

Értelmezhetőségi tényezők a vulgáris latin nyelvtani nemi rendszer vizsgálatában

Bevezetés, problémafelvetés

A nyelvtani nemi rendszernek a vulgáris latin nyelvfejlődés során végbement átalakulása egy kifejezetten érdekes terület nyelvészeti szempontból. Az erről a változásról információval szolgáló adatok ugyanis nem kifejezetten gyakoriak, de megfelelő eszközök alkalmazásával sok információval szolgálhatnak a folyamat pontos menetéről. Jelen dolgozat célja, amint arra a címe is utal, egy olyan változó megvizsgálása és kontrollálása, ami befolyásolhatja az erről az átalakulásról nyert képünket. Emellett bemutatunk és kiértékelünk egy lehetséges statisztikai modellt ezeknek az adatoknak a vizsgálatára.

A nyelvtani nemi (a továbbiakban: genusz¹) rendszer átalakulásáról több munka is született az utóbbi időben. Ezek közül a legfontosabb Loporcaro (2018) monográfiája, amelyben kiemelkedő részletességgel vizsgálja a genusz-rendszert a latinnal kezdve, de az egyes újlatin nyelveket is bemutatva. Különös figyelmet érdemelnek itt azok az általa bemutatott újlatin dialektusok, ahol valamilyen (esetlegesen megváltozott) formában megmaradt a semlegesnem (2018:12-14). Ez a nem ugyanis javarészt eltűnt az újlatin nyelvekben (2018:63-70). Emellett viszont a hazai kutatásban is jelentkezett érdeklődés ez iránt a változás iránt és két tanulmány is született, amik ezzel az átalakulással foglalkoznak (Szlovicsák 2022a és 2022b). Ezen tanulmányokban a szerző Loporcaro monográfiájával szemben már a feliratos anyag, és a Császárkori Latin Feliratok Számítógépes Nyelvtörténeti Adatbázisa² (a továbbiakban: Adatbázis) segítségével vizsgálja a genusz-rendszer átalakulását. Jelen dolgozatban követni fogjuk ezt a tanulmányt, és az Adatbázis segítségével nyerhető adatok mentén fogjuk vizsgálni a változást. Az említett tanulmányokkal szemben viszont törekedni fogunk egy zavaró változó kiszűrésére, ami lehetővé teszi az adatok biztosabb értelmezését, és ezen egyéb hatások jelentőségének megvizsgálását.

A feliratos anyag és az Adatbázis alkalmazása több szempontból is igen hasznos a jelen kutatás kapcsán. Egyrészt, amint erre Szlovicsák (2022a:47) is kitér, a feliratok megfelelő idő-

¹ Szemben Szlovicsák (2022a és 2022b) tanulmányaival, ahol nyelvtani nemi rendszer szerepel.

² Ld.: lldb.elte.hu (utolsó hozzáférés: 2023. 01. 05.). A dolgozatban szereplő adatok az Adatbázis 2022. 11. 20.-i állapotát tükrözik.

és térbeli meghatározottsága lehetővé teszi dialektológiai különbségek és nyelvtörténeti változások megfigyelését. Ezeken túl az is egy kifejezetten hasznos részét jelenti ennek az Adatbázisnak, hogy adott esetben nem csak egy, hanem akár két különböző kódon is rögzítésre kerülhet ugyanaz az adat. Azaz, ha egy genuszkeverés adott esetben mássalhangzó-keverésként is értelmezhető, akkor ez a tény rögzítésre kerül.³ Kiemelendő viszont, hogy a döntés, hogy egy adott értelmezés a kód vagy az alternatív kód mezőbe kerül nem jelent preferenciát.⁴ Az alternatív kód ugyanannyira érvényes, mint a fő kód,⁵ csak éppen ez került a második helyre. Ebből fakadóan az adatok egyszerűbb kezelése miatt az elemzés során a vizsgált genuszkeverési kódokat kezeltem a fő kódként, a további lehetséges értelmezést pedig alternatív kódként. A keltezéssel kapcsolatban kiemelendő még, hogy nagyon sok esetben pontos keltezés nem, csak egy intervallum adható meg. Ezen intervallumok esetén az adatokat, az Adatbázis legfrissebb datálási funkciójához⁶ hasonlóan az intervallum határpontjainak számtani közepét kezeltem az adatok keletkezési idejeként. Ezután a vizsgálatot azokra az adatokra szűkítettem le, amelyeknek az így kapott keltezése beleesik a Kr.u. 1. századtól a 7. század végéi tartó időszakba. Szlovicsák (2022b) tanulmányával ellentétben viszont nem kategorikus, hanem folytonos változóként kezeltem az adatok keltezését, ez így összetettebb statisztikai modellek alkalmazását is lehetővé tette.⁷ Mivel ez a típusú keltezés eltér az Adatbázis esetén eddig követett eljárástól,⁸ az eredmények értékelésekor be fogjuk mutatni ennek az előnyeit, megmutatjuk a korábbi eredményekkel való konzisztenciáját is, illetve kitérünk az egyes fenntartásokra is.⁹

Az alternatív kódok tehát az említett okokból jelentősen megnehezíthetik az adatok vizsgálatát és érvényességét. Tegyük fel ugyanis, hogy a 3. jegyzetben említett példához hasonlóan minden hím- és semlegesnem közötti keverést megmagyarázhatunk mássalhangzó keverésként. Ekkor, ha a vizsgálatba nem vesszük be az alternatív kódokat, úgy tűnhetne, mintha például területi vagy időbeli különbségekkel függene össze a genuszkeverések

³ Pl.: LLDB-110320: VEX[IL]LVM ARGENT | INSIGNEM = *vexillum argento insigne*. Ennek az adatnak a kódja: *masc. pro neutr.*, az alternatív kódja: *-ø > -m*.

⁴ Erről lásd az Adatbázis gyűjtési utasítását: II/1.2 https://lldb.elte.hu/admin/doc_guidelines.php (utolsó elérés: 2023. 01. 05.)

⁵ A *fő kód* kifejezést felváltva használjuk a *kód* kifejezéssel, különösen akkor, ha szeretnénk kihangsúlyozni, hogy nem az alternatív kódról van szó. Ez éppen akkor fontos, amikor kifejezetten a genuszkeverésekről van szó.

⁶ Ld.: az *Időszak [A]* funkciót: https://lldb.elte.hu/admin/search_2.php (utolsó elérés: 2023. 01. 05.)

⁷ Kategorikus változó szerepelne jelen esetben, ha a Keltezés változót különböző időszeitekre osztanánk, ld. pl. 8. jegyzet. Ha pedig folytonos változóként kezeljük a Keltezést, akkor feltesszük, hogy az 1 és 700 között bármilyen értéket felvehet.

⁸ Lásd például Adamik (2019:23-27) tanulmányát, ahol szintén két korszakra (Korábbi korszak: Kr.u. 1-3. század, Későbbi korszak: 4-7. század) bontva kerül megvizsgálásra a nyelvi változás.

⁹ Ezzel kapcsolatban például ld. Van de Velde–Petré (2020:346-347)

eloszlása, pedig világos, hogy nem erről lenne szó, hanem arról, hogy mássalhangzókeverések területi különbségeiből fakadna a genuszkeverések különbsége és nem magukból a területi különbségekből.¹⁰ Az alternatív keverések nélküli modell viszont nem tudná kimutatni az eltérések valódi magyarázatát. Amennyiben kiderülne, hogy tényleg csak az alternatív kódok határozzák meg a genusz-keverések eloszlását az is releváns eredmény volna, de nem egyenértékű azzal, amit például Szlovicsák (2022b:266) tanulmányában találunk, miszerint összefüggés volt megfigyelhető a területi eloszlás és a genuszkeverések eloszlása között. Ezen dolgozat célja tehát az alábbi: statisztikai eszközök segítségével kontrollálni az alternatív kódok hatását és ennek révén megvizsgálni, hogy érvényben maradnak-e a korábbi eredmények.

A genuszkeveréseken belül Szlovicsák (2022b:256-257) két fajta lehetséges értelmezést különböztet meg, az egyikre hiperkorrekt olvasatként, míg a másikra nem hiperkorrekt olvasatként hivatkozik.¹¹ Jelen dolgozatban szintén ezeket az elnevezéseket fogjuk követni, viszont elsődlegesen a nem hiperkorrekt értelmezéssel foglalkozunk és csak a korábbi eredményekkel való összevetés miatt mutatjuk be a hiperkorrekt értelmezést, és annak lehetséges problémáit. Így tehát három szintet különböztetünk meg a főkódon belül: *Masc/Fem*, *F/M > Neutr* és *N/F > Neutr*, ahol az első szint azokat a keveréseket tartalmazza, amelyek a hím- és a nőnem közötti genuszkeverések, a második szinten nem semlegesnemű szavak kerültek semlegesnembe, míg a harmadik szint nem más, mint a semlegesnem leépülésének valós jele, ahol semlegesnemű szavak más genuszba kerülnek.¹² A hiperkorrekten értelmezett főkódon belül megegyezik ezzel az első szint, viszont a másik kettő módosul *Masc/Neutr* és *Fem/Neutr* szintekre, ahol az első esetben a semlegesnem a hímnemmel keveredik, míg a második esetben a nőnemmel.¹³

¹⁰ Annyiban kiemелendő ez a hipotetikus példa, hogy a vizsgált alternatív kódok eloszlása nem változott jelentősen területenként. Ettől viszont eltérő jellemző volt megfigyelhető a hiperkorrekt módon értelmezett adatok esetén.

¹¹ Például a LLDB-111677: HOC TITVLVM = *hunc titulum* adat főkódja: *neutr. pro masc.*, hiszen a hímnemű *titulus* alak mellett a mutató névmás semlegesnemű *accusativusa* (illetve lehet *nominativus* is) áll. Az ilyen típusú adatok (és hasonlóan a *neutr. pro fem.* kódú adatok) értelmezhetőek hiperkorrekt és nem hiperkorrekt módon is. A hiperkorrekt értelmezés szerint azért áll semlegesnem a hímnem helyett, mert gyenge a megkülönböztetés a két nem között, hiszen a vulgáris latin során a semlegesnem eltűnt (ezt az olvasatot Loporcaro [2018: 12–14] is nehezményezte). Ezzel szemben egy nem hiperkorrekt olvasat elfogadja, hogy az ilyen típusú genuszkeverések (ti. *neutrum* áll valamelyik másik nem helyett) egy valóban létező, megmaradási tendenciából fakadhattak, ahol a semlegesnem egyes esetekben a várttal ellentétben új adatokkal is bővíthetett és nem tűnt el. Ez a második olvasat annak fényében különösen fontos, hogy Loporcaro számos dialektust fedezett fel, ahol a semlegesnem megkülönböztetése valamilyen módon megmaradt. Ld.: Loporcaro (2018:60)

¹² Az adatbázis szerinti kódok alapján: *Masc/Fem* = *fem. pro masc.*, *masc. pro fem.* és *fem. pro communi.*; *F/M > Neutr* = *neutr. pro masc* és *neutr. pro fem.*; *Neutr > F/M* = *masc. pro neutr.* és *fem. pro neutr.*

¹³ Az adatbázisbeli kódok szerint tehát: *Masc/Fem* = *fem. pro masc.*, *masc. pro fem.* és *fem. pro communi.*; *Masc/Neutr* = *neutr. pro masc* és *masc. pro neutr.*; *Fem/Neutr* = *neutr. pro fem.* és *fem. pro neutr.*

Az alternatív kódok kezelésével kapcsolatban több lehetséges megoldás kínálkozik. Az eljárás minden esetben azzal kezdődik, hogy megvizsgáljuk, van-e összefüggés az alternatív kódok és a fő kódok között. Amennyiben van, ezt a változót mindenképpen meg kell vizsgálnunk és valahogyan figyelembe kell venni a vizsgálat során. Ezek után felmerül az a kérdés is, hogy van-e összefüggés az alternatív kód és egyes változók között, illetve hogy az alternatív kód különböző szintjei mentén van-e összefüggés a Provinciák és a fő kód között. Az első esetben két olyan változó, a területi eloszlás és az alternatív kódok, amelynek nem szabadna összefüggnie, összefügg egymással.¹⁴ A vizsgált adatok között nem találunk példát ilyen jellegű összefüggésre, azaz az egyes területeken megegyezik az alternatív kódok eloszlása, nincs olyan terület, ahol valamely alternatív kód jelentősen gyakoribb vagy ritkább lenne, mint a többin. A hiperkorrekt értelmezés esetén viszont előfordul olyan eset, ahol az alternatív kód egyes szintjein belül más-más a területenkénti eloszlása a fő kódoknak (erről ld. lentebb). Ez a tény arra utal, hogy a két változó (ti. a provincia és az alternatív kód) interakcióban állnak, az egyik (pl. az alternatív kód) hatással van arra, hogy a másik változó (pl. a provincia) mit határoz meg a genuszkeverések (az ún. függő változó) eloszlásáról. Ebben az esetben a modellben figyelembe kell venni ezt az interakciót.

Dolgozatunkban a genuszkeverések vizsgálatához és leírásához Agresti (2013:293-300) és Levshina (2015:277-288) nyomán egy multinomiális logisztikus modellt alkalmaztunk. A fő kódon (azaz a függő változón) belül a „Masc/Fem” típusú keverések jelentették az alapszintet, amihez a modellben a másik két keverési típus gyakoriságát hasonlítottuk. Ezeket az arányokat utána pedig a modell segítségével a vizsgált változók (a provincia, az alternatív kód és a keltezés) segítségével próbáltuk meg a legjobban leírni. A végeredmény egy olyan modell lett, ahol kontrollálni lehetett az alternatív kódok hatását, így megállapítható a korábbi eredmények¹⁵ ezek melletti érvényessége.

A vizsgált adatok

A dolgozatban vizsgált adatokat az Adatbázisból kaptuk meg, még hozzá úgy, hogy exportáltunk minden olyan adatot, aminek vagy a fő- vagy az alternatív kódja genuszkeverés volt (ld. 11. jegyzet) és ezeket után már az R, illetve az RStudio segítségével kezeltük az adatokat. A diagramok készítéséhez és a statisztikai vizsgálatokhoz is ezt a programot, illetve egyes segédkönyvtárakat használtunk.¹⁶ Az adatbázisból kinyert adatok fő kódjait, alternatív

¹⁴ Ebben az esetben például egy kevert modell alkalmazása segíthetne, ld. Winter–Grice (2021:1255-1257)

¹⁵ Khi-négyzet próbák segítségével Szlovicsák (2022b) tett több megfigyelést.

¹⁶ R 4.2.2 verzió, RStudio 2022.12.0+353 "Elsbeth Geranium" verzió, a felhasznált csomagok: *tidyverse*, *readxl*, *writexl*, *RColorBrewer*, *nnet*, *stargazer*, *broom*, *gofcat*.

kódjait, illetve az egyes Provinciákat is csoportosítottuk. Az így kapott területi egységek: *Africa, Gallia és Germania, Hispania, Dél-Italia, Észak-Italia, Illyricum, Róma*.¹⁷ Minden egyéb területet kizártunk a vizsgálat alól. Ez és a vizsgált időszak Kr.u. 1-7. századra történő korlátozása után végül 643 megfigyelésünk maradt. Egy kezdetleges próbálkozás után az Alternatív kódot végül négy szintre osztottuk, mivel csak így kaptunk megfelelő mennyiségű elemszámokat az egyes szinteken belül, illetve így mind az adatok időbeli, mind a területi eloszlása konzisztens lett. A négy végső szint: *Phonologia, Nominalia, Syntactica etc.* és *Nincs*.¹⁸ Ezek közül az utolsó magától értetődő, a maradék három pedig megfelel a különböző nyelvi szinteknek, amik magyarázhatják a genuszkeveréseket. Ezek közül a különösen a *phonologia* és a *nominalia* szintek esetén várhatunk nagy hatást, gyakran felmerül ugyanis az a lehetséges magyarázat a semlegesnem gyengülésére, hogy ez (legalábbis részben) a hangtani és az egyéb szótani kategóriák átalakulása miatt indult meg a vulgáris latinban, így például Väänänen (1981³:102-103) és Herman (2003:57) összefoglaló művében is találunk ehhez hasonló felvetést. Az így kapott alternatív kódok eloszlását mutatja be az 1. és a 2. Diagram.

Az első diagram provinciánként mutatja be az egyes alternatív kódok eloszlását, míg a második már a nem hiperkorrekt értelmezés alapján csoportosított főkédek eloszlását mutatja be az egyes alternatív kódokon belül. Ami az első diagrammal kapcsolatban kiemelendő, hogy ha provincia szerint csoportosítjuk az adatokat, nem fedezünk fel összefüggést eközött a két

¹⁷ Az egyes területek az alábbi provinciákat tartalmazzák: Afrika (Afrika proconsularis, Numidia, Mauretania Caesariensis, Mauretania Tingitana), Gallia és Germánia (Aquitania, Belgica, Lugdunensis, Gallia Narbonensis, Germania superior, Germania inferior), Hispánia (Baetica, Hispania citerior, Lusitania), Illyricum (Dacia, Dalmatia, Moesia Inferior, Moesia Superior, Noricum, Pannonia Inferior, Pannonia Superior, Raetia), Észak-Itália (Regiones VI–XI), Közép- és Dél-Itália (Regiones I–V, Róma nélkül).

¹⁸ A *Nominalia* és a *Syntactica etc.* kategóriák azért kerültek használatra, mert ezek maguk az Adatbázisban előforduló tág kódolási kategóriák. A *Syntactica etc.* szintbe bevettük az *Errores non grammaticae* típusú kódokat is, mivel a két kategória meglehetősen alacsony elemszámú volt, de arányaikban nem tértek el (tehát nem mutatott különbséget ezeken a kódokon belül a genuszkeverések eloszlása). A *Phonologia* kategóriában azokat a genuszkeveréseket vontam össze, amelyek az Adatbázisban vagy a *Vocalismus* vagy a *Consonantismus* kategóriába esnek, azaz ezek magán- és mássalhangzó keverések. Az összevonást itt is az alacsony elemszám, de hasonló viselkedés is indokolta. Továbbá az egyes kategóriák az alábbi hibákat tartalmazzák (csak azokat soroljuk fel, amik előfordultak az adatok között), az adatbázisban használt kódjaik szerint: *Phonologia* = *-s > ø, -s > ø elisa, -m > -ø, -ø > -m, -m > ø elisa, i > E, i > E, i: > E, i: > E, ae > I, e > I, ae > I, ú > O, u (+ voc) > ø, a > E, a/á > O, commutationes vocalium variae, c > Q/QV/CV, n (+ cons.) > ø*. *Syntactica etc.* = *ablativus absolutus pro participio coniuncto vel appositione, hypercorrectio, permixtio syntagmatum, varia ad congruentiam nomin. et adiect. pertinentia, onomastica (nomina grammaticae conspicua), Graecismus, sing. pro plur., plur. pro sing., litterae omissae, litterae perperam incisae, litterae superfluae, abbreviatio insolita*. *Nominalia* = *nom. pro acc., nom./abl. pro gen., nom./acc. pro abl., acc. pro nom., acc. pro gen., commutatio vel permixtio casuum aliorum, dat./abl. pro gen., gen. pro nom., decl. I pro II, decl. I pro III, decl. II pro I, decl. III pro I, decl. III pro II, commut. in formatione pronominum, praep. > casus sine praep., cett. ad usum pron. pertinentia, commut. in decl. pron. hic, commut. in decl. pron. ille, commut. in decl. pron. relat.* Illetve ide kerültek még azok a genuszkeverési kódok is, amik az alternatív kódok között szerepeltek. Egy ilyen furcsa helyzet úgy jöhetett létre, hogy egyes adatok két féle genuszkeverésként is értelmezhetők, erre példa: LLDB-28296: GRADA = *gradus*, főkédek az Adatbázisban: *neutr. pro masc.*, alternatív kódja pedig *fem. pro masc.* Az ilyen esetekben nem változtattuk meg a kódok sorrendjét, így megtartottuk az adatgyűjtő által választott sorrendet, még ha ennek nem is volt jelentősége.

változó között, azaz nincsen szükség olyan lineáris modell alkalmazására, ami kezelni tudja ezeket az összefüggő változókat (ahol a *független* változók függenek össze). A ránézésre sejtett összefüggés hiányát egy Fisher-féle egzakt próba sem cáfolta meg, a kapott p -érték az 1. Diagram adatai esetén 0,426, azaz közel sem szignifikáns.¹⁹ A 2. Diagram adatai esetén

¹⁹ Jelen dolgozatban a kapott p -értékeket minden esetben akkor tekintjük szignifikánsnak, ha 0,05-nél kisebbek.

1. Táblázat

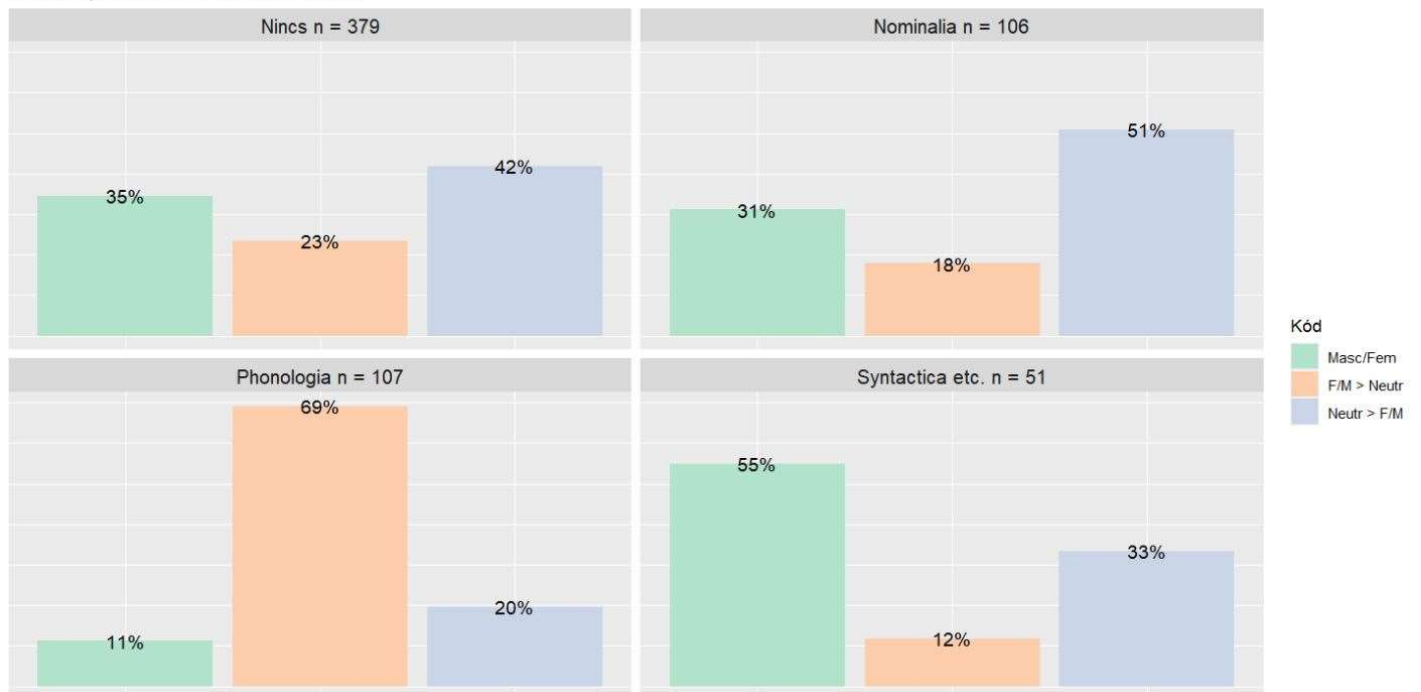
1. Diagram

A különböző alternatív kódok Provinciánkénti eloszlása



2. Diagram

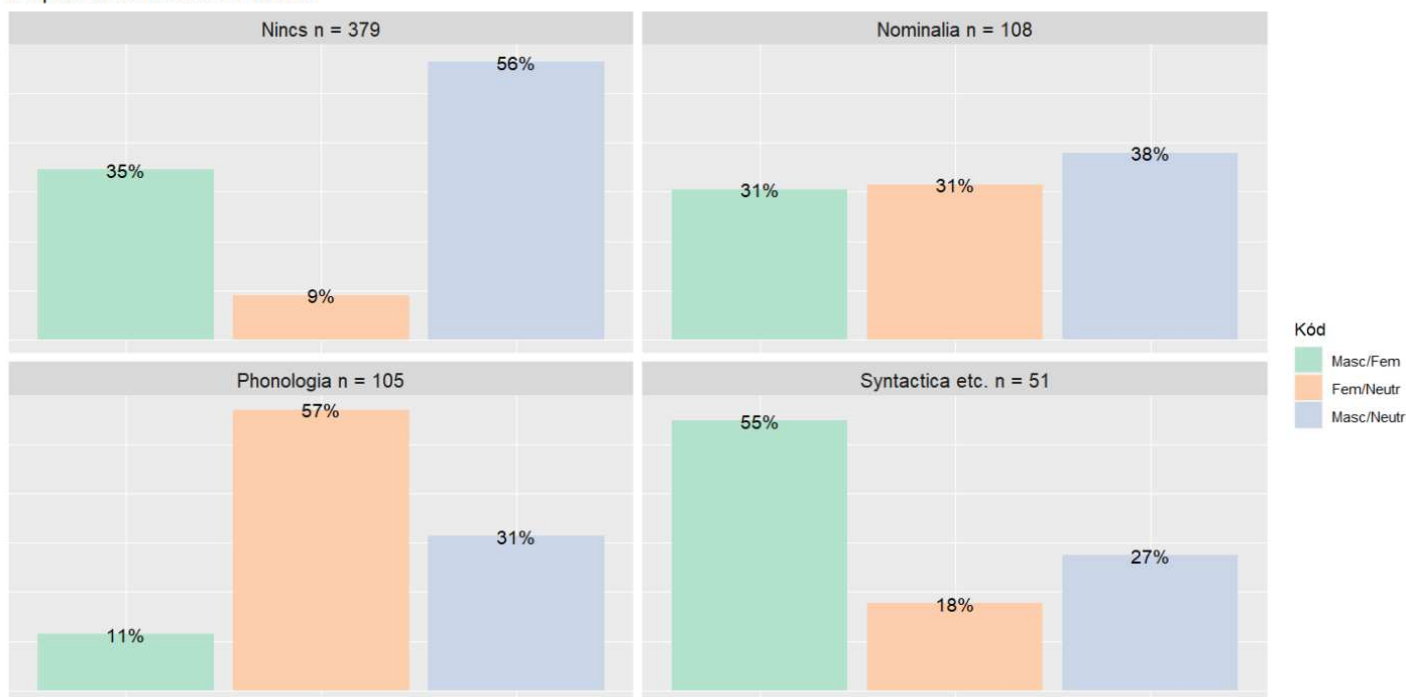
A különböző alternatív kódok melletti végleges eloszlás, a nem hiperkorrekt értelmezés mellett



természetesen nem számítunk függetlenségre, hiszen épp azért van szükség az alternatív kód kezelésére, mert van közöttük és a főkédek eloszlása között összefüggés, ami látszik is a diagramon. Míg ugyanis például az alternatív kóddal nem rendelkező, azaz a diagramon „Nincs” kóddal jelölt adatok eloszlása nagyjából egyenletes, addig a *Phonologia* szintet uralják a semlegesnem esetleges stabilitására utaló keverések, a *Nominalia* szinten belül pedig éppen a semlegesnem eltűnésére utaló keverésekből van a legtöbb, bár nem olyan arányban, mint a hangtani keverések között. Kiemelendő még a mondattani szintű és egyéb keverések kategóriája, amiben egyrészt a hím- és nőnem közötti keverések aránya a legmagasabb, amely genuszkeverés egyébként elsődlegesen összehasonlításként szerepel a vizsgált adatok között, másrészt pedig itt a legalacsonyabb az adatok száma, emiatt ez a kategória még az összevonások után is kissé problémás. A 2. Diagram párjaként a 3. Diagram a hiperkorrekt módon értelmezett adatok kapcsolatát mutatja be az alternatív kódokkal.

3. Diagram

A különböző alternatív kódok melletti végleges eloszlás, a hiperkorrekt értelmezés mellett



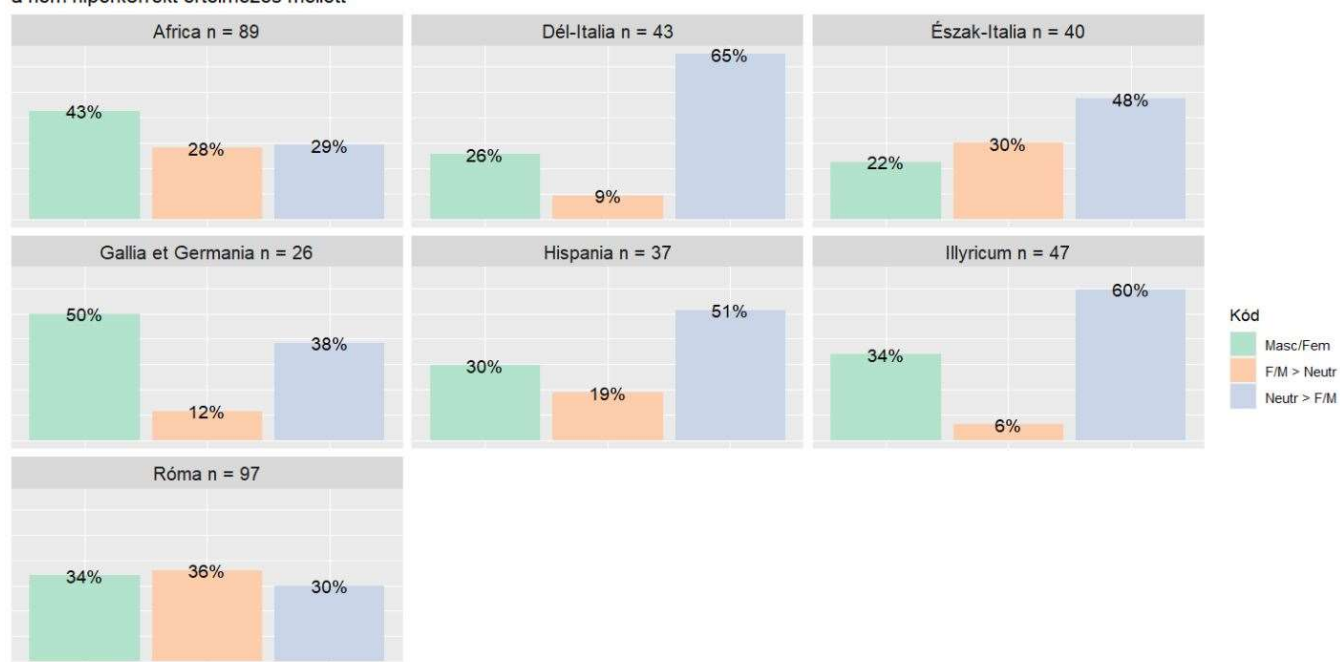
Ebben az esetben az alternatív kóddal nem rendelkező adatok eloszlása közel sem olyan egyenletes, mint az a 2. Diagramon volt, hiszen itt a középso, nő- és semlegesnem közötti keveréseket tartalmazó kategória aránya meglehetősen kicsi. Jelen diagram esetén ezeknek a nőnemet érintő keveréseknek a jelentős része a fonológiai alternatív kóddal rendelkező adatok között található, amiből azt sejthetjük, hogy a nő- és semlegesnem közötti keverések

létrejöttében igen nagy szerepük volt egyéb nyelvtani átalakulásoknak és különösen a hangtani keveréseknek. Feltűnő még, hogy a *Nominalia* alternatív kódú adatok eloszlása meglehetősen egyenletes, azaz a hiperkorrekt értelmezés szerinti adatok eloszlására nem voltak különösebb hatással ezek a keverések, míg a nem hiperkorrekt adatok esetén felfedezhető volt a különbség ezek eloszlásában (ld. 2. Diagram). Az elemszámok természetesen nem térnek el a korábban látottaktól és emiatt a hím- és nőnem közötti keverések aránya se változott, így a *Syntactica etc.* szinten belül továbbra is ez a legnagyobb arányú keverés, emiatt is alkalmas viszonyítási alapnak, mivel amikor megváltoztatjuk a semlegesnem keverések csoportosítását, ezek a keverések nem változnak. Ezek a táblázatok tehát tisztáztak egy lehetséges problémaforrást az Alternatív kód változóval kapcsolatban, még hozzá azt, hogy nem függ össze a Provincia változóval.

2. Táblázat

4. Diagram

A főkédek területenkénti eloszlása, az alternatív értelmezés nélküli adatokon belül a nem hiperkorrekt értelmezés mellett



5. Diagram

A főkédek területenkénti eloszlása, a phonológiai keveréseken belül a hiperkorrekt értelmezés mellett



Az viszont még elképzelhető, hogy interakció fennáll eközött a két változó között, azaz előfordulhat, hogy az alternatív kód változó valamely szintjén belül eltérés figyelhető meg a főkédek területenkénti eloszlásában, azaz az alternatív kód főkédre gyakorolt esetleges hatása

területenként változik.²⁰ Az adatok megvizsgálása után (ehhez Fisher-féle egzakt próbákat használtunk),²¹ kiderült, hogy valóban fellép interakció az alternatív kódok és a provinciák között, de egyetlen esetben se lépett fel minden szinten. Az érintett Alternatív kódok közül egyet-egyet szemléltetünk a 4. és az 5. Diagramok segítségével, amik bár nem fednek le minden alternatív kódot, de így is szemléltetik az összefüggés jellegét, és az egyes problémákat. Ezen túl könnyen átláthatóak az egyes területi különbségek, miközben mindkét lehetséges értelmezés hatását bemutatják. A 4. diagram azt mutatja, hogy a hiperkorrekt értelmezés esetén az interakció az alternatív kódok és a provinciák között, a *Nincs* szinten lépett fel, azaz az alternatív értelmezés nélküli fő kódok eloszlása összefüggött a provinciákkal. A Fisher-próbával kapott *p*-érték kisebb, mint 0,001.²² Ez a megfigyelés több szempontból is figyelemre méltó. Egyrészt mivel azt mutatja, hogy az alternatív értelmezéssel nem rendelkező fő kódok is függenek a területi eloszlástól, azt a sejtést igazolja, hogy a genusrendszer átalakulása azokban az esetekben is különböző volt területenként, amikor semmilyen alternatív módon nem lehetett magyarázni a nemek közötti bizonytalanságot. Feltűnő viszont, hogy az így szűrt adatok között a semlegesnem megmaradására utaló keverések közel sem olyan gyakoriak Rómában vagy Afrikában, mint azt például Szlovicsák (2022b:261 és 266) tanulmányában láttuk.²³ Ez megfigyelésileg azzal magyarázható, hogy ezeket a keveréseket jelentősen befolyásolhatta a *phonologiai* keverések jelenléte, ahogyan az a 2. Diagramon is látható, és amit a későbbiekben igazolunk is. Másrészt kiemelendő még, hogy ezen a szintjén az alternatív kódnak nem találtunk szignifikáns különbséget a hiperkorrekt adatokon belül, azaz nem zárható ki, hogy ezen értelmezés mellett nem volt jelentős az összefüggés a területek és a fő kód eloszlása között, ha az adatnak nem volt alternatív értelmezése. Más szóval, amennyiben azt vizsgáljuk, hogy melyik genuszal keveredik jobban a semlegesnem, azaz melyik nembe kerülnek a semlegesnemű szavak, akkor a területi különbségek feltehetően az alternatív kódokból fakadnak.

²⁰ Ld. Agresti (2013:53)

²¹ Ezzel el lehetett dönteni, hogy az alternatív kód egyes szintjei mentén létrehozott táblázatok változói, amikből a 4-5. diagramok egyet-egyet szemléltetnek, függetlenek-e egymástól vagy sem. Tehát ha interakció áll fent, akkor azt figyeljük meg, hogy az alternatív kód megváltoztatására megváltozik a fő kódok területenkénti eloszlása. Illetve szimmetrikusan, provinciánként változik, hogy a fő kód eloszlására milyen hatást gyakorolnak az alternatív kódok. Ezt tudjuk azzal ellenőrizni, hogy az alternatív kód leszűkítése mellett megvizsgáljuk, hogy van-e összefüggés a területi eloszlás és a fő kódok eloszlása között.

²² A maradék három szint esetén nem lépett fel interakció, a Fisher-féle egzakt teszttel kapott *p*-értékek: *Synctactica* etc: 0,151, *Phonologia*: 0,107, *Nominalia*: 0,154.

²³ Ahol például, a későbbi korszakban, Rómában több mint négyszer annyi *F/M > Neutr* típusú keverés volt, mint *Neutr. > F/M*.

Az 5. Diagram esetében megegyező megfigyeléseket tehetünk azzal, amit az előző diagram kapcsán írtunk. Az összefüggés itt is fennállt, hasonlóan Fisher-féle egzakt teszttel a p -érték 0,0015 volt. Kijelenthető tehát, hogy ezen a szinten, a hiperkorrekt módon értelmezett adatok esetében fellépett összefüggés, továbbá e mellett az értelmezés mellett a *Nominalia* szinten is fellépett interakció, a Fisher-féle egzakt teszttel kapott p -érték: 0,00015. A maradék két szinten itt sem lépett fel interakció.²⁴ Azaz a hiperkorrekt értelmezés mellett a területi különbségek forrása nem maguknak a genuszkeveréseknek a tulajdonsága, hanem összefügg esetleges alternatív értelmezésekkel, a nyelvi változás más szintjei befolyásolják. Ez jelentősen eltér attól, amit a nem hiperkorrekt esetben figyeltünk meg, a 4. Diagramon, hiszen ott a területi különbségek éppen az alternatív módon nem értelmezhető, azaz *tisztán* genuszkeveréses adatokból származtak. Ez persze egyáltalán nem zárja ki, hogy történetileg ott is meghatározták a nyelvi változás más szintjei ezeknek a keveréseknek az alakulását, csak ott éppen már ezektől függetlenül is kivehetőek a területi különbségek.

A regressziós modell

Az adatok leírására jelen dolgozatban, mindkét értelmezés mellett egy multinomiális logit²⁵ modellt alkalmaztunk, méghozzá a függő változón belül egy alapszintet használva.²⁶ A modellbe két kategorikus és egy folytonos változót vettünk be, ezek rendre a Provincia, Alternatív kód és Keltezés voltak. Az így kapott modellt összevetettük azzal a modellel, ahol csak az Alternatív kód magyarázta a fő kódokat, azaz azt tudtuk megvizsgálni, hogy a bővített modell, ahol már a területi és időbeli különbségeket is figyelembe vesszük, jobb-e annál, mintha csak az alternatív kód segítségével magyaráznánk a fő kód eloszlásának különbségeit. Ezzel el tudtuk dönteni, hogy az adatokban megmutatkozó különbségek és jellemzők valóban a területi és időbeli különbségekből, vagy csupán más nyelvtani bizonytalanságokból fakadnak. Azaz ellenőrizni tudtuk Szlovicsák (2022b:266) eredményeit.

Mivel a korábbi eredmények azt mutatták, hogy az egyes területek között mutatkozott eltérés abban is, hogy ott hogyan alakult a semlegesnem az időben, azaz interakció volt felfedezhető az idő- és térbeli eloszlás között, emiatt a modellbe is bevettük ezt a felfedezett interakciót. A másik interakciót viszont a két említett változó, a Provinciák és az Alternatív

²⁴ A Fisher-féle egzakt teszttel kapott p -értékek: *Synctactica etc*: 0,211, *Nincs*: 0,326.

²⁵ Azaz a függő változó egyes szintjeinek alapszinthez viszonyított bekövetkezési esélyének a természetes logaritmusát írják le a független változók.

²⁶ Szemben azzal a lehetséges modellel, ahol minden egyes szint valószínűségét vetik össze a többi szinttel, ld. Levshina (2015:283-288).

kódok között több okból kifolyólag sem vettük be a végső modellbe. Egyrészt mindkét esetben igaz volt, hogy az alternatív kódoknak csak korlátozott számú szintjén lépett fel az interakció, azaz, nem volt teljesen kiterjedt, másrészt az első, azaz a nem hiperkorrekt esetben csak az alternatív módon nem értelmezhető adatokat érintette ez az interakció és a modellben az alternatív kódokon belül ez a szint a viszonyítási alap, ezért az ezen a szinten jelentkező különbségek meg fognak jelenni a területek között, az egy szinten jelentkező interakció vizsgálata tehát nem indokolt. Mindkét esetre igaz továbbá, hogy amennyiben felállítjuk az interakciót, akkor ott az egyes szintek között nagyon sok alacsony elemszám fog megjelenni, ez jól látható az 5. diagramon is, ami könnyen okozhat gondot a modellben, ahogyan azt Agresti (2013:229) is bemutatja. Ennek fényében tehát a végleges modellbe három változó került be, illetve bekerült a Provincia és a Keltezés változók interakciója is. A nem hiperkorrekt értelmezés mellett kapott eredményeket a 3. Táblázat mutatja be.²⁷

3. Táblázat

	<i>Dependent variable:</i>			
	F/M > Neutr (1)	Neutr > F/M (2)	F/M > Neutr (3)	Neutr > F/M (4)
Konstans	-0,387*** (0,137)	0,194 (0,118)	-1,796*** (0,500)	-0,218 (0,421)
Dél-Italia konstans			1,981** (0,838)	0,822 (0,712)
Észak-Italia konstans			2,894*** (1,016)	0,136 (1,017)
Gallia et Germania konstans			1,323 (0,867)	-0,752 (0,795)
Hispania konstans			0,310 (0,825)	-0,916 (0,720)
Illyricum konstans			0,358 (1,051)	-0,184 (0,756)
Róma konstans			0,652 (0,711)	1,259** (0,637)
Együttható			0,004*** (0,001)	-0,0004 (0,001)
Phonologia	2,206*** (0,340)	0,366 (0,381)	2,349*** (0,354)	0,361 (0,394)
Nominalia	-0,166 (0,319)	0,299 (0,250)	-0,070 (0,332)	0,431 (0,265)
Syntactia etc.	-1,154** (0,470)	-0,693** (0,329)	-1,049** (0,488)	-0,732** (0,356)
Dél-Italia együttható			-0,008** (0,004)	0,001 (0,003)
Észak-Italia együttható			-0,006** (0,003)	0,003 (0,003)
Gallia et Germania együttható			-0,005* (0,003)	0,003 (0,002)
Hispania együttható			-0,0001 (0,002)	0,005** (0,002)
Illyricum együttható			-0,003 (0,004)	0,005 (0,003)
Róma együttható			-0,0005 (0,002)	-0,005** (0,002)
Akaike Inf. Crit.	1 315,731	1 315,731	1 258,012	1 258,012
<i>Note:</i>			*p<0,1; **p<0,05; ***p<0,01	

²⁷ A multinomiális modell együtthatóit az *nnet* csomagból származó *multinom* függvény segítségével kaptuk meg. A táblázatot pedig a *stargazer* csomag segítségével állítottuk elő.

Jelen táblázatban az első két oszlopban az első modell, azaz a csak az alternatív kódokat tartalmazó modell együttthatói szerepelnek, míg a második két oszlopban a bővített, azaz a keltezés és Provinciákat is figyelembe vevő modell együttthatói kerültek. A négy oszlop közül az első és a harmadik azt mutatja be, hogy az egyes változók szintjei mennyiben növelik vagy csökkentik annak az esélyét, hogy egy adott genuszkeverési hiba ne hím- és nőnem közötti, hanem a semlegesnem stabilitására utaló hiba legyen. Ehhez hasonlóan a 2. és a 4. oszlop arra utal, hogy ha az adott sorban szereplő feltétel teljesül, akkor mennyivel fog nőni vagy csökkenni azon esély természetes logaritmus, hogy egy adott esetben a genuszkeverési hiba hím- és nőnem közötti keverés helyett a semlegesnem eltűnésére utaló hiba legyen. Fontos tehát figyelni arra, hogy itt a bekövetkezési esélyek logaritmusai szerepelnek, azaz amennyiben arra vagyunk kíváncsiak, hogy valójában mennyivel lesz valószínűbb az egyik esemény a másikhoz képest egy változó megváltoztatásának hatására, akkor még a táblázatban szereplő kitevőre kell emelnünk az e számot, ahol e a természetes logaritmus alapja, az Euler-féle szám. A területek között azért nem szerepel Africa provincia, mert a többi területet ehhez viszonyítjuk, így az egyes területek esetén látott értékek azt jelentik, hogy ennyiben tér el az ott megfigyelt érték az Africa területén megfigyelt értékhez képest. Ehhez hasonlóan az alternatív kódok között a három szint együttthatói azt mutatják, hogy mennyiben térnek el az így is értelmezhető adatok azoktól, amik nem rendelkeznek alternatív értelmezéssel.

Továbbá, bár nem teljesen pontos, de a megértést könnyebbé teszi, ezért a területek között azok, amik után szerepel a „konstans” kifejezés, azok a területek általános különbségei Africa provinciához képest, az időbeli változástól függetlenül. Azok a provincianevek pedig, amik után az „együttható” kifejezés szerepel az adott provincia és az időbeli változás összefüggését mutatják, méghozzá, hogy mennyire tér el, az „Együttható” névvel jelölt, Africa esetén megfigyelhető időbeli változástól.²⁸ Az „Együttható” az Africában az évenkénti változást mutatja, azaz, hogy egy év alatt megközelítőleg mennyit változik a korábbi értékhez képest az adott keverés viszonyított bekövetkezési esélyének természetes logaritmus. Így például a „Dél-Italia konstans” sorban az első érték, az 1,981 azt jelenti, hogy itt az általánoshoz (-1,796) képest ennyivel gyakoribbak kezdetben a semlegesnem megmaradására utaló keverések a hím- és nőnem közötti keveréseknél. Ugyanebben az oszlopban, a „Dél-Italia együtttható” sorban látható érték pedig azt mutatja, hogy az „Együttható” névvel jelölt africai évenkénti növekedéshez képest (0,004), itt 0,008-cal kevesebb az esély éves növekedése, azaz

²⁸ Valójában a kapott értékek mindkét esetben együttthatók, pusztán az „együttható” névvel jelöltek függenek az idő változásától, míg a „konstans” névvel jelöltek nem.

egy csökkenés figyelhető meg ($0,004 - 0,008 = -0,004$). Emellett az esély kezdeti értékét, azaz azt az értéket, amit a „konstans” szóval megjelölt tagok mérnek, még módosíthatja, növelheti az egyes alternatív kódok fellépése. A „kezdeti érték” elnevezés arra utal, hogy a konstans tagok azt becsülik meg, hogy a vizsgált főkódok valószínűségei hogyan viszonyultak egymáshoz az 1. század elején. Az egyes együtthatók természetesen csak a minta alapján végzett becslései a valós együtthatóknak. A becslés bizonytalanságát a zárójelben szereplő sztenderd hiba méri, illetve amennyiben a kapott becslés szignifikánsan eltér a nullától, akkor azt a táblázatban két, illetve három csillag jelöli.²⁹ Így tehát például az első sor és első oszlop találkozásánál álló érték, a $-0,367$ azt jelenti, hogy az alternatív értelmezéssel nem rendelkező adatok között szignifikánsabb alacsonyabb volt a semlegesnem stabilitására utaló keverések aránya, a hím- és nőnem közötti keverések arányánál. Az első sorban és második oszlopban álló érték megbecsüli, hogy gyakoribbak lehettek a semlegesnem eltűnésére utaló keverések a hím- és nőnem közötti keveréseknél, de ez az eltérés nem szignifikáns, így nem tudjuk azt valószínűsíteni, hogy ez valóban így volt.

Az eredmények kapcsán először is érdemes leszögezni, hogy a 3. Táblázat alján látott érték, az Akaike Információs Kritérium (röviden: AIC) azt segít eldönteni, hogy melyik modell írja le jobban a megfigyelt adatokat, amint azt például Levshina (2015:149-150) bemutatja. Két érték közül mindig az alacsonyabb utal a jobb modellre, azaz esetünkben a két modell közül a bővebb írja le jobban a megfigyelt adatokat, így tehát indokolt a vizsgálatba bevinni a területi és időbeli különbségeket is, mivel ezek érdemben segítenek a függő változók leírásában. Azaz Szlovicsák (2022b:266) korábbi eredményeinek megfelelően, kimutatható az összefüggés a genuszkeverések eloszlása, és a térbeli különbségek és az idő változása között, az alternatív kódok esetleges zavaró hatása mellett is. Ez a modell viszont nem csak erre alkalmas, azzal ugyanis, hogy megbecsüljük az idő hatását a változásra és az egyes területek különbségeit, pontosabban le tudjuk írni az ezek közötti különbségeket. Illetve azt is jól látjuk, hogy mely változók okoztak jelentős eltérést, és melyek hatása elhanyagolható, emiatt pedig az értelmezésnél csak a szignifikáns együtthatókra térünk ki, hiszen ezekről tudjuk kellő bizonyossággal felfedezni az összefüggést.³⁰ Továbbá jelen esetben arra, hogy mennyire írja le a modell a megfigyelt adatokat az Ugba és Gertheiss-féle pszeudo- R^2 mértéket használtuk, az

²⁹ Jelen dolgozatban az egy csillaggal jelölt, azaz 0,1 és 0,05 közötti p -értékeket még nem tekintjük szignifikánsnak.

³⁰ Sokszor persze az is releváns, hogy mi nem szignifikáns, viszont ekkor a becsült értékeket nem érdemes meglehetősen komolyan venni.

így kapott érték 0,332 volt.³¹ Amint ezt a szerzők, Ugba és Gertheiss bemutatták (2022:15) ez a mérték, más pszeudo- R^2 értékeknél jobban tart a valódi értékhez, így az interpretációja is hasonló. Ennek fényében ez a modell a teljes szórásnak körülbelül a 33%-át magyarázza meg, ami nem feltétlenül rossz, de nem kizárt, hogy elérhető egy ennél jobb modell is.

A 3. Táblázat második modelljének együtthatói közül elsőként a hármas számú oszlopra fókuszálva megfigyelhető, hogy itt több szignifikáns eredmény szerepel, mint az utolsó oszlopban, azaz a semlegesnem esetleges stabilitására utaló hibák sokkal változatosabbak voltak a különböző változók mentén, mint a semlegesnem eltűnésére utaló keverések. Ezek között jól látható, hogy a „Konstans” névvel jelölt sorban³² a semlegesnem stabilitására utaló keverésekből jelentősen kevesebb van, mint a hím- és nőnem közötti keverésekből, amennyiben a kelezést nem vesszük figyelembe. Ez a kiinduló helyzet, amihez képest az idő folyása okozhat változást. Az „Együttható” sor esetén látható is, hogy ezek a keverések az idő előrehaladtával folyamatosan nőttek Africa provinciában, itt már az alternatív kódoktól függetlenül. Az itt, az „Együttható” sorban látott érték alacsonynak tűnhet, de ha figyelembe vesszük, hogy ez az évenkénti várható változás, akkor világossá válik, hogy ez a tag a 6-7. századra meg fogja haladni a kezdeti értéket, azaz azt várjuk, hogy Africa területén a semlegesnem stabilitására utaló keverések arányban fölülmúlják a hím- és nőnem közötti keveréseket. Minden más ebben (3.) az oszlopban látott szignifikáns együttható az ettől a tendenciától és kezdeti értéktől való valamiféle eltérésre utal. Az eltérések közül kiemelkedő, hogy mind Észak-, mind Dél-Itália esetén egy hasonló, szignifikáns eltérés figyelhető meg. Mindkét területen gyakoribbak voltak kezdetben a vizsgált semlegesnem-keverések, viszont a későbbi korszakra ezek jelentősen lecsökkentek. Eltérés viszont, hogy míg Dél-Itáliában a kezdeti állapotban nagyjából megegyezett a semlegesnem stabilitására utaló keverések és a hím- és nőnem közötti keverések aránya,³³ addig Észak-Itáliában meglehetősen gyakoribbak voltak a semlegesnem-keverések és ezek arányának csökkenése se volt olyan gyors, mint Dél-Itáliában. Semelyik más terület esetén sem látunk szignifikáns eltérést az alapszinthez, Africához képest. A többi területi egység közül egyik esetén sem kizárható az, hogy az africai trendhez hasonlóan alakult ennek a semlegesnem keverésnek az aránya a hím- és nőnem keveréseivel képest. Azaz minden más területen elképzelhető, hogy a két keverési szint

³¹ Ezt az értéket az Ugba által fejlesztett *gofcat* csomag segítségével kaptuk meg, amelynek két kimenete közül, ez a gyökösen büntetett érték, ezzel kapcsolatban ld. Ugba–Gertheiss (2022:4-5).

³² Ezek alternatív értelmezéssel nem rendelkező (Alternatív kód = „Nincs”), africai adatok

³³ Ha a Dél-Itáliához tartozó értéket hozzáadjuk a „Konstans” sorban látható értékhez, akkor egy nulla közeli számot kapunk, ami éppen azt jelenti, hogy nem volt jelentős eltérés a *F/M > Neutr* típusú és a *Masc/Fem* típusú keverések valószínűségei között.

egymáshoz viszonyított esélye az africában látott módon alacsonyan indult és az idő előrehaladtával folyamatosan nőtt.

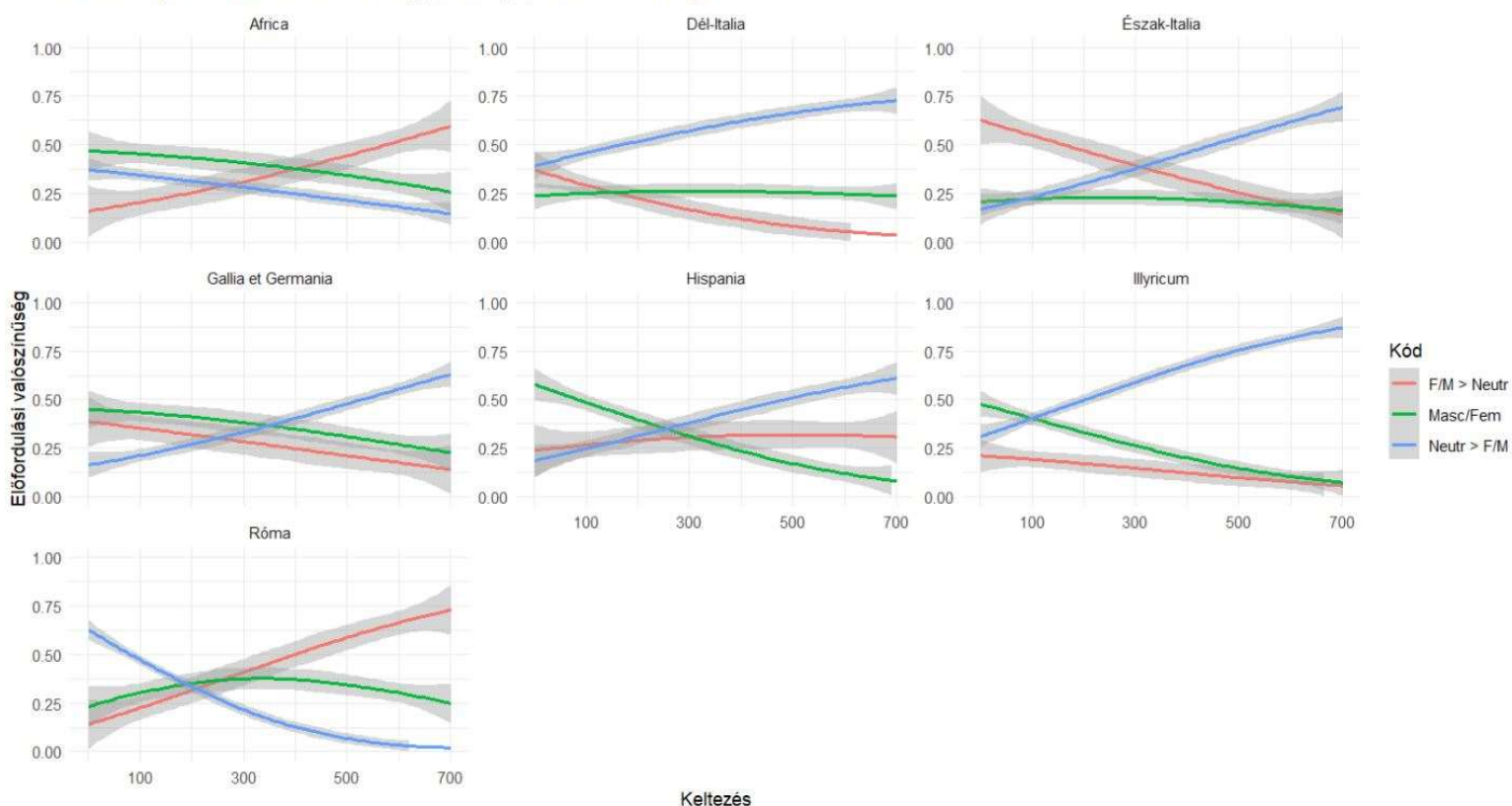
Az alternatív kódok hatása kizárhatatlan. Ezeken belül a *Nominalia* szintnek meglepő módon nem volt jelentős hatása, a másik két szintnek viszont volt.³⁴ A *Phonologia* kategóriába tartozó alternatív értelmezéssel rendelkező genuszkeverések a modell alapján igen nagy eséllyel lesznek a semlegesnem megmaradására utaló keverések, míg, ha egy adat a *Syntactica etc.* kategóriából kerül ki, akkor még az alapszintnél is valószínűtlenebb, hogy a semlegesnem stabilitására utal. Ezek a különbségek már a 2. Diagramon is látszottak, viszont most a mértékükre is kaptunk egy becslést. Az látszott tehát, hogy míg hangtani keverések igen nagy eséllyel jelezték előre a semlegesnem stabilitására utaló keveréseket, azaz a hangtani keverések megléte mellett igen valószínű volt, hogy stabil semlegesnem fordult elő. Ezek alapján feltehető, hogy a különböző hangtani keverések nemhogy a semlegesnem leépülését segítették, hanem éppen ellenkezőleg, lehetséges, hogy lehetővé tették a semlegesnem további stabilitását.

Az utolsó oszlop együtthatóira térve látható, hogy itt az alapszint se a konstans, se az időbeli növekedés esetén nem tért el jelentősen egytől, azaz a semlegesnem eltűnésére utaló keverések Africa területén nem mutattak jelentős eltérést, se kezdetben, se az idő előrehaladtával a hím- és nőnem közötti keverésektől. Róma esetén viszont egyértelmű eltérések figyelhetők meg. Itt még az első század elején viszonylag gyakoriak lehettek a semlegesnem bizonytalanságára utaló hibák, de ezek aránya fokozatosan csökkent, amint azt a „Róma együttható” értéke mutatja. Ezek alapján tehát Rómában feltételezhető, hogy míg a semlegesnem stabilitására utaló hibák Africához hasonlóan nőttek, addig itt egyértelműen csökkent a semlegesnem bizonytalanságára utaló hibák esélye. Így alakulhat ki egy olyan kép Szlovicsák (2022b:266) tanulmányában Rómáról, hogy itt a semlegesnem még a 7. században is viszonylagos stabilitást mutatott. Hispánia esetén látható még egy eltérés az átlagos adatoktól, itt azt látjuk, hogy ezek a semlegesnemkeverések az idő előrehaladtával egyre nagyobb eséllyel következtek be a hím- és nőnem közötti keverésekhez képest. Továbbá a *Syntactica etc.* alternatív kóddal rendelkező adatok esetén itt is arra számíthatunk, hogy nagyobb eséllyel figyelünk meg hím- és nőnem közötti keverést, mint semlegesnemkeverést. Ezeket az eredményeket a 6. Diagram segít vizualizálni.

³⁴ Fentebb például Herman (2005:57) alapján felvetettük a lehetőséget, hogy ez az alternatív kód is jelentős hatással fog bírni. A sejtés viszont csak a *phonologiai* keverésekre igazolódott be.

6. Diagram

A nem hiperkorrekt értelmezés alapján előrejelzett valószínűségek



Ezen a Diagramon a modell alapján előre jelzett valószínűségeket ábrázoltuk. Azaz, azt mutatja meg, hogy a kapott együtthatók alapján milyen trendeket várhatunk az egyes provinciákon belül a különböző genuszkeverések valószínűségének időbeli alakulásában. A görbék körüli sáv a különböző alternatív kódokból fakadó eltéréseket segít bemutatni. Ezek között jól látható, hogy Africa és Róma estén számíthatunk a semlegesnem esetleges stabilitására, a többi területen inkább az eltűnés valószínű, különösen Itália és Hispánia esetén, ahol egyértelműen szignifikáns eredményeit figyeltük meg ezeknek a különbségeknek. A nem hiperkorrekt értelmezés eredményeihez hasonlóan, a hiperkorrekt értelmezés eredményeit a 4. Táblázat mutatja be, azon belül pedig a 7. Diagram tartalmazza becslések alapján sejtett változást. Az itteni eredményekhez hasonlóan ott is, csak a szignifikáns eredmények esetén állítható biztosan az alakulás ábrázolásának megbízhatósága.

A hiperkorrekt adatok esetén a kapott AIC érték azt mutatja, hogy itt is érdemes bevenni a modellbe az extra változókat, azaz a provinciákat és az időbeli változást. Emellett a modell szórását leíró tulajdonságát mérő Ugba és Gertheiss-féle pszeudo- R^2 érték itt 0,323, azaz valamivel kisebb, mint a nem hiperkorrekt értelmezés esetén kapott érték, de nem jelentősen. Ha a hiperkorrekt értelmezés melletti becsült együtthatókat vesszük szemügyre, akkor rögtön

feltűnő lehet, hogy egy olyan eset van csak, ahol szignifikáns időbeli változást figyelhetünk meg, ez pedig Dél-Itália, ahol a nő- és semlegesnem közötti keverések esélye jelentősen csökken az időben a hím- és nőnem közötti keverésekhez képest. Ezen túl viszont nem az idő magyarázza meg itt a különbségeket, hanem a területek közötti különbségek és az alternatív kódok hatása. Látható ugyanis, hogy a nő- és semlegesnem közötti keverések esetén ezek esélye a hím- és nőnem közötti keverésekhez képest igen alacsony a provinciák alapszintje, azaz Africa esetén, ami különösen meglepő lehet Szlovicsák (2022a:53) eredményei alapján, aki Africa területén ezeknek jelentős előfordulását figyelte meg. Ezen eredmények alapján viszont kijelenthető, hogy az ott megfigyelt területek közötti hasonlóság okát a különböző alternatív értelmezések jelentették, például Afrika és Hispania között a kapcsolat magyarázata, ahogyan az Szlovicsák (2022a:53) tanulmányában is felmerül, a két terület esetrendszerének Adamik (2019:29-30) által megfigyelt hasonlósága lehetett. Jelen eredmények fényében nem tudjuk egyértelműen kijelenteni, hogy az említett tanulmányban megfigyelt hasonlóság valóban kizárólag az Alternatív kódok hasonlóságának tudható be, de ez feltételezhető. Az ugyanis kijelenthető, hogy nem maguknak a területi különbségeknek a hatása. A további különbségekre kitérve Dél-Itália esetén nem csak a változás esetén figyelhető meg egy jelentős különbség, de az is feltűnő, hogy itt az 1. század elején még gyakoriak lehettek a nő- és semlegesnem közötti keverések, viszont ezek aránya jelentősen lecsökkent a többi genuszkeveréshez képest az idő teltével. A maradék három szignifikáns érték mind az alternatív kódok eltéréséből származik, míg a *Phonologia* és a *Nominalia* szintek a nő- és semlegesnem közötti keverések esélyét növelik a hím- és nőnem közötti keverésekhez képest, addig a *Syntactica etc.* alternatív értelmezéssel rendelkező adatok között jóval ritkábbak voltak a hím- és semlegesnem közötti keverések, mint a hím- és nőnem közötti keverések. Ezeket az eredményeket megfigyelhetjük a 7. Diagramon, de természetesen az időbeli változás egyáltalán nem egyértelmű és csak a kapott értékeket segít vizualizálni.

4. Táblázat

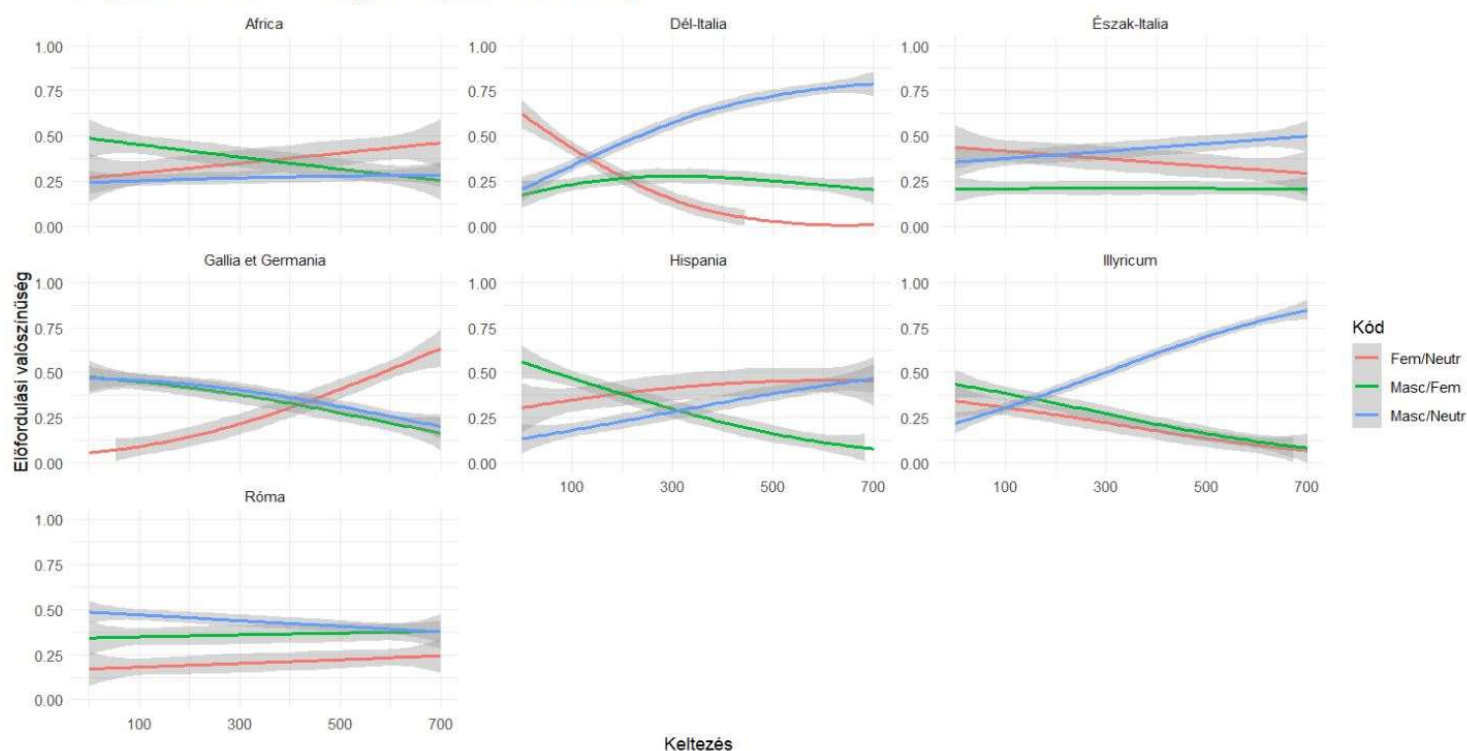
	Dependent variable:			
	Fem/Neutr	Masc/Neutr	Fem/Neutr	Masc/Neutr
	(1)	(2)	(3)	(4)
Konstans	−1,349*** (0,192)	0,491*** (0,111)	−2,110*** (0,531)	−0,430 (0,406)
Dél-Italia konstans			2,344*** (0,909)	0,999 (0,717)
Észak-Italia konstans			1,655 (1,136)	1,345 (0,916)
Gallia et Germania konstans			−1,779 (1,117)	0,637 (0,713)
Hispania konstans			0,012 (0,820)	−0,728 (0,740)
Illyricum konstans			0,466 (1,071)	0,055 (0,759)
Róma konstans			−0,126 (0,804)	1,078* (0,576)
Együttható			0,002 (0,002)	0,001 (0,001)
Phonologia	2,958*** (0,370)	0,521 (0,355)	3,239*** (0,391)	0,553 (0,364)
Nominalia	1,379*** (0,311)	−0,274 (0,259)	1,677*** (0,330)	−0,227 (0,269)
Syntactia etc.	0,214 (0,429)	−1,184*** (0,346)	0,251 (0,454)	−1,150*** (0,364)
Dél-Italia együttható			−0,010** (0,004)	0,0003 (0,003)
Észak-Italia együttható			−0,003 (0,003)	−0,001 (0,003)
Gallia et Germania együttható			0,004 (0,003)	−0,001 (0,002)
Hispania együttható			0,002 (0,002)	0,004* (0,002)
Illyricum együttható			−0,002 (0,004)	0,003 (0,003)
Róma együttható			−0,002 (0,002)	−0,002 (0,002)
Akaike Inf. Crit.	1 235,788	1 235,788	1 218,170	1 218,170

Note:

*p<0,1; **p<0,05; ***p<0,01

7. Diagram

A hiperkorrekt értelmezés alapján előrejelzett valószínűségek

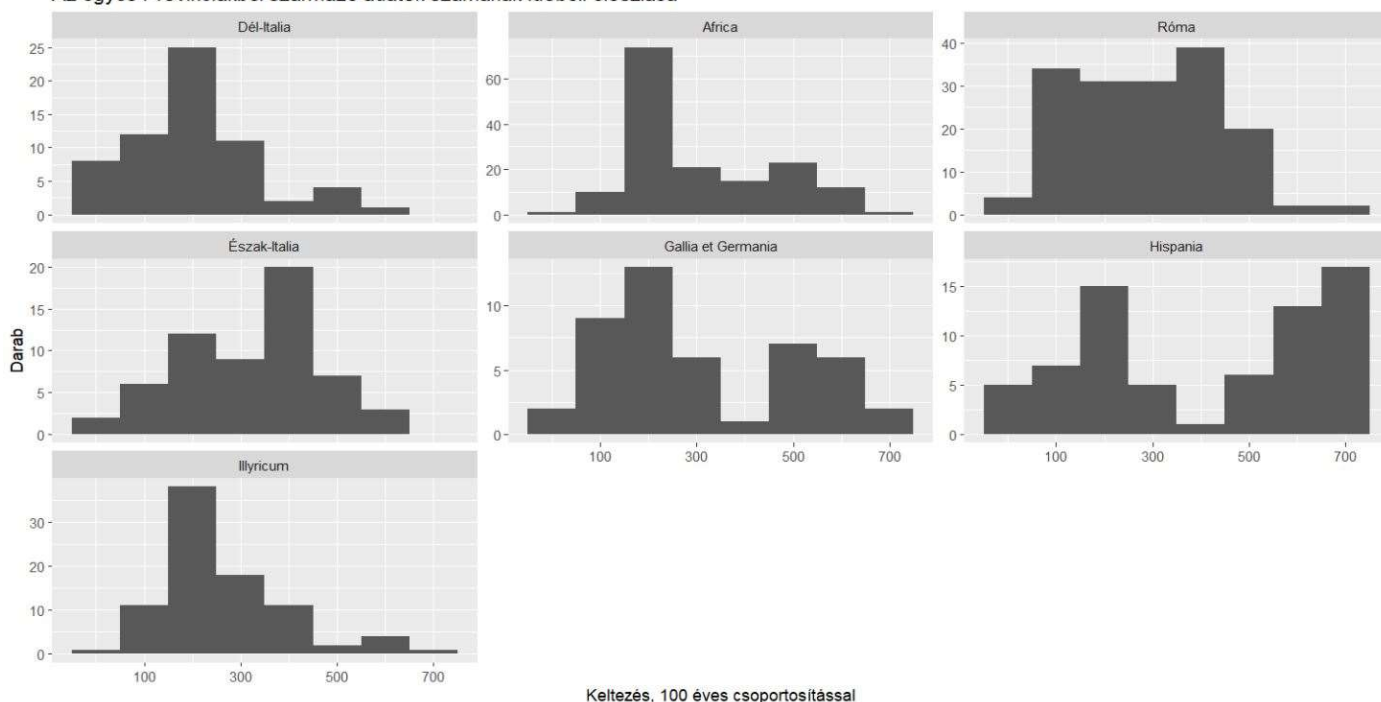


Konklúzió és a kutatás további iránya

A dolgozat azon célja miszerint statisztikai eszközökkel kívántuk megvizsgálni az alternatív kódok két hazai tanulmány eredményeire gyakorolt lehetséges hatását, teljesült. Az alkalmazott modell közelítőleg megfelelő módon tudta leírni a korábbi eredményeket, miközben kimutatta azok esetleges problémáit. Megmutattuk ugyanis, hogy Szlovicsák (2022b:266) eredményei alapvetően jól mutatták be a területi különbségeket és hozzá hasonlóan itt is Africa és Róma esetén lehetett egy stabilitási tendenciát megfigyelni. Kiemelendő volt az is, hogy míg Szlovicsák (2022a:53) megfigyelései szerint a hiperkorrekt értelmezés mellett jelentős különbségek léteztek az egyes területek között, addig itt azt láthattuk, hogy ezen különbségek nagy része eltűnik, ha figyelembe vesszük az alternatív kódokat. Kiderült, hogy ezek az eredmények és különbségek inkább álltak kapcsolatban az alternatív kódokkal, mint a területi különbségekkel. Érdekes volt továbbá, hogy a *Nominalia* alternatív kóddal rendelkező adatok nem voltak jelentős hatással a nem hiperkorrekt módon kódolt adatok eloszlására.

8. Diagram

Az egyes Provinciákból származó adatok számának időbeli eloszlása



Jelen dolgozat eredményeit két fontos kérdés miatt előzetes megállapításoknak és nem végleges konklúzióknak érdemes tekinteni. Egyrészt többször felmerült mint probléma az

alternatív kódok nagyon alacsony száma, amennyiben ugyanis egyszerre nézzük az alternatív kódok és provinciák hatását, akkor meglehetősen alacsony lesz az egyes területekről származó adatok száma és nagyon nehéz lesz statisztikai eszközök segítségével biztos jellemzőket kivenni. Ezen problémát a jövőben vagy kis elemszámú mintákra kifejlesztett statisztikai eszközök segítségével lehet kiküszöbölni, vagy érdemes lesz áttérni kvalitatív eszközök alkalmazására a különböző Provinciákban előforduló lehetséges alternatív értelmezések hatásának tisztázására. Mindemellett jelen kutatás nem vette figyelembe, de nem is tudta figyelembe venni a vizsgált alternatív kódok sajátos, provinciánként eltérő eloszlását és jellemzőit.

Mindemellett, amint arra Van de Velde és Petré (2020:346) is felhívják a figyelmet, jelen vizsgálat alatt közel sem biztos, hogy a legjobb döntés volt az idő folytonosként való kezelése, ennek fényében és az így kapott eredmények gyakori bizonytalansága miatt a jövőben mindenképpen érdemes lesz leszűkíteni a vizsgálatot korszakok segítségével, maradni tehát a kutatás korábban bevett döntéseinél. Emellett viszont kiderült az eredmények értékelése alatt, hogy a genuszkeveréses adatok számának időbeli eloszlása a várttal ellentétben nem független az időtől, azaz egyes provinciák esetén, amint ezt a 8. Diagram is tükrözi, az adatok száma lecsökken a 2-3. század után, míg más provinciákban, például Hispaniában még növekedés is látható, Rómában pedig hosszabb időn át konstans az adatok száma, azon kívül pedig meglehetősen alacsony. Ezt az egyenlőtlen eloszlást az se oldja meg, ha az adatokat Korábbi (1-3. század) és Későbbi (4-7. század) korszakra osztjuk, a különbségek ugyanis ekkor is megmaradnak. Emiatt ezt az eltérést valamilyen módon orvosolni kell a későbbiekben, hogy megfelelő következtetések legyenek levonhatóak provinciánként, a jelenlegi modell ugyanis ezt nem vette figyelembe, ahogyan a többi említett tanulmány se, ami ezeket az adatokat vizsgálta.

Emellett a továbbiakban fontosnak tartjuk szemügyre venni azt is, hogy a jelen tanulmányban alapszintként használt *Masc/Fem* genuszkeverések száma az időben mennyire változékony, és erre milyen hatással vannak a területi és egyéb különbségek, a genuszkeveréseken kívül. Ugyanis akkor lehet ennek a szintnek a mentén jó összehasonlításokat végezni, ha ez az adat nem függ a vizsgált változóktól, így ugyanis választ kaptunk arra, hogy ha egy feliraton genuszkeverés szerepel, akkor az mekkora valószínűséggel tartozik a vizsgált kategóriák egyikébe, de arra nem tudtunk választ adni, hogy a genuszkeverések általános bekövetkezési valószínűsége hogyan változik az időben. Erre véleményünk szerint a jövőben a nem nyelvtani hibák (*Errores non grammaticae*) alkalmas

megoldást kínálhatnak, amennyiben kiderül, hogy területileg nem állandó a *Masc/Fem* keverések aránya, ugyanis az sejthető, hogy a nem nyelvtani hibák eloszlása nem függ dialektológiai és nyelvtörténeti különbségektől. A megfelelő viszonyítási alap ugyanis könnyen segíthet az eltérő elemszámokból eredő esetleg problémák kiküszöbölésében. Ennek a felvetésnek megvizsgálása viszont már a kutatás folytatására marad.

Összességében tehát sikeresen megmutattuk a multinomiális logit regresszió alkalmazásának lehetőségeit és nehézségeit a vizsgált genuszkeverések elemzésére és leírására. Ennek révén több olyan részeredményt és javaslatot tudtunk megfogalmazni, ami a későbbiekben mindképpen a kutatás előrehaladását fogja szolgálni.

Bibliográfia

- ADAMIK Béla (2012), „A császárkori feliratok vulgáris latin nyelvi adatainak dialektológiai érvényessége”, *Antik Tanulmányok* 56, 91–105.
- ADAMIK Béla (2019), „Az esetrendszer átépülése az afrikai latinságban a feliratok tükrében”, *Publicationes Universitatis Miskolcensis, Sectio Philosophica* 23.1, 11–34.
- AGRESTI, Alan (2013³), *Categorical Data Analysis*, Wiley Series In Probability And Statistics, John Wiley & Sons, Hoboken.
- HERMAN József (2003), *Vulgáris latin. Az újlatin nyelvek kialakulásának útja*, Tinta Könyvkiadó, Budapest.
- LEVSHINA, Natalia (2015), *How to do Linguistics with R. Data Exploration and statistical analysis*, John Benjamins Publishing Company, Amsterdam/Philadelphia.
- LOPORCARO, M. (2018), *Gender from Latin to Romance. History, Geography, Typology*, Oxford Studies in Diachronic and Historical Linguistics 27. Oxford.
- SZLOVICSÁK Béla (2022a), „A semlegesnem leépülésének problémája a vulgáris latinban a feliratos források tükrében”. *Ókor* XXI/1, 46–55.
- SZLOVICSÁK Béla (2022b), „A nyelvtani nemi rendszer dialektológiai jellemzői a vulgáris latinban”. *Antik Tanulmányok* 66/2, 255–267.
- UGBA, Ejike R – GERTHEISS, Jan (2022), „A Modification of McFadden’s R^2 for Binary and Ordinal Response Models”. [arXiv:2204.01301v1](https://arxiv.org/abs/2204.01301v1) [stat.ME] (utolsó elérés: 2023. 01. 08.)
- VÄÄNÄNEN, Veikko (1981³), *Introduction au latin vulgaire*, Klincksieck, Paris, 101–106.
- VAN DE VELDE, Freek – PETRÉ, Peter (2020), „Historical Linguistics” In: *The Routledge Handbook of English Language and Digital Humanities*, szerk. ADOLPHS, Svenja – KNIGHT, Dawn, Routledge, Oxon, 328–352.
- WINTER, Bodo – GRICE, Martine (2021), „Independence and generalizability in linguistics”. *Linguistics* 59/5, 1251–1277.