Kulcsfogalmak jelentésváltozása a Kádár-korszak politikai diskurzusában

Ring Orsolya[1], Kmetty Zoltán[1], Szabó Martina Katalin[1,2], Kiss László[1], Nagy Balázs[2], Vincze Veronika[3]

[1]Társadalomtudományi Kutatóközpont, CSS-RECENS Kutatócsoport
1097 Budapest, Tóth Kálmán u. 4.
[2]Szegedi Tudományegyetem, Informatikai Intézet
6720 Szeged, Árpád tér 2.
[3]MTA-SZTE Mesterséges Intelligencia Kutatócsoport
6720 Szeged, Tisza Lajos körút 103.
{kiss.laszlo, kmetty.zoltan, ring.orsolya}@tk.mta.hu,
{bnagy, martina, vinczev}@inf.u-szeged.hu

Kivonat A jelen dolgozatban a Magyar Szocialista Munkáspárt Központi Bizottságának (MSZMP KB) hivatalos havilapját, a *Pártélet* című kiadványt elemezzük néhány korabeli kulcsfogalom időbeli szemantikai változása szempontjából. A vizsgálatokhoz a korpuszt magunk hoztuk létre a lap teljes, digitalizált képként rendelkezésre álló anyagából. A korpusz egyedülálló, hiszen tudomásunk szerint nincs másik olyan digitalizált adatbázis, amely a Kádár-korszak longitudinális szövegbányászati elemzését lehetővé tenné. Mivel a *Pártélet* az állampárt hivatalos lapja volt, szövegeinek elemzése révén a korabeli politikai diskurzus megismerése, változásainak feltárása válik lehetővé. Jelen kutatásunk középpontjában a *döntés* és az *irányítás*, illetve a velük kapcsolatban álló fogalmak szemantikai tartalmának időbeli változása állt.
Kulcsszavak: korpuszépítés, információkinyerés, szóbeágyazási modellek, történeti diskurzuselemzés, számítógépes történettudomány

1. Bevezetés

A magyar történelem 1956 és 1989 közötti időszaka a történettudományban és a társadalomtudományban is a gyakran vizsgált korszakok közé tartozik. Politikai diskurzusának nyelvi jellemzői azonban eddig nem képezték elemzés tárgyát. E probléma okán korpuszt építettünk a *Pártélet* című folyóiratból, az MSZMP KB hivatalos ideológiai lapjából, majd azt NLP-módszerekkel feldolgoztuk és elemeztük.

A *Pártélet* című lap lehetővé teszi az országot irányító állampárt hivatalos diskurzusának elemzését. A kiadvány célja a politikai ideológia terjesztése, tehát a közvetlen agitáció és propaganda volt. A *Pártélet*, amely 54 150 példányban jelent meg, elsősorban nem az átlagemberekhez, hanem az állampárt különböző tisztségviselőihez szólt. A lap a párthierarchia egészét célozta, az első lapszámban megjelent ajánlás szerint ezen belül is elsősorban a pártfunkcionáriusok, az

aktivisták, valamint a propagandatevékenységért felelős pártmunkások számára íródott.

Az utóbbi években a magyarországi társadalmi és politikai folyamatokhoz kapcsolódó diskurzusok kvalitatív történelmi elemzése egyre gyakoribbá vált, ugyanakkor ezek az elemzések leginkább a hagyományos történeti diskurzuselemzés módszereit alkalmazzák, azaz a kortárs dokumentumok kvalitatív és manuális elemzésén alapulnak (Szabó, 2007; Pap, 2017; Gyáni, 2016). A szövegbányászati módszerek alkalmazását, illetve a kvantitatív elemzéseket alapvetően a digitalizált szövegkorpuszok hiánya akadályozta, ami pedig a digitális formátumú szövegek hiányával, valamint a történeti korpuszok építésének tipikus technikai problémáival mutat szoros összefüggést. Felismerve ezt a jelentős hiányt kezdtünk bele egy nagyméretű, digitalizált szövegkorpusz létrehozásába. A Pártéletkorpusz egyedülálló lehetőséget kínál számos, eleddig kivitelezhetetlen vizsgálat elvégzésére. Így például a segítségével elemezhető a korszak politikai diskurzusában zajló időbeli változások dinamikája (Xu és Kemp, 2015; Jatowt és Duh, 2014; Kulkarni és mtsai, 2014; Hamilton és mtsai, 2016a,b; Garg és mtsai, 2018).

Dolgozatunkban egy esettanulmányon, néhány kulcsfogalom időbeli dinamikájának a vizsgálatán keresztül mutatjuk be a korpuszunk és a kvantitatív szövegelemzés hasznosságát a történeti diskurzuselemzés számára. A munka során a szóbeágyazás módszerét alkalmazzuk, ami a természetesnyelv-feldolgozás (NLP) és a gépi tanulás területén gyakorta használt eszköz bármely két szó szemantikai kapcsolatának feltárására, illetve dinamikus perspektívába helyezve az időbeli szemantikai változások mérésére.

Esettanulmányunk célja a korszak különböző kulcsfogalmaival kapcsolatos politikai diskurzus változásainak azonosítása a Kádár-korszak éveiben Magyarországon. Ezen kulcsfogalmakat történelmi és szociológiai kritériumok alapján választjuk ki, és azt vizsgáljuk, hogy hogyan változik közöttük az időben a szemantikai kapcsolat. Mindehhez hat, történettudományi szempontból elkülönülő alkorszakot definiálunk, és az egyes korszakok vektorainak összehasonlításával kiszámítjuk a fogalmak időbeli dinamikáját.

Megvizsgálva a választott kulcsfogalmaknak a politikai diskurzusban betöltött szerepét, új, kvantitatív kutatási eredményekkel egészítjük ki és pontosítjuk a korábban e témában született kvalitatív eredményeket.

2. Történelmi háttér

Az 1956-os forradalom, majd az azt követő megtorlás időszaka után a Kádár-korszak a társadalom konszolidálását tűzte ki célul. A konszolidációs politika lényege a társadalom „lecsendesítése", a politikától, a politikai gondolkodástól való eltávolítása volt. A konszolidáció központi elemét képezte a fogyasztásra helyezett hangsúly, a társadalom széles rétegei számára elérhető „második gazdaságbeli" termelési formák tolerálása, idővel támogatása. Természetesen a sikeres konszolidációs politika az előző, Rákosi-korszakkal való viszonylagos szembehelyezkedést is szükségessé tette, akárcsak a hruscsovi Szovjetunió számára a sztálini előzményekkel való leszámolást. A már 1962-ben megindult új gazda-

ságpolitikai intézkedések előkészítették a terepet az 1968-ban kihirdetett gazdasági reformprogramnak, az „új gazdasági mechanizmusnak". Ennek keretében az egyes gazdasági szereplők, vállalatok a korábbinál lényegesen nagyobb önállóságra tehettek szert, döntési jogkörük és a központi irányítástól való függetlenségük megnőtt. Jelentősen megerősödött a „második szektor" (a saját fogyasztásra és értékesítésre termelő háztáji és kisegítő gazdaságok, a gazdasági munkaközösségek stb.), valamint a legális magánszektor is. A társadalom egyre jelentősebb rétegei érezhették úgy, hogy fogyasztási színvonaluk és életszínvonaluk javul. A reformfolyamat „keményvonalas" ellenzői azonban 1972-re megbuktatták a „mechanizmust", a gazdaságban ismét erőteljes központosítást indítottak. A döntések ismét centralizáltak lettek, a gazdasági szektor központi irányítása fokozódott. Az 1979-es „második olajárrobbanás" után újra bevezették az 1968. évi reform néhány elemét. Csökkent a központi irányítás szerepe és újra erősebben támogatták a lakosság „második gazdaságban" való részvételét. Mindez természetesen ismét csak a konszolidációs társadalompolitikával hozható összefüggésbe. A Rákosi-korszak kvázi háborús ideológiájával, háborús készülődésre utaló társadalompolitikájával szemben a Kádár-rendszer a „békés" szocialista fejlődést, a magas (illetve magasan tartott) fogyasztási színvonalat, valamint a depolitizálást tűzte zászlajára. A fogyasztás fokozása, a fogyasztási színvonal magasan tartása jelentette a Kádár-rendszer legfőbb erejét – egyfelől távol tudta tartani a társadalom jelentős rétegeit az aktív politizálástól (ez természetesen igen komoly ideológiai háttérmunkát igényelt), másfelől a rendszer „hatékonyságának" is bizonyítékául szolgált.

3. A korpusz létrehozása

3.1. Korpuszépítés, előfeldolgozás

A vizsgálatokhoz használt korpuszt, a Pártélet című lap számait az Arcanum Digitheca[1] oldalról töltöttük le. A lap szkennelt, PDF-formátumú oldalait a letöltés után további komplex feldolgozási folyamatoknak vetettük alá, amelyek eredményeképpen megkaptuk a szövegek elemezhető és megfelelő minőségű nyers változatait.

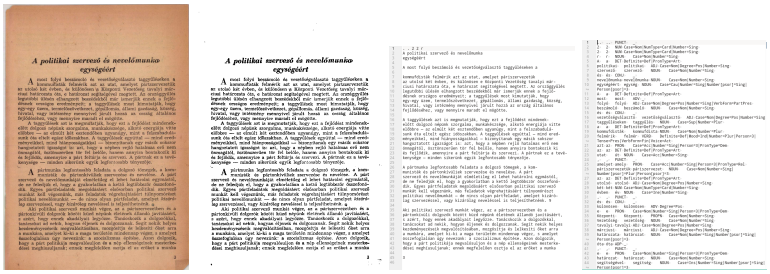
Először, mivel az optikai karakterfelismerő eszköz (Optical Character Recognition, OCR) képfájlokkal működik, az egyes PDF-oldalakat képi formátumba (PNG) konvertáltuk a pdftoppm konverter segítségével.

Második lépésként a PNG fájlokat binarizáltuk, vagyis fekete-fehér képekké alakítottuk át az ImageMagick[2] nevű eszközzel, amely a pdftoppm konverterhez hasonlóan ugyancsak minden Linux disztribúcióban elérhető. 50%-os küszöbértéket alkalmaztunk, ami azt jelenti, hogy minden ezen érték feletti pixelt feketére, a többit fehérre állítottuk. Ez a technika növeli az OCR-folyamat hatékonyságát azáltal, hogy növeli a kontrasztot a szöveg és a háttér között.

[1] https://adtplus.arcanum.hu/en/collection/Partelet/
[2] https://imagemagick.org

Az eredményfájlokon ezután a *tesseract* nevű nyílt forráskódú OCR eszközzel dolgoztunk tovább[3]. Az OCR segítségével az oldalfényképeket géppel olvasható szövegekké alakítottuk, alkalmassá téve őket a további gépi feldolgozásra. Végül a nyers szövegekből eltávolítottuk az oldalszámokat, az üres sorokat és kezeltük az elválasztásokat. Mindehhez saját bash és Python szkripteket használtunk.

A munkafolyamat egyes lépéseit az alább ábrák szemléltetik.



1. ábra: A fájlok állapota az egyes feldolgozói szakaszokban

A kapott szöveget a magyarlanc nyelvi elemző eszközzel[4] (Zsibrita és mtsai, 2013) dolgoztuk fel, amelynek segítségével a korpusz szövegeit először mondatokra bontottuk, tokenizáltuk és lemmatizáltuk. Ezután eltávolítottuk az írásjeleket és a stopszavakat. A stopszavak szűrése a további, szóbeágyazási modellel végzett szemantikai vizsgálatok szempontjából fontos lépés volt, hogy elkerüljük a nagyon gyakran előforduló szavak által okozott zajt az eredményekben.

A magyarlanc eszköz körülbelül 22000 szónak „ismeretlen szófaj" jelölést adott, azaz nem tudta meghatározni azok szófaját és morfológiai sajátságait, nagyrészt az OCR-hibáknak köszönhetően. Ezek kezelésére az alábbi lépéseket követtük. Először is kigyűjtöttük azokat az ismeretlen elemzésű szavakat, amelyek legalább hússzor előfordultak a korpuszban, továbbá minimum három karakterből álltak (pl. *imperializmus*). A következő lépésben ezeket kézi erővel javítottuk, majd az eredeti szövegben lecseréltük az eredeti alakokat a javított változatokra, végül újraelemeztük a korpuszt a magyarlanccal. Ezzel a módszerrel az ismeretlen szavak 64%-át sikerült kijavítanunk.

3.2. Alapvető korpuszadatok

A teljes *Pártélet* folyóirat összesen 33 évfolyamból áll, amelyeket 1956 és 1989 között publikáltak, évente 12 számmal. A végleges korpuszunk összesen 13 185 200 tokent tartalmaz. A tokenek megoszlása a korpuszban kiegyensúlyozottnak tekinthető az egyes évek között, tehát a tokenek száma nagyjából megegyezik minden évben. Megjegyezzük azonban, hogy az 1956-os novemberi és decemberi

[3] https://github.com/tesseract-ocr

[4] http://www.inf.u-szeged.hu/rgai/magyarlanc

számok nem jelentek meg, ezért hiányoznak az összeállításunkból. A folyóirat utolsó száma 1989 áprilisában jelent meg.

4. A korpusz feldolgozása

Célunk a kiválasztott fogalmak szemantikai változásának feltárása volt, amihez a szóbeágyazás módszerét alkalmaztuk.

A szóbeágyazás alapvetően egy adott szótár szavainak vektorszerű ábrázolását jelenti, ahol a szóvektor dimenziójának alacsonyabbnak kell lennie, mint maga a szótár elemeinek a száma. A szótár egy adott dokumentumot vagy egy adott korpuszt reprezentál.

Az egyes nyelvi elemek vektorai alapján kiszámíthatjuk az egyes vektorok közötti távolságot, képet kapva ezáltal az adott két szó közötti szemantikai hasonlóságról, illetve különbségről. Egy adott beágyazási modellben ugyanis a hasonló kontextusban szereplő szavak vektorai hasonlóan helyezkednek ez az adott vektortérben, és a disztribúciós hipotézis alapján a szemantikailag hasonló szavak hasonló disztribúciós sajátságokkal (Harris, 1954), ezáltal pedig hasonló vektorreprezentációval rendelkeznek. Az elmondottakkal összefüggésben, a szóvektorok az időbeli szemantikai változások feltérképezésére is jól használhatóak. Amennyiben ugyanis a szavak vektorait különböző időszakokat reprezentáló korpuszok alapján készítjük el, azok összehasonlításával megkaphatjuk azok dinamikus változásait (Bamler és Mandt, 2017). A módszerrel többek között reprezentálhatóvá válhatnak egyes kulcsfogalmakat, illetve társadalmi csoportokat érintő változások az adott történelmi korszakok folyamatában (Hamilton és mtsai, 2016a; Garg és mtsai, 2018).

Mielőtt kiválasztottuk a jelen feladathoz legmegfelelőbb algoritmust, több megoldást is teszteltünk (Word2vec (Mikolov és mtsai, 2013a,b), FastText[5], GloVe (Pennington és mtsai, 2014)), amelyek közül a GloVe-t találtuk az adott vizsgálati cél szempontjból a legjobban működőnek. A módszerek kiértékelésekor elsősorban kvalitatív eszközökre támaszkodtunk és általunk kiválasztott kulcsfogalmak közelségét/távolságát vizsgáltuk. A Word2vec és a Glove hasonló eredményeket adott, a FastText betű alapú modellje, azonban történelmileg nagyon távol álló, de hasonló betűkből álló szavakat is egymáshoz közeli térbe helyezett el. A beágyazási módszerek és a tesztelési eljárások részleteinek tárgyalása e cikk keretein kívül esik, az alábbiakban a feldolgozási lépésekre összpontosítunk.

Annak céljából, hogy a feldolgozás számítási igényeit csökkentsük, első lépésben redukáltuk a vizsgálatba vont szavak számát: azokat, amelyek kevesebb mint ötször jelentek meg az a kiinduló szövegkorpuszban, töröltük.

Ezt követően a GloVe modellt a korpusz szavainak globális együtt-előfordulási statisztikája alapján tanítottuk. 10-es méretű ablakot használtunk az együttelőfordulási mátrix felépítéséhez, ami azt jelenti, hogy a célszó előtti és utáni 10 szót kezeltük a szó kontextusaként. 300 dimenziós beágyazási méretet választottunk, amely a legtöbb beágyazási algoritmus, köztük a word2vec és a GloVe

[5] https://fasttext.cc

alkalmazása esetében az alapértelmezett választás. Ennek megfelelően minden szót egy 300 valós számból álló vektor reprezentál. A tanításhoz az R-nyelvű text2vec csomagba (Selivanov és Wang, 2016) implementált GloVe algoritmust használtuk, ahol az iterációk maximális száma 10 volt.

A szavak hasonlóságát a szóvektorok koszinusz hasonlóságával számoltuk ki, ami a leggyakrabban használt metrika a beágyazáson alapuló elemzésekben. A maximális koszinusz hasonlóság 1, ami abban az esetben teljesül, ha két szóvektor orientációja teljesen azonos egymással, azaz pontosan ugyanabba az irányba mutatnak; 0, ha a vektorok merőlegesek egymásra; végül -1, ha a két vektor ellentétes irányba mutat, egymással 180 fokos szöget zár be.

A teljes korpuszt hat különböző, ugyanakkor részben átfedő időszakra osztottuk, ami fontos lépés volt az időbeli megközelítésünk szempontjából (Kozlowski és mtsai, 2018). Nyilvánvaló, hogy a szóbeágyazás minősége nagyban függ a korpusz minőségétől és méretétől. Mivel az időbeli változás tanulmányozása érdekében hat kisebb időszakra kellett felosztanunk az eredetileg viszonylag nagy méretű korpuszunkat, az egyes alkorpuszok mérete, amelyeken végül dolgoztunk, jelentősen kisebb volt. Ennek okán döntöttünk úgy, hogy az alkorpuszokban átfedő időszakokat is megengedünk. Fontos ugyanakkor hangsúlyozni, hogy, bár az időszakok átfedésben vannak, mindegyiknek megvan a saját, egymást nem átfedő vektortere, azaz minden egyes alkorpuszra készítettünk egy egyedi GloVe modellt. Az időszakok a következőek voltak: 1956-1965 (2 510 565 token), 1962-1968 (2 065 400 token), 1965-1972 (2 377 305 token), 1968-1976 (2 672 386 token), 1972-1982 (3 257 968 token), 1976-1989 (3 848 622 token).

Az alkorpuszok meghatározását követően minden időtartamra meghatároztuk ugyanannak a szónak az egyedi vektorát.

A beágyazások reprodukálhatóságát is teszteltük, a következőképpen: a tanítási folyamatot többször megismételtük egy-egy kiválasztott alkorpuszra, és csupán minimális eltéréseket tapasztaltunk a tesztelés eredményei között. Ugyanakkor úgy döntöttünk, hogy egy robusztus, statisztikailag megbízhatóbb eredmény elérése érdekében a következő megoldást alkalmazzuk: 20 beágyazási modellt készítünk minden időszakra, majd a 20 különálló vektor mindegyike esetében kiszámítjuk a kiválasztott fogalmakat reprezentáló vektorok közötti koszinusztávolságot. Végül, a 20 egyedi hasonlóság átlagával kapjuk a vizsgált fogalmak közötti tényleges hasonlósági mutatót (Antoniak és Mimno, 2018). Tesztjeink azt mutatták, hogy az alkalmazott megoldás az esetünkben stabil és megbízható eredményhez vezetett.

5. Eredmények

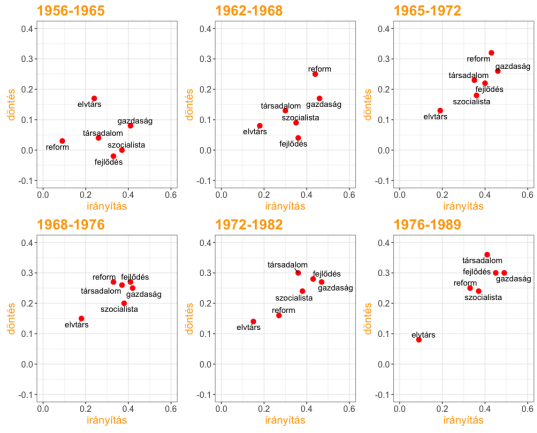
Elemzésünk során elvégeztük a vizsgált fogalmak gyakoriságának vizsgálatát, amelynek eredményét az alábbi szófelhők segítségével mutatjuk meg.

Vizsgálatunk jól mutatja a szavak gyakoriságának és ezen keresztül a korszak diskurzusában megjelenő kifejezések szerepének változását. Ezek közül kiemelendő a *döntés* szó gyakoriságának növekedése, amelynek következtében a húsz szó

2. ábra: A vizsgált szavak gyakorisága időszakonként

között az első időszak 19. helyéről a hatodik időszakban a 11. helyre lépett előre, míg az *elvtárs* szó az első periódus 6. helyéről a hatodik periódus 19. helyére esett vissza. Ha az első és az utolsó időszak leggyakoribb tíz-tíz kifejezését vetjük össze egymással, azt is kimutathatjuk, hogy az első tíz kifejezés egy kivételével ugyanaz, csak a szavak sorrendje változik. Egy lényeges kivételt tapasztalunk, az első időszakban még itt szereplő szavak közül kikerül az *elvtárs* kifejezés, ami az azonos politikai elveket vallókat jelöli és helyét a *társadalom* kifejezés veszi át, ami jól jelzi a diskurzus irányának megváltozását.

A következőkben a vizsgált fogalmak szemantikai összefüggéseit elemezzük. A 3. ábrán a fogalmak koszinusz közelségét mutatjuk meg az általunk meghatározott hat időszakban.



3. ábra: A vizsgált fogalmak koszinusz hasonlósága hat időszakban.

Eredményeink jól szemléltetik, hogy a *döntés* és *irányítás* szavak kapcsolata a vizsgált fogalmakkal jelentős változáson ment át. Az első két periódusban fogalmaink az *irányítás* szóhoz állnak közelebb, ami azt jelzi, hogy a diskurzus inkább direktívákat, mintsem alternatívákat is jelző döntési folyamatokra való utalásokat tartalmaz. Történetileg ez a magyar gazdaságpolitika azon időszaka, amikor a beruházások további finanszírozása és ezzel egyidejűleg az életszínvonal emelése komoly problémává vált. A források optimalizálása érdekében megindult a vállalatok összevonása, a „trösztösítés", az ország ipari vállalatait 15 nagyüzembe vonták össze. Felduzzadt a központi irányító apparátus, a helyi üzemegységek saját döntési lehetőségei viszont megszűntek.

Az 1960-as évek közepétől a kifejezések egyre közelebb kerülnek a *döntéshez*, ami a diskurzusban megjelenő alternatívákat jelzi. Az 1968-as reform, az „új gazdasági mechanizmus" értelmében nőtt az egyes vállalatok önállósága, a vállalatok saját megtermelt nyereségük egy részének beruházásáról maguk dönthettek. A mezőgazdaságban is növekedett a termelőszövetkezetek mozgástere, többek között engedélyezték számukra a jövedelmező melléküzemágak létesítését. A *társadalom* kifejezés például az első két időszakban meglehetősen gyenge kapcsolatban áll a *döntés* kifejezéssel (0.04-0.13), viszont erős a kapcsolata az *irányítással* (0.26- 0.30) a következő négy periódusban viszont, bár erős kapcsolata marad az *irányítással* (0.35-0.41), ugyanolyan erős kapcsolatba kerül a *döntéssel* (0.23-0.36)is. Érdekes tendenciát figyelhetünk meg a *reform* kifejezésnél is, amelynek a koszinusz közelsége az első időszakban mind a *döntéshez* (0.03), mind az *irányításhoz* (0.09) alacsony volt, a többi periódusban viszont magas, csak az 1970-es években tapasztalható a *reform-döntés* kapcsolatban némi gyengülés (0.16), ami magyarázható az 1968-as gazdasági reform ebben az időben történő leállításával. A *reform-irányítás* esetében minden időmetszetben 0.2 feletti küszöbértéket tapasztalhatunk. Szintén a korabeli gazdasági intézkedésekkel magyarázhatjuk a *gazdaság* kifejezéssel kapcsolatos eredményeinket. A *gazdaság* és az *irányítás* viszonyában minden vizsgált időszakban magas koszinusz küszöbértéket tapasztaltunk(0.41-0.49), míg a *döntéssel* csak az 1960-as évek második felétől. Ugyancsak figyelemre méltóak az *elvtárs* szóval kapcsolatos eredményeink, amely esetében minden periódusban mind a *döntéshez*, mind az *irányításhoz* viszonyítva 0.2 alatti küszöbértéket látunk.

6. Konklúzió

Dolgozatunkban a Kádár-korszak állampártjának hivatalos lapja alapján, történeti és szociológiai szempontok alapján kiválasztott kifejezések alapján az 1956 és 1989 közötti politikai diskurzus dinamikájának változásait vizsgáltuk. Elemzésünk fő célja az volt, hogy a korszak diskurzusának hosszanti, számítógépes és automatizált szövegelemzésére tegyünk kísérletet. Ehhez első lépésként összeállítottunk egy nagyméretű, 13 millió tokent tartalmazó, digitalizált korpuszt a *Pártélet* című folyóirat 379 számából. A korpusz nyers szövegeinek előfeldolgozását követően az adatokat szóbeágyazási módszerrel dolgoztuk fel. Ez a módszer lehetővé tette a vizsgált diskurzus néhány kulcsfogalma dinamikus változásainak

elemzését. Célunk az volt, hogy megvizsgáljuk a szóbeágyazás módszerének hasznosságát a történeti diskurzuselemzés számára. Elemzésünk megmutatta,hogy a hasonló korpuszok építése és kvantitatív elemzése hozzájárulhat a különféle történelmi korszakok diskurzusának mélyebb megértéséhez.

A jövőben célunk kutatásunkat más fogalmakra is kiterjeszteni, valamint ezen fogalmak dinamikus változásait is elemezni. A korpuszt további tisztítás és címkézés után terveink szerint elérhetővé tesszük a szélesebb kutatói közönségnek is. Végül, de nem utolsósorban, szeretnénk megvizsgálni a *Pártélet* és más célközönségnek szóló sajtótermékek diskurzusának dinamikai jellemzői közötti különbségeket és hasonlóságokat.

Köszönetnyilvánítás

A kutatást részben az Emberi Erőforrások Minisztériuma támogatta (TUDFO/47138-1/2019-ITM).

Hivatkozások

Antoniak, M., Mimno, D.: Evaluating the stability of embedding-based word similarities. Transactions of the Association for Computational Linguistics 6, 107–119 (2018)

Bamler, R., Mandt, S.: Dynamic word embeddings. In: ICML (2017)

Garg, N., Schiebinger, L., Jurafsky, D., Zou, J.: Word embeddings quantify 100 years of gender and ethnic stereotypes. In: Proceedings of the National Academy of Sciences of the United States of America (2018)

Gyáni, G.: A történelem mint emlék(mü). Kalligram, Budapest (2016)

Hamilton, W.L., Leskovec, J., Jurafsky, D.: Cultural Shift or Linguistic Drift? Comparing Two Computational Measures of Semantic Change. Proceedings of the Conference on Empirical Methods in Natural Language Processing. Conference on Empirical Methods in Natural Language Processing 2016, 2116–2121 (2016a)

Hamilton, W.L., Leskovec, J., Jurafsky, D.: Diachronic word embeddings reveal statistical laws of semantic change. ArXiv abs/1605.09096 (2016b)

Harris, Z.S.: Distributional structure. WORD 10(2-3), 146–162 (1954), https://doi.org/10.1080/00437956.1954.11659520

Jatowt, A., Duh, K.: A framework for analyzing semantic change of words across time. In: IEEE/ACM Joint Conference on Digital Libraries. pp. 229–238 (2014)

Kozlowski, A.C., Taddy, M., Evans, J.A.: The geometry of culture: Analyzing meaning through word embeddings. ArXiv abs/1803.09288 (2018)

Kulkarni, V., Al-Rfou', R., Perozzi, B., Skiena, S.: Statistically significant detection of linguistic change. In: Proceedings of WWW (2014)

Mikolov, T., Chen, K., Corrado, G.S., Dean, J.: Efficient estimation of word representations in vector space. CoRR abs/1301.3781 (2013a)

Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Proceedings of NIPS (2013b)

Pap, M.: A népitöl a szocialista demokráciáig A korai Kádár-korszak demokráciafogalma a pártfolyóiratok tükrében. Múltunk 1, 202–226 (2017)

Pennington, J., Socher, R., Manning, C.D.: Glove: Global vectors for word representation. In: Proceedings of EMNLP (2014)

Selivanov, D., Wang, Q.: text2vec: Modern text mining framework for r. Tech. rep. (2016), computer software manual](R package version 0.4. 0). Retrieved from https://CRAN. R-project. org/package= text2vec

Szabó, M.: A dolgozó mint állampolgár. Fogalomtörténeti tanulmány a magyar szocializmus három korszakáról. Korall 27, 151–171 (2007)

Xu, Y., Kemp, C.: A computational evaluation of two laws of semantic change. In: Proceedings of CogSci (2015)

Zsibrita, J., Vincze, V., Farkas, R.: magyarlanc: A Tool for Morphological and Dependency Parsing of Hungarian. In: Proceedings of RANLP (2013)