

## Szó, beszéd – avagy hogyan kommunikálunk egymásról

Üveges István<sup>1,2</sup>, Szabó Martina Katalin<sup>2,3</sup>, Vincze Veronika<sup>4</sup>

<sup>1</sup>Szegedi Tudományegyetem, Nyelvtudományi Doktori Iskola  
6722 Szeged, Egyetem utca 2.

<sup>2</sup>Szegedi Tudományegyetem, Informatikai Intézet  
6720 Szeged, Árpád tér 2.

<sup>3</sup>Társadalomtudományi Kutatóközpont, CSS-RECENS  
1097 Budapest, Tóth Kálmán utca 4.

<sup>4</sup>MTA-SZTE Mesterséges Intelligencia Kutatócsoport  
6720 Szeged, Tisza Lajos körút 103.  
uvegesistvan898@gmail.com  
{martina,vincze}@inf.u-szeged.hu  
Szabo.Martina@tk.hu

**Kivonat** A dolgozatban egy nagy méretű, magyar beszélt nyelvi adatbázis (HuTongue) segítségével azt vizsgáljuk, hogyan kommunikálnak egymásról a diskurzusok résztvevői, és megszólalásaikhoz milyen szentimentek és emóciók kapcsolódnak. A vizsgálathoz használt korpusz beszélt nyelvi szövegeket tartalmaz, amely manuálisan leiratozott és annotált. A korpusz létrehozásának elsődleges célja az volt, hogy a kutatócsoport elsősorban szociológiai kutatási kérdéseikhez megfelelő vizsgálati anyagot biztosítson. A csoport arra a kérdésre keresi a választ, hogy zárt közösségeken belül az informális kommunikáció és a megbecsültség hogyan befolyásolja a közösség működését és normarendszerét. A korpusz annotációja lehetővé teszi a csoporttagok egymásra vonatkozó közléseinek a vizsgálatát más szemantikai és pragmatikai sajátságokkal összefüggésben. A jelen dolgozatban kifejezetten ezekre az interakciókra fókuszálunk, és azokat a korpusz szentiment- és emóciótartalmaival összefüggésben elemezzük.

**Kulcsszavak:** magyar, diskurzus, korpuszelemzés, nyelvi jellemzők, kommunikáció, hálózat, hálózatelemzés, csoportkohézió, pletyka, szentiment- és emócióelemzés

### 1. Bevezetés

A dolgozatban a HuTongue korpusz alapján olyan magyar diskurzusok szemantikai és pragmatikai sajátságait tárgyaljuk, amelyben a résztvevők közlései valamely jelen nem levő személyre vagy személyekre irányulnak.

A HuTongue korpusz szövegei hétköznapi szituációkban, külső ingerektől elszigetelt környezetben keletkeztek (Galántai és mtsai, 2018; Gulyás és mtsai,

2018). A korpusz létrehozásának elsődleges célja az volt, hogy megfelelő kutatási anyagot nyújtson a pletyka elemzéséhez az emberi interakcióban, ugyanakkor a benne foglalt anyagok keletkezési körülményei, mennyisége és feldolgozásának módja miatt az adatbázis számos társadalomtudományi és nyelvészeti vizsgálat elvégzésére lehetőséget adhat.

Bár a pletyka fogalmának a szakirodalom alapján több, különféle megközelítése is ismert (Emler, 1994; Grosser és mtsai, 2012), a HuTongue korpusz annotálásához a résztvevő szakértők egy tág definíciót határoztak meg: az annotációs fázis során minden olyan megnyilatkozás ebbe a kategóriába kellett, hogy tartozzon, amelynek tárgya egy vagy több jelen nem levő személy volt. Ennek a meghatározásnak köszönhetően egy olyan korpusz jött létre, amely lehetővé teszi az emberi kommunikáció e szegmensének a széleskörű, valamely definíció által nem korlátozott vizsgálatát.

Az elmúlt évtizedekben több beszélt nyelvi korpusz keletkezett több nyelven (Crowdy, 1993; Hemphill és mtsai, 1990; Maekawa és mtsai, 2000; Oostdijk, 2000; Van Bael és mtsai, 2007). Neuberger és mtsai (2014) alapján, közöttük találunk agglutináló nyelveket reprezentáló korpuszokat is, például török (Mengusoglu és Deroo, 2001) és finn (Seppänen és mtsai, 2003). Azonban a nyilvánvaló változatoság ellenére is csak néhány olyan korpusz létezik, amely a hangzó szövegek mellett azok gépelt leiratait, valamint annotációs szinteket is tartalmaz. Ez az átírási eljárás magas munkaerő- és költségigényével magyarázható.

Különösen csekély a magyar beszélt nyelvet reprezentáló korpuszok száma (Bodó és mtsai, 2017; Gósy, 2013). Közülük a Budapesti Egyetemi Kollégiumi Korpusz (BEKK) (Bodó és mtsai, 2017) esetében az interakciókat a résztvevők saját telefonjaikon rögzítették, ezért tulajdonképpen társalgásrészleteket tartalmaz. A BEszélt nyelvi Adatbázis (BEA) (Gósy, 2013) létrehozóinak fő célja az volt, hogy fonetikai (és nem szemantikai vagy pragmatikai) vizsgálatokat tegyen lehetővé, ezért a korpuszban alkalmazott annotációt is ennek megfelelően alakították ki. (A kutatócsoport egy újabb, hasonló korpuszról I.Szabó és mtsai (2021).) A fentebbi korpuszok tehát nem reprezentálják tökéletesen a teljes élőbeszédet, illetve feldolgozásuk módjával, annotációjukkal elsősorban fonetikai, és nem egyéb (pl. szemantikai vagy pragmatikai) vizsgálatokat kívánnak támogatni.

Mindezekkel összefüggésben a HuTongue hiánypótlónak tekinthető, hiszen nemzetközi szinten is rendkívül ritka a nagyméretű, informális beszélgetéseket tartalmazó korpusz. Tudomásunk szerint a HuTongue az első olyan magyar beszélt nyelvi korpusz, amely több száz órányi beszélgetésből áll és tartalmazza az összes beszélgetés részletesen annotált leiratát. Jelen dolgozatban a korpusz újabb, szűrt változatára támaszkodunk, melynek részletesebb bemutatását I.Vincze és mtsai (2021).

Mivel megfelelő magyar nyelvű adatbázisok hiányában eddig nem volt lehetőség a magyar beszélt nyelv beható korpuszalapú vizsgálatára, a jelen dolgozatban tett megállapítások értékes adalékul szolgálhatnak a magyar nyelvű beszélgetések néhány aspektusának vizsgálatához.

## 2. Kapcsolódó irodalom

Magunk a jelen dolgozattal azokhoz a kutatásokhoz szeretnénk hozzájárulást tenni, amelyek azokat a diskurzusokat vizsgálják, ahol egy adott csoport tagjai jelen nem levő személyekről beszélgetnek (részben tehát a pletykakutatásokhoz).

Az emberi informális kommunikáció jelentős részét, egyesek szerint kétharmadát, jelen nem levő, más személyekről folytatott értékelő tartalmú beszélgetés teszi ki (Dunbar, 1996, 2004). Jó pletykát hallani és pletykálni mindenki szeret, mégis magához a pletykához a köznapi értelemben pejoratív konnotációkat fűzünk, elítéljük azt. Miért létezik akkor, és miért olyan elterjedt a pletyka? Miért használunk ki szinte minden alkalmat arra, hogy másokat a hátuk mögött kibeszéljünk? (Galántai és mtsai, 2018).

Az erre vonatkozó ismeretanyag viszonylag szerény, és kvalitatív tapasztalatokra hagyatkozik. A pletyka mérését korábbi kutatások többnyire kvalitatív vagy kérdőíves kvantitatív módszerekkel (pl. Eckhaus és Ben-Hador (2019), írott szövegek alapján (Mitra és Gilbert, 2012) vagy megfigyelésen keresztül végezték (Dunbar, 2004).

A pletyka számos diszciplína által megközelített, ám eltérő mélységben vizsgált és különféleképpen meghatározott jelenség (Kisfalusi és Takács, 2018). Ezzel összefüggésben a pletyka fogalmának a szakirodalom alapján többféle megközelítése is ismert (Emler, 1994; Grosser és mtsai, 2012). A pletyka hagyományos definíciója szerint jelen nem levő személyről vagy személyekről folytatott értékelő tartalmat hordozó beszélgetéseket takar, melyekben egy értékelő tartalmat megfogalmazó személy és legalább egy hallgatója jelen van (Emler, 1994; Grosser és mtsai, 2012).

Míg a legtöbb tanulmány a pletyka egy vagy két céljára összpontosít, valójában – mint egy sokoldalú szociológiai jelenség – sok más társadalmi célja is lehetséges (Giardini és Wittek, 2019; Pápay, 2019). Azok az esetek például, amikor az üzenet feladója titkot árul el a pletyka vevőjének a pletyka tárgyáról, a diskurzus a feladó és a vevő közötti kapcsolat erősödhet, miközben a tárgy reputációja romlik. Ezzel összefüggésben, bár a pletyka fogalma gyakran negatív konnotációt hordoz, azonban fontos szerepet tölthet be az információáramlásban, a személyek közti kapcsolatok megerősítésében, a csoportnormák fenntartásában és betartatásában vagy szelepként szolgálhat a felgyülemlett negatív érzelmek „kiengedésében” (Grosser és mtsai, 2012).

Ami a HuTongue korpusz eddigi elemzési eredményeit illeti, egy korábbi dolgozatban a korpusz topikmodellezése történt meg az annotációval összefüggésben (Pápay és mtsai, 2018). A kutatók azt vizsgálták, hogy a pletykataggal ellátott, valamint pletykaként nem annotált diskurzusok topikjaiban milyen eltérések mutatkoznak. Egy a csoport egy másik dolgozatában a pletyka és a nem pletyka szövegek szófaji eloszlási különbségeit vizsgálta, valamint azt, hogy milyen grammatikai konstrukciókkal (névmások és igei morfémák) utal a pletyka feladója a pletyka tárgyára az üzenetében (Pápay, 2019).

A jelen dolgozat főbb kutatási kérdései a következők: Egyrészt azt szeretnénk feltérképezni, hogy azokban a diskurzusokban, amikor a csoporttagok egymásról beszélnek, mennyire dominálnak a különböző érzelmek és értékelő tartalmak.

Ezt a kutatási kérdést azért is tekintjük fontosnak, mert rámutathat a diskurzus motivációs tényezőire. A fentebbivel összefüggésben szeretnénk megtudni, hogy az, amikor a diskurzusok résztvevői egymásról beszélnek, mennyire gyakran teszik ezt negatívan. Megállapítható-e, hogy az egymásról való kommunikációt (beleértve a pletykát) javarészt negatív megállapítások, értékelések alkotják. Harmadrészt, szeretnénk megismerni az egymásról beszélés és a csoportdinamika összefüggéseit. Vajon hogyan mutat rá az egymásról beszélés a társas kapcsolatok minőségére, szorosságára?

A fentebbiekkel kapcsolatosan azt feltételezzük, hogy a beszéd mennyiségével nem feltétlenül fog korrelálni a negatív tartalom mennyisége. Másképpen, lehetséges, hogy valaki az összes megnyilatkozásának mennyiségéhez képest sokat beszél valakiről, és ezeknek a megnyilatkozásoknak éppen a pozitív értékelő tartalma van túlsúlyban a további tartalmakhoz képest. Ezzel összhangban azt is feltételezzük, hogy az egymásról beszélés pusztán mennyisége, valamint a pozitív és a negatív értékelő tartalmak mennyisége alapján nem azonos kapcsolatháló rajzolódik ki.

### 3. A vizsgálati korpusz

Ebben a részben bemutatjuk röviden a vizsgálati adatbázist. A korpusz szövegei hétköznapi kommunikációs helyzetekben keletkeztek, egy külső ingerektől elzárt környezetben (Szabó és Galántai, 2017; Galántai és mtsai, 2018; Gulyás és mtsai, 2018). A magas minőségű hanganyagot egy szórakoztatóipari cég rögzítette egy televíziós műsor készítése céljából. A hang rögzítése napi 24 órában történt, a korpusz pedig összesen 8 egymást követő nap felvételeit tartalmazza. A felvételek készítése során a megfelelő hangminőség biztosítása érdekében minden érintett mikroportot viselt <sup>1</sup>.

A korpusz létrehozásának elsődleges célja az volt, hogy megfelelő kutatási anyagot nyújtson a pletyka elemzéséhez az emberi interakcióban. A HuTongue egy egyedülálló adatbázis: legjobb tudomásunk szerint ez az első olyan nagyméretű magyar beszédkorpusz, amelyet teljes egészében kézzel írtak le és annotáltak.

A leiratozással párhuzamos annotációs folyamat kardinális lépése volt azoknak a diskurzusoknak az annotálása, amelyben a résztvevők jelen nem levő személyről vagy személyekről beszéltek. A kutatás fő tárgya a pletykaszövegek vizsgálata, azonban a korpusz annotálásához a kutatók előzetesen egy tág definíciót határoztak meg a pletyka fogalmára vonatkozóan. A munkadefiníció szerint az annotátoroknak minden olyan megnyilatkozást jelölniük kellett, amelynek tárgya egy vagy több jelen nem levő személy. Amennyiben a diskurzusban megemlített, de jelen nem levő személy az annotátor számára beazonosítható volt, akkor ezt egy megfelelő annotációs címkével kellett feltüntetnie. A tág definíció

<sup>1</sup> Mivel ily módon a szövegek keletkezési körülményei (a résztvevők motivációi, valamint az alkalmankénti külső irányítás) befolyásolhatták a beszélői megnyilatkozásokat, a korpuszt félig vagy részlegesen spontánnak kell tekintenünk (Szabó és Szvetelszky, 2019).

célja az volt, hogy a lehető legkevésbé befolyásolják az annotátorokat a munkájukban azok pletykával kapcsolatos előzetes képzetei. Ennek köszönhetően tehát – ahogyan azt a bevezetőben is említettük – egy olyan korpusz jött létre, amely lehetővé teszi az emberi kommunikáció e szegmensének valamely definíció által előre nem korlátozott vizsgálatát.

A dolgozat további részében az annotáció alapján tett megállapításainkat tárgyaljuk a korpusz szentiment- és emóciótartalmaival összefüggésben.

#### 4. A vizsgálatok és eredményei

A diskurzusokat a korpuszban elhelyezett annotációs tagekre támaszkodva elemeztük, vagyis a részletesebb vizsgálatokban csakis azon szövegrészletek vettek részt, amelyet az annotátorok a korábbi, tág meghatározás alapján jelen nem levő harmadik személyről szóló megnyilatkozásként jelöltek meg. A fenti elvárásnak megfelelő szövegrészletek a teljes, 8 nap hangfelvételeit tartalmazó korpuszból (főleg annak mérete miatt) automatikus eszközökkel lettek kiválogatva.

Az 1. táblázat azt mutatja meg, hogy a tokenek számában mérve ki kiről mennyit beszélt a korpuszban. A korpusz anonimizálása céljából a neveket természetesen mindenhol egységesen más nevekkal helyettesítettük.

ki/kiről	András	Dani	Gabi	Kornél	Sanyi	Vanda
András		3644	7088	2016	2797	2417
Dani	5131		4058	1412	2997	3438
Gabi	15489	5889		6514	5350	7395
Kornél	1925	905	2931		3114	2313
Sanyi	20838	7965	7856	10438		6304
Vanda	5098	6234	5125	2609	4233	

1. táblázat. A szereplők egymásról használt szavainak száma.

Elemzésünk fő irányát a szentiment- és emóciótartalmak vizsgálata képezte. Szentiment alatt azokat a szemantikai tartalmakat értjük, amelyek valamilyen (pozitívan vagy negatívan) értékelő tartalmat hordoznak, míg az emóciók valamely érzelem (pl. öröm, bánat, düh stb.) létere utalnak a szövegben (Zhao és mtsai, 2016). A szentiment- és emóciótartalmakat szótárak segítségével azonosítottuk a szövegekben. A feldolgozáshoz két, általános célra készült szentimentszótárt, valamint egy emóciószótárt használtunk. A szentimentszótárak közül az egyik lista (Liu, 2012) angol nyelvű szentimentszótárának fordítása volt, a másik egy bővebb, jelentősebb kiegészítést, így szleng szavakat is tartalmazó szótár (Szabó, 2015). Az emócióelemzéshez (Szabó és mtsai, 2016) szólistáit alkalmaztuk.

A vizsgálatok során számos összefüggést megvizsgáltunk:

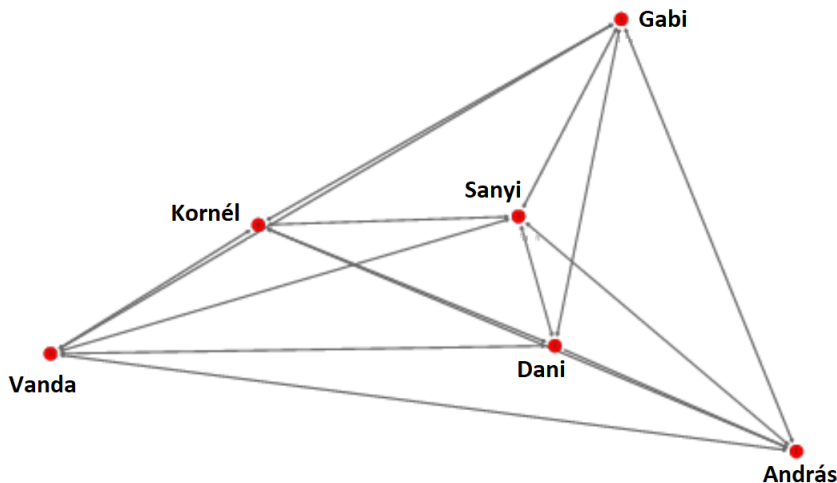
- az egyes emóciók és szentimentek gyakorisági előfordulása megszólalótól függetlenül,

- az egyes emóciók és szentimentek gyakorisági előfordulása megszólalónként,
- az egyes emóciók és szentimentek gyakorisági előfordulása megszólalónként a témaszemélyek szerinti bontásban,
- mindezeket kiszámoltuk úgy is, hogy az általunk vizsgált 8 emóciókategória pozitív és negatív elemeit összevontuk egymással (pl. öröm és szeretet, düh és bánat), és azokat a szentimentszótárakkal kapott eredményekkel összevontuk. Azt reméltük ugyanis, hogy ezzel a módszerrel pontosabb elemzési eredményeket érhetünk el, mintha pusztán a szentimentszótárakat használtuk volna fel.

Ezekből a jelen dolgozatban csupán néhány, érdekesebb összefüggés bemutatására vállalkozhattunk, ezért itt csak a pozitív és a negatív tartalmakra fókuszálunk (amelyeket a lista utolsó pontjának megfelelően az eredmények összevonásával nyertünk ki).

A bemutatott ábrákon a hálózati pontokat az egyes résztvevők adják, míg az éleket az, hogy az adott résztvevőről mennyit beszélnek az összes tokenzám viszonylatában, valamint, hogy az adott résztvevő mennyit beszél a többi résztvevőről az ő összes tokenzámaéhoz képest. Egy-egy résztvevőpár közötti él hosszát tehát a kettejük közötti kapcsolat két iránya együtt adja.

Első lépésben megnéztük tehát, hogy milyen kapcsolatháló rajzolódik ki kizárólag az egymásról beszélés mennyisége alapján. Itt tehát szemantikai tartalmakat még nem vizsgáltunk. Az így kapott hálózatot az 1. ábra mutatja be.



1. ábra: A szereplők szociális távolsága az egymásról használt szavak tükrében.

Ami az ábrán talán a legszembeötlőbb, az az, hogy Sanyi és Dani a vizsgált szempont szerint központi szerepet foglal el a hálózatban: a rövidebb élek

alapján róluk is sokat beszélnek, valamint ők is viszonylagosan sokat beszélnek másokról, és e további résztvevők és közöttük ez a mennyiség nagyjából arányosan oszlik meg. Az is látható, hogy ez a két résztvevő egymásról kiemelten sokat beszél. Hozzájuk képest például Vandának vagy Andrásnak nincs központi szerepe, csupán egy-két emberrel van szorosabb ilyen jellegű kapcsolatuk.

Következő lépésben megnéztük, az egymásról szóló beszéd tartalmak hány százaléka tartalmaz értékelő megnyilatkozást, azaz mekkora arányban találhatók bennük pozitív, illetve negatív töltetű szavak. Az eredményeket a 2. és 3. táblázatok mutatják.

Ki/Kiről	András	Dani	Gabi	Kornél	Sanyi	Vanda
András		21,90	23,38	20,44	25,28	17,91
Dani	22,88		23,63	17,49	22,86	20,77
Gabi	26,48	22,13		23,86	22,95	22,07
Kornél	19,01	10,39	21,08		19,52	17,42
Sanyi	20,84	17,31	18,79	18,68		17,20
Vanda	19,52	19,02	18,56	18,59	15,57	

2. táblázat. A résztvevők egymásról használt negatív szavainak aránya az összes egymásról használt szóhoz viszonyítva.

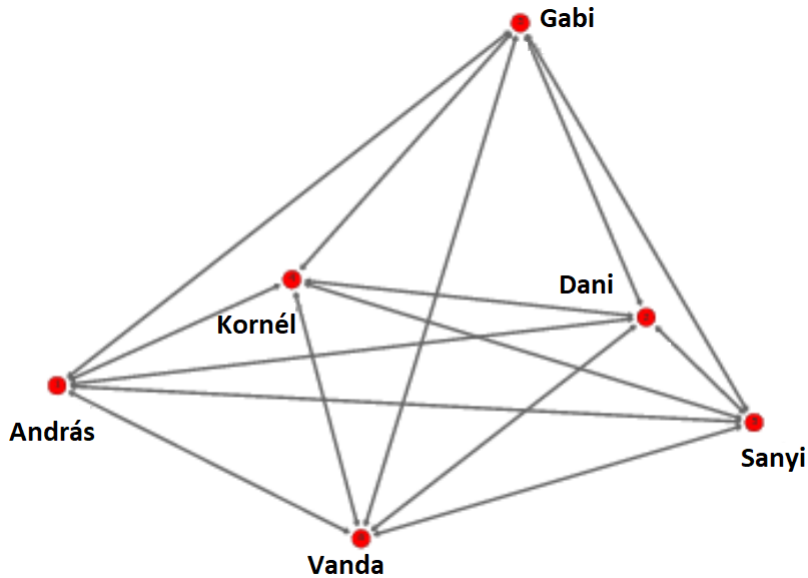
Ki/Kiről	András	Dani	Gabi	Kornél	Sanyi	Vanda
András		34,77	18,34	19,64	14,80	17,91
Dani	19,55		15,72	17,14	13,55	18,09
Gabi	17,41	19,58		19,36	19,96	21,54
Kornél	17,61	13,59	14,60		14,07	13,92
Sanyi	15,35	15,46	19,14	17,26		20,65
Vanda	15,06	14,90	21,62	16,33	18,99	

3. táblázat. A résztvevők egymásról használt pozitív szavainak aránya az összes egymásról használt szóhoz viszonyítva.

Az eredmények azt mutatják, hogy a csoport tagjai kimagasló számban használtak egymásról értékelő megnyilvánulásokat. Egyes esetekben ez akár a szavak 50%-a is lehet, például András Daniról szóló szavainak több mint fele sorolható be valamilyen érzelmi tartalmat hordozó csoportba, ami mindenképpen alátámasztja kettejük szorosabb kapcsolatát.

Ezek után figyeljük meg a negatív tartalmak alapján kirajzolódó kapcsolati hálót!

Azt látjuk, hogy a hálózat jelentős részben átrendeződik, amennyiben csupán a negatív tartalmakra fókuszálunk. Gyakorlatilag az előző ábrához képest csupán Dani őrzi meg valamelyest a pozícióját, ő a negatív tartalmak tekintetében is viszonylagosan sok szoros kapcsolattal rendelkezik. Ami a többi résztvevőt il-



2. ábra: A szereplők szociális távolsága az egymásról használt negatív szavak tükrében.

leti, Sanyi szerepe feltűnően csökken, Danival való szoros kapcsolata viszont a negatív kapcsolatrendszerben is megmarad. Érdekes viszont, hogy Vanda a negatív hálóban sokkal fontosabb szerepet tölt be, több szoros kapcsolattal a nem negatív háléhoz viszonyítva.

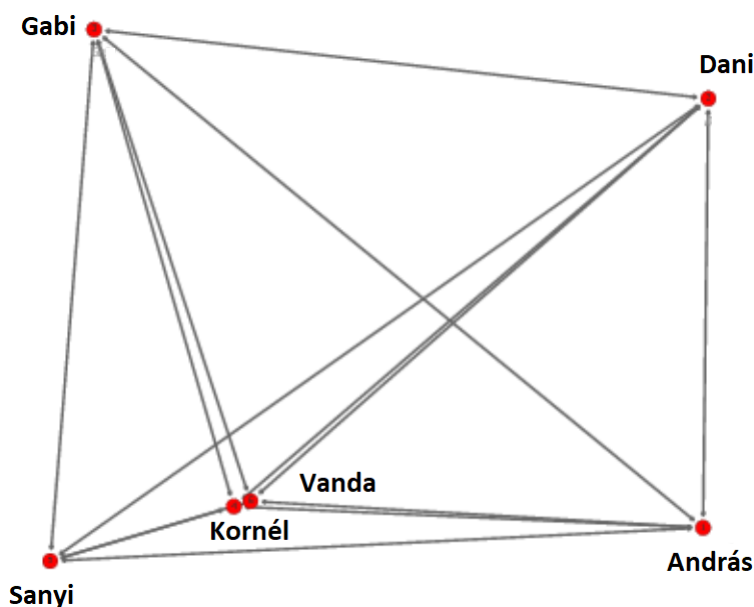
Végül, figyeljük meg a pozitív tartalmak alapján kirajzolódó kapcsolati hálót!

A pozitív kapcsolatháló felépítése feltűnően eltér mind a negatív, mind az általános hálózattól. Mindenekelőtt, a pozitív hálózati kapcsolatok jelentősen kiegyenlítetlenebbek mindkét vizsgált kapcsolatrendszerhez képest. Azt látjuk, hogy Kornél, Vanda és Sanyi kapcsolatai a legerősebbek egymással (közülük is leginkább az utóbbi két résztvevőé), ugyanakkor másoktól nagyon távol esnek. Ha a többi résztvevőt nézzük, megállapítható, hogy azok kapcsolatai ebben a hálózatban rendre kifejezetten gyengék. Azt is látjuk, hogy amíg Sanyi és Dani pontjai a negatív hálóban nagyon közel voltak egymáshoz, a pozitív hálózatban távol esnek egymástól. Ez azt jelenti, hogy egymással kapcsolatos kommunikációjuk erősen negatív tartalmú.

## 5. Összegzés

A dolgozatban egy nagy méretű, magyar beszélt nyelvi adatbázis (HuTongue) segítségével azt vizsgáljuk, hogyan kommunikálnak egymásról a diskurzusok résztvevői, és megszólalásaikhoz milyen szentimentek és emóciók kapcsolódnak. Bár a





3. ábra: A szereplők szociális távolsága az egymásról használt pozitív szavak tükrében.

szakirodalom alapján az emberi informális kommunikáció jelentős részét, egyesek szerint kétharmadát, jelen nem lévő, más személyekről folytatott értékelő tartalmú beszélgetés teszi ki, az erre vonatkozó ismeretanyag viszonylag szerény, és kvalitatív tapasztalatokra hagyatkozik. Dolgozatunkkal ehhez a kutatási irányhoz kívántunk néhány korpuszalapú hálózatvizsgálati eredménnyel hozzájárulni.

Összefoglalva a vizsgálati tapasztalatokat a következő legfontosabb megállapításokat tehetjük: Mindenekelőtt, az egymásról beszélés mennyisége, valamint értékelő szemantikai tartalma alapján egészen eltérő kapcsolati hálót lehetséges rajzolni. Mindez egybevág kezdeti hipotézisünkkel (1. 2. fejezet). Emellett lehetséges, hogy valaki (jelen esetben Sanyi) központi szerepet tölt be a hálózatban az alapján, hogy ő mennyit beszél másokról és róla mennyit beszélnek, azonban mind a pozitív, mind a negatív hálózatban inkább távolabbi kapcsolatokkal rendelkezik. Ez arra mutat, hogy az egymásról beszélés fontos csoportszervező funkciót tölthet be úgy is, hogy nem irányul a targetek negatív vagy pozitív értékelésére. A fenti tapasztalat ugyancsak egybevág a vizsgálatokat megelőző hipotézisünkkel. Azt feltételeztük ugyanis, hogy a beszéd mennyiségével nem feltétlenül fog korrelálni a negatív (és egyáltalán az értékelő) tartalom mennyisége. Másképpen, lehetséges például, hogy valaki az összes megnyilatkozásának mennyiségéhez képest sokat beszél valakiről, és ezeknek a megnyilatkozásoknak a pozitív értékelő tartalma túlsúlyban van a negatív tartalmakhoz képest. Az látszik tehát, hogy az egymásról beszéléssel nem nő szükségképpen a negatív érté-

kelés mennyisége. Mindebből a fordított irányban az is következik, hogy pusztán az egymásra vonatkozó beszéd mennyiségéből nem következtethetünk a vizsgált személyek társas kapcsolatainak tartalmi viszonyaira.

## Köszönetnyilvánítás

A kutatást az Európai Kutatási Tanács (European Research Council), az Európai Unió Horizont 2020 kutatási és innovációs programjának keretében támogatta (ERC\_CoG\_2014\_648693 sz. szerződésben), a kutatás vezetője Takács Károly.

## Hivatkozások

- Bodó, Cs., Kocsis, Zs., Vargha, F.: A Budapesti Egyetemi Kollégiumi Korpusz. Elméleti és módszertani kérdések. In: Benő, A., Fazakas, N. (szerk.) *Élőnyelvi kutatások és a dialektológia: Válogatás a 19. Élőnyelvi Konferencia - Marosvásárhely, 2016. szeptember 7-9. - előadásaiából.* pp. 169–177 (2017)
- Crowdy, S.: Spoken corpus design. *Literary and Linguistic Computing* 8(4), 259–265 (1993)
- Dunbar, R.I.: *Grooming, Gossip and the Evolution of Language.* Harvard University Press, Cambridge, MA (1996)
- Dunbar, R.I.: Gossip in evolutionary perspective. *Review of General Psychology* 8(2), 100–110 (2004)
- Eckhaus, E., Ben-Hador, B.: Gossip and gender differences: a content analysis approach. *Journal of Gender Studies* 28(1), 97–108 (2019)
- Emler, N.: Gossip, reputation and social adaptation. In: Goodman, R.F., Ben-Ze'ev, A. (szerk.) *Good gossip.* University Press of Kansas, Lawrence (1994)
- Galántai, J., Pápay, B., Kubik, B.G., Szabó, M.K., Takács, K.: A pletyka a társas rend szolgálatában-az informális kommunikáció struktúrájának mélyebb megértéséért a computational social science eszközeivel. *Magyar Tudomány* 179(7), 964–976 (2018)
- Giardini, F., Wittek, R.: *The Oxford Handbook of Gossip and Reputation.* Oxford University Press (2019)
- Gósy, M.: Bea—a multifunctional hungarian spoken language database. *Phonetician* 105, 50–61 (2013)
- Grosser, T., Kidwell, V., Labianca, G.J.: Hearing it through the grapevine: Positive and negative workplace gossip. *Organizational Dynamics* 41, 52–61 (2012)
- Gulyás, A., Galántai, J., Szabó, M.K., Szebeni, Z.: A HuTongue spontán beszélt nyelvi korpusz leiratozásának és annotálásának minőségbiztosítási munkálatai. In: MSZNY 2018 - XIV. Magyar Számítógépes Nyelvészeti Konferencia. pp. 317–330 (2018)
- Hemphill, C.T., Godfrey, J.J., Doddington, G.R.: The atis spoken language systems pilot corpus. In: *Speech and Natural Language: Proceedings of a Workshop Held at Hidden Valley, Pennsylvania, June 24-27, 1990* (1990)
- Kisfalusi, D., Takács, K.: A pletyka és a reputáció összefüggései középiskolai osztályközösségekben. *Szociológiai Szemle* 28(1), 83–104 (2018)

- Liu, B.: Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies* 5(1), 1–167 (2012)
- Maekawa, K., Koiso, H., Furui, S., Isahara, H.: Spontaneous speech corpus of japanese. In: *LREC*. pp. 947–9520. Citeseer (2000)
- Mengusoglu, E., Deroo, O.: Turkish lvcsr: Database preparation and language modeling for an agglutinative language. In: *IEEE International Conference on Acoustics Speech And Signal Processing*. vol. 6, pp. 4018–4018. IEEE; 1999 (2001)
- Mitra, T., Gilbert, E.: Have you heard?: How gossip flows through workplace email. In: *ICWSM* (2012)
- Neuberger, T., Gyarmathy, D., Grácsi, T.E., Horváth, V., Gósy, M., Beke, A.: Development of a large spontaneous speech database of agglutinative hungarian language. In: *International Conference on Text, Speech, and Dialogue*. pp. 424–431. Springer (2014)
- Oostdijk, N.: The spoken dutch corpus. overview and first evaluation. In: *LREC*. pp. 887–894. Athens, Greece (2000)
- Pápay, B.: *The Purpose and Types of Organizational Gossip*. Doctoral Dissertation (2019)
- Pápay, B., Kubik, B.Gy., Galántai, J.: Gossip is more than just story telling: Topic modeling and quantitative analysis on a spontaneous speech corpus. In: Jorge, A., Campos, R., Jatowt, A., Nunes, S. (szerk.) *Proceedings of the Text2StoryIR’18 Workshop, Grenoble, France, 26-March- 2018*. CEUR-WS.org (2018)
- Seppänen, T., Toivanen, J., Väyrynen, E.: Mediateam speech corpus: a first large finnish emotional speech database. In: *Proceedings of the Proceedings of XV International Conference of Phonetic Science*. pp. 2469–2472. Citeseer (2003)
- Szabó, M.K., Galántai, J.: Egy magyar nyelvű spontán beszélt nyelvi korpusz (HuTongue) létrehozásának tapasztalatai. In: *XXVI. MANYE Kongresszus konferenciakötete*. Pécs (2017)
- Szabó, M.K., Szvetelszky, Zs.: Részlegesen spontán körülmények között keletkezett pletykaszövegek pragmatikai szempontú vizsgálata. *Nyelvtudományi Közlemények* 115, 317–343 (2019)
- Szabó, M.K., Vincze, V., Ring, O., Üveges, I., Vit, E., Samu, F., Gulyás, A., Galántai, J., Szvetelszky, Zs., Bodor-Eranus, E.H., Takács, K.: *StaffTalk: magyar nyelvű spontán beszélgetések korpusza*. In: *XVII. Magyar Számítógépes Nyelvészeti Konferencia*. Szegedi Tudományegyetem, Szeged (2021)
- Szabó, M.: Egy magyar nyelvű szentimentlexikon létrehozásának tapasztalatai és dilemmái (‘Experiences and dilemmas of the creation of a Hungarian sentiment dictionary’). *Nyelv, kultúra, társadalom. Segédkönyvek a nyelvészet tanulmányozásához*. 177, 278–285 (2015)
- Szabó, M., Vincze, V., Morvay, G.: *Magyar nyelvű szövegek emócióelemzésének elméleti nyelvészeti és nyelvtechnológiai problémái*. Reményi, AÁ, Sárdi Cs., Tóth Zs.(szerk.) *Távlatok a mai magyar alkalmazott nyelvészetben*. Budapest: Tinta pp. 282–292 (2016)
- Van Bael, C., Baayen, R.H., Strik, H.: Segment deletion in spontaneous speech: a corpus study using mixed effects models with crossed random effects. In: *INTERSPEECH*. pp. 2741–2744 (2007)

Vincze, V., Üveges, I., Szabó, M.K., Takács, K.: A magyar beszélt és írott nyelv különböző korpuszainak morfológiai és szófaji vizsgálata. In: XVII. Magyar Számítógépes Nyelvészeti Konferencia. Szegedi Tudományegyetem, Szeged (2021)

Zhao, J., Liu, K., Xu, L.: Sentiment analysis: mining opinions, sentiments, and emotions (2016)