

A kvantitatív szövegelemzés mint előrejelzési lehetőség - Pénzügyi Szemle folyóirat

2014. június 02. 10:30

A kvantitatív szövegelemzés mint előrejelzési lehetőség

Rendkívül fontos, hogy ki alakítja az ún. „uralkodó diskurzust”, akár a politikában, akár az élet más területén. Miről és hogy lehet beszélni? Ki határozza meg a fő témát, a közbeszéd főirányát? Ki az, aki irányít, és ki az (melyik politikai erő), aki csak reagál? A szövegek elemzése leleplezi az uralkodó diskurzust, azonosítása és elemzése révén a társadalom, a hatalom működési mechanizmusairól tudhatunk meg többet. Közgazdász kutatók egy csoportja a fenti kérdésekre keresi a választ a társadalomtudományokban használt kvantitatív szövegelemzési eljárások segítségével.

Birher Nándor - Boros István - Kocziha Mihály - Kolozsi Pál - Stummer Attila - Sztráray-Kézdy Éva

Kvantitatív szövegelemzést jogszabályi környezetben, a társadalmi diskurzus, jogalkotás, jogalkalmazás felosztás szerint (normatív szövegek életciklusa) - a kapcsolati mintákat is figyelembe véve - először jelen tanulmány szerzői végeztek el (1). A társadalomtudományokban használt kvantitatív szövegelemzési eljárások a különböző közleményekben, szövegekben előforduló szavak, kifejezések, esetleg témák gyakoriságát, ill. együttes megjelenését és egymáshoz való viszonyát elemzik. Ennek során a ténylegesen megjelenő manifeszt tartalmakon túl, a nyíltan ki nem mondott, csak az üzenet megszerkesztettségéből kiolvasható látens tartalmakat is vizsgálják. A módszer a gyakorlatban a szöveg megfigyelési egységeinek előzetesen megállapított kategóriák szerinti, kategóriaszótár segítségével történő kódolását, majd mennyiségi elemzését jelenti. (2)

A komplex vizsgálat leírása

A tisztán nyelvészeti megközelítéssel szemben a társadalomtudományos kutatásokban a szövegek elemzésével nem pusztán az adott közleményről szeretnénk valamit megtudni, hanem a szövegkorpusz társadalmi kontextusában történő tartalmi és formai vizsgálatával bizonyos társadalmi jelenségeket szeretnénk jobban megérteni (3).

A jogalkotás, a jogszabályok valamint a társadalmi közbeszéd szövegelemzési módszerrel történő vizsgálata során abból a filozófiai és jogszociológia elméletből indulunk ki, mely szerint a jogalkotás és a társadalmi folyamatok egymással kölcsönhatásban állnak, másrészt a különböző nyilvánosságot kapó szövegek, közlemények részben tükrözik a társadalmi valóságot, részben maguk is részt vesznek a társadalmi jelenségek alakításában.

A kutatás során világossá vált, hogy a társadalmi diskurzus nyomon követhető a normatív szövegekben is (4). Egyértelmű összefüggés van tehát aközött, amiről a parlamenti vita folyik, ami jogszabályszöveg lesz, és amit a bírói gyakorlat mond. Az összefüggés azonban nem jelent egyértelmű korrelációt. Gyakran csak távoli a kapcsolat azok között a kulcsszavak között, amelyeket a bírói gyakorlat használ, és amelyek a jogszabályok korpuszában megjelenik. Pontosan azért, hogy jobban megértsük a viszonyt a közbeszéd és a normatív szövegek előállítása, ill. használata között, szükséges a kvantitatív tartalomelemzés ezeken a területeken.

Azon túlmenően, hogy ennek jelentős jogelméleti értéke lehet, a módszer alkalmas arra is, hogy előre jelezzen olyan társadalmi folyamatokat, amelyek a jogalkotást is érintik, vagy fordítva, a jogalkotásból kiindulva jósolni lehessen a különböző témák kapcsán feléledő társadalmi diskurzust. Ugyanis a szövegelemzési eljárások egyik fontos funkciója, azon túl, hogy - sokszor más módszerrel nem is elemezhető módon - a jelen állapotról tegyen megállapításokat, a szövegben előforduló szabályszerűségek alapján előre tudjon jelezni bizonyos társadalmi jelenségeket, trendeket.

Módszerünk, bár tisztán kvantitatív jellegű, megközelítésünkben felhasználtunk egy kvalitatív szövegelemzési módszert, a kritikai diskurzuselemzést, ill. elsősorban annak filozófiai, elméleti hátterét.

„A módszernek (critical discourse analysis - CDA) célja, hogy a szöveg elmélyült formai és tartalmi elemzése során kapott eredményeket összefüggésbe állítsa a szöveg kontextusát adó társadalmi, kulturális, politikai rendszerrel. Bár szövegelemzési módszer, végső célja nem az, hogy elsősorban a szövegről, hanem - a szöveg által - a társadalomról, annak működési mechanizmusairól, ezen belül is a társadalomban működő hatalmi viszonyokról tudjunk meg többet (5). A módszer kiindulópontja szerint a szöveg egyrészt leképeződése annak a társadalomnak, amelyben létrejött, másrészt a szöveg is aktívan formálja, alakítja ezt a társadalmi kontextust.

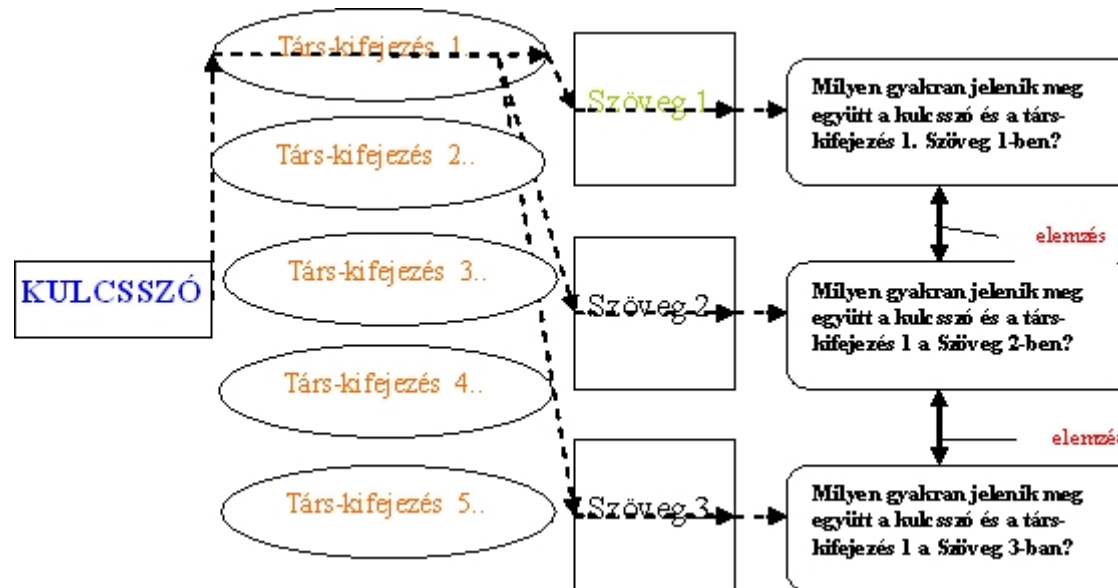
A módszer filozófiai hátteréül szolgált Michel Foucault, francia posztmodern filozófus elmélete, mely szerint korszakunkban a hatalom gyakorlásának legfőbb eszköze a nyelv (szemben korábbi korszakokban uralkodó fizikai erővel). Így rendkívül fontossá vált, - főleg a tömegkommunikáció elterjedésével -, hogy ki alakítja az ún. „uralkodó diskurzust”, akár a politikában, akár az élet más területén. Miről és hogy lehet beszélni? Ki határozza meg a fő témát, a közbeszéd főirányát? Ki az, aki irányít, és ki az (melyik politikai erő), aki csak reagál? A szövegek elemzése leleplezi az uralkodó diskurzust, azonosítása és elemzése révén a társadalom, a hatalom működési mechanizmusairól tudhatunk meg többet.” A későbbiekben érdemes lenne a mélyebb összefüggések feltárása érdekében a kvantitatív elemzések mellett a kritikai diskurzuselemzés elméleti kiindulópontján túl, a módszer gyakorlatát is bevonni az elemzésbe.

Visszatérve az alkalmazott kvantitatív módszerhez, elemzésünk többlet tartalma az, hogy nem egyszerűen a közbeszédet, hanem a cselekvéseket közvetlenül alakító normatív tartalmat alakító szövegeket is elemezzük a parlamenti naplókban.

A kifejlesztett módszertan alapján lehetőség van:

- nagy tömegű,
- szöveges adatbázisban,
- szavak és szókapcsolatok előfordulási gyakoriságának mérésére, illetve ennek elemzésére.

Módszerünk szerint meg kell határozni (1) a kulcsszót, (2) a társ-kifejezéseket és (3) rendelkezni kell megfelelő formában az elemzendő szövegekkel is. A kutatás újdonsága, hogy a normatív szövegek életciklusai szerinti kapcsolati hálót rajzol az egyes társ-kifejezések között. Ez a kapcsolati háló mutatja a társ-kifejezések normatív életciklusban való elhelyezkedését. Ennek a folyamata látható az alábbi ábrán.



Ahhoz, hogy a Codex (6) különböző területeinek szövegei között kapcsolati hálót (mintát) tudjunk leírni, szükséges az egyes területek, úgymint a jogalkotás előkészítése (parlamentari vita), a jogalkotás és a joggyakorlat (bírói határozatok) szövegeinek elemzése, az abban szereplő kulcsszavak azonosítása. Ennek a műveletnek az elvégzésében az Atlas Ti. program (7) nyújt segítséget.

Munkánkban követjük a Klaus Krippendorff által meghatározott módszertant (8), amely a következő hat lépésből áll:

1. Adatkészítés: a szöveg kijelölése, a normatív tartalmú textus meghatározása (időbeli, hatálybeli szűkítés, címekre való szűkítés etc.).
2. Mintavételi egység meghatározás: ilyen egységek lehetnek például a jogszabály-előkészítés (parlamentari naplók, parlamentari bizottsági jegyzőkönyvek), jogszabályok, bírói határozatok szövegei.
3. Mintavétel: a meghatározott szűkítések szerinti keresés, a releváns szövegek összegyűjtése.
4. Kódolás és kategóriák: A kiválasztott minta adatrögzítési egységeit (9) meghatározott kategóriákba soroljuk. A kategóriák egy előzetes hipotézis szerint előre meghatározottak, de a kutatás közben is bővíthetők, alakíthatók.
5. Adatredukció és kapcsolati rendszerek felvázolása: Az eredeti adatok csoportosítása és kapcsolati rendszerük felvázolása, ez utóbbi tevékenység saját módszertan alapján történik.
6. Elemzés, következtetés.

A módszer alkalmazásával meg lehet mondani, hogy egyes szavak milyen gyakorisággal fordulnak elő más szavak környezetében, azaz egyes kifejezések milyen kontextusban fordulnak elő adott szövegekben. A módszer lehetővé teszi a szövegelemzésen túlmenően a hálózati kapcsolati vonatkozások feltérképezését is, a kinyert adatok alapján a szavak között ugyanis kapcsolati hálókat lehet felvázolni, ami utal a szó „centralitására (10)” (mennyire központi helyet foglal el, azaz mennyire igaz, hogy áthatja a joganyagot).

A módszer azért hasznos, mert elemezhetővé teszi, hogy bizonyos kiemelt fogalmaknak mi a tényleges relevanciája egyes témákkal kapcsolatban a különböző vizsgált szinteken. Egy-egy témának ugyanis sokféle aspektusa van, de a módszerrel ezek közül ki lehet szűrni, hogy melyek a valóban relevánsak a különböző szinteken. Ennek kiemelt jelentősége lehet a munkánk hasznosulása szempontjából, hiszen akkor várhatunk magas fokú hasznosulást, ha társadalmilag releváns kérdésekre koncentrálnunk. A társadalmilag releváns kérdések esetében pedig reális feltételezés, hogy átlagot meghaladó mértékben beszélnek és írnak róla.

A módszer arra is lehetőséget teremt, hogy összevessük pl. a társadalmi közbeszéd és a jogalkotás témáit, azaz elemezhetjük, hogy a jogalkotás valóban azokkal a kérdésekkel foglalkozik-e, amivel a társadalom szélesebb rétegei. Természetesen mód van az egyes témák kontextusbeli változásának időbeli nyomon követésére is (mi volt fontos régebben, és mi fontos ma?).

Kiemelten elemezhető, hogy mivel foglalkoznak az országgyűlési képviselők, milyen témák jönnek elő a parlamenti vitákban. Rendelkezünk, illetve igény esetén rendelkezhetünk az országgyűlési viták jegyzőkönyvével, de ezek elemzése a terjedelem miatt egyszerű átolvasással szinte megoldhatatlan. Megvizsgálható, hogy milyen kérdések vannak egyes témákkal kapcsolatban a módosító indítványok fókuszában. Elvileg lehetőség van a szakirodalom elemzésére is, azaz hogy megnézzük, egy-egy témával kapcsolatban mely területeket kutatják a szakemberek, és vajon ezzel a kérdéssel foglalkozik-e a jogalkotás is?

Egy konkrét példa

Birher Nándor kutatása során többek között elemezte, hogy a „roma” és a „cigány” szavak milyen egyéb szavakkal párban szerepelnek a (1) jogszabály-alkotási dokumentumokban, azaz elsősorban a parlamenti viták jegyzőkönyveiben, (2) a különböző szintű jogszabályokban, illetve (3) a bírósági határozatokban (normatív életciklus). Azt vizsgálta, hogy három időpontban hatályos szövegekben (1998, 2010, 2013) milyen gyakorisággal fordulnak elő a „roma” és „cigány” szavak környezetében a „munka”, a „foglalkoztatás”, az „alkalmazás”, az „oktatás”, a „képzés”, illetve az „egészség” szavak vizsgálatával:

- be tudta mutatni, hogy bizonyos kifejezések miképp terjedtek el a cigánysággal foglalkozó joganyagokban, hogyan kerültek a perifériáról a centrumba (mikor kerültek be, honnan, milyen közvetítőkön keresztül kerültek be a joganyagba);
- követhető, hogy az idő előrehaladtával mely kifejezések váltak hangsúlyossá, és melyek nem (például a cigánysággal kapcsolatos szövegekben nem szerepel az alkotmányosság és az egészség kérdése, pedig a közbeszéd szerint ennek lehetne relevanciája);
- be tudta mutatni, hogy vannak-e olyan kifejezések, amelyek a vitákban (társadalmi közbeszédben), vagy akár a bíróságokban előjönnek, de azokkal a joganyag nem foglalkozik igazán (cigányság és egészség pl.)

A komplex vizsgálat része, a kvantitatív szövegelemzés

Ahogy azt korábban már jeleztük, a teljes Codex corpuszának vizsgálata előtt el kell végezni az egyes területek önálló kvantitatív szövegelemzését, azaz meg kell találni a kulcsszavakat, a kapcsolódó társkifejezéseket az egyes, relevánsnak tekintett szövegekben. Ennek a műveletnek az elvégzése a szógyakoriság

mérésre, ill. az egyes szavak előfordulása közti korreláció elemzésre koncentrálnak. A kvantitatív szövegelemzés mint módszer alkalmas arra, hogy a parlamenti jegyzőkönyvek alapján előre jelezzen olyan kockázatokat, amelyek jó eséllyel fognak megjelenni a társadalmi diskurzusban is.

A szövegelemzési módszerek alkalmazása a társadalomtudományok körében az utóbbi évtizedekben egyre jobban terjed, a tudományos élet fő irányához tartozik, kreativitása és innovatív jellege folytán egyre bővülő kutatási kérdések megválaszolására használják. Az elemzés bevezetése jelentősen növeli a parlamenti folyamatok nyomon követésének hatékonyságát azon szervezeteknél, amelyek jogalkotást és joggyakorlatot érintő különböző ellenőrzésre, vizsgálatokra kötelezettek. (Folyamatosan és objektíven követhetővé válik, hogy az ellenőrzést végző számára meghatározott feladatok hogyan jelennek meg a „parlamenti beszédben”, a jegyzőkönyvekben).

A korábbi vizsgálatok alapján világossá vált, hogy a jogalkotás során a társadalomban tematizált események megjelennek a jogszabály előkészítésben, a jogalkotás különböző szintjein, illetve a joggyakorlatban. Az események nevesítésének változásait (szavak előfordulási gyakorisága és helye, normatív életciklus) meghatározza a szó jelentőségét, illetve az időbeli vizsgálat megmutatja, hogy ez a jelentőség a vizsgálat szempontjából növekvő, vagy csökkenő-e?

A vizsgálat elvégzésének az alábbi aspektusait vehetjük figyelembe:

1, A parlamenti viták (országgyűlési napló), mint a politikai közbeszéd szintere időrendi vizsgálata kulcsszavak előfordulási gyakoriság szerint. A kulcsszavak előfordulási gyakoriságának vizsgálata automatizálható. A komplex elemzés elvégzéséhez a későbbiekben a kulcsszavak vizsgálhatók a bizottsági jegyzőkönyvekben, jogszabályokban és bírósági határozatokban is. A kulcsszavak előfordulási gyakoriságának időbeli változása egyszerűen mérhető. A kulcsszavakat a korábbi más jellegű tapasztalatok alapján célszerű meghatározni.

2, Fontos az új kulcsszavak bevezetése. Ezeket a korábban, az 1. pont szerint elvégzett vizsgálatokban kapott fogalmak asszociációs terének elemzésével kaphatjuk meg.

3, Az egyes kulcsszavakat hozzá lehet rendelni az azokat használó személyekhez.

4, Következő lépésben, immár a komplex elemzés részeként le lehet írni a szavak kapcsolati hálójának dinamikáját, meg kell vizsgálni, hogy a szavak helye a kapcsolati hálóban hogyan változik.

A kulcsszavak kijelölésének módszertana az alábbi lépésekből áll:

(a) meg kell határozni azokat a kifejezéseket, amelyek a működés jogszerűségét/jogszerűtlenségét érintik, mint például: hatékony, átlátható, számon kérhető, közpénz, mutyi, csalás, bűn, haver, közbeszerzés, ...

(b) meg kell határozni azokat a kifejezéseket, amelyeknél az ellenőrzés jogszabály erejénél fogva követelmény (társkifejezés) mint például: önkormányzat, oktatás, etc....

(c) meg kell vizsgálni, hogy a két fogalmi kör között milyen korreláció van, ez hogyan változik az időben, akár a jövőre nézve is (forecast).

A vizsgálat elvégzését követően várható eredmények: a társadalom különböző diskurzusai által tematizált tartalmak jelzik a társadalomban megjelenő

feszültségeket, fókuszokat. Az így azonosított területek lehetnek a későbbi részletes vizsgálatok célzott terepei.

A szövegelemzés leginkább a széles területet lefedő, jellemzően évi gyakorisággal végzett ellenőrzések fókuszálásában lehet alkalmas eszköz. Ezzel biztosítható, hogy a fókuszok ne esetlegesen legyenek, hanem módszertanilag megalapozottabbá váljanak, illetve hogy az időről időre megújuló témacsoportos ellenőrzések fókusza illeszkedjen a társadalmi érdeklődéshez.

A szövegelemzéssel tervezési szempontból befoghatóvá, elemezhetővé válnak mindenképp az országgyűlési viták és a jogalkotás, illetve a bírósági gyakorlat szövegei, amelyek eddig is rendelkezésre álltak, de méretük, kiterjedésük, nagyságuk miatt nem kerültek részletes elemzésre.

Összegezve megállapíthatjuk, hogy az itt felvázolt módszer tisztán kvantitatív elemzési része már önmagában is komoly lehetőségeket rejt magában. Alkalmas arra, hogy az egyes országgyűlési, vagy bizottsági ülések anyagát folyamatosan, objektíven elemezze. Az adatbázisok rendelkezésre állása lehetővé teszi, hogy ezt a munkát költséghatékonyan lehessen végezni. További lépésként pedig az imént leírt komplex elemzés is elvégezhető, amely alkalmas a normák érvényesülésének vizsgálatára, a társadalmi diskurzus empirikus leírására a normatív tartalmú kapcsolati minták felvázolásán keresztül.

Jegyzetek:

(1) Birher N., Bene D., Rendszerhibák és javításuk, in Szak- és Felnőttképzés 2013. 11-12 szám, 16-20.

(2) Antal László: A tartalomelemzés alapjai, Bp. 1976.

(3) Kmetty Zoltán - Sztárayné Kézdy Éva: [A tartalomelemzés. In: Kutatásmódszertan.](#)

(4) Diskurzus alatt a Michel Foucault, francia filozófus által bevezetett fogalmat értjük, mely egy intézményesült gondolkodási módot jelent. „A diskurzus látszólag talán jelentéktelen dolog, de jellemző tilalmai hamar felfedik kapcsolatát a vággyal és a hatalommal. [...] A történelem szüntelenül arra oktat minket, hogy a diskurzus nemcsak egyszerűen tolmácsolja a küzdelmeket és az uralmi rendszereket, hanem érte folyik a harc, általa dúl a küzdelem; tehát a diskurzus az a hatalom, amelyet az emberek igyekeznek megkaparintani.” (Foucault, 1998. 51.o.)

(5) Kmetty Zoltán - Sztárayné Kézdy Éva: [A tartalomelemzés. In: Kutatásmódszertan.](#)

(6) A Codex a lex validus és a jus vivens összessége, a hatályos jog, a bírói jog, a gyakorlat szerződési joga együtt. Összességében elmondható, hogy arról a tényleges szabályhalmazról van szó, amely a valóságban érvényesülhet.

(7) www.atlasti.com

(8) Krippendorff Klaus: A tartalomelemzés módszertanának alapjai, Budapest, 1996

(9) A mintavételi egység és az adatrögzítési egység nem ugyanaz. Mintavételi egység lehet pl. a parlamenti napló, azonban az adatrögzítési egységek pl. az abban szereplő szavak.

(10) A centralitás fogalma a hálózati csomópontokkal kapcsolatban is többféle lehet, megkülönböztethetjük többek között a közelségi centralitás, közbülsőségi centralitás, sajátérték centralitás, fokszám centralitás fogalmait. A kutatás jelenlegi stádiumában ezek a számítások egyelőre nem relevánsak, pontos használatuk a nagyobb halmazokon végzett mérések értékelését követően lesz lehetséges.

Irodalom:

Antal László: A tartalomelemzés alapjai. Magvető, Budapest, 1976.

Fairclough, Norman: [Language and Power. Longman, London, 1989.](#)

Foucault, Michel: A diskurzus rendje. In: Foucault, Michel: A fantasztikus könyvtár. Vál. és ford.: Romhányi Török Gábor. Pallas Stúdió - Attraktor Kft., Budapest, 1998.: 50-74.

Foucault, Michel: The Archeology of Knowledge and the Discourse on Language. New York, Pantheon, 1972.

Kmetty Zoltán - Sztárayné Kézdy Éva: [A tartalomelemzés. In: Kutatásmódszertan.](#)

Krippendorff Klaus: A tartalomelemzés módszertanának alapjai. Balassi Kiadó, Budapest, 1996.

Wodak, Ruth et al.: The Discursive Construction of National Identity. Edingburgh University Press, Edingburgh, 1999.

Block, N.: The Mind as the Software of the Brain. In Smith, D. and Osherson, E. (eds.): Invitation to Cognitive Science: Thinking. 2nd ed. MA: MIT Press, Cambridge, 1995.

Bays, T.: On Floyd and Putnam on Wittgenstein on Gödel In: The Journal of Philosophy 101(4) 2004.: 197-210.