkről ilyenkor semmit nem tudunk, akkor sem, ha rákérdezzük, hiszen nem tudjuk ellenőrizni, hogy valójában ki töltötte ki a kérdőívet; ellenőrizhetetlen, hogy mennyire értették a kérdéseket, mennyire vették komolyan a kérdőívet, és lehet, hogy mindannyian egy-két baráti közösségből kerülnek ki, de még azt sem tudjuk, hogy köztük mi a viszony. Ezért az így gyűjtött adatok csak rendkívül korlátozottan használhatók – leginkább arra lehetnek alkalmasak, óvatosan és megszorításokkal –, hogy nagyobb, jobban megtervezett kutatás előkészítésében fölhasználjuk őket.

## Az adatok földolgozása

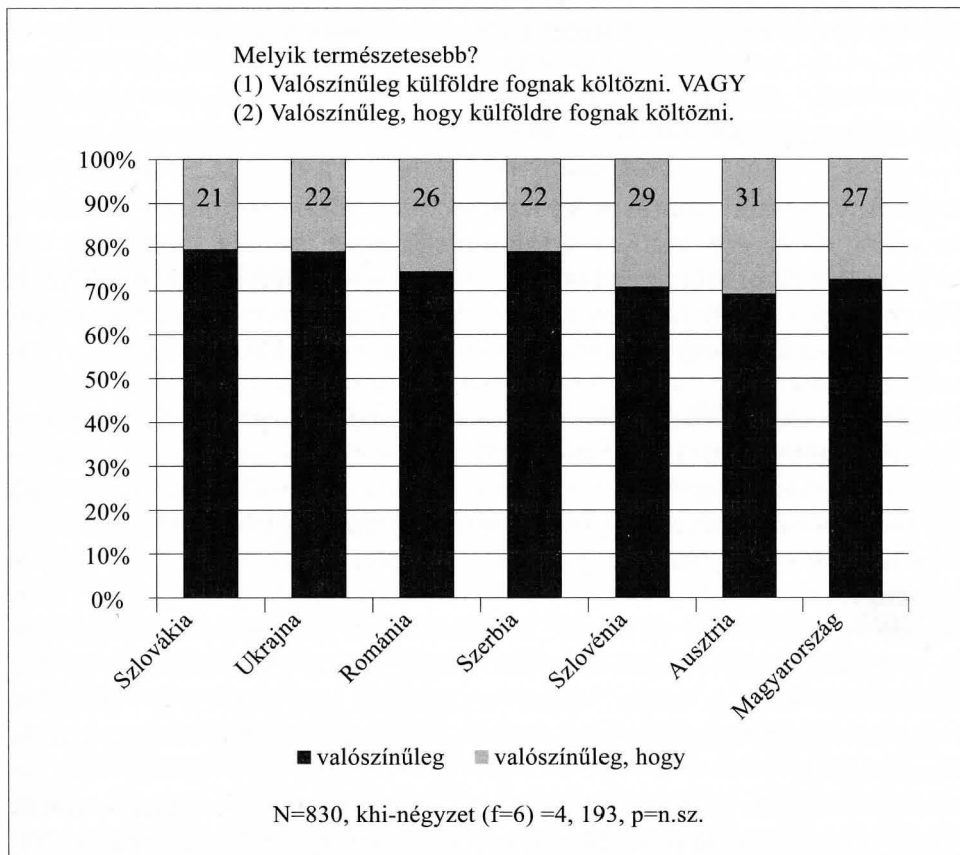
Az összegyűjtött adatokat értelemszerűen rögzíteni és kódolni kell ahhoz, hogy később elemzéseket végezhesünk a korpuszon.

## Lejegyzés, kódolás és statisztikai elemzés

Az adatok rögzítése különböző, akár kifejezetten a kutatáshoz tervezett adatbázis-kezelő szoftverekkel történik. Az interjúk lejegyzésében többféle lejegyzési hagyományt követhetünk, akár módosíthatjuk is valamelyiket. A szövegben ilyenkor a vizsgálat szempontjából fontos formai vagy tartalmi elemeket kódokkal látjuk el, ez jelentősen megkönnyíti később az elemzést, de természetesen némiképpen korlátoz is, mert előfordulhat, hogy az azonos kódok miatt egymással nem tökéletesen azonos elemek is azonos kategóriába kerülnek.

A statisztikai elemzést a szociológusok által is használt szoftverek, leggyakrabban az SPSS (*Statistical Package for the Social Sciences*) nevű program segítségével szokták végezni. A statisztikai elemzés a társadalomtudományokban elfogadott, a matematikainál valamivel lazább szabályok szerint történik (a legtöbbször a sokkal kisebb elemszámok miatt kell engedelményeket tenni). De még ez is azt jelenti, hogy egy-egy összefüggést csak akkor tartunk érvényesnek, ha a megfelelő statisztikai próba (pl.  **$\chi^2$ -próba**, azaz khi-négyzet-próba) szerint 5 százaléknál kisebb az esélye annak, hogy

az adatok együttesfordulása a véletlen műve. Másképp fogalmazva: legalább 95 százalék az esélye annak, hogy a függő változó viselkedése és az adott független változó között összefüggés van.



3. ábra:

A „valószínűleg, hogy” változóra vonatkozó elemzések egyikének grafikonja (a változó megítélése a magyarországi és a határontúli magyar változatokban) az MNSZV alapján [forrás: Kontra Miklós (szerk.) *Nyelv és társadalom a rendszerváltás kori Magyarországon*, 2003: 155]

A statisztikai elemzés persze jóval több, mint az egyszerű gyakoriság megállapítása (azaz hogy például hányszor fordul elő a vizsgált változó egyik vagy másik változata), és nemcsak azt vizsgálja, hogy egy-egy független változó van-e hatással a függő változó viselkedésére, hanem azt is, hogy

vannak-e kölcsönhatások az egyes független változók között – ezt sokszor **keresztáblával** mutatják be. Az egyes változók jellemző együttlőfordulási csoportjainak föltérképezésére szokták alkalmazni a **faktoranalízist**, ennek segítségével a már meglévő független változók jellemző csoportjait egy-egy újabb, egyesített változóvá is összevonhatjuk.

## Következtetések

A társasnyelvészeti kutatások igyekeznek minél több szempontot figyelembe venni, de az eredmények értelmezésében ezt mindig szem előtt kell tartanunk, amikor általánosítva fogalmazunk – vagy amikor ilyen elemzést értelmezünk –, hogy minden ember élettörténete más, mások a személyes viszonyulásaik egy-egy beszédhelyzethez, nyelvváltozathoz. Ehhez hasonlóan, amikor a kutatás eredményeit statisztikai elemzés után közöljük, szintén nem kategorikus állításokat teszünk: ha jól végeztük a vizsgálatot és az elemzést, akkor a kapott összefüggések érvényesek, de nem kizárólagosak: mindig átlagokat kapunk. Azaz, amikor például azt mondjuk, hogy egy tényező hatással van valamelyik nyelvi forma megjelenésének gyakoriságára, az nem azt jelenti, hogy mindig, minden helyzetben, minden egyes beszélő viselkedésére hatással van, hanem hogy tendenciaként igaz, amit mondunk, de lehetnek eltérések. Természetesen azt sem jelenti, hogy csak ez a tényező van hatással az adott nyelvi változó viselkedésére. Fontos lehet tudnunk, hogy milyen nagyok az átlagtól való eltérések – lehetnek csekélyebbek és akár igen nagyok is –, ennek átlagos mértékét a **szórás** mutatja meg.

Nagyon fontos, hogy az adatokban megmutatkozó **együttlőfordulás** nem jelent egyben föltétlenül ok-okozati összefüggést is. Ha tehát azt tapasztaljuk, hogy egy nyelvi változó egyik változata szignifikánsan (statisztikai értelemben érvényesen) gyakrabban fordul elő együtt (korrelál) a független változók valamelyik változatával, akkor ez nem jelenti azt, hogy a független változó adott változata *okozza* a függő változó viselkedését. Sokszor tapasztaljuk, hogy valójában mindkét változó viselkedését egy harmadik, rejtett változó okozza.