

Tényfeltáró szövegek dinamikájának kvantitatív és szemantikai elemzése

Quantitative and semantic analysis of the dynamics of investigative texts

Dr. TÓTH Erzsébet¹, Dr. TERDIK György², Dr. GÁL Zoltán³

Debreceni Egyetem, Informatikai Kar, Debrecen, Magyarország
4028 Debrecen, Kassai út 26., Magyarország,
{¹toth.erzsebet, ²terdik.gyorgy, ³gal.zoltan}@inf.unideb.hu,
<http://www.inf.unideb.hu>

Abstract

Our research aims to quantitatively and semantically analyze the dynamics of several dozen political and military texts from the Central Intelligence Agency (CIA). We characterized the texts and their constituent entities using 17-dimensional feature vectors (FVs) representing the distribution of parts of speech within a given text unit. Word frequencies conformed to Zipf's law, but in the investigative texts examined, the value of the power-law exponent ($b \approx 0.83$) deviated from the theoretical value of approximately 1. We divided the texts into 80 entities and calculated the Cosine, Hellinger, and Euclidean distances between the FVs of individual entities and the text FV-s. We defined text dynamics using metrics for Thematic Cohesion (TC) and Topic Drift (TD), along with their coefficient of variation. Our results indicate that Hellinger distance is the most suitable metric for capturing text dynamics. Based on the calculated thematic cohesion characteristics, cosine distance proved most effective at detecting variations within the text.

Keywords: textual dynamics; semantic and qualitative text analysis; Hellinger distance; cosine distance; Euclidean distance.

Kivonat

Kutatásunk célja a Central Intelligence Agency (CIA) politikai és katonai témájú, több tucat szövege dinamikájának kvantitatív és szemantikai elemzése. A szövegeket és az entitásait 17 dimenziós tulajdonságvektorokkal (FV) írtuk le, amelyek a szófajok megoszlását mérik az adott szövegegységben. A szógyakoriságok a Zipf-törvényhez illeszkedtek, de a vizsgált tényfeltáró szövegekben a hatványkitevő értéke ($b \approx 0,83$) eltért az elméleti, közel 1-es értéktől. A szövegeket felosztottuk 80 szövegentitásra, az egyes entítások FV-jei és a teljes szöveg FV-jei között koszinusz, Hellinger, és euklideszi távolságot számoltunk ki. A szövegek dinamikáját a tematikus kohézió (TC), a témaváltás intenzitás (TD) mérőszámokkal és azok variációs együtthatójával definiáltuk. Eredményül azt kaptuk, hogy a Hellinger hasonlóság alkalmas leginkább arra, hogy kifejezze a szövegek dinamikáját. A kiszámított tematikus kohézió jellemzők szerint a koszinusz távolság képes leginkább felismerni a szövegben lévő eltéréseket.

Kulcsszavak: szövegdinamika; szövegek szemantikai és kvantitatív elemzése; Hellinger távolság; koszinusz távolság; euklideszi távolság.

1. BEVEZETÉS

Dinamikusan fejlődő terület napjainkban a kvantitatív nyelvészet, amelynek fő célja, hogy olyan egyetemes törvényszerűségeket és matematikai modelleket találjon, amelyek az emberi természetes nyelvek működését magyarázzák. Ez a kutatási terület méri és számszerűsíti a különböző nyelvi egységeket, mint például a szavak gyakoriságát, a mondatok hosszát és az emberi hang különféle mintázatait.

A kvantitatív nyelvészet számos törvényt fogalmazott meg a szöveges adatok vizsgálata során, amelyek közül az egyik legalapvetőbb a Zipf törvény. Rijsbergen hangsúlyozza Luhn munkáját automatikus szövegelemzés területén, amely feltételezte, hogy a szavak előfordulási gyakoriságai alapján bizonyos szavak és mondatok kiemelhetők a szövegből, amelyekkel egy dokumentum tartalma leírható. Zipf törvénye kimondja, hogy

a szavak előfordulási gyakorisága fordítottan arányos a szavak rangsorával. Luhn ezt a törvényt nullhipotézisként használta, hogy meghatározza a szavak rangsorának felső és alsó küszöbértékét. A szavak szignifikanciáját úgy értelmezte, hogy azok mennyire képesek kifejezni a szöveg témáját. Az önkényesen meghatározott küszöbértékek segítségével elhagyta a ritka és a gyakori szavakat a szavak rangsorából. Ilyen módon be tudta azonosítani azokat a fontos szavakat, amelyek a szöveg tartalmát, illetve témáját írják le [1-3]. Az ókori görög irodalmi szövegek esetében a Zipf törvénnyel kapcsolatos kutatási eredményeinket lásd [4-5]-ben; a horvát és az angol nyelvű jogi korpusz esetében ugyanezt lásd [6-7]-ben.

A szövegdinamika a narratíván belül zajló folyamatokat jelenti, amelyek segítségével a szerző eléri céljait, míg az olvasói dinamika az ezekre adott kognitív, etikai, affektív és esztétikai reakciókat foglalja magában [8]. Jelen vizsgálatban a szövegdinamika alatt a szövegben bemutatott események és történések egymásra épülő sorozatát értjük, amely a cselekmény fokozatos kibontakozását eredményezi.

A dolgozat további részének felépítése a következő: a CIA (Central Intelligence Agency) szövegek dinamikájának elemzési módszerei kerülnek bemutatásra a második részben. Az összefoglalás és a kapott eredmények a dolgozat végén vannak kifejtve.

2. AZ ALKALMAZOTT MÓDSZEREK ÉS EREDMÉNYEK

A szófaji címkézés (POS-címkézés) a természetes nyelvfeldolgozás (NLP) egyik alapvető feladata, amely során a szöveg minden egyes szavát ellátják a megfelelő szófaji címkével (például főnév, ige, melléknév stb.). A szófaji címkézés kulcsfontosságú számos NLP-feladatnál, például a szintaktikai elemzés, az információkinyerés és a hangulatelemzés szempontjából. A továbbiakban egy egyszerűsített áttekintést adunk arról, hogyan működik egy tipikus szófaji címkéző algoritmus:

Tokenizálás: A bemeneti szöveget először különálló szavakra vagy tokenekre bontják. Ez a lépés elengedhetetlen ahhoz, hogy a szöveget kezelhető egységekre lehessen osztani.

Jellemzők kinyerése: Kinyerik az egyes szavak különböző jellemzőit, amelyek segítenek meghatározni a szófajukat. Ezek a jellemzők lehetnek maga a szó, a mondatbeli környezete, a szomszédos szavak, a toldalékok és egyebek.

Gépi tanulási modellek: A szófaji címkézés gyakran gépi tanulási algoritmusokat alkalmaz, különösen felügyelt tanulási módszereket, mint például a rejtett Markov-modellek (HMM-ek) vagy a feltételes véletlen mezők (CRF-ek), illetve mélytanulási modelleket, mint a rekurrens neurális hálózatok (RNN-ek) vagy a Transformer-alapú architektúrák.

Felügyelt tanulás: A felügyelt tanulási megközelítések során az algoritmust egy címkézett adathalmazon tanítják be, amelyben minden szó el van látva a helyes szófaji címkével. A modell megtanulja megjósolni az egyes szavak legvalószínűbb szófaját a tanító adatokból kinyert jellemzők alapján.

Felügyelet nélküli tanulás: A felügyelet nélküli tanulási módszerek olyan technikákat alkalmaznak, mint a klaszterezés vagy az eloszlásbeli hasonlóság, hogy csoportosítsák a hasonló környezetben előforduló szavakat. Bár a szófaji címkézéssel kevésbé elterjedtek, a felügyelet nélküli módszerek hasznosak lehetnek olyan esetekben, amikor kevés a címkézett adat.

Értékelés: A szófaji címkéző algoritmus teljesítményét olyan mérőszámok alapján értékelik, mint a pontosság (accuracy), a precizitás (precision), az újrachívás (recall) és az F1-érték. A modell előrejelzéseit összevetik a tesztadatkészletben található „arany standard” (hiteles) annotációkkal.

Iteratív finomítás: A szófaji címkéző modelleket gyakran iteratív módon finomítják a tesztadatokon nyújtott teljesítményük alapján. Ez magában foglalhatja a modellparaméterek finomhangolását, a szakterület-specifikus tudás beépítését vagy a tanító adatok további példákkal való bővítését.

Összességében a szófaji címkéző algoritmusok nyelvi ismereteket, statisztikai modelleket és gépi tanulási technikákat használnak fel egy adott szöveg szófajainak elemzésére és címkézésére. Bár egyetlen megközelítés sem tökéletes, az NLP területén elért fejlődés folyamatosan javítja a szófaji címkéző rendszerek pontosságát és hatékonyságát.

2.1. A vizsgált szöveghalmaz jellemzői

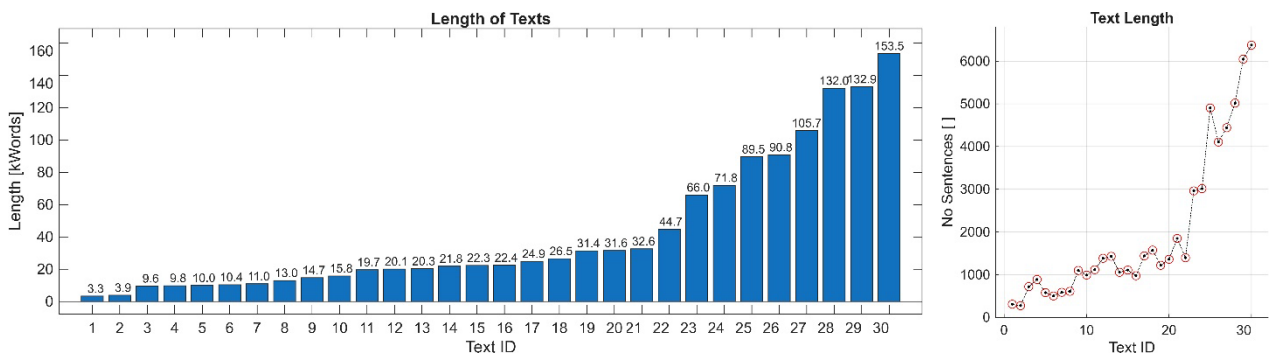
A Central Intelligence Agency (CIA) szervezet Digitális Könyvtárából 30 darab szöveget töltöttünk le, amelyek mintegy 50 évvel ezelőtt történt politikai és katonai eseményeket mutatnak be. A hasonló témájuk miatt a szövegek homogén korpuszt alkotnak. A vizsgált szövegeket az 1. táblázat tartalmazza a szavak száma szerinti növekvő sorrendben.

Az elemzett CIA szövegek

1. táblázat

	Szerző(k)	Cím
1.	Central Intelligence Agency	Memorial Wall Publication
...	...	...
5.	Central Intelligence Agency	Our First Line of Defense: Presidential Reflections on US Intelligence
...	...	...
15.	Michael Warner, J. Kenneth McDonald	US Intelligence Community Reform Studies since 1947
...	...	...
30.	L. Britt Snider	The Agency and the Hill: CIA's Relationship with Congress, 1946-2004

A vizsgált szövegek terjedelme 3,3 és 153,5 ezer szó között változik, és 70%-uk 35 000 szónál rövidebb. Ez lehetőséget biztosít a szövegméretek három lineáris értéktartományban történő vizsgálatára (lásd 1. ábrát bal oldalon).



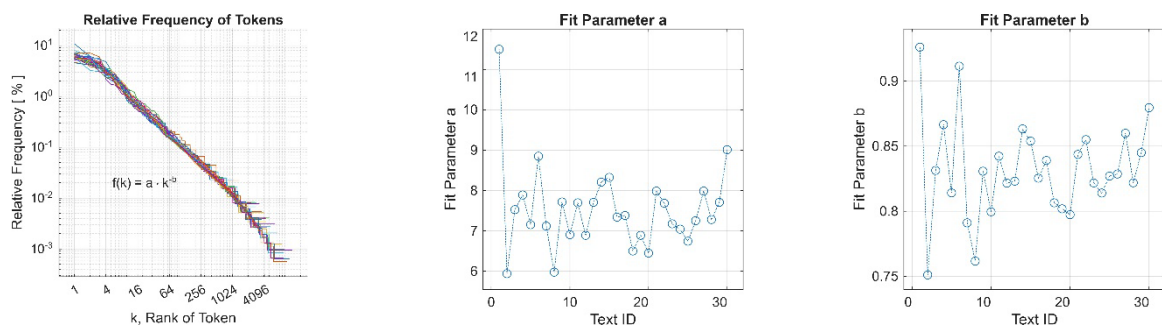
1. ábra. Az elemzett szövegek szavainak (bal), illetve mondatainak száma (jobb).

A vizsgált szövegek nagy része kevesebb mint 2 000 mondatból áll. Minimálisnak tekinthető a 6 000 mondatnál hosszabb terjedelmű szövegek száma (lásd 1. ábrát jobb oldalon).

Az elemzett szövegek szavainak gyakorisága az előfordulási szám szerint csökkenő sorrendben rendezve illeszkedik a Zipf-törvényhez, de speciális hatványkitevő értékekkel. Mint ismeretes, a Zipf-törvény az alábbi állítja:

$$f(k) = a \cdot k^{-b} \quad (1)$$

ahol f a szavak gyakorisága, k a sorszám, míg (a, b) illesztési paraméterek. Megjegyezzük, hogy esetünkben a tényfeltáró szövegeknél a hatványkitevő $b \approx 0,83$, ami a törvény szerint állított $b \approx 1$ értéktől eltér.



2. ábra. A szavak előfordulási gyakorisága Zipf-törvény szerint: gyakoriságok (bal), „a” illesztési paraméter (közép), „b” illesztési paraméter (jobb).

A CIA szövegeket tokenekre (szószetonokra) bontottuk, majd az egyes tokeneket a szöveggörnyezet figyelembevételével a megfelelő szófaji kategóriába soroltuk. A tokenek szófaji besorolását a MATLAB programozási környezet beépített függvényeivel végeztük el.

A tulajdonságvektor token kategóriái

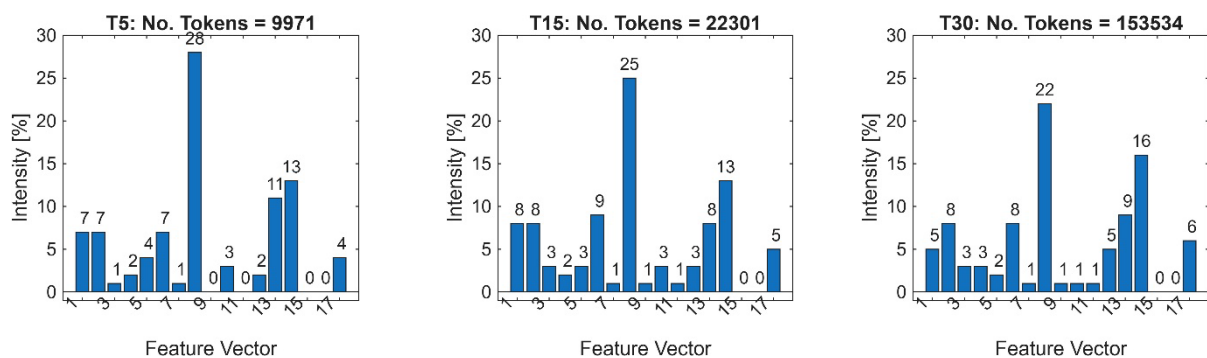
2. táblázat

ID	Token kategória	ID	Token kategória	ID	Token kategória	ID	Token kategória
1	melléknév	6	elválasztószó	11	viszonyszó	16	szimbólum
2	értelmező	7	indulatszó	12.	névmás	17	ige
3	határozószó	8	főnév	13	tulajdonnév		
4	segédige	9	számnév	14	írásjel		
5.	kötőszó	10	egyéb	15	alárendelőszó		

Vizsgálatunk során összesen 17 tokenkategóriát alkalmaztunk (lásd 2. táblázat), amelyek közül az „egyéb” kategória a nemzeti nyelvek sajátos, külön kategóriába nem sorolható jellemzőinek kezelését teszi lehetővé. Adott szöveg FV (Feature Vektor) tulajdonságvektorának az egyes szófajok relatív gyakoriságát tekintjük:

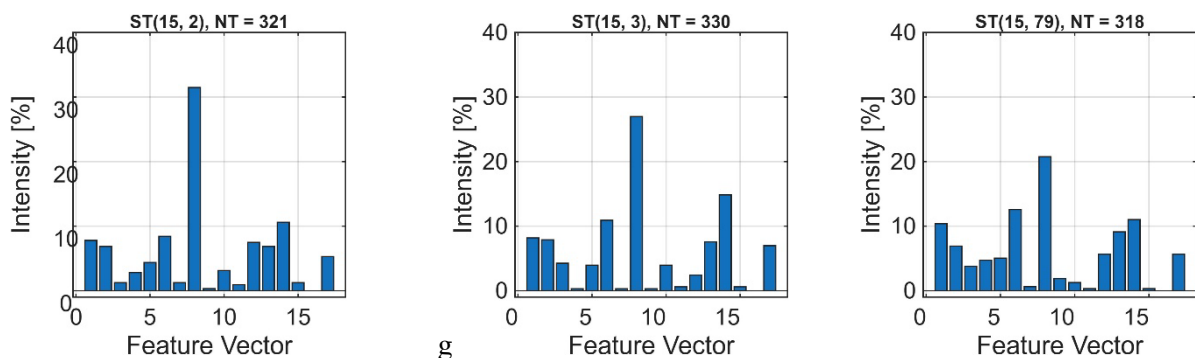
$$FV = [POS_1, POS_2, \dots, POS_{17}] / N \in [0, 1]^{17} \quad (2)$$

ahol POS_i az i -edik szófajhoz tartozó szavak száma és N a szövegrész szavainak száma. Megjegyezzük, hogy a FV tulajdonságvektor tetszőleges méretű szöveg esetén meghatározható, így nemcsak egy teljes dokumentum, hanem annak egy tetszőleges darabja, mint szövegintitás rendelkezik FV tulajdonságvektorral. Ezt később is fel fogjuk használni az egyes szövegrészek közötti eltérések jellemzésére. Három szöveg (T5, T15, T30) esetén a tulajdonságvektorokat a 3. ábra szemlélteti.



3. ábra. Három példaszöveg globális tulajdonságvektora: T5, T15, T30.

Megállapítható, hogy azonos témájú szövegek is különböző tulajdonságvektor mintázatokat mutathatnak, amelyeket többek között a nyelv sajátosságai, a szöveg stílusa és a szerző személye befolyásolhat. Az intenzitás értékeket a $[0, 100]$ intervallumban lefelé kerekítettük. Mindhárom diagramon a főnévi token kategória mutatja a legnagyobb intenzitást, míg a legritkábban előforduló kategóriák a számnév, a viszonyzó, az alárendelőszó és a szimbólum. A T15 esetén $m = 80$ egyenlő szószámú szövegintításra bontás után a ST2, ST3 és ST79 entitások tulajdonságvektorait a 4. ábra mutatja be.



4. ábra. A T15 szöveg 2 (bal), 3 (közép), 79 (jobb) entitásainak tulajdonságvektora.

NT az entitáson belüli tokenek számát jelenti

Szembevetendő, hogy a szöveg elején egymás után sorba következő entitások FV-jei között nincs jelentősebb különbség. Ezzel szemben a szöveg végén lévő utolsó előtti entitás FV-je nagy mértékben eltér egyedi

mintázatát tekintve az elől található 2-es és a 3-as entitás FV-jétől. Általában a szövegek legelső, illetve legutolsó entitása tartalomjegyzéket, valamint irodalomjegyzéket foglal magában. Mivel ezek lényegesen eltérnek a tartalmi mondatoktól, ezért nem meglepő az érintett FV-k közötti nagy különbség. Megállapítható, hogy a T15-ös szöveg entitásainak FV-jei eltérnek egyedi mintázatukat tekintve a szülő szöveg FV-jétől.

Két tetszőleges FV tulajdonságvektor között távolságmétrikával számszerűsíthető az eltérés. Esetünkben három metrikát alkalmaztunk: Hellinger, koszinusz és euklideszi távolságot. Ezeket az alábbiak szerint definiáljuk:

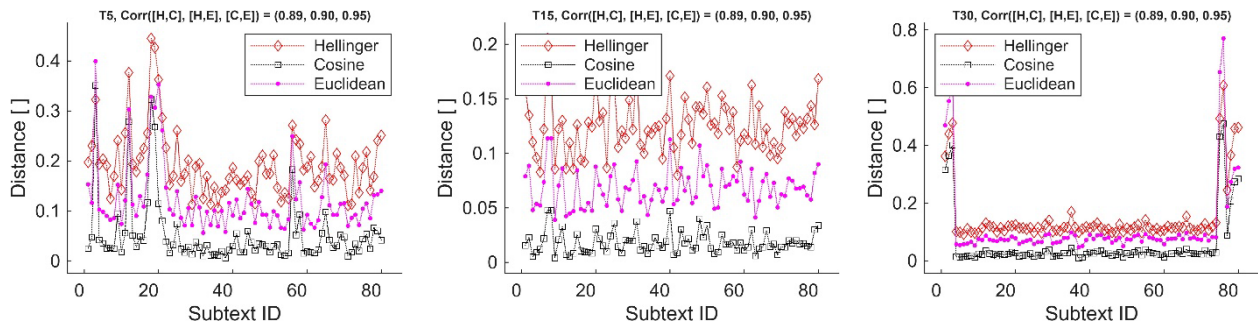
$$d_H(\bar{u}, \bar{v}) = \frac{1}{2}(\sqrt{\bar{u}} - \sqrt{\bar{v}})(\sqrt{\bar{u}} - \sqrt{\bar{v}})^T \quad (3)$$

$$d_C(\bar{u}, \bar{v}) = 1 - \frac{\bar{u} \cdot \bar{v}}{\|\bar{u}\| \|\bar{v}\|} \quad (4)$$

$$d_E(\bar{u}, \bar{v}) = \sqrt{(\bar{u} - \bar{v})(\bar{u} - \bar{v})^T} \quad (5)$$

ahol d_H, d_C, d_E rendre az $\bar{u}, \bar{v} \in [0, 1]^{17}$ tetszőleges tulajdonságvektorok közötti Hellinger, koszinusz, illetve euklideszi távolság, $\bar{u}^T$ az $\bar{u}$ vektor transzponáltja, $\cdot$ a skaláris szorzás, $\|\bar{u}\|$ pedig az $\bar{u}$ vektor hosszát jelöli. Megjegyezzük, hogy mivel az $FV = \bar{u}$ tulajdonságvektor relatív gyakoriságokat tartalmaz, ezért minden esetben az $\bar{u}$ vektor koordinátáinak összege 1, vagyis bármely tulajdonságvektor a Hellinger metrika szerinti 17 dimenziós térben a 0,5 sugarú gömb felületére mutat, azaz $d_H(\bar{u}, 0) = \frac{1}{2}\sqrt{\bar{u}}\sqrt{\bar{u}}^T = 0,5$. Ebből következik, hogy bármely két tulajdonságvektor közötti Hellinger távolság leg több 1 lehet. Ez akkor következhet be, ha a gömb középpontjára szimmetrikus pontokra mutat a két vektor.

Adott szöveg esetén m darab egyenlő hosszúságú entitásnak az $FV_i, i = 1, 2, \dots, m$; $FV_i \in \mathbb{R}^{17}$ tulajdonságvektorai és a szöveg tulajdonságvektora közötti d_H, d_C, d_E távolságokat az 5. ábrán láthatjuk.



5. ábra. Az entítások és a teljes szöveg tulajdonságvektorai közötti távolságok a T5 (bal), T15 (közép) és T30 (jobb) szövegek esetén. Az egyes távolságpárok közötti korreláció az ábra címében szerepel.

Megfigyelhető, hogy minden esetben $d_H \geq d_E \geq d_C$, ami abból adódik, hogy az egynél kisebb pozitív számokat a négyzetgyök függvény erőteljesebben felnagyítja 0 közelében, mint az identitás függvény. A koszinusz távolság ugyanakkor kisebb, mint az identitás függvény (mivel $1 - x \leq \cos(x)$, ha $x \in [0, 1]$).

2.2. Tematikus kohézió (Thematic Cohesion), témaváltás intenzitás (Topic Drift) és szövegdinamika

Mivel a $d(FV_i, FV) \in [0, 1], i = 1, \dots, m$ ezért az $s_i = d(FV_i, FV) \in [0, 1], i = 1, \dots, m$ értékek a tulajdonságvektorok hasonlósága (s – similarity). Így két hasonló vektor közötti távolság kicsi, míg két nemhasonló, vagyis eltérő vektor közötti távolság nagy. A tematikus kohézió (thematic cohesion) adott szöveg esetében az egyes szövegentítások és a teljes szöveg FV-jei közötti hasonlóságok átlaga, míg a témaváltás intenzitás (topic drift) ezen FV-k közötti hasonlóságok szórása. A kettőből számolt variációs együttható kevésbé függ a szöveg méretétől, így a szöveg dinamikáját az alábbiak szerint definiáljuk:

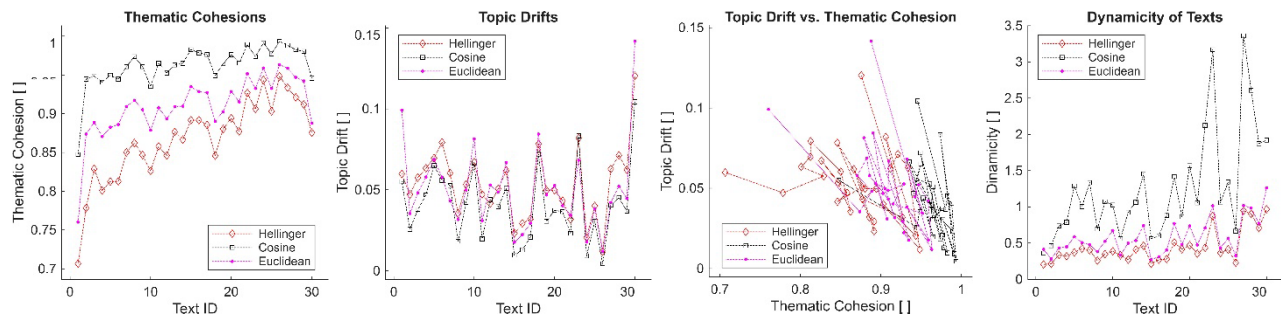
$$\mu = \frac{1}{m} \sum_{i=1}^m d(FV_i, FV) \quad (6)$$

$$TC = 1 - \mu \quad (7)$$

$$TD = \sqrt{\frac{\sum_{i=1}^m (d(FV_i, FV) - \mu)^2}{m}} \quad (8)$$

$$D = \frac{TD}{1-TC} \quad (9)$$

ahol $d(FV_i, FV)$ adott szöveg i -edik entitása és a szöveg közötti távolság. Az elemzett szövegekre kiszámolt jellemzők és ezek közötti összefüggést a 6. ábra mutatja be. A kiszámított tematikus kohézió jellemzők rávilágítanak arra, hogy a koszinusz távolság tükrözi leginkább a szövegen belüli eltéréseket, míg a Hellinger távolság a legkisebb értékeket adja. A kalkulált témaváltás intenzitás jellemzői kevésbé függenek az alkalmazott távolságmérika típusától (ld. 6.b ábra).



6. ábra. Tematikus kohézió (a), Témaváltás intenzitás (b) az elemzett szövegek esetében.
A tematikus kohézió és a témaváltás intenzitása közötti szórásdiagram (c),
illetve a szövegek dinamikája (d).

A szövegekre kiszámított két jellemző közötti szórásdiagram mindhárom távolságnak az átlagértéktől való eltérését mutatja a 6.c ábra. Itt látható, hogy a távolságok szerinti állapotpontok klasztereződést mutatnak, azaz beazonosíthatók a három távolság szerinti csoportosulások. A szövegek dinamikája a $d(FV_i, FV)$, $i = 1, \dots, m$ távolságok variációs együtthatója. Ennek nyomon követésére és kifejezésére leginkább a Hellinger-távolság alkalmas, mivel kevésbé függ a szöveg méretétől. Ez jól megfigyelhető a 6.d ábrán, ahol az értékek a $(0, 1)$ intervallumban maradnak, vagyis az entitások viszonylag szabályosabbak, egyenletesebbek és kiszámíthatóbbak, mint egy teljesen véletlenszerű (Poisson) folyamat esetében. Ez várhatóan a szövegek tényfeltáró jellegéből származik.

3. ÖSSZEFOGLALÁS ÉS KÖVETKEZTETÉSEK

Vizsgálatunk 30 darab, a CIA Digitális Könyvtárából származó, hasonló témájú politikai és katonai szövegre terjedt ki. A szövegekből kinyertük a tokeneket, majd 17 különböző szófajba soroltuk, és az egyes tokenkategóriák relatív gyakoriságából 17 dimenziós tulajdonságvektort (FV) képeztünk. Elemzésünk megmutatta, hogy az azonos témájú szövegek FV-jeinek egyedi mintázatai is eltérhetnek, amit többek között a nyelv, a stílus és a szerző sajátosságai befolyásolhatnak. A szövegek belső szerkezetének vizsgálata ugyanakkor azt jelezte, hogy az egyes szövegentitások és a teljes szöveg tulajdonságvektorai eltérnek egymástól mintázatukat tekintve. A szógyakoriságok a Zipf-törvényhez illeszkedtek, de a vizsgált tényfeltáró szövegekben a hatványkitevő értéke ($b \approx 0,83$) eltért az elméleti, közel 1-es értéktől.

A szövegrészek FV-jei és a teljes szöveg FV-jei között Hellinger, koszinusz és euklideszi távolságot számoltunk, amelyek korrelációban voltak, miközben minden vizsgált esetben $d_H \geq d_E \geq d_C$ sorrend adódott. A tematikus kohéziót (TC) a szövegentitások FV-jei és a teljes szöveg FV-jei közötti hasonlóságok átlaga alapján, a témaváltás intenzitását (TD) pedig ezek szórása alapján határoztuk meg. A szövegek dinamikáját a variációs együttható mérőszámmal definiáltuk. A kapott tematikus kohézió jellemzők szerint a koszinusz hasonlóság tükrözte leginkább a szövegekben lévő eltéréseket, míg a szövegen belüli dinamika nyomon követésére a Hellinger hasonlóság bizonyult a legalkalmasabbnak.

KÖSZÖNETNYILVÁNÍTÁS

Ezt a munkát a Debreceni Egyetem QoS-HPC-IoT Laboratórium és a TKP2021-NKTA-34 projekt támogatta. A TKP2021-NKTA-34 számú projekt a Kulturális és Innovációs Minisztérium Nemzeti Kutatási Fejlesztési és Innovációs Alapból nyújtott támogatásával, a TKP2021-NKTA pályázati program finanszírozásában valósult meg.

IRODALMI HIVATKOZÁSOK

- [1] Rijsbergen, van C. J.: Information Retrieval. 2nd ed. London: Butterworths, 1979. URL: <https://www.dcs.gla.ac.uk/Keith/Preface.html> (Utolsó letöltés: 2026-09-15)
- [2] Baayen, R. H.: Word Frequency Distribution. Dordrecht: Kluwer Academic Publishers, cop. 2001.
- [3] Desagulier, G.: Statistics for text analysis. In: International Encyclopedia of Language and Linguistics, 3rd Edition, Nesi, H., Milin, P. (eds.), vol. 14. Elsevier, 2026.
- [4] Gál, Z., Tóth, E.: Deep learning-based analysis of ancient Greek literary texts: A statistical model based on word frequency for the classification of texts. In: Proc. of the 12th IEEE International Conference on Cognitive Infocommunications: CogInfoCom 2021. Ed.: Jan Nikodem, Ryszard Klempous, IEEE-INST Electrical Electronics Engineers INC, Piscataway, 529-535, 2021. ISBN: 9781665424950
- [5] Gal Z., Tóth E. Deep Learning-Based Analysis of Ancient Greek Literary Texts in English Version: A Statistical Model Based on Word Frequency and Noise Probability for the Classification of Texts. Infocommunications Journal, Joint Special Issue on Cognitive Infocommunications and Cognitive Aspects of Virtual Reality, 2024, 2–11, <https://doi.org/10.36244/ICJ.2024.5.1>.
- [6] Tóth E., Gál Z. Multilabel clustering analysis of the Croatian-English parallel corpus based on Latent Dirichlet Allocation Algorithm. In: Proc. of the 14th IEEE International Conference on Cognitive Infocommunications: CogInfoCom 2023, Piscataway (NJ): IEEE-INST Inc, 2023, 25–32, ISBN 97983503256
- [7] Tóth E., Gal Z. Optimizing Text Clustering Efficiency through Flexible Latent Dirichlet Allocation Method: Exploring the Impact of Data Features and Threshold Modification. Infocommunications Journal, Joint Special Issue on Cognitive Infocommunications and Cognitive Aspects of Virtual Reality, 2024, 58–66, <https://doi.org/10.36244/ICJ.2024.5.7>
- [8] Phelan, J., Rabinovitz, P. J. Narrative as Rethoric. In: Herman, D. et al. (eds.): Narrative Theory: Core Concepts and Critical Debates, Columbus, The Ohio State University, 2012, 3-7.