

Ring Orsolya

Társadalomtudományi Kutatóközpont, Politikatudományi Intézet

CSS-RECENS Kutatócsoport

ring.orsolya@tk.hu

Kiss László

Társadalomtudományi Kutatóközpont

CSS-RECENS Kutatócsoport

kiss.laszlo@tk.hu

Agrárpolitikai kihívások és jogszabályalkotás a korai Kádár-korban Történeti források szövegelemzése és szövegbányászati vizsgálata*

Tanulmányunkban a magyarországi kollektivizálás második hullámát egy, a társadalomtudományban használatos módszer adaptálásának segítségével vizsgáljuk, rámutatva, hogy a nagy adatbázisok elemzése milyen nagyobb, általánosabb történettudományi összefüggések levonására kínál lehetőségeket. Majd ugyanezt a kutatási kérdéskört egy szövegbányászati módszer, az úgynevezett topikmodellezés alkalmazásával elemezzük. Írásunk célja annak illusztrálása, hogy az egyre nagyobb számban elérhető digitális gyűjteményeknek a hazai történettudomány számára még részben újfajta módszerekkel történő elemzése milyen lehetőségeket és korlátokat rejt magában. Elemzésünk a korai Kádár-korszak agrárpolitikai jogszabályalkotásának vizsgálatát végzi el, arra a fő kutatási kérdésre keresve a választ, hogy milyen módon tükröződik vissza a jogszabályalkotásban a korszak gazdaságpolitikája, milyen módon reagált a jogszabályalkotás a korszak gazdaság- és társadalompolitikai változásaival kapcsolatban felmerülő szabályozási igényekre. A témaalapú osztályozás mellett szövegeinket szövegbányászati módszerekkel is vizsgálat alá vesszük, elemzésünk harmadik részében pedig kísérletet teszünk a két módszer kombinálására – azzal a céllal, hogy összefüggést találjunk a CAP-kódkönyvben meghatározott közpolitikai topikok, konkrétan az agrárpolitikai altopikok, valamint a topikmodellezés során kapott témák között.

Kulcsszavak:

digitális bölcsészet, történettudomány, szövegbányászat, osztályozás, topikmodellezés

* A kutatást az FK-131826 témaszámú NKFIH projekt (Az 1945 és 1989 közötti sajtó társadalomtörténeti elemzése a természetesnyelv-feldolgozás [NLP] módszereivel) támogatta.



1. Bevezetés

Napjainkra hatalmas mennyiségű történeti forrásanyag vált elérhetővé és kereshetővé különböző digitális gyűjteményekben, így például az Arcanum Digitális Tudománytárban,¹ a Hungaricana² vagy az Adatbázisok Online³ felületén. A digitalizált szövegek és az ezekhez rendelt metaadatok elemzése nagyban megkönnyíti a forrásokhoz való hozzáférést, de egyben új kihívásokat is támaszt, hiszen lehetőséget ad a *big data/big text* jellegű megközelítés történettudományi kutatásokra történő kiterjesztésére, amihez azonban a források újfajta megközelítése, feldolgozásukhoz pedig újfajta módszertan szükséges.

Gyáni Gábor szerint az állam és a politika múltját vizsgáló történetírás a 20. század második-harmadik évtizedéig tisztán narratív volt, alkalmazói egyáltalán nem éltek a kvantifikáció eszközével. A két világháború között következett be az, az *Annales* folyóirat vezető történészei által kezdeményezett paradigmaváltás, aminek következtében a történetírás szakított a szorosán vett filológiai, szövegkritikai módszertannal, és elkezdte kihasználni a mérésben rejlő elemzési lehetőségeket, elsődlegesen a gazdaság- és társadalomtörténet-írás területén a különböző statisztikai adatsorok elemzésbe való bevonásával. A kimondottan kvantitatív történeti kutatási programok először az 1970-es években öltöttek komolyabb méreteket, több nagy kutatócsoport is alakult a 19–20. századi, nagy tömegű, formalizálható adatokat magukban rejtő forrásanyagok gépi feldolgozására.⁴

Jean Marczewski a kvantitatív módszereknek három gyakorlatát írta le. Az első az a megszokott történetírói gyakorlat, amely a történelmi események leírását – illusztrációképpen – statisztikai adatokkal egészíti ki; a második az úgynevezett kliometrikus történetírás, ahol a történész közgazdasági magyarázó modelleket tesztel történeti anyagon.⁵ Az ezt a módszert alkalmazó történész modelleket szerkeszt, melyekkel a történeti valóságot értelmezi, bemutatja a változók közötti lehetséges viszonyokat, megkeresi a változóknak a történeti múltban megfelelő értékeket, és az ebből következő függő változók értékét összeveti a változók tényleges értékével. A harmadik az a fajta kvantitatív történetírás, amely az időben visszamenőlegesen hozzáférhető nemzetgazdasági mérlegekre és statisztikákra épít, de nem csupán retrospektív mérlegeket konstruál, hanem „lényeges szerepet tulajdonít az adott időben és adott helyen lejátszódott történeti eseményeknek is”.⁶

¹ Hozzáférés: 2021.02.10, <https://adtplus.arcanum.hu/hu/>.

² Hozzáférés: 2021.02.10, <https://hungaricana.hu/hu/>.

³ Hozzáférés: 2021.02.10, <https://adatbazisokonline.hu/>.

⁴ Gyáni Gábor, „Egy új történeti forrás: az adatbank,” *Történelmi Szemle* 24, 1. sz. (1981): 94–99, 94; Gyáni Gábor, „A statisztika mint konstruált és/vagy objektív történeti adat,” *Statisztikai Szemle* 96, 3. sz. (2018): 245–254, 246, <http://dx.doi.org/10.20311/stat2018.03.hu0245>.

⁵ Eddie Scott M., *Ami „köztudott”, az igaz is? Bevezetés a kliometrikus történetírás gondolkodásmódjába* (Debrecen: Csokonai Kiadó, 1996).

⁶ Jan Marczewski, „Kvantitatív történetírás,” in *Történetelméleti és módszertani tanulmányok*, szerk. Glatz Ferenc (Budapest: Gondolat Kiadó, 1977), 291–303, 294. Gyáni, „A statisztika,” 249–250.

A fenti igen vázlatos áttekintés is illusztrálja, hogy a kvantitatív történettudomány elsősorban a forrásokban rejlő strukturált, számszerű(síthető) adatok – statisztikák, összeírások stb. – elemzésében ragadható meg. Az 1970-es években a hazai történettudományban is megkezdődött a diskurzus a számítógépek kutatási alkalmazásainak lehetőségeiről. Már ekkor felismerték, hogy a számítógép használata nemcsak olyan források feldolgozását teszi lehetővé, amire manuálisan nem nyílta lehetőség, de egyben a forrásokból kinyerhető információk körét is bővíti.⁷ A számítógépek és a történettudomány kapcsolatáról folyó diskurzus párhuzamosan zajlott és összekapcsolódott a történettudományi kutatások kvantifikálásáról szóló eszmecserevel.⁸

A társadalomtudományokban az 1980–1990-es évek fordulóján jelentek meg az első olyan statisztikai eljárások, amelyek nagyobb szövegtörzseket voltak képesek komplex módon elemezni, de a módszer használatának felfutása csak a 2000-es években indult meg.⁹ Az automatizált szövegelemzési eljárások megjelentek más tudományterületeken, így például az irodalomtudományban¹⁰ és a történettudományban is.¹¹ Bár a digitalizáció révén egyre több számítógéppel is elemezhető, strukturálatlan történeti adathalmaz (dokumentum) áll rendelkezésünkre, amely alkalmas mind a számítógéppel támogatott kvalitatív, mind a kvantitatív szövegelemzésre, a nagy szövegtörzsek számítógépes elemzésének elterjedését nagyban nehezítette és nehezíti, hogy az elemzendő szövegek nem digitálisan keletkeztek, azokat visszamenőleg digitalizálni kell, az így kapott szövegek minősége pedig gyakran csak sok előkészítő, tisztító munka után alkalmas az automatizált szövegelemzésre. Bár az OCR-szoftverek az utóbbi években jelentős fejlődésen mentek keresztül, az eredetileg sokszor rossz minőségű papíron keletkezett, vagy nem megfelelő felbontásban szkennelt

⁷ Gyáni, „Egy új történeti forrás,” 94. Granasztói György, „A számítógépek a történettudományban,” *Történelmi Szemle* 15, 1–2. sz. (1972): 29–47, 36; lásd továbbá „Számítógépek és a történettudomány,” *Történelmi Szemle* 15, 3–4. sz. (1972): 526–583 – lásd Katus László *A számítógép alkalmazásáról a kapitalista korszak kutatásában* című előadásáról szóló részt: 535–536.

⁸ Benda Gyula, „Kvantifikáció és történelem. (Hozzászólás Granasztói György *A történész és a mérés – egy modell korlátai* című írásához),” *Történelmi Szemle* 22, 1. sz. (1979): 105–108.

⁹ Gary Miner, John Elder IV, Andrew Fast, Thomas Hill, Robert Nisbet and Dursun Delen, *Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications* (Cambridge: Academic Press, 2012), <https://doi.org/10.1016/C2010-0-66188-8>; Bing Liu, *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions* (Cambridge: Cambridge University Press, 2015), Jun Zhao, Kang Liu, and Liheng Xu, „*Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Bing Liu (University of Illinois at Chicago) Cambridge University Press, 2015, 381 pp.; hardcover, ISBN 9781107017894, \$80,” *Computational Linguistics* 42, 3. sz. (2016): 595–598, https://doi.org/10.1162/COLI_r_00259; Németh Renáta, Katona Eszter és Kmetty Zoltán, „Az automatizált szövegelemzés perspektívája a társadalomtudományokban,” *Szociológiai Szemle* 30, 1 (2020): 44–62, 45.

¹⁰ Franco Moretti, *Distant Reading* (New York–London: Verso Books, 2013); Péter Róbert, „A Big Data kihívás és lehetőség a bölcsészettudományokban: Digitális szövegek és metaadatok távoli olvasása,” *Magyar Tudomány* 177, 11. sz. (2016): 1323–1330.

¹¹ Shawn Graham, Ian Milligran and Scott Weingart, *Exploring Big Historical Data: The Historian’s Macroscopic* (London: Imperial College Press, 2015), <https://doi.org/10.1142/p981>; B. Alex, C. Grover, E. Klein, C. Llewellyn and R. Tobin, „User-Driven Text Mining of Historical Text,” in *Working with Text: Tools, Techniques and Approaches for Text Mining*, eds. Emma L. Tonkin and Gregory J. L. Tourte (Oxford: Chandos Publishing, 2016), 209–230, <https://doi.org/10.1016/B978-1-84334-749-1.00009-3>.

dokumentumok szövegbányászati elemzésre való előkészítése még mindig jelentősen több időráfordítást igényel, mint amire a *webscraping* segítségével gyűjthető szövegek esetében van szükség.¹²

A kvantitatív (pl. statisztika) és kvalitatív (pl. szöveges dokumentum) forrástípusok, a kvantitatív és kvalitatív történeti elemzés mellett különbséget tehetünk a szövegek kvantitatív és kvalitatív elemzési módszertana között is. Kvalitatív módszertan (kvalitatív szöveg-, illetve tartalomelemzés, számítógéppel támogatott szövegelemzés) alkalmazása esetén a szöveget jegyzetekkel és kódokkal látjuk el, amit végezhetünk gépi támogatással vagy anélkül, de a szövegen magán nem végzünk beavatkozást.¹³ Kvantitatív szövegelemzés (szövegbányászat, szöveganalitika) esetén viszont az eredendő kvalitatív információforrást kvantitatív módon képezzük le.¹⁴

A szöveges történeti források kvantitatív szövegbányászati és statisztikai módszerek használatával történő elemzése révén lehetőséget kaphatunk a szövegek közötti kapcsolatok, a bennük rejlő mintázatok feltárására és elemzésére, olyan módon, ahogyan azt a szövegek egyenkénti olvasásával nem tudnánk megtenni, s ezáltal új összefüggések feltárására nyílik lehetőségünk. Írásunk célja annak illusztrálása, hogy az egyre nagyobb számban elérhető digitális gyűjteményeknek a hazai történettudomány számára még részben újfajta módszerekkel történő elemzése milyen lehetőségeket és korlátokat rejt magában.¹⁵

Esettanulmányunk a magyarországi kollektivizálás második hullámát (1957–1963) járja körül, ami egy viszonylag jól feldolgozott téma; célunk éppen annak bemutatása, hogyan lehet új elemzési eszközökkel, más tudományágak által alkalmazott modellek segítségével tovább mélyíteni a folyamatról korábban alkotott képet. Az elemzés

¹² Daniel Lopresti, „Optical Character Recognition Errors and their Effects on Natural Language Processing,” *International Journal on Document Analysis and Recognition (IJ DAR)* 12, 3. sz. (2009): 141–151, <https://doi.org/10.1007/s10032-009-0094-8>.

¹³ Roel Popping, *Computer-Assisted Text Analysis* (London–California–New Delhi: Sage, 2000), <https://dx.doi.org/10.4135/9781849208741>; Sebők Miklós, Zorigt Burtejin és Zágoni Bella, „A szövegek társadalomtudományi elemzése: Egy elméleti áttekintés,” in *Kvantitatív szövegelemzés és szövegbányászat a politikatudományban*, szerk. Sebők Miklós (Budapest: L’Harmattan Kiadó, 2016), 15–22, 20.

¹⁴ Justin Grimmer and Brandon Stewart M., „Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts,” *Political Analysis* 21, 3. sz. (2013): 267–297, <https://doi.org/10.1093/pan/mps028>; Sebők, Zorigt és Zágoni, „A szövegek társadalomtudományi elemzése,” 20–21.

¹⁵ Hasonló történeti szövegtörzsek NLP (Natural Language Processing) módszerekkel történő elemzésével foglalkoznak például az alábbi munkák is: Martina Katalin Szabó, Orsolya Ring, Balázs Nagy, László Kiss, Júlia Koltai, Gábor Berend, László Vidács, Attila Gulyás and Zoltán Kmetty, „Exploring the Dynamic Changes of Key Concepts of the Hungarian Socialist Era with Natural Language Processing Methods,” *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 2020. szept. 23., <https://doi.org/10.1080/01615440.2020.1823289>; Zoltán Kmetty, Veronika Vincze, Dorottya Demszky, Orsolya Ring, Balázs Nagy and Martina Katalin Szabó, „Pártélet: A Hungarian Corpus of Propaganda Texts from the Hungarian Socialist Era,” in *Proceedings of the 12th Language Resources and Evaluation Conference*, eds. Nicoletta Calzolari et al., (Marseilles: European Language Resources Association, 2020), 2381–2388; Ring Orsolya, Kmetty Zoltán, Szabó Martina Katalin, Kiss László, Nagy Balázs és Vincze Veronika, „Kulcsfogalmak jelentésváltozása a Kádárkorszak politikai diskurzusában,” in *XVI. Magyar Számítógépes Nyelvészeti Konferencia*, szerk. Berend Gábor, Gosztolya Gábor és Vincze Veronika (Szeged: Szegedi Tudományegyetem, 2020), 333–342.

során a kollektivizálás mint történelmi eseménysor politikai vetületére koncentrálnak mutatjuk be az agrárszektor átalakításának megjelenését a Kádár-korszak jogszabályalkotásában, az MSZMP Politikai Bizottsága napirendjében, kitérve a kollektivizálással kapcsolatos egyéb társadalom- és gazdaságpolitikai témák megjelenésére, felfutására, ezek összefüggéseire is.

2. Források és adatbázisok

A *Magyar Közlönyben* megjelent jogszabályok adatbázisát az Optijus anyagainak felhasználásával állítottuk össze.¹⁶ A jogalkotás vizsgálata során a *Magyar Közlönyben* publikált nyilvános anyagok mellett a Minisztertanács által kiadott titkos határozatok kódolt adatbázisát is bevontuk az elemzésbe. Az MSZMP Politikai Bizottsága napirendi jegyzéke esetében a Magyar Nemzeti Levéltár napirendi jegyzékeit és adatbázisait használtuk, ez alapján épült fel az MSZMP PB napirendjének adatbázisa.¹⁷

Adatbázisaink alap metaadat-struktúrája a jó összehasonlíthatóság érdekében minden esetben ugyanazt a rendszert követi. Minden egyes adatbázisrekord egyedi azonosítót tartalmaz, valamint az adott rekord dátumát, illetve ha rendelkezésre áll, a korabeli egyedi azonosítószámát, egy részletes leíró címet és az adott rekord közpolitikai kódszámát.

A *Magyar Közlönyben* megjelent jogszabályok szövegbányászati elemzéséhez szövegeinket PDF-ből TXT-formátumra konvertáltuk, a visszakereshetőség érdekében a fájlok elnevezésében szerepeltettük a kódolt adatbázis „id” oszlopában lévő egyedi azonosítószámot.

3. Módszertan

Tanulmányunkban először kvalitatív módszertannal nagy adatbázisokon végzett kutatást ismertetünk, egy a társadalomtudományban használatos módszer adaptálására, és annak segítségével történettudományi kérdések megválaszolására teszünk kísérletet. Célunk annak bemutatása, hogy a nagy adatbázisok elemzése milyen nagyobb, általánosabb összefüggések levonására ad lehetőséget, hogyan használható arra, hogy megtaláljuk azokat a területeket, ahol érdemes a kutatás során mélyfúrásokat végezni, továbbá hogyan segítheti a történelmi kutatásokat az eredmények számszerűsítésével. A választott módszertan egy deduktív eljárás, az osztályozás vagy klasszifikálás, amely során az elemezni kívánt szövegtörzsek elemeit előzetesen kialakított kategóriákba

¹⁶ Hozzáférés: 2021.02.10, <https://optijus.hu/jogi-informacios-szolgalattasok/optijus>.

¹⁷ A mostani elemzésbe bevont jogszabályok a teljes korpusznak csak egy részhalmazát teszik ki. A jelenleg rendelkezésre álló teljes adatbázis a *Magyar Közlönyben* megjelent 1957 és 1989 közötti jogszabályok esetében 13918, a Minisztertanács titkos határozatai esetében 5228 kódolt rekordot tartalmaz. A titkos határozatok sora nem teljes, mivel ezek csak néhány éve vesztették el titkos minőségüket, így a belőlük való adatbázis-építés jelenleg is folyamatban van, jelenleg az 1957 és 1967 közötti titkos jogszabályok feldolgozása történt meg. Az 1957 és 1989 közötti időszak MSZMP PB adatbázisa 19843 kódolt napirendi pontot tartalmaz.

rendeztük, azaz előre tisztában voltunk a sokaság rendezésének alapjául szolgáló jellemzőkkel.¹⁸

Elemzésünk során a politikatudományban használt, úgynevezett CAP-kódolás (Comparative Agendas Project) módszerét alkalmaztuk, hogy meghatározzuk adatbázisainkban az egyes közpolitikai témákat, és számszerűsítsük előfordulásukat. Frank R. Baumgartner és Bryan D. Jones amerikai politológusok a közpolitikai változás mintázatait vizsgálva figyeltek fel arra, hogy ezek a változások általában fokozatosan következnek be, de időnként felgyorsulnak és a kormányzat látszólag váratlanul kezd reformokba egy-egy közpolitikai területen. Baumgartner és Jones szerint a közpolitikai változás részben magából a politikai-közpolitikai rendszer működéséből ered mivel a közpolitikai aktorok (pártok, civil szervezetek) időnként képesek arra, hogy egy-egy kérdést a közpolitikai napirendekre felvigyenek, és ha eléri, hogy a nagypolitika figyelme is ráirányuljon az adott problémára, akkor ezáltal akár jelentős közpolitikai változásokat idézhetnek elő.¹⁹

Baumgartner és Jones elemzésükhöz kialakítottak egy kódkönyvet, amelyben eredetileg 19 közpolitikai főcsoportkódot (pl. makrogazdasági politika) és ezeken belül összesen 225 altémát (pl. költségvetési politika, monetáris politika stb.) határoztak meg. A kódok kizárólagosak, tehát minden megfigyelés csak egy kódot kaphat. Ez a kódolás lehetővé teszi például annak vizsgálatát, hogy egy adott közpolitikai témáról mikor beszélt sokat a sajtó, mikor tartották az emberek is fontosnak, és mikor született belőle jogszabály.²⁰ Baumgartner és Jones projektje (Comparative Agendas Project)²¹ mára a nemzetközi politikatudomány egyik jelentős adatbázisprojektjévé nőtte ki magát, amelyben az első kelet-közép-európai résztvevő a Hungarian Comparative Agendas Project²² lett. Ennek célja a magyar közpolitika dinamikájának feltárása, amihez konzisztens adatbázisokat hoznak létre a magyar közpolitikai folyamatok meghatározó aktorai – így a parlament, a média és a közvélemény szereplői – számára meghatározó ügyekről. A módszer megbízhatóságát az egységes és pontosan lefektetett kódolási szabályok garantálják, amelyek révén a különböző közpolitikai területek prioritásainak változásai időben és területenként is nyomon követhetők.²³

Az osztályozást a nemzetközi CAP-projekt módszertana alapján kettős vak kézi kódolással végeztük, azaz minden esetben két kutató egymástól függetlenül határozta meg, hogy az adott megfigyelés melyik közpolitikai topikba vagy altopikba tartozik.

¹⁸ *Szövegbányászat*, szerk. Tikk Domonkos (Budapest: Typotext Kiadó, 2007); Molnár Csaba, „Osztályozás (klasszifikáció),” in Sebők, *Kvantitatív szövegelemzés*, 65–72.

¹⁹ Boda Zsolt, „A közpolitikai változás elméletei és a megszakított egyensúly,” in *A magyar közpolitikai napirend: Elméleti alapok, empirikus eredmények*, szerk. Boda Zsolt és Sebők Miklós (Budapest: MTA TK PTI, 2018), 15–23.

²⁰ Frank R. Baumgartner and Bryan D. Jones, *Agendas and Instability in American Politics* (Chicago: University of Chicago Press, 2010); Boda Zsolt, „A közpolitikai változás elméletei és a megszakított egyensúly,” in Boda és Sebők, *A magyar közpolitikai napirend*, 15–23; Sebők Miklós, „A magyar Comparative Agendas Project módszertana,” in *uo.*, 23–30.

²¹ Hozzáférés: 2021.02.10, <https://www.comparativeagendas.net/>.

²² Hozzáférés: 2021.02.10, <https://cap.tk.mta.hu/>.

²³ Boda Zsolt és Sebők Miklós, „Előszó: a Hungarian Comparative Agendas Project bemutatása,” *Politikatudományi Szemle* 24, 4. sz. (2015): 33–40.

Miután a két független kódolást összevetettük, a vitás esetekben konszenzusos kódokat osztottunk ki.²⁴

Ezt követően ugyanazt a kutatási kérdéskört kvantitatív módszertannal is megvizsgáltuk az egyik legelterjedtebb szövegbányászati módszer, az úgynevezett csoportosítás vagy klaszterezés és azon belül a topikmodellezés alkalmazásával. Ebben az esetben a korpusz elemeinek rendezéséhez nem határozunk meg előzetesen csoportokat. A topikmodellezés olyan felügyelet nélküli tanulási módszer, amelynek során az alkalmazott algoritmus a dokumentum tulajdonságait és a modell becsléseit felhasználva hoz létre különböző kategóriákat, amelyekhez később hozzárendeli a szöveget.²⁵ Az általunk alkalmazott topikmodellezési eljárás, a látens Dirichlet-allokáció (LDA) alapja az a feltételezés, hogy minden korpusz topikok/témák keverékéből áll, ezen témák pedig statisztikailag a korpusz szókészlete valószínűségi függvényeinek (eloszlásának) tekinthetők.²⁶ Az LDA a korpusz dokumentumainak csoportosítása során az egyes dokumentumokhoz topikszavakat rendel, a topikok megbecsléséhez pedig a szavak együttes megjelenését vizsgálja a dokumentum egészében. Az LDA algoritmusnak előzetesen meg kell adni a keresett klaszterek (azaz a keresett topikok) számát, ezt követően a dokumentumhalmazban szereplő szavak eloszlása alapján az algoritmus azonosítja a kulcsszavakat, amelyek eloszlása kirajzolja a topikokat.²⁷

A klaszterezés vagy csoportosítás egy induktív kategorizálás, ami akkor hasznos, amikor nem állnak a kutató rendelkezésére előzetesen ismert csoportok, amelyek szerint a vizsgált dokumentumokat rendezni tudná. Hiszen ebben az esetben a korpusz elemeinek rendezéséhez nem határozunk meg előzetesen csoportokat, hanem az eljárás során olyan különálló csoportokat hozunk létre a dokumentumokból, amelyek tagjai valamilyen szempontból hasonlítanak egymásra. A csoportosítás legfőbb célja, hogy az egy csoportba kerülő szövegek minél inkább hasonlítsanak egymásra, miközben a különböző csoportba kerülők minél inkább eltérjenek egymástól. Azaz klaszterezésnél nem egy-egy szöveg jellemzőire vagyunk kíváncsiak, hanem arra, hogy a szövegek egy-egy csoportja milyen hasonlóságokkal bír.²⁸

Az általunk választott osztályozás és klaszterezés természetesen egymástól függetlenül is alkalmazható eljárások, de jól ki is egészíthetik egymást.

²⁴ A CAP-kódolás nemcsak kézzel, hanem felügyelt gépi tanulással is végezhető, 1990 utáni magyar nyelvű dokumentumokon tanítva jelenleg is használatban van egy felhőalapú klasszifikációs eljárás, amely nagy hatékonysággal képes a CAP (fő)kategóriáinak osztályozására. Miklós Sebők and Zoltán Kacsuk, „The Multiclass Classification of Newspaper Articles with Machine Learning: The Hybrid Binary Snowball Approach,” *Political Analysis*, 2020. nov. 11., 1–14, <https://doi.org/10.1017/pa.n.2020.27>.

²⁵ Grimmer and Stewart, „Text as Data,” 15.

²⁶ Németh, Katona és Kmetty, „Az automatizált szövegelemzés,” 55; David M. Blei, Andrew Y. Ng and Michael I. Jordan, „Latent Dirichlet Allocation,” *Journal of Machine Learning Research* 3, 1. sz. (2003): 993–1022.

²⁷ Zorigt Burtejin, „Csoportosítás (klaszterezés),” in Sebők, *Kvantitatív szövegelemzés*, 85–101, 98.

²⁸ Tik, *Szövegbányászat*, 145.

4. Elemzés

4.1. Az agrárszektor helyzete a korai Kádár-korszakban

Az 1956-os forradalom eredményeként 1957. január 1-ig az 1956 októberében működő tsz-ek 62%-a felbomlott.²⁹ A forradalom leverését követően hatalomra kerülő Kádár János érzékelve az agrárirányítás Rákosi-korszakbeli csődjét, új irányvonalat határozott meg. A mezőgazdasági termelőket sújtó begyűjtési rendszer 1956. novemberi „hivatalos” eltörlése ugyan még inkább csak jelképes politikai aktus volt (a begyűjtési rendszer eddigre már rég összeomlott), az MSZMP 1957 nyarán közzétett Agrárpolitikai Tézisei azonban már részletesen bemutatták a meghirdetésre kerülő új mezőgazdasági és szövetkezeti politikát.

1957-ben a mezőgazdasági termények zömét még magángazdaságok állították elő. Az 1959-ben megindult kollektivizálás eredményeként 1961-re az ország mezőgazdaságilag hasznosítható területének 80%-a került a tsz-ek használatába.³⁰ Az egyéni gazdaságok száma 165000-re csökkent, s a helyüket 271 állami, és körülbelül 4200 termelőség szövetkezeti gazdaság vette át.³¹ A számok mögött látható a hagyományos paraszti társadalom erőszakos (és csaknem teljes) felszámolása, a parasztság termelőszövetkezetekbe kényszerítése.³²

A jogalkotás új feladata lett az állami tulajdonban lévő állami gazdaságok működésének szabályozása és a termelőszövetkezeti működés irányítása – ennek fő eszközei a tervezés- és támogatáspolitikai, a beruházások irányítása, az ár- és hitelpolitika, valamint a felvásárlási rendszer kiépítése. Az első átfogó szabályozást az 1959. évi 7. törvényerejű rendelet a mezőgazdasági termelőszövetkezetekről és a termelőszövetkezeti csoportokról (márc. 19.) adta.

A legfontosabb feladat a mezőgazdaság beszolgáltatási kötelezettségének visszafogásából adódóan az agrártermékek állami felvásárlásának, kvázi-kereskedelmének szabályozása lett. Az állam, illetve a továbbra is államilag szervezett, központosított élelmiszer-gazdaság csak úgy tudott agrártermékhez, alapanyaghoz jutni, ha valamilyen árrendszer bevezetésével és alkalmazásával megvásárolta azt a termelőktől – az egyénileg gazdálkodóktól, illetve a szövetkezetektől.³³ Közvetlen állami irányítás alatt, centralizálva működtek a gépállomások is, amelyek szerepe a szövetkezetek saját gépekkel való ellátásának lehetővé tételével párhuzamosan fokozatosan csökkent – ezen területek működésének újraszabályozása ugyancsak a korszak jogalkotásának fő feladata lett.³⁴

²⁹ Valuch Tibor, „Agrárkérdések és a magyar falu 1956–57-ben,” in *Évkönyv VIII.*, szerk. Kőrösi Zsuzsanna, Standeisky Éva és Rainer M. János (Budapest: 1956-os Intézet, 2000), 286–302, 300, 303.

³⁰ Fukker Valéria, „A pártoló tag – epizód a tsz-szervezések történetéből,” *ArchivNet* 6, 2. sz. (2006), https://archivnet.hu/politika/a_partolo_tag_epizod_a_tszszervezesek_tortenetebol.html.

³¹ Varga Zsuzsanna, „Az agrárpiac rehabilitálása Magyarországon 1956 után,” *Metszetek* 2–3. sz. (2012): 83–92, 88.

³² Vö. Ö. Kovács József, *A paraszti társadalom felszámolása a kommunista diktatúrában: A vidéki Magyarország politikai társadalomtörténete 1945–1965* (Budapest: Korall Kiadó, 2012).

³³ Varga, „Az agrárpiac rehabilitálása,” 87.

³⁴ Juhász Pál, „Agrárpiac, kisüzem, nagyüzem: (Agrárszociológiai tanulmányok),” *Medvetánc* 2, 1. sz. (1982): 117–139, 120.

Az élelmiszer-ellátás biztosítása, a lakossági fogyasztás biztosítása, és lehetséges volumenének növelése komplex politikai kérdésként jelentkezett – a megkezdődő kádári konszolidációs folyamat egyik alapelemévé vált. A kádári életszínvonal-politika fontos eleme volt az első szándékból a termelészövetkezeti parasztság fogyasztási színvonalának emelését célzó háztáji gazdálkodás bevezetése. A háztáji már rövidtávon komoly hatást ért el a termelészövetkezeti parasztság életszínvonalának növelésében, kontraproduktív volt ugyanakkor maguknak a termelészövetkezeteknek a működése szempontjából. A termelészövetkezetek nem megfelelő jövedelemelosztása, a szakemberhiányból fakadó sok vezetési probléma, de vélhetően az erőszakos kollektivizálás „utóhatása”, a tsz-szel mint termelési egységgel való ellenségesség is hozzájárult ahhoz az érdektelenséghez, ami a tsz-ekben végzett munkát jellemezte. A szövetkezetek működése akadozott, az élelmiszer-ellátásban betöltött szerepük elégtelen volt. Már 1959-től kezdve megjelent a „gyengén teljesítő” szövetkezetek megerősítésének, a szövetkezeti működés megszilárdításának jogszabályi igénye, amely összekapcsolódott a termelészövetkezetek szakemberellátásának fokozásával.

1961 végén az MSZMP Politikai Bizottsága átfogó agrárpolitikai reformmunkába kezdett, amely három nagy területre terjedt ki: az agrárszektor új ár-, adó- és pénzügyi rendszerének kidolgozására, a mezőgazdasági irányítás felülvizsgálatára és az új termelészövetkezeti törvény megalkotására. Bár 1963-ra el is készültek a mezőgazdasági reformjavaslatok, a bevezetésüket a gazdaságban jelentkező problémák miatt 1963–64 fordulóján elhalasztották, és csak 1967-ben született meg az átfogó újraszabályozást tartalmazó – 1967. évi III. törvény a mezőgazdasági termelészövetkezetekről, majd az 1967. évi IV. törvény a földtulajdon és földhasználat továbbfejlesztéséről.³⁵

4.2. Osztályozás

A szocialista korszak(ok) igazgatását a megkettőzött struktúra, a párt- és az államigazgatási irányító szervek egymás melletti, egymás tevékenységét átfedő, alapvetően a pártszervek dominanciája mellett zajló tevékenység jellemezte. A Rákosi- és a Kádár-korszak igazgatási szabályozásának, ezen belül jogforrásainak rendszere meglehetősen bonyolult – a jogszabályok körét a törvények, törvényerejű rendeletek mellett országgyűlési, miniszteri, minisztertanácsi rendeletek és határozatok, utasítások sokasága alkotta.

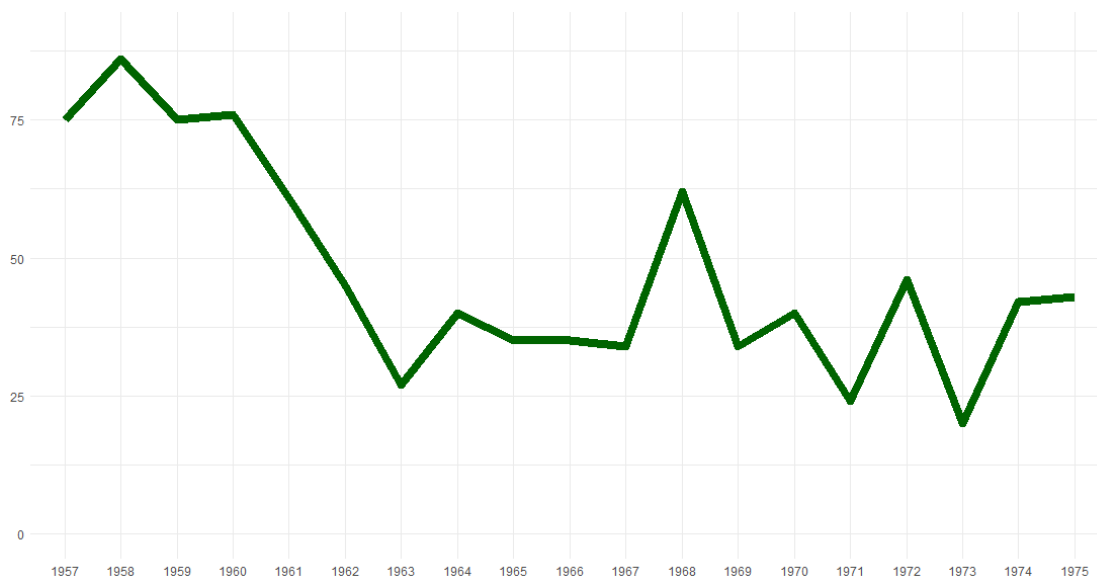
A jogforrások hierarchikus rendszerét a jogszabályok közzétételéről és hatálybalépéséről szóló 1954. évi 26. számú törvényrendelet alapozta meg, később pedig az 1974. évi 24. sz. törvényerejű rendelet szabályozta átfogó jelleggel. A szabályozás egyik központi kérdése volt a jogszabály kihirdetési kötelezettsége, vagy annak hiánya – a *Magyar Közlönyben* kellett kihirdetni azokat a jogszabályokat, amelyek az állampolgárokra közvetlenül kötelezettséget róttak, az állampolgárok jogait vagy kötelességeit érintették, illetve amelyek állami szervekre, szövetkezetekre és egyéb gazdálkodó szervekre vonatkozóan általános érvényűek voltak.³⁶ A *Magyar Közlönyben* kihirdetésre nem kötelezett, de titkosnak sem minősített jogszabályok egy részét az ugyancsak hi-

³⁵ Varga, „Az agrárpiac rehabilitálása,” 90.

³⁶ Ádám Antal, „Jogalkotás, jogforrások és a jogszabályok kihirdetése,” *Jogtudományi Közlöny* 30, 6. sz. (1975): 301–312, 303.

vatalos lapnak számító *Határozatok Tárában*, illetve a különböző ágazati értesítőkből, közlönyökben hirdették ki, sok jogszabálytípus esetében azonban felderíthetetlen volt a kihirdetési rend, sőt egyes, elvben nem titkosnak minősített jogszabályok semmilyen hivatalos fórumon nem kerültek nyilvánosságra.³⁷

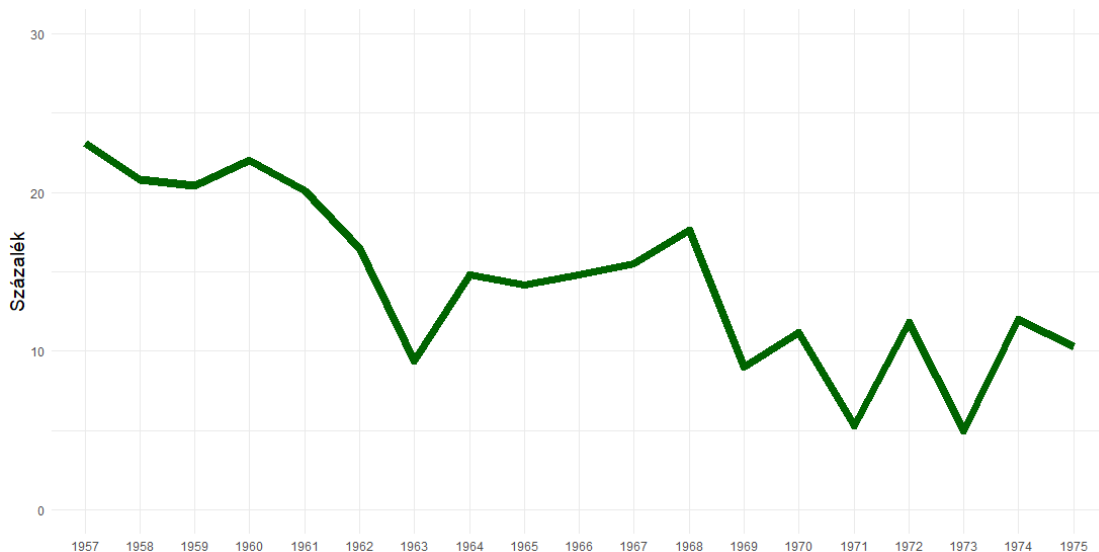
Ha a *Magyar Közlönyben* megjelent jogszabályok CAP-főtopikok szerinti megoszlását vizsgáljuk, jól láthatjuk, hogy a mezőgazdasági szektort érintő jogszabályok száma és aránya 1957 és 1961 között kiemelkedően magas. Az abszolút számokat tekintve csak az 1968-as „reformév” hozott hasonló értéket, az agrárszektor jogszabályainak az összes megjelent jogszabályhoz viszonyított aránya azonban még az „új gazdasági mechanizmus” bevezetésekor sem ért el olyan magas értéket, mint az 1950-es és 1960-as évek fordulóján.³⁸



1. ábra. A *Magyar Közlönyben* megjelent agrárpolitikai jogszabályok száma, 1957–1975

³⁷ Ádám, „Jogalkotás, jogforrások,” 305.

³⁸ Bár elemzésünk az 1957 és 1963 közötti időszak jogszabályalkotását, valamint közpolitikai napirendjét vizsgálja, a kontextusba helyezés végett az agrárpolitikai jogszabályok, titkos határozatok és MSZMP PB napirendi pontok arányát az adott év teljes jogszabályi, határozati, illetve napirendi anyagán belül hosszabb időkeretben, 1957 és 1975 között mutatjuk be.



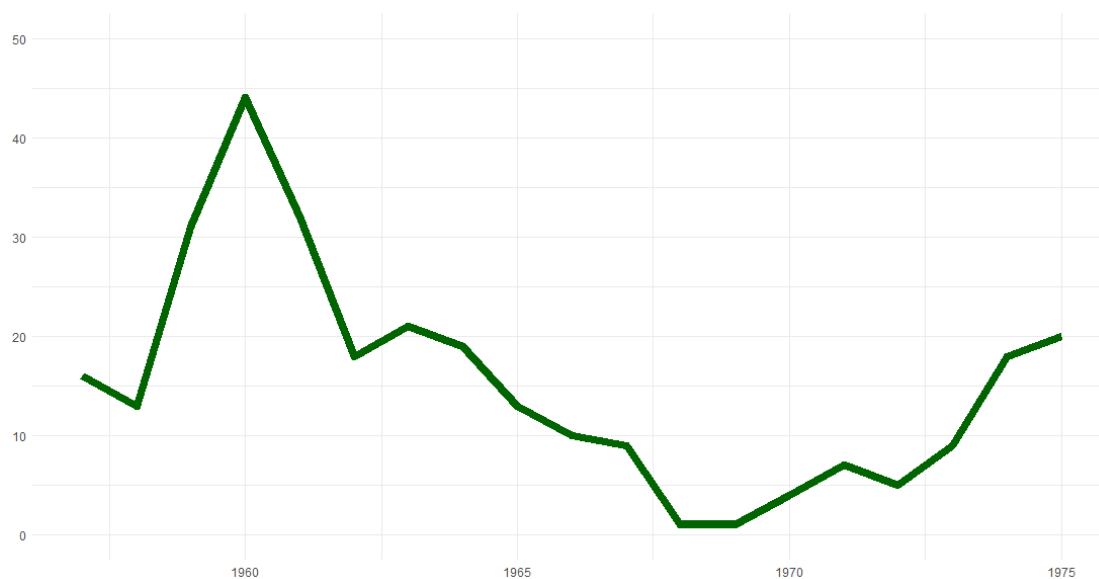
2. ábra. A Magyar Közlönyben megjelent agrárpolitikai jogszabályok összes megjelent jogszabályon belüli aránya (N=6494)

A mezőgazdaság kérdéskörének felértékelődését jól jelzi az MSZMP Politikai Bizottságának fokozott érdeklődése is. Az MSZMP által kiadott párthatározatok, pártutasítások, irányelvek stb. nem számítottak hivatalosan önálló jogforrásnak, a rendszer működéséből fakadóan azonban komoly jelentőséggel bírtak a tényleges jogszabályalkotás folyamatára. A párthatározatok a „demokratikus centralizmus” értelmében kötelező érvénnyel bírtak a párt szervezetre, valamint a pártfegyelem révén a párt teljes tagságára nézve. „Miután a többség elfogadta a párthatározatot, annak pontos végrehajtása a pártélet feltétlen törvénye, amely kötelező minden párttagra nézve, bárhol dolgozik, és bármilyen tisztséget tölt be.”³⁹ A párthatározatok és az államigazgatási szervek által kiadott jogszabályok témák szerinti egymásra épülése ugyan nem feltétlenül valósult meg minden esetben, de kifejezetten gyakorinak számított, hogy a párthatározatokban, pártutasításokban megfogalmazott elvek, előírások egy későbbi hivatalos állami jogszabály szövegében is visszaköszöttek. Sajátos szereppel bírtak a Minisztertanács által alkotott, de kifejezetten titkosnak minősített határozatok. Azon túl, hogy egyes speciális témák szabályozását bizonyos időszakokban kizárólag titkos határozatok útján végezték, sok esetben itt is látható egyfajta egymásra épülés a nyilvános határozatokkal – a titkos határozatokban megalapozott, elindított folyamat végterméke sok esetben nyilvános jogszabály lett.

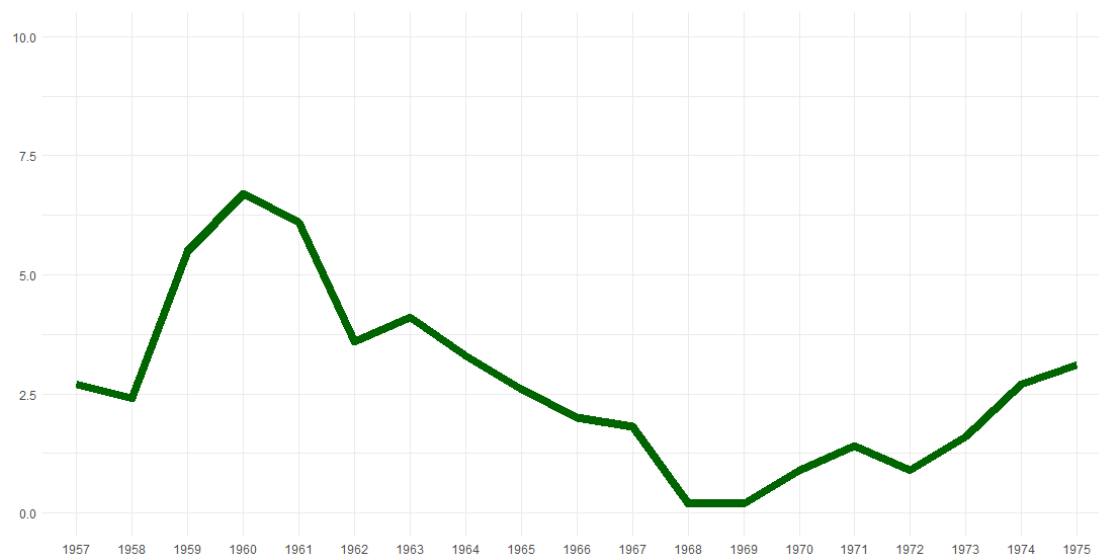
Az MSZMP által hozott párthatározatok száma 1957 és 1963 között 54 volt – ebből kifejezetten agrárpolitikával 15 határozat (27,8%) foglalkozott. Bár ez az arány kifejezetten magas, az abszolút szám relatíve alacsony, így a korpusz további bontásokra, altémakörökbe sorolásra nem alkalmas. A párthatározatok helyett így egy nagyobb korpuszon, az MSZMP Politikai Bizottság üléseinek napirendi pontjain végeztük el az elemzést. 1959 és 1961 között az ülések napirendi pontjai között évente 30–40 agrártéma szerepelt, míg más években ez az érték a 20-at sem nagyon érte el. A Politikai Bizottság tevékenységét ugyan a hatvanas évek elején is a kül- és a belpolitika

³⁹ Milko Balev, „A párthatározat,” *Béke és szocializmus* 25, 4. sz. (1982): 41–42.

dominálta, de a kollektivizálás időszakában 5–7%-ot ért el az összes napirendi ponton belül az agrárpolitika aránya.



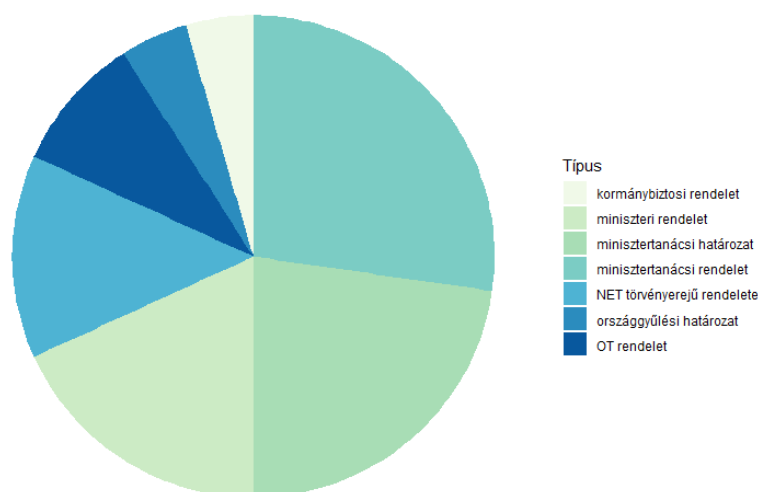
3. ábra. Az MSZMP Politikai Bizottság agrárpolitikai témájú napirendi pontjainak száma



4. ábra. Agrárpolitikai témájú napirendi pontok aránya az MSZMP Politikai Bizottságának napirendjén (N=10191)

A *Magyar Közlöny*ben megjelentetett, agrárpolitikai témájú jogszabályok jogforrások szerinti megoszlását vizsgálva a miniszteri rendeletek túlsúlya fedezhető fel. Az összes agrárpolitikai jogszabály mintegy háromnegyede az ágazatot felügyelő miniszter saját vagy más miniszterekkel közösen kiadott rendelete. A jogforrás-hierarchia magasabb

szintjén elhelyezkedő, az Elnöki Tanács által kiadott törvényerejű rendeletek a vonatkozó jogszabályok 7%-át, a Minisztertanács által alkotott rendeletek és határozatok pedig azok 9, illetve 8%-át tették ki. Az 1957 és 1963 közötti teljes *Magyar Közlönyben* megjelent joganyaggal összevetve az agrárpolitikai szabályozást, a területen a miniszteri rendeletek erőteljes felül-, míg a többi jelentősebb jogforrástípus közepes alulreprezentáltságát állapíthatjuk meg.



5. ábra. A *Magyar Közlönyben* megjelent agrárpolitikai témájú jogszabályok a jogforrás típusa szerint (N=445)

Az adatok alapján több vizsgálatra érdemes kérdést is megfogalmazhatunk: egyrészt milyen, a kollektivizáláshoz kapcsolódó folyamatok indokolták ezt a kiemelkedő méretű mezőgazdasági témájú jogszabályalkotást, másrészt milyen típusú jogszabályok születtek, az agrárszektor működésének mely szegmensei, milyen részterületei kerültek szabályozásra. A két kérdés összekapcsolásaként pedig megvizsgálhatjuk, milyen módon reagált a jogszabályalkotás a kollektivizálással kapcsolatban felmerülő szabályozási igényekre. A jogi szabályozási feladatok áttekintése után nézzük meg, hogyan alakult részleteiben az 1957 és 1963 között az agrárszektorra irányuló politikai figyelem, és milyen részterületekre fókuszáltak a szektort érintő jogszabályok. Az elemzéshez a CAP közpolitikai kódjait használjuk, ezek ugyanis a közpolitikai főtémacörök meghatározásán túl alkalmasak az altémakörök elkülönítésére is. Az „agrárpolitika” főtémacörön belül a CAP-kódkönyv kilenc altémakört tartalmaz:

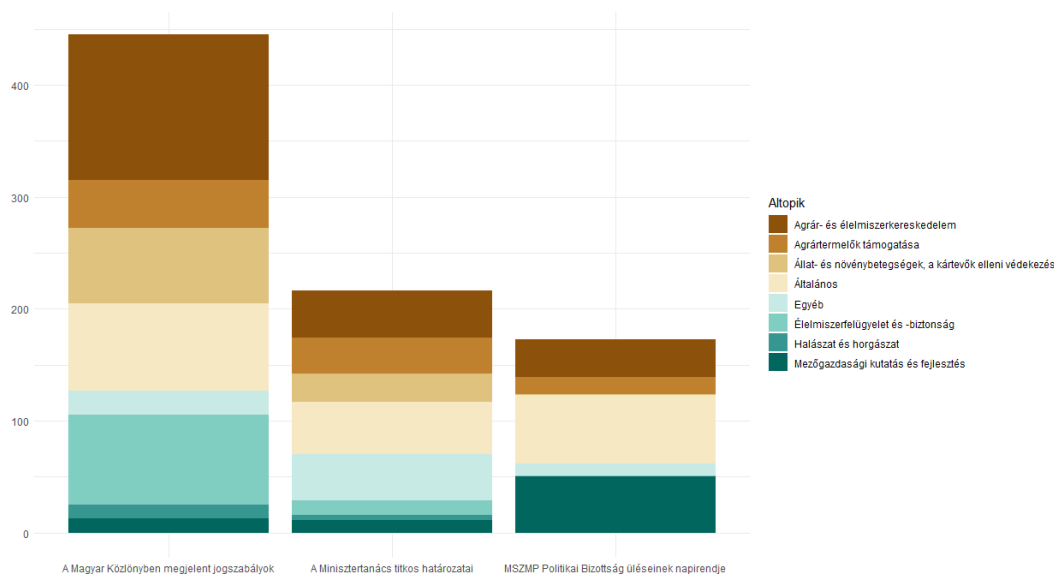
- általános (több altéma kombinációját tartalmazza)
- agrár- és élelmiszer-kereskedelem
- agrártermelők támogatása, intervenciók alapok és biztosítás
- élelmiszer-felügyelet és -biztonság
- agrármarketing és piackutatás
- állat- és növénybetegségek, a kártevők elleni védekezés
- halászat és horgászat

- mezőgazdasági kutatás és fejlesztés
- egyéb

Bár a CAP-kódrendszer nem minden esetben illeszthető könnyen a történeti kutatások szempontrendszeréhez, és az eltérő történelmi korszakok közpolitikájának tagolása, az egyes kategorizálendő elemek témakörökbe sorolása (különösen az altémakörök szintjén) is okozhat nehézségeket, mégis alapvetően alkalmas arra, hogy a közpolitikai figyelem alakulásának történeti elemzéséhez felhasználjuk.

Ha a teljes vizsgált időszakban nézzük az agrárpolitikai jogalkotást, valamint a Politikai Bizottság agrárpolitikával kapcsolatos napirendi pontjait, jól látható a különbség a jogszabályok és a napirendi pontok altémakörök szerinti megoszlása között. A párt és az állam agrárpolitikai tevékenysége eltérő irányultságú, más és más irányítási szegmensekre, területekre helyezi a hangsúlyt. A Politikai Bizottság napirendi pontjai között kiemelkedően sok foglalkozik általános agrárpolitikai témával (35%), valamint mezőgazdasági kutatással és fejlesztéssel (29%), tehát a testület munkáját inkább egyfajta makroszintű szemlélet jellemzi, míg a megjelent jogszabályok esetében az „operatív témák” dominálnak – köztük a párt érdeklődési köréből szinte teljességgel hiányzó állat- és növényvédelem, valamint az élelmiszer-felügyelet. A jogszabályi szint viszont alig érinti a láthatóan a párt hatáskörébe tartozó agrárfejlesztéseket.

Ha visszautalunk az előzőekben kifejtett politikai figyelmet indukáló, illetve jogszabályalkotást szükségessé tevő gazdasági és társadalmi folyamatokra, láthatjuk, hogy az agrártermékek kereskedelmével kapcsolatos árképzési, árusítási, felvásárlást szabályozó kérdések a jogszabályok, a minisztertanácsi titkos határozatok, valamint a Politikai Bizottság napirendi pontjai között is kiemelkedő arányban képviseltetik magukat. A nyilvános agrárpolitikai jogszabályok 30, a titkos határozatok 20%-a foglalkozik az agrárkereskedelemmel (ezen belül elsősorban az árszabályozással), és a Politikai Bizottság agrár tematikájú napirendi pontjainak is egyötödét teszik ki az ebbe a CAP-altémakörbe sorolható elemek.

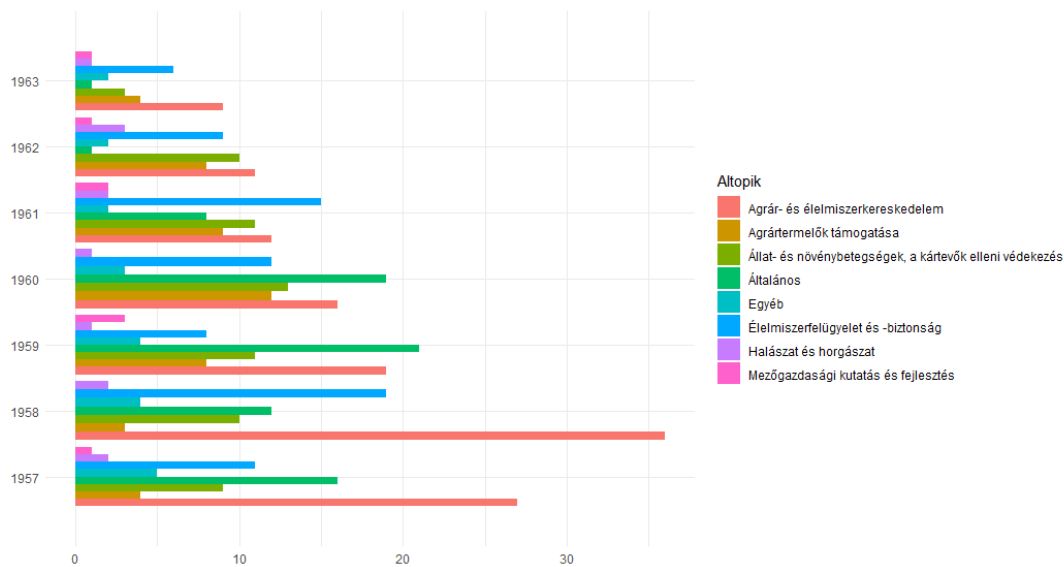


6. ábra. Agrárpolitikai témák altópikok szerinti megoszlása

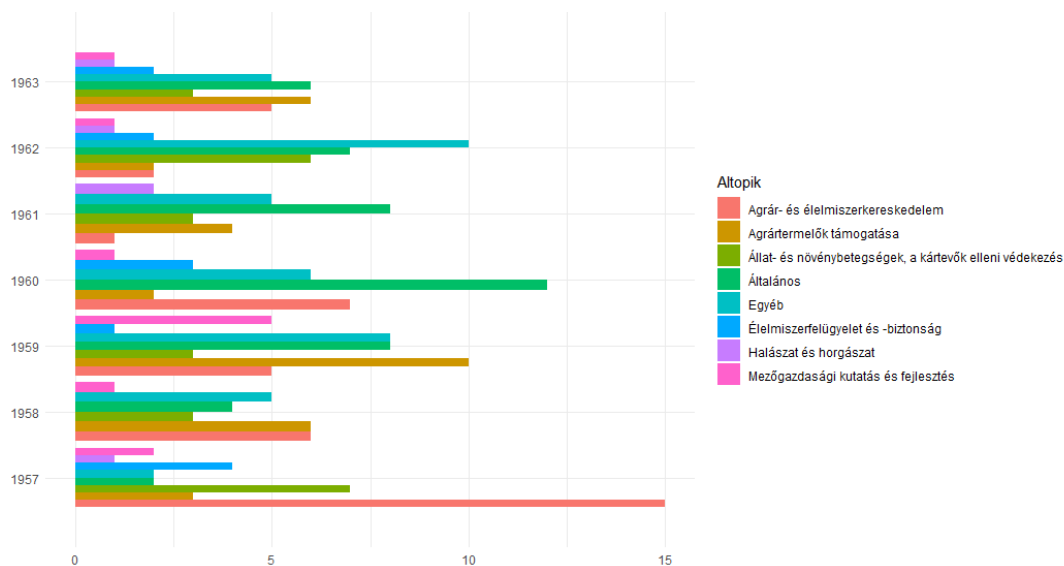
A teljes vizsgált korszakra vonatkozó összesített megoszlásoknál jobban mutatja a közpolitikai figyelem változását az egyes altémakörökbe sorolható elemek évenkénti előfordulása az agrárpolitikai jogszabályokon és napirendi pontokon belül.

A nyilvános jogszabályok, valamint a titkos minisztertanácsi határozatok adatsorai-ból kiválóan látszik, hogy az agrárkereskedelem szabályozása a korszak első éveiben adta a legtöbb munkát az államigazgatásnak – összhangban a korábbi történeti elemzés során említett ténnyel, amely szerint a begyűjtési rendszer kiiktatása következtében az igazgatás elsődleges feladatává vált az új árrendszer és az erre alapuló felvásárlási rendszer kialakítása. Az altémakör az egész vizsgált korszakban az agrárpolitika fontos területe maradt, az ide sorolható jogszabályok száma azonban az 1957-es és 1958-as csúcsevek után fokozatosan lecsökken.

A termelőszövetkezetek (és a termelők) megerősítésének politikai igénye is jól látszik az „agrártermelők támogatása” altémakörbe tartozó jogszabályok számának 1958–59-es megemelkedésében – a terület hangsúlyossá válása különösen a titkos minisztertanácsi határozatoknál szembeötlő, de 1959-ben a területhez kapcsolódó nyilvános határozatok száma is megnőtt. Ennél az altémakörnél a titkos határozatok és a kihirdetett jogszabályok egymásra épülésének sajátossága is megfigyelhető: 1958 elején a Minisztertanács a 3004/1958. A termelőszövetkezetek gazdasági megerősítését és fejlesztését szolgáló intézkedésekről titkos határozatban rendelte el a tsz-ek megerősítését szolgáló komplex (növénytermesztési és állattenyésztési fejlesztéseket, az értékesítés javítását, hitelelengedési kedvezményt, az állami szolgáltatások díjainak kedvezményét, szakemberek delegálását stb. tartalmazó) intézkedéssorozatot, egyben felhatalmazta, illetve megbízta az egyes szakterületi minisztereket a szükséges rendeletek, határozatok meghozatalával. A *Magyar Közlönyben* 1958 októberében jelent meg a titkos kormányhatározat kihirdetett végrehajtási utasítása, majd a következő év elejétől kerültek kihirdetésre a titkos határozatban előirányzott jogszabályok – többek közt a 6/1959. (II. 12.) FM-rendelet a termelőszövetkezetek gépi munkadíjkedvezményéről, a 13/1959. (III. 15.) *Kormányrendelet a mezőgazdasági termelőszövetkezetekbe lépő termelők szerződéses kötelezettségének rendezéséről*, vagy az 1010/1959. (IV. 3.) *Kormányhatározat a termelőszövetkezeti mozgalom fejlődésével kapcsolatos szakember-átcsoportosítás egyes kérdéseiről*.



7. ábra. A Magyar Közlönyben kihirdetett agrárpolitikai jogszabályok altopikok szerint, évenként (N=445)

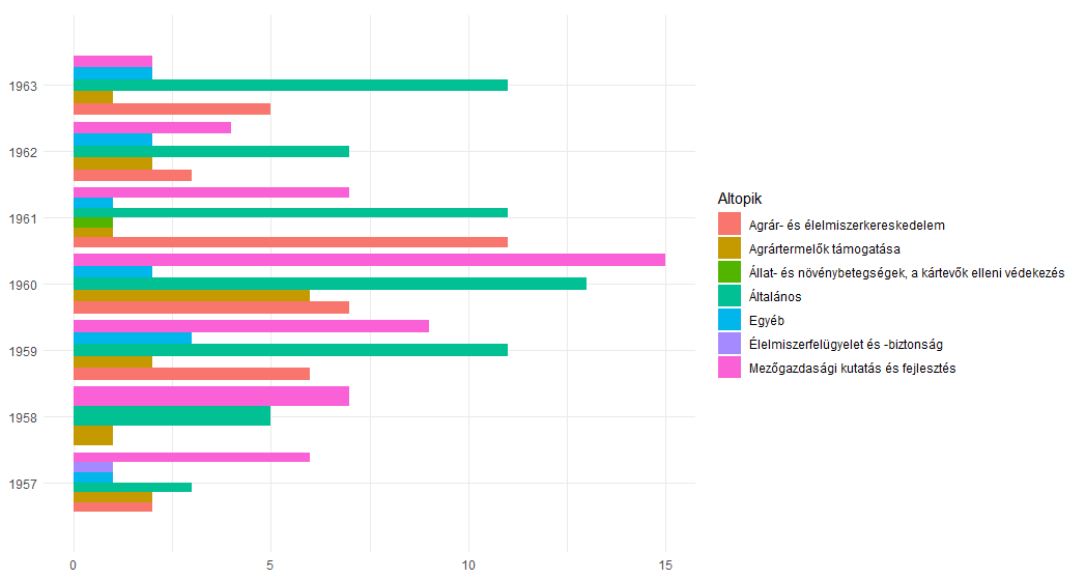


8. ábra. A Minisztertanács titkos agrárpolitikai hálózatai altopikok szerint, évenként (N=215)

Ha a Politikai Bizottság napirendi pontjainak altémakörök szerinti alakulását is megnézzük, sajátos fáziskésést figyelhetünk meg egyes nagy elemszámú topikok, így elsősorban az agrárkereskedelem témakör esetében. Az egyes konkrét előterjesztéseket megvizsgálva az adatokból a Politikai Bizottság speciális szerepét, ellenőrző, nyomon követő funkcióját mutathatjuk ki. A bizottság által kezdeményezett intézkedések mellett (pl. 1957. december 17. *Határozati javaslat a mezőgazdasági felvásárlási rendszer további javítására*; 1959. június 30. *Javaslat az 1959. évi bortermés elhelyezése*

érdekében szükséges intézkedésekre) nagy számban jelennek meg az ülések napirendjén olyan jelentések, beszámolók, amelyek a korábban jogszabályban vagy más utasításban elrendelt tevékenységek végrehajtásáról, a szabályozott folyamatok állásáról tájékoztatták a testületet (többek között 1960. szeptember 20. *Fehér Lajos tájékoztatója a burgonya felvásárlásával kapcsolatos intézkedésekről*; 1960. november 22. *Fehér Lajos tájékoztatója az őszi mezőgazdasági munkák és a felvásárlás helyzetéről*; 1961. október 17. *Jelentés a felvásárlás első háromnegyedévi tapasztalatairól*).

Míg az agrárkereskedelem esetében a Politikai Bizottság ellenőrző, „monitorozó” jellege tűnik erősebbnek, addig például a mezőgazdasági kutatás és fejlesztés területén egyértelműen kezdeményező szerepben van a testület – ez talán a terület nagyobb fokú generalitásával magyarázható. A fejlesztési irányvonalak általános („politikai szintű”) meghatározása, illetve az ezzel kapcsolatos javaslatok jóváhagyása egyértelműen a bizottság feladatkörébe tartozott.



9. ábra. Az MSZMP Politikai Bizottsága üléseinek agrárpolitikai napirendjei altópikok szerint, évenként (N=173)

4.3. Szövegbányászati vizsgálat

Az előzőekben bemutatott témaalapú osztályozás mellett, amikor is az egyes dokumentumokat kézi kódolás segítségével klasszifikáltuk, szövegeinket szövegbányászati módszerekkel is vizsgálat alá vettük. Ahhoz, hogy egy szöveghalmazt közvetlenül kvantitatív módon elemezzünk, szükséges annak adattá alakítása. Ezt követhetik a különböző szövegbányászati vizsgálatok, amelyek segítségével kvantitatív tartalomelemzést végezhetünk, megvizsgálva például a szövegekben előforduló különböző szavak, illetve témák gyakoriságát vagy együttes előfordulását. Mindezek révén lehetőségünk nyílt a szöveghalmaz belső struktúrájának feltárására.⁴⁰

⁴⁰ Németh, Katona és Kmetty, „Az automatizált szövegelemzés,” 55.

A következőkben *Natural Language Processing (NLP)* azaz a természetes nyelvfeldolgozás során alkalmazott egyik eljárást, a topikmodellezést hívtuk segítségül a mezőgazdasággal kapcsolatos szövegeink elemzéséhez. A természetes nyelvfeldolgozás során az alacsony mértékben strukturált szöveghalmazunkat kisebb egységekre bontottuk, és kapcsolatokat kerestünk a szöveg különböző elemei között. Ezzel az egyes szövegektől eltávolodva olyan ismeretek birtokába jutottunk, amelyek a szövegek egymás utáni olvasásával, illetve kvalitatív elemzéssel nem állnának rendelkezésünkre.⁴¹

Az elemzésbe a *Magyar Közlönyben* megjelent, a vizsgált korszakra vonatkozó és a CAP mezőgazdasági főtopikjába tartozó 445 darab jogszabály szövegét vontuk be. A szövegeket egyesével, jogszabályonként külön PDF-ben mentettük le az Optijus adatbázisából, majd ezeket az *ABBYY FineReader 14* optikai karakterfelismerő szoftver segítségével konvertáltuk TXT-formátumúvá, ügyelve arra, hogy a szövegek karakterkódolása UTF-8 legyen. Az *ABBYY FineReader* automatikusan állapítja meg a kódlapot, de manuálisan átállíthatjuk azt a kívánt kódtípusra. Erre azért szükséges figyelni, mert az UTF-8 kódolás egymagában képes az összes nyelv összes karakterét ábrázolni, így nem ütközhetünk abba a problémába, hogy mivel a program által automatikusan kiválasztott karakterkódolás nem képes a szövegben lévő egyes karakterek megjelenítésére, értelmetlen betűkapcsolatokat hozzon létre értelmes szöveg helyett.⁴²

Az így keletkezett TXT-eket egységes struktúrájú elnevezéssel láttuk el, ami tartalmazta az egyes jogszabályok CAP-kódolt metaadattáblában szereplő azonosítóját. Mindez a későbbi visszakereshetőség szempontjából nagyon fontos lépés. A következőkben ezen szövegeinket kellett adatokká alakítanunk, amihez az *R* szoftverkörnyezetet használtuk.⁴³

A szövegbányászati elemzések egyik igen fontos eleme az „előfeldolgozás” (*preprocessing*), ami több lépésből áll; ezek kiválasztása és elvégzésük sorrendje nagyrészt kutatói döntésen alapul. Mivel az előfeldolgozás folyamata során részben felesleges információktól szabadítjuk meg a szöveget, könnyen belátható, hogy az egyes lépéseket gondosan és a kutatási kérdésünknek megfelelően kell megterveznünk. Az előfeldolgozás első lépését még nem *R*-ben végeztük. A szöveget a *magyarlanc* morfológiai elemző segítségével szótöveztük, és meghatároztuk a benne szereplő szavak szófaját.⁴⁴ A szótövezés a statisztikai elemzések elvégzésének előfeltétele. A magyar nyelvben

⁴¹ Bolonyai Flóra és Sebők Miklós, „Kvantitatív szövegelemzés és szövegbányászat,” in *Empirikus jogi kutatások: Paradigmák, módszertan, alkalmazási területek*, szerk. Jakab András és Sebők Miklós (Budapest: Osiris Kiadó, 2020), 361–380, 364.

⁴² Hozzáférés: 2021.02.10, <http://www.cs.bme.hu/~egmont/utf8/>.

⁴³ Hozzáférés: 2021.02.10, <https://www.r-project.org/>. Az *R*-ben végzett szövegbányászati elemzések támogatására számos leírás érhető el, bár ezek használatát némiképp nehezíti, hogy gyakorlati példák angol nyelvű korpuszokra vonatkoznak, így a magyar nyelvű szövegek speciális elemzési kihívásaira nem reagálnak. A magyar nyelvű szövegek *R*-ben történő itt bemutatott elemzéséhez nagy segítségünkre voltak Labádi Gergely publikációi (hozzáférés: 2021.02.11, <http://labadiger.gy.github.io/>). Elemzésünk során az alábbi *R*-csomagokat használtuk: *tm*, *NLP*, *topicmodels*, *quanteda*, *tidytext*, *tidyr*, *ggplot*, *RTextTool*, *reshape2*, *dplyr*, *textmineR*, *LDavis*, *doParallel*, *scales*, *slam*.

⁴⁴ Hozzáférés: 2021.02.10, <https://github.com/zsibritajanos/magyarlanc>; János Zsibrita, Veronika Vincze and Richárd Farkas, „magyarlanc: A Toolkit for Morphological and Dependency Parsing of Hungarian,” in *International Conference Recent Advances in Natural Language Processing'2013: Proceedings*, eds. Ruslan Mitkov, Galia Angelova and Kalina Bontcheva (Shoumen: IN-COMA Ltd., 2013), 763–771.

a mondatokban a szavakhoz végződéses járulnak, illetve a szavaknak vannak elő- és utótagjai. Ezek eltávolításával (lemmatizálás) megkapjuk a szavak szótári alakját, azaz egységes alakra hozzuk az egy tőből különböző módon képzett vagy ragozott szavakat. A szófaj-meghatározás a *magyarlanc* futtatása során automatikusan megtörténik, így ha szükséges, később dönthetünk arról, ha az elemzésbe csak bizonyos szófajú szavakat szeretnénk bevonni. A természetes nyelvi feldolgozásban a szófaji egyértelműsítésnek gyakran fontos szerepe van, hiszen az ugyanolyan alakú szavak több jelentéssel is bírhatnak, ami egy adott aktuális kontextusban kulcsfontosságú lehet az elemzés szempontjából. A következő lépésben a korpuszt „stopszavaztuk”, azaz eltávolítottuk azokat a kifejezéseket, amelyek nem rendelkeznek tartalmi jelentéssel, ezek az úgynevezett funkciószavak (*function words*), például a kötőszavak.⁴⁵ Ehhez az *R* program *tm* programcsomagjában lévő, beépített magyar nyelvű stopszósótárát használtuk, amit a korpuszra jellemző stopszólistával egészítettük ki.⁴⁶ Az eddigiek eredményeként a korpuszunkban tehát a lemmatizált tartalomsszavak (*content words*) maradtak. Az előfeldolgozás következő lépése azon kifejezések kizárása az elemzésből, amelyek nagyon kevés, vagy nagyon sok dokumentumban szerepeltek. Például említhetjük a törvény, rendelet stb. szavakat, amelyek korpuszunkban értelemszerűen nagyon gyakoriak, de az elemzésbe nem tartottuk érdemesnek bevonni őket. Ehhez szógyakorisági listát készítettünk, ami a korpusz összes szavának gyakoriságát mutatta meg. Emellett készítettünk egy listát arról, hogy egy-egy adott kifejezés hány dokumentumban fordul elő. Végül az alsó eltávolítási küszöbértéket 0,1-ben a felsőt 0,4-ben határoztuk meg.⁴⁷

Ugyancsak az előkészítő műveletek közé tartozik a névelem-felismerés és a szignifikáns *n*-gramok megkeresése. Ezek olyan több szóból álló kifejezések, amelyeket az elemzés során együtt szeretnénk kezelni. Ugyancsak eltávolítottuk a korpuszból a központosítást és egyéb speciális karaktereket, a letöltés következtében a szövegekbe került url-címeket, elvégeztük a szövegek kisbetűsítését. Az előkészítő lépések eredményeként korpuszunk 600797 tokent tartalmazott.

A korpuszépítés során a szövegek adattá konvertálására egy dokumentumkifejezés-mátrixot (*document-term matrix, DTM*) hoztunk létre, amelyben minden dokumentum egy megfigyelés, azaz sor, és a dokumentumban szereplő minden kifejezés egy oszlop.⁴⁸

A dokumentum kifejezés-mátrix létrehozásával tulajdonképpen adattá alakítottuk szövegeinket. Ezután ugyancsak *R*-ben illesztettük az *LDA*-t, hogy ennek segítségével feltárjuk a mezőgazdasággal kapcsolatos jogszabályszovegek látens belső struktúráját.⁴⁹ Bár a topikok számának meghatározása kutatói döntésen alapul, és a modell

⁴⁵ Tikk, *Szövegbányászat*, 35; Burtejin, „Csoportosítás (klaszterezés),” 90.

⁴⁶ Hozzáférés: 2021.02.10, <https://cran.r-project.org/web/packages/tm/index.html>.

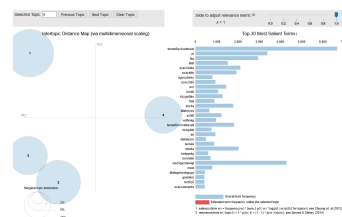
⁴⁷ Az előfeldolgozott szövegeket lásd a tanulmány mellékleteként (1030_data.zip): <http://ojs.elte.hu/digitalisbolcseszett/article/view/1030/1808>. A szövegkorpusz, a stopszavak, illetve a metaadatok elérhetők a GitHubon: https://github.com/ringorsolya/lda_historical_sources_hun.

⁴⁸ Hozzáférés: 2021.02.10, <https://www.rdocumentation.org/packages/tm/versions/0.7-7/topics/TermDocumentMatrix>.

⁴⁹ Hozzáférés: 2021.02.10, <https://www.rdocumentation.org/packages/lda/versions/1.4.2/topics/lda.collapsed.gibbs.sampler>.

futtatása során bevett gyakorlat a különböző topikértékekkel való kísérletezés, miután elkészült az elemzés több lehetőségünk is van az elkészült modell kiértékelésére. A zavarodottságfüggvény (*perplexity*) például a topikok által reprezentált elméleti szóeloszlásokat hasonlítja össze a szavak tényleges eloszlásával a dokumentumokban. Értéke nem önmagában értelmezendő, hanem két modell összehasonlításában, ahol a legalacsonyabb zavarodottságértékkel rendelkező modellt tekintik a legjobbnak. Elemzésünk során két *LDA*-módszert alkalmaztunk, az *LDA Gibbs*-et és a *VEM*-et, a validálás során ezek különböző topikszámokra adott zavarodottságeredményeit vetettük össze egymással. Az elemzés során 3 és 10 közötti topikszámokkal kísérleteztünk. A modell illesztése után minden esetben kiírtuk az egyes topikokba tartozó leggyakoribb 30 szót és megvizsgáltuk, hogy mely topikszám esetén kapunk jól interpretálhatóan egymástól elkülönülő topikokat. Az elemzésünket az *LDAvis* segítségével vizuálisan is megjelenítettük, ami nagymértékben segítette ezen eredmények értelmezését. Mindezzel négy egymástól jól elkülöníthető nagy témát sikerült azonosítanunk.

Hogy elemzésünket kvalitatív módon is kiértékeljük az *R* program segítségével azt is meghatároztuk, melyik jogszabály mely topikba tartozik inkább, majd ezt összevetettük a jogszabályok CAP-altopikba sorolásának eredményével is. Az összevetést interaktív HTML-es vizualizáció segíti.



A dinamikus HTML-tartalom megtekintéséhez kattintson a képre
<http://ojs.elte.hu/digitalisbolcseszlet/article/view/1030/1810>

Az első topikot elsődlegesen a mezőgazdasági termelés és termelési körülmények (a vetőmagellátás, szerződéses termelés, terményfelvásárlás) fogalmai határozzák meg. A második topik a domináns fogalmak alapján a termelőszövetkezeti működés, a harmadik topik a mezőgazdasági terményforgalom és ezen belül a jogszabályokban kiemelt figyelemmel övezett bortermelés és borkereskedelem, míg a negyedik az állat- és növény-egészségügyi védekezés témaköréhez kapcsolható.

Elemzésünk harmadik részében kísérletet tettünk a két eddig alkalmazott vizsgálati módszer, az előre megadott kategóriák mentén történő kézi osztályozás, valamint a szövegbányászat kombinálására. Célunk, hogy összefüggést találjunk a CAP-kódkönyvben meghatározott közpolitikai topikok, konkrétan az agrárpolitikai altopikok, valamint a topikmodell során kapott témák között.

A két eljárás összekapcsolása érdekében a jogszabályokat a fentiekben már jelzettek szerint hozzárendeltük a topikmodell-elemzés során azonosított négy topikhoz. Mivel a jogszabályok CAP-kódolása során a CAP-altopikba rendelést már elvégeztük, így az egyes jogszabályok esetében lehetségessé vált a CAP-altopik és a topikmodell szerinti besorolás együttes vizsgálata.

	Mezőgazdasági termelés	Termelősövetkezeti működés	Mezőgazdasági terményforgalom	Állat- és növényegészségügyi védekezés	Összesen
Általános	20,5%	9,0%	55,1%	15,4%	100,0%
Agrár- és élelmiszer-kereskedelem	35,4%	21,5%	20,8%	22,3%	100,0%
Agrártermelők támogatása	12,5%	16,7%	50,0%	20,8%	100,0%
Élelmiszer-felügyelet és -biztonság	16,3%	36,3%	13,8%	33,8%	100,0%
Állat- és növénybetegségek	20,9%	17,9%	20,9%	40,3%	100,0%
Halászat és horgászat	16,7%	25,0%	25,0%	33,3%	100,0%
Mezőgazdasági kutatás és fejlesztés	12,5%	12,5%	62,5%	12,5%	100,0%
Egyéb	18,2%	18,2%	31,8%	31,8%	100,0%
Összesen	22,9%	20,7%	30,1%	26,3%	100,0%

1. táblázat. CAP-altopik-besorolás és a topikmodell szerinti besorolás együttes vizsgálata

CAP-altopik	Elemszám	Százalék
Általános	78	17,5%
Agrár- és élelmiszer-kereskedelem	130	29,2%
Agrártermelők támogatása	48	10,8%
Élelmiszer-felügyelet és -biztonság	80	18,0%
Állat- és növénybetegségek	67	15,1%
Halászat és horgászat	12	2,7%
Mezőgazdasági kutatás és fejlesztés	8	1,8%
Egyéb	22	4,9%
Összesen	445	100,0%

2. táblázat. A vizsgált jogszabályok megoszlása CAP-altopikok szerint

LDA-topik	Elemszám	Százalék
1	102	22,9%
2	134	30,1%
3	92	20,7%
4	117	26,3%
Összesen	445	100,0%

3. táblázat. A vizsgált jogszabályok megoszlása a topikmodell egyes topikjai szerint

Látható, hogy míg a topikmodell-vizsgálat során négy, közel azonos elemszámú topikot kaptunk, addig a CAP-altopikok elemszáma nagyobb eltéréseket mutat. Kiemelkedő jelentőségű az „agrár- és élelmiszer-kereskedelem” altopik (a jogszabályok közel 30%-a tartozik ide), míg a „halászat és horgászat”, a „mezőgazdasági kutatás és fejlesztés”, valamint a nem egyértelműen besorolható elemeket tartalmazó „egyéb” altopikok együttesen sem teszik ki az összes vizsgált jogszabály 10%-át.

Már a két klaszterváltozó egyszerű keresztmegoszlása alapján is kirajzolódni látszik a CAP-altopikok és a topikmodell eredményei közötti összefüggés. Az „általános” elnevezésű altopikba sorolt jogszabályok 55%-a, az „agrártermelők támogatása” altopik jogszabályainak fele a második topikba került, hasonlóan a kis elemszámú „mezőgazdasági kutatás és fejlesztés” altopik elemeinek közel egyharmadához, míg az „állat- és növénybetegségek” altopik jogszabályainak 40%-a a negyedik topikba sorolódott.

Az „agrár- és élelmiszer-kereskedelem” altopik jogszabályainak 35,4%-a került az első topikba, míg az „élelmiszer-felügyelet és -biztonság” altopik esetében közel azonos, 33% illetve 36% a harmadik és a negyedik topikba sorolt jogszabályok aránya.

A két eljárás topikjainak eltérő elemszáma természetesen befolyással van a százalékos értékek alakulására, így a pontosabb eredmény miatt az adjusztált standardizált reziduálisok vizsgálatával cellánként is megkerestük a CAP-altopikok és a topikmodell topikjai közötti szignifikáns összefüggéseket. Az „általános” altopik jogszabályai a második topikba, az „agrár- és élelmiszer-kereskedelem” jogszabályai az első topikba sorolódtak a függetlenség esetén várható eloszláshoz képest szignifikánsan nagyobb arányban. Ezt az összefüggést mutatja az „agrártermelők támogatása” valamint a „mezőgazdasági kutatás és fejlesztés” altopik és a második topik, az „élelmiszer-felügyelet és -biztonság”, továbbá a harmadik topik, az „állat- és növénybetegségek” altopik és a negyedik topik is. A „halászat és horgászat” esetében nem találtunk elég erős összefüggést a topikmodell egyes topikjaival.

A magyarországi agrárpolitika történeti bemutatásánál ejtettünk szót a termelőszövetkezetek működési problémáiról, a „gyengén teljesítő” termelőszövetkezetek kérdéséről, ezek támogatásának agrárpolitikai vonatkozásairól. Ha az „agrártermelők támogatása” CAP-altopik és a topikmodell eredményeinek összefüggését nézzük, egyértelműen kirajzolódik az a tendencia, hogy vizsgált időszakunkban az agrártermelői támogatás elsődlegesen a tsz-ek működésének támogatásával volt azonos. Az „általános” altopik és a topikmodell „termelőszövetkezeti” topikjának erős kapcsolata jelzi az agrárpolitikai igazgatás fő irányát, az újonnan kialakított termelőegységek működésének, működési környezetének erőteljes szabályozását.

Kiemelendő, hogy a topikmodell-elemzés során „mezőgazdasági terményforgalom”-ként azonosított topik az „élelmiszer-felügyelet és -biztonság” CAP-altopikkal mutat

erős kapcsolatot, míg az „agrár- és élelmiszer-kereskedelem” altopik a „termelés és termelési körülmények” topikkal. A mezőgazdasági termékek forgalmazásánál a két eljárás összefüggése szerint elsősorban az élelmiszer-biztonsági, egészségügyi vonatkozások domináltak, míg a növénytermesztés általános szabályozása leginkább a termékek felvásárlásának, felvásárlási árának, a termelési körülményeknek, a termelőegységek azonosításának a kérdéskörével foglalkozott. Mindez egyértelműen összhangban van korábbi megállapításunkkal a mezőgazdasági termelési és termékkereskedelmi feltételrendszer kialakításának szükségességéről.

A nagy méretű szövegtörzsek elemzésére több lehetőség is rendelkezésünkre áll. Ilyen lehetőség a korpusz különböző szakaszainak címkékkel való ellátása, amire tanulmányunkban a CAP-kódolás alkalmazását hoztuk példaként. Ennek során deduktív módon, előre meghatározott kategóriákat használtunk és kézi kódolást végeztünk. Az ilyen típusú osztályozás előnye lehet, hogy a kutatási kérdéstől függően mélyebb elemzést tesz lehetővé, ellenben jelentős emberierőforrás-igénye van, képzett kutatók szükségesek hozzá, és még így is számolni kell az inkonzisztencia lehetőségével. Az osztályozás végezhető gépi támogatással is, vagy felügyelt gépi tanuló algoritmusok segítségével. Ezek jóval kisebb emberi erő bevonását teszik szükségessé, magát az elemzést felgyorsítják, de jelentős előkészítő munkát igényelnek, például tanulókorpusz létrehozását, eredmények validálását. Az ilyen előre meghatározott kategóriák előnye, hogy az egységesen kódolt korpuszok segítségével lehetőséget ad a folyamatok hosszú időintervallumban történő vizsgálatára, hátrányuk ellenben, hogy kevésbé igazíthatóak egy-egy korpusz egyedi jellemzőihez. Bár a CAP-kódrendszer nem minden esetben illeszthető könnyen a történeti kutatások szempontrendszeréhez, és az eltérő történelmi korszakok közpolitikájának tagolása, az egyes kategorizálandó elemek témakörökbe sorolása (különösen az altémakörök szintjén) is okozhat nehézségeket, alapvetően alkalmas arra, hogy a közpolitikai figyelem alakulásának történeti elemzéséhez felhasználjuk. Ezen altémakörök szintjén megjelenő kategorizálási nehézségekre jelenthet megoldást a dokumentumok klaszterezése.

Az esettanulmányunk során bemutatott felügyeletlen gépi tanulás körébe tartozó topikmodellezés ugyanis az adott korpusz jellegzetességeiből indul ki és segít eligazodni azokban a korpuszokban, amelyek tartalmáról olyan keveset tudunk, hogy máskülönben érdemi vizsgálatuk megkezdéséhez hosszas tanulmányozásukra lenne szükségünk. Emellett még az előkészítő lépésekkel együtt is lényegesen gyorsabb elemzést tesz lehetővé, mint a kézzel végzett osztályozás.

5. Összegzés és további kutatási irányok

Tanulmányunkban a magyarországi kollektivizálás második hullámát egy, a társadalomtudományban használatos módszer adaptálásának segítségével vizsgáltuk meg. Bemutattuk, hogy a nagy adatbázisok elemzése milyen nagyobb, általánosabb összefüggések levonására kínál lehetőségeket. Ugyanezt a kutatási kérdéskört kvantitatív szövegelemzési módszertannal is megvizsgáltuk, az egyik legelterjedtebb szöveg-bányászati módszer, az úgynevezett csoportosítás vagy klaszterezés, és azon belül a topikmodellezés alkalmazásával. Célunk annak illusztrálása volt, hogyan lehet új

elemzési eszközökkel, más tudományágak által alkalmazott modellek segítségével tovább mélyíteni egy történelmi folyamatról korábban megalkotott képet.

Egyetértve Z. Karvalics Lászlóval, úgy véljük, hogy önmagában sem a *big data* jelenség és annak a történelmi kutatásokra gyakorolt hatása, így a nagy digitális gyűjtemények létrejötte és folyamatos gyarapodása, sem azok elemzésének új módszerei nem hoznak ugyan paradigmátikus változást a történelmi megismerés módjában, viszont számtalan új módszertani lehetőségnek nyitnak utat.⁵⁰ Ezek az új módszerek azonban nem helyettesítik a hagyományos kutatási formákat és módszereket, inkább kiegészítik, támogatják azokat. A kvantitatív és kvalitatív források és módszertanok mellett a szövegelemzés kvantitatív és kvalitatív formái is kiegészítik egymást. De a választott módszertannak mindig a kutatói döntésen és a kutatási kérdéseken, valamint a rendelkezésre álló forrásokon kell alapulnia.

Az egyre gyarapodó digitalizált dokumentumállomány ellenére a szövegbányászati módszerek széles körű történettudományi alkalmazása előtt jelenleg nagy gátat jelent a jó minőségű, nagy méretű korpuszok és az egységesen metaadatolt adatbázisok hiánya. Mint rámutattunk, önmagában a digitális állományok megléte nem teszi azokat szövegbányászati vizsgálatra alkalmassá, az egyik legnagyobb leküzdendő akadályt a források szöveggé alakítása előtt az optikai karakterfelismerés hiányosságai jelentik. A papíron keletkezett dokumentumok sokszor csak hosszadalmas manuális tisztítás után válnak szövegbányászati módszerekkel elemezhetővé.

Ahogy arra korábbi tanulmányában Péter Róbert is rámutatott, olyan kifinomult kutatási eszközökre és szoftverekre is szükség van, amelyek úgy képesek megbirkózni hatalmas mennyiségű szöveggel és metaadattal, hogy az elemzésnél a bölcsészettudományi kutatásban hangsúlyos részletek sem vesznek el.⁵¹ Ehhez nagy segítséget jelenthet a különböző tudományterületek összefogása, az interdiszciplináris kutatócsoportok működése. Mindamellet azon a véleményen vagyunk, hogy megalapozott történettudományi kutatás digitális bölcsészeti módszerekkel csak úgy végezhető, ha megfelelő képzés is járul hozzá, azaz a módszer iránt érdeklődő történészek számára lehetőség nyílik a bölcsészettudományi kutatások kérdésfeltevéseihez és módszertanához igazodó digitális bölcsészeti képzettség megszerzésére, ami egyben azt is lehetővé tenné, hogy szakmai diskurzus induljon e módszerek használatáról, amelyek történettudományi kérdésfeltevés és releváns források ismerete mellett azonban kiválóan egészíthetik ki a hagyományos kutatási módszerek alkalmazását.

⁵⁰ Z. Karvalics László, „A nagy adat és digitális történelem: Egy izgalmas házasság múltja, jelene és jövője,” *Magyar Tudomány* 179, 5. sz. (2018): 670–671, <https://doi.org/10.1556/2065.179.2018.5.10>.

⁵¹ Péter, „A Big Data kihívás,” 1328–1329.

Agricultural Policy Challenges and Legislation in the Early Kádár Era: Text Analysis and Text Mining of Historical Sources

In our study, we examine the collectivization in Hungary by adapting a method used in political science and analyse large databases. In addition we examine the same question using text mining method. The aim of our paper is to illustrate the possibilities and limitations of the analysis of digital historical collections with new research methods. Our analysis examines the agricultural legislation of the early Kádár era. In addition to the classification, we also examine our texts using text mining methods, and in the third part of our analysis we try to combine the two methods – to find a correlation between agricultural policy subtopics and topics obtained in the topic model.

Keywords:

digital humanities, history, classification, text mining, topic modelling

<MŰHELY>

