

Budapesti Corvinus Egyetem
Tudományos Diákköri Konferencia

Magyar pártos polarizáció elemzése gépi tanulási módszerrel

Konzulens: dr. Simonovits Gábor

Szerzők: Novák Levente Péter (Közgazdasági elemző I.)

Kaló Eszter (Közgazdálkodás és közpolitika II.)

2025 tavasz

Absztrakt

A dolgozat célja a magyarországi politikai polarizáció, azon belül a pártos polarizáció vizsgálata az 1998 és 2018 közötti időszakban. A kutatás természetes nyelvfeldolgozáson alapuló gépi tanulási módszert alkalmaz az MSZP és a Fidesz képviselőinek parlamenti felszólalásaira. Az elemzés kiterjed a teljes szövegtre, valamint külön a kulturális és gazdasági tartalmú felszólalásokra is. A tematikus besorolás a poltextLAB Babel machine szolgáltatásának segítségével történt. Az eredmények alapján a kulturális témák esetében a pártos polarizáció növekvő trendet mutatott a 2010–2014-es ciklusban, ami tovább erősödött a 2014-2018-as időszakban. A gazdasági témák mentén jelentkező megosztottság ezzel szemben 1998 és 2014 között stabilan alacsony mértékű volt és csak a későbbi ciklusban mutatott hangsúlyos emelkedést. Az empirikus elemzés továbbá arra is rámutat, hogy az MSZP a vizsgált időszakban jóval kevésbé volt képes saját narratívák meghatározására a parlamenti diskurzusban, mint a Fidesz.

Tartalomjegyzék

I. Bevezetés, relevancia	1
II. Elméleti keret.....	2
II. 1. A politikai polarizáció	2
II. 2. Kulturális polarizáció és gazdasági konvergencia a kínálati oldalon (elméleti predikciók)	4
II. 3. A politikai polarizáció operacionalizálása természetes nyelvfeldolgozásra épülő gépi tanulási módszerrel	6
III. Elemzés előkészítése	8
III. 1. Adatbázis forrása és adatok előfeldolgozása	8
III. 2. Kategorizálás	10
IV. Eredmények.....	11
V. Konklúzió	20
Hivatkozásjegyzék	23
Melléklet.....	26
Melléklet I.	26
Melléklet II.	26

Ábrajegyzék

1. ábra: Pártos polarizáció alakulása a teljes korpuszon (saját ábra)	13
2. ábra: Pártos polarizáció alakulása a kulturális korpuszon (saját ábra)	14
3. ábra: Pártos polarizáció alakulása a gazdasági korpuszon (saját ábra).....	14

1. táblázat: Irodalom összehasonlítás a polarizáció dimenzióiban (saját táblázat)	3
2. táblázat: Választási eredmények (1998-2014), forrás: valasztas.hu (Saját szerkesztés)	6
3. táblázat: Ciklusonkénti leíró statisztikák (saját táblázat a teljes korpuszon).....	11
4. táblázat: Ciklusonkénti leíró statisztikák (saját táblázat a kulturális korpuszon).....	12
5. táblázat: Ciklusonkénti leíró statisztikák (saját táblázat a gazdasági korpuszon)	12
6. táblázat: Top 10 magyarázóerővel bíró bigram pártonként, ciklusonként (1998-2018) (saját táblázat a teljes korpuszon)	15
7. táblázat: Top 10 magyarázóerővel bíró bigram pártonként, ciklusonként (1998-2018) (saját táblázat a gazdasági korpuszon).....	17
8. táblázat: Top 10 magyarázóerővel bíró bigram pártonként, ciklusonként (1998-2018) (saját táblázat a kulturális korpuszon)	19
9. táblázat: Nem nulla magyarázóerővel bíró bigramok száma (saját táblázat)	20

I. Bevezetés, relevancia

A politikai polarizáció napjaink egyik legmeghatározóbb társadalmi és politikatudományi jelensége, fogalma. A közbeszédben és a tudományos diskurzusban egyaránt egyre gyakrabban jelenik meg a társadalmi megosztottság, az egymással párhuzamos valóságokat élő választói csoportok problémája. A szavazók nem csupán eltérő vélemények mentén szerveződnek, hanem az identitásalapú törésvonalak mentén egyre erőteljesebben el is zárkoznak egymástól. A közösségi média platformjai, az algoritmus vezérelt tartalomszűrésük miatt, tovább erősítik ezt a széttagoltságot, a különböző véleménybuborékokban élő felhasználók nem pusztán mást gondolnak a világról, hanem elképzelhető, hogy másként is látják azt.

Ebben a kialakult, fragmentált közegben a politikai aktorok felelőssége különösen jelentőssé válik. A szavazók és a politikai szereplők helyzete elképzelhető egy kereslet-kínálat típusú interakcióként (Gennaioli és Tabellini, 2023), ahol a politikai elit diskurzusa, ideológiai kínálata és retorikája nem csupán reagál a választói igényekre, hanem alakítja is azokat – a polarizációs folyamat aktív szereplőiként. A modern politikai kommunikáció keretrendszerében a parlamenti felszólalások továbbra is a politikai nyilvánosság meghatározó részét képezik. Bár közvetlen fogyasztásuk mértéke csökkenni látszik, ezek a beszédek továbbra is gyakran szolgálnak alapul későbbi médiatartalmakhoz, kampányüzenetekhez, illetve a pártok narratíváinak felépítéséhez.

Elemzésünk a magyar parlament 1998 és 2018 közötti időszakának két domináns politikai szereplője, a Fidesz és az MSZP felszólalásaira fókuszál. A kulturális törésvonalak, mint a migráció, nemzeti identitás vagy vallás előtérbe kerülése, valamint a gazdasági kérdésekben mutatkozó retorikai konvergencia egyaránt utalnak arra, hogy a politikai verseny a kínálati oldalon is jelentős átalakuláson ment keresztül (Gennaioli és Tabellini, 2023; Krekó & Enyedi, 2018). Mindez különösen fontos kontextust biztosít annak megértéséhez, hogyan alakulhatott ki egy olyan magyar politikai rendszer, amelyben a Fidesz a nemzeti szuverenitás és a kulturális homogenitás védelmére építi diskurzusát, míg az MSZP, bár ezen párt narratívája kevésbé meghatározott, főként ügyek mentén domborodik ki, megértésünkben továbbra is inkább a szolidaritás nyelvén szólal meg (Bíró-Nagy, 2017; Hargitai, 2002).

Célunk nem elsősorban a konkrét szakpolitikai álláspontok különbségeinek feltárása, hanem annak vizsgálata, hogy a parlamenti felszólalások hogyan strukturálódnak ideológiai és retorikai dimenziók mentén, a gazdasági és kulturális polarizáció kontextusában. Továbbá célunk megvizsgálni, hogy a parlamenti felszólalásokban megjelenik-e a politikai polarizáció,

empirikus elemzés által megfigyelhető-e az a jelenség, amit a média, közvélemény és az elméleti politikatudomány írásai feltételeznek.

A pártos polarizációt Gentzkow és sztsai (2019) gépi tanulási módszere alapján úgy operacionalizáltuk, hogy a magyar parlamenti felszólalásokon egyetlen bigramja alapján az a megfigyelő, aki ismeri a modellt, milyen helyesen tudja prediktálni azt, hogy az adott képviselő a Fidesz vagy az MSZP politikusa. Gennaioli és Tabellini (2023) keretrendszerét követve a teljes korpuszunkat kulturális és gazdasági alkorpuszokra is felbontottuk. A kapott eredmények alapján a kulturális témákban egy ciklussal korábban, a 2010-14-es ciklustól kezdve már jelen volt növekvő mértékű politikai polarizáltság. A gazdasági témák mentén 1998 és 2014 között stabilan alacsony szintű politikai polarizáció volt jellemző, amely a 2014–2018-as ciklusban emelkedni kezdett. Továbbá kiemelendő empirikusan alátámasztott eredményünk, hogy az MSZP jelentősen kevesebb témát tudott tematizálni a magyar országgyűlésben a vizsgált két évtized bármelyik ciklusában, mint a Fidesz.

Dolgozatunk a következőképpen épül fel. A következő fejezetben ismertetjük az elemzés elméleti keretét, valamint az ebből levezetett predikciókat és a politikai tartalom operacionalizálásához alkalmazott módszertant. Ezután bemutatjuk az elemzés előkészítésének lépéseit, majd részletesen tárgyaljuk az empirikus eredményeket. A dolgozat zárásaként összefoglaljuk a legfontosabb megállapításokat.

II. Elméleti keret

A fejezet célja egyrészt átfogóan bemutatni a polarizáció fogalmának különböző értelmezéseit és kutatási irányait, másrészt megalapozni azt az elméleti háttérrel, amelyre a dolgozat empirikus elemzése támaszkodik. A fejezet három nagy alfejezetre tagolódik, az első áttekinti a fogalom elméleti alapjait, dimenzióit és főbb kutatási irányait, a második bemutatja, hogy a kulturális identitás elsődleges törésvonallá válik a gazdasági preferenciák helyett, a harmadik pedig bemutatja a dolgozatban használt természetes nyelvfeldolgozási módszertant, amivel operacionalizáljuk a polarizációt.

II. 1. A politikai polarizáció

A politikai polarizáció a politikatudomány egyik legszéleskörűbben vizsgált jelensége, azonban többnyire elméleti fókuszban esik szó róla, a szubjektivitásnak nagyobb teret engedve ezzel (pl. Palonen 2009). Jelen fejezet a polarizáció témájában készít irodalomösszehasonlítást és három fő tengely mentén mozog. Elsőként a fogalmi alapokat igyekszik meghatározni, ezt

követően a polarizáció dimenzióival és típusaival foglalkozik, valamint módszertani megközelítéseket térképez fel, különösen a szöveg, mint adat használatát politikai kommunikációban.

Míg a szakirodalom egy része a társadalmi polarizációra, a szavazók befolyásolására összpontosít, felmerül a kínálati oldal, az hogy a politikai elit, a közbeszéd, a különböző szereplők hogyan konstruálják és erősítik a megosztottságot a közéleti diskurzus különböző formáin keresztül. Higley és társai (2002) már a fogalom bevezetésekor a politikai elit oldaláról megközelítve definiálják a politikai polarizációt.

Amerikai, empirikus tanulmányok, mint például McCarty, Poole és Rosenthal (2006) könyve, főként az amerikai kongresszuson belüli, a politikai elit által képviselt ideológiák versenyére összpontosítottak a pártok programjai közötti növekvő eltérésekre fókuszálva. Megalkottak módszereket az ideológiai távolságok mérésére, és ezzel az ideológiai polarizációt mérhető jelenséggé határozták meg. Emellett Iyengar, Sood és Lelkes (2012) tanulmánya az affektív polarizációval foglalkozik, a pártosodás érzelmi összetevőjére, ami a csoporthoz tartozás érzetében, bizalmatlanságban és az erkölcsi fölény érzésében, ez általi különbségalkotásban nyilvánul meg. Bár a definíciók eltérőek, egyetértés mutatkozik abban, hogy ideológiai eltérést és affektív távolságtartást egyaránt magában foglal, nem csupán a politikai kérdésekben felmerülő nézetkülönbségeket, hanem a szembenálló politikai oldal, az ellenzék ellenséges és moralizált megítélésének kialakulását is (Iyengar, Sood és Lelkes, 2012).

Dimenzió	Fókusz	Példa
Tömeg vagy elit	Ki polarizálódik, a közvélemény vagy a politikai elit?	McCarty et al. (2006)
Ideológia vagy identitás	A különbségek a politikákról vagy az identitásokról/érzelmekről szólnak?	Iyengar et al. (2012)
Probléma- vagy identitás fókusz	Tartalmiak vagy értékbeliek a törésvonalak?	Dias & Lelkes (2021)
Vélt vagy valós	Túlbecsülik az emberek a polarizációt?	Levendusky & Malhotra (2016)

1. táblázat: Irodalom összehasonlítás a polarizáció dimenzióiban (saját táblázat)

A kortárs szakirodalom nagy része a dimenziók közötti kapcsolatokat, hatásokat vizsgálja, például, hogy hogyan erősíti az elit retorikája, ami a kínálati oldalt jelenti, a választók identitását, ami pedig a keresletit, növelve a polarizációt.

Az irodalomelemzés egyik legfontosabb felismerése, hogy a polarizáció mérésének módja alapvetően befolyásolhatja a következtetéseket. Kérdőíves és felmérés alapú kutatások főként a választói attitűdök és az érzelmi reakciók tendenciáit vizsgálják. Ezeknek az erőssége lehet, hogy hosszú távon is követhetők, lehetőséget adnak akár hatáskövetésre is, valamint standardizált összehasonlításokra, míg gyengeség lehet a válaszadási hajlandóság okozta torzítás a mintában és nincs jelen diszkurzív aspektus. A viselkedésalapú és kísérleti megközelítések közé tartoznak többek között az előítéleteket mérő tesztek, valamint a pártpreferencia szerinti besorolást vizsgáló feladatok (Iyengar és Westwood, 2014). Gyakran az eredmények nem általánosíthatóak nagyobb, vagy más társadalmi környezetre, erősen kontextusfüggőek lehetnek. Szavazási eredmények és pártprogramok elemzésének úttörői közé sorolhatjuk McCarty et al. (2006) (DW-NOMINATE modell), ami az elit szintű ideológiai polarizáció kutatásának fő módszere volt, azóta szélesebb körben is elterjedt, pártprogramok, manifesztumok elemzésén keresztül.

A magyarországi politikai polarizációval kapcsolatos kutatások az utóbbi években lendületet és teret nyertek, tükrözve az ország egyre inkább kettéágazó politikai tájképét, a változó, de hosszú ideig stabil politikai erőket. Bíró-Nagy (2017) leírja, hogy a Fidesz hogyan használta az egyre nacionalistább és populistább retorikát a kulturális szakadék elmélyítésére, különösen 2010 után. Ezzel szemben az MSZP továbbra is a társadalmi és gazdasági igazságosságot helyezte előtérbe, politikai diskurzusát a befogadás és a szolidaritás köré keretezve (Krekó & Enyedi, 2018). Az elmúlt években a politikai diskurzusban egyre inkább előtérbe kerültek a kulturális narratívák, különösen olyan témák mentén, mint az identitáspolitiká, a nemzeti szuverenitás, a bevándorlás és a nemzeti identitás. Ezeket a narratívákat elsősorban a Fidesz határozza meg és irányítja (Bíró-Nagy, 2017). A kormány sikeresen keretezte például a bevándorlással és a szuverenitással kapcsolatos vitákat a nemzeti túlélés szempontjából, ellenfeleit „az állam ellenségeinek” állítva be (Bajomi-Lázár, 2015).

II. 2. Kulturális polarizáció és gazdasági konvergencia a kínálati oldalon (elméleti predikciók)

Az amerikai politikában a 2010-es évek elejétől kezdve megfigyelhető egy átalakulás abban, hogy a választók a korábbi gazdasági újraelosztási kérdések helyett inkább kulturális témákra, mint a bevándorlás, vallás vagy emberi jogok, helyeznek nagyobb hangsúlyt politikai véleményük kialakításakor (Bonomi et al. 2021). Hasonló tendenciák figyelhetők meg a magyar belpolitikában is, ahol egyre nagyobb figyelmet kapnak a nem-gazdasági jellegű törésvonalak (Bíró-Nagy, 2017).

Ennek következményeként a hagyományos, csupán közgazdaságtani racionalitáson alapuló modellek nem képesek kielégítően megmagyarázni, hogy miért kezdtek el az alacsonyabb iskolázottságú és alacsony jövedelmű szavazók egyre nagyobb arányban olyan politikai erőket támogatni – például a republikánus pártot az Egyesült Államokban –, amelyek kisebb állami újraelosztást hirdetnek. Ezzel párhuzamosan a magasabb iskolázottságú és magasabb jövedelmű választók egyre inkább olyan pártok felé mozdulnak, amelyek nagyobb szerepet szánnak az állami újraelosztásnak (Gethin et al., 2022; Gennaioli és Tabellini, 2023).

Bódi és Kovalcsik (2023) elemzése alapján Magyarországon is megfigyelhető egy, a nemzetközi trendekhez illeszkedő átalakulás. Eredményeik szerint a Fidesz szavazótáborának társadalmi összetétele 2002 és a 2018-as választások között alapvetően stabil maradt. Ugyanakkor 2014-től kezdődően egyre inkább megfigyelhető, hogy a legmagasabb iskolai végzettséggel rendelkező választók a baloldali pártok felé fordulnak. Ezzel párhuzamosan a 2018-as és 2022-es választások során a Fidesz jelentős támogatottság-növekedést könyvelhetett el az alacsonyabb iskolai végzettséggel és szerényebb jövedelemmel rendelkező választók körében. (Bódi és Kovalcsik, 2023)

Gennaioli és Tabellini (2023) egy formális politikai gazdaságtani modellt dolgoznak ki, amely a társadalmi identitás szerepén keresztül magyarázza, hogyan torzulhatnak a szavazók gazdaságpolitikai preferenciái. A modell szerint, ha a választók erős identitásalapú kötődést alakítanak ki bizonyos kulturális témák mentén, akkor ez a kötődés képes felülmúlni a klasszikus értelemben vett gazdasági érdekeik mentén való politikai viselkedést. Így válik érthetővé, hogy miért támogathatják az alsóbb jövedelmi decilisekbe tartozók az érdekeikkel ellentétesen kisebb újraelosztást, míg a felsőbb decilisekbe tartozók akár nagyobb mértékű állami beavatkozást is. Az általuk felvázolt modell képes megmagyarázni olyan stilizált tényeket is, amikre korábbi kutatások modellezései nem adtak megfelelő választ, illetve empirikus tesztek alapján a modell felépítéséből következő predikciók is igaznak bizonyulnak amerikai kontextust tekintve.

Mivel Magyarországon is egyre inkább megfigyelhető, hogy a politikai közbeszédben a kulturális kérdések – például nemzeti identitás, migráció vagy vallás – központi szerepet kapnak, valamint ezzel párhuzamosan nő a politikai polarizáció mértéke is (Krekó & Enyedi, 2018), különösen indokoltá válik annak vizsgálata, hogy Gennaioli és Tabellini (2023) modellje milyen mértékben alkalmazható a hazai kontextusban. A modell által felvetett elméleti predikciók tesztelése segíthet mélyebb megértést nyerni arról, hogyan és miért alakult ki a

jelenlegi politikai törésvonalak rendszere Magyarországon, és miként strukturálódnak a pártok közötti verseny kulturális és gazdasági dimenziók mentén.

A dolgozatban Gennaioli és Tabellini (2023) két predikcióját vizsgáljuk magyar kontextusban. Modelljük alapján, mivel az elmúlt évtizedben a kulturális identitás dimenziója vált meghatározóbbá a gazdasági kérdésekkel szemben, azt feltételezzük, hogy a pártok politikai álláspontjai kulturális kérdésekben egyre inkább eltávolodnak egymástól, míg gazdasági (újraelosztási) kérdésekben közeledés figyelhető meg.

1998 és 2018 között a magyar politikai életet jelentős mértékben a Fidesz és az MSZP dominanciája jellemezte. Bár ez a két párt nem érte el az amerikai kétpártrendszer – a Republikánusok és a Demokraták közötti – versengés intenzitását, mégis számottevően megközelítette azt. Az öt, ezen időszakban megrendezett parlamenti választás során a Fidesz és az MSZP együttes társadalmi támogatottsága rendre meghaladta a 62%-ot (lásd 2. táblázat), ami jól mutatja a két párt központi szerepét. E dominancia indokoltá teszi, hogy az amerikai kétpártrendszer dinamikájára megalkotott formális modellt a két magyar párt viszonyrendszerében teszteljük.

Párt/Választási év	1998	2002	2006	2010	2014
Fidesz (%)	29,5	41,1	42	52,7	44,9
MSZP (%)	32,9	42,1	43,2	19,3	25,6

2. táblázat: Választási eredmények (1998-2014), forrás: valasztas.hu (Saját szerkesztés)

II. 3. A politikai polarizáció operacionalizálása természetes nyelvfeldolgozásra épülő gépi tanulási módszerrel

A természetes nyelvfeldolgozás (NLP) területén a politikai polarizáció operacionalizálásának egyik elterjedt megközelítése, hogy a politikusok vagy választók által elmondott beszédek, illetve az általuk írt bejegyzéseket klasszifikációs problémaként értelmezik (Németh, 2023). Ha az alkalmazott osztályozó modell magas pontossággal képes megkülönböztetni az eltérő csoportokhoz (például politikai pártokhoz vagy szavazótáborokhoz) tartozó megszólalásokat, az a polarizáció magas fokára utal. Ennek oka, hogy az egyes csoportok nyelvhasználata jól elkülönül. Ezzel szemben, ha az osztályozás pontossága alacsony, az arra utal, hogy a csoportok hasonló témákról és hasonló nyelvezetben nyilvánulnak meg, így a modell nem tudja őket megbízhatóan szétválasztani (Németh, 2023). Ezt az operacionalizálási elvet alkalmazza többek között Goet, (2017); Bayram és sztsai (2019), valamint Taddy (2013a) és (2015) és Gentzkow és sztsai (2019).

A dolgozatban Gentzkow és szerzőtársai (2019) gépi tanulási megközelítését alkalmazzuk a Gennaioli és Tabellini (2023) által megfogalmazott elméleti predikciók magyar parlamenti beszédeken való teszteléséhez. Gentzkow és szerzőtársai (2019) kritikával illetik a korábbi, hasonló logikára épülő módszereket, mivel azok nem kezelik megfelelően a végeselem-minta torzítást (finite sample bias). Ez a probléma abból adódik, hogy például egy parlamenti beszéd során a felszólaló több ezer lehetséges kifejezés közül választhat, de valójában csak néhány száz szót használ. Így előfordulhat, hogy bizonyos kifejezések csupán véletlenszerűen jelennek meg kizárólag az egyik párt beszédeiben, miközben nem tükröznék valódi pártos polarizációt.

Ennek kezelésére Gentzkow és szerzőtársai (2019) regularizált osztályozó modellt alkalmaznak, amely korlátozza a modell komplexitását, és elkerüli, hogy az algoritmus véletlen zajokra tanuljon rá, például olyan kifejezésekre, amelyeket az egyik párt véletlenszerűen használ, míg a másik nem. Így csökkenthető a túlillesztés (overfitting) veszélye, és megbízhatóbb polarizációs becslés érhető el.

Gyakorlatban egy lasszó-típusú regularizációt használnak modell paramétereikhez, amivel lehetőség nyílik a végeselem-minta torzításokra kontrollálásra vagy legalábbis azok csökkentésére. Intuitívan minden egyes bigram (két szóból álló szókapcsolat) magyarázóerővel bír annak vonatkozásában, hogy mennyire utal egy felszólaló pártjára (frakciójára). Azoknak a bigramoknak van nullától eltérő magyarázóereje (koefficiense), amelyeket az egyik párt képviselői jelentősen többször használtak, mint a másik párté. A bigramok koefficiensének számít az iránya és nagysága. Pozitív irányú koefficiensok fideszes, negatív irányú koefficiensok MSZP-s felszólalót sejtetnek. Minél nagyobb a koefficiens abszolút értéke, annál biztosabban utal az adott képviselő frakciójára. Példának okáért, ha egy szókapcsolatot rendszeresen használnak a Fidesz képviselői egy adott ciklusban, de az MSZP képviselői nem (pl. migrációs válság), akkor annak magas pozitív koefficiense lesz, míg, ha egy kifejezést egy ciklusban úgy használnak gyakran az MSZP képviselői, hogy közben a Fidesz képviselői nem, akkor annak a bigramnak magas, negatív értékű koefficiense lesz (pl. magyar felsőoktatás). A regularizációs tagon keresztüli koefficiens nullára csökkentésével nyílik lehetőség kontrollálni arra, hogyha egy-egy párt képviselői a pusztán véletlen okán használnának többször egy-egy szófordulatot, mint a másik párt képviselői.

A dolgozatban tehát a politikai polarizációt Gentzkow és szerzőtársai (2019) nyomán úgy operacionalizáljuk, hogy megbecsüljük, egy megfigyelő, aki ismeri a modellt, milyen pontossággal tudná megállapítani egy parlamenti felszólaló párthovatartozását (Fidesz vagy

MSZP) pusztán egyetlen bigram alapján. Ugyanezt a módszertani megközelítést alkalmazza például Green és sztsai (2020) amerikai politikusok Covid-19-cel kapcsolatos tweetjein, Fiva és sztsai (2021) norvég parlamenti beszédeken, illetve Simola és sztsai (2023) finn parlamenti adatokon. A modell szempontjából nem jelent torzítást az, hogyha valamelyik párt képviselői összesen több felszólalást mondtak el egy adott ciklusban, hiszen a megfigyelő neutrális előzetes várakozásokkal becsüli meg a felszólaló pártját (Gentzkow és sztsai, 2019). Fontos kiemelni, hogy a modell magyarázóereje jelentősen eltér attól a predikciótól, amit egy statisztikus vagy egy ökonóméter tudna becsülni egy felszólalás alapján az illető frakciójáról, ugyanakkor a pusztán statisztikai alapú becslés nem lenne mentes a végelem-minta torzítástól (Gentzkow és sztsai, 2019).

Magyar kontextusban ismereteink szerint eddig csak Buda és Németh (2024) publikált olyan tanulmányt, amely hasonló elvre építve vizsgálta a szövegalapú politikai polarizációt. Ugyanakkor az ő céljuk elsősorban a felügyelt gépi tanulás alkalmazhatóságának bemutatása volt egy esettanulmányon keresztül. Ezt támasztja alá például az, hogy modelljük érzékenyen reagált az ellenzéki-kormánypárti szerepkörökre jellemző mintázatokra, nem pedig kifejezetten a nyelvi polarizációra tanult rá. Emellett a tanulmányban nem szerepelnek olyan validációs eljárások, amelyek az eredmények érvényességét erősítenék. Fontosnak tartjuk azonban hangsúlyozni, hogy tanulmányuk értékes inspirációként szolgált a jelen dolgozat számára.

III. Elemzés előkészítése

A következő fejezet a dolgozat empirikus elemzésének módszertani alapjait mutatja be. Az első alfejezet az alkalmazott szövegtörzset, tehát a parlamenti felszólalások forrását, előkészítését és szűrését ismerteti, valamint bemutatja a bigramokra épülő szövegelemzés lépéseit. A második alfejezet a témakódolás, kategorizálás módszertanára, a felszólalások kulturális vagy gazdasági törésvonalakhoz történő hozzárendelésére fókuszál.

III. 1. Adatbázis forrása és adatok előfeldolgozása

A dolgozathoz használt adatbázis parlamenti felszólalásokat tartalmazó nyílt hozzáférésű, és ingyenesen letölthető a parllawspeech.org/data oldalról. Az adatbázist Schwalbach és szerzőtársai (2025) állították össze. Jelen dolgozatban a magyar Országgyűlés 1998 és 2018 közötti öt parlamenti ciklusának felszólalásait használjuk. Az adatbázis szűrése során csak azokat a felszólalásokat tartottuk meg, amelyeket a Fidesz vagy az MSZP frakció tagjai mondtak el, nem az ülésvezető, valamint amelyek legalább 51 szóból álltak. Továbbá kizártuk azokat a felszólalásokat, amelyek nem rendelkeztek releváns politikai tartalommal a

Babel machine klasszifikációs algoritmus szerint. A Babel machine algoritmusról a későbbiekben részletesen is írunk.

A polarizáció operacionalizálásához használt gépi tanulási algoritmus alkalmazása előtt a szövegeket előfeldolgoztuk. Első lépésként a felszólalások szavait lemmatizáltuk, vagyis szótári alakra hoztuk. Ehhez a `hugspacy` Python-csomag `hu_core_news_lg` nevű magyar nyelvi modelljét használtuk (Orosz et al., 2023). Ezt követően eltávolítottuk a zárójeles részeket (például: „Taps a kormánypárt soraiból”), az írásjeleket, valamint a szöveget kisbetűsre alakítottuk. Az előfeldolgozás részeként elvégeztük a stopszavazást is, amelyhez a `HunMineR` R-csomag `data_stopwords_extra` listáját használtuk (Poltextlab, 2025). A stopszavas listát manuálisan kiegészítettük a magyar parlamentre jellemző eljárási szavakkal (például: képviselőtárs, napirend, padsor stb.). A teljes, a stopszavakhoz manuálisan hozzáadott lista a mellékletben elérhető.

Az előfeldolgozást követően a szövegeket bigramokká alakítottuk, majd eltávolítottuk a ritkán használt bigramokat. Az automatikus eltávolítás során Gentzkow és szerzőtársai (2019) módszertanához képest enyhébb küszöbértékeket alkalmaztunk, mivel korpuszunk rövidebb időszakot ölel fel. Csak azokat a bigramokat tartottuk meg, amelyek legalább az egyik ciklusban legalább öt különböző felszólaló által használva voltak, a teljes időszakot tekintve összesen legalább öt felszólaló használta őket, és előfordulásuk száma elérte legalább a tízet. Ezt követően manuálisan eltávolítottuk azokat a bigramokat, amelyeket a parlament, mint intézmény sajátosságai alapján procedurálisnak, illetve alacsony szemantikai jelentéssel bíró kifejezéseknek tekintettünk. Mind a végső, megtartott bigramok listája, mind az eltávolított bigramok listája a mellékletben feltüntetett GitHub-repozitóriumban elérhető.

Gentzkow és szerzőtársai (2019) mintájára mi is létrehoztunk egy adatbázist, amely az összes képviselő–ciklus párt tartalmazza, vagyis minden egyes képviselő adott ciklusban betöltött mandátumát. Az adatbázis a képviselő és a ciklus azonosítóján túl tartalmazza azt is, hogy a képviselő a Fidesz vagy az MSZP frakció tagja volt-e, kormánypozícióban volt-e az adott ciklus során (dummy változóként), valamint a nemét (szintén dummy változóként). Az eredeti kutatásban kontrollváltozóként alkalmaztak továbbá államra és régióra vonatkozó kategóriákat, valamint – az amerikai Kongresszus intézményi felépítéséből fakadóan – azt is, hogy a képviselő a Szenátus vagy a Képviselőház tagja volt-e az adott ciklusban. Az ilyen jellegű képviselő–ciklus párokhoz rendelt változók célja, hogy kontrollálni lehessen azokat a nyelvhasználati eltéréseket, amelyek például a kormánypárti vagy ellenzéki szerepkörből, a nemi hovatartozásból, vagy a képviselő regionális háttéréből fakadhatnak (Gentzkow és szerzőtársai,

2019). Mivel Magyarország földrajzi és intézményi szempontból jelentősen eltér az Egyesült Államoktól – különösen olyan, nyelvi szempontból releváns regionális törésvonalak, mint az amerikai Észak–Dél viszony, nálunk nem jellemzők –, az ilyen típusú kontrollváltozók hiánya nem jelent számottevő torzítást az elemzés szempontjából.

III. 2. Kategorizálás

A kutatás célja, hogy a magyar parlamenti felszólalások polarizációját a gazdasági és kulturális törésvonalak szerint vizsgálja. A kategorizálás során a végső célunk a teljes korpuszból kiindulva létrehozni két alkorpuszt is. Egyet, amiben csak a kulturális tematikájú felszólalások vannak és egyet, amelyben a gazdasági, újraelosztási kérdésekkel foglalkozó felszólalások. A felszólalás tartalmától függően egy felszólalás egyszerre jelen lehet mindkét alkorpuszban vagy éppen egyikben sem.

A kategorizálás első lépésként a felszólalások automatizált témakódolását a Comparative Agendas Project (CAP) 21 közpolitikai kategóriáját tartalmazó rendszerének alkalmazásával végeztük el, a PoltextLAB által fejlesztett Babel Machine szolgáltatás segítségével. A szolgáltatás keretében minden egyes parlamenti felszólaláshoz hozzá lett rendelve a tartalma szerint leginkább ráillő közpolitikai kategória. A Babel Machine egy nyílt forráskódú, finomhangolt XLM-RoBERTa alapú nyelvi modell, amely a magyar nyelvű parlamenti felszólalásokon a modell tanításakor 0,84-es macro F1-score teljesítményt ért el (Sebők és sztsai, 2024). A rendszer determinisztikus működésének köszönhetően a kategóriák hozzárendelése teljes mértékben replikálható, megbízhatósága esetenként meghaladja az ember általi kódolásra jellemző pontosság szintjét is (Sebők és sztsai, 2024).

A dimenziócsökkentő eljárás keretében a felszólalásokat gazdasági és kulturális korpuszba rendeztük. A gazdasági korpuszba azok a felszólalások kerültek, amelyeket elsősorban a következő címkék valamelyikével láttak el: makroökonómia, pénzügy és belföldi kereskedelem, külkereskedelem, közlekedés, energia, egészségügy, mezőgazdaság, munkaügy, szociálpolitika, lakhatás vagy környezetvédelem. A kulturális korpuszba pedig azok a felszólalások kerültek, amelyek a Babel machine szolgáltatás alapján az alábbi kategóriák valamelyikéhez tartoztak: emberi jogok, nemzetközi kapcsolatok, kultúra, jog és bűnözés, honvédelem, oktatás, szociálpolitika, lakhatás vagy környezetvédelem. Tehát azok a felszólalások, amik tartalmukat tekintve elsődlegesen a szociálpolitika, a lakhatás vagy a környezetvédelem kérdéséről szóltak mindkét korpuszhoz hozzá lettek rendelve. Az elsődlegesen technológia, közigazgatás vagy állami területek témaköréhez tartozó

felszólalásokat egyik alkorpuszhoz sem adtuk hozzá, ugyanakkor a teljes korpusznak részét képezik. A Babel machine kategorizáló szolgáltatása a politikai tartalommal nem rendelkező felszólalásokat külön jelöli, ezeket a felszólalásokat a teljes korpuszból is töröltük és sem a kulturális, sem a gazdasági korpuszhoz nem adtuk hozzá.

IV. Eredmények

Az előszűrést követően az elemzéshez használt teljes korpusz összesen 15.552 féle különböző bigramot tartalmaz, amelyet az 5 vizsgált parlamenti ciklus alatt 699 különböző képviselő összesen 1.592.337 alkalommal használt. Az elemzés során három különböző korpusz lett használva: teljes korpusz, gazdasági korpusz és kulturális korpusz. A gazdasági korpusz és a kulturális korpusz alhalmazai (alkorpuszai) a teljes korpusznak és nem kölcsönösen kizáróak, tehát egy-egy felszólalás szerepelhet mindkét korpuszban. Arról, hogy az egyes felszólalás beletartozik-e egy adott alkorpuszba az a Babel machine szolgáltatása által hozzáadott CAP kódtól függ, amiről részletesebben az előző fejezetben írtunk.

A három korpusz leíró statisztikáit a 3., 4. és 5. táblázat mutatja be. Minden esetben az első oszlophoz az adott ciklus időtartama tartozik, második és harmadik oszlopban annak a száma van, hogy a Fidesz és az MSZP frakciók hány képviselője mondott el legalább egy olyan felszólalást, amely az adott korpuszhoz tartozik. A negyedik oszlop azt mutatja, hogy összesen az adott ciklusban hány egyedi bigramot használtak a képviselők. Az ötödik és hatodik oszlop azt jelzi, hogy összesen hányszor használták ezeket az egyedi bigramokat az adott ciklusban a pártok képviselői. Az utolsó oszlopban az összes, adott korpusz feltételeinek megfelelő felszólalás száma látható. Fontos kiemelni a táblázatok értelmezésekor, hogy a 2014-es választáson már megközelítőleg csupán csak fele annyi országgyűlési képviselő lett megválasztva, mint korábban.

Ciklus	# fidesz képviselők	# mszp képviselők	Egyedi bigramok	Összes kifejezés		
				Fideszes	MSZPs	Összes felszólalás
1998-2002	149	137	12315	127795	108082	14845
2002-2006	178	185	13596	165551	199754	23659
2006-2010	150	190	13593	124516	187271	17818
2010-2014	232	61	14204	258732	129229	22384
2014-2018	113	31	13541	191817	99590	13203

3. táblázat: Ciklusonkénti leíró statisztikák (saját táblázat a teljes korpuszon)

Ciklus	# fidesz képviselők	# mszp képviselők	Egyedi bigramok	Összes kifejezés		
				Fideszes	MSZPs	Összes felszólalás
1998-2002	143	123	9082	41601	32521	14845
2002-2006	161	173	10567	49716	65142	23659
2006-2010	133	168	10320	34677	64902	17818
2010-2014	223	59	11353	93758	44323	22384
2014-2018	110	31	10506	62131	35404	13203

4. táblázat: Ciklusonkénti leíró statisztikák (saját táblázat a kulturális korpuszon)

Ciklus	# fidesz képviselők	# mszp képviselők	Egyedi bigramok	Összes kifejezés		
				Fideszes	MSZPs	Összes felszólalás
1998-2002	142	129	10522	74116	65747	14845
2002-2006	169	177	11929	100561	120712	23659
2006-2010	144	180	11975	81796	110163	17818
2010-2014	222	57	12275	124250	67326	22384
2014-2018	109	31	11655	101078	54252	13203

5. táblázat: Ciklusonkénti leíró statisztikák (saját táblázat a gazdasági korpuszon)

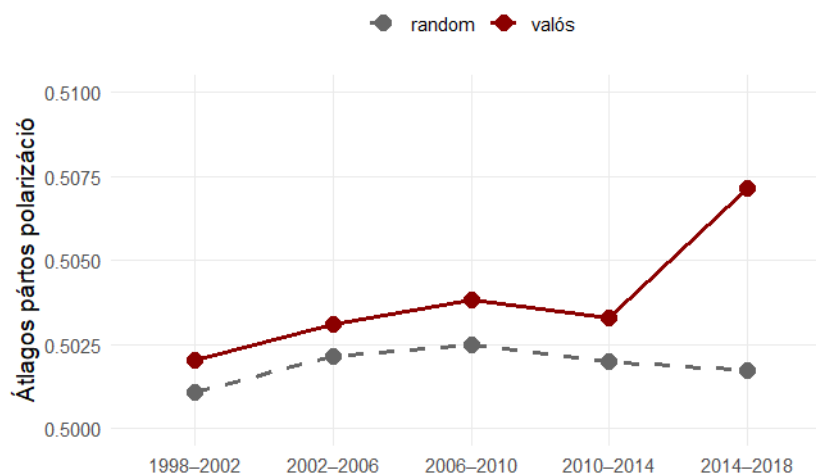
A korábban bemutatott polarizáció operacionalizálásához alkalmazott gépi tanulási számításokat a distrom R-csomag (Taddy, 2013b) segítségével végeztük el. A kutatás reprodukálhatósága érdekében a szükséges kódokat és adatbázisokat a mellékletben megjelölt GitHub-repozitóriumban tettük közzé. A preferált regularizációs taggal számolt eredmények a teljes korpuszon az 1. ábrán láthatók. Minden ciklushoz tartozik egy átlagos pártos polarizációs érték. Folytonos vonallal vannak jelölve a valós adatokon számolt eredmények. Ez az átlagos pártos polarizációs érték azt jelöli, hogy egy személy, aki ismeri a pontos modellt milyen pontossággal tudná helyesen megtippelni egy felszólaló pártját csupán egyetlen bigram alapján (Gentzkow és sztsai, 2019). Nagyobb érték, nagyobb polarizáltságra utal.

A gépi tanulási technikával számolt eredmények esetén mindig fontos lépés az eredmények validálása. Az 1., 2. és 3. ábrán szürke színnel látható a Gentzkow és sztsai (2019) által is használt egyik validációs eljárás alkalmazásából származó eredmények. A validációs folyamat abból áll, hogy minden egyes felszólalóhoz random módon újraosztásra kerül az MSZP és Fidesz pártcímke. Egy adott felszólaló akkora eséllyel kaphatja meg egy ciklusban a Fidesz címkéjét, amilyen a fideszes és mszp-s képviselők aránya az adott ciklusban. Példának okáért egy képviselőhöz a 2002-2006-os ciklusban 49,04% eséllyel kerül fideszes címke és 50,96% eséllyel mszp-s címke, hiszen a teljes korpusz adatbázisban 178 fideszes és 185 mszp-s képviselő volt (2. táblázat). A random módon újraosztott címkéket követően a torzítatlan becslés várható értéke 0,5. Az, hogy a randomizált címkéjű eredmény trendszerűen eltér ettől

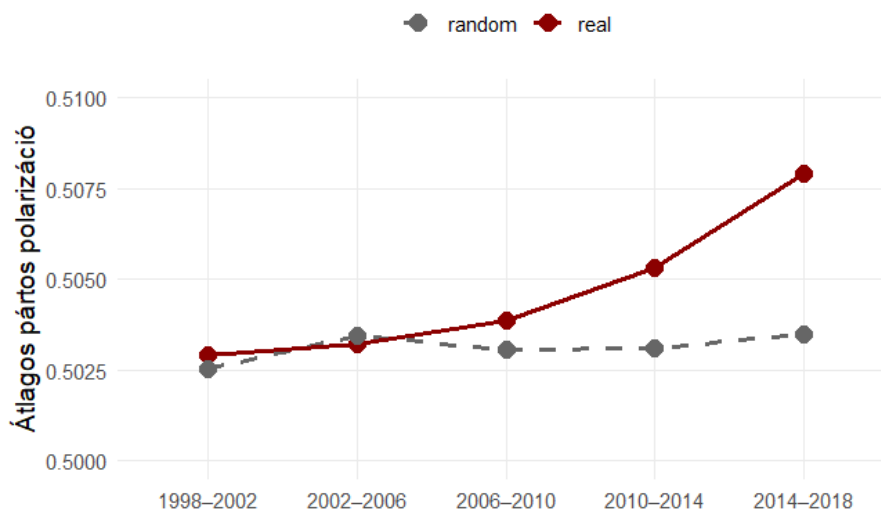
az értéktől az azt jelenti, hogy a becslés nem teljesen mentes a torzításoktól. Mind az alacsony mértékű 0,5-től való eltérés, mind a valós és a random adatpontok közötti pozitív korreláció összhangban van Gentzkow és sztsai (2019) eredményeivel.

A teljes korpuszon kapott eredmények (1. ábra) úgy interpretálhatók, hogy a vizsgált adatokon 1998 és 2014 között alacsony mértékű tendenciózus pártos polarizáció figyelhető csak meg, hiszen a valós adatok eltérése a 0,5 értéktől jelentős részt a mintában maradt torzítás eredménye. Ugyanakkor a 2014-2018-as ciklusban már polarizáltabb párt dinamika figyelhető meg. A kapott eredmények ilyen tekintetben is hasonlítanak az amerikai kongresszusban megfigyelhető mintázatra, viszont van bennük egy körülbelül 20 éves késleltetés. Gentzkow és szerzőtársai (2019) a 150 évnyi kongresszusi beszédek vizsgáló elemzésükben azt találták, hogy az 1990-es évekig nem figyelhető meg számottevő pártos polarizáltság, viszont a '90-es évek elején a pártos polarizáció mértéke jelentősen megnőtt és ez a megnövekedett pártos polarizáció azóta is jelen van.

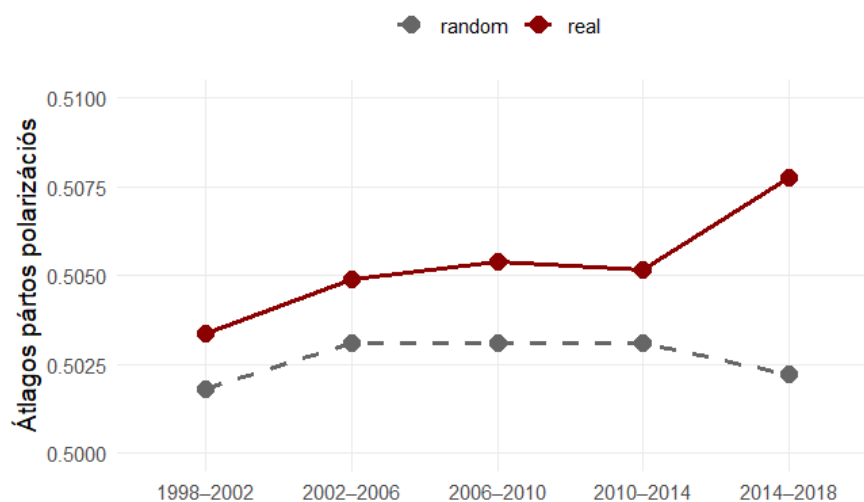
A redukált, gazdasági és kulturális korpusz eredményei a 2. és 3. ábrán láthatók. A randomizált validációs eljárások módja megegyezik a teljes korpuszon végzettel. A gazdasági korpusz eredményeit tekintve a teljes korpusz vizsgálatához hasonló eredmények láthatók: a model alacsony mértékű torzítással terhelt, a 1998-2014-es időszakban alacsony mértékű pártos polarizáció figyelhető meg, viszont 2014-2018-as ciklusban szignifikánsan megnő a pártos polarizáció mértéke. A kulturális korpuszt tekintve (2. ábra) ugyanakkor már a teljes korpusztól és a gazdasági korpuszon számítottaktól eltérő eredmények jelennek meg. A kulturális témájú felszólalások körében az 1998-2010-es időszak pártos polarizációtól mentes, ugyanakkor a 2010-2014-es időszakban megfigyelhető egy jelentős mértékű emelkedés pártos polarizáltság tekintetében, ami a 2014-2018-as időszakban tovább nőtt.



1. ábra: Pártos polarizáció alakulása a teljes korpuszon (saját ábra)



3. ábra: Pártos polarizáció alakulása a kulturális korpuszon (saját ábra)



2. ábra: Pártos polarizáció alakulása a gazdasági korpuszon (saját ábra)

Az eredmények módszertani validációján túl tartalmi validációs módszert is végeztünk. A tartalmi validáció érdekében kigyűjtöttük korpuszonként, ciklusonként és frakciókként a modell szempontjából legmeghatározóbb kifejezéseket. Ez az eljárás a pártos polarizáltság pontosabb megértését is segíti.

A 6. táblázat azt mutatja, hogy mely kifejezések bírtak a legnagyobb magyarázóerővel ciklusonként a teljes korpuszon. A „#F” az adott bigram fideszes képviselők általi előfordulásainak számát jelzi az adott ciklusban, míg a „#M” az MSZP-s képviselők általi használat gyakoriságát. A táblázat bal oldalán a leginkább fideszesnek tekinthető kifejezések szerepelnek – ezek rendelkeznek a legnagyobb pozitív koefficienssel –, míg a jobb oldalon a legjellemzőbben MSZP-s szókapcsolatok láthatók, amelyek negatív koefficiense a legnagyobb mértékű.

Fideszes (1998-2002)	#F	#M	MSZPs (1998-2002)	#F.	#M.
polgári koalíció	225	11	ápv rt	133	336
cw ag	44	0	egészségügyi hozzájárulás	84	213
európai mű	31	0	mezőgazdasági termelő	43	138
járó adókedvezmény	112	19	magyar vidék	9	61
olimpia megrendezés	42	5	fogyatékos élő	9	45
nemzeti stratégia	103	10	következő időszak	41	65
budapest bank	33	1	családi pótlék	415	601
professzionális haderő	40	0	egyértelmű tesz	58	78
kémiai biztonság	54	10	jelentős rész	224	322
rendezési terv	64	11	agrargazdaság fejlesztés	23	103

Fideszes (2006-2010)	#F	#M	MSZPs (2006-2010)	#F.	#M.
onnantól fogva	71	2	magyar honvédség	67	310
regionális vízmű	73	8	munka világ	47	144
térségi társulás	69	0	munka törvénykönyv	52	134
vasúti szárnyvonal	71	5	magyarország fejlesztési	55	318
többszázéves modell	48	5	magas jövedelmű	1	25
tőkés társaság	79	11	következő időszak	96	389
ajak kristály	40	3	akar mond	31	59
életképességi határ	36	2	százalékos mértékű	7	44
adócsökkentési program	62	8	fő ügyész	30	98
pannon pum	42	0	fontos tart	388	1047

Fideszes (2002-2006)	#F	#M	MSZPs (2002-2006)	#F.	#M.
kicsi település	83	3	munka világ	58	291
kirobbanó sikeres	50	0	munka törvénykönyv	101	301
airport rt	56	4	lépés program	38	265
polgári oldal	68	6	következő időszak	71	274
nemzetközi című	36	0	egészségügyi hozzájárulás	48	220
dunatisza köz	54	7	magyar felsőoktatás	32	78
békés megye	246	43	magas jövedelmű	10	50
politikai rendőrség	37	5	szolgálati idő	35	98
ingyen gyógyszer	34	0	fő ügyész	119	237
környezeti hatásvizsgálat	54	9	társadalmi csoport	45	84

Fideszes (2010-2014)	#F	#M	MSZPs (2010-2014)	#F.	#M.
helyi földbiztonság	107	1	helyzetű térség	65	167
nemzeti vidékstratégia	75	2	érdekegyeztető tanács	64	73
arányos egykulcsos	47	2	családi pótlék	112	160
tanyafejlesztési program	62	2	magyar felsőoktatás	35	126
nemzeti hitvallás	128	10	szociális partner	7	19
arányos adórendszer	65	4	személyi jövedelemadó	182	255
elő nemzetiség	73	4	egykulcsos adó	49	126
alaptörvény elfogadás	99	8	magyar közoktatás	17	49
közepes családi	63	0	munkaerőpiaci alap	50	96
tanya élő	89	2	magyar köztársaság	366	341

Fideszes (2014-2018)	#F	#M	MSZPs (2014-2018)	#F.	#M.
alap megtárgyal	83	0	fő ügyész	83	379
összegző jelentés	447	11	családi pótlék	32	155
bank elszámoltatás	97	2	nemzeti bank	559	911
illegális bevándorló	187	5	roma önkormányzat	11	92
migrációs válság	86	2	országos roma	6	82
biztonsági határzár	60	0	magyar nemzeti	706	844
kötelező betelepítési	118	4	szociális ellátórendszer	21	94
betelepítési kvóta	137	5	adott eset	459	651
kötelező kvóta	91	3	magyar felsőoktatás	28	83

6. táblázat: Top 10 magyarázóerővel bíró bigram pártonként, ciklusonként (1998-2018) (saját táblázat a teljes korpuszon)

A teljes korpusz kifejezései alapján (6. táblázat) az alacsony pártos polarizációval rendelkező ciklusokban (1998-2014) a legpártosabb kifejezések között megjelennek kulturális és gazdasági, újraelosztási vonatkozású szókapcsolatok is. Ez egy jó indikátora annak, hogy a teljes korpuszban valóban hasonló fontossággal bírnak mindkét tematikájú felszólalások. Ugyanakkor fontos kiemelni, hogy valós pártos polarizáló hatásuk a modell szempontjából alacsony, hiszen az eredmények alapján a vizsgált első 4 ciklusban nem volt még számottevő polarizáció a teljes korpuszt tekintve.

A 2014-2018-as ciklusban viszont a teljes korpusz tekintetében már megjelent növekedés a pártos polarizáltságban, így ennek leképződését érdemes pontosabban is megvizsgálni. A Fidesz oldalán erősen dominánsnak tűnnek a kulturális vonatkozású, azon belül is migráció témájú bigramok (“illegális bevándorló”, “migrációs válság”, “biztonsági határzár”, “kötelező betelepítési [kvóta]”). Az MSZP által (és Fidesz által nem használt) bigramokat tekintve viszont nem ennyire egyértelmű az elkülönülés. Megjelennek kulturális tematikájú (“fő ügyész”, “országos roma [önkormányzat]”) és gazdasági, újraelosztás tárgykorú (“családi pótlék”, “szociális ellátórendszer”) szókapcsolatok is.

A gazdasági kategóriába sorolt felszólalások legmeghatározóbb bigramjai a 7. táblázatban láthatók. Hasonlóan a teljes korpuszhoz, a gazdasági alkorpuszban is stabilan alacsony mértékű pártos polarizáció van jelen 1998 és 2014 között (3. ábra). Ezen időszakban a legmeghatározóbbnak tűnő szókapcsolatok többnyire valóban gazdasági, újraelosztási tematikájúak, ami megerősíti, hogy sikeres volt a felszólalások gazdaságának kategorizálása. Az utolsó vizsgált, 2014-2018-as ciklusban, amikor a modell eredményei alapján már jelen volt növekedő mértékű pártos polarizáció, ezt a jelenséget valóban mindkettő párt részéről gazdasági relevanciájú kifejezések idézték elő. A Fidesz oldalán ide sorolható kifejezések példának okáért az “otthon meleg[e program]”, “munkahelyvédelmi akció”, “családi adórendszer”, “munkavégzés tilalom”. Az MSZP részéről a legnagyobb magyarázóerővel bíró, gazdasági témájú bigramok többek között a “családi pótlék”, a “jóléti kiadás” és a Magyarországon hatályban lévő nyugdíjminimumra utaló “28 500” kifejezés.

Fideszes (1998-2002)	#F	#M	MSZPs (1998-2002)	#F.	#M.
cw ag	43	0	ápv rt	75	219
polgári koalíció	138	9	mezőgazdasági szövetkezet	2	84
postabank rt	24	0	mezőgazdasági termelő	40	126
munkaügyi konferencia	22	1	agrargazdaság fejlesztés	23	98
olimpia megrendezés	41	5	egészségügyi hozzájárulás	84	210
nemzetközi munkaügyi	44	6	magyar vidék	1	47
műös autópálya	38	0	következő időszak	33	46
kémiai biztonság	54	10	fogyatékos élő	7	40
csatlakozás időpont	44	6	társadalombiztosítás költségvetés	23	131
budapest bank	23	1	munka világ	53	142

Fideszes (2006-2010)	#F	#M	MSZPs (2006-2010)	#F.	#M.
életképességi határ	35	1	magas jövedelmű	1	24
szellemi infrastruktúra	38	2	magyarország fejlesztési	42	229
reform nevezett	34	0	kollektív szerződés	7	67
népegészségügyi adat	30	2	munka világ	38	121
megélhetési válság	28	1	munka törvénykönyv	38	97
ajak kristály	35	3	munka program	23	137
károly róbert	29	2	százalékos mértékű	6	37
vasúti szárnyvonal	58	2	akar mond	17	36
pannon pum	34	0	nyugdíjkorrekciós program	3	69
többszörös modell	44	5	következő időszak	60	285

Fideszes (2002-2006)	#F	#M	MSZPs (2002-2006)	#F.	#M.
kirobbanó sikeres	41	0	lépés program	25	223
kicsi település	55	2	munka világ	45	239
nemzetlő című	27	0	magas jövedelmű	10	48
ingyen gyógyszer	28	0	munka törvénykönyv	88	264
békés megye	208	31	egészségügyi hozzájárulás	45	201
elő szervezet	36	5	szolgálati idő	29	88
innovációs járulék	39	1	társadalmi csoport	27	60
kilométer út	35	3	következő időszak	44	185
kilométer autópálya	122	28	százalékos mértékű	15	51
szociális bérlakás	62	13	gyógyászati segédeszköz	18	136

Fideszes (2010-2014)	#F	#M	MSZPs (2010-2014)	#F.	#M.
helyi földbizottság	73	1	helyzetű térség	47	121
tanyafejlesztési program	61	2	családi pótlék	90	145
nemzeti vidékstratégia	62	1	érdekegyeztető tanács	43	66
tanya élő	84	1	szociális partner	4	15
arányos egykulcsos	47	2	személyi jövedelemadó	134	230
otthonvédelmi akcióterv	35	0	radikális adócsökkentés	8	77
szolgáltatói szektor	40	2	egykulcsos személyi	31	93
népegészségügyi termékadó	48	2	egykulcsos adó	47	105
együttműködés program	36	2	forint emelkedik	36	32
agrár széchenyikártya	28	0	igény vesz	156	238

Fideszes (2014-2018)	#F	#M	MSZPs (2014-2018)	#F.	#M.
bank elszámoltatás	74	2	családi pótlék	28	145
otthon meleg	41	0	jóléti kiadás	6	61
alap megtárgyal	39	0	szociális ellátórendszer	13	79
összegző jelentés	217	8	nemzeti bank	500	679
munkahelyvédelmi akció	68	0	szociális ellátás	23	83
meleg program	35	0	magyar nemzeti	484	597
paralimpiai játék	48	1	szolgálati idő	11	71
családi adórendszer	46	1	szociális támogatás	14	67
munkavégzés tilalom	34	1	28 500	4	68
tag megtárgyal	28	0	érdek szolgál	61	102

7. táblázat: Top 10 magyarázóerővel bíró bigram pártonként, ciklusonként (1998-2018) (saját táblázat a gazdasági korpuszon)

A kulturális kategóriába sorolt felszólalások legnagyobb magyarázóerővel bíró bigramjai a 8. táblázatban láthatók. Ezen korpusz eredményeiről is elmondható, hogy többségében kulturális vonatkozású szókapcsolatok az okozói. Mivel a kulturálisnak minősített felszólalások korpuszában, a teljes korpusztól és a gazdasági korpusztól eltérően, már a 2010-2014-es ciklustól kimutatható növekvő pártos polarizáltság, így ebben az esetben két periódusnyi bigramot is érdemesebb tüzetesebben megvizsgálni. A 8. táblázat legjellemzőbb fideszhez köthető kifejezései például a “nemzeti hitvallás”, az “összetartozás nap”, és a “közbiztonság helyreállítás”. A 2010-14-es ciklusban a leginkább MSZP-hez köthető szókapcsolatok között dominálnak a közoktatással kapcsolatosak (“nemzetvédelmi egyetem”, “magyar felsőoktatás”, “állami finanszírozás”, “magyar közoktatás”, “felsőoktatási intézmény”).

A 2010-14-es ciklushoz képest kulturális polarizáltság dimenziójában a 2014-18-as időszakban kimutatható a Fidesz oldalán egy erőteljes elmozdulás a nemzeti identitással kapcsolatos témákról a bevándorlással kapcsolatosak felé (“reintegrációs őrizet”, “rendvédelmi feladat”, “migrációs válság”). Az MSZP-re jellemző kifejezések között van jogállamisági aggályokra utaló bigram (“fő ügyész”), kisebbségi kérdéssel kapcsolatos szókapcsolat (“országos roma [önkormányzat]”) és, hasonlóan a 2010-14-es időszakhoz, a közoktatás ügyéhez is (“magyar felsőoktatás”, “magyar közoktatás”, “oktatási rendszer”).

Az eredmények értelmezése során felfedeztünk egy, a kutatási célhoz nem szorosan kapcsolódó, viszont politikatudományi szempontból releváns empirikus eredményt. A korábban taglaltak alapján, a használt modell pozitív koefficiens értéket rendel azokhoz a bigramokhoz, amelyek egy adott ciklusban jelentősen gyakrabban vannak használva a Fidesz képviselői által, mint az MSZP-s felszólalók által. Hasonló módon, negatív koefficiens értékkel látja el azokat szókapcsolatokat, amiket jellemző módon az MSZP képviselői használnak, miközben a Fidesz felszólalásokban nem hangzik el. Ugyanakkor, mivel a lehetséges kifejezések többségét mindkét párt rendszeresen használja, a legtöbb bigram, a regularizációs tagok hatására, nulla értékű koefficienssel rendelkezik, tehát semmilyen magyarázóerővel nem bír a modell szempontjából.

Fideszes (1998-2002)	#F	#M	MSZPs (1998-2002)	#F.	#M.
polgári koalíció	71	3	fogyatékos élők	6	37
veszélyes anyag	44	3	gazdasági társaság	12	27
történeti hivatal	31	7	rendkívül fontos	58	93
professzionális haderő	40	0	családi pótlék	75	116
szervezett alvilág	16	0	jelentős rész	61	88
professzionális hadsereg	24	0	probléma megoldás	22	42
tagjelölt ország	20	0	rész vesz	127	119
kémiai biztonság	52	8	adófizető pénz	7	15
szolgálat időtartam	16	1	következő időszak	7	14
önkéntes lakásfoglaló	20	1	áll rendelkezés	56	56

Fideszes (2006-2010)	#F	#M	MSZPs (2006-2010)	#F.	#M.
alternatív energia	24	2	mentelmi jog	9	121
onnantól fogva	24	0	magyar honvédség	41	232
globális felmelegedés	33	5	honvédelmi minisztérium	17	72
natura 2000-es	28	0	magyar felsőoktatás	46	118
időügyi charta	17	1	áll rendelkezés	37	173
térségi társulás	22	0	magyar közoktatás	21	105
regionális vízmű	15	0	önkéntes tartalékos	11	76
kínai népköztársaság	15	1	következő időszak	17	84
szociális válság	37	10	egyesült állam	17	108
tandíj bevezetés	33	9	büntető törvénykönyv	39	142

Fideszes (2002-2006)	#F	#M	MSZPs (2002-2006)	#F.	#M.
politikai rendőrség	31	5	magyar felsőoktatás	26	67
erdős ákos	20	0	következő időszak	15	67
princz gábor	23	0	rendszeres gyermekvédelmi	25	102
történeti hivatal	42	6	társadalmi csoport	12	30
polgári oldal	24	0	fő ügyész	90	145
jóléti rendszerváltás	25	3	döntés születik	6	30
élő szervezet	26	2	felsőoktatási intézmény	92	194
felzárkóztató oktatás	53	9	helyzetű térség	9	25
állambiztonsági tevékenység	19	2	túli magyarság	31	100
fejlesztési központ	15	0	emberi jog	26	73

Fideszes (2010-2014)	#F	#M	MSZPs (2010-2014)	#F.	#M.
tanyafejlesztési program	40	0	nemzetvédelmi egyetem	3	56
nemzeti hitvallás	120	5	magyar felsőoktatás	19	109
alaptörvény elfogadás	55	2	állami finanszírozott	10	59
összetartozás nap	29	1	zrínyi miklós	6	45
közbiztonság helyreállítás	33	0	magyar közoktatás	13	42
magyarország alaptörvény	110	6	operatív program	33	20
közrend közbiztonság	56	3	fejlesztési program	18	25
élő nemzetiség	53	2	felsőoktatási intézmény	98	115
szociális családügyi	36	2	magyar köztársaság	146	174
előzetes jognyilatkozat	49	1	magyar honvédség	261	378

Fideszes (2014-2018)	#F	#M	MSZPs (2014-2018)	#F.	#M.
összegző jelentés	146	3	fő ügyész	56	317
reintegrációs őrizet	41	1	családi pótlék	3	46
bank elszámoltatás	29	0	magyar felsőoktatás	12	53
szakképzési hídprogram	29	0	roma önkormányzat	10	51
alap megtárgyal	29	0	országos roma	6	45
rendvédelmi feladat	37	1	idős ember	14	68
migrációs válság	26	0	magyar közoktatás	6	30
köznevelési kerekasztal	26	0	nemzeti bank	64	116
kulturális alapellátás	26	0	bírói kar	3	39
foglalt követelmény	35	0	oktatási rendszer	59	74

8. táblázat: Top 10 magyarázóerővel bíró bigram páronként, ciklusonként (1998-2018) (saját táblázat a kulturális korpuszon)

A 9. táblázat a nem nulla magyarázóerővel bíró bigramok számát mutatja ciklusonként, pártonként és korpuszonkénti bontásban. A táblázatból jól látható, hogy az MSZP, akár a teljes korpuszt, akár az alkorpuszokat vizsgálva, lényegesen kevesebb olyan kifejezéssel rendelkezik, amely szorosan köthető a párt képviselőinek felszólalásaihoz, mint a Fidesz. Ez arra utal, hogy kevés olyan bigram akad, amely elsősorban az MSZP-s beszédekben jelenik meg, míg a fideszesek vagy egyáltalán nem, vagy csak jóval ritkábban használták őket.

A 2010-14-es és 2014-18-as ciklusban ez fakadhat abból, hogy jelentősen kevesebb MSZP-s parlamenti képviselő volt, mint fideszes és így sokkal kisebb számú kifejezést is mondtak el összességében (3., 4., 5. táblázat). Viszont az 1998-2002, 2002-2006 és 2006-2010-es ciklusok során a képviselők száma és a használt kifejezésszámok is kiegyenlítettebbek voltak, vagyis ez az első 3 vizsgált ciklus esetében nem magyarázza a látott anomáliát. A kapott eredmények arra utalnak, hogy az 1998–2010 közötti három parlamenti ciklusban, hogy a Fideszhez viszonyítva, kifejezetten kevés olyan téma volt, amelyet az MSZP egyértelműen és egyedül tematizált volna a parlamenti felszólalások szintjén.

Ciklus	Teljes korpusz		Gazdasági korpusz		Kulturális korpusz	
	Fidesz	MSZP	Fidesz	MSZP	Fidesz	MSZP
1998-2002	167	39	92	26	48	4
2002-2006	293	80	193	57	64	12
2006-2010	239	60	167	38	51	12
2010-2014	419	98	191	55	145	24
2014-2018	489	82	243	45	137	23

9. táblázat: Nem nulla magyarázóerővel bíró bigramok száma (saját táblázat)

V. Konklúzió

A dolgozat célja a magyar politikai polarizáció alakulásának elemzése volt az 1998 és 2018 közötti időszakban. Az elemzéshez természetes nyelvfeldolgozást használtunk a magyar parlament Fidesz és MSZP frakcióhoz tartozó képviselők felszólalásain. A politikai polarizáció Genztkow és szerzőtársai (2019) gépi tanulási technikája alapján úgy operacionalizáltuk, hogy, a megfigyelő, aki pontosan ismeri a modellt egyetlen bigram alapján milyen pontosan tudná helyesen prediktálni, hogy az adott kifejezést használó képviselő melyik pártnak a tagja. A választói csoportok megfigyelhető átrendeződésének okán az elemzésünket Gennaioli és Tabellini (2023) keretrendszerét követve a teljes korpuszunkat felbontottuk kulturális és gazdasági témákkal foglalkozó felszólalásokat tartalmazó alkorpuszokra is és a pártos polarizációt, a teljes korpusz mellett, az alkorpuszokon is elemeztük.

A teljes és a gazdasági témákat tartalmazó korpuszokat tekintve a 2014-ig változatlan mértékű alacsony pártos polarizáció mutatható ki. A modell eredményei azonban arra utalnak, hogy a 2014-2018-as ciklusban erősödött a pártos polarizáció. A validációs eredmények alapján, megnézve a legnagyobb magyarázóerővel bíró kifejezéseket kijelenthető, hogy a két párt közötti polarizált narratíva valóban redisztribúciós kérdések mentén történik ezekben a korpuszokban.

A kulturális korpuszt vizsgálva 1998 és 2010 közötti időszak stabilan pártos polarizáltságtól mentes volt. A 2010–2014-es ciklusban már megjelent a pártos polarizáció, amely a 2014–2018-as időszakban tovább növekedett. A validációs eljárás alapján megállapítottuk, hogy a 2010–2014-es ciklusban a megosztottságot jelentős mértékben a Fidesz részéről a nemzeti identitás kérdésének hangsúlyozása, míg az MSZP oldaláról a közoktatás kiemelt kezelése táplálta. A 2014–2018-as ciklusban a tartalmi fókuszokban markáns elmozdulás figyelhető meg: a Fidesz ekkor a bevándorlás kérdésére helyezte a hangsúlyt, míg az MSZP több irányban osztotta meg figyelmét, főként jogállamisági, kisebbségi és közoktatási témákkal foglalkozva.

A módszertani validációt jelentő pártcímkék random újraosztása során Gentzkow és szerzőtársai (2019) eredményeihez hasonló értékeket kaptunk, vagyis a mi eredményeinkben is csupán alacsony mértékű kontrollálatlan torzítás maradt. Tartalmi szempontból hasonlítva az eredményeinket az övékéhez az mondható el, hogy Magyarországon körülbelül 20 évvel később, a kettőezertízes évek közepén jelent meg az a típusú növekvő trendet mutató pártos polarizáció, ami már az 1990-es évek elejétől jellemzi az amerikai Kongresszusi beszédeket.

Továbbá a kutatási célhoz nem szorosan kapcsolódó empirikus eredményünk, hogy bármelyik korpuszt és bármelyik ciklust tekintve az MSZP-s képviselők kifejezetten kevés olyan szókapcsolatot használnak, amelyet a fideszes képviselők ne használnának. Az MSZP szempontjából darabszámra kevés magyarázóerővel bíró bigram jelenléte azt jelenti, hogy alig volt olyan témája a vizsgált két évtizedben, amelyet az MSZP úgy tudott volna egyértelműen tematizálni a magyar parlamentben, hogy a fideszes képviselők ne beszéltek volna róla szintén. Példának okáért ezzel szemben az első három ciklusban, amikor nagyságrendileg azonos számú országgyűlési képviselője volt a két pártnak, így az összehasonlításnak van értelme, a fideszes politikusok átlagosan négyszer annyi magyarázóerővel rendelkező bigramot használtak. Ez azt jelenti, hogy bármelyik ciklust tekintve azonosíthatók olyan témák, amelyeket a fideszes képviselők úgy tematizáltak sikeresen a parlamenti felszólalások keretében, hogy az MSZP-s képviselők alig vagy egyáltalán nem használták a hozzájuk tartozó kifejezéseket.

A dolgozat hozzájárul a magyar politikai nyelvhasználat és pártdinamikák mélyebb megértéséhez, valamint újabb érvekkel szolgál amellet is, hogy a politikai polarizáció vizsgálata során indokolt a kulturális és gazdasági dimenziók szétválasztása. Tudomásunk szerint korábban nem született olyan publikált empirikus vizsgálat, amely magyar nyelvű természetes nyelvi adatokon alapulva, politikatudományi szempontból, szisztematikusan elemezte volna a pártos polarizációt, a kulturális és gazdasági tematikák szerinti megosztottságot.

Hivatkozásjegyzék

1. Bajomi-Lázár, P. (2014). *Party Colonisation of the Media in Central and Eastern Europe*. Central European University Press. Budapest.
<https://www.aup.nl/en/book/9789633860410/party-colonisation-of-the-media-in-central-and-eastern-europe>
2. Bayram, U., Pestian, J., Santel, D., Minai, A.A. (2019). What's in a word? Detecting partisan affiliation from word use in congressional speeches. *In: Paper presented at the International Joint Conference on Neural Networks (IJCNN)*, Budapest, Hungary, 14–19th July 2019.
3. Bíró-Nagy, András. (2017). Illiberal Democracy in Hungary: The Social Background and Practical Steps of Building an Illiberal State.
https://www.researchgate.net/publication/316994178_Illiberal_Democracy_in_Hungary_The_Social_Background_and_Practical_Steps_of_Building_an_Illiberal_State
4. Bódi, M., & Kovalcsik, T. A Fidesz társadalmi bázisának változása 1998-2022 között. (2023). *Földrajzi Közlemények*, 147(1), 40–52. <https://doi.org/10.32643/fk.147.1.3>
5. Bonomi, Giampaolo, Nicola Gennaioli, and Guido Tabellini (2021), “Identity, Beliefs, and Political Conflict”, *The Quarterly Journal of Economics*, vol. 136, no. 4, pp. 2371–2411.
6. Buda, J. M., & Németh, R. (2024). A felügyelt gépi tanulás alkalmazási lehetőségei szöveges adatokon. A magyar országgyűlésben 1998–2018 között elhangzott beszédek elemzése. *Statisztikai Szemle*, 102(11), 1087–1103. <https://doi.org/10.20311/stat2024.11.hu1087>
7. Dias, N., & Lelkes, Y. (2021). The Nature of Affective Polarization: Disentangling Policy Disagreement from Partisan Identity. *American Journal of Political Science*, 66(3), 775–790. <https://doi.org/10.1111/ajps.12628>
8. Fiva, Jon H.; Nedregård, Oda; Øien, Henning (2021). Polarization in Parliamentary Speech, *CESifo Working Paper*, No. 8818, Center for Economic Studies and Ifo Institute (CESifo), Munich
9. Gennaioli, N., & Tabellini, G. (2023). Identity Politics. *Social Science Research Network*.
<https://doi.org/10.2139/ssrn.4395173>
10. Gentzkow, M., Shapiro, J. M., & Taddy, M. (2019). Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional speech. *Econometrica*, 87(4), 1307–1340. <https://doi.org/10.3982/ecta16566>

11. Gethin, Amory, Clara Mart'inez-Toledano, and Thomas Piketty (2022), "Brahmin Left Versus Merchant Right: Changing Political Cleavages in 21 Western Democracies, 1948–2020", *Quarterly Journal of Economics* 137 (1), pp. 1-48.
12. Goet, N.D., (2017). Measuring polarization with text analysis: Evidence from the UK House of Commons, 1811–2015. In: *Paper prepared for the Polarization, Institutional Design and the Future of Representative Democracy workshop*, Berlin, Harnack Haus.
13. Green, J., Edgerton, J., Naftel, D., Shoub, K., Cranmer, S.J. (2020). Elusive consensus: Polarization in elite communication on the COVID-19 pandemic. *Science Advances* 6 (28). <https://doi.org/10.1126/sciadv.abc2717>
14. Hargitai L., (2002). Érez vagy gondolkodik a magyar választópolgár? (*Médiakutató*). (n.d.). https://mediakutato.hu/cikk/2002_02_nyar/01_magyar_polgar
15. Higley, J., Pakulski, J., Wesolowski, W. (2002). Introduction: Elite Change and Democratic Regimes in Eastern Europe. In Higley, John – Pakulski, Jan – Wesolowski, Włodzimierz (eds.): *Postcommunist Elites and Democracy in Eastern Europe*. Chippenham, Wiltshire: Palgrave. 1–33.
16. Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, not ideology. *Public Opinion Quarterly*, 76(3), 405–431. <https://doi.org/10.1093/poq/nfs038>
17. Iyengar, S., & Westwood, S. J. (2014). Fear and Loathing across Party Lines: New Evidence on Group Polarization. *American Journal of Political Science*, 59(3), 690–707. <https://doi.org/10.1111/ajps.12152>
18. Krekó, P., & Enyedi, Zs. "Explaining Eastern Europe: Orbán's Laboratory of Illiberalism." *Journal of Democracy* 29, no. 3 (2018): 39-51.
19. Levendusky, M., & Malhotra, N. (2015). Does media coverage of partisan polarization affect political attitudes? *Political Communication*, 33(2), 283–301. <https://doi.org/10.1080/10584609.2015.1038455>
20. McCarty, Nolan M., Keith T. Poole, and Howard Rosenthal. (2006). *Polarized America: The dance of ideology and unequal riches*. MIT Press. Cambridge.
21. Németh, R. A scoping review on the use of natural language processing in research on political polarization: trends and research prospects. *J Comput Soc Sc* 6, 289–313 (2023). <https://doi.org/10.1007/s42001-022-00196-2>

22. Orosz, G., Szabó, G., Berkecz, P., Szántó, Z., & Farkas, R. (2023, August). Advancing Hungarian text processing with HuSpaCy: Efficient and accurate NLP pipelines. In International Conference on Text, Speech, and Dialogue (pp. 58-69). Cham: Springer Nature Switzerland.
23. Palonen, Emilia 2009. Political Polarisation and Populism in Contemporary Hungary. In *Parliamentary Affairs* 2009/2. 318–334.
24. Poltextlab. (2025). HunMineR: Companion package for the Hungarian text mining textbook [Computer software]. GitHub. <https://github.com/poltextlab/HunMineR>
25. Sebők, M., Máté, Á., Ring, O., Kovács, V., & Lehoczki, R. (2024). Leveraging Open Large Language Models for Multilingual Policy Topic Classification: The Babel Machine Approach. *Social Science Computer Review*, 0(0). <https://doi.org/10.1177/08944393241259434>
26. Simola, A., May, V., Olakivi, A., & Wrede, S. (2022). On not ‘being there’: Making sense of the potent urge for physical proximity in transnational families at the outbreak of the COVID-19 pandemic. *Global Networks*, 23(1), 45–58. <https://doi.org/10.1111/glob.12382>
27. Schwalbach, Jan; Hetzer, Lukas; Proksch, Sven-Oliver; Rauh, Christian; Sebők, Miklós (2025), “ParlLawSpeech”; *GESIS, Cologne*. doi:10.7802/2824.
28. Taddy, M. (2013a). “Distributed Multinomial Regression.” *Annals of Applied Statistics*. <https://cran.r-project.org/web/packages/distrom/citation.html>
29. Taddy, M. (2013b). Multinomial inverse regression for text analysis. *Journal of American Statistical Association*, 108(503), 755–770. <https://doi.org/10.1080/01621459.2012.734168>
30. Taddy, M. (2015). Distributed multinomial regression. *Annals of Applied Statistics*, 9(3), 1394–1414. <https://doi.org/10.1214/15-aoas831>

Melléklet

Melléklet I.

A kutatás reprodukálhatósága érdekében a szükséges kódokat és adatbázisokat az alábbi GitHub-repozitóriumban tettük közzé.

https://github.com/nlemu/politikai_polarizacio

Melléklet II.

Stopszavak listája:

ajánlási, albizottság, alelnök, általános, asszony, benyújtott, bizottság, bizottsági, csengő, dr, ellenzék, ellenzéki, elfogad, elnézés, elmúlt, elnök, elnöki, érdekes, fidesz, fideszmagyar, gyurcsány, ház, határozat, határozati, időkeret, jelez, jobbik, javaslat, képviselő, képviselőtárs, kormánypárt, kormánypárti, kormány, köztársasági, köszön, medgyessy, megkocogtatás, miniszter, miniszterelnök, módosítás, módosító, mszp, napirend, napirendi, orbán, országgyűlés, padsor, sor, szabaddemokrata, szeret, szó, szocialista, szocialistaszabad, szóló, szövetség, taps, tisztelt, törvény, törvényjavaslat, úr, ülés, választási, válaszol, vonatkozó, zaj