

Záróbeszámoló

az OTKA K-112172. sz. kutatási projekt eredményeiről

A 2014-2018 ciklusban *A tisztességes eljárás nyelvi aspektusai. A jogi nyelv hatása a joghoz való hozzáférés érvényesülésére* címmel folytatott kutatás eredményeit összegző beszámolóinkat I. a pályázati anyagban felvázolt kutatási tervben meghatározott célkitűzések menetében haladva; II. azok évenkénti teljesüléséről adott részbeszámolókat felidézve; majd III. az eredményeket áttekintően vizsgálva terjesztjük elő; IV. külön kitérve a teljesületlen tervekre és a kutatási irány továbbvitelének lehetőségeire.

I. A 2014-ben elfogadott célkitűzések és kutatási terv

“I. A kutatás összefoglalója, célkitűzései szakemberek számára

A kutatás alapvető célkitűzései két kulcsszó köré rendezhetőek. A tisztességes (fair) eljárás számos részjogosítványban fejeződik ki. A büntető eljárásjog alapvető szintű követelménye, hogy a bűnösség megállapítására és a szankciók kiszabására csak olyan eljárás eredményeként kerülhessen sor, amely védelmet biztosít az állami túlhatalommal szemben. A joghoz való hozzáférés arra a – közhelyszerű – tényre világít rá, hogy bár a jogszabályok kihirdetése révén a joganyag és a jogérvényesítés eszközrendszere elvileg és formailag bárkinek egyformán rendelkezésére áll, e lehetőséggel nem egyenlő esélyekkel élünk. Az egyenlőtlenséget a vagyon/jövedelem és a műveltség/iskolázottság független szociológiai változói magyarázzák a legkielégítőbben.

Mindeddig kevés figyelem irányult ezen elvárások nyelvi dimenziójára, arra, hogy a jogi eljárások tisztességes volta és a joghoz való hozzáférés esélye a nyelv dimenziójában is értelmezhető és értelmezendő. A büntető eljárásjog egyes rendelkezései érzékelik és tükrözik a nyelvhasználat jelentőségét, de nem vetnek számot a nyelv belső rétegzettségével. A szociolingvisztika vizsgálja a társadalmi és nyelvi egyenlőtlenségek összefüggéseit, kutatásunk pedig ebbe az összefüggésrendszerbe célozza beemelni a jogi egyenlőtlenségek dimenzióját. A nyelvi kompetencia nem homogén: az „anyanyelvi kompetencia” nem foglalja magában a „(jogi) szaknyelvi kompetenciát”. Feltevésünk szerint a jogi eljárásban zajló sorozatos intralingvális átfordítások (a laikus és a jogi nyelv kódja között) jelentősen rontják a laikusok esélyeit annak megértésében, hogy mi zajlik az eljárás folyamán, így növelik kiszolgáltatottságukat, csökkentik a joghoz való hozzáférésüket.

Mi a kutatás alapkérdése?

Kiinduló tételünk szerint a tisztességes eljárás és a joghoz való hozzáférés biztosításának követelménye azt a kívánalmat is magában foglalja, hogy a jogi eljárás laikus résztvevői ne kerüljenek kiszolgáltatott, alárendelt helyzetbe pusztán azért, mert nem értik azt a (szak)nyelvet, amin a professzionális résztvevők hozzájuk beszélnek. A kiinduló tételből következő kérdés az, hogy 1. kimutatható-e olyan különbség a jogi és a laikus nyelvhasználat között, amely akadályt gördít a megértés elé; 2. ha vannak ilyen akadályok, azok eredményez(het)ik-e a tisztességes eljárás és a joghoz való hozzáférés követelményének

sérülését; 3. ha ezen veszélyek realitása igazolható, milyen javaslatok tehetők azok kikerülésére és az eljárások korrektségének, elérhetőségének fokozottabb biztosítására.

Az alapkérdést instrumentálisan szolgáló előkérdés annak empirikusan alátámasztott megválaszolása, hogy a magyar jogi nyelv milyen specifikus jellemzőkkel bír a hétköznapi nyelvhez képest, és milyen különbségek tárhatók fel a jogi illetve a hétköznapi kontextusban zajló diskurzusok között. A magyar jogi nyelv (beszélt és írott változatainak) rendszeres, szisztematikus leírása jelenleg nem áll rendelkezésre; ha e munka befejezésére nem is, de megkezdésére és megalapozására reménnyel lehet vállalkozni. A rendelkezésünkre álló korpuszok (jogszabálytár, döntvénytár, Magyar Nemzeti Szövegtár, egyéb, már meglevő s még megszerzendő beszélt nyelvi szöveganyagok) elemzése elsősorban számítógépes és hagyományos, kvalitatív és kvantitatív nyelvészeti elemzésekkel, másodsorban történeti, összehasonlító és (jogi) analitikai módszerek alkalmazásával tehető meg.

II. Munkaterv

1. év

2014. szeptember 1. – 2015. augusztus 31.

A kutatás első féléve (2014. szeptember 1. – 2015. február 28.) a felkészülést, az elméleti alapot és a „csapatépítést” szolgálja. A kutatásban összesen kilenc kutató vesz részt a következő tudományterületekről: jogásként Szabó Miklós, Vinnai Edina, Gedeon Magdolna, Pázmándi Kinga és Zódi Zsolt, aki egyúttal informatikai ismeretekkel is rendelkezik, ugyancsak informatikusként csatlakozik Szendi Attila, nyelvészektől pedig Vincze Veronika, Kurtán Zsuzsa és Sajgál Mónika. A kutatás eredményes megvalósítása a szakismeretek három körének összedolgozását kívánja meg: a jogi, a nyelvészeti és az informatikai szakértelmét. A kutatás résztvevőinek – saját szakterületükben való jártasságukon túl – értőnek és fogékonyak kell lenniük a másik kettő felé is. Ennek elérését a 2014. november közepére „A jog nyelvi dimenziója” címmel tervezett konferenciától várjuk (ide értve a felkészülést és a konferencia anyagának közzétételét is). Ugyanakkor kell kiépítenünk az együttműködés kereteit azokkal, akiktől a kutatásaink alapjául szolgáló adatbázisok megszerzését, illetve továbbfejlesztését várjuk (Borsod-Abaúj-Zemplén Megyei Rendőr-főkapitányság, Miskolci Törvényszék, ügyvédi irodák, civil szervezetek, tudományos és üzleti kutatóhelyek).

Az első félévben elkezdve, de a második szemeszterben (2015. március 1. – 2015. augusztus 31.) is folytatva a kialakított együttműködések keretében megkezdjük az adatbázisok alkalmassá tételét a kutatás számára. A „törvénytár” és „döntvénytár” esetében elegendő az adatbázis (a gyűjtemények) digitális és elemezhető formában való megszerzése; a laikus beszélt nyelv (kihallgatások, meghallgatások, beadványok stb.) esetében a korábbi kutatásunk során kialakított adatbázisunk továbbfejlesztésére is szükség van. Elsősorban e körben számítunk a külső alkalmazásban résztvevő kutató, Sajgál Mónika és a hallgatók segítségére, miközben a korpusz formálás felé tett lépések megtétele a kutatókra vár. Ez utóbbi teendő végigkíséri majd a kutatás egész időtartamát.

Ugyancsak a kutatás első évére esik a Német Jog és Társadalom Egyesület (Vereinigung für Recht und Gesellschaft) által két évente megrendezésre kerülő konferenciák következő eseménye 2014 szeptemberében Berlinben, ahol Sajgál Mónika fog részt venni (feltéve, hogy lesz még lehetőség regisztrálni a kutatás költségvetésének megnyílását követően) – ő részt vett a legutóbbi Bécsi konferencián is és hasznos szakmai kapcsolatokat sikerült kialakítania.

2. év

2015. szeptember 1. – 2016. augusztus 31.

Az első év során szakmai konzultáció keretében véglegesítjük az informatikai eszközök alkalmazásához szükséges teendőket. Ennek alapját nyitott forráskódú számítógépes program kiválasztása (pl. Textrend, Textplore) képezi, esetleg magyar jogi nyelvi lexikonnal való kiegészítése, a szótövesítés elvégzése, amely lehetővé teszi akár más nyelvekre kidolgozott kvantitatív elemzések lefuttatását is. Bizonyos, hogy az informatikai eszköztár tekintetében az előkészületek megtétele a legfáradságosabb és legidőigényesebb. A kutatás ideje alatt folyamatosan figyelemmel kísérjük a nyelvtechnológia fejlődését, és – ha szükséges – konzultáció formájában vagy akár konkrét kutatási feladatok elvégzésére bevonunk olyan szakembert, vagy egymást követően több különböző szakembert, akik ezeknek az újonnan kifejlesztett programoknak az alkalmazásában segítségünkre lehetnek.

A kutatás második évében folytatni tervezzük az adatgyűjtést és adatbázis-fejlesztést, megkísérelve néhány ügy végigkövetését az eljárás kezdetétől annak jogerős befejezéséig. Ezzel párhuzamosan – ezen időszak második felétől – folyamatosan törekszünk a hanganyagok transzkribálására, a korpuszok egységesítésére és az adatbázis rendezésére külső, a kutatásban megbízási szerződéssel foglalkoztatott szakemberek segítségével. Ekkor válik esedékessé a kutatási honlap létrehozása, feltöltése és folyamatos szerkesztése is a kutatás végéig.

A kutatás teljes ideje alatt folyamatosan figyelemmel kísérjük az angol és német nyelvterületen a témában szervezett konferenciákat, és lehetőséget kívánunk biztosítani a kutatásban résztvevő valamennyi kutató számára, hogy egy alkalommal részt vegyen egy Európán belül szervezett rendezvényen, elősegítve ezzel egyrészt azt, hogy tudomást szerezzünk az aktuális kutatási irányokról, másrészt azt, hogy a hasonló kutatásokkal foglalkozó külföldi szakemberek is megismerjék Magyarországon egyedülálló kutatásunkat. Ugyancsak a második évben tervezzük megjelentetni a 2014 novemberében megrendezett konferencia tanulmánykötetét.

3. év

2016. szeptember 1. – 2017. augusztus 31.

A kutatás harmadik évétől az elemzések megkezdése és az eredmények tesztelése kerül a kutatás homlokterébe. A rendelkezésre álló adatbázisokon a rendelkezésre álló módszerekkel első megközelítésű eredményekre számítunk, amelyek további teendőket tehetnek szükségessé az adatgyűjtés és adatbázis-fejlesztés területén, az elemzési módszerek tekintetében, esetleg újabbak bevonása iránt. Ugyanekkor meg kell kezdeni az eredmények értelmezését, a hipotézisekkel való szembesítést és az elméleti modellbe szervezhetőség vizsgálatát.

Az előzetes és részeredmények várakozásunk szerint már lehetővé és szükségessé teszik azok közzétételét, hogy a szakmai közönség is tudomást szerezzon kutatásunk témájáról és fő irányvonalairól. Amellett, hogy vizsgálataink eredményeit folyamatosan publikáljuk kutatási honlapunkon, a kínáló konferenciákon és nyelvészeti és jogi folyóiratokban, 2017 tavaszán tervezzük útjára indítani (a „Magyar orvosi nyelv” mintájára) a „Magyar jogi nyelv” című periodikát. Évi két lapszám keretei között, megfelelően kialakított rovatokkal, s a tárgy iránt érdeklődő közönséget is bevonva reméljük tartós fórumot kínálni a tárgynak. Kézenfekvő, hogy az induló lapszám a kutatásról adjon számot.

4. év

2017. szeptember 1. – 2018. augusztus 31.

A kutatás utolsó évében már nem tervezzük az adatbázisok további fejlesztését, hanem az addig összegyűlt anyagok kiértékelésére, az eredmények összegzésére helyezzük a hangsúlyt. A rendelkezésünkre álló korpuszok (törvénytár, döntvénytár, szövegtár) számítógéppel is támogatott vizsgálatában három fő szakaszt különítünk el. Az első szakaszban kvantitatív elemzések segítségével különböző nyelvi jelenségek és mintázatok előfordulási gyakoriságát kívánjuk vizsgálni és számszerűsíteni. Ezt követően jogi, jogelméleti, kognitív nyelvészeti és pragmatikai kategóriákra támaszkodva arra igyekszünk magyarázatot találni, hogy az adott szövegtípus miért éppen az adott nyelvi mintázatokkal jellemezhető. A vizsgálat harmadik szakaszában arra keressük a választ, hogy a jogi szövegtípusokra jellemző mintázatok mennyiben szolgálják vagy gátolják a tisztességes eljárás és a joghoz való hozzáférés biztosítását. A nagy mennyiségű adat miatt a törvénytárak és bírósági ítéletek vonatkozásában célszerűnek tűnik statisztikai módszerek alkalmazása is.

Vizsgálataink eredményeit minél szélesebb körben igyekszünk megismertetni a szakmai és a laikus közönséggel, ezért – az előzőekben említetteken túl – tervezzük egy záró konferencia vagy workshop szervezését, valamint eredményeink egységes tanulmánykötetben való közzétételét is. Ugyancsak fontos célkitűzésünk, hogy a kutatás megállapításai mind a jogi, mind az alkalmazott nyelvészeti egyetemi tananyag részévé válhasson, és ezáltal kutatásunk hozzájárulhasson mindkét tudományterület fejlődéséhez. ”

II. Időközi szakmai részbeszámolók

“1. évi részbeszámoló

A kutatócsoport a tervezettek szerint összeállt. Nem tervezett módon, de a kutatás hírére érkezett jelentkezések nyomán három „pártoló közreműködőt” is befogadtunk: Dobos Csilla (docens, ME), Farkasné Puklus Márta (nyelvtanár, ME), Markovich Réka (tanársegéd, BME – PhD hallgató, ELTE). Egyetemi hallgatók bevonására az előzetes terv szerint – a projekt késedelmes indulása miatt – nem került sor. Búzás Barbara és Petró Angéla, jelenleg harmadéves joghallgatókkal sikerült olyan kapcsolatot kialakítani, hogy bevonásuk megtörtént, s megbízási szerződésük előkészítése folyamatban van.

2014. november 14-15-én, az MTA Miskolci Akadémiai Bizottságával együttműködve, „A jog nyelvi dimenziója” címmel megtartottuk a tervezett nyitókonferenciát. A kétnapos konferencia első napján – meghívott előadók közreműködésével – a kutatás megalapozásához, valamint a jog és a nyelv összefüggéseinek tisztázásához szükséges előadásokra és megvitatásokra került sor. A második nap szolgált a kutatói team tagjainak bemutatkozó előadásaira, a kutatással kapcsolatos észrevételeik és javaslataik megvitatására, valamint a csoporton belüli munkamegosztás és az együttműködés rendjének kialakítására. A konferencia programja:

NOVEMBER 14.: Szigeti Péter: A jog tárgyspecifikus kategóriái: jogrend – törvényesség – alkotmányosság; Badó Attila: Társadalmi státusz és tisztességes eljárás; Bencze Mátyás: A bírói indokolás minőségének vizsgálata; Takács Péter: A magyar állam átnevezése; Hidasi Viktória: A nemzeti beszédmód; Varga Csaba: Jog és nyelv? Jog mint nyelv?; Hegyi Szabolcs: Tényállítást vagy értékítéletet? – Egy alapjogi kérdés nyelvi megközelítése; Vita, összegzés

NOVEMBER 15.: Dobos Csilla: Nyelven belüli fordítás a szakemberek és a laikusok közötti kommunikációban; Gedeon Magdolna: Latin jogi terminusok használata a középkori Magyarországon; Kurtán Zsuzsa: A szöveg érthetőségének biztosítása: elméletek és modellek a jelentés; Markovich Réka: Kondicionálisok a jogszabályszovegekben; Pázmándi Kinga: A köznapi nyelv és a jogi szaknyelv távolsága, értelmezési dimenziói; Puklus Márta: Jogi szaknyelv és szakfordítás; Sajgál Mónika: Társadalmi cselekvési lehetőségek és nyelvi pozicionálás a jogban; Szabó Miklós: A jogi nyelv erőtere; Szendi Attila: A jog történeti szövegrétegei és számítógépes feldolgozásuk; Vincze Veronika: A jogi szövegek nyelvi sajátosságai és korpuszalapú vizsgálatuk lehetőségei; Vinnai Edina: A tájékoztatáshoz való jog szabályozása és gyakorlata; Zódi Zsolt: Alkalmas-e a Google a jogi problémák megoldására?

2015. május 11-én a kutatócsoport Kurtán Zsuzsa kezdeményezésére és javaslata alapján munka-megbeszélést tartott Miskolcon, a nyelvi elemzéshez szükséges adatgyűjtés, adatbázisok helyzetének áttekintése és a kutatás számára való alkalmassá tétele céljából. Megbeszélés és közös döntés tárgyát képezte az adatgyűjtés és az adatbázisok köre; a kutatással megválaszolható kérdések köre; a jogi szaknyelv leírásának lehetséges mélysége és szempontjai. Eldöntöttük, hogy – lehetőségeink korlátai között – a jogi szaknyelv leírásához: kutatási témánk, a jogi szakemberek és a laikusok közötti interakciók nyelvhasználatának összevetéséhez indikátorként a funkcionálisan tágran értelmezett tájékoztatás műfajaira, szövegeire koncentráljuk a figyelmünket. Ez az egyik, jól körülhatárolható „ütközőzóna” ugyanis, ahol a laikus és professzionális nyelvhasználat konfrontálódik és tetten érhető a tisztességes eljárás és a joghoz való hozzáférés nyelvi feltételrendszere.

2015. június 19-én kutatócsoportunk megtartotta második munkamegbeszélését Miskolcon. Az első év addigi eredményeinek és nehézségeinek áttekintése és a második év feladatainak előzetes áttekintése alapján kialakítottuk a kutatócsoporton belüli munkamegosztás struktúráját és felelősöket rendeltünk az egyes részprojektek mellé. Ennek során nevesítettük a részfeladatokat: A „leíró magyar jogi szaknyelv” kereteinek, struktúrájának és fogalmainak kijelölése (Kurtán Zsuzsa); a jogi adatbázisokon végzendő kvantitatív, számítógépes elemzések feltételeinek biztosítása és irányítása (Vincze Veronika); az elemzések alapjául szolgáló adatbázisok körének kijelölése, kialakításuk, feldolgozásra előkészítésük (Zódi Zsolt); ezek közül a hangfelvételek rendbetétele, szükséges kiegészítése, fejlesztése, átírása, s az elérhető más hangzó anyagok begyűjtése, adatbázisba szervezése (Sajgál Mónika); a joghoz való hozzáférés és a tisztességes eljárás jogi–alapjogi–alapelvei feltételeinek és e feltételek érvényesülésének elemzése (Pázmándi Kinga: a magánjog, Vinnai Edina: a büntetőjog területén); a jogi (jogalkotói és jogalkalmazói) nyelv történeti előzményeinek vizsgálata, történeti adatbázis-metszetek összeállítása és történeti–összehasonlító módszerek alkalmazása (Gedeon Magdolna és Szendi Attila); a kiadványaink szerkesztése és a Magyar Jogi Nyelv c. folyóirat kiadásának előkészítését (Szabó Miklós). Természetesen a részfelelősök által irányított és felügyelt tevékenységek elvégzése a kutatócsoport minden tagjának feladatát képezi.

Megkezdődött a kapcsolatok kialakítása a kutatási program teljesítésének feltételeit megteremtendő. Zódi Zsolt révén sikerült Az Opten Informatikai Kft.-től az előre meghatározott kritériumrendszer szerint megszerezni és a nyelvi elemzők rendelkezésére bocsátani a magyar jogi korpuszt reprezentáló mintegy 15.000 jogszabály, bírósági ítélet és felsőbbbírósági irányítási eszköz szövegét. Felvettük a kapcsolatot jogalkalmazó szervekkel is, azonban velük érdemi együttműködésre még nem került sor. A csoport tagjai rendszeresen részt vettek és vesznek olyan konferenciákon, tapasztalatcseréken, műhely-megbeszéléseken, amelyektől hasznosítható ismeret vagy tapasztalat volt remélhető a kutatás számára; ezek a csoport tagjaival – szóban, illetve a kutatás számára kialakított Google Drive felületen – megosztják. Ennek részét képezi a már korábban felvett hangzó anyagok egységesítése. A kb.

470 percnyi rendőrségi kihallgatásból 186 percnyi anyag került átírásra a beszélt nyelvi korpuszokban használt GAT (Gesprächsanalytisches Transkriptionssystem) átírási rendszer alapján. Ezen kész transzkribált hanganyagok feltöltésre kerültek a hanganyagokkal együtt a kutatás során használt gdrive felületre, így hozzáférhető és elemezhető a többi részt vevő számára is. A továbbiakban Sajgál Mónika feladata, hogy a korpuszt egységesítse a kutatók által a munkamegbeszélés által meghatározott elvárások szerint. Úgyszintén megkezdjük a további munka feltételeként vagy elősegítőjeként szolgáló dologi, irodai és informatikai eszközök, könyvek, adatforrások stb. beszerzését – a pénzügyi beszámoló által részletezett módon.

A kutatás menetében, az első év folyamányaként, megkezdődött a második évi munkaterv végrehajtásának előkészítése is. Így:

- Összeállt és megszerkesztve a kiadóhoz került a konferenciakötet, a Prudentia Iuris sorozat részeként; megjelenése a közeljövőben várható.
- Megtörtént a kapcsolatfelvétel és a tapasztalatokról való konzultáció Dr. Bósze Péterrel, a Magyar Orvosi Nyelv c., mintaadó szakfolyóirat főszerkesztőjével.
- Megtörtént a kutatási honlap szerkezetének és arculatának kialakítása, s az informatikai megvalósítás előkészítése.

2. évi részbeszámoló

Fő kutatási témánk alapja a jogi szakemberek és a laikusok közötti beszédhelyzetek nyelvhasználata. A kutatócsoport elmúlt évi két megbeszélésén (2015. november 11. Veszprém, Akadémiai Bizottság, 2016. június 22. Budapest, BME) áttekintettük a jogi szövegek példaalapú szabályszerűségeit feltáró és leíró elemzés szükséges lépéseit. Főbb célkitűzéseink és eredményeink a következők voltak:

A szaknyelv – jelen kutatásban a jogi szaknyelv – leírásához célszerű a tevékenységeknek megfelelő tipikus beszédhelyzetek nyelvhasználatát megjelenítő szövegeket választani, melyek elemzését lehatároljuk a fő célkitűzéssel összhangban álló részcélok rendszerében. A kutatásban a következő kérdésekre keressük a választ:

- (a) Milyen nyelvi minták mutathatók ki, amikor laikusok szembesülnek a jogi nyelvhasználattal?
- (b) Milyen nyelvi minták mutathatók ki, amikor jogi szakembereknek kell értelmezniük a laikusok beszédét?
- (c) Milyen nyelvi jelenségek okozzák az értelmezési problémát a laikusok és a jogi szakemberek perspektívájából?
- (d) Érzékelik-e a beszélők a nyelvhasználati problémát? Ha igen, minek a hatására?
- (e) Milyen nyelvi tevékenységet végeznek a beszélők a probléma kiküszöbölése érdekében?
- (f) Milyen problémakezelési lehetőségek léteznek, és mi jellemzi ennek igénybe vétele során a nyelvhasználatot?

A fenti kérdésekre az alábbi szövegtípusok vizsgálatával keressük a válaszokat: laikusok és szakemberek találkozásainak dokumentumai, a ténykérdés megállapítására irányuló szövegek, a jogkérdés megállapításához kapcsolódó szövegek, ítéletek, bírósági határozatok, interjúk, résztvevő megfigyelések, kihallgatási jegyzőkönyvek, átírások, szóbeli tervezetlen szövegek és átváltott szóbeli/írásbeli tervezett formai változatai, törvények, rendeletek, jogász nyelvész szakértők szövegei, kérdőíves felmérések, internetes keresőprogramok, jogi tájékoztatás szövegei, oktatási anyagok.

A fenti kérdések vizsgálatát egybekötöttük a szövegminták többdimenziós tanulmányozásával, továbbá a folyamatosan bővülő korpusz automatikus elemzésének értékelésével. A már elvégzett kutatások eredményeit, szövegtörzsét is figyelembe véve és azokat tovább fejlesztve olyan nyelvi korpuszt építünk, mellyel az a célunk, hogy digitálisan rendelkezésünkre álljanak olyan írott jogi professzionális szövegek, melyeket a kutatási célkitűzéseinknek megfelelően statisztikai vizsgálattal határozunk meg. A meghatározás alapja azoknak a szövegeknek a köre (például jogszabályok, törvények), amelyekre a határozatokban/ítéletekben a leggyakrabban találunk hivatkozást.

A jogi szaknyelv leírásához – kutatási témánk, a jogi szakemberek és a laikusok közötti interakciók nyelvhasználatának összevetéséhez – előtérbe helyezzük a funkcionálisan tágran értelmezett tájékoztatás és figyelmeztetés műfajait, szövegeit. Tanulmányozzuk, hogy milyen írott és beszélt nyelvhasználati követelményekkel és elvárásokkal szembesülnek a beszédhelyzetek résztvevői. Meghatároztuk, hogy a tartalmi logikai kapcsolódásokat és a nyelvhasználati jellemzőket valamennyi nyelvi síkon (lexikai, morfológiai, szintaktikai, szó szerkezeti és mondatszinten) az automatikus korpuszelemzés példái alapján tárjuk fel és írjuk le.

A példaalapú szabályszerűségek bemutatásával alapot teremtünk ahhoz, hogy a jogi nyelv sajátosságait elkülöníthessük más regiszterektől, a köznyelvtől. Ehhez megállapodtunk a jogi nyelv leírásának főbb csomópontjaiban, melyek keretében bemutatjuk a jellemző műfajok és szövegtípusok egyes nyelvi jelenségeit: állítmányi szerkezetek, alanyi szerkezetek, mellérendelés, alárendelés, központosítás.

A 2015/16-os évre, a kutatási program második évére a munkatervünk a következőket irányozta elő:

I. A nyitókonferencia anyagának kötetben való kiadása. Ez megtörtént. A bibliográfiai adatok: Szabó Miklós (szerk.): *A jog nyelvi dimenziója*. Miskolc: Bíbor, 2015. [Prudentia Iuris sorozat 31.] 296 o. ISBN 978-615-5536-07-6

A kötet tartalomjegyzékét és a kutatócsoport tagjai és közreműködői írásait nyilvánossá tettük a kutatási honlapunkon. A megjelent példányokat a társintézmények és kutatóik, valamint jogalkalmazó szervek számára is eljuttattuk.

II. Terveztük a nyelvi elemzés informatikai eszközeinek alkalmazásához szükséges teendők elvégzését. Ennek eredményeként a Szegedi Egyetemen kidolgozott és használt, de a mi céljainknak is megfelelő „Magyarlanc” szoftvert választottuk, s használhatjuk külön díj fizetése nélkül. A „magyarlanc” elemző program alkalmas txt formátumú szövegek automatikus nyelvi előelemzésének elvégzésére. Az elemzés eredményeképpen a szöveget mondatra, illetve szavakra bonthatjuk, a szavakhoz morfológiai elemzést rendelhetünk, illetve a mondatokhoz szintaktikai (függőségi) elemzést is társíthatunk. A program ismertetéséhez hivatkozás: Zsibrita, J., Vincze, V., Farkas, R.: magyarlanc: A toolkit for morphological and dependency parsing of Hungarian. In: Proceedings of RANLP. (2013) 763–771. Linkek a tájékozódáshoz: <http://www.inf.u-szeged.hu/rgai/magyarlanc> (Online) és a bemutatáshoz: <http://rgai.inf.u-szeged.hu/magyarlanc-service/> (demo). E lépések megtételéhez külső szakember(ek) bevonása nem vált szükségessé.

III. Az összesen 18 óra 36 percnyi hanganyagból és a már meglévő konverzáció analitikai transzkripciók alapján az elmúlt évben körülbelül 10 órányi hanganyag került átírára két joghallgató közreműködésével olyan normalizált átíratban, amely a Magyarlanc szintaktikai, morfológiai és szófaj egyértelműsítő programban elemezhető. Megtörtént ezen szövegek formai egységesítése, a konverzáció analitikai transzkripciók és a normalizált átíratok összevetése és javítása.

IV. Vállalt feladatunk volt az érdemi elemzéseket megalapozó korpusz kialakítása, rendezése. Ez eltérő teendőket kívánt meg (A) az írásos és (B) a hangzó adatbázisok esetében.

(A) A kutatás egyik fontos része egy szövegtörzs létrehozása, amelyen a statisztikai és kvalitatív vizsgálatokat elvégezzük. Ez a szövegtörzs hat rész-törzsből áll, amelyek mindegyike nagyjából 300.000 és 400.000 szó között mozog. Az egyes szövegtörzsek 1-től 5-ig a „paradigmatikus”, „jogászjogi” nyelvtől, vagyis a kódexek nyelvétől távolodnak kifelé, míg a laikus nyelvhasználatig (jogi tájékoztató fórumok szövegeiig) elérnek. Az egyes rész-törzsek tartalma az alábbi:

1. „Jogászjog”: Ptk., Btk., Pp., Be., Ket.
2. „Egyéb” magyar jogszabályok: 13 db tv., 60 db Korm. r., 156 db miniszteri rendelet
3. Bírósági ítéletek: nem szerkesztett, ún. anonimizált ítéletek (72 ítéletábrái, 27 járásbírói, 28 törvényszéki ítélet)
4. Jogmagyarázó szövegek:
 - a) tankönyv – preferáltan az 1. pont jogszabályaihoz kapcsolódó tankönyvek (Bevezetés a jogtudományba, Büntetőjog különös rész, Kötelmi jog általános szabályai, Öröklési jog)
 - b) törvények indokolásai (1. és 2. ponthoz kapcsolódó indokolások), összesen 55 törvény indokolása
5. Laikus jogmagyarázó, tájékoztató szövegek
 - a) magyarorszag.hu ügyleírásai
 - b) jogiforum.hu fórum-beszélgetései
6. Beszélte nyelvi törzs: kihallgatások leírt szövegei

A szövegek közül a jogszabályokat, a bírósági ítéleteket és a törvények indoklását egy jogi adatbázisokkal foglalkozó kiadótól kaptuk meg, txt állományok formájában.

A többi adatbázis-elem közül a tankönyveket (4a) a Miskolci Egyetem társ-tanszékeitől kapjuk meg, az 5. adatbázisokat pedig az internetről töltjük le. A vonatkozó webhelyek az alábbiak:

5. a)-hoz (hivatalos tájékoztatás) <https://ugyintezes.magyarorszag.hu/> (Ez kb. 300 ún. ügyleírás)
5. b)-hez fórumok: a jogiforum.hu-ról a következő fórumokat tervezzük letölteni (ez még folyamatban van): Munkajogi kérdések; Munkavállalói rendes felmondás; Jogi tanács kérése; Rágalmazás, személyiségi jogok megsértése; Devizahitel megtámadása; Elévülés erkölcsi bizonyítvány; Büntetőjog; Öröklés.

(B) A hangzó adatbázisok (fenti A/6. rész-törzs) megalapozása már a jelen kutatása előtt megtörtént az FKFP 0653/2000. számú kutatási program támogatásával. Megkezdődött, s immár e projekt részeként befejeződött a nagyrészt rendőrségi, kisebb részben bírósági kihallgatások hangfelvételeinek transzkripciója (vö.: III. pont). Tervezzük a hanganyagtörzs kibővítését további felvételek készítésével. Ennek érdekében felvettük a kapcsolatot a Borsod-Abaúj-Zemplén megyei rendőrfőkapitánnyal, de – legalábbis eddig – nem leltünk pozitív fogadtatásra. Optimistábbak vagyunk a bírósági kihallgatásokkal kapcsolatban. Hozzájárulást kértünk, s kaptunk az Országos Bírói Hivatal elnökétől; ennek alapján konkrét kutatási engedélyt kértünk a Törvényszék elnökétől, akinek támogatása mellett (reményeink szerint) folyamatban van az engedély kiadása. A sikerre számítva beszerzünk egy korszerű hangrögzítő berendezést, s – hallgatói közreműködéssel – folytatjuk a felvételek átírását majd elemzését. Ez alapvetően kvalitatív elemzés lesz.

V. A kutatási honlap kialakítása és feltöltése is a második évre volt ütemezve. Külső szakember bevonásával, de nyitott felhasználású szerkesztő program felhasználásával sikerült kialakítani, s – bár ez még folyamatban van – feltölteni a honlapot. A magyar és angol nyelvű

honlap immár nyilvánosan is elérhető a <https://sites.google.com/site/otkamiskolc2015/> címen, az adatokkal, információkkal való feltöltése folyamatos. A kutatás céljainak, résztvevőinek, előzményeinek és eredményeinek bemutatása mellett, illetve részeként ez kínál felületet a kutatók publikációinak PDF formában való elérhetővé tétele számára is.

VI. Terveink között szerepelt végezetül a tudományos közösség hasonló tudományterületen munkálkodó kutatóival való kapcsolatfelvétel a távlati együttműködés kapcsolati hálójának kialakítása végett. Erre a célra saját rendezvényeinkre (így különösen a 2016. június 22-i budapesti találkozónkra) való meghívást, illetve a kutatóinkhoz intézett konferencia-meghívások elfogadását (pl. „Jog és stílus” Pázmány Péter Katolikus Egyetem, 2015. május 20., „Jogi érvelés és érvelélmélet” BME, 2016. június 3.), azon való részvételt és bemutatkozási lehetőséget használtuk fel.

3. évi részbeszámoló

A munkatervben vállalt feladatok három fő cél szolgálatában álltak: 1. a magyar jogi szaknyelv sajátosságainak feltárása annak megválaszolása céljából, hogy annak mely sajátosságai képeznek akadályt a joghoz való hozzáférés számára; 2. a „Magyar jogi nyelv” című folyóirat útnak indítása; 3. a párhuzamos előrehaladás egyrészt az előzetes eredmények feldolgozásában, másrészt azok disszeminálásában.

1. Nyelvészeti elemzések

A kutatócsoport nyelvész és nyelvész-informatikus szakembereiből szerveződött „alcsoport” vállalkozott a fő kutatási témánk alapját képező, a jogi szakemberek és a laikusok között létrejövő beszédhelyzetek nyelvhasználatának vizsgálatára. E csoport – az elsősorban a nyelvészeti elemzésekre irányuló – egyeztető megbeszélései során (Miskolc, 2016. szeptember 13-14., Budapest, 2017. január 19., Eger, 2017. május 30., Budapest, 2017. június 10., Budapest, 2017. július 24.) Kurtán Zsuzsa irányítása mellett elvégezte a jogi szövegek példaalapú szabályszerűségeit feltáró és leíró elemzés szükséges lépéseinek megtervezését, valamint a részelemzések összehangolását. Kutatócsoportunkhoz, s azon belül a nyelvészek csapatához csatlakozott az év során Balogh Dorka szakfordító-tolmács, a PPKE Jog- és Államtudományi Kar oktatója.

Főbb célkitűzések és eredmények a tárgyévben a következők voltak:

(A) A jogi nyelv rétegzettségének megfelelő részkorpuszok összehasonlító nyelvészeti elemzése

A jogi korpusz építéséhez kutatócsoportunk olyan szövegek kiválasztása mellett döntött, amelyek a jogi eljárás, vagyis a ténymegállapítás, a jogkérdés megállapítása és eldöntése, továbbá az ítélelhozatal folyamataihoz köthető beszédhelyzetekben jellemzőek. Figyelembe vettük például a jogi szaknyelvi regiszter rétegzettségét, a nyelvhasználat hierarchikus jellegét, a nyelvi egyeztetések láncolatait, a jellemző aszimmetrikus helyzetek sorát, a tervezettség szintjét, az elvárt interakció mértékét. Az elemzések során pedig előtérbe helyezük a tájékoztatás és figyelmeztetés műfajait, szövegeit.

Az összeállított korpusz a következő írott szövegtípusokat tartalmazza:

(1) a jogforrások szövegtípusai, ezen belül a kódexjog alapvető, jellemző szövegei (például Polgári törvénykönyv, Büntető törvénykönyv, Polgári perrendtartás), valamint egyéb jogszabályok (további törvények, kormányrendeletek, miniszteri rendeletek);

(2) a működési funkciót betöltő szövegtípusok: bírósági ítéletek járásbírósági, törvényszéki és ítéletábrái szintről;

(3) a magyarázó szövegtípusok: jogszabályok indoklásai és a leendő jogi szakembereknek szánt tankönyvek, valamint laikusok számára készülő jogmagyarázó internetes fórumszövegek.

Az írott jogi korpusz szavainak összege: 2.198.435. Az elsődleges automatizált szövegfeldolgozás adatai szerint, a szófaji megoszlás alapján legmagasabb az aránya a főneveknek (27,9%), az igék 7,8 %-ot, míg a szintaktikai függőségi viszonyok alapján (csak példaként kiemelve) a mellérendelést jelző kötőszavak 3,6%-ot tesznek ki.

Célunk, hogy a számítógépes korpuszelemzés eredményeinek (morfológiai és szintaktikai jellemzők gyakorisága, függőségi viszonyok) értékelésével – a korpusz példái alapján – a köznyelvtől elkülöníthető főbb tartalmi, logikai kapcsolódásokat és nyelvhasználati jellemzőket valamennyi nyelvi síkon feltárjuk és leírjuk. A jogi nyelv rétegzettségének összehasonlító elemzésével vizsgáljuk a köznyelv és a jogi nyelv elhatárolhatóságát, valamint az egyes szövegtípusok jellemzőit a következő főbb területeken: állítmányi és alanyi szerkezetek, mellérendelés, alárendelés, központosítás.

(B) A beszélt jogi nyelv vizsgálata

Az írott szövegek mellé beszélt szövegeket is gyűjtött a kutatócsoport. Egy korábbi FKFP kutatás során megkezdett korpusz-építő munkát terveztük folytatni rendőrségi kihallgatások, bírósági tárgyalások hangfelvételeivel, s ezek átírt szövegeivel, melyeket jogászokkal és laikusokkal készített interjúk egészítenek ki. Míg a 2000-2003 közötti kutatási program során főleg rendőrségi kihallgatások anyagait tudtuk összegyűjteni, a jelen kutatási időszakban – többszöri próbálkozásunk ellenére – hasonló együttműködési megállapodást nem tudtunk kötni a B.-A.-Z. Megyei Rendőr-főkapitánysággal. Sikerral jártunk ugyanakkor a bíróságoknál: megkapva a kutatási engedélyt az Országos Bírói Hivatal elnökétől, meg tudtunk állapodni a Miskolci Törvényszék elnökével, és az önkéntesen vállalkozó bírók, továbbá a tárgyaláson résztvevők eseti engedélyével a 3. kutatási év során folytatni tudtuk a hangfelvételek készítését. Ennek eredményeként a beszélt nyelvi korpusz felépítése jelenleg a következőképpen áll:

2000-2003 között készült felvételek:

Nyomozati eljárások során készült felvételek (gyanúsított és tanúkihallgatások, szembesítések, mélyinterjúk): összesen 868 perc (14 h 28 min). Ebből normalizált és konverzációanalitikai átírással feldolgozva: 733 perc (12 h 21 min), ami 84 %-os feldolgozottságot jelent. (A különbség abból adódik, hogy nem minden mélyinterjú került átírásra, valamint hangminőségi probléma miatt 22 perc hiányzik.) Szószámban kifejezve ez jelenleg: 54.144 szó.

Bírósági tárgyalásokon készült 248 perc (4 h 08 min) hangfelvétel, melyből át van írva konverzációanalitikai és normalizált átírás szerint: kb. 190 perc (3 h 10 min), ami szószámban kb. 15 500 szó.

Ügytípusokat tekintve a nyomozások során testi sértés, súlyos testi sértés, lopás, garázdaság, szeméremszétség, önbíráskodás miatti eljárásokban készültek a felvételek; ezen belül 2 tanúkihallgatás, 11 gyanúsított kihallgatás, 10 szembesítés, 20 interjú, 1 gyám meghallgatása, illetve kötetlen beszélgetések); a bírósági tárgyalások lopás, csalás, garázdaság, szabálysértés ügyében folytak, és tanúkihallgatás, ítélethirdetés, vádlotti meghallgatás, vádbeszéd, védőbeszéd alkalmával készültek.

A 2016-2017 kutatási évben készült felvételek:

Büntető és polgári ügyszakban zajló bírósági tárgyalásokon mintegy 37 óra 46 percnyi felvétel készült, amiben az eljárás professzionális és laikus résztvevőivel készült interjúk is szerepelnek – ezek feldolgozása és átírása még folyamatban van. Eddig – hallgatói közreműködéssel – átírásra került kb. 9 órányi felvétel (572 min), amely 66.362 szószámot jelent. Az átírás gyorsítása érdekében a 4. kutatási évben külön kérelemben kezdeményezni fogjuk egy olyan, fejlesztés alatt álló beszédfelismerő szoftver beszerzését, amely a hangzó anyagból nyers leíratot képes előállítani. Végső célkitűzésként szeretnénk elérni egy, az elemzést szolgáló, aránylag homogén korpusz összeállítását, ami más kutatók és kutatócsoportok számára is, előre pontosan meghatározott feltételek mellett elérhető reprezentatív korpusz kialakítását jelentené.

(C) A magyar jogi nyelv leírásához egyes nyelvi jelenségek bemutatásának megtervezése valamennyi nyelvi síkon – elsősorban morfológiai és szintaktikai szinteken – a teljes korpuszban részkorpuszonként.

A kommunikációs zavarok és félreértések áthidalásához szükséges magyarázatok akkor sikeresek, ha nyelvhasználatuk megfelel az adott kontextus követelményeinek, amihez az alapvető nyelvhasználati jellemzők tudatosítása szükséges. A szaknyelvek feltárásához és leírásához nem csak a köznyelvtől különítjük el a tudományterületek és szakmák regisztereit, hanem a rétegzettségnek megfelelő sajátosságok alapján is. Mindezek háttéréből a jogi szövegekben jellemző magyarázatokat vizsgáljuk a következő kutatási kérdés középpontba helyezésével: milyen nyelvi minták mutathatók ki, amikor laikusok és jogi szakemberek szembesülnek értelmező magyarázatokkal? A jogi nyelv rétegzettségének megfelelő szövegtörzsek elemzésének példáival (jogsabályok, tankönyvek, jogi tájékoztatás stb.) mutatjuk be a magyarázatok nyelvi jellemzőit, középpontba helyezve a magyarázó/következtető mellérendelésekre utaló kötőszavakat.

2. Magyar Jogi Nyelv

A 3. kutatási évre tervezett folyóirat első lapszáma (2017/1.) megjelent. Minthogy évi két megjelenést tervezünk és hirdettünk meg, eredeti terveinktől eltérően a 2017/2. lapszám kiadását is szeretnénk a pályázatból biztosítani – az ennek támogatásához szükséges pénzügyi átcsoportosítás engedélyezését fogjuk kérelmezni a 4. kutatási évben. Ezt követően a lapot nem tudjuk támogatni, ezért olyan kiadót kellett keresni, amely vállalni tudja annak fenntartását. Ezt a HVG-Orac Kiadó, több jogi szakfolyóirat gondozója fogja biztosítani. Az első lapszám nyomtatott változatának kiadása és tiszteletpéldányként való szétküldése mellett elkészült és egyidejűleg szabadon hozzáférhetővé vált annak digitális változata is (<https://joginyelv.hu>), elkezdve a fenntartáshoz szükséges előfizetői kör toborzását is. A szerkesztőbizottság tagjai részben a kutatócsoportból kerültek ki: Dobos Csilla, Kurtán Zsuzsa, Láncos Petra Lea, Szabó Miklós, Tamás Dóra, Tóth Judit, Vinnai Edina.

3. Előrehaladás, disszemináció

A fentiekkel nagyrészt megteremtődtek azok a feltételek, amelyek az elemzések elvégzéséhez és az eredmények összeállításához szükségesek. Mindez természetesen a kutatással párhuzamosan, annak részeként ebben az évben is zajlott. A közzétételre konferenciákon és publikációk révén kerülhetett sor; ezek felsorolását a publikációs lista tartalmazza. Emellett a kutatócsoport két tagja, Vinnai Edina és Zódi Zsolt, az Országos Bírósági Hivatal által szervezett, „Közérthetőség és szakszerűség a bírósági jogalkalmazásban” című kötelező szakmai képzés keretében előadást tartott több törvényszéken is a kutatás témájához kapcsolódóan, ezzel gyakorló jogászok szakmai köreiből is ismertté téve a kutatás témáját és célkitűzéseit. ”

4. évi részbeszámoló értelemszerűen nem készül, hanem helyette a záróbeszámoló vált esedékessé. Pusztán a teljesség kedvéért érdemes felidézni, hogy az utolsó évre mértéktartó célkitűzéseket és terveket tűztünk magunk elé. Ezek egyik csoportja az összeállított és korpuszba szervezett nyelvi adatok elemzése, értékelése és a következtetések levonása; a másik pedig – ennek következményeként – egy záró konferencia vagy workshop szervezése, valamint eredményeink egységes tanulmánykötetben való közzététele.

Az elemzések mindenre kiterjedő elvégzése két okból nem volt lehetséges. Az egyik ok az, hogy az adatgyűjtés és -feldolgozás, elemzésre előkészítés párhuzamosan zajlott a kutatás egész tartama alatt. Különösen igaz ez a hangzó korpusz fejlesztésére: a hangfelvételek előkészítése (engedélyek, egyeztetések), a felvételek rögzítése, a rögzített felvételek átírása – az összesen 60 órányi hanganyagot tekintve – tartós feladatnak bizonyult. Az írott szöveges részkorpuszok terjedelme elvileg korlátlanul növelhető, azonban statisztikai elemzésre való előkészítésük idő- és munkaigényes. A másik ok az, hogy a grammatikai szintű statisztikai elemzésen túlmutató szemantikai szintű számítógépes elemző program (T-Lab) birtokbavétele – kiválasztása, beszerzése, kézikönyvének megismerése, a kínált eszközök használatának megismerése, kipróbálása – hasonlóképpen idő- és munkaigényes feladatnak bizonyult. Mindez csak azt jelenti, hogy a nyelvi korpuszokban és a rendelkezésünkre álló módszerekben rejlő lehetőségeket nem tudtuk teljesen kimeríteni; azt nem, hogy lemondanánk ezekről. Az alábbi IV. részben felvázoltak szerint folytatni és kiteljesíteni szándékozunk a kínálkozó elemzési irányokat.

A kutatás lezárásához méltó, egész napos konferenciát sikerült szervezni az MTA Miskolci Akadémiai Bizottsága székházában 2018. május 25-én. A konferencia résztvevői – a nyitó konferencia szélesebb körű meghívottaival ellentétben – a kutatási program résztvevő vagy támogató kutatói voltak. Annak a célnak az elérése érdekében, hogy a tanulmányra fejlesztett konferencia-előadásokat befoglaló majdani kötet az integráltságnak a megszokottnál magasabb szintjét érje el, az absztraktok és munkaanyagok egymás közti köröztetésével igyekeztünk alaposan előkészíteni a konferenciát. Ez különösen sikeresnek mondható az első, nyelvészeti blokkban, de jellemzi a kötet második részét is. Egy hiányt kellett tudomásul vennünk: a fent jelzett okok miatt nem készült el időre, s kimaradt a kötetből Szendi Attila „A jogi nyelv érthetőségének történeti dimenziói új elemzés módszertani megközelítésben” című írása. A téma túlnőtt egy tanulmány keretein, s nagyobb formát igényel, de a jelen beszámoló Függelékében megtalálható az előadás jelentősen kibővített írott anyaga. A megjelent kötet írásai tehát a következők:

Vincze Veronika: *A Miskolc Jogi Korpusz nyelvi jellemzői*

Dobos Csilla: *Állítmányi szerkezetek a magyar jogi nyelvben*

Balogh Dorka: *Alanyi szerkezetek a magyar jogi nyelvben*

Kurtán Zsuzsa: *Mellérendelés a magyar jogi nyelvben*

Sajgál Mónika: *Alárendelés a magyar jogi nyelvben*

Kurtán Zsuzsa: *Központozás a magyar jogi nyelvben*

Markovich Réka: *A jogszabályok logikai mélystruktúrája*

Gedeon Magdolna: *A magyar jogi nyelv történetéről*

Farkasné Puklus Márta: *Tolmácsolás a tárgyalóteremben. »Elvesztettem a fonalat, hol tartunk?«*

Zódi Zsolt: *A jog érthetőségének határai. Meg tudják-e oldani a nyelvészek a jogi szövegek érthetlenségének problémáját?*

Pázmándi Kinga: *A jog nyelvi dimenziói és a fogyasztóvédelem. »Devizahiteles« perek szereplőinél azonosítható nyelvi attitűdök*

Vinnai Edina: *»Megértette a figyelmeztetést?« A figyelmeztetések és tájékoztatások közlése a büntetőeljárásokban*

Szabó Miklós: *A szavak ura: Pragma vs. praxis*

III. A teljesítmények tematikus áttekintése

1. Időközi kérelmek és NKFIH engedélyek a pénzügyi tervtől való eltérésre

A kutatás előrehaladtával, különösen az utolsó évben szembesültünk azzal, hogy eredeti célkitűzéseink közül melyek azok, amelyek nem, vagy nem a tervezett mértékben valósulnak meg (pl. külföldi konferenciákon való részvétel) és a pénzügyi terv áttekintése alapján személyi és dologi kiadásaink átcsoportosítását kértük az NKFI Hivataltól, melyekre az alábbi engedélyeket kaptuk:

A **2017.10.31-i** kérelme és annak kiegészítései alapján az NKFI Hivatal kutatás-fejlesztésért felelős Elnökhelyettese engedélyezte:

- Az Imre és Társa Bt.-vel való szolgáltatási szerződések megkötését a kérelemben leírtaknak megfelelően,
- A Magyar Jogi Nyelv folyóirat 2. lapszámának megjelenését,
- a 2018. májusi konferencia zárókötetének költségeit,
- a tervezett szálláshely költségek megemelését
- egy szövegelemző szoftver (TLAB) beszerzését, és a beszédfelismerő-szövegszerkesztő szoftver licenzének megvásárlását.
- Farkasné Puklus Márta, Dobos Csilla és Balogh Dorka csatlakoztatását a projekthez.

A **2018. május hó 22.** napján kelt kérelméhez az NKFI Hivatal Kutatás-fejlesztési Elnökhelyettese hozzájárult, az alábbiak szerint:

1. anyanyelvi lektor bevonása és finanszírozása Zódi Zsolt előadás tanulmányához, mely megjelenik az International Journal of Law in Context folyóiratban.
3.1. soron „külföldi utazások” fedezet rendelkezésre áll a szükséges kb. 50e Forinthez.
2. a 3.1. és a 3.3. sorokról a 3.2. sorra (készletbeszerzés) lektorálási díj, illetve további könyv- és irodaszerek beszerzése számolható el ezáltal.

Vonatkozó szabályzat: K_2014 típusú TSz. 4.2.15. A költségterv egyes jogcímei közötti átcsoportosítás előzetes engedélykérési kötelezettség esetén lehetséges

A **2018. május hó 30.** napján kelt kérelméhez az NKFI Hivatal Kutatás-Fejlesztési Elnökhelyettese hozzájárult az alábbiakhoz:

- technikai munkatárs csatlakoztatása
- név: Flaskó László
- Intézmény: Miskolci Egyetem, Informatikai Szolgáltató Központ
- beosztás: számítástechnikai munkatárs
- Keresetkiegészítés formájában
- 1x 200.000,- Ft értékben
- 1.2.2. klts.soron.

Vonatkozó szabályzat: K_2014 típusú PSz 6.1. pont: A kutatással kapcsolatos feladatokra alkalmazhatók: hallgatók, doktorjelöltek, posztdoktorok és technikai segédmunkatársak.

6.1.4. pont: Előre meghatározott, tételesen kijelölt és elszámolható technikai feladatra történő kifizetés az intézménnyel munkaviszonyban állók részére kereset-kiegészítés formájában történhet (1.2., 1.5. sorok).

A fedezet rendelkezésre áll.

A **2018. június hó 20.** napján kelt kérelmére vonatkozóan tájékoztatom, hogy az NKFI Hivatal Kutatás-fejlesztési Elnökhelyettese hozzájárult az alábbiakhoz:

záró kötet lektorálása

- Dr. Bodnár Ildikó által
 - Miskolci Egyetem Modern Filológiai Intézetének volt munkatársa
 - szakmailag elismert docens
 - megbízási szerződés keretében
 - személyi soron fedezet rendelkezésre áll
 - 150.000 Ft értékben,
- valamint 2018. 08. hó 31. napjáig résztvevőként regisztrálható.

Vonatkozó szabályzat: K_2014 típusú TSz 6.3.1. Fordítási, nyelvi és szakmai lektori költségek minden esetben csak az OTKA illetékes kollégiumának előzetes engedélyével számolhatók el.

6.1.5. Az intézménnyel munkaviszonyban nem állók (nem kutatók) részére technikai feladatokkal kapcsolatos munkák (lásd 6.1.4.) megbízási szerződéssel vagy számla ellenében fizethetők ki. Azokat a költségterv egyéb megbízási (1.4.) vagy egyéb költség (3.3.) során kell elszámolni.

Dr. Bodnár Ildikó EPR azonosítója: 996280

Dr. Bodnár Ildikó csatlakoztatása a projekthez folyamatban van, 0,1 > minimális FTE értéken.

2. Kutatócsoport szervezése, fenntartása

Minden további eredmény elérésének legelső feltétele munkamegosztásra és együttműködésre egyaránt képes és hajlandó csapat megszervezése és összetartása. Az induló kutatócsoport tagjai voltak: jogászként Szabó Miklós, Vinnai Edina, Gedeon Magdolna, Pázmándi Kinga és Zódi Zsolt, aki egyúttal informatikai ismeretekkel is rendelkezik, történész-informatikusként Szendi Attila, nyelvész-informatikusként Vincze Veronika, nyelvészként pedig Kurtán Zsuzsa és Sajgál Mónika (akadémiai háttér hiányában mindketten a kutatás teljes ideje alatt megbízási, illetve szolgáltatási szerződés keretében működtek közre a vállalatok teljesítésében). A nemcsak szakmailag-tudományosan, hanem intézményi háttérben is heterogén csoport integritását a feladatok és eredmények áttekintésére irányuló rendszeres megbeszélések szolgálták. Ezeket félévente legalább egyszer, változatos helyszíneken szerveztük meg. Alkalmanként – kapcsolatépítés és látókörünk tágítása érdekében – szélesebb körből is meghívtunk hasonló érdeklődésű szakembereket.

A kutatócsoportot elindító tagok mindvégig a csoport tagjai maradtak, ha saját intézményes háttérük időközben módosult is. Ami változás történt, az – szinte a legelső időponttól – új jelentezők csatlakozása volt, akik vagy „hivatalosan” is a kutatócsoport közreműködőivé váltak, vagy „csak” külső pártolóként, de a többiekénél nem kisebb aktivitást mutatva vettek részt a munkánkban: Dobos Csilla, Markovich Réka, Farkasné Puklus Márta, Balogh Dorka. Innen „sarjadt ki” az időközben megalapított Magyar Jogi Nyelv szerkesztőbizottsága is, a kutatócsoport egyes tagjain (Kurtán Zsuzsa, Dobos Csilla, Vinnai Edina, Szabó Miklós) túl befogadva másokat is (Tóth Judit, Tamás Dóra, Láncoz Petra Lea). A kapcsolatépítést segítette a MANYE és MANYSI rendezvényein, konferenciáin rendszeressé vált részvételünk is.

3. Intézményes kapcsolatépítés

Kutatási tervünk egy részének megvalósításához együttműködésre volt szükség más – hatósági, bírósági – intézményekkel is. A vizsgált jogi eljárások konkrétan (nyomozó)

hatósági, bírósági eljárásokat jelentenek, amelyek szenzitív jellege miatt a hozzáférésünk engedélyekhez volt kötve. Míg a továbbfolytatni szánt FKFP kutatásunk során együttműködési megállapodást tudtunk kötni a Borsod-Abaúj-Zemplén Megyei Rendőrfőkapitánysággal, amit viszont nem sikerült megtennünk a bírósági szervezettel, addig a jelen kutatásunk előkészítéseként többszörös kísérlet ellenére sem sikerült támogatást kapnunk a rendőrségtől, azonban az OBH elnökének támogatásával kutatási engedélyt kaptunk a Miskolci Törvényszék elnökétől. Ez a beszélt jogi nyelvi részkorpusz hangzó anyagainak felvételét érintette – oly módon, hogy a részkorpusz nyomozási kihallgatásokról készült felvételei mintegy másfél évtizeddel korábban kerültek rögzítésre, mint a bírósági tárgyalásokról készült felvételek többsége. Következésképpen azonban e részleges sikertelenség azt is okozta, hogy nem sikerült a – most is tervbe vett – longitudinális felvételeket elkészíteni, a hangfelvételekhez tartozó jegyzőkönyveket beszerezni. Ügyvédi irodákkal és civil szervezetekkel nem került sor kapcsolattépítésre.

Más természetű – piaci alapú – együttműködést igényelt munkatervünk azon elemeinek realizálása, ahol a kellő kompetenciával vagy kapacitással a csoporton belül nem rendelkezünk. Ide sorolhatjuk a kutatási honlap kialakítását (ebben a Miskolci Egyetemi Könyvtár munkatársa, Molnár Csaba működött közre), a hangfelvételek utolsó adagjának transzkribálását beszédfelismerő szoftver segítségével (szolgáltatási szerződés keretében a Belux Kft. végezte el a feladatot), az elemzendő szótár lemmatizálását (a Miskolci Egyetem Informatikai Szolgáltató Központ munkatársa, Flaskó László volt segítségünkre).

4. Jogi nyelvi korpusz kialakítása

Célul tűztük ki a kutatócsoportunk által a 2000-es évek elején végzett kutatás (FKFP–0653/2000) folytatását, kiteljesítését. Ez lényegében a jogi (nyomozati és bírósági) eljárásokról készült hangfelvételekből és azok transzkripciójából összeállított beszélt jogi nyelvi korpusz felépítését és további elemzésekre alkalmassá tételét jelentette. Ez – egyetemi hallgatók bevonásával – lényegében sikerült: létrejött egy 60 órányi hangfelvételtől és annak transzkripciójából álló beszélt jogi nyelvi korpusz. A grammatikai elemzéshez természetesen – a jogi nyelv számára elsődleges – írott nyelvi (rész)korpuszokra is szükség volt. Ezek három szövegtípusba sorolhatók: a tételes jogi normaszövegek (azon belül kódexjog és más jogszabályok); a jog működési szövegei (azon belül a jogalkotást képviselő, törvényjavaslatot előterjesztő miniszteri indoklás és a jogalkalmazás során születő bírói ítéletek szövegei); s a jogi szövegeket magyarázó szövegek (ilyenek a tankönyvek és az internetes jogi fórumok). Az így előállt teljes (írott és beszélt nyelvi) korpusz – a Miskolc Jogi Korpusz – mindösszesen 2.386.902 szövegszót és 147.209 mondatot tartalmaz.

A korpusz továbbépíthető és továbbhasznosítható. A továbbépítés már eddig is megtörtént, valahányszor egy készülő dolgozat specifikusan valamely, a korpuszban nem szereplő jogszabály vagy ítélet elemzését kívánta meg. A beszélt nyelvi részkorpuszon kívüli, írott nyelvi részkorpuszok bővítése csupán technikai erőfeszítéseket igényel. A továbbhasznosítás saját terveink között is szerepel, de más kutatások számára is lehetővé kívánjuk tenni, élve a kutatási honlap kínálta lehetőségekkel.

5. A magyar jogi nyelv elsődleges grammatikai leírása

Fő kutatási célunk – a tisztességes eljárás feltételeként a joghoz való hozzáférés érvényesülése – annak vizsgálatára irányult, hogy a jog nyelvi megformáltsága miképpen befolyásolja a jog mint társadalmi intézményrendszer hatékony és korrekt működését. Vizsgálati tárgyként ez lényegében a jogi nyelv (köz)érthetőségének kérdését jelenti. E kérdés egyik előkérdése

annak empirikusan alátámasztott megválaszolása, hogy a magyar jogi nyelv milyen specifikus jellemzőkkel bír a hétköznapi nyelvhez képest, és milyen különbségek tárhatók fel a jogi illetve a hétköznapi kontextusban zajló diskurzusok között. A magyar jogi nyelv (beszélt és írott változatainak) rendszeres, szisztematikus leírása eddig nem állt rendelkezésre; e munka megkezdésére és megalapozására vállalkoztunk. A jogszabályi és jogalkalmazói nyelvhasználat kvantitatív és kvalitatív elemzésével meg kívántuk adni a standard magyar jogi nyelv jellemzőit, amelyhez képest a joggyakorlaton belüli egyes nyelvhasználati módok jellemezhetővé válnak.

A válaszhoz kellett felépíteni a Miskolc Jogi Korpuszt, amelyet a köznyelvi korpuszok közül a Szeged Korpuszsal egybevetve váltak kimutathatóvá a jogi nyelv megértést nehezítő jegyei. Ezek anyagának számítógépes statisztikai elemzése tette lehetővé a magyar jogi nyelv elsődleges leírását a legfőbb grammatikai jegyekre kiterjedően. Ezt a munkát kutatócsoportunk nyelvész tagjaiból álló csapata Kurtán Zsuzsa vezetésével végezte el. Az elemzés és a kidolgozás legapróbb részleteit – a példák jelölésétől a szövegezés szerkezetéig – egyeztetve és egymással folyamatosan konzultálva, megbeszéléseket tartva mintegy két év alatt jutottak el megállapításaik megfogalmazásáig. Ahogy a zárókötet első hat tanulmánya mutatja, a kutatási program egészén belül is sikerült önálló egységet képezni ebből a részfeladatból. A magyar jogi nyelvről szólva nem kell immár intuitív vagy konszenzuális tételekre hagyatkozni, hanem egzakt módon megalapozott állításokra támaszkodhatunk (még ha ezek némelyike tartalmilag egybe is esik a korábbiakkal).

6. Disszemináció (1) Nyitókonferencia

A következőkben a kutatásunk láthatóvá tétele, (rész)eredményeink megismertetése, a jog és nyelv kutatásokba való bekapcsolódásunk érdekében tett lépéseinket foglaljuk össze.

2014. november közepén „A jog nyelvi dimenziója” címmel tartottuk meg a kutatási projektet megnyitó konferenciát, egyrészt megosztva és megvitatva kutatási elgondolásainkat, másrészt muníciót várva és kérve a tágabb jogelméleti közösségtől a jog és nyelv kapcsolatának releváns aspektusairól. Az ekként két blokkból álló, kétnapos konferencia tanulmányokká érlelt anyagát a tervezettek szerint közzétettük a Szabó Miklós szerkesztette *A jog nyelvi dimenziója* című kötetben (Miskolc: Bíbor, 2015. [Prudentia Iuris sorozat 31.] 296 o. ISBN 978-615-5536-07-6). A nyomtatott köteteket tiszteletpéldányként megküldtük annak a szakmai körnek, amelynek érdeklődése közel áll kutatási témánkhoz. A kötet anyagát PDF formátumban feltöltöttük a kutatási honlapunkra, ahol bárki számára hozzáférhető.

7. Disszemináció (2) Kutatási honlap

Egyik első vállalkásunk volt a kutatási honlap létrehozása, feltöltése és folyamatos szerkesztése is a kutatás végéig. Ez megtörtént; azonban vállalkásunkat kiegészítjük a honlap további fenntartásával. Mint utaltunk rá, sejtető volt, s bizonyossá vált, hogy bár a kutatási projekt lezáródik, maga a kutatás mehet és mennie is kell tovább. A kutatás identitását, elérhetőségét, szócsövét ez a honlap tudja fenntartani egyéb intézményes keret hiányában. Táblázatok, grafikonok, elemzések, irodalomjegyzékek formájában megszülető és közzétehető adatok és következtetések – gyakorlatilag terjedelmi korlát nélkül – megoszthatók és diskurzus tárgyává tehetők. Erre a lehetőségre utaltunk a fenti 3. pontban is. Szándékunk kiterjed az eddigi, s a jövőben megszülető anyagokra is. A <https://sites.google.com/site/otkamiskolc2015/> címen elérhető honlap kialakításához külső szakértő segítségére volt szükség, azonban annak egyszerűsége nem igényelt jelentősebb forrást, fenntartása pedig semmilyen költséget sem.

8. Disszemináció (3) Magyar Jogi Nyelv

Ugyancsak az eredeti vállalásainknak megfelelően indítottuk útjára 2017 tavaszán a „Magyar Jogi Nyelv” című periodikát. A periodika ötletadója a „Magyar Orvosi Nyelv” című folyóirat volt, jóllehet annak ars poeticáját – az orvosi nyelvben oly gyakori latin és angol szavak magyar nyelvű szakkifejezésekkel való felváltását és ezek meghonosítását – nem vettük át és nem is kellett átvennünk. A jogi nyelvünkben ugyanis a magyar államnyelvvé tétele által kikényszerítve ez a magyarítás lényegében a XIX. századdal lezajlott – még ha az angol nyelv térhódítása a többi szaknyelvhez hasonlóan a magyar jogi nyelvet is fenyegeti. Saját ars poeticánknak tehát nem a nyelvhelyességi útmutató vagy a „nyelv-szépítészet” vállalását tesszük meg, hanem annak tudatosítását mind a laikus, mind a nyelvész vagy jogász szakember olvasóban, hogy milyen jelenségei és következményei vannak a jog és a nyelv interferenciájának, hogy mi a jelentősége a nyelvi megformáltság mikéntjének a jog társadalmi intézményrendszere elvárt vagy remélt működése számára. Kutatási projektünk az első, majd – kérelmünknek eleget téve – a második (2017/1., 2017/2.) lapszám kiadását és mutatványszámként való szétküldését finanszírozta. Olyan kiadót kellett tehát találnunk, amely a lap fenntartását a projekt lezárulta után is vállalja; ez a – jogi szakfolyóiratokat széles körben gondozó – HVG-Orac Kft. lett, amely a 2018/1. lapszámot már megjelentette, s a 2018/2. lapszám szerkesztése most van folyamatban.

9. Disszemináció (4) Konferenciák

Az, hogy lehetőséget kívánunk biztosítani a kutatásban részt vevő valamennyi kutató számára, hogy egy alkalommal részt vegyen valamely releváns konferencián vagy más szakmai fórumon, két célt kívánt szolgálni. Egyrészt azt volt hivatott elősegíteni, hogy tudomást szerezzünk az aktuális kutatási irányokról, a kutatási terület formálódásáról, jelentős szerzőkről és művekről. Másrészt azonban azt is lehetővé kívánta tenni, hogy a hasonló kutatásokkal foglalkozó külföldi szakemberek is megismerjék Magyarországon egyedülálló kutatásunkat, vagyis nemzetközi fórumokon is meg tudjunk jelenni eredményeinkkel. Ez a törekvésünk sem maradt sikertelen; számos hazai szereplés mellett a következő idegen nyelvű fellépések idézhetők ide:

2014

- Vinnai Edina: *Forensic Linguistics in Hungary*. Mainz, Németország, 2014. szeptember 4-6. (Angol nyelvű előadás, a Germanic Society for Forensic Linguistics „2014 Roundtable” című konferenciája)

2015

- Szabó Miklós: *Deep Structure of Language(s) of Law*. Washington D.C. USA, 2015. július 31. (Angol nyelvű előadás, az IVR [Nemzetközi Jogfilozófiai Társaság] XXVII. „Law, Reason, Emotion” című világtalálkozója, Georgetown Conference Center)

2017

- Vinnai Edina: *Interdisciplinary Research on the Judiciary in Hungary*. Wrocław, Lengyelország, 2017. április 26-28. (Angol nyelvű előadás, a Central and Eastern European Forum of Young Legal, Political and Social Theorists, a Centre for Legal Education and and Social Theory (CLEST) és a University of Wrocław „9th Forum, Democracy, Academia and the Political” című konferenciája)
- Szabó Miklós: *Linguistic conditions of access to law and fair process*. Lisszabon, Portugália, 2017. július 19. (Angol nyelvű előadás, az IVR [Nemzetközi Jogfilozófiai Társaság] XXVIII. világtalálkozóján)

- Zódi Zsolt: *Judicial Style and Understandability of Legal Documents*. Lisszabon, Portugália, 2017. július 19. (Angol nyelvű előadás, az IVR [Nemzetközi Jogfilozófiai Társaság] XXVIII. világkongresszusán)

2018

- Vinnai Edina: *The fulfilment of right to information in Hungarian legal practice*. Koppenhága, Dánia, 2017. (Angol nyelvű előadás, a Germanic Society for Forensic Linguistics „2017 Roundtable” című konferenciája, 2017. augusztus 24-27.)
- Farkasné Puklus Márta: *Gerichtsdolmetschen in Ungarn*. Montpellier, Franciaország, 2018. (Német nyelvű előadás, a Gesellschaft für Sprache und Sprachen (GeSuS) „Traditionen und Erneuerungen” című konferenciája, 2018. április 5-7.)
- Vinnai Edina: *Plain Language in Law in Hungary*. Montpellier, Franciaország, 2018. (Angol nyelvű előadás, a Gesellschaft für Sprache und Sprachen (GeSuS) „Traditionen und Erneuerungen” című konferenciája, 2018. április 5-7.)

10. Disszemináció (5) Publikációk

A grammatikai analízis mellett immár a joggyakorlatot is vizsgáló kvalitatív, kisebb részben kvantitatív szemantikai és pragmatikai elemzéseket is végeztek kutatócsoportunk tagjai. Ezek kiterjedtek az esettanulmányoktól az elméletképzésig, a nyelvtörténeti kutatásoktól a fordítás és tolmácsolás problémáira való kitekintésig. Vizsgálataink eredményeit folyamatosan publikáltuk kutatási honlapunkon, a kínálkozó konferenciákon, a nyelvészeti és jogi folyóiratokban, szakkönyvekben, alkalmi előadásokban. Ezek közül azokat, amelyek megfeleltek a publikációként elismerés kritériumainak, évente feltöltöttük a publikációs tárhelyre, a záróbeszámoló idejére pedig összesen közel 100 tételből áll a kutatócsoport közleményeinek listája (az elektronikus rendszerből kinyomtatva kerül csatolásra).

11. Disszemináció (6) Zárókonferencia

A terveinknek megfelelően 2018. május 25-én tartottuk meg kutatásunk záró konferenciáját, „A jogi nyelv érthetősége és a joghoz való hozzáférés” címmel a Miskolci Akadémiai Bizottság dísztermében. Mint jeleztük, külső előadót nem hívtunk meg; az előadásokat az egynapos konferencia három blokkjában kutatócsoportunk tagjai tartották meg:

Vincze Veronika: *A Miskolc Jogi Korpusz jellemzői*

Dobos Csilla: *Állítmányi szerkezetek a magyar jogi nyelvben*

Balogh Dorka: *Alanyi szerkezetek a magyar jogi nyelvben*

Kurtán Zsuzsa: *Mellérendelés a magyar jogi nyelvben*

Sajgál Mónika: *Alárendelés a magyar jogi nyelvben*

Kurtán Zsuzsa: *Központosítás a magyar jogi nyelvben*

Gedeon Magdolna: *A magyar jogi nyelv története*

Szendi Attila: *A jogi nyelv érthetőségének történeti dimenziói új elemzésmódszertani megközelítésben*

Farkasné Puklus Márta: *Tolmácsolás a tárgyalóteremben*

Markovich Réka: *A jogszabályok logikai mélystruktúrája*

Zódi Zsolt: *A Plain English Movement és a jogi szövegek érthetőségének határai*

Pázmándi Kinga: *A jog nyelvi dimenziói és a fogyasztóvédelem*

Vinnai Edina: „*Megértette a figyelmeztetést?*” *A figyelmeztetések és tájékoztatások közlése a büntetőeljárásokban*

Szabó Miklós: *A szavak ura: Pragma vs. praxis*

A konferencia időpontja megengedte a tanulmánykötetként megformált anyagának megjelentetését a projekten belül. Ez Szabó Miklós és Vinnai Edina szerkesztésében *A törvény szavai* címmel és *Az OTKA-112172 kutatási zárókonferencia anyaga, Miskolc, ME – MAB, 2018. május 25.* alcímmel jelent meg (Miskolc: Bíbor, 2015. [Prudentia Iuris sorozat 33.] 313 o. ISBN 978-615-5536-61-8). A nyomtatott példányok ezúttal is tiszteletpéldányként szolgálnak, de PDF-ként egyidejűleg elérhetővé tettük a kutatási honlapunkon is.

IV. Elmaradások, jövőre vonatkozó tervek

1. Kisebb elmaradások

Az eredetei vállalásaink közül nem tudtuk valóra váltani a *mediációs eljárások* bevonását kutatási programunkba. Ez azt szolgálta volna, hogy ezek kutatásunk tárgyává tételével összehasonlító vizsgálat alá vehessük a konfliktuskezelés formális jogi (hatósági és bírósági) és a közvetítői (alternatív) modelljét, melynek révén értékelhetővé vált volna a fair eljárás nyelvi feltételeinek különbsége is. Hamar kiderült azonban, hogy ez a vállalás túlnőtt volna kutatásunk keretein, s annak kis leágazása helyett valójában külön, önálló kutatási programot érdemelne. Kapacitásaink hiányában tehát erről le kellett mondanunk.

Nem kapacitásunk szűkössége, hanem a kutatási részprogramok párhuzamossága miatt maradt el a *szóbeli és írásbeli nyelvhasználat különbségeinek* vizsgálata, amit a joghoz való hozzáférés és jog előtti egyenlőség szempontjából tarthatunk kiemelkedő fontosságúnak. Azonban a szóbeli részkorpusz kialakítása – az engedélyektől az átírásig – párhuzamosan zajlott a kutatás többi elemével, s mire elemezhető állapotba került, az összehasonlító elemzésre már nem maradt kellő idő. A későbbiekben, szabadulva az idő szorításából, természetesen erre a kérdésre is vissza tudunk térni.

Ismét más okból kellett lemondanunk arról az elképzelésről, hogy megkíséreljük néhány ügy végigkövetését az eljárás kezdetétől annak jogerős befejezéséig. A *longitudinális vizsgálat*hoz több jogalkalmazó szerv egymást követő engedélyére lett volna szükség, mégpedig a kihallgatáson vagy tárgyaláson való jelenléten túl az iratokba való betekintés és másolatok készítésének engedélyére is. Ez elérhetetlen volt, s a jövőben sem gondoljuk, hogy elérhető lesz, így erről a tervről le kellett mondanunk.

Folyamatban van még azon vállalásunk teljesítése, hogy a beszélt jogi nyelvi részkorpuszt hangfelvételek és transzkripciók rendszerezett együtteseként a kutatási honlapunkon keresztül más kutatók számára is *hozzáférhetővé* tegyük. Első lépésként egy rövidebb hangfelvétel-részletet és az ahhoz tartozó átírást kívánjuk nyilvánossá tenni a kutatási honlapon, majd érdeklődő kutatók részére – az adatvédelmi szabályok betartása és betartatása mellett – közvetlen hozzáférést biztosítanánk további elemzések elvégzése érdekében.

2. PAPERPILE: bibliográfiai hivatkozáskezelő használata az OTKA projektben

Kutatócsoportunk a kutatás menetében információcserére, közös munkavégzésre, adatok, források, szakirodalom megosztására a Google Drive felülete által kínált lehetőséget vette igénybe. Ez hozta magával azt a javaslatot az Egyetemi Könyvtár részéről, hogy közös adatbázis – elsősorban a jog és nyelv tárgykörébe sorolható – szakirodalom gyűjtését és virtuális szakkönyvtárként való kialakítását a PAPERPILE nevű szoftver segítségével oldjuk meg. Ez, továbbgondolva, lehetővé teszi az OTKA-projekt lezárulta után is ezen források fenntartását, fejlesztését és ennek révén az ugyanezen tárgyat kutatók erejének, információinak egyesítését. Ekként bázisként szolgálhat a folytatni tervezett kutatások számára is. Az ehhez szükséges támogatást és segítséget könyvtárunk szakembereitől megkaptuk, s arra a jövőben is számíthatunk. A szoftverről a következőket érdemes tudni:

A Miskolci Egyetem Könyvtár, Levéltár, Múzeum referencia menedzser szoftvereken keresztüli digitális metaadat és teljes szöveges adatok kezelése, szolgáltatása egyetemi oktatók és kutató csoportok számára nyújtott szolgáltatása 2017 nyara óta érhető el. Ehhez lehetőséget, alapot a felhő alapú szoftverek használata biztosít, pontosabban a Paperpile, a Google Drive és a Facebook for Workplace (ezen két utóbbi szoftver a felsőoktatási intézmények számára ingyenesen használható).

A könyvtár számára értéknövelt szolgáltatás biztosítására ad lehetőséget, amely alapvető igény lehet a digitális dokumentumok korában. Fontos megjegyezni, hogy ennek a szolgáltatásnak az igénybe vétele önkéntes, egyetlen egyetemi kutató vagy kutatócsoport sincs kötelezve a használatára. A Könyvtár az egyéni kutatókat ugyanúgy támogatja, mint a kutatócsoportokat, és ugyanúgy segítik/támogatják mintha egyéb, hagyományos könyvtári szolgáltatásokat vennének igénybe.

A szolgáltatás használatához felhasználói fiókok kerültek kialakításra, többek között az OTKA pályázatban résztvevő oktatók, kutatók számára is. Ehhez a fiókhoz a kutatási projekt összes tagja hozzáfér.

A közösen egyeztetett kutatói igények – kutatás témája, irodalomkutatás időintervalluma, nyelvi kritériumok, kapcsolódó magyar és idegen nyelvű kereső kifejezések, stb.) – megbeszélése után a könyvtáros az elérhető forrásokban (előfizetési adatbázis, nyílt források, stb.) elvégzi az irodalomkutatást és egy bibliográfiát állít össze a kutató/kutatócsoport számára. Ők ezt ellenőrzik, majd az esetleg szükséges pontosítások, újabb keresések után a kutatók felől érkező pozitív jóváhagyás után a kiválasztott tételeket, lehetőség szerint teljes szöveggel teszi elérhetővé a kutatók számára.

A könyvtár / könyvtáros emellett menedzseli az intézményi G-Drive felületet is (ezen keresztül történik a hozzáférés biztosítása), kezeli a Paperpile előfizetéseket (igény szerint), a Facebook for Workplace segítségével intézi a szükséges adminisztrációt, a kutatócsoportoknak közös kommunikációs felületet biztosít, felveszi/letiltja a felhasználókat, monitorozza a szükséges feladatokat/munkavégzést stb.) valamint szintén menedzseli a felhős „infrastruktúrát” is.

A felhő alapú segédeszközök használatának előnyei:

- nem szükséges újabb program telepítése (csak egy kiegészítőt kell a Google Chrome böngészőbe beilleszteni)
- eltérő informatikai háttérrel is képes mindenki kezelni (megfelelő, előzetes bemutatás után)
- segíti a megosztott közösségi munkát

A Paperpile alkalmazás bemutatása:

A Paperpile online, web/felhő alapú kereskedelmi referencia-kezelő szoftver, Google Chrome böngésző bővítményeit használva, valamint Google Dokumentumok és a Google Scholar integrációval. A 2012-ben alakult, és a Paperpile LLC gyártja. Használható Windows, Linux, MacOS, és Chrome OS platformokon. HTML5, Javascript, jQuery és Ext JS eszközöket használ. A Google Chrome alatt automatikusan frissül.

A Paperpile adatokat importál a kiadók webhelyeiről és adatbázisokból, mint a PubMed, a Google Tudós, a Google Könyvek és az arXiv. A Paperpile segítségével visszakereshető és tárolhatóvá válnak a közzétett PDF-fájlok, amelyek átkerülnek a felhasználó Google Drive-fiókjába. Segítséget nyújt a Google Dokumentumokban található hivatkozások és bibliográfiák elkészítésében, formázásában, ezzel lehetővé téve a dokumentumok együttes szerkesztését. Paperpile BibTeX és RIS formátumokban képes exportálni és importálni az adatokat, és migrálhatja a Mendeley, a Zotero és a Papers adatait is.

A kutatási munkához gyorsan és egyszerűen hozzáadhatóak a citációk és bibliográfiák. Egyszerű, helyes, gyors és teljes mértékben együttműködő.

Fontosabb tulajdonságai:

1. tudományos kutatócsoportok megsztott tudásbázisa
2. online referencia menedzser program
3. tárhelyként is használható a teljes szöveges dokumentumokhoz
4. beépített pdf olvasóval rendelkezik
5. jegyzetelhető
6. exportálható
7. metaadatok kezelésére is alkalmas
8. szabványos metaadatok áttemelése megoldott
9. webintegráció
10. teljes szöveg egyben legrabbelhető
11. Google Drive-val való szinkronizálás
12. alacsony ár
13. bibliográfiai formátumok ismerete és kezelése
14. Google Docs-ba integráltság:
 - a. hivatkozások egyszerű beillesztése, és kezelése,
 - b. felhasznált irodalom készítése egyszerű
 - c. közös dokumentum szerkesztési lehetőség
 - d. egyszerű publikálás
15. Csak Google Chrome böngészőben működik – sajnos nem működik a Firefox, IE alatt.
16. Csak hivatkozásokat / hivatkozási listákat helyez be a Google Dokumentumokba
17. Kiváló a referenciák és a hozzá kapcsolódó PDF-dokumentumok gyűjtése és kezelése olyan környezetben, amely nem függ attól, hogy Ön egy adott helyen tartózkodik-e, vagy egy adott számítógépen
18. Mivel az internetes használatra tervezték, a Paperpile kiválóan alkalmas az együttműködésre
19. A gyűjtött referenciák exportálhatók más referencia-kezelő alkalmazásokra, köztük az EndNote és a BibTeX (a LaTeX dokumentumok esetében)
20. A Google Dokumentumok funkciói használhatóak, és a végeredmény bármikor exportálható Word vagy PDF formátumban, és a Google Drive szinkronizálással bárhol elérheti őket
21. Létrehozható egy tökéletesen formázott bibliográfia
22. A hivatkozásokat egyszerre lehet beilleszteni

23. Támogatja a szövegekőzi és a lábjegyzetekbeli hivatkozásokat.
24. Több mint 7000 folyóirat-specifikus idézési stílus pl. APA, MLA, Chicago, s tb
25. Az online referenciák megkereshetőek és beilleszthetőek a dokumentumok / publikációk írása közben
26. Automatikusan keresi és kezeli a PubMed ID-eket és DOI-kat
27. Egyidejűleg több társszerző is használhatja
28. Hivatkozások Exportálhatóak RIS vagy BibTeX formátumban.
29. EndNote® kompatibilis: exportálhatja a dokumentumot Word és EndNote számára
30. LaTeX kompatibilis: exportálja a dokumentumot a LaTeX használatához
31. Tiszteletben tartja a magánéletét: A kiegészítőnek engedélye van a dokumentumok kezelésére, de mindenkor magánál marad. Soha nem fér hozzá, nem továbbítja vagy tárolja az Ön adatait, kivéve a citációk és bibliográfiák formázásához szükséges idézési metaadatokat.
32. FONTOS: Csak azok a felhasználók férhetnek hozzá, akik rendelkeznek a megfelelő jogosultsággal (felhasználói név és jelszó) – ennek a kezelését / kiosztását a Könyvtár menedzseli (előfizetés, felhasználói nevek kiosztása, stb.)
33. szerzői jogi problémák megoldására nem alkalmas > DE a dokumentumok harmadik fél számára nem kerülnek továbbadásra
34. EISZ adatbázis szolgáltatói közül egyre több engedélyezi a teljes szöveg saját G-Drive-ba történő mentését (pl. Proquest)

[A Paperpile egy 100%-os felhőalapú referencia-kezelő alkalmazás, amely az egyetem minden tagjának ingyenes a Google-fiókján keresztül.]

Jelenleg (2018. szept. 26-án) 237 dokumentum bibliográfiai adatait tartalmazza az a fiók, mely dokumentumok jelentős része érhető el teljes szöveggel is a felhasználói fiókhoz kapcsolódó G-Drive fiókban.

https://paperpile.com/app/folder/1b10af10-78ab-04ca-95d2-69ead4c79509

Alkalmazások YouTube In MP4 - Do Post to CiteULike Save to RefWorks Capture Reference Create GQueues Task Amazon Cloud Drive További könyvjelzők

Paperpile Klmkutato8@uni-miskolc.hu

+ Add Papers Search your papers Select

ALL PAPERS (232)

STARRED

FOLDERS

- International Journal of Legal Disc
- lista.ris (2)
- Narrativity (20)
- OTKA Könyvek (8)**
- proba (1)
- Routledge_Corpus Linguistic (13)
- Routledge_Discourse Analysis (1)
- Routledge_Forensic Linguistic (3)
- STUDIES IN LOGIC, GRAMMAR AN

LABELS

SHARED FOLDERS

- SzA (8)

TRASH (11)

OTKA Könyvek

The Language of the Law ★
Mellinkoff D
Resource Publications, 2004 – Book, filed in OTKA Könyvek
View PDF

Mellinkoff 2004 - The Language of the Law.pdf
33.66 MB | Rename | Remove | Save to disk | ✓ Is main PDF X
Attach supplementary PDF or other files...
Browse journal website for supplementary files

Book review: Help for innocent ears: Language crimes: The use and abuse of language evidence in the courtroom by Roger W. Shuy ★
Doyle JM
J Crim Law Criminol, 1993 – Journal Article, filed in OTKA Könyvek
View PDF

Being Pragmatic about Forensic Linguistics Symposium ★
Cheng EK
JL & Pol'y, 2012 – Journal Article, filed in OTKA Könyvek
View PDF

The Oxford Handbook of Language and Law ★
Tiersma PM, Solan LM
Oxford University Press, 2012 – Book, filed in OTKA Könyvek
View PDF

Common Law Legal English and Grammar: A contextual approach ★
Riley A, Sours P
Hart Publishing, 2014 – Book, filed in OTKA Könyvek
View PDF


Lawyer's language. How and why legal language is different ★
Phillips A
Routledge, 2003 – Book, filed in OTKA Könyvek
View PDF

8 Papers found
8 shown, 0 selected

Sort
Date Added
Publication Date
Title
Journal
Author

Filter
Has PDF (8)

Type
Book (5)
Journal Article (2)
Miscellaneous (1)



A Paperpile kezelőfelülete a böngészőben

Browser address bar: <https://app.metapdf.com/g/1ugh3QWOCKwFUHQ83VigMe6OwBwpU9Ns5>

Navigation: Back, Forward, Home, Refresh, Search, Print, Share, Download, etc.

Annotations: Open, More, Search annotations, Comment, Fontos!

Document Header:

DE GRUYTER OPEN

STUDIES IN LOGIC, GRAMMAR AND RHETORIC 41 (54) 2015

DOI: 10.1515/slogr-2015-0018

Author: Chris Baber, University of Birmingham, Great Britain

Title: **THINKING THROUGH TOOLS: WHAT CAN TOOL-USE TELL US ABOUT DISTRIBUTED COGNITION?**

Abstract: In this paper, I question the notion that tool-use must be driven by an internal representation which specifies the "motor program" enacted in the behaviour of the tool-user. Rather, it makes more sense to define tool-use in terms of characteristics of the dynamics of this behaviour. As the behaviour needs to be adjusted to suit changes in context, so there is unlikely to be a one-to-one, linear mapping between an action and its effect. Thus, tool-use can best be described using concepts from Nonlinear Dynamics. Such an approach can be used to create a sort of cybernetic model of tool-use. However, there is a danger that such a model can either lead us back to internal representations (in that the comparator used to evaluate feedback during behaviour could be assumed to be pre-defined) or could fail to capture cognitive aspects of behaviour. In particular, the question of how the craftworker's intent can be enacted in the use of tools to produce a specific object seems to be lost in the cybernetic account. My solution is two-fold. First, the "model" is created on-the-fly and adapted through moment-by-moment interactions in the system of tool-user-tool-material-environment. This means that, rather than assuming a pre-defined internal representation that drives behaviour, I propose that cognition involves the selection of salient parameters that characterize the behaviour and the continued monitoring and management of behaviour in terms of these parameters. Second, intent is only loosely defined a priori but crystallizes through the continued interactions between craftworker and object through a process in which the affordances of the object become apparent to, and respond to by, the craftworker.

Keywords: affordance, tool-use, internal representation, nonlinear dynamics

Section: 1. Craftwork and Tool-Use

Text: For the craft worker, a tool is the means by which intention is imposed on raw material; the tool disappears from immediate awareness in the smooth, skilful performance of the craft-worker's art. For me, this raises two fundamental questions relating to the use of tools and the relation between tool-use and cognition. The first concerns what we might mean by intention

Page 1 / 16

A pdf megjelenítés és jegyzetelés a böngészőben

https://drive.google.com/drive/u/2/folders/1A5RRXt73jt77vw7hYE21IMVTmd8pSVLM

Alkalmazások YouTube In MP4 - Do Post to CiteULike Save to RefWorks Capture Reference Create GQueues Task Amazon Cloud Drive További könyvjelzők

Drive

Keresés a Drive-on

Saját meghajtó > Paperpile > All Papers > A

Név ↑	Tulajdonos	Utolsó mód...	Méret
Aalto-Heinilä 2016 - Fairness in statutory interpretation ...	én	2018. márc. 18.	2 MB
Ágota 2005 - Hat terminológiai lecke.pdf	én	2018. ápr. 4.	6 MB
Alessandro Capone 2016 - Pragmatics and Law - Philos...	én	2018. máj. 19.	3 MB
Andone 2014 - Maneuvering with the Burden of Proof - ...	én	2018. márc. 18.	181 kB
Andruszkiewicz 2016 - On Some of the Aspects of the L...	én	2018. márc. 18.	216 kB
Arcimaviciene and Jonaitiene 2015 - Metaphor Evaluati...	én	2018. márc. 18.	386 kB

Új

- Saját meghajtó
- Csapatmeghajtók
- Velem megosztva
- Legutóbbi
- Csillagozva
- Kuka
- Tárhely
2,8 GB van használatban

A Paperpile pdf fájlok a Google Drive-ban

https://paperpile.com/shared/Wa9UHJ

Alkalmazások YouTube In MP4 - Dc Post to CiteULike Save to RefWorks Capture Reference Create GQueues Task Amazon Cloud Drive További könyvjelzők

Paperpile [Open Paperpile](#)

Kutato8 KIm has shared 12 papers with you

Justice, fairness and juridical perfectibility
 Petrilli S
International Journal of Legal Discourse, 2016 – Journal Article [Paperpile](#) [f](#)
[View PDF](#) [Website](#) [Abstract](#)

Justice and the right to non-functionality
 Ponzio A
International Journal of Legal Discourse, 2016 – Journal Article [Paperpile](#) [f](#)
[View PDF](#) [Website](#) [Abstract](#)

How not to do things with words: On conceptions of language and freedom of speech
 Schaumburg-Müller S
International Journal of Legal Discourse, 2016 – Journal Article [Paperpile](#) [f](#)
[View PDF](#) [Website](#) [Abstract](#)

The Changing Face of the English Court in the Early Twenty-First Century: A Neo Pluralistic Approach
 Roberts S
International Journal of Legal Discourse, 2016 – Journal Article [Paperpile](#) [f](#)
[View PDF](#) [Website](#) [Abstract](#)

A study of legal translation from the perspective of error analysis
 Hu PC, Cheng L
International Journal of Legal Discourse, 2016 – Journal Article [Paperpile](#) [f](#)
[View PDF](#) [Website](#) [Abstract](#)

Communication of an interpreter and fair trial under Nigerian criminal justice system
 Michael CO
International Journal of Legal Discourse, 2016 – Journal Article [Paperpile](#) [f](#)
[View PDF](#) [Website](#) [Abstract](#)

Bailiff enforcement in Poland – Standards of practice, Polish and Swedish legal terms and collocations in the area of debt enforcement
 Hadryan M
International Journal of Legal Discourse, 2016 – Journal Article [Paperpile](#) [f](#)

A bibliográfiai tételek megosztása a böngészőben a Paperpile-ba való bejelentkezés nélkül

The screenshot shows the Paperpile web application interface. At the top, there is a browser address bar with the URL <https://paperpile.com/app/folder/34b8fd4e-66da-0c96-a316-784f6f1d1dc4>. Below the browser, the Paperpile header includes a navigation bar with a search bar and various action buttons like 'Select', 'Trash', 'Edit', 'Cite', and 'Share'. A notification at the top indicates 'Copied 1 citation to the clipboard'. The main content area displays a list of papers under the folder 'International Journal of Legal Discourse'. One paper is selected: 'Justice, fairness and juridical perfectibility' by Petrilli S, published in 2016 in the *International Journal of Legal Discourse*. A dialog box titled 'Formatted Citations' is open in the foreground, showing the formatted citation: 'Petrilli, Susan. 2016. "Justice, Fairness and Juridical Perfectibility." *International Journal of Legal Discourse* 1 (1): 1-22.' The dialog box has 'Copy to Clipboard' and 'Close' buttons.

Idézéshez formázott bibliográfiai tétel exportja a Paperpile-ból

V. Függelék

1. Az írott részkorpuszok összetétele

Az elemzés számára 'txt' formában letöltött jogi szövegek forrását a jogszabályok és bírói ítéletek nyilvános, hivatalos adatbázisai szolgáltatták. A korpusz összeállításának megköltése az volt, hogy alkorpuszonként közel azonos terjedelmű, véletlen kiválasztással kijelölt elemek kerüljenek be. Ehhez program megírására volt szükség, amelyre külső, informatikus szakértőt kellett bevonnunk. Számára a „megrendelés” a következő elvárásokat tartalmazta:

1. A jogszabályok szűrése tekintetében:

- a) A korpusz csak az érdemi, vagyis szabályozási célzatú jogszabályok halmazára terjedjen ki; azaz ne tartalmazzon módosító, hatályon kívül helyező vagy nemzetközi szerződést kihirdető jogszabályt.
- b) Letöltésre a lekérdezés napján hatályos szöveg kerüljön, ne keresse vissza az eredeti szöveg-állapotot.
- c) Az ennek megfelelő jogszabály-tömegnek (kb. 4.000 db) az 50%-a (2000 db) kerüljön be a korpuszba, mégpedig géppel, véletlenszerűen válogatva (pl. minden második fájl).
- d) Az egyes fájlok metaadatként csak a típust rögzítsék, mely három értéket vehet fel: törvény (tv), kormányrendelet (kr), miniszteri rendelet (mr). A metaadatot a fájlnevezés tükrözze, pl. tv0001.txt, kr0002.txt stb.
- e) A letöltött jogszabály ne tartalmazza annak – gyakran több száz, akár ezer – oldalas mellékleteit.

Az első szűrés alapján kiderült, hogy a „minden második fájl” a számuk alapján nem volt értelmezhető a program számára. Ezért a megrendelést úgy módosítottuk, hogy a törvény és a kormányrendelet közül mindegyik kerüljön be a korpuszba, a miniszteri rendeletek közül pedig annyi, hogy összesen 2.000 fájl legyen. Ennek végeredménye:

Tv – 326 db

Kr – 563 db

Mr – 1111 db

Ez együtt adja az autoritativ jogi normaszövegek halmazát, s képezte alapját két alkorpusz elkülönítésének: a „kódexjog” (a nagy, klasszikus törvénykönyvek, amelyek a jogi nyelvet a legtisztábban hordozzák) és az egyéb, nem-jogi szaknyelvi elemekkel kisebb-nagyobb mértékben keveredő, inkább technikai szabályozó előírások.

2. A bírósági határozatok szűrése tekintetében:

A bírósági határozatok két fő fajtájának megjelenítése volt szükséges: egyrészt a bíróságokra nézve (részben kötelező) iránymutatást megfogalmazó ún. irányítási eszközöké; másrészt az egyedi ügyekben, jogvitákban hozott és csak a felekre kötelező egyedi határozatoké, ítéleteké.

a) Az irányítási eszközök három leglényegesebb típusa került be a korpuszba: a Kúria által közzétett jogegységi határozatok, a kollégiumi vélemények és az (elvi) bírósági határozatok. Ezek szerepeltetése:

A Jogegységi határozatok mindegyike, összesen 184 db (fájlnevezés: je001.txt)

A Kollégiumi vélemények mindegyike, összesen 359 db (fájlnevezés: kv001.txt)

A Bírósági Határozatok című periodikából géppel válogatott véletlenszerű minta, a teljes (20.975 darabos) halmaz 20%-a vagyis 4000 db határozat (fájlnevezés bh0001.txt)

b) Az egyedi (anonimizált) határozatok a Kúriánál alacsonyabb szintű bíróságok (járásbíróság, törvényszék ítélőtábla) konkrét ügyekben hozott ítéletei. Ezek kb. 90.000 tételre rúgó teljes halmazából géppel véletlenszerűen válogatott része, a teljes 10%-a (9.000 db) került a korpuszba.

Az egyedi érdemi határozatok három típusa és fájlnevezése:

Járás-/városi bíróság határozata: Jb – 3000 db – jb0001.txt

Törvényszék határozata: Ts – 3000 db – ts0001.txt

Ítélet-tábla határozata: It – 3000 db – it0001.txt

3. A „magyarázó szövegek” tekintetében:

Magyarázó szövegnek nevezzük a nem a jogalkotás vagy jogalkalmazás során keletkező, de azok szöveg-produktumait feldolgozó és ismertető szövegtípusokat. Ezeknek két alapvető – hasonló célú, de lényegesen eltérő jellegű – csoportját különítettük el: a szaknyelvi magyarázó szövegeket, vagyis a szakemberek által leendő szakemberek számára készített tankönyveket, egyetemi jegyzeteket; valamint a laikusok által laikusok számára köznyelven írt magyarázatokat az internetes jogi fórumok felületén. A kettő között helyezkedik el, de a laikus nyelvhasználathoz igazodik a szakemberek által laikusok számára nyújtott hivatalos tájékoztatások szövegtömege. Ezek forrása a korpuszban a következő:

a) A tankönyvek a részben országosan, részben a Miskolci Egyetem folyó jogi képzésben használatos bevezető-áttekintő jellegű, valamint a nagyobb jogterületeket (polgári jog, büntetőjog, közjog) lefedő tankönyvek/jegyzetek véletlenszerűen kiválasztott részeit.

b) A hivatalos tájékoztatás szövegei egy honlapról letöltött kb. 300 ún. tájékoztató célú ügyleírás anyagából álltak össze:

<https://ugyintezes.magyarorszag.hu/>

c) A laikus internetes fórumok forrása is egy honlap: a jogiforum.hu, ahol azonban egyes specifikus kérdések mentén szerveződő rész-fórumok szolgáltatták a mintát; így:

Munkajogi kérdések

<http://www.jogiforum.hu/forum/16/33457>

Munkavállalói rendes felmondás

<http://www.jogiforum.hu/forum/16/140>

Jogi tanács kérése

<http://www.jogiforum.hu/forum/20/10880>

Rágalmazás, személyiségi jogok megsértése

<http://www.jogiforum.hu/forum/20/15572>

Devizahitel megtámadása

<http://www.jogiforum.hu/forum/20/28906>

Elévülés erkölcsi bizonyítvány

<http://www.jogiforum.hu/forum/21/9835>

Büntetőjog

<http://www.jogiforum.hu/forum/21/33744>

Öröklés

<http://www.jogiforum.hu/forum/35/18667>

2. A beszélt részkorpusz összetétele

A beszélt jogi nyelvi korpusz kuriozitása és nehéz előállíthatósága azt kívánja meg, hogy elemzését más kutatók, kutatócsoportok számára is tegyék lehetővé. Ezért amellet, hogy e részkorpuszt a hangfelvételek és transzkripciók rendszerezett együtteseaként a fenntartott kutatási honlapon bemutatjuk, tartalmát ismertetjük és kisebb mintát – annak átiratával együtt – mintaként bemutatunk, azt – kutatási engedéllyel – a terület más kutatói számára is hozzáférhetővé kívánjuk tenni.

A Miskolc Jogi Korpusz beszélt nyelvi alkorpusza két részből tevődik össze. Egyrészt tartalmazza a 2000 és 2003 között folytatott FKFP kutatás során rögzített mintegy 18,5 órás hanganyagot, melynek nagyobb része, 14,5 óra rendőrségi kihallgatáson, kisebb része, 4 órányi felvétel pedig bírósági, azon belül is elsősorban büntető bírósági tárgyalásokon készült. Az alkorpusz másik része a jelenlegi OTKA kutatás keretében 2016. október és 2017. július között készült hangfelvételekből áll. Ezúttal rendőrségi kihallgatásokon nem volt lehetőségünk felvételeket készíteni, de az Országos Bírósági Hivatal, valamint a Miskolci Törvényszék elnökének engedélye alapján közel 40 órányi polgári és büntető tárgyalást rögzítettünk, továbbá az eljárások professzionális (bíró, ügyész, ügyvéd, ügyvédjelölt) és laikus résztvevőivel mélyinterjúkat is készítettünk. Az engedélyek birtokában végül öt (három büntető és két polgári szakos) bíró vállalta a kutatásban való részvételt, és összesen 30 (18 büntető és 12 polgári) ügy tárgyalásán készültek hangfelvételek. A kutatás lezárásának idejére szinte a teljes beszélt nyelvi alkorpusz átírt, szöveges formában is rendelkezésre áll (leszámítva néhány rosszabb minőségű felvételt és a mélyinterjúkat). Az átírásban emberi és „gépi” erőt is igénybe vettünk: egyrészt két joghallgató folyamatosan dolgozott a korábbi és a jelen kutatásban készült hangfelvételek átírásán, másrészt az utolsó kutatási évben megvásároltuk egy hangfelismerő és azt írott formába konvertáló szoftver licenzét. Valamennyi hangfelvétel és azok átiratai a kutatócsoport tagjai számára elérhető D-Drive felületen hozzáférhetőek. Az alábbiakban a 2016-2017 folyamán készült hangfelvételek adatainak összefoglalása látható:

Dátum	bíró	ügy	időtartam
2016.10.14.	Horváth Dóra	életveszélyt okozó testi sértés	93:06 min, 39:36 min
2016.10.17.	Tuzáné Papp Szilvia	tulajdonjog megállapítása	20:29 min
2016.10.19.	Tuzáné Papp Szilvia	többlethasználati díj megfizetése	02:21 min, 05:55 min
2016.10.19.	Tuzáné Papp Szilvia	gyermektartásdíj megállapítása	44:05 min, 13:28 min
2016.10.24.	Tuzáné Papp Szilvia	gyermektartásdíj felemelése	30:26 min, 37:04 min
2016.11.02.	Tuzáné Papp Szilvia	jognyilatkozat pótlása	50:15 min, 08:01 min
2016.11.02.	Tuzáné Papp Szilvia	gondnokság alá helyezés	27:25 min
2016.11.02.	Tuzáné Papp Szilvia	bontóper (békítő)	17:28 min
2016.11.09.	Horváth Dóra	vesztegetés	96: 37 min
2017.01.11.	Tuzáné Papp Szilvia	tulajdonjog használatának rendezése	39:29 min
2017.01.18.	Horváth Dóra	emberölés	142:31 min, 61:40 min, 92:34 min, 161:59 min, 96:31 min,34:36 min,93:48 min
2017.01.25.	Horváth Dóra	emberölés kísérlete (közmunkások)	41 min
2017.02.01.	Horváth Dóra	csecsemőgyilkosság	41 min
2017.03.07.	Kaáli Éva	Közérdekű adat kiadása	60:17 min, 16:02 min
2017.03.27.	Horváth Dóra	halált okozó testi sértés, rablás	109 min, 82 min
2017.05.18.	Kaáli Éva	vállalkozói díj megfizetése	101:46 min
2017.05.18.	Kaáli Éva	szülői felügyelet gyakorlásának rendezése	69 min
2017.05.18.	Kaáli Éva	lakáshasználati díj megfizetése	69:03 min
2017.05.18.	Kertész Gábor	motorszám hamisítás	10:00 min, 6:14 min
2017.05.18.	Kertész Gábor	lopás	10:40 min
2017.05.25.	Kertész Gábor	garázdaság	-
2017.06.13.	Kertész Gábor	közúti baleset gondatlan okozása	126:27 min
2017.06.14.	Horváth Dóra	vesztegetés (tárgyalás és ítélet)	97:42 min, 11:41 min
2017.06.19.	Horváth Dóra	kábítószer fogyasztása és terjesztése	35:49 min
2017.06.20.	Kertész Gábor	lopás és családon belüli erőszak	61:12 min
2017.06.20.	Kertész Gábor	garázdaság	45:27 min
2017.06.22.	Kertész Gábor	tartási kötelezettség elmulasztása	13:48 min, 4:11
2017.06.22.	Kertész Gábor	lopás kísérlete (tárgyalás és ítélet)	22:48 min, 11:55 min
2017.06.22.	Kertész Gábor	ittas járművezetés vétsége (tárgyalás és ítélet)	14:54 min, 5:15 min
2017.07.04.	Kertész Gábor	állatkínzás vétsége	12:08 min
2017.07.11.	Horváth Mária	életveszélyt okozó testi sértés	16:55 min, 89:14 min, 14:42 min
			2266:34 min, 37 óra 46 min

3. A Magyar Jogi Nyelv megjelent lapszámainak tartalomjegyzéke

Eredeti vállalásunknak megfelelően útjára indítottuk a Magyar Jogi Nyelv című folyóiratot, melynek elektronikus és nyomtatott kiadását a HVG-ORAC Kiadó vállalta. A létrejött szerkesztőbizottság szándéka szerint nem alaptudományi, hanem olyan alkalmazott tudományi igényt támaztó periodika kíván lenni, amely szakmai-közéleti fórumként szélesebb – nemcsak jogász, hanem nyelvész – közönséghez szeretne eljutni. A megjelenő írások, tanulmányok színvonalát anonim lektorok bevonásával biztosítjuk.

Eredetileg egy lapszám megjelenését terveztük finanszírozni a pályázat költségvetéséből, de a rendelkezésre álló források még egy lapszám megjelentetését is lehetővé tették, és erre engedélyt is kaptunk az NKFI Hivataltól. Az alábbiakban az első három lapszám tartalomjegyzéke olvasható, a tanulmányok címére kattintva pedig elérhető a szöveg is.

2017/1.

Beköszöntő

Bősze Péter: A Magyar Jogi Nyelv című új szaklap köszöntése

Szendi Attila: A 19. századi magánjogi szakszavaink előfordulása korabeli folyóiratokban (1830–1850)

Tóth Judit: Mi fán terem a (kényszer)betelepítés?

Pátkai Nándor: A KÖZÉRTHETŐ JOGALKOTÁSRÓL – Az általános közigazgatási rendtartás nyelvezeté

Láncos Petra Lea: A jogi szaknyelvi lektorálás funkciói és kihívásai az ÁKR. példáján

Tóth Judit – Kurtán Zsuzsa: A jogszabályok szövege – avagy a jogalkotók szövegelése

Farkasné Puklus Márta: EULITA Konferencia Bécsben: a jogi tolmácsolás és fordítás sokrétűsége

2017/2.

Seresi Márta – Láncos Petra Lea: Tolmácsolás az Európai Unió Bíróságán – a gyakorló tolmácsok szemével

Horváth Ildikó: A tolmácsoláshoz való jog és a bírósági tolmácsolás

Tóth Judit: Az akadálymentes jogalkalmazói kommunikációról

Zódi Zsolt: Jogi szövegtípusok

Veress Emőd: Jogászképzés Sziszüphosz egyetemén

Láncos Petra Lea – Balogh Dorka: Jog és nyelv konferencia Białystokban

Balogh Dorka: Recenzió Tamás Dóra Mária Bevezetés a jogi terminológiába a terminológus szemüvegén át című kötetéről

2018/1.

Nyelvtörténet

Gedeon Magdolna: A magyarországi bányajogi szaknyelv történeti áttekintése

Tanulmányok

Majoros Tünde: Intralingvális fordítások a bírósági nyelvhasználatban

Minya Károly – Vinnai Edina: Hogyan írjunk érthetően?

Horváth Péter Iván: A fogalmazás apró műhelytitkai a hiteles fordításban

Szoták Szilvia: A Termini Magyar Nyelvi Kutatóhálózat terminológiai tevékenysége

Moldova Zsófia: „Terhelt tájékoztatás”

Fórum

Szitás Benedek: Jogalkotásunk közérthetősége

Kurtán Zsuzsa: Hogyan élte a nyelv a jog világát?

Tóth Judit: Hogyan készül a bírósági tolmács?

Miklódy Dóra: XX. Fordítástudományi Konferencia – TransELTE2018

Jogi terminológia konferencia

4. A kutatás ideje alatt beszerzett magyar és idegen nyelvű könyvek listája

Magyar nyelvű:

Horváth Dóra, Mitev Ariel: *Alternatív kvalitatív kutatási kézikönyv*. Alinea Kiadó, 2015. (ISBN 9786155303821)

Idegen nyelvű:

1. Ruth Wodak, Michael Meyer: *Methods of Critical Discourse Studies*. SAGE, 2016. (ISBN 978-1-4462-8240-3 ISBN 978-1-4462-8241-0)
2. Sarah Pink, Heather Horst, John Postill, Larissa Hjorth, Tania Lewis, Jo Tacchi: *Digital Ethnography. Principles and Practice*. SAGE, 2016. (ISBN 978-1-4739-0237-4 ISBN 978-1-4739-0238-1)
3. Alison Riley, Patricia Sours: *Common Law Legal English and Grammar. A Contextual Approach*. Hart Publishing, 2014. (ISBN 978-1-84946-576-2)
4. Peter M. Tiersma, Lawrence M. Solan (eds): *The Oxford Handbook of Language and Law*. Oxford University Press, 2012. (ISBN 978-0-19-957212-0)
5. Roger Shuy: *Linguistics in the Courtroom. A Practical Guide*. Oxford University Press, USA, 2006. (ISBN-10: 0195306643 ISBN-13: 9780195306644)
6. Roger Shuy: *Fighting over Words. Language and Civil Law Cases*. (ISBN-13: 978-0195328837 ISBN-10: 0195328833)
7. Roseann Duenas Gonzalez, Victoria F. Vasquez, Holly Mikkelson: *Fundamentals of Court Interpretation: Theory, Policy and Practice. Summer Institute for Court Interpretation*. University of Arizona (ISBN-13: 978-0890894149 ISBN-10: 0890894140)
8. Michal Bobek (ed): *Central European Judges Under the European Influence. The Transformative Power of the EU Revisited*. 2015 (ISBN 9781849467742)
9. Howard Gregory: *Language and Logics. An Introduction to the Logical Foundations of Language*. (ISBN 9780748691630)
10. Györfi Tamás: *Against the New Constitutionalism*. (ISBN: 978 1 78347 300 7)
11. John Forrester: *Thinking in Cases*. (ISBN 978-1-5095-0862-4)
12. Badó Attila (szerk.): *Fair trial and Judicial Independence. Hungarian Perspective*. (ISBN 978-3-319-01215-5)
13. Hertogh, Marc: *Nobody's Law - Legal Consciousness and Legal Alienation in Everyday Life*. Palgrave Macmillan UK, 2018. (ISBN 1137603968)
14. Roger Cotterrell: *Law's Community. Legal Theory in Sociological Perspective*. Clarendon Press, Oxford Socio-Legal Studies, 1997. (ISBN: 9780198264903)
15. Roger Cotterrell: *The Politics of Jurisprudence. A Critical Introduction to Legal Philosophy*. 2018. (ISBN: 9780406930552)
16. Bongiovanni, G., Postema, G., Rotolo, A., Sartor, G., Valentini, C., Walton, D. (Eds.): *Handbook of Legal Reasoning and Argumentation*. Springer, 2018. (ISBN 978-90-481-9451-3)
17. Cominelli, Luigi: *Cognition of the Law. Toward a Cognitive Sociology of Law and Behavior*. Springer (ISBN 978-3-319-89347-1)
18. Rebecca Clift: *Conversation Analysis. Textbook*. Part of Cambridge Textbooks in Linguistics. 2016 (ISBN: 9780521157193)
19. Douglas A. Kibbee: *Language and the Law. Linguistic Inequality in America*. 2016. (ISBN 9781107623118)

20. Anna M. Babel (ed.): *Awareness and Control in Sociolinguistic Research*. 2016. (ISBN 9781107072381)
21. Maria Aloni, Paul Dekker (eds.): *The Cambridge Handbook of Formal Semantics*. Part of Cambridge Handbooks in Language and Linguistics. 2016. (ISBN 9781107028395)
22. Nikolas Coupland (ed.): *Sociolinguistics. Theoretical Debates*. 2016. (ISBN 9781107635753)
23. Martin Hogg: *Obligations. Law and Language*. 2018. (ISBN 9781107458215)
24. H. Patrick Glenn, Lionel D. Smith (eds.): *Law and the New Logics*. 2018. (ISBN 9781107514539)
25. Michael Hanne, Robert Weisberg (eds.): *Narrative and Metaphor in the Law*. 2018. (ISBN 9781108422796)
26. Harold J. Berman: *Law and Language. Effective Symbols of Community*. 2016. (ISBN 9781316619339)
27. Lloyd L. Weinreb: *Legal Reason. The Use of Analogy in Legal Argument*. (2nd Edition) 2016. (ISBN 9781316607329)
28. Marcus Galdia: *Legal Linguistics*. Frankfurt a. M.: Peter Lang, 2009.
29. Marcus Galdia: *Legal Discourses*. Frankfurt a. M.: Peter Lang, 2014.
30. Theo Öhlinger (ed.) *Recht und Sprache*. Fritz Schönherr – Gedächtnissymposium. MANZ Verlag Wien, 1986. (ISBN 978-3-214-05230-0)
31. William Crawford, Csomay Enikő: *Doing Corpus Linguistics*. Routledge, Taylor & Francis, 2015. (ISBN13: 9781138024618, ISBN10: 1138024619)
32. Stukker, Spooren, Steen: *Genre in Language, Discourse and Cognition*. Walter de Gruyter, 2016. (ISBN 9783110467703)
33. Mellinkoff, David: *The Language of the Law*. Little, Brown and Company, 1963. (ISBN 9781592446902)
34. Daubach, Sprick: *Der Zivilprozess: Eine Einführung für Gerichtsdolmetscher und – übersetzer*. BDÜ Weiterbildungs- und Fachverlagsgesellschaft, 2010. (ISBN 9783938430316)
35. Kock, Rainer: *Der Strafprozess: Eine Einführung für Gerichtsdolmetscher und – übersetzer*. BDÜ Weiterbildungs- und Fachverlagsgesellschaft, 2013. (ISBN 9783938430521)
36. Thormann, Hausbrandt: *Rechtssprache*. BDÜ Weiterbildungs- und Fachverlagsgesellschaft, 2016. (ISBN 9783938430781)
37. Daum, Ulrich: *Gerichts- und Behördenterminologie*. BDÜ Weiterbildungs- und Fachverlagsgesellschaft, 2017. (ISBN 9783938430910)
38. Daum, Hansmeyer: *Arbeitsbuch zur Gerichts- und Behördenterminologie*. BDÜ Weiterbildungs- und Fachverlagsgesellschaft, 2013. (ISBN 9783938430392)
39. Schlüter-Ellner, Corinna: *Juristendeutsch verständlich gemacht und Treffende Verben in der deutschen Rechtssprache*. BDÜ Weiterbildungs- und Fachverlagsgesellschaft, 2012. (ISBN 9783938430453)

5. Összefoglaló a T-LAB programról

SZENDI ATTILA*

A T-LAB KVANTITATÍV SZÖVEGELEMZŐ SZOFTVER - A JOGI NYELV ÉRTHETŐSÉGÉNEK DIMENZIÓI ÚJ ELEMZÉSMÓDSZERTANI MEGKÖZELÍTÉSBEN

Bevezető

A szakszókincs elemzéséhez számítógépes programot mint elemző eszközt már a magyar nyelvű kutatásokban is többször sikerrel alkalmaztak. Ugyanezt a módszert a szemantikai kapcsolatok elemzésére még csak az utóbbi évtizedben és csak kísérleti jelleggel vezették be. Ettől eltekintve a "jog és nyelv" témakörében végzett OTKA kutatás¹ kezdetén is megfogalmazódott az informatikai elemző eszközök alkalmazásának szükségessége/lehetősége. Ez ösztönözte a harmadik szempontunkat, vagyis a számítógépes kvantitatív szövegelemzés bevezetését, melynek a magyar nyelvű jogtudományi és társadalomtudományi kutatásokban való alkalmazása 2014-ben - a szakirodalom tanúsága szerint is - úttörő jellegű kezdeményezés volt.²

Tanulmányomban szükségesnek tartom röviden összefoglalni a későbbiekben részletesen kifejtésre kerülő elemzőmódszertan elvi egységeit. Mindenekelőtt meg kell határozni az elérni kívánt új célt, a tárgyat, a rendelkezésre álló vagy alkalmazni kívánt eszközöket, eljárásokat. Az eredményekből, ha negatívak is következtetéseket kell levonni, melyeknek hatására módosításokat, javaslatokat lehet megfogalmazni nem megriadva az új szempontok megfogalmazásától sem.

Az új módszer célja, hogy a jogi szövegek szemantikai hálóját jobban megértsük, és a történeti dimenzió segítségével a változások eddig nem látható szegmenseit felfedjük. A módszer tárgyát jelen esetben a jogi szövegek képezik, melyek korpuszként a Miskolci Jogi Korpusz név alatt kerültek összeállításra.³ A kutatás módszertanára is hatással van, de elsősorban az eszközeire, hogy több tudományterület szempontrendszerait, fogalmait is alkalmaztuk, így a jogtudományi, nyelvtudományi, szociológiai, pszichológiai, történettudományi, de az informatikai és a matematikai-statisztikai ismeretek alkalmazásával tesszük még sokoldalúbban egymáshoz illeszthetővé az elemzések tárgyát, eszközeit és eljárásait.

A konkrét eszközök között kiemelten foglalkozunk az informatikai eszközökkel. Ez az információ a tárgykörben publikált tanulmányok esetében minimális dokumentálást kapott, pedig a módszer szempontjából megismerésük, megtanulhatóságuk a nem informatikus kutatók számára alapvető kérdés. A probléma jelentősége két szakirodalmi említéssel is alátámasztható. Az egyik egy külföldi példa, a Massachusetts Institute of Technology 2016-ban publikált, a könyvtárak jövőjéről szóló egyik dokumentuma, mely az egyetemi könyvtárak szolgáltatási körébe sorolja be a „felhasználóbarát, használatra kész, grafikus, ám univerzális adat- és szövegbányászati eszközt”.⁴ A másik egy magyar példa, a Sebők Miklós által szerkesztett, a kvantitatív szövegelemzésről szóló kötet kitekintő tanulmányából való megállapítás: „Márpedig a megfelelő szoftver, illetve applikáció kiválasztása egy olyan döntés, mely akár a kutatási eredményekre is kihatással lehet.”⁵

* Főigazgató, Miskolci Egyetem, Könyvtár, Levéltár és Múzeum

¹ OTKA K-112172. sz. „A tisztességes eljárás nyelvi aspektusai: A jogi nyelv hatása a joghoz való hozzáférés érvényesítésére”

² Jelen kötet két tanulmánya is foglalkozik kvantitatív szövegelemzéssel (Vincze Veronika és Zódi Zsolt).

³ Részletesen lásd. Vincze Veronika tanulmányát

⁴ Massachusetts Institute of Technology tanulmánya: Institute-wide Task Force on the Future of Libraries Preliminary Report. 2016. 6. old. (<https://future-of-libraries.mit.edu/sites/default/files/FutureLibraries-PrelimReport-Final.pdf>)

⁵ Sebők 2016 - Kvantitatív szövegelemzés és szövegbányászat a politikatudományban, 148

A tanulmány címében jelzett tartalom kifejtésekor az elmúlt négy év kutatási tapasztalatainak összefoglalását tűztem ki célul, de a munka során egyre inkább dominánssá vált az elemzés módszertanának a megismertetése. Ennek következtében a megjelölt témakörök (új jogi, nyelvészeti, történeti elemzési módszerek) bemutatása csak váza egy teljességre törekvő ismertetésnek, de a témában eddig napvilágot látott tanulmányok, kutatások alapján szükségesnek látjuk, hogy egy multidiszciplináris kutatás rendszerét is leírjuk, mert az új lehetőségek és az alkalmazott módszerek ennek fényében lesznek értékelhetők, követhetők vagy módosíthatók.

A 2015-ben megjelent “A jog nyelvi dimenziója” című kötet két tanulmányában is megfogalmazódtak olyan szempontok, amelyek a témánkhoz kapcsolhatók. Szabó Miklós “A jogi nyelv erőtere” című tanulmányában⁶ a *jogtörténeti hatóerőt* a jogi szaknyelv magyarrá válásának vizsgálatában látja. Ez inspirált bennünket a történeti dimenzió felvállalására, jóllehet ennek egy másik olvasata valósult meg a jelen tanulmányban: az összehasonlító jogtörténeti nézőpont.⁷ Kurtán Zsuzsa, a nyelvész oldaláról, - nyelvészeti szempontból - ugyanebben a kötetben - “A magyar jogi szaknyelv leírásának kutatási programjában”⁸ a megértést befolyásoló tényezők között említi a *szemantikai kapcsolatokat*, valamint a *statisztikai, kvantitatív módszerek* használatát.

A Miskolci Jogi Korpusz kvantitatív szövegelemző szoftverének technikai lehetőségei

A bevezetőben említettem, hogy módszertani kísérletünkben kiemelt helyen szerepelt a kvantitatív szövegelemzéshez használt szoftver kiválasztása, tesztelése. Ennek a helyességére a már korábban említett szakirodalom mellett az újabbak is erre hívták fel a figyelmünket.⁹

Három szempont vezetett bennünket a szoftver kiválasztásánál: 1. Az elméleti munkákban jelzett funkcionalitás kiemelt szempont volt. 2. Ezzel egyenértékű volt, hogy az *alkalmazás grafikus felülete* tegye lehetővé a nem informatikus-matematikus kutatóknak a szoftver használatát. 3. A kvantitatív szövegelemzéshez nélkülözhetetlen, hogy a korpusz előfeldolgozási fázisában az elemző szoftver *magyar nyelvű szótövesítést* tudjon alkalmazni. Nem szakmai szempontként kellett figyelembe vennünk az intézményi, kutatói pénzügyi beszerezhetőség korlátait is, mely miatt egy ideig az ingyenes, nyílt forráskódú alkalmazásokat helyeztük figyelmünk középpontjába.¹⁰

Az OTKA kutatás kezdetétől az együttműködés informatikai támogatására használtunk felhő alapú tárolóhelyeket. Ezért a felhőalapú (cloud) alkalmazás igénye is felvetődött, de 2014-ben még a kevésbé megvalósítható tervek kategóriájába soroltuk. Az elmúlt négy év informatikai fejlődése nemcsak a kereskedelmi programok kategóriájában, de a szabadon hozzáférhető felhőszolgáltatások formájában nagy előrelépés történt és ez is megvalósult.¹¹ Jelenleg ezek a szabadon hozzáférhető felhőalkalmazások azonban a szövegelemzés funkcionalitásának csak kis részét kínálják.

Az ingyenes felhőalapú szövegelemző alkalmazások néhány éven belül a tervezett fejlesztések után, jól használható, sok funkciós eszközökké válhatnak, azonban addig is lehetőséget szerettünk

⁶ Szabó M. 2015, 182.

⁷ A magyarrá váló szaknyelv folyamatát Gedeon M tanulmánya ebben a kötetben részletesen kifejti.

⁸ Kurtán Zs 2015, 200

⁹ Sántha Kálmán. 2013. Multikódolt adatok kvalitatív elemzése. Budapest: Eötvös J. Kvk., Sántha Kálmán. 2016. “Metaforaháló és kognitív térkép a tanári nézetek vizsgálatában.” *Magiszter*, no. 2: 4–11.

¹⁰ A kipróbált szoftverek: 1. Kvantitatív és kvalitatív szövegelemző: KHCoder (<http://khcoder.net/en/>).

2. Szemantikus hálózatelemző: WORDij (<http://wordij.net/>). 3. Konkordanciák és kollokáció készítő: AntConc (<http://www.laurenceanthony.net/software/antconc/>), mlct [MultiLingual Coprus Tool] (<https://sites.google.com/site/scottpiaosite/software/mlct>). LancsBox [LAncester University corpus toolbox] (<http://corpora.lancs.ac.uk/lancsbox/index.php>)

¹¹ TANIT [Text ANalysis Tools] (<http://dighum.bibl.u-szeged.hu/tanit/>), ITMS [Interactive Text Mining Suite] (<https://languagevariationsuite.shinyapps.io/TextMining/>), természetesen a felhőalkalmazások mögött parancsori szövegelemző, szótövesítő programok működnek, mint a Mallet szövegelemző és a magyarlanc professzionális szótövesítő, az R statisztikai program és a SnowBall szótövesítő.

volna biztosítani a kvantitatív szövegelemzés több funkciójának is a használatára. Ezért döntöttünk a T-LAB plus 2018¹² kereskedelmi szoftver megvásárlása mellett. Pozitív döntésünket lehetővé tette az, hogy 2017-től a T-LAB szoftver a szótövesítésében támogatja magyar nyelvet is. Ezzel az egyik legfontosabb akadály hárult el egy professzionális kvantitatív szövegelemző program alkalmazása elől a jog és nyelv kutatásának szemantikai, jogelméleti, jogszociológiai és jogtörténeti szövegei esetében.¹³ A T-LAB szoftver működésével kapcsolatban eddig szerzett tapasztalatainkat röviden összefoglaljuk.

A T-LAB szoftverben 15 féle elemzése szövegelemzési funkciót találunk melyek 3 fő csoportba vannak osztva. Ezen kívül a korpusz előkészítésnek fázisaiban is statisztikai adatokat kapunk szövegeinkről. Mielőtt a három fő elemzési csoport funkcióinak ismertetésére térnék szükséges beszélni a program logikájáról. A programnak vannak elemzési egységei, funkciói, melyekben statisztikai algoritmusok dolgoznak. Az eredményeket táblázatok és grafikonok formájában kapjuk meg, melyek értelme nem látványelem, hanem a táblázatokban található adatok értelmezésének egyetlen módja az eredményeket felhasználni kívánó ember számára. A program további feldolgozásra kimenet elmenthetővé teszi táblázatos formában és további adatvizualizációs programokban való feldolgozásra.

A program a szöveget *kontextuális egységekre* és *lexikai egységekre* bontja. A kontextuális egység döntésünk szerint lehet a bekezdés vagy a mondat vagy szövegcsont. A lexikai egységek a kulcsszavak, melyek a szótövezett szavakból kapunk meg három módszer alapján (TF-IDF¹⁴, Khí négyzet próba, vagy előfordulási gyakoriság) illetve az n-gramok, melyeket a szoftver generál. Ezt kulcsszólistát is maximalizálhatjuk abszolút számban és közvetlenül szerkeszthetjük is, de készíthetünk többszavas kifejezéseket is belőle és tovább szűrhetjük tiltószavak listájával is. A kulcsszavak gondos előkészítése a későbbi elemzések eredményeit nagy mértékben javíthatják, mert ezek a kulcsszavak képezik az alapját a későbbi szövegelemző műveleteknek.

A szövegfájlokat különböző egységekre bonthatjuk, melyek változóként funkcionálnak és kategóriákat is rendelhetünk a ezekhez a szövegrész jelölésekhez. (Ezt majd a jól strukturált jogszabályok esetében könnyen ki lehet használni.)

A szoftver alapvetően asztali számítógépeken való működésre készült, így az elemezhető korpusz méret 90 MB-ban korlátozott. Ez azonban a formázatlan szövegre vonatkoztatva kb. 90 Ptk. méretű szöveggel egyenlő.

A szöveg előkészítési folyamatában több ponton is be lehet avatkozni. Azonban a szótövesítő eljárásnál fajtájánál nem (stemming, toldaléklevágás). A kvantitatív szövegelemző program hiteles működéséhez nélkülözhetetlen a szótövező képesség. Ez magyar nyelven is működik, de egy olyan változata, mely toldaléklevágást végez. Ez egy progarmozástechnikailag gyors algoritmus, minden szóval megbírkózik, de olykor az esetek 10-30%-ban hibás szótövet kapunk, mely egyértelműsítésre szorul vagy egyáltalán nem tudunk vele szemantikailag mit kezdeni. Ennek a problémának a kiküszöbölésére azt az eljárást dolgoztuk ki, hogy a T-LAB szoftver lehetővé teszi a magyar nyelvű szótövesített szólistának a szerkesztését, hibajavítását. egészen addig, hogy saját szótövesítő párokból álló fájlt tölthetünk be. Ezt ki is használtuk, mivel a szótövesítés eredményével az egyes elemzéseknél nem voltunk megelégedve. A betöltendő listafájl úgy kaptuk meg, hogy a parancssori magyarlanc szoftver segítségével az elemezni kívánt szöveget lemmatizáltuk.¹⁵

¹² <https://tlab.it/en/presentation.php>. Sajnos egyéb adminisztrációs okokból csak néhány hónap állt rendelkezésre a megvásárolt szoftver megismerésére a gyakorlatban.

¹³ 2017 óta több kereskedelmi szövegelemző szoftver is rendelkezik magyar szótövező modullal. (pl. Provalis Research WordStat, MaxQDA)

¹⁴ Term frequency–inverse document frequency

¹⁵ A magyarlanc kimenetként kapott lemmatizált szövegfájlon még több normalizálást kellett elvégezni. Ezt a eljárást a későbbiekben automatizálni fogjuk.

A program funkcionális csoportosításának köszönhetően könnyebben jellemezni tudjuk a 15 funkciót. Az egyes elemzések eredményeit más elemzésekhez kapcsolódhatnak (be vagy kimenete lehet) ezzel az adatok láncolatát hozzuk létre az egyébként elkülönülő elemzési funkciók számára.

A *tematikus elemzéscsoport* funkciói tartalmazzák a klasszikus kvantitatív/kvalitatív szövegelemző eljárásokat: *csoportosítás, osztályozás, szótár alapú osztályozás, témamodellelés* (topic modelling).

Az *összehasonlító elemzéscsoport* alapját a *kontingenciatáblázat* adatai adják, melyek a kulcsszavak előfordulási gyakoriságaiból, vagy a kulcsszavak együttes előfordulásai gyakoriságából képezik a korpusz egészére vagy részszövegeire vonatkozóan. A *korrespondencia analízis* a kontingencia táblázat nagyszámú értékeinek elemzéséhez nyújt segítséget grafikonok formájában. A korrespondencia elemzés adatait az itt található klaszteranalízissel további elemzések alá vethetjük. A *specifikusság elemzéssel* a legjellemzőbb kulcsszavakat és leginkább alulreprezentált szavakat vethetjük össze, melyek a korpusz szövegfájljainak részegységeiben (subsets) találhatóak.

Az elemzési csoportok közül a legérdekesebb, melyek majd a mi jogszabályi elemzéseink gerincét is adják az *együttes előfordulás elemzéscsoport* tagjai. Az együttes előfordulás (co-occurrence) a kulcsszavakra vonatkozik és mátrix alapon rendezve elemez. A szavak együttes előfordulásának jelentősége abból az elméletből fakad miszerint azon szavak, melyek egy adott korpusz kontextuális egységeiben szignifikánsan gyakran fordulnak elő a szavak kölcsönös szemantikai terében vannak.¹⁶ Az így előállított szemantikai tér kulcsszavainak *asszociációs tere* megjeleníthető valamint kirajzolható belőle egy *konceptiós térkép* (a kulcsszavak klaszterezésével). Végül a *szekvenciális elemzéssel* megjeleníthetjük a dokumentum kulcsszavainak *hálózatát, szemantikai terét* a kulcsszavak egymáshoz viszonyított relatív pozíciójával a segítségével.

Ez utóbbi elemzéscsoport alapját a grammatikai/lexikai elemzések *kollokációs* eljárásai mintájára dolgozták ki, de összetettebb algoritmusok, eljárások keretében. Nem véletlen, hogy az elemzéscsoport részét képezi egy *konkordancia* funkció is.

A szoftver lehetőségeit nagyon leegyszerűsített csoportosításban írtuk le. A program 15 fő elemzési funkcióval rendelkezik, mely a opciók, az elemzési egységek, a változók és kategóriáik valamint az egymásba ágyazott funkciók révén egy összetett elemzésre képes eszközt kapunk, amelynek azonban minden lehetőségét kihasználva inkább a bonyolultsága kerül előtérbe és csak hosszabb gyakorlati tapasztalati idő után tűnik vonzó (rugalmas és átlátható) lehetőségnek.

Szakirodalmi kitekintés

A bevezetőben is említett szempontok miatt célszerűnek látok egy rövid időrendi áttekintést adni a szövegelemzésről (text mining) megjelent hazai és külföldi publikációikat illetően, kiemelten a társadalomtudományi alkalmazási lehetőségeiről. Ezzel az áttekintéssel szemléltetni szeretném a téma multidiszciplinaritását, és egyben az egyes tudományterületek változó mértékű előrehaladását a szövegelemzés módszerének alkalmazásában.

A tartalomelemzés tárgykörében már korábban is jelentek meg publikációk, mi a 80-as évektől vesszük fel a fonalat, mivel ekkortól számíthatjuk a magyar nyelvű szakirodalom megjelenését is. A kiemelkedő tartalomelemzés módszertanáról szóló művek sorát Klaus *Krippendorff* 1980-ban megjelent könyve¹⁷ nyitja meg, melyet 1995-ben magyarul is kiadtak “*A tartalomelemzés módszertanának alapjai*” címmel¹⁸. Jelentősége, hogy fejezetet szentelt a számítógéppel végezhető szövegelemzésnek és szoftvereinek. A számítógéppel végzett szöveggel végezhető eljárásokról Gerard

¹⁶ Lancia coo 2005

¹⁷ Klaus Krippendorff 1980, *Content Analysis; An Introduction to its Methodology*, Sage

¹⁸ Krippendorff et al. 1995 - *A tartalomelemzés módszertanának alapjai*

Salton 1988-ban jelentetett meg könyvet.¹⁹ 1989-ben Prószéky Gábor a magyar nyelvű és főleg a magyar nyelvről is számítógépes nyelvész szemével írt monográfiát “*Számítógépes nyelvészet*” címmel.²⁰ Prószéky gyakorlatilag minden fő elemzési irányt leírt. Már ekkor terjedelmes fejezetet szánt a szemantikus hálók kérdésének.²¹

1995-ben jelent meg magyarul *Earl Babbie* - azóta több magyar nyelvű kiadást is megért - könyve, *A társadalomtudományi kutatás gyakorlata*.²² A könyv jelentősége nemcsak az adatelemzési eljárások és informatikai szoftverek bemutatásában rejlik, hanem a matematikai-statisztikai algoritmusok, és azok értelmezésének a leírásában is. 1996-ban a történettudomány kutatóitól, a *Mawdsley-Munck* szerzőpárostól is megjelent egy könyv a *Számítógép a történettudományban* címmel.²³ E mű bár nagy részében az adatbáziskezeléssel foglalkozik, fejezetet szentel a szövegelemzés akkor elérhető szoftvereinek is. 10 évvel Prószéky monográfiájának megjelenése után, 1999-ben látott napvilágot a számítógépes nyelvészet nagyközönség számára is értelmezhető, aktualizált része.²⁴

A társadalomtudományi kutatások Babbie által megírt statisztikai elméletéhez jól kapcsolódott 2002-ben Székelyi Mária és Barna Ildikó SPSS-hez többváltozós elemzési technikákról írt könyve társadalomkutatók számára,²⁵ mely elsősorban a strukturált számszerűsített adatok elemzését mutatja be, de módszeres magyar nyelvű leírását is adja mindazoknak a statisztikai algoritmusoknak, melyek a kvantitatív szövegelemzés alapjait is képezik.

A 2000-es évek legkiemelkedőbb tudományos tevékenysége a számítógépes nyelvészet területén az volt, hogy 2003-tól évente megrendezésre kerültek a Magyar Számítógépes Nyelvészeti Konferenciák, melyeknek később megjelentek a konferenciakötetei. Azóta is ezekben kerülnek nyilvánosságra a legújabb módszertani és kutatási eredmények. Szirmai Mónika a korpusznyelvészetről írt összefoglalót²⁶, amely az idegennyelv-oktatásban is jól használható konkordancia programokat ismerteti. 2006-2007-ben több adatelemzéssel foglalkozó kötet is kiadásra került²⁷. Ezek közül kiemelendő az informatikusok által szerkesztett *Adatbányászat a hatékonyság eszköze* és a *Szövegbányászat* című két kötet.

Ebben az időben látott napvilágot Sántha Kálmán pedagógiai tárgyú tanulmányai, könyvei, melyek az első kísérletek a kvalitatív szövegelemzés alkalmazására nem nyelvész szakemberek által.²⁸

Inkább elemzésmódszertani mű Corbin Juliet M. *A kvalitatív kutatás alapjai: a Grounded Theory elemzési módszer technikája és eljárásai*, a L'Harmattan kiadó által megjelent munkája, melynek eredeti kiadása 2008-ban látott napvilágot nálunk pedig 2015-ben. Azonban a MAXQDA programra alapozva gyakorlati tanácsokat is ad.

2010 után a külföldi szakirodalomban is több monográfiát jelentettek meg a szövegelemzésről a társadalomtudományokban.²⁹ Ebben az időszakban adták ki az irodalomtudományban használható

¹⁹ Gerard Salton 1988: *Automatic Text Processing: The Transformation Analysis and Retrieval of Information by Computer* (Addison-Wesley series in computer science)

²⁰ Prószéky Gábor: *Számítógépes nyelvészet* (Számítástechnika-Alkalmazási Vállalat, 1989)

²¹ Prószéky im. 208-236

²² Earl Babbie: *A társadalomtudományi kutatás gyakorlata* 1995

²³ *Számítógép a történettudományban* · Evan Mawdsley – Thomas Munck 1996

²⁴ Prószéky and Balázs 1999 - *Számítógéppel emberi nyelven.pdf*

²⁵ Székelyi Mária - Barna Ildikó. 2002. *Túlélőkészlet az SPSS-hez: Többváltozós elemzési technikákról társadalomkutatók számára.*

²⁶ Szirmai 2005 - *Bevezetés a korpusznyelvészetbe*

²⁷ Eisenhart 2006 - *Kvalitatív adatelemzés módszere.pdf*, Informatikai alapismeretek történelem alapszakos hallgatóknak 2006, Abonyi János (szerk.) 2006: *Adatbányászat a hatékonyság eszköze. Útmutató kezdőknek és haladóknak.* Budapest, Computer Books Kiadói Kf., Tikk Domonkos (szerk.) 2007: *Szövegbányászat.* Budapest, Typotex Kiadó.

²⁸ Sántha Kálmán. 2006. *Mintavétel a kvalitatív pedagógiai kutatásban.* [Budapest]: Gondolat. Sántha Kálmán. 2009. *Bevezetés a kvalitatív pedagógiai kutatás módszertanába: [Oktatási Segédanyag].* Budapest: Eötvös J. Kvk., Sántha Kálmán. 2011. *Abdukció a kvalitatív kutatásban: Bizonytalanság vagy stabilitás?* Budapest: Eötvös J. Kvk., Sántha Kálmán. 2013. *Multikódolt adatok kvalitatív elemzése.* Budapest: Eötvös J. Kvk., Sántha, Kálmán. 2016. “Metaforaháló és kognitív térkép a tanári nézetek vizsgálatában.” *Magiszter*, no. 2: 4–11.

szövegelemzésről szóló monográfiát és tanulmányt is.³⁰ Ezek közül Jockers kötete az irodalom szakos egyetemi hallgatóknak szól a szkript alapú R nyelv szövegelemzési képességeiről. Ezen törekvéseknek is a hatására szerveződnek konferenciák a digitális bölcsészet (digital humanities, DH) elnevezése köré 2015-ben, Számítógép az irodalomtudományban címmel.³¹ Ugyanebben az évben a Szegedi Egyetemen is konferenciát rendeznek a digitális bölcsészet tárgykörében.³² A konferenciáknak köszönhető születnek meg Labádi Gergely irodalomtudományi szövegelemzési tanulmányai³³ és jön létre a jövőt szimbolizáló webportál, mely online szolgáltatásként bármely magyar nyelvű szöveg grammatikai és bizonyos kvantitatív elemzéseit elvégzi.³⁴

A magyar szakirodalomban a kvantitatív szövegelemzés legteljesebb monográfiáját 2016-ban Sebők Miklós állította össze, a politikatudományban való alkalmazásra koncentrálva.³⁵ 2014-ben és 2018-ban módszertanilag és az alkalmazás szempontjából is kiemelkedő jelentőségű publikációk láttak napvilágot a politikatudomány, valamint a törvényhozás munkájának elemzése témakörében.³⁶

Nyelvészeti területen létrejött az e-magyar.hu digitális nyelvfeldolgozó rendszer, mely a magyar számítógépes nyelvészet egyik legnagyobb teljesítménye, és amelyet később részletesen ismertetünk.³⁷

A jogtudomány területén Markovich Réka és szerzőtársai publikáltak a jogszabályszovegek gépi elemzéséről több tanulmányt is, melyekre még később visszatérünk.³⁸ Jelen kötetben Vincze Veronika tanulmánya tartalmaz számítógépes nyelvészeti elemzéseket, pontosabban grammatikai, kvalitatív és kvantitatív elemeket, és Zódi Zsolt tanulmánya kvantitatív számítógépes szövegelemzést.³⁹

Nem lenne teljes a kép, ha nem említenénk meg az egyik legújabb tanulmányt, amely a kapcsolatháló elemzéssel foglalkozik, és módszertani alapjául is szolgált saját tanulmányunknak.⁴⁰ Ez a történeti tárgyú elemzés olyan kapcsolatháló elemzést ír le, mely technológia a szemantikus hálózat feltérképezésére is használható.

A kvantitatív szövegelemzés a társadalomtudományokban

Az előbbi részben láthattuk, hogy 2010 után is több külföldi és hazai munka foglalkozik a szövegelemzéssel (text mining) a társadalomtudományokban, sőt a digitális bölcsészetben is. Ezen belül is a Sebők Miklós által szerkesztett mű⁴¹ kifejezetten a kvantitatív szövegelemzéssel foglalkozik a politikatudományban. A fő elemzési típusok ismertetésén kívül inkább az elemzések fontos

²⁹ Mayring 2014 - Qualitative content analysis - theoretical foundation, basic procedures and software solution.pdf, Brendan 2014 - Statistical Text Analysis for Social Science.pdf, Wiedemann 2015 - Text Mining for Qualitative Data Anal... - Text Mining for Qualitative Data Analysis in the Social Sciences - SpringerLink

³⁰ Jockers 2014 - Text Analysis with R for Students of Literature.pdf, Clement, Tanya (2014): Text Analysis, Data Mining, and Visualizations in Literary Scholarship. In: Price, Kenneth M. – Siemens, Ray (eds.): Literary Studies in the Digital Age. Modern Language Association of America

³¹ DHU 2015 Számítógép az irodalomtudományban <http://dhu2015.mit.bme.hu/file/repository/dhu2015konferenciakotet.pdf>, <http://www.matud.iif.hu/2016/11/06.htm>

³² <http://digibolcsesz.ek.szte.hu/>

³³ Labádi 2016 - Latra vetett szavak - A kőszívű ember fiai kiadásairól és átiratairól

³⁴ <http://dighum.bibl.u-szeged.hu/tanit/>, Labádi et al. 2018 - TANIT – Magyar nyelvű szövegeket elemző eszköz összehasonlító digitális bölcsészeti feladatokhoz.pdf

³⁵ Sebők 2016 - Kvantitatív szövegelemzés és szövegbányászat a politikatudományban

³⁶ Birher et al. 2014 - A kvantitatív szövegelemzés mint előrejelzési lehetőség.pdf politológia 2014 Pénzügyi Szemle, Sebők et al. 2018 - A törvények formális minősége - empirikus vázlat

³⁷ Váradi et al. 2017 - Az e-magyar digitális nyelvfeldolgozó rendszer.pdf

³⁸ Markovich et al. 2015 - Jogszabályszovegek gépi elemzésének tanulságai.pdf

³⁹ Vincze Veronika 2018 - A Miskolc Jogi Korpusz nyelvi jellemzői.pdf 9-36, Zódi Zsolt 2018 A jog érthetőségének határai. Meg tudják-e oldani a nyelvészek a jogi szövegek érthetőségének problémáját. 257-276

⁴⁰ Gulyás Attila, Szabó Martina Katalin, Ifj. Boros István, Havadi Gergő. 2018. "A Rákosi-éra pártjegyzőkönyveinek feldolgozása, elemzése és vizualizációja szövegalapú kapcsolatháló-elemzési módszerekkel." XIV. Magyar Számítógépes Nyelvészeti Konferencia, 109–24.

⁴¹ Sebők 2016

összefüggéseire kívánunk rámutatni, melyek azok a feltételek, amelyek nélkül nem lehet hatékony kvantitatív szövegelemzést végezni.

Két nagy csoportot különböztetünk meg a szövegelemzésben: *kvalitatív* és *kvantitatív*. A kvalitatív szövegelemzés előzetes kódolásra vagy szótáralapú sémarendszerre épül. A kódok kvantitatív elemzésével éri el a kutató a szöveg szemantikai üzenetének elemzését. A kvantitatív elemzésnél mind a kódolás, mind a szótár alapú kategóriaépítés elmarad. A szöveg elemzését az algoritmusokra bízuk építőkövei a szövegegységek és lexémák, bizonyos elemzéseknél finomhangolásként tanulási folyamatot is megengedve a végrehajtó algoritmusnak. A vizsgálati egységek létrehozásához komoly grammatikai, számítógépes elemző algoritmusokra is szükség lesz. Ezek biztosítják az algoritmusok minél pontosabb elemzését a szövegen. Az algoritmusokról és a grammatikai folyamatokról a későbbiekben írok.

A *kvantitatív szövegelemzésnek* is két nagy típusát és több kisebb elemzési módját különböztetjük meg, melyeknek jó része a két nagy típusra épül. Az egyik nagy típus az *osztályozás* (classification). Az osztályozás felügyelt tanulás - a rendelkezésre álló elemek előre meghatározott osztályokba való sorolása, amelynek során azt tanulja meg az algoritmus folyamatos pozitív és negatív megerősítések segítségével, hogy mi alapján lehet az egyes elemeket a megfelelő csoportba, osztályba tenni. Az osztályozásban az általunk ismert tematikai csoportokba kívánjuk rendezni a dokumentumokat, a szótár-alapú osztályozási eszköz. Az osztályozásra a leggyakrabban a Naiv Bayes algoritmus alkalmazzák.

A másik nagy típusnál a dokumentumok *csoportosításánál* (clustering) éppen fordítva azt várjuk, hogy a program hasonlóság alapján csoportosítson tematikai szövegegységeket. Csoportosítás: *felügyelet nélküli tanulási* módszerek, amelyekben a szövegek csoportokba rendezéséhez nem tudjuk előre a csoportcímkéket. Szegmentálás/klaszterezés: felügyelet nélküli tanulás - a rendelkezésre álló elemek csoportosítása valamilyen tulajdonság alapján. Az osztályozással szemben itt nincs megerősítés, az algoritmus maga keres kohéziót az adatok között, illetve közvetlenül nem befolyásolható, hogy milyen tulajdonság alapján alakuljon ki a csoport.

További elemzési lehetőségek a *témamodellezés* (topic modelling) létrehozása. A szövegklasszifikációs feladat speciális esete a topic felismerés. Itt egy *bizonyos kategóriába tartozó* (pl. *adott témáról szóló*, innen az elnevezés) dokumentumokat kell elkülöníteni az összes többit.

A *vélemény vagy emóció tartalom elemzés* (sentiment analysis) is alapvetően szótárra épülő elemzési típus, mely a szövegben található szavak alapján kívánja feltárni a szövegben megfogalmazott vélemény jellemzőit (pl. pozitív, negatív minősítők). Speciális osztályozási feladat.

További lehetőség a *névelemfelismerésen* (named entity recognition) alapuló elemzés, mely lehetővé teszi a személynevek, helyek, szervezetek feltárását a korpuszon belül. Ez a felismerés lehet az elemzés tárgya vagy szűrés alkalmazása a szövegen, ha nem kívánjuk a névelemeket is bevonni.

Hálózatépítés: kapcsolatok feltárása, azonosítása nevek, fogalmak között szemantikus térkép kialakítása. A szövegalapú kapcsolatháló-elemzés egy fontos eredménye a szövegstruktúra olyan módon történő felrajzolása, amely rámutat a szövegben rejlő témákra és az ehhez tartozó szövegrészekre.

Szekvencia/sorrendiség meghatározás: a szavak sorrendiségének elemzése. Szekvenciális minták keresése, elsősorban a párbeszédék feltárására, diskurzusok tanulmányozására.

Koreferencia elemzése: Azt a jelenséget, amikor ugyanarra az entitásra (személy, hely, szervezet stb.) különböző nyelvi jelölőkkel hivatkozhatunk, koreferenciának nevezzük.

Regresszió: általában létező, megismerhető, de hiányzó értékek becslése más paraméterek segítségével (pl. valakinek a korát becsüljük egyéb jellemzőiből).

Faktoranalízis: Célja a közvetlenül meg nem figyelhető összefüggések azonosítása megfigyelhető változók alapján. Kiszámú faktort próbál azonosítani, mely faktorok az összefüggő változók közötti kapcsolatot jellemzik.

A természetesnyelv-feldolgozás (natural language processing, NLP) egyre több vizsgálati szempontot tesz lehetővé a szövegelemzés számára. Az egyre tökéletesebb algoritmusok egyre több eljárás hatékonyságát arra a szintre emelték, ahol már érdemes vele elemzéseket végezni. A Szegedi Egyetemen fejlesztett "Magyarlanc" szoftver tudása mögött óriási adatbázisok találhatóak: függőségi fabank, tulajdonnevek lemmatizálása, többszavas kifejezések, magyar szóegyértelműsítő korpusz.

A kvantitatív szövegelemzés elméleti háttere: statisztikai algoritmusok

A számadatok elemzése a statisztika tudományában különböző algoritmusok segítségével történik. Az algoritmusok bizonyos logikai eljárásokat végeznek el a számok amelyek különböző témákhoz tartozhatnak. A kvalitatív tartalomelemzésben a kódokon ugyanilyen algoritmusokat lehet futtatni a logikai lehetséges eredmények kinyerésére, melyet utána értelmezni lehet. 2011 után a kvantitatív algoritmusokkal magán a vizsgálandó szöveg szavain, kulcsszavain történnek meg ezek a statisztikai műveletek. Ezek eredményeit kell értelmezni a szavakat számokként tekintve az értelmezett szövegek tematikájának szaktudományai szempontrendszerei szerint (jogtudomány, politológia, antropológia, irodalomtudomány, történettudomány stb.)

A szövegek (kvantitatív) adattá, azaz számokká való átalakítására azért van szükség, hogy matematikai, illetve statisztikai műveleteket tudjunk végezni velük.

A kvantitatív szövegelemzés az utóbbi tíz esztendőben nagy áttöréssel ment keresztül a számítógépes nyelvészet és a társadalomtudományi kutatók körében. Az elmúlt harminc év nyelvészetét a „racionalista” (szabály-alapú, szimbólumkezelő) és az „empirista” (statisztikai alapú, gépi tanulásos) nyelvészeti modellek harca jellemezte. Az utolsó néhány év mélytanulmányos forradalmának köszönhetően megfordult, és egyértelműen az empirista paradigma a jellemző. Tudományszociológiai szemszögből a racionalista és az empirista kutatási modellek közti különbség lényege a felülről vezérelt (top down) és az alulról kiinduló (bottom up) keresési stratégia.

Az első áttörést Collobert 2011-ben hozták meg, akik egyszerre, ugyanazon vektorok felhasználásával, tudtak javítani több olyan klasszikus feladat addigi legjobb eredményén, mint a szófaj szerinti címkézés (part of speech tagging), a névelem-felismerés (named entity recognition), a sekély mondattani elemzés (chunking), és a szemantikai szerep felismerése.

Egyre növekszik a fabank (treebank) korpuszok száma, ezek egységesített (univerzális) szófaj- és függőség-tipológián alapulnak, és ezzel nagyban elősegítik a minden emberi nyelvre kiterjedő univerzálé-kutatást.

Az ún. rejtett Markov Modellek (hidden Markov Model, HMM) felhasználásával kapott először fontos szerepet a valószínűségi nyelv- modellezés (language modeling).

A szóvektorok általánosításaival mátrix- és projektívtér-modellekre. A mátrix beágyazások esetében egy szóhoz nem egy vektort, hanem egy mátrixot rendelünk. Mint ismeretes, a gépi tanulást (machine learning) máig a címkézett adatokon alapuló ún. felügyelt tanulás (supervised learning) dominálja.

A nyers adatokon alapuló felügyeletlen (unsupervised) tanulás nagy erővel kutatott terület, ahol komoly eredményekről csak az utóbbi 10 évben beszélhetünk, és az igazi áttörés, a felügyeletlen struktúra-tanulás, még várat magára.⁴²

⁴² Ács et al. 2018 - Hibrid nyelvtechnológiák

A kvantitatív szövegelemzés elméleti háttere: grammatikai elemzők

Az e-magyar.hu digitális nyelvfeldolgozó rendszer mintáján keresztül ismertetem.⁴³

Az emToken - tokenizáló Az eszköz megállapítja, hogy a bemenetként megadott magyar nyelvű szövegben hol található a mondat- és szóhatárok. Az emMorph - morfológiai elemző Az elemzőlánc morfológiai elemző része a szöveg minden egyes szóalakjához hozzárendeli az összes lehetséges morfológiai, morfoszintaktikai elemzését. Az emLem⁴⁴ szótövesítő A morfológiai elemző az egyes morfémák szintjéig menő részletes elemzést ad az egyes szóalakokra: a toldalékok azonosításán túl összetételi tagokra bontja a szót, képzőket azonosít. Az emTag - egyértelműsítő A program a betanult anyag alapján meghatározza a korábban tokenekre bontott mondat minden tokenjének szótövét és szófaját, majd ezt címkével jelöli is. Az emDep - függőségi elemző Az eszköz a mondatok szerkezeti egységei (szavak, többszavas kifejezések) közötti függőségi viszonyokat tárja fel. Az emCons - összetevős elemző Mondatok összetevős szerkezeti elemzése azt tárja fel, hogy a szavak egymással kombinálódva milyen kifejezéseket alkotnak, illetve hogyan állnak össze egy mondattá. Az emChunk - főnévi csoport (frázis) felismerő Az emChunk modul a szövegben a maximális NP-ket azonosítja, vagyis olyan NP-ket, melyek nem részei egy magasabb szintű NP-nek sem. Az emNer automatikus tulajdonnév-felismerő rendszer azonosítja a folyó szövegben található tulajdonneveket, és besorolja őket az előre meghatározott névkategóriák valamelyikébe (személynév, intézménynév, földrajzi név, egyéb).

⁴³ Sass et al. 2017 - Az e-magyar rendszer GATE környezetbe integrált magyar szövegfeldolgozó eszközlánca.pdf, Váradi et al. 2017 - Az e-magyar digitális nyelvfeldolgozó rendszer.pdf

⁴⁴ Zsibrita et al. 2013 - magyarlanc - A Toolkit for Morphological and Dependency Parsing of Hungarian.pdf

Jogtörténeti elemzési kísérlet

A kvantitatív szövegelemzés módszertanának általános bemutatása után a konkrét elemzések menetét és eredményét mutatjuk be.

Elemzésünk korpusza - a módszertani jelleg miatt - a Miskolci Jogi Korpuszból csak egy kódex szövege: a Polgári törvénykönyv (2013. évi 5. tv, Ptk.). Jogtörténeti párhuzamként pedig az 1928. évi magánjogi törvényjavaslat (Mjt.) szövege. Ez utóbbit digitális formában, hibátlan szöveggel az “Ezer év törvényei között” érhattunk el.⁴⁵

Korábban kísérletképpen próbálkoztunk 19. századi magánjogi jogszabály szövegekkel, de a nyomtatott könyvek OCR-ezett szövegei túlságosan hibásnak bizonyult az elemző szoftver számára: több szó egybeírva került a korpuszba, ill. hibás betűkkel. A korabeli szóhasználat miatt a lemmatizálás is sok hibával működött. Ezért úgy döntöttünk, hogy 100 éven belülről meggyünk csak vissza az időben.

Következő kérdésünk az volt, hogy két kódex szövegére milyen kvantitatív elemzést tudunk futtatni. A két törvény szövege jól struktúrált, gyakorlatilag a tartalomjegyzékek alapján tökéletes tématerkép lehet összeállítani.⁴⁶ Ezért az elemzések közül a szavak együttes előfordulására építő elemzések közül az Asszociációt és a Szekvenciális és hálózati elemzést választottuk ki. Az asszociációs elemzés azonban elsősorban az egy az egyhez típusú elemzést, ezért itt kiválasztott szavakra is szükség volt. Ennek a kiválasztásnak szempontja, hogy mindkét szövegben szerepeljen és a gyakoriságban szignifikáns eltérés legyen. A következő öt kulcsszót került elemzésre: elévülés, elidegenítés, haszonélvezet, jognyilatkozat, köteleusrész, vagyoni jog.

Az 1928. évi magánjogi törvényjavaslat 12.166 egyedi szóból áll, 67.807 előfordulással, melyekből 5.700 egyszer előforduló szó (hapax legomai). A 2013. évi Polgári törvénykönyv 12120 egyedi szóból áll, 75.341 előfordulással, melyből 5.207 egyszer előforduló szó (hapax legomai). A T_LAB szövegelemző szoftver Khí négyzet alapú algoritmus alapján 1812 kulcsszót választott ki. Az elemzési szövegegységeket - automatikusan - szövegcsomokban határozta. “Subset”-et, változókat nem alkalmaztunk, automatikus stopszavakat a szoftver hozott létre. A szavakat külső szoftverrel lemmatizált alakokkal szótövesítettünk.

A két jogszabály szigorú szabályok szerint tagolt részekből áll. A Ptk. 8 könyvből, 64 részből, 120 címből és 175 fejezetből áll. A könyvek tematikája: az ember mint jogalany, a jogi személy, családjog, dologi jog, kötelmi jog, öröklési jog. Az 1928. évi magánjogi törvényjavaslat 4 részből, 32 címből, 101 fejezetből áll. A részek tematikája: személyi és családi jog, dologi jog, kötelmi jog, öröklési jog.

A szöveg elemzését nem a szakszókincs vagy a grammatikai elemzésekhez is gyakorta alkalmazott kollokációval végeztük, hanem a szavak egy szövegegységen belüli előfordulásának vizsgálatával, valamint az egy az egyhez hozzárendelésének kapcsolatát. Történeti szempontból mindenképpen érdekes a szókincs változása, de ez alapvetően a megváltozott jogrendszerhez ill. az általános nyelvi változások miatt mutathatóak ki. Ennél sokkal jelentősebb a kiválasztott szavak együttes előfordulásának vizsgálata és az ebből levonható szemantikai kapcsolat elemzése. A Mellékletben az összes kiválasztott szó együttes előfordulásának grafikus és táblázatos eredményeik közöljük. Szövegesen csak egy szó elemzésére kínálkozik hely.

Említettük, hogy a szavak együttes előfordulása nem a konkordanciák, a kollokációk értelmében vesszük, hanem az egy szövegegységen belüli előfordulást függetlenül a szavak egymáshoz viszonyított elhelyezkedésétől. A szavak együttes előfordulásának elmélete szerint ugyanis, ha két vagy több szó rendszeresen ugyanabban a nyelvi kontextusban vannak, azok meghatároznak egy szemantikai teret.

⁴⁵ <https://net.jogtar.hu/ezer-ev-torvenyei>

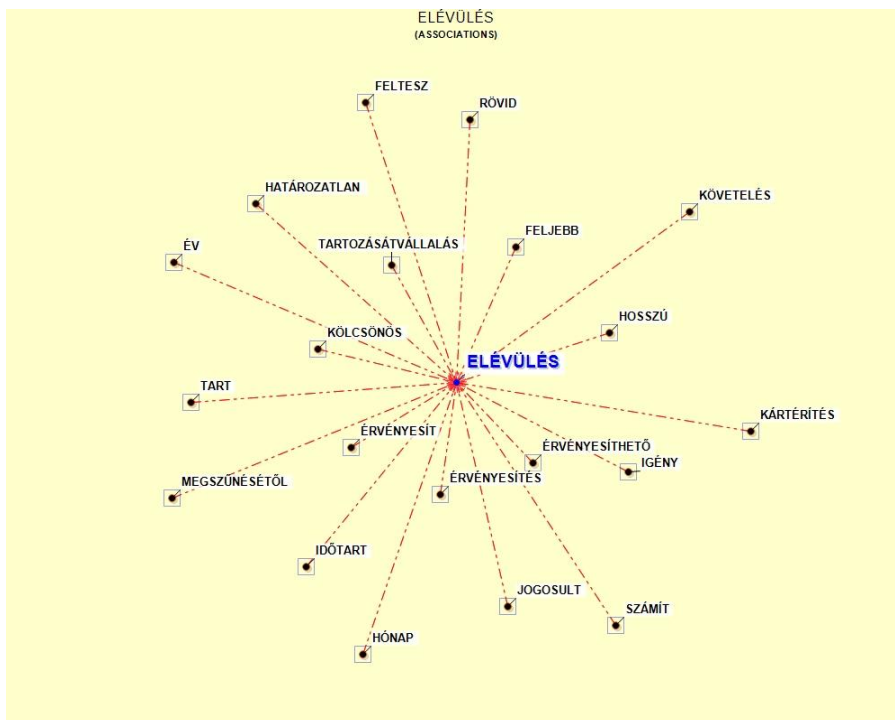
⁴⁶ melyeket a T-LAB saját topic modellező algoritmusával is bizonyított

Kiválasztott szavunk az elévülés. Elévülés: a római jogban praescriptio; az idő múlásának hatása valamely jog, kötelezettség vagy igény érvényesíthetőségének, kikényszeríthetőségének vagy létezésének megszűnésére.

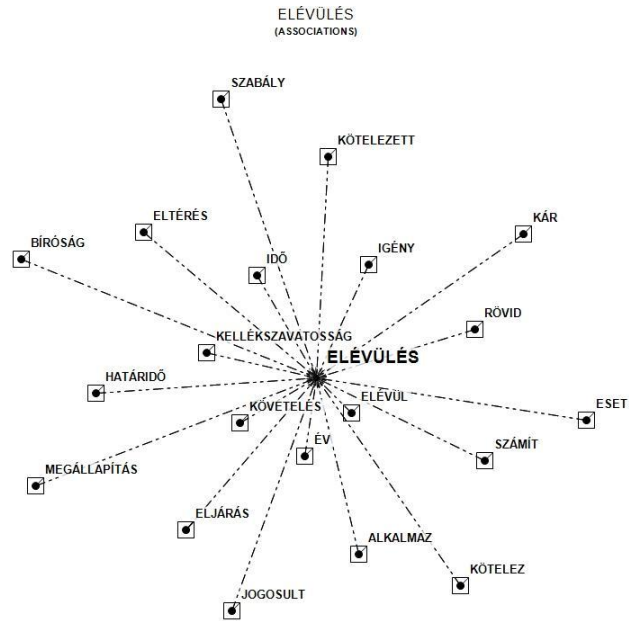
A mellékletben látható szókapcsolati grafikonok és táblázatok alapján a Ptk.-ban leggyakoribb szavak az együttes előfordulásra: *év, idő, rövid*, valamint az *elévül, igény, követelés, kellékszavatosság*. Ugyanez a szókapcsolati előfordulás a Mjt-ben: *történt, befejezés, félbeszakadás, nyugvás* valamint *követelés, keresetindítás*.

Ebből látszik, hogy a Ptk-ban az idő fogalmi körébe tartozó szavak dominálnak a szókapcsolati háló egyik csoportjában, míg a Mjt-ben inkább az időhöz kapcsolódó folyamatok szavai szerepelnek. Ennek a változásnak a további részletezésére egy további elemzést is elvégeztünk a szókapcsolatokon belül az ún. “second order”, mely a szókapcsolatok kapcsolatainak statisztikai leképezése. A Ptk-ban: *hosszú, feljebb, érvényesíthető, érvényesítés, érvényesít, kölcsönös, tartozásátvállalás*. A Mjt-ben: *új, történt, nyugvás, megállapít, befejezés, keresetindítás, per*.

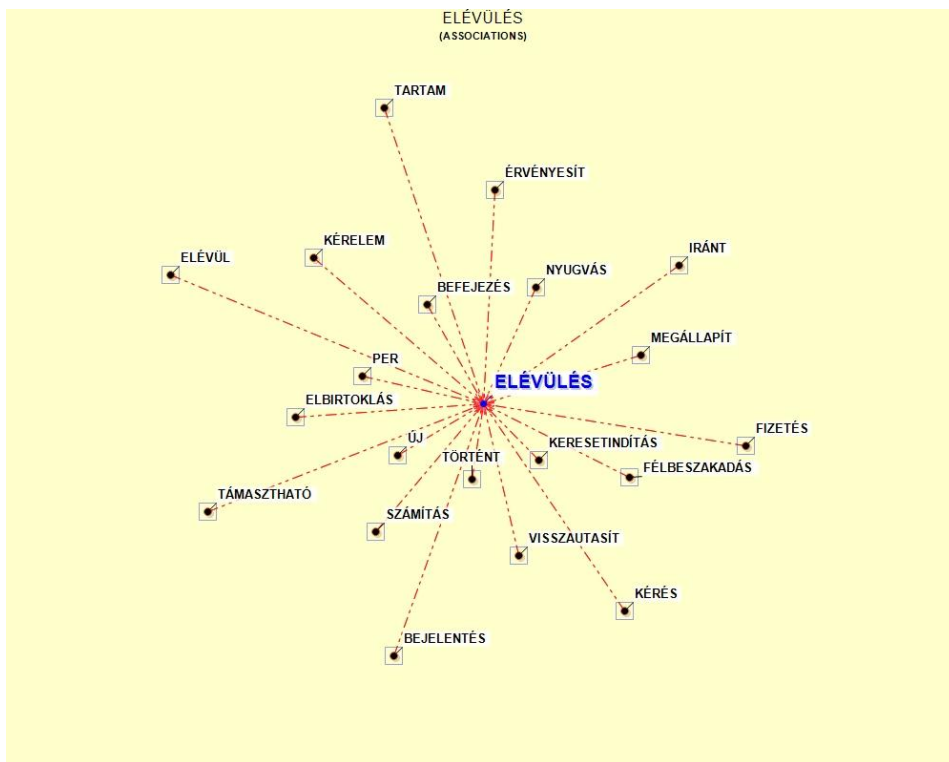
Ezek alapján arra következtethetünk, hogy a Mjt-ben az elévülés ténymegállapítása a domináns, míg a Ptk-ban már az idővel párhuzamosan megtehető érvényesítése a meghatározó.



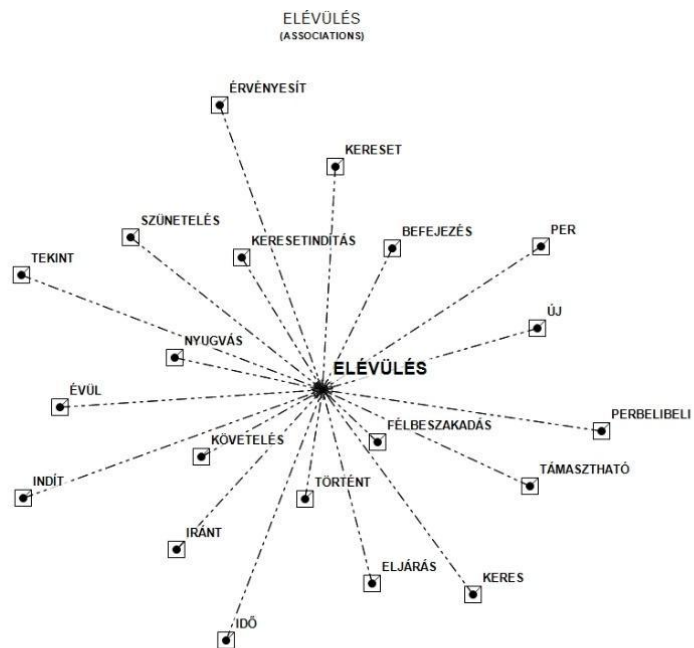
Szókapcsolati grafikon: elévülés (Ptk. 2013) “second order”



Szókapcsolati grafikon: elévülés (Ptk. 2013) “first order”



Szókapcsolati grafikon: elévülés (Mjt 1928) “second order”



Szókapcsolati grafikon: elévülés (Mjt 1928) “first order”

Összefoglalás

Korábban már írtunk arról, hogy a T-LAB beállításai több összetett opció található. Ezek között az egyik kiemelő lehetőség “subset”, mely alatt szövegrészeket és változókat is értünk. A “subset” beállítását azonban a korpusz építése közben kell elvégezni. Ezért a T-LAB szoftver használatának további lépéseit elsősorban ebben az irányban kell folytatni. További elemzési lehetőség a szókapcsolatok feltárásánál a “2nd order”, azaz a másodlagos kapcsolatok elemzése. Ennél az elemzési módnál hangsúlyozzuk ki a kvantitatív szövegelemzés összetettségét. A T_LAB szoftverben található statisztikai algoritmusok alkalmazása csak az adott tudományterület kutatója számára hoz értelmezhető eredményt és nem önmagában kész eredményt ad az algoritmus futtatásakor. A változók és egyéb opciók vagy akár az algoritmusok egymásba ágyazott eredményeivel való vizsgálódás hasonlítható a műszaki tudományok laboratóriumaiban végzett rendkívül időigényes kísérletekhez.

Összességében megállapítható, hogy a T-LAB szövegelemző szoftverrel eddig végzett kísérleteink a úttörő jellegűek annak minden nehézségeivel, de továbbmutató eredményeivel is.

A további kutatások lehetőségét a 2097/2017. (XII. 28.) Korm. határozat adja meg, mely “a Miskolci Egyetemen létrehozandó Tudományos Adatelemző Módszertani Központ és Levéltári Nyílt Kutatási Adatforrástár létrehozása előkészítésének a Beruházás Előkészítési Alapból történő támogatásáról” szól. Ennek jogtudományi és nyelvtudományi becsatornázását éppen ezek a kutatások teszik lehetővé.

MELLÉKLETEK

LEMMA	PTK 2013	MJT 1928
JOG	1487	1131
SZERZŐDÉS	1324	538
SZEMÉLY	985	588
TAG	822	150
JOGOSULT	769	355
ESET	680	626
KÖTELES	597	398
TÁRSASÁG	597	80
VAGYON	563	339
GYERMEK	497	380
KÖTELEZETTSÉG	457	286
DOLOG	452	688
SZÜLŐ	447	339
BÍRÓSÁG	443	193
SZABÁLY	438	471
HÁZASTÁRS	434	256
TELJESÍTÉS	409	252
MEGFELELŐ	403	325
MEGHATÁROZ	373	151
RENDELKEZÉS	369	131
RÉSZVÉNY	365	6
KÖVETELÉS	344	658
KÖZÖS	326	136
VONATKOZÓ	320	123
HOZZÁJÁRULÁS	290	97
ALKALMAZ	274	94

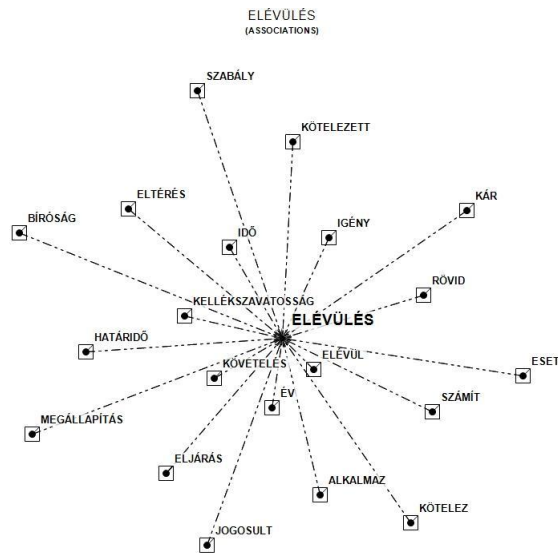
Összehasonlító táblázat a leggyakoribb kulcsszavak előfordulására

LEMMA	SUB	TOT	CHI2
TAG	822	969	394,19
TÁRSASÁG	597	677	340,06
RÉSZVÉNY	365	371	308,94
SZERZŐDÉS	1324	1862	251,32
VEZETŐ	194	195	170,14
IGÉNY	191	198	151,07
MEGSZÜNÉS	270	314	138,33
SZÖVETKEZET	150	142	133,73
KÖZGYŰLÉS	242	278	130,43
KÖVETŐ	183	196	129,04
TULAJDONJOG	156	161	125,24
NYILVÁNTARTÁS	146	148	124,48
TEVEKENYSÉG	148	153	118,14
ZÁLOGTÁRGY	151	159	113,09
BIZTOSÍTÓ	153	162	112,38
JOGOSULT	769	1124	109,59
TAGSÁG	141	149	104,28
SZERINTI	128	132	103,01
JOGSZABÁLY	147	159	99,97
ALAPÍTÓ	177	201	99,90
FELÜGYELET	163	187	88,27
RENDELKEZÉS	369	500	88,02
NÉLKÜLI	136	150	85,98
BIZTOSÍT	212	88	85,78
TARTÁS	269	347	84,68
IGAZGATÓSÁG	97	98	83,59
ELTÉRŐ	152	175	81,12
JOGNYILATKOZAT	217	271	80,50
SOR	104	109	76,65
HATÁROZAT	266	349	76,39
HOZZÁJÁRULÁS	290	387	75,67

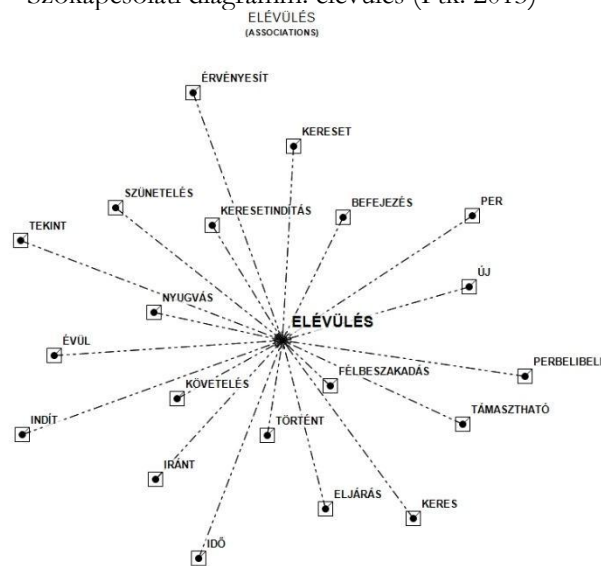
Felülreprezentált szavak viszonyítása a teljes korpuszhoz, Ptk. 2013

LEMMA	SUB	TOT	CHI2
OLY	432	438	468,23
AMENNYI	393	408	397,89
HITTELEZŐ	520	630	318,99
ADÓS	337	380	264,42
JELZÁLOGJOG	266	290	232,22
KÖVETEL	390	500	192,26
KÉTSÉG	178	188	171,03
HACSAK	141	142	155,40
TELEK	168	221	149,46
HAGYATÉK	318	414	147,09
ELŐBB	129	172	145,00
TULAJDONOS	431	603	144,40
TÁRGY	352	469	142,13
KÖVETELÉS	658	1002	139,63
VÉGETT	123	126	129,16
GYÁM	294	374	124,16
ÁLL	348	426	121,06
TEKINTET	303	349	119,84
BÍRÓ	126	134	118,54
JELZÁLOG	108	109	118,29
JOGÜGYLET	127	136	117,02
ELTARTÁS	116	121	115,58
ÖRÖKHAGYÓ	339	476	111,48
CSELEKMÉNY	124	135	108,60
ÖRÖKSÉG	139	158	106,05
ÖRÖKÖS	300	416	104,77
ÉPÍTMÉNY	93	94	101,42
HAGYOMÁNY	131	152	93,27
IDŐ	186	239	90,68
FOG	142	174	81,48
DOLOG	688	1140	80,97
TÖRVÉNYES	291	424	79,11

Felülreprezentált szavak viszonyítása a teljes korpuszhoz Mjt. 1928



Szókapcsolati diagramm: elévülés (Ptk. 2013)



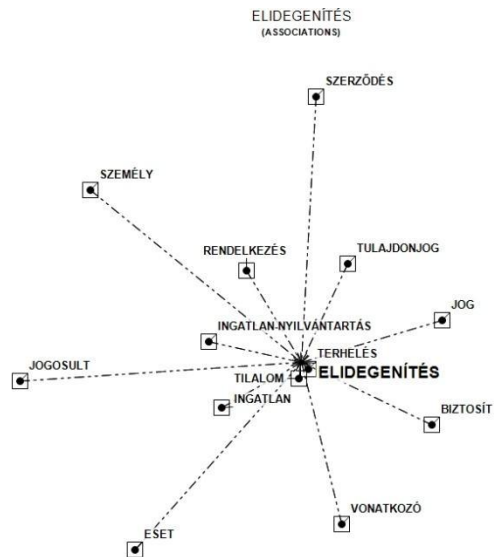
Szókapcsolati diagramm: elévülés (Mjt. 1928)

LEMMA_B	COEF F	EC_B	EC_AB	CHI2
ELEVÜL	0,29	14	7	340,59
ÉV	0,17	98	11	105,93
KÖVETELÉS	0,17	242	17	94,13
KELLÉKSZAVATOSSÁG	0,14	20	4	73,26
IDŐ	0,12	138	9	43,93
IGÉNY	0,11	158	9	36,56
RÖVID	0,09	25	3	30,73
SZÁMÍT	0,08	181	7	15,70
ALKALMAZ	0,08	270	8	11,23
ELJÁRÁS	0,07	117	5	13,02
HATÁRIDŐ	0,07	172	6	11,21
ELTÉRÉS	0,06	51	3	12,43
KÖTELEZETT	0,06	230	6	6,36
KÁR	0,06	173	5	6,51
ESET	0,06	603	9	1,72
KÖTELEZ	0,06	191	5	5,29

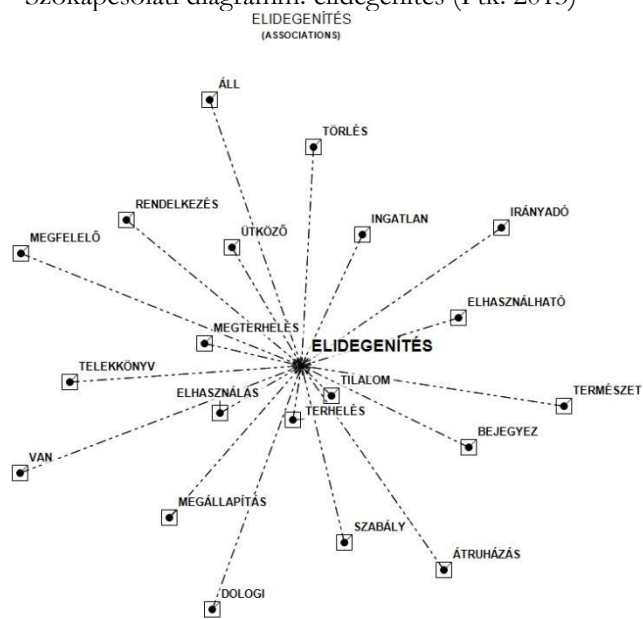
Szókapcsolat táblázata: elévülés, (Ptk. 2013)

LEMMA_B	COEF F	EC_B	EC_AB	CHI2
FÉLBESZAKADÁS	0,42	33	22	793,84
TÖRTÉNT	0,26	11	8	314,86
KÖVETELÉS	0,24	493	49	207,76
NYUGVÁS	0,23	8	6	243,74
KERESÉTINDÍTÁS	0,20	20	8	166,40
BEFEJEZÉS	0,19	12	6	158,67
ÚJ	0,16	8	4	105,69
TÁMASZTHATÓ	0,15	35	8	88,80
ELJÁRÁS	0,15	56	10	83,01
IRÁNT	0,15	46	9	83,27
ÉVÜL	0,13	24	6	73,73
SZÜNETELES	0,13	11	4	74,78
KERESÉT	0,13	86	11	60,14

Szókapcsolat táblázata: elévülés, (Mjt. 1928)



Szókapcsolati diagramm: elidegenítés (Ptk. 2013)



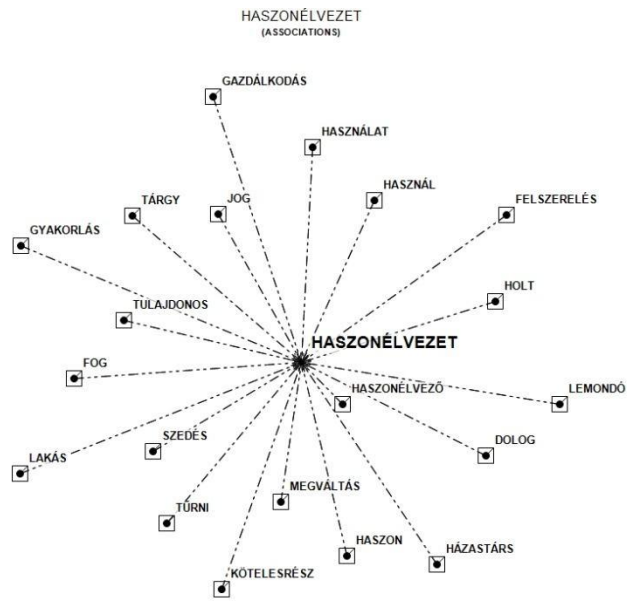
Szókapcsolati diagramm: elidegenítés (Mjt. 1928)

LEMMA_B	COEFF	EC_B	EC_AB	CHI2
TERHELÉS	0,91	11	10	3 468,43
TILALOM	0,51	28	9	1 096,52
INGATLAN	0,11	72	3	42,75
INGATLAN-NYILVÁNTARTÁS	0,11	72	3	42,75
RENDELKEZÉS	0,08	323	5	22,15
TULAJDONJOG	0,08	124	3	22,76
JOG	0,07	961	7	10,38
BIZTOSÍT	0,07	178	3	14,41
VONATKOZÓ	0,05	302	3	6,66
ESET	0,04	603	3	1,50
JOGOSULT	0,04	625	3	1,34
SZEMÉLY	0,03	675	3	1,03
SZERZŐDÉS	0,03	780	3	0,55

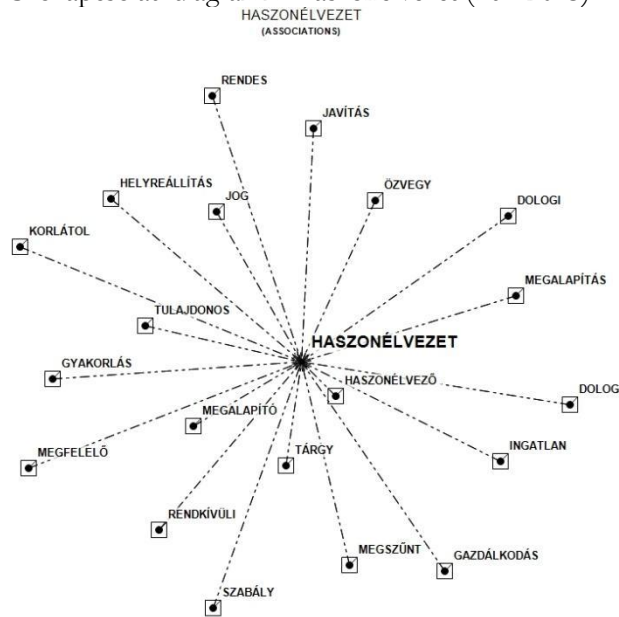
Szókapcsolat táblázata: elidegenítés, (Ptk. 2013)

LEMMA_B	COEFF	EC_B	EC_AB	CHI2
TILALOM	0,48	39	19	1 024,38
TERHELÉS	0,34	35	13	527,75
ELHASZNÁLÁS	0,22	8	4	220,06
MEGTERHELÉS	0,22	13	5	209,54
ÜTKÖZŐ	0,14	11	3	87,39
INGATLAN	0,13	168	11	63,61
ELHASZNÁLHATÓ	0,13	13	3	73,08
BEJEGYEZ	0,11	69	6	48,63
SZABÁLY	0,10	443	14	28,84
MEGÁLLAPÍTÁS	0,10	40	4	38,15
TELEKKÖNYV	0,09	226	9	25,93
RENDELKEZÉS	0,09	110	6	26,77
TÖRLÉS	0,09	30	3	28,55

Szókapcsolat táblázata: elidegenítés (Mjt. 1928)



Szókapcsolati diagramm: haszonélvezet (Ptk. 2013)



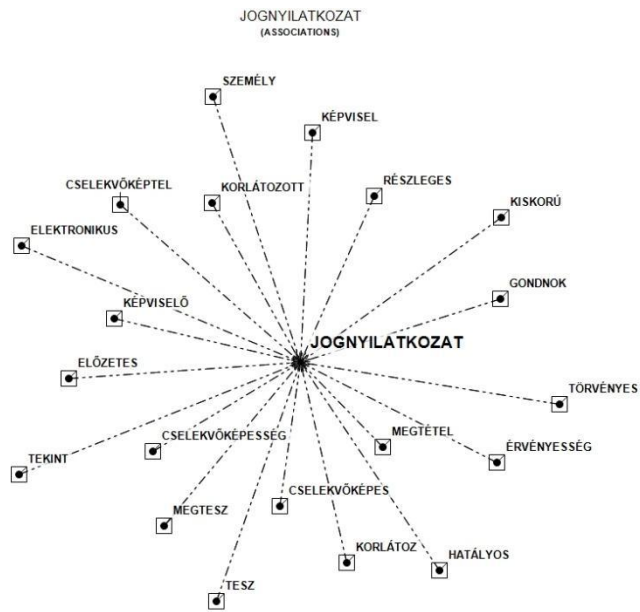
Szókapcsolati diagramm: haszonélvezet (Mjt. 1928)

LEMMA_B	COEFF	EC_B	EC_AB	CHI2
HASZONÉLVEZŐ	0,50	25	19	1 046,44
MEGVÁLTÁS	0,20	7	4	162,96
SZEDÉS	0,18	14	5	123,83
TULAJDONOS	0,17	133	15	100,95
JOG	0,17	961	40	73,22
HASZNÁL	0,16	67	10	93,60
HOLT	0,15	12	4	91,89
DOLOG	0,15	286	19	64,04
HASZON	0,14	42	7	74,26
TŰRNI	0,14	8	3	78,19
FOG	0,14	33	6	70,31
TÁRGY	0,13	97	10	59,42
HASZNÁLAT	0,13	147	12	52,69

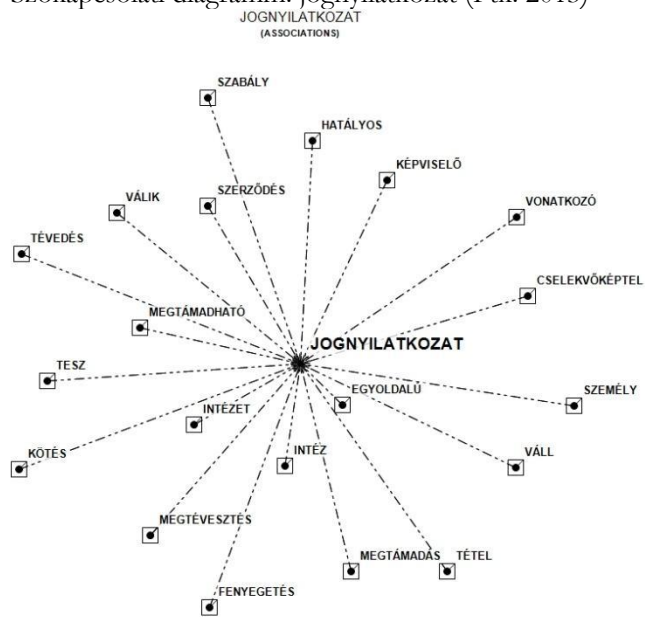
Szókapcsolat táblázata: haszonélvezet, (Ptk. 2013)

LEMMA_B	COEFF	EC_B	EC_AB	CHI2
HASZONÉLVEZŐ	0,57	81	55	1 482,37
TÁRGY	0,25	291	46	232,76
MEGALAPÍTÓ	0,23	20	11	232,83
TULAJDONOS	0,19	351	37	104,54
JOG	0,16	847	49	48,61
ÖZVEGY	0,15	31	9	92,59
MEGALAPÍTÁS	0,13	51	10	63,76
INGATLAN	0,13	168	18	50,05
MEGSZŰNT	0,13	92	13	53,81
RENDKÍVÜLI	0,13	20	6	64,01
GYAKORLÁS	0,12	103	13	45,80
HELYREÁLLÍTÁS	0,11	11	4	53,24
JAVÍTÁS	0,11	11	4	53,24

Szókapcsolat táblázata: haszonélvezet (Mjt. 1928)



Szókapcsolati diagramm: jognyilatkozat (Ptk. 2013)



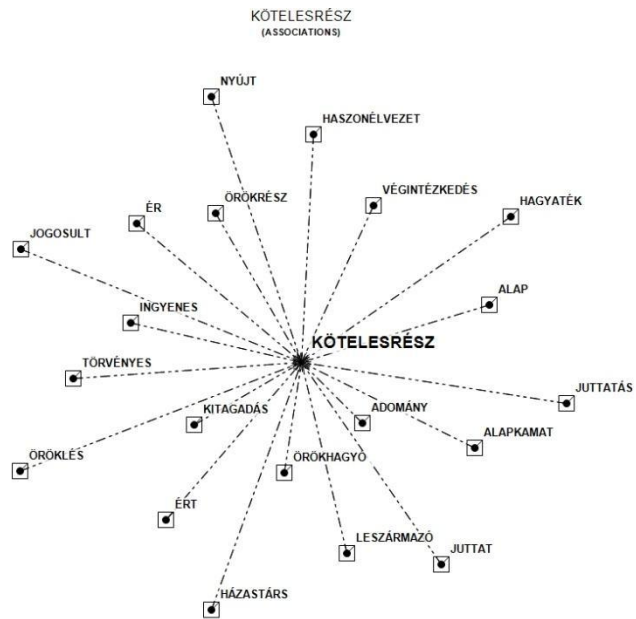
Szókapcsolati diagramm: jognyilatkozat (Mjt. 1928)

LEMMA_B	COEFF	EC_B	EC_AB	CHI2
MEGTÉTEL	0,31	55	28	367,98
CSELEKVŐKÉPES	0,24	26	15	225,75
CSELEKVŐKÉPESSÉG	0,22	74	23	168,22
KÉPVISELŐ	0,20	87	23	137,15
KORLÁTOZOTT	0,20	21	11	148,15
RÉSZLEGES	0,20	51	17	134,91
GONDNOK	0,18	58	17	115,00
ÉRVÉNYESSÉG	0,17	33	12	105,48
KORLÁTOZ	0,17	83	19	93,40
MEGTESZ	0,17	23	10	108,57
ELOZETES	0,17	54	15	94,63
CSELEKVŐKÉPTELEL	0,16	34	11	83,78
KÉPVISEL	0,15	35	11	80,82

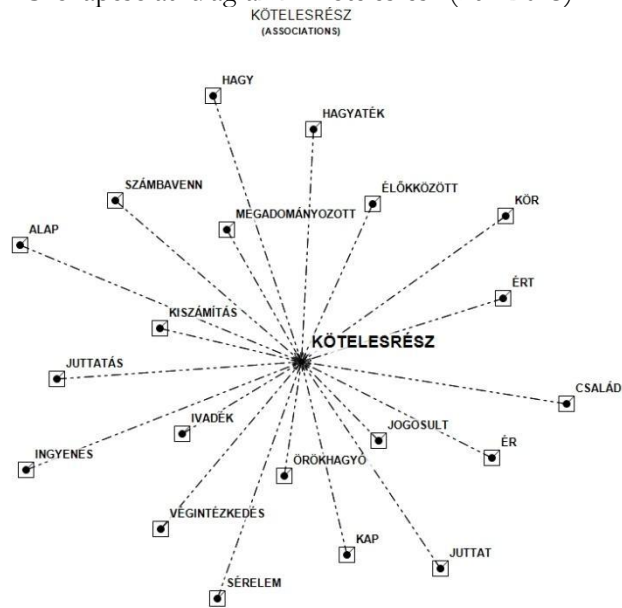
Szókapcsolat táblázata: jognyilatkozat, (Ptk. 2013)

LEMMA_B	COEFF	EC_B	EC_AB	CHI2
EGYOLDALÚ	0,46	20	14	950,50
INTÉZ	0,24	92	16	250,28
INTÉZET	0,23	68	13	224,86
MEGTÁMADHATÓ	0,18	17	5	136,80
SZERZŐDÉS	0,14	422	20	64,01
KÉPVISELŐ	0,13	85	8	60,69
CSELEKVŐKÉPTELEL	0,12	23	4	61,63
VÁLL	0,12	123	9	49,89
MEGTÁMADÁS	0,12	55	6	54,17
MEGTÉVESZTÉS	0,12	14	3	58,18
TESZ	0,12	159	10	45,53
VÁLIK	0,12	130	9	46,41
HATÁLYOS	0,11	61	6	47,82

Szókapcsolat táblázata: jognyilatkozat (Mjt. 1928)



Szókapcsolati diagramm: kötelesrész (Ptk. 2013)



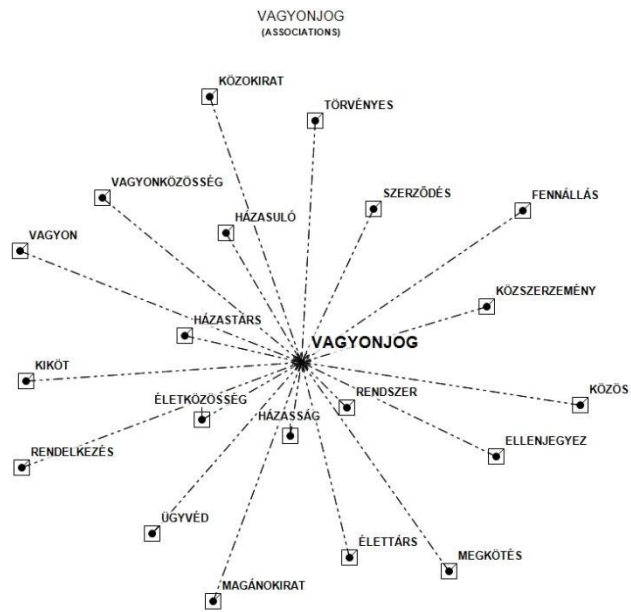
Szókapcsolati diagramm: kötelesrész (Mjt. 1928)

LEMMA_B	COEFF	EC_B	EC_AB	CHI2
ADOMÁNY	0,34	14	7	465,16
ÖRÖKHAGYÓ	0,24	107	14	228,37
KITAGADÁS	0,24	5	3	239,90
INGYENES	0,18	51	7	118,83
ÖRÖKRÉS	0,16	11	3	105,96
VÉGINTEZKEDÉS	0,16	20	4	101,74
ALAP	0,16	65	7	90,65
ALAPKAMAT	0,16	65	7	90,65
LESZÁRMAZÓ	0,14	60	6	71,25
ÉRT	0,14	138	9	65,13
TÖRVÉNYES	0,14	112	8	64,41
ÉR	0,14	142	9	62,89
HASZONÉLVEZET	0,12	57	5	50,89

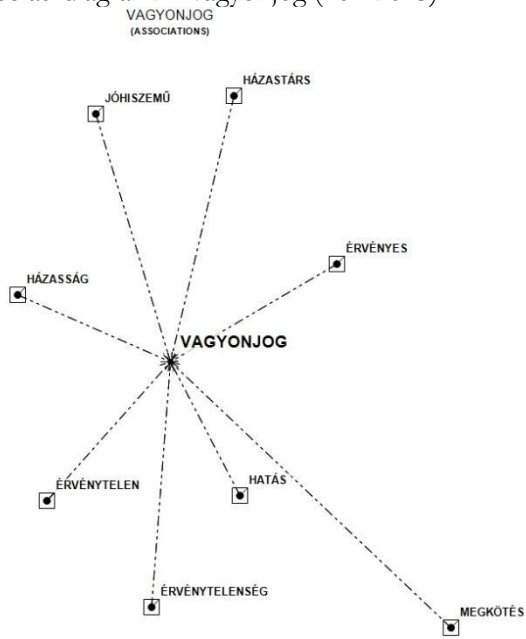
Szókapcsolat táblázata: kötelesrész, (Ptk. 2013)

LEMMA_B	COEFF	EC_B	EC_AB	CHI2
JOGOSULT	0,31	315	49	392,24
ÖRÖKHAGYÓ	0,28	266	41	320,83
IVADÉK	0,26	113	24	267,14
KISZÁMÍTÁS	0,25	13	8	281,72
MEGADOMÁNYOZOTT	0,22	21	9	215,71
ÉLŐKKÖZÖTT	0,19	17	7	160,52
ÉRT	0,17	193	21	102,72
ÉR	0,17	202	21	96,66
KAP	0,17	104	15	104,12
VÉGINTEZKEDÉS	0,16	63	11	95,87
JUTTATÁS	0,15	28	7	92,37
SZÁMBAVENN	0,14	16	5	84,65
HAGYATÉK	0,14	210	18	62,91

Szókapcsolat táblázata: kötelesrész (Mjt. 1928)



Szókapcsolati diagramm: vagyonjog (Ptk. 2013)



Szókapcsolati diagramm: vagyonjog (Mjt. 1928)

LEMMA_B	COEFF	EC_B	EC_AB	CHI2
RENDSZER	0,46	17	11	867,68
HÁZASSÁG	0,37	123	24	551,84
ÉLETKÖZÖSSÉG	0,27	59	12	284,16
HÁZASTÁRS	0,27	240	24	267,69
HÁZASULÓ	0,19	13	4	145,73
SZERZŐDÉS	0,17	780	27	83,87
KÖZSZERZEMÉNY	0,16	10	3	106,34
ELLENJEGYEZ	0,14	13	3	80,50
ÉLETTÁRS	0,14	56	6	69,32
ÜGYVÉD	0,13	16	3	64,36
KIKÖT	0,12	35	4	49,55
VAGYONKÖZÖSSÉ G	0,12	35	4	49,55
TÖRVÉNYES	0,11	112	7	42,40
FENNÁLLÁS	0,11	83	6	43,44
KÖZÖS	0,11	231	10	37,70
MEGKÖTÉS	0,11	84	6	42,80

Szókapcsolat táblázata: vagyoni jog, (Ptk. 2013)

LEMMA_B	COEFF	EC_B	EC_AB	CHI2
HATÁS	0,35	23	5	552,56
ÉRVÉNYTELENS ÉG	0,25	16	3	284,69
ÉRVÉNYTELEN	0,25	29	4	277,92
HÁZASSÁG	0,23	165	9	243,80
JÓHISZEMŰ	0,22	56	5	222,70
HÁZASTÁRS	0,21	154	8	205,12
ÉRVÉNYES	0,21	41	4	194,77
MEGKÖTÉS	0,11	77	3	55,25

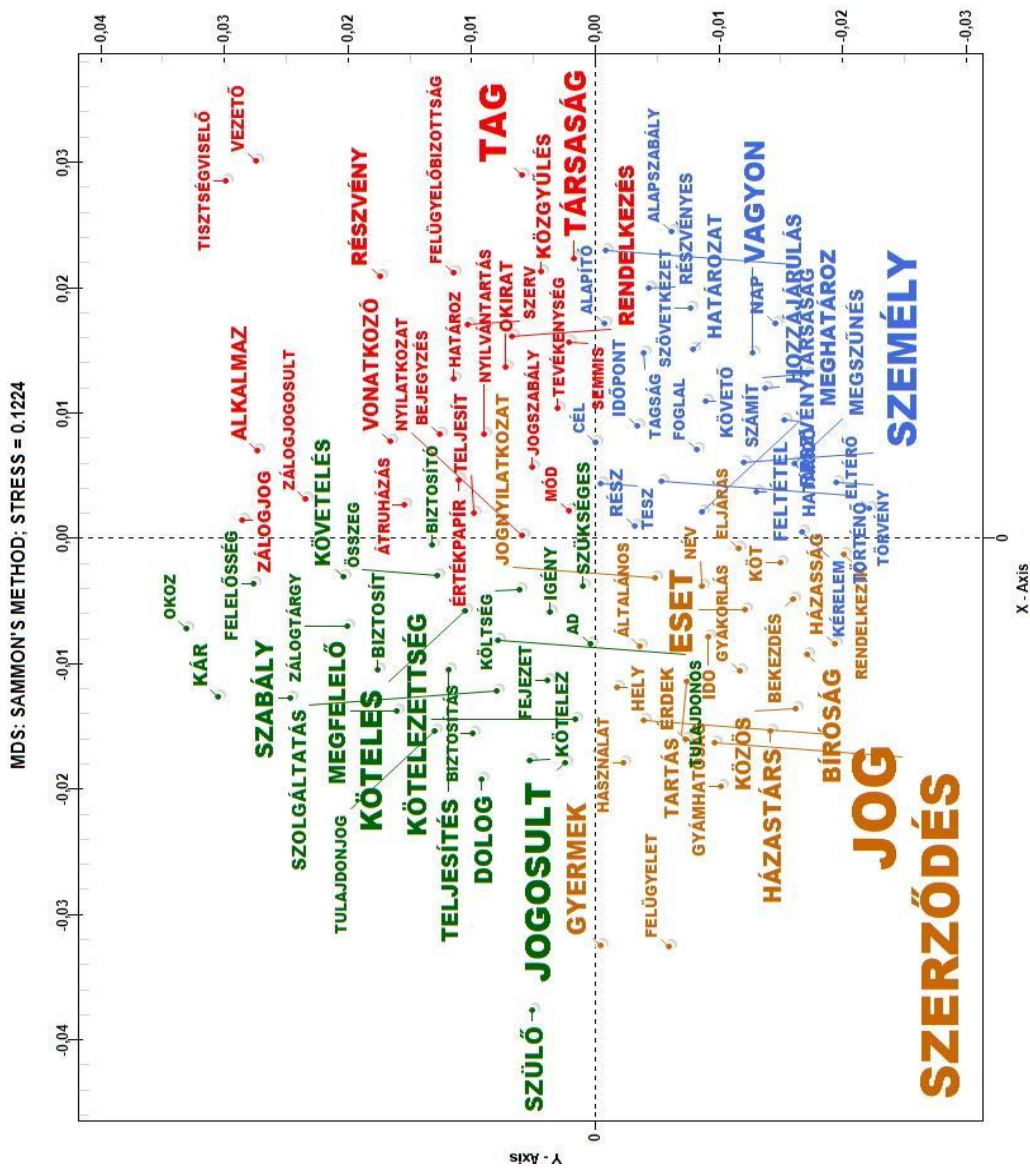
Szókapcsolat táblázata: vagyoni jog (Mjt. 1928)

PRED	SUCC	TOT
ESET	ESETT	683
JOG	SZEMÉLY	326
DOLOG	DOLOGI	312
KÖTELEZ	KÖTELEZETT	222
VEZETŐ	TISZTSÉGVISELŐ	168
ÉR	ÉRT	163
HATÁROZ	HATÁROZAT	146
MEGFELELŐ	ALKALMAZ	144
VAGYON	HOZZÁJÁRULÁS	144
SZÜLŐ	FELÜGYELET	139
VÁLIK	VÁLL	132
MÉRT	MÉRTÉK	132
ÁLL	ÁLLAM	127
BÁRMELY	BÁRMELYIK	114
SZABÁLY	MEGFELELŐ	112
TEKINT	TEKINTET	107
OKOZ	KÁR	103
IRÁNT	IRÁNTI	92
JOG	GYAKORLÁS	92
GAZDASÁG	TÁRSASÁG	90
VONATKOZÓ	SZABÁLY	89
KERÜL	SOR	84
EGYESÜL	EGYESÜLET	81
LÉTESÍTŐ	OKIRAT	78
VALAMELY	valamelyik	76
ALAP	ALAPKAMAT	75
NYILATKOZAT	NYILATKOZIK	74
VÉTEL	VÉTELÁR	73
SZERZŐDÉS	FELTÉTEL	71

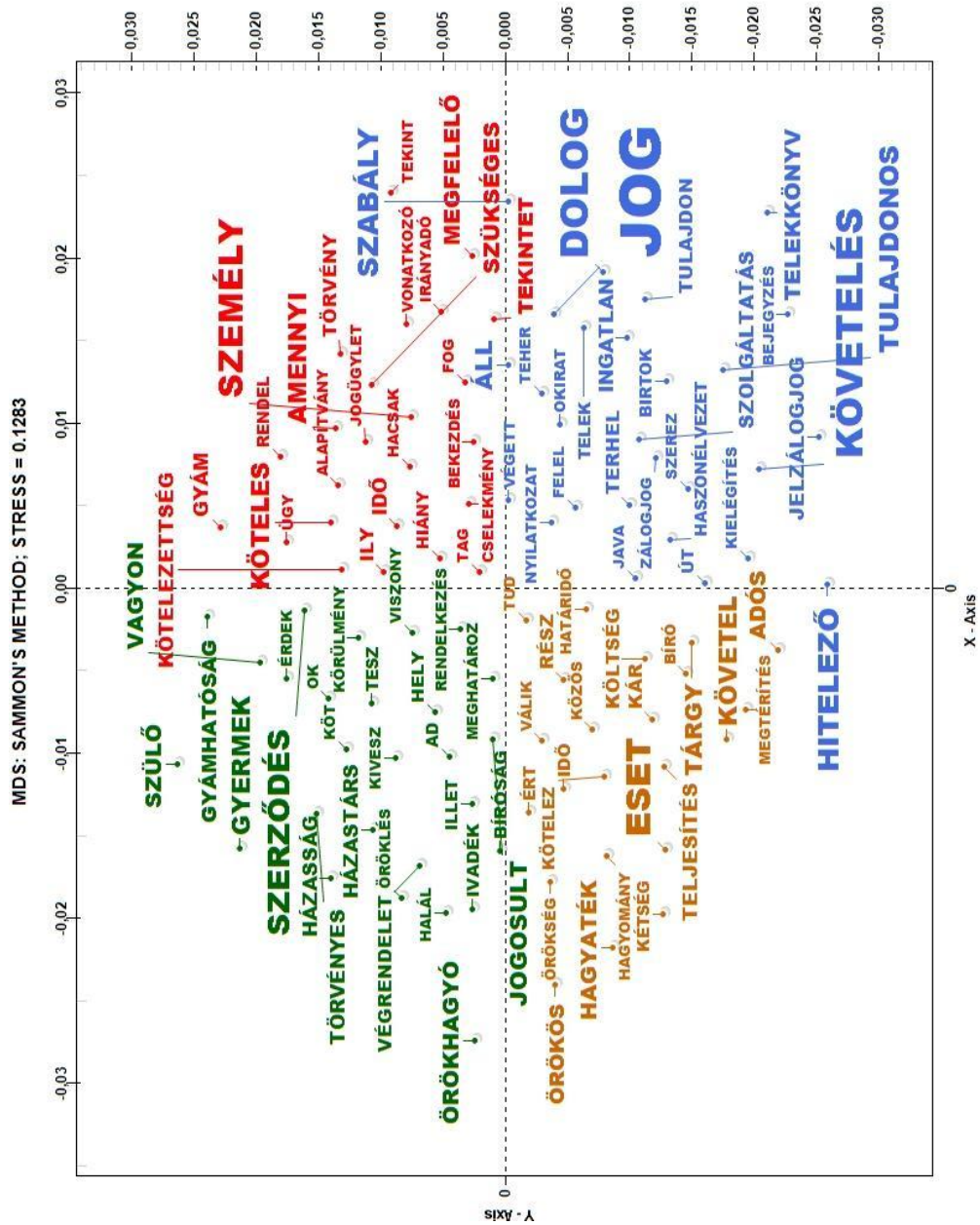
Szavak sorrendisége 2013

PRED	SUCC	TOT
ESET	ESETT	638
DOLOG	DOLOGI	493
TELEK	TELI	348
TEKINT	TEKINTET	304
ÁLL	ÁLLAM	296
ILY	ilyen	215
ÉR	ÉRT	206
KÉTSÉG	ESET	171
VALAMELY	valamelyik	133
SZABÁLY	MEGFELELŐ	131
VÁLIK	VÁLL	126
KÖTELEZ	KÖTELEZETT	107
KÖVETKEZIK	KÖVETKEZTÉBEN	100
JÁR	JÁRUL	98
KERES	KERESET	90
SZAB	SZABAD	88
FELADAT	FELÜL	86
VÉTEL	VÉTELÁR	84
SZÜLŐ	HATALOM	81
TELEKKÖNYV	BEJEGYZÉS	80
BÁRMELY	BÁRMELYIK	75
INTÉZ	INTÉZET	75
JELZÁLOGOS	HITELEZŐ	69
INGÓ	DOLOG	68
ÉL	ÉLET	67
SZOLGÁL	SZOLGÁLAT	67
VONATKOZÓ	SZABÁLY	67
TISZT	TISZTA	66
TÖRVÉNYES	KÉPVISELŐ	64
DOLOGI	JOG	63
SZERZŐDÉS	MEGKÖTÉS	62
MEGFELELŐ	ALKALMAZ	61

Szavak sorrendisége 1928



Multidimenzionális skála Ptk 2013.



Multidimenzionális skála Mtt 1928

A T-LAB SZÖVEGLEMLZŐ MAGYAR NYELVŰ LEMMATIZÁLÓ SCRIPT

A T-LAB szoftver magyar nyelvhez csak toldaléklevágó szótövesítővel rendelkezik (stemming). A kvantitatív szövegelemzéshez azonban a magyar nyelv esetében ez nem megfelelő, ezért egy külön előfeldolgozó scriptet készítettünk mely a korpusz szövegeiből lemmatizált szólistát készít, melyet a T-LAB

A PTK. ÉS A MJT. LEMMATIZÁLT SZÓJEGYZÉKÉNEK RÉSZLETE

ABBA;ABBA
ABBAHAGYÁS;ABBAHAGYÁSÁT
ABBAN;ABBAN
ABBANHAGY;ABBANHAGYJA
ABBANHAGYÁS;ABBANHAGYÁS
ABBANHAGYÁS;ABBANHAGYÁSA
ABBANHAGYÁS;ABBANHAGYÁSÁHOZ
ABBANHAGYÁS;ABBANHAGYÁSÁT
ABBANHAGYÁS;ABBANHAGYÁSÁVAL
ABBANHAGYÁS;ABBANHAGYÁSTÓL
ABBANHAGYHAT;ABBANHAGYHATJA
ABBANHAGYJ;ABBANHAGYJON
ABBANHAGYT;ABBANHAGYTA
ABBANHAGYV;ABBANHAGYVA
ABBÓL;ABBÓL
ABLAK;ABLAK
ABLAK;ABLAKTÓL
ABLAKNYÍLÁS;ABLAKNYÍLÁSON
ABLAKSZOLGALOM;ABLAKSZOLGALOM
ÁBRÁZOL;ÁBRÁZOLT
AD;AD
AD;ADJA
AD;ADJÁK
AD;ADJANAK
AD;ADJON
AD;ADNA
AD;ADNAK
AD;ADNI
AD;ADNIA
AD;ADHAT
AD;ADHATJA
AD;ADHATNAK
AD;AD-E
AD;ADOTT
AD;ADTAK
AD;ADTÁK
AD;ADVA
ADANDÓ;ADANDÓ
ADÁS;ADÁS
ADÁS;ADÁSA
ADÁS;ADÁSÁBAN
ADÁS;ADÁSÁHOZ
ADÁS;ADÁSÁIG
ADÁS;ADÁSÁKOR
ADÁS;ADÁSÁNAK
ADÁS;ADÁSÁNÁL
ADÁS;ADÁSÁRA
ADÁS;ADÁSÁT
ADÁS;ADÁSÁVAL
ADÁSVÉTEL;ADÁSVÉTEL
ADÁSVÉTEL;ADÁSVÉTELE
ADÁSVÉTEL;ADÁSVÉTELÉNEK
ADÁSVÉTEL;ADÁSVÉTELÉRE
ADÁSVÉTEL;ADÁSVÉTELI
ADAT;ADAT
ADAT;ADATAI

ADAT;ADATAIBAN
ADAT;ADATAIBÓL
ADAT;ADATAINAK
ADAT;ADATAIRÓL
ADAT;ADATAIT
ADAT;ADATHAT
ADAT;ADATOK
ADAT;ADATOKAT
ADAT;ADATOKKAL
ADAT;ADATOKNAK
ADAT;ADATOKRÓL
ADAT;ADATOT
ADAT;ADATOTT
ADAT;ADATRÓL
ADATBEVITEL;ADATBEVITELI
ADATIK;ADATHATJA
ADATIK;ADASSÉK
ADDIGI;ADDIGI
ÁDER;ÁDER
ADHATÓ;ADHATÓ
ADÓ;ADÓT
ADÓ;ADÓTÓL
ADÓ;ADÓ
ADÓ;ADÓK
ADOMÁNY;ADOMÁNNYAL
ADOMÁNY;ADOMÁNY
ADOMÁNY;ADOMÁNYAIK
ADOMÁNY;ADOMÁNYAIKNAK
ADOMÁNY;ADOMÁNYÁNAK
ADOMÁNY;ADOMÁNYKÉNT
ADOMÁNY;ADOMÁNYNAK
ADOMÁNY;ADOMÁNYOK
ADOMÁNY;ADOMÁNYOKAT
ADOMÁNY;ADOMÁNYOKKAL
ADOMÁNY;ADOMÁNYOKON
ADOMÁNY;ADOMÁNYT
ADOMÁNY;ADOMÁNYUK
ADOMÁNY;ADOMÁNYUL
ADOMÁNYOZÁS;ADOMÁNYOZÁS
ADOMÁNYOZÁS;ADOMÁNYOZÁSI
ADOMÁNYOZÁS;ADOMÁNYOZÁSKOR
ADOMÁNYOZÁS;ADOMÁNYOZÁSTÓL
ADOMÁNYOZÓ;ADOMÁNYOZÓ
ADOMÁNYOZÓ;ADOMÁNYOZÓNAK
ADOMÁNYOZÓ;ADOMÁNYOZÓT
ADOMÁNYOZÓ;ADOMÁNYOZÓTÓL
ADOMÁNYOZOTT;ADOMÁNYOZOTT
ADOMÁNYTÁRGY;ADOMÁNYTÁRGYAKAT
ADÓS;ADÓS
ADÓS;ADÓSA
ADÓS;ADÓSHOZ
ADÓS;ADÓSNAK
ADÓS;ADÓSNÁL
ADÓS;ADÓSOK
ADÓS;ADÓST
ADÓS;ADÓSTÓL
ADÓSLEVÉL;ADÓSLEVÉL
ADÓSLEVÉL;ADÓSLEVÉLBE
ADÓSLEVÉL;ADÓSLEVELET
ADÓSLEVÉL;ADÓSLEVELEK
ADÓSLEVÉL;ADÓSLEVELEKET
ADÓSLEVÉL;ADÓSLEVELEKRE
ADÓSLEVÉL;ADÓSLEVELET
ADÓSLEVÉL;ADÓSLEVÉLNEK
ADÓSLEVÉL;ADÓSLEVÉLRE
ADÓSLEVÉLRŐL;ADÓSLEVÉLRŐL
ADÓSSÁG;ADÓSSÁG

ADÓSSÁG;ADÓSSÁGA
ADÓSSÁG;ADÓSSÁGAIK
ADÓSSÁG;ADÓSSÁGAI
ADÓSSÁG;ADÓSSÁGÉRT
ADÓSSÁG;ADÓSSÁGNAK
ADÓSSÁG;ADÓSSÁGOK
ADÓSSÁG;ADÓSSÁGOKNAK
ADÓSSÁG;ADÓSSÁGOT
ADÓSTÁRS;ADÓSTÁRSOK
ADÓSTÁRS;ADÓSTÁRSKÉNT
ADOTT;ADOTTNAK
ADOTTSA;ADOTTSAI
ADOTTSA;ADOTTSAINÁL
ADÓZIK;ADÓZOTT
ADTA;ADTA
AFELŐL;AFELŐL
ÁG;ÁG
ÁG;ÁGAI
ÁG;ÁGAK
ÁG;ÁGAKAT
ÁG;ÁGÁN
ÁG;ÁGON
ÁG;ÁGRÓL
ÁGBELIBELI;ÁGBELI
AGGÁLY;AGGÁLYA
AGGÁLY;AGGÁLYT
ÁGI;ÁGI
ÁGY;ÁGYTÓL
AGYAG;AGYAG
AHÁNYADIK;AHÁNYADIK
AHHOZ;AHHOZ
AHOGYAN;AHOGYAN
AHOVÁ;AHOVÁ
AJÁNDÉK;AJÁNDÉK
AJÁNDÉK;AJÁNDÉKÁT
AJÁNDÉK;AJÁNDÉKBA
AJÁNDÉK;AJÁNDÉKNÁL
AJÁNDÉK;AJÁNDÉKOK
AJÁNDÉK;AJÁNDÉKOKAT
AJÁNDÉK;AJÁNDÉKOT
AJÁNDÉKOZ;AJÁNDÉKOZ
AJÁNDÉKOZ;AJÁNDÉKOZHAT
AJÁNDÉKOZÁS;AJÁNDÉKOZÁS
AJÁNDÉKOZÁS;AJÁNDÉKOZÁSAI
AJÁNDÉKOZÁS;AJÁNDÉKOZÁSÁRA
AJÁNDÉKOZÁS;AJÁNDÉKOZÁSI
AJÁNDÉKOZÁS;AJÁNDÉKOZÁSOKAT
AJÁNDÉKOZÁS;AJÁNDÉKOZÁSRA
AJÁNDÉKOZÁS;AJÁNDÉKOZÁSSAL
AJÁNDÉKOZÁS;AJÁNDÉKOZÁST
AJÁNDÉKOZÓ;AJÁNDÉKOZÓ
AJÁNDÉKOZÓ;AJÁNDÉKOZÓNAK
AJÁNDÉKOZÓ;AJÁNDÉKOZÓT
AJÁNDÉKOZÓ;AJÁNDÉKOZÓTÓL
AJÁNDÉKOZOTT;AJÁNDÉKOZOTT
AJÁNLI;AJÁNLI
AJÁNLI;AJÁNLIOTTA
AJÁNLI;AJÁNLIHAT
AJÁNLI;AJÁNLIJA
AJÁNLIÁS;AJÁNLIÁS
AJÁNLIÁS;AJÁNLIÁSTÓL
AJÁNLIAT;AJÁNLIAT
AJÁNLIAT;AJÁNLIATA
AJÁNLIAT;AJÁNLIATÁBAN
AJÁNLIAT;AJÁNLIATÁHOZ
AJÁNLIAT;AJÁNLIATÁNAK
AJÁNLIAT;AJÁNLIATÁT
AJÁNLIAT;AJÁNLIATÁTÓL
AJÁNLIAT;AJÁNLIATBAN

AJÁNLAT;AJÁNLATI
 AJÁNLAT;AJÁNLATNAK
 AJÁNLAT;AJÁNLATOK
 AJÁNLAT;AJÁNLATOT
 AJÁNLAT;AJÁNLATRA
 AJÁNLAT;AJÁNLATRÓL
 AJÁNLAT;AJÁNLATTAL
 AJÁNLAT;AJÁNLATTÓL
 AJÁNLATKÉRŐ;AJÁNLATKÉRŐ
 AJÁNLATSZÁM;AJÁNLATSZÁMBA
 AJÁNLATTÉTEL;AJÁNLATTÉTEL
 AJÁNLATTÉTEL;AJÁNLATTÉTELHEZ
 AJÁNLATTÉTEL;AJÁNLATTÉTELI
 AJÁNLATTÉTEL;AJÁNLATTÉTELRE
 AJÁNLATTEVŐ;AJÁNLATTEVŐ
 AJÁNLATTEVŐ;AJÁNLATTEVŐHÖZ
 AJÁNLATTEVŐ;AJÁNLATTEVŐK
 AJÁNLATTEVŐ;AJÁNLATTEVŐNEK
 AJÁNLATTEVŐ;AJÁNLATTEVŐVEL
 AJÁNLATTEVŐTŐL;AJÁNLATTEVŐTŐL
 AJÁNLKOZIK;AJÁNLKOZOTT
 AJÁNLÓ;AJÁNLÓJA
 AJÁNLÓ;AJÁNLÓJÁNAK
 AJÁNLÓ;AJÁNLÓT
 AJÁNLÓTT;AJÁNLÓTT
 AKAD;AKAD
 AKADÁLY;AKADÁLY
 AKADÁLY;AKADÁLYA
 AKADÁLY;AKADÁLYÁNAK
 AKADÁLY;AKADÁLYÁT
 AKADÁLY;AKADÁLYBA
 AKADÁLY;AKADÁLYRÓL
 AKADÁLY;AKADÁLYT
 AKADÁLYKÖZLÉS;AKADÁLYKÖZLÉSI
 AKADÁLYOZ;AKADÁLYOZNA
 AKADÁLYOZ;AKADÁLYOZNAK
 AKADÁLYOZ;AKADÁLYOZTA
 AKADÁLYOZ;AKADÁLYOZVA
 AKADÁLYOZ;AKADÁLYOZZA
 AKADÁLYOZ;AKADÁLYOZZÁK
 AKADÁLYOZÁS;AKADÁLYOZÁSA
 AKADÁLYOZÁS;AKADÁLYOZÁSÁÉRT
 AKADÁLYOZÁS;AKADÁLYOZÁSÁVAL
 AKADÁLYOZÓ;AKADÁLYOZÓ
 AKADÁLYOZOTT;AKADÁLYOZOTT
 AKADÁLYOZTAT;AKADÁLYOZTATVA
 AKADÁLYOZTATOTTSÁG;AKADÁLYOZTATOTTSÁGA
 AKAR;AKAR-E
 AKAR;AKARJA
 AKAR;AKARNA
 AKAR;AKART
 AKAR;AKARTA
 AKAR;AKARTÁK
 AKAR;AKART-E
 AKAR;AKAR
 AKARAT;AKARAT
 AKARAT;AKARATA
 AKARAT;AKARATÁBÓL
 AKARAT;AKARATÁHOZ
 AKARAT;AKARATÁN
 AKARAT;AKARATÁNAK
 AKARAT;AKARATÁRA
 AKARAT;AKARATÁT
 AKARAT;AKARATÁTÓL
 AKARAT;AKARATÁVAL
 AKARAT;AKARATI
 AKARAT;AKARATTAL
 AKARAT;AKARATUKAT
 AKARATBELIBELI;AKARATBELI

AKARATELHATÁROZÁS;AKARATELHATÁROZÁSÁRA
AKARATELHATÁROZÁS;AKARATELHATÁROZÁSRA
AKARATHIÁNY;AKARATHIÁNY
AKARATHIBA;AKARATHIBA
AKARATKIJELENTÉS;AKARATKIJELENTÉS
AKARATNYILATKOZAT;AKARATNYILATKOZAT
AKARATNYILATKOZAT;AKARATNYILATKOZATA
AKARATNYILVÁNÍTÁS;AKARATNYILVÁNÍTÁSÁT
AKÉP;AKÉPEN
AKI;AKIÉ
AKI;AKIHEZ
AKI;AKIT
AKI;AKIKET
AKI;AKIKKEL
AKI;AKIKNEK
AKI;AKIKRE
AKI;AKINEK
AKI;AKINÉL
AKI;AKIRE
AKI;AKIVEL
AKIK;AKIK
AKIRÓL;AKIRÓL
AKTÓL;AKTÓL
AKKÉNT;AKKÉNT
AKKORI;AKKORI
AKKORTÓL;AKKORTÓL
AKTUS;AKTUSA
ALÁBB;ALÁBB
ALÁBBI;ALÁBBI
ALACSONY;ALACSONYABB
ALACSONY;ALACSONYABBAN
ALACSONY;LEGALACSONYABB
ALÁÍR;ALÁÍRHATJA
ALÁÍR;ALÁÍRJA
ALÁÍR;ALÁÍRJÁK
ALÁÍR;ALÁÍRNI
ALÁÍR;ALÁÍRNIA
ALÁÍR;ALÁÍRNIUK
ALÁÍR;ALÁÍRTA
ALÁÍRÁS;ALÁÍRÁS
ALÁÍRÁS;ALÁÍRÁSA
ALÁÍRÁS;ALÁÍRÁSÁNAK
ALÁÍRÁS;ALÁÍRÁSÁRA
ALÁÍRÁS;ALÁÍRÁSÁT
ALÁÍRÁS;ALÁÍRÁSÁVAL
ALÁÍRÁS;ALÁÍRÁSSAL
ALÁÍRÁS;ALÁÍRÁST
ALÁÍRÁS;ALÁÍRÁSUKKAL
ALÁÍRÓ;ALÁÍRÓJÁT
ALÁÍRT;ALÁÍRT