

Idői struktúrák feltárása kvalitatív és kvantitatív szövegelemzéssel

Huszár Zsuzsanna¹, Dr. Sramó András²

¹PTE BTK Tanárképző Intézet, 7624 Pécs, Ifjúság útja. 6.

huszped@tki.pte.hu

²PTE IGYFK Gazdaságtudományi Intézet, 7100 Szekszárd, Rákóczi út 1.

sramo@igyfk.pte.hu

Kivonat. Az előadás megkísérli egy PhD kutatás kezdeti lépéseit felvázolni a számítógépes nyelvészet kvantitatív technikáinak és a kvalitatív tartalom-elemzési adatok számítógépes feldolgozási technikáinak ötvöztetésével, a tartalomelemzés szempontjainak bemutatásával és a szóstatistikai vizsgálatok néhány eredményének közzétételével a 10-16 éves korosztály iskolai fogalmazásainak 600 ezer szavas reprezentatív mintájának bázisán.

Kulcsfogalmak: szövegelemzés, szóstatistika, Zipf-törvény, kvalitatív tartalomelemzés, idői struktúra

1 Bevezetés

A folyamatban lévő PhD-kutatás, amelyről előadásunk szól, két doktori programhoz is szoros szálakkal kötődik: egyrészt a Szegedi Tudományegyetem korábbi Neveléstudományi Doktori Programjához, s ennek Cs. Czachesz Erzsébet által irányított, a kommunikatív kompetencia fejlődésével és fejlesztésével foglalkozó alprogramjához, másrészt a Pécsi Tudományegyetem Alkalmazott Nyelvészeti Doktori Iskolájához.

A kutatás szövegbázisa a 10-16 éves korosztály iskolai fogalmazásainak a Szegedi Tudományegyetem kutatásai számára rögzített, 2170 fogalmazásból álló, mintegy 600 ezer szavas, országos, reprezentatív mintája.

2. A kutatási elképzelések leírása

A kutatás az idő, ill. az idői struktúrák nyelvi megjelenésére, nyelvi létezés módjára irányul. Vizsgálatunkban a fizikai időfogalommal szemben az idő társadalmi vetülete (tempus) kerül előtérbe. Alapkérdésünk, hogy hogyan bukkan elő, és milyen kontextusba ágyazottan az idő, illetve az időiség a gyermek-fogalmazásokban, hogyan ragadható meg strukturáló ereje a gondolkodásra nézve. A kutatás célja az időiség nyelvi megjelenésének komplex vizsgálata, s az elemzési célokkal adekvát

kritériumrendszer és alkalmas számítógépes szövegfeldolgozó eljárás kidolgozása, ennek segítségével egy ellenőrizhető fogalmazástipológia kialakítása, majd ezen tipológia összevetése a kutatás független változóival, végül az eredmények interpretálása a neveléstudomány és az alkalmazott nyelvészet számára.

Független változóink: a településtípus, az iskolatípus, az évfolyam, a nem valamint a fogalmazások típusa, érvelő vagy elbeszélő jellege. Független változók a kvantitatív és kvalitatív szövegelemzés által feltárt szövegjellemzők. Jelen előadásunkban a kvantitatív technikák közül a szóstatistikát, a kvalitatív technikák közül pedig a tartalomelemzést emeljük ki, utalva a tartalomelemzés számítógépes lehetőségeire [4] is. A tartalomelemzés hazai irodalmában az idő önálló szempontként mint a narratív pszichológia kompozíciós elve nyer kifejtést [2].

A készülő disszertáció pedagógiai szempontból is értelmezhető diagnózis lenne arról, hogy miféle idői perspektívában gondolkoznak a különböző életkorú tanulók, és arról, hogy megjelenik-e, s hogyan jelenik meg iskolai fogalmazásaikban, érvrendszerük mögöttesében a gondolkodás távlatossága. A megismeréstudomány és a nyelvészet közös metszetében elemezhető volna, hogy van-e összefüggés a nyelv absztrakciós szintje és a gondolkodás távlatossága között.

2.1 Előfeltevések

Az időnek strukturáló szerepe van. Kulturális szempontból az idői struktúrák lényegiek. A grammatikai struktúrákon túl, illetve azoktól némiképp elkülöníthetően egy szöveg idői struktúrája mint a múlt, jelenre, jövőre vonatkozó tartalmak megkülönböztetéséből és arányaiból adódó struktúra is értelmezhető és elemezhető. Az időhöz kötöttség nem feltétlenül szándékolt eleme egy szövegnek, amennyiben a szöveg idővonatkozásai, időarányai nem előre eltervezettek. Az egyes fogalmazásokon túl a szöveg-egészben megjelenik valamiféle időszemlélet, amely az idővel való társadalmi bánásmódként válik azonosíthatóvá. A szövegek idői struktúrája a fenti értelemben nem más, mint az időhöz való társadalmi viszony kifejeződése.

2.2 Alaphipotézisek

Alaphipotézisünk szerint a nyelv univerzális jellege az általunk vizsgált szövegmintában is tükröződik. Az univerzalitást mint a Zipf-törvény [5] érvényesülését vizsgáljuk előadásunkban. Feltevésünk szerint a fogalmazások idői struktúrái a társadalmi szinten szabályozott "racionális" időfelhasználást tükrözik, vagyis az időnek instrumentális jelentősége van., s az idő lineáris ill. logaritmus szemlélete domináns. Feltételezzük továbbá, hogy az időszemlélet nyelvi megragadása világképre utaló jelentőséggel bír.

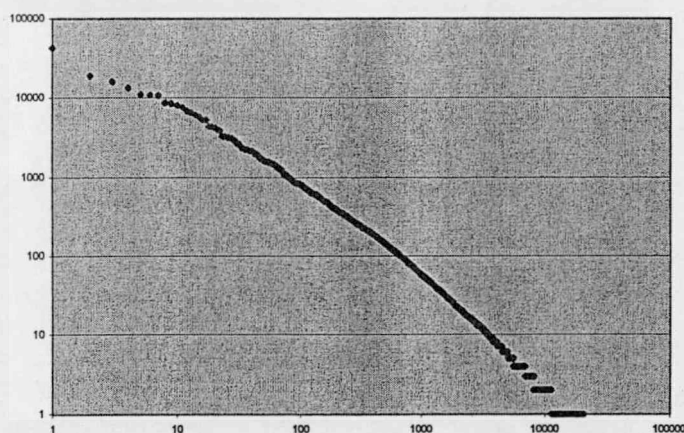
2.3 A kvalitatív elemzés tartalmi szempontjai

1. Előremutató és visszautaló idői perspektíva és ezek aránya a szövegekben.
2. A szövegben foglalt tartalmak átfogó idői tartománya.

3. A múlthoz, a jelenhez, és a jövőhöz kapcsolódó kijelentések aránya.
4. Múlthoz, jelenhez, jövőhöz kapcsolódó kijelentések értékelítettsége: pozitív, negatív, semleges tartalma. Az átértékelés időhöz kötött mozzanatai és jellegzetes tartalmai.
5. A pozitív, negatív és semleges tartalmakhoz kapcsolódó földrajzi, topológiai kategóriák; az értéktartalmak s a térbeli viszonyok egymásra vetítése.
6. Szubjektív és objektív időészlelés. A szubjektív időélmény jellemzői. Explicit hiányérzetek: mire nincs idő?
7. A fogalmazások alapritmusát kijelölő időbeli viszonyok, napi, heti, éves stb. ritmusok megjelenési gyakorisága.
8. Az iskolához és az iskolában töltött időhöz valamint a feltételezett tanulói szerepelvárásokhoz való érzelmi viszony.
9. Az iskola funkciói a tanulók deklarációiban.
10. Az érvek forrása; írásbeliségre és szójhagyományra hivatkozó utalások aránya.
11. A személyes hivatkozási rendszer időbeli jellemzői: intragenerációs, intergenerációs, többgenerációs utalások és ezek aránya. Az eleven emlékezet terjedelme.

3. Kezdeti eredmények

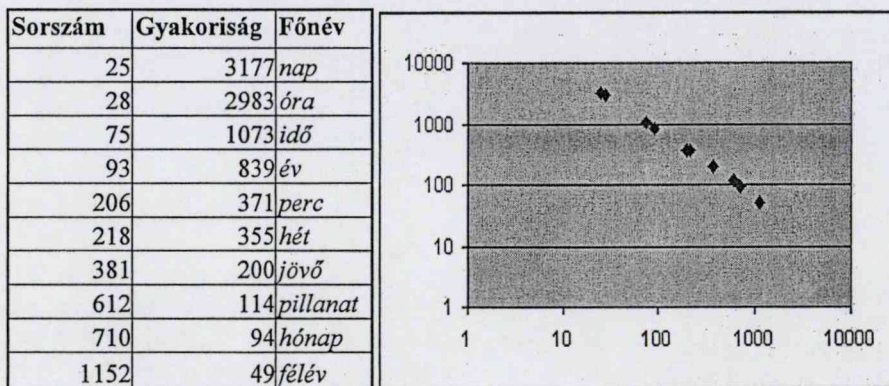
Az említett korpusz önálló szóstatistikai feldolgozását nem végeztük el, mivel rendelkezésünkre állt a korpusz gyakorisági szótára [1]. Ennek felhasználásával azt vizsgáltuk, hogy az időre vonatkozó, az időiséget kifejező, adott szófajú szavak hol helyezkednek el a szógyakoriságok által kijelölt sorrendben. Eredményeinket a Zipf-törvény segítségével mutatjuk be.



1. ábra. A teljes mintában előforduló különböző szavak gyakorisága, pozíciója a gyakorisági sorrendben (20 300 szó)

A Zipf-törvény alkalmazásával kapcsolatban meg kell említenünk, hogy a korpusz szószáma (601 135) lényegesen meghaladta az általunk ismert minták szószámát, és így egy analitikusan "szépen" viselkedő görbét kaptunk, amely a középső szakaszán jól közelíthető egy loglineáris egyenessel. A pontdiagram egyértelműen mutatja a törvény érvényesülését a vizsgált mintára, vagyis annak az univerzális szabályszerűségnek a megjelenését, amely szerint egy adott mintában az egyes szavakhoz tartozó előfordulási gyakoriságok és az ezáltal kijelölt sorrendi pozíciójuk, sorszámuk szorzata megközelítőleg stabil. Ezt a szabályszerűséget mint univerzális sajátosságot tartja számon a szakirodalom, s ennek alapján kimondható, hogy az iskolai fogalmazások is hordozzák ezen univerzális jellegzetességet.

A következőkben a teljes anyag leggyakoribb időre utaló főneveinek felsorolását adjuk meg gyakoriságuk és gyakorisági sorrendjük feltüntetésével



2. ábra. A leggyakoribb időre vonatkozó főnevek gyakorisága és előfordulási sorrendi pozíciója a teljes mintában (10 szó).

A leggyakoribb idői lépték, konkrét időtartam, amely a fogalmazásokban feltűnik, a *nap*, ezt követi az *óra*. A *nap* és az *óra* a teljes korpusz leggyakoribb főnevei között a második és a harmadik helyen szerepel, ez is megerősíti, hogy mint a fogalmazások meghatározó időegységeit tartjuk számon őket.

Az idői kifejezések gyakoriságát mint a korosztály időszemléletére utaló egyik lehetséges utalást tekintjük, s ennek alapján valószínűsítjük, hogy a korosztály gondolkodásmódjában a jelenorientáltság dominál.

A teljes mintában megfigyelt szabályszerűség az időhatározók tekintetében is érvényesül. Az időhatározók sorában kiugró gyakorisággal mutatkozik a *mindig* kifejezés (1669 említés), ami feltevésünk szerint mint állandóság és rendszeresség akár az iskolás életforma jelzője is lehet. Ezzel az uralkodó folyamatossággal szemben a végességre és az időtlenségre utaló kifejezések (*utoljára*, *örökre*) feltűnően kis (23-22) gyakorisággal, a többi időhatározótól erősen leszakadva jelennek csak meg. Talán mert a metafizikai dimenziókra való nyitottság nem lehet, s nem is jellemzője ennek a korosztálynak.

4. Egy konkrét példa

Bár jelen írás keretei között a tartalomelemzés osztályozási rendszerének, s a kialakított kategóriáknak és kódolási technikának a bemutatására nem vállalkozunk, előrebocsátjuk, hogy elemzésünk során lépcsőzetes osztályozási rendszert használunk, és Vieikko Pietila [6] nyomán tárgyi és viszonyulási tartalmakat különböztetünk meg. Végezetül egy konkrét fogalmazás szöveges jellemzését kíséreljük meg.

"Szeretek iskolába járni, mert sok mindent meg lehet tanulni a felnőttektől, amit magamtól nem tudnék megérteni." (Fogalmazáskód: 8552107418)

A fogalmazás egyetlen összetett mondat, három tagmondatból áll. Az elemzés alapjául az ebben foglalt négy kijelentési egység szolgál. A felütés pozitív, és miután a szöveg egyetlen mondat, a felütés gyakorlatilag a konklúzióval azonos. Az iskolában megélt idő értékelése pozitív, deklaráltan csak pozitívum jelenik meg, viszont a megoldás terjedelme, a feladathoz való közömbös esteleg negatív viszonyra utal. Az érvelés "minimalista", a feladat formális teljesítése a felkínált tanulószerep hártására utal. Az érvelésmód racionális, a szerkezet logikai, ok-okozati viszonyt mutat. A szöveg tartalmi hangsúlya a tanulásra esik. Az iskola funkciójára vonatkozóan kognitív szempontok hangsúlyosak; a tanulás, a tudásátadás és a megértés. A szövegnek nincs kitéüntetett időbeli kezdőpontja, sem pedig kitéüntetett végpontja, átfogó idői tartománya nem explicit, hanem az iskola intézményéhez való egyéni kötődés időszakaként értelmezhető, valószínűsíthető. A fogalmazás meghatározó ritmusát az iskolába járás (napi) szokása és a felnőtt-gyerek viszony „generációs képlete” adja. Idői perspektíváját tekintve inkább jelenorientált, de utal a közeljövőre is.

Hasonlóképpen leírhatóak a többi fogalmazás jellemzői is. Ezen jellemzők kódolása, digitalizálása, táblázatba rendezése lehetővé teszi a kvalitatív elemzéssel kapott adatok kvantitatív vizsgálatát. Egy ellenőrizhető fogalmazástípológia kialakításához a továbbiakban clusteranalízis és a faktoranalízis kínálkozik célravezető eljárásnak.

Hivatkozások

1. Cs. Czachesz Erzsébet – Csirik János: 10-16 éves tanulók írásbeli szókinccsének gyakorisági szótára. BIP, 2002.
2. Ehmann Bea: A szöveg mélyén. A pszichológiai tartalomelemzés. Új Mandátum, Budapest, 2002.
3. Kornai, András: Zipf's law outside the middle range. In: Proc. Sixth Meeting on Mathematics of Language. University of Central Florida, 1999. 347-356. o.
4. Krippendorff, Klaus: A tartalomelemzés módszertanának alapjai. Balassi Kiadó, Budapest, 1995.
5. Manning, Christopher D. – Schütze, Hinrich: Foundations of Statistical Natural Language Processing. The MIT press, Cambridge, 1999.
6. Pietila, Veikko: Tartalomelemzés. Tömegkommunikációs Kutatóközpont, Budapest, 1979.
7. Powers, David M. W.: Applications and Explanations of Zipf's Law. In D. M. W. Powers (ed.) NeMLaP3/CoNLL98: New Methods in Language processing and Computational Natural Language Learning, ACL, 1998. 151-160. o.