

## A PEDAGÓGIAI SZÖVEGEK ANALITIKUS VIZSGÁLATA – A SZAVAK SZINTJE

**Kojanitz László**

*Tankönyv és Taneszköz Kutató és Fejlesztő Intézet*

A pedagógiai szövegek analitikus vizsgálatának legfőbb terepe mindeztáig a tankönyvek világa volt. Leginkább a tankönyvek vizsgálata során dolgozták ki azokat a módszereket, amelyeket egy új kutatás megtervezésekor érdemes lehet figyelembe venni. E tanulmányban igyekszem részletes áttekintést adni a pedagógiai szövegek szókinését, szóhasználatát vizsgáló kutatási eljárásokról és tapasztalatokról.

### **A pedagógiai szöveganalízisek tárgya, célja, szintjei**

Az ismeretek továbbadásához, közvetítéséhez iskolában használt szövegek pedagógiai célt szolgálnak. Ilyen értelemben pedagógiai szövegnek tekintendők nemcsak a tankönyvek és munkafüzetek, hanem az elektronikus tananyagok, a videofilmek, az Internetről gyűjtött háttéranyagok, vagy újságokból származó cikkek is. Az utóbbi kettőt az különbözteti meg a többitől, hogy eredetileg nem az oktatás céljából készítették őket. Az ilyen életből vett ismeretforrások egyre inkább felértékelődnek az iskolai gyakorlatban, azonban csak akkor tudják hasznosan betölteni a szerepüket, ha megfelelnek a pedagógiai szövegekkel szemben támasztott igényeknek. Ezért a tanulhatóság vizsgálata szempontjából érdemes a pedagógiai szövegek fogalmát tágan értelmezni (*Selander*, 1991).

A tankönyvkutatáson belül régóta önálló kutatási területet jelentenek a tankönyvanalízisek. Közös lényegük, hogy a tankönyvek objektív elemzéséhez és értékeléséhez előre meghatározott szabályok szerint számszerűsíthető adatokat gyűjtenek a tankönyvek különböző összetevőiről (*Mikk*, 2000). Ezek az analízisek nagyon sokfélék lehetnek, s egymástól jelentősen eltérő kutatási célokat szolgálhatnak. Például a tankönyvek tartalmának összehasonlításakor mérhetővé lehet tenni, hogy a különböző témák milyen sűrűséggel (*frequency-analysis*), milyen jellemző kapcsolódásokkal (*contingency-analysis*), milyen terjedelemben (*room-analysis*), milyen értéktartalommal (*value-analysis*) jelennek meg (*Laubig* és mtsai, 1986; *Dárdai*, 2002). A tankönyvek illusztrációinak összehasonlításakor is érdemes számszerűsített eredményekre támaszkodni (*Kojanitz*, 2003b). Analízissel pontosan fel lehet tárni, könnyen értelmezhető módon be lehet mutatni, hogy a tankönyvek tartalma, kérdései, feladatai mennyire állnak összhangban az előírt tanulási követelményekkel (*Kojanitz*, 2003a). Hasznos lehet ez a vizsgálati módszer azokban az

esetekben is, amikor a kutató arra keresi a választ, hogy a tankönyv minden tőle elvárható funkciójának megfelel-e. Az ismeretátadáson kívül megfelelő feltételeket nyújt-e az ismeretek alkalmazásának gyakorlásához, a tanulás tanulásához, az értékre neveléshez vagy a gondolkodás fejlesztéséhez is.

E tanulmány elősorban az érthetőség, tanulhatóság problémájához kapcsolódó analitikus kutatásokról, kutatási módszerekről ad áttekintést. Az ezen irányú kutatások azt igyekeznek feltárni, miként lehet a pedagógiai szövegek tulajdonságait mérhetővé tenni, illetve, hogy melyek azok a mérhető szövegtulajdonságok, amelyek leginkább hatással vannak a tanulók olvasásértési és tanulási teljesítményére. Ezen összefüggések feltárása, s egyes bemérhető szövegtulajdonságok indikátorokként történő használata segítséget adhat annak eldöntéséhez, hogy egy szöveg jól tanulható-e, illetve, hogy összhangban van-e a tanulók képességeivel. Az ilyen típusú elemzések és értékelések nagy előnye, hogy lehetőséget ad a hibák kijavítására még azt megelőzően, hogy a tanulók elkezdjenék a kész tankönyveket használni.

Bármilyen pedagógiai szöveg eredményes használatának elengedhetetlen feltétele, hogy azt a tanulók megértsék. A szövegben olvasottak alapján helyesen össze tudják rakni a szerző által közvetíteni kívánt gondolatokat, gondolatsorokat. Ez a rekonstrukció sohasem lehet tökéletes, de a pedagógiai szövegek esetében elvárható, hogy csak nagyon kevés gondolat maradjon kétértelmű vagy homályos az olvasó számára. A pedagógiai szövegek tanulhatóságának vizsgálata alatt elősorban a szövegek érthetőségét befolyásoló összetevők feltárását és értékelését értjük. A jó pedagógiai szöveg azonban nemcsak érthető, hanem érdekes is, azaz szükség van azon kutatások áttekintésére is, amelyek a tanulói érdeklődés felkeltése és fenntartása szempontjából vizsgálják a szövegeket.

Az a szöveg, amit nem értünk, nem lehet érdekes sem. De a túlzottan könnyen érthető szöveg sem az. Az érthetőség és érdekesség együttes igényének kielégítése nem könnyű feladat. A szerzőnek ezért ügyelnie kell arra, hogy a tartalom és fogalmazásmód ne legyen se túl bonyolult, se túl egyszerű, mert mindkettő az érdekesség, az „optimális hír-érték” rovására megy (Mikk, 2000; Nagy, 2000). Az érthetőséget is túlzásba lehet vinni, erre a problémára a pedagógiai szöveganalízisen alapuló értékelések és ajánlások megfogalmazásakor is érdemes odafigyelni.

A pedagógiai szövegek analizésén alapuló adatok nélkül nem képzelhető el tudományos értékű kísérleti tankönyvkutatás. Ahhoz, hogy a különböző tankönyveket használó tanulók teljesítménykülönbségeit helyesen értelmezni lehessen, szükség van a kísérletben kipróbált tankönyvek szövegjellemzőit tükröző adatsorokra.

A tanulói kísérletekkel összekapcsolt kutatásoknak többféle módszere is kialakult. A leggyakoribb ezek közül az, amikor a kísérletben felhasznált szövegvariációk jellemzőit (pl. az alárendelő összetett mondatok aránya, a mondatok és szavak átlagos hosszúsága, az elvont jelentésű főnevek sűrűsége) előre bemérik. S miután a tanulók e szövegekkel különféle feladatokat végeznek el, sor kerülhet a tanulói teljesítmények és a szövegvariációk jellemzőit tükröző adatok statisztikai módszerekkel történő összevetésére. Csak ezen az úton ellenőrizhető azon feltételezések helyessége, amelyek például az alárendelő és mellérendelő mondatok olvasásértés befolyásoló szerepével kapcsolatban is sokszor megfogalmazódnak. A tanulói teljesítmények és a különböző szövegjellemzők adatai között kiszámolt korrelációs koefficiensek összehasonlítása során kiderülhet, melyek a

szövegértést leginkább befolyásoló szövegjellemzők. Ezek felhasználásával úgynevezett olvashatósági képlet (readability formula) is készülhet (Flesch, 1946; Mikk, 2000). Az olvashatósági képlet egyrészt meghatározza, hogy milyen szövegtulajdonságok bemérése lehet hasznos egy pedagógiai szöveg nehézségi szintjének megállapításához. Másrészt egy speciális számítási módszert is ad, amelynek révén az általunk vizsgált szövegre jellemző mérőszámhoz (index) juthatunk, s amelyet a korábbi tanulói kísérletek eredményeként meghatározott határértékekkel is összehasonlíthatunk. Az olvashatósági képlettel kiszámolt mérőszám segítséget ad ahhoz, hogy eldöntsük: a szöveg komplikáltsága összhangban áll-e a tanulók életkorával, képességeivel, vagy sem (Fry, 1977). Sok elméleti és gyakorlati kritika is megfogalmazódott már ezzel az értékelési módszerrel kapcsolatban. Több kutató (Baker, Atwood és Duffy, 1988; Bruce és Rubin, 1988) azért kérdőjelezi meg az olvashatósági képletek alkalmazhatóságát, mert úgy gondolják, hogy a tanulói eredmények alakulásában az adott olvasási (tematikus és olvasói) kontextusoknak döntő szerepe van. A kritikák másik része az olvashatósági képletek alkalmazásának kedvezőtlen hatásaival függ össze. Az olvashatósági képletek gyakorlati alkalmazása az Egyesült Államokban vált a legáltalánosabbá. Több államában az ilyen olvashatósági képletek meghatározó szerepet kaptak a tankönyvi jóváhagyásban is. Sok szakértő szerint (Thomas B. Fordham Institution, 2004) ennek hatására a tankönyvek nyelvezete túlzottan leegyszerűsödött, s a tartalmi elemekről a formai elemekre helyeződött a hangsúly a tankönyvek készítésekor.

Az optimálisnak tekinthető szöveg jellemzőinek meghatározása történhet úgy is, hogy először kiválasztják azt a szövegváltozatot, amellyel a tanulók a legjobb eredményt tudták elérni. E szöveg analízissel bemért paraméterei szolgálhatnak később viszonyítási pontul más szövegek alkalmasságának megítéléséhez (Klare, 1985).

Egy másik kutatási irányzat a tanulók által készített szövegek komplikáltságát veszi viszonyítási alapul. Abból a feltételezésből kiindulva, hogy a tanulók által használt mondatstruktúrák nem jelenthetnek különösebb problémát olvasás közben sem. Az aprólékos szöveganalízis az ilyen típusú összehasonlítások során is fontos szerepet kap (B. Fejes, 2002).

Annak, aki tanulói teljesítmények mérésén alapuló kísérleti tankönyvkutatásokat akar végezni, érdemes számba vennie mindazokat a megfontolásokat és módszereket, amelyek a szövegjellemzők adatszerű mérése és rögzítése terén már kialakultak. A kutatás egyik legfontosabb döntése lehet az, hogy milyen jellegű szöveganalízis alkalmazása mellett döntünk. Érdemes kritikusan megvizsgálni a lehetőséget, mert bár a kipróbált megoldási variációk száma igen nagy, egyáltalán nem biztos, hogy ezek mindegyik helyes is. A szövegminőség kvantitatív adatokkal történő megragadása ugyanis rendkívül összetett problémát jelent.

Az elmúlt évtizedekben rendkívül sok és sokféle pedagógiai szövegelemzés (pedagogic text analysis) zajlott le a nagyvilágban. Ezekről egy kitűnő, angol nyelvű áttekintést készített Jaan Mikk észt kutató (Mikk, 2000). Magyarországon is sor került már ilyen típusú vizsgálatokra (B. Fejes, 2002; Dárdai, 2002; Kojanitz, 2003a, 2003b, 2004a). Vannak olyan hazai szöveganalízisen alapuló vizsgálatok is, amelyek eredetileg nem a pedagógiai szövegekhez kötődtek, de amelyek módszerei és eredményei ugyancsak hasznosíthatók az iskolában használt szövegek vizsgálata során is (Deme, 1971;

Pléh, 1998). A következő fejezetben ismertetem, hogy milyen szervezőelv alapján lehet mindezeket áttekinthető módon bemutatni.

## A szöveganalízisek bemutatásának lehetséges útjai

Az egyik lehetséges megoldás, ha azon szempontok köré szervezve mutatjuk be a pedagógiai szöveganalízis eredményeit, amelyek a hétköznapi értékelésben is fontos szerepet játszanak. Ilyen szempont lehet a tartalom, a strukturáltság és a szövegezés. A tanulók számára a még érthető, de már érdekes szöveg az optimális. A tartalmi egységek logikusan kapcsolódnak egymáshoz, s a szerző által követett rendező elv az olvasó számára is könnyen felismerhető. A szövegezés komplikáltsága pedig megfelel a diákok olvasási szintjének, anélkül, hogy megengedhetetlen tartalmi leegyszerűsítésekhez vezetne.

A vizsgálatok bemutatásának másfajta rendezőelvét kínálja a pedagógiai szövegektől elvárt funkciók sora: motiválás, ismeretátadás, rendszerezés, problémafelvetés, magyarázat. A szöveganalízisek célja lehet az is, hogy azonosítsák azokat a szövegtulajdonságokat, amelyek meghatározó szerepet játszanak a különböző funkciók megvalósulásában. A korábban említetteken kívül ilyen szövegtulajdonság lehet még a szöveg konkrétsága és elvontsága, a szerző és az olvasó közötti kommunikációs viszony, a tételmondatok és az érvelő mondatok aránya, a megértést támogató szövegstruktúrák (comprehension-enhancing structures) típusai: időrenden alapuló, tulajdonságokat hozzárendelő, állításokat ütköztető, kauzális, problémafelvető (*temporal, attributive, adversative, covariance, responsive*) (Julkunen, 1991).

A szöveganalízisek bemutatásához egy harmadik megoldási módot választottam. Ehhez az ötletet és az ösztönzést egy érdekes kutatási eredmény szolgáltatta. Geiling 1987-ben német nyelvű szövegekkel végzett kutatások alapján arra a következtetésre jutott, hogy a megértési nehézségek 13%-a volt visszavezethető a szavak szintjére, 43%-a a mondatok szintjére és 44%-a a szöveg szintjére (Geiling, 1987). Vagyis van, amikor a megértési problémát a túl sok ismeretlen szó, van, amikor a túlzottan komplikált mondatstruktúrák, s van, amikor a megfelelő tartalmi kifejtés hiánya vagy a logikailag rendezetlen szövegstruktúra okozza. A megértési problémák ilyen jellegű színteztettségét az idegennyelvek tanulása közben mindnyájan átéljük. Például, amikor egy mondat tartalmát még akkor sem értjük, amikor már az összes szó jelentését külön-külön ismerjük. A szövegek megértésének szavak, mondatok, szövegek szintjére bontott differenciálását nemcsak a mindennapi tapasztalat, hanem a neurofiziológiai kutatások is alátámasztották (Mikk, 2000). Egyes agysérültek esetében megfigyelték, hogy a baleset hatására az egyik szint megértése szinte teljesen megszűnt, míg a másik két szinten semmi gond nem mutatkozott. Collins és Smith (1982) a szöveges információk megértésének nehézségeit ugyanakkor nem három, hanem négy szinten tartották jól megragadhatónak: a szavak, a mondatok, a mondatok közötti kapcsolatok és a szöveg nagyobb egységeinek kapcsolata szintjén.

A pedagógiai szöveganalízisek bemutatása során is érdemes e négy szint szerint előrehaladni. Ez a megoldás lehetővé teszi, hogy az egyszerűbb problémáktól haladjunk a

komplexebbek felé. Ráadásul így könnyebben észrevehetők a hasonlóságok és a különbségek a kutatók és a kutatási módszerek között tekintetben, hogy ki mit tart fontosnak a szóhasználat, a mondatszerkezetek, a mondatközi kapcsolatok és a szövegtani sajátosságok vizsgálatakor.

## A szavak használatának gyakorisága

A megfelelő tartalom, szóhasználat, stílus és strukturáltság együttesen biztosíthatja, hogy egy pedagógiai szöveg alkalmas legyen a tanulásra, hogy megfelelőképpen motiváljon, biztosítsa a zavartalan ismeretátadást, hogy lehetővé tegye az új információk rendszerezését. Sok minden azonban már a szavak szintjén eldőlt. A nem megfelelő szavak használata önmagában is a szöveg megértésének és az érdeklődés felkeltésének akadályává válhat. A tanulók érdeklődéséhez és szókincséhez alkalmazkodó szóhasználat viszont megkönnyítheti a tanulást a számukra.

A tanulóknak nehézséget jelenthetnek mindazok a szavak, amelyeket még nem ismernek, vagy amelyek pontos megértése még gondot okoz a számukra. Ilyenek a mindennapi beszédben ritkán előforduló szavak, a régies kifejezések, a tudományos szakszavak és az idegen (alakú) szavak. A tankönyvanalízis során az e típusokba sorolható szavak számának mérése, sűrűségének megállapítása mindenképpen hasznos lehet a szövegzés komplikáltságának összehasonlításához vagy jellemzéséhez.

A tanulók számára ismerős, a mindennapi nyelvben is sűrűn előforduló szavak elolvasása szignifikánsan gyorsabb, ahogyan azt kísérleti kutatások is bizonyították (*Howes és Solomon, 1951; McGinnies, Comer és Lacey, 1952; Davey, 1987; Plourde és Besner, 1997*).

A könnyebb megértés szempontjából még fontosabb, hogy a mindennapi életben leggyakrabban előforduló szavakat a diákok is rendszeresen használják. Így az e szavakhoz kapcsolódó tapasztalatok sokféle élethelyzetből és kontextusból származnak, s nagyon sokféle jelentéskapcsolat alakul ki közöttük. E kapcsolatok aktivizálódása nagy szerepet játszik a megértésben (*McCann, Besner és Davelass, 1988*).

*Nagy József* az olvasáskészség, a szövegértés alapvető feltételeként írja le a leggyakoribb 5000 szó rutinszerű felismerését. Olvasás közben e szavak felismeréséhez ebben az esetben a tanulónak már nincs szüksége a szavak kibetűzésére, hanem a felismerés szinte automatikus, rutinszerű. Ez teszi lehetővé, hogy a tanuló ilyenkor már a szöveg tartalmára tud figyelni, s a hosszabb szövegek elolvasása sem jelent túlzott erőfeszítést számára. Így kell tudni olvasni már egy 5. osztályos diáknak is ahhoz, hogy esélye legyen a sikeres tanulásra. A leggyakoribb 1000 szó egy átlagos nehézségű szöveg szavainak 75%-át, a leggyakoribb 5000 szó a 96%-át jelenti. Ezt a számítást folytatva azt mondhatjuk, hogy egy alapvető szórutinnal olvasó tanuló esetében egy szövegoldalón kb. 5–10 olyan szó lehet, amelyet még nem ilyen automatizmus szerint ismer fel (*Nagy, 2004*). A pedagógiai szöveganalízis hasznos részévé válhat annak ellenőrzése, hogy a leggyakoribb 1000 és 5000 szó a szövegszavak hány százalékát teszi ki. Tíz-tizenkét éves korig inkább az 1000 leggyakoribb szó, később pedig az 5000 leggyakoribb szó

százalékos aránya lehet egy olyan jelzőszám, amely a szöveg érthetőségének hasznos mutatójává válhat.

Hasonló megfontolások alapján a külföldi kutatók is fontos adatként kezelik a beszélt nyelvben leggyakrabban használt szavak arányát és sűrűségét. Rendkívül magas lehet ugyanis a korrelációs koefficiens ( $r = 0,56$   $p < 0,95$ ) szövegek használata alapján mért tanulói eredmények és ezen adat között (Mikk, 2000).

A pedagógiai szakszövegek értékelésekor a tartalomhordozó szavak (főnév, ige, melléknév, számnév) esetében külön is szükséges kiszámolni, hogy ezek hány százaléka szerepel, illetve nem szerepel a gyakran használt szavak között. Általában a főnevek és a melléknevek közül kerül ki a legtöbb olyan szó, amely ritkán fordul elő a mindennapi nyelvhasználat során.

Ha a szavak „ismerőssége” (familiarity of words) felől akarjuk megközelíteni a megértés problémáját, a pedagógiai szövegek szóanyagának vizsgálatokor hasznosak lehetnek a következő adatok: azoknak a szavaknak a százalékos aránya, amelyek nem szerepelnek a gyakran használt szavak szótárában; a szövegben előforduló szavak ismétlődésének mérőszáma. Ugyanezeket az értékeket ki kell számolni szófajonként is, külön a főnevekre, melléknevekre, igékre stb.

## A szavak hosszúsága és szófaja

A gyakrabban használt szavak hosszúsága általában rövidebb. Ezzel szemben az elvont jelentésű szavak általában hosszabbak. A szavak hosszúságát növelik a különböző elő- és utótagok, képzők. Az összetett szavak pedig különösen hosszúak lehetnek. Ilyen esetekben ráadásul sok információ zsúfolódik össze a szavakban. A szöveg szavainak átlagos hosszúsága ezért általában összefügg a szöveg tartalmi komplikáltságával. A statisztikai módszerekkel végzett vizsgálatok igazolták ezt a gondolatmenetet. A pedagógiai szövegek olvashatóságának (readability) és érthetőségének (comprehensibility) vizsgálatokor ezért a legtöbb esetben felhasználják a szöveg szavainak átlaghosszúságára vonatkozó adatokat is (Mikk, 2000). Baker és Stoke (1968) vizsgálatuk során elsősorban a hosszú szavak (pl. a 4–5 szótagonál hosszabb szavak) százalékarányát használták a szövegek érthetőségének összehasonlításához.

A szöveg érthetősége szempontjából a szavakat szófajuk szempontjából is szükséges megvizsgálni. Empirikus kutatások alapján általánosan elfogadott vélemény, hogy a mondatok legnehezebben érthető alkotóelemei az új tartalmat leginkább hordozó főnevek. Elley (1969) kutatásában szoros összefüggés ( $r = 0,90$ ) mutatkozott a szöveg nehézségi szintje és a főnevek előfordulásának sűrűsége között. Más vizsgálatok a melléknevekkel kapcsolatban is jeleztek hasonló összefüggéseket (Klein-Braley, 1985). Ehhez hasonló kutatási eredményekre támaszkodva többféle képlet is készült a szövegek nehézségét előrejelző mérőszám kiszámításához. Van, aki a főnevek és az igék hányadosát tartja hasznos eszköznek: úgynevezett főnév index (index of substantivity). Van ennél összetettebb képlet is. A szófajkomplikáltsági index (index of word class complicity) kiszámításához a főnevek és igék számán kívül szükség van a melléknevek és a határozószavak

számára is: főnév + melléknév/ige + határozószó. Ezen indexek kiszámítása nagyon jó előrejelzést adott a természettudományi tankönyvek érthetőségéről (Mikk, 1997).

## A szavak jelentésének elvontsága

A megértés szempontjából az is számít, hogy mennyire elvont a szöveg tartalma. Minél konkrétabb, annál érthetőbb, minél elvontabb, annál nehezebb (Begg és Paivo, 1969; Klare, 1985; Eye, Dixon és Krampen, 1989). Az elvont tartalmat felidézni is jóval nehezebb. Paivo (1969) ezt a kísérleti eredményekkel is alátámasztott jelenséget azzal magyarázta, hogy a konkrét tartalmú mondatok tartalmát, mint képzeteket, az elvont jelentésű mondatokét pedig, mint verbális információkat tároljuk. Az elvontság mértékének vizsgálata fontos lehet a pedagógiai szövegek esetében. Az ilyen típusú különbségek analitikus feltárását a szavak szintjén is el lehet végezni.

Flesch (1950) pontosan definiált grammatikai tulajdonságok alapján azonosította a konkrét jelentésű szavakat. Ezen a módon megmérte, hogy 100 szóra mennyi ilyen konkrét jelentésű szó jut. Majd ezt az adatot felhasználva számította ki a szövegek elvontsági szintjét. Angol, német és észt nyelvterületen kísérletek történtek arra is, hogy az elvont jelentésre utaló toldalékok alapján gyűjtsék ki a szövegekből az ilyen főneveket, s ezek számát használják a szöveg elvontságának kiszámításához (Gillie, 1957; Groeben, 1982). A grammatikai szabályokon és sajátosságokon alapuló vizsgálati módszerek mellett a szavak jelentését figyelembe vevő elvontsági skálák is készültek. Wellington (1983) a természettudományi tantárgyak esetében négy szintet különböztetett meg a szavak elvontságának megállapításához: a dolgok nevei (nominative words), folyamatok nevei (names of processes), fogalmak (notions), matematikai szakszavak és szimbólumok (mathematical terms and symbols). Mikk (2000) a főneveket sorolta három csoportba: (1) főnevek, amelyek közvetlenül érzékelhető tárgyakat jelölnek; (2) főnevek, amelyek közvetlenül érzékelhető folyamatokat, jelenségeket jelölnek; (3) főnevek, amelyek közvetlenül nem érzékelhető fogalmakat jelölnek. Az utolsó csoportba tartozókat tekintette elvont főneveknek, s ezek számának ismeretében számította ki a szövegek elvontsági szintjét. Tanulói kísérleteket végezve Kukemelk és Mikk (1993) azt állapították meg, hogy az így kapott érték az egyik legjobb előrejelzője volt a szöveg nehézségi szintjének. A tanulók ismeretelsajátításának szintje és a szöveg főneveinek elvontsági szintje között  $r = -0,53$  és  $r = -0,66$  (mindkettő  $p < 0,05$  szinten szignifikáns) korreláció volt. Az elvont jelentésű főnevek közül nagyon sok a szakszó, így ezzel is magyarázható ez a negatív hatás a tanulók teljesítményére.

Magam a szakiskolai földrajztankönyvek vizsgálatakor megpróbáltam kombinálni a főnevek ismerőssége és elvontsága szerinti analízis módszerét. A megértés nehézsége szempontjából kétféleképpen is elválaszthatjuk egymástól a szövegekben előforduló főneveket: közvetlenül érzékelhető dolgok – közvetlenül nem érzékelhető fogalmak, illetve a mindennapi nyelvhasználatban gyakran vagy ritkán használt főnév. Ennek alapján problémamentesnek tekinthetjük a közvetlenül érzékelhető tárgyak, tevékenységek, jelenségek mindennapi nyelvhasználatban is előforduló megnevezésére szolgáló főneve-

ket, illetve szóösszetételeket. Ha az ilyen problémamentesnek tekinthető főnevek száma magasabb, mint a problémásaké, a tankönyvi magyarázó szöveg könnyebben érthető a tanulók számára. A szöveg nehézségi fokának értékét megadhatjuk a problémamentes és a problémás főnevek hányadosával (Kojanitz, 2003c).

### **Az idegen szavak és a szakszavak**

Az idegen szavak és szakszavak különös figyelmet érdemelnek a pedagógiai szövegekben. A legtöbb megértési és tanulási problémát ugyanis a szakszavak szokták okozni a tanulóknak. Igen magas negatív korrelációkat ( $r = -0,64$  és  $r = -0,56$  mindkettő  $p < 0,05$ ) állapítottak meg a szakszavak sűrűsége és aközött, hogy a tanulók mennyire tartották érdekesnek és érthetőnek az egyes szövegeket (Mikk, 2000). A sok szakszó kétféle módon is hátráltathatja az olvasottak befogadását. Egyrészt ilyenkor nagyobb az esélye annak, hogy ezek között a tanulók számára ismeretlen szavak is előfordulnak. Másrészt, ha túlzottan sűrűn fordulnak elő, meghosszabbítják, és bizonytalanná teszik az új ismeretek megértését és feldolgozását. Mindezt a tanárok és a szülők is érzékelik. A tankönyvekkel kapcsolatos legtöbb kritika a túlzottan sok szakszóra vagy a kellően mértékben nem megmagyarázott szakszavakra vonatkozik (Chambliss és Calfee, 1998; Merzyn, 1996; Kojanitz, 2004a). Sok esetben a szakszavak definíciói sem mindig érthetőek a diákok számára. Egyebek mellett azért, mert ezek a definíciók is újabb magyarázatra szoruló szakszavakat tartalmaznak.

A szövegek készítőinek ügyelniük kell arra, hogy a szakszavak és az idegen szavak száma és sűrűsége ne legyen túl magas. Ezek megértése és befogadása tulajdonképpen ugyanolyan feladatot jelent, mint az idegen nyelvek tanulása során az új szavak megtanulása. Az elsajátításukhoz időt, érthető magyarázatot és minél több konkrét példát kell adni a tanulók számára. A túlzottan sok idegen szó és szakszó viszont éppen ezt nem teszi lehetővé.

A vizsgálatok szempontjából az idegen szavak azonosítása nem jelent különös problémát. A legtöbb ilyen jellegű szöveganalízis az idegen szavak szótárát használja ehhez a feladathoz. Ha egy szó benne van a szótárban, akkor azt mindenképpen figyelembe eszik a mérésnél, s nem mérlegelik, hogy egy közismert idegen szóról, vagy egy tudományterület speciális szakszaváról van-e szó. Ez a differenciálás általában csak a szakszavak kiválasztásakor történik meg az idegen szavak esetében is.

A szakszavakra vonatkozó analízisek elvégzése viszont nem könnyű feladat. Nehéz ugyanis egyértelmű szabályt alkotni arról, hogy a szöveg szavai közül melyik minősüljön szakszónak. Például a „zivatar” és a „csapadék”, a „hőmérséklet” és a „hőingás” szavak esetében háromféle döntést is hozhatunk: egyiket se tekintjük szakszónak; mindegyiket szakszónak tekintjük; a csapadékot és a hőingást szakszóként kezeljük, a zivatart és a hőmérsékletet pedig nem. A szakszavak vizsgálata során a speciális jelentéssel összefüggő további differenciálás igénye is jogosan felmerül: pl. „torkolat” és „tölcsértorkolat”; „állattenyésztés” és „alpi tehenészet”.



A kutatók többféle megoldással próbálkoznak a szakszavakkal kapcsolatos mérések elvégzésekor. Vannak, akik a különböző tudományterületek lexikonjait használják a feladathoz (*Kukemelk és Mikk, 1993*). Például a biológiatankönyvek esetében minden olyan szót szakszónak tekintenek, amely külön szócikként szerepel a természettudományi lexikonban. Ennek a megoldásnak a legfőbb előnye, hogy a mérést végzők számára egyértelmű szabályt lehet adni. Nem áll fenn annak a veszélye, hogy valamelyik szó az egyik tankönyvben szakszóként lesz azonosítva, a másokban viszont nem. Ilyen megoldás segítségével korrelációkat lehet kiszámítani a tanulók olvasásértési teljesítményei és aközött, hogy a feladatokhoz használt szövegekben milyen százalékarányban fordultak elő a lexikonban szereplő vagy nem szereplő főnevek (*Kukemelk és Mikk, 1993*).

Egy másik megoldás az, amikor csak azokat a szakszavakat veszik figyelembe, amelyekhez valamilyen definíció is található a tankönyvben. Ez a mérőszám elsősorban a tankönyvben lévő új ismeretek mennyiségére vonatkozóan adhat támpontot (*Mikk, 2000*).

A kutatók többsége tartalmuk és használatuk szempontjából különböző csoportokba sorolja a szövegekben talált szakszavakat, majd az adatokat külön-külön értelmezik, annak megfelelően, hogy a különböző típusba sorolt szakszavakról mi a véleményük a szöveg érthetősége szempontjából (*Graf, 1989; Baumann, 1974; Elts, 1992; Kojanitz, 2004a*). A biológia tankönyvek esetében megkülönböztethetjük egymástól a rendszertani kategóriákat (emlős, növény), a növény- és állatneveket (kékbálna, szarvas), az organizmusok részeit (szem, klorofill), az olyan fogalmakat, amelyek a biológiai folyamatokhoz kapcsolódnak (szaporodás, élet) és a biológia elméleti fogalmait (törzsfajlás, alkalmazkodás). A hazai alsó tagozatos természetismeret tankönyvek vizsgálata során azt tapasztaltuk, hogy jelentős különbségek vannak közöttük tekintetben, hogy e biológiai fogalomfajták milyen számban és milyen arányban szerepelnek a szövegükben. Ezek az adatok a tankönyvek tartalmi különbségeiről is sokat elárultak. Ahol magas volt a rendszertani szakszavak száma, ott a 3. osztályosok számára készült tankönyv túlzott mértékben is a szakszövegek stílusát és szabatoságát próbálta meg követni. Ahol a növény- és állatnevek száma volt magas, ott a természet konkrétumainak a bemutatása dominált az elméleti ismeretekhez képest. Árukodó volt az is, ha az organizmusok részei szerepeltek kiugróan magas számban. Ez azt mutatta meg, hogy abban a tankönyvben jobban előtérbe kerültek a növények és állatok külső és belső felépítéséről szóló biológiai ismeretek. Az elemzések arra is rávilágítottak, hogy a biológiai folyamatokhoz kapcsolódó szakszavak és az elméleti fogalmak még meglehetősen ritkán fordulnak elő a 3. és 4. osztályos tankönyvekben (*Kojanitz, 2004c*).

A történelemtankönyvek szakszavainak vizsgálatakor szükséges különbséget tenni aszerint, hogy azok között vannak általános jelentésű társadalomtudományi szakszavak (állam, társadalmi viszonyok, politika), és konkrét történelmi eseményekhez kötődő szakszavak (felvilágosodás, New Deal, harmadik rend) is. Az ilyen típusú megkülönböztetés lehetősége összefügg a szakszavak azon tulajdonságával, hogy mennyire kötődnek bizonyos kontextushoz. *Barsalau (1982)* nevéhez kötődik a fogalmak ilyen jellegű különbségeinek feltárása. Úgy találta, hogy e megkülönböztetés azért is fontos, mert a kontextusfüggő és kontextusfüggetlen fogalmak megértése és felidézése között jellegzetes különbségek figyelhetők meg, s ezeket a különbségeket a pedagógiai munka szem-

pontjából is figyelembe kell venni. A kontextusfüggetlen fogalmak könnyen aktivizálódnak önmagukban állva is, míg a kontextusfüggő szavak esetében ehhez szükség van a megfelelő szövegekörnyezetre.

A szakszavak nehézségi szintjeit a mindennapi nyelvhez viszonyítva is meg lehet határozni: közismert, a mindennapi nyelvben is gyakran használt szakszó; a mindennapi nyelvben ritkán használt szakszó; csak egy vagy két tudományterület által használt speciális szakszó. Az utóbbi csoportba tartozó tudományos szakszavak mennyiségének és előfordulási sűrűségének megállapítása nagyon fontos adat lehet a tankönyvi szövegek érthetőségének összehasonlításakor. Az ilyen típusú szakszavak nagy mennyisége önmagában is sok gondot okozhat a tanulóknak. A nagy mennyiségű szakszó használata következhet a túlméretezett ismeretanyagból és a túlzottan tudományoskodó fogalmazásmódból is. Ezért a szakszavakra vonatkozó adatok fontos problémákra hívhatják fel a készítőket és a használókat figyelmét. A tankönyvek összehasonlításához mérőszámot is használtunk. *Baumann* a szövegben található szakszavak számát a szövegben található összes szó számával elosztva képzett ilyen jelzőszámot (*Baumann*, 1974). Ehhez hasonló megoldás, amikor a szakszavak százalék arányát határozzák meg a tankönyvekben. Ezt az adatot használták fel például a kutatók a finn közismereti tankönyvek pedagógiai szempontú összehasonlításakor (*Julkunen*, 1991).

A szakszavakkal kapcsolatos szöveganalízisek kiterjedhetnek a szövegben előforduló szakszavak ismétlődésének vizsgálatára is. Az ismeretek megértése szempontjából ugyanis gyakran elhangzó kritika, hogy a tankönyvek az új fogalmakat és szakszavakat nem használják elég alkalommal ahhoz, hogy ezek beépülhessenek a tanulók fogalomrendszerébe (*Graf*, 1989). Fontos mutatószám lehet a szöveg tanulhatósága szempontjából a szakszavak összes előfordulásának és az egymástól eltérő szakszavak számának hányadosa, amely így kifejezi, hogy átlagban egy szakszó hányszor ismétlődik meg a vizsgált szövegrészben. Az ilyen jellegű hazai vizsgálat azt mutatta, hogy a felsős és középiskolai tankönyveink többségében a túl sok szakszó használata és a szakszavak kevés alkalommal történő megismétlődése egyszerre nehezíti meg a tanulók dolgát (*Kojanitz*, 2004b).

## Az érdekesség problémája

A szavak szintjén végzett szöveganalízisek bemutatását azokkal fejezzük be, amelyek az érdekesség szempontjából igyekeztek összehasonlítható adatokhoz jutni. Már a bevezetőben is említettem, hogy az érdekesség mennyire fontos összetevője lehet a tanulhatóságnak. *Anderson*, *Mason* és *Shirey* (1984) valamint *Klare* (1985) kutatási eredményei szerint a tanulók teljesítménye szempontjából az érdekesség az érthetőséggel egyenrangú szerepet játszik a tankönyvi szövegekkel végzett tanulás eredményességében.

Ha analízisen és statisztikai adatok összevetésén alapuló módszerrel vizsgálják a szövegjellemzők és az érdekesség összefüggéseit, általában a következőképpen járnak el. A tanulóknak számszerű osztályzattal kell értékelniük a szövegek érdekességét. Ezek után szignifikáns korrelációkat keresnek a tanulói értékelések eredményei és az egyes

szövegek jellegzetességeiről felvett számszerűsített adatok között, így választva ki azokat a szövegjellemzőket, amelyek valószínűleg az átlagnál nagyobb szerepet játszanak abban, hogy a tanulók mennyire tartanak érdekesnek egy szöveget. Egy érettségivizsgálatban (Mikk, 2000) 9–10. évfolyamos tanulóknak fizikai témájú szöveget kellett értékelniük abból a szempontból, hogy mennyire találták azokat érdekesnek. A vizsgálat tanúsága szerint pozitív korreláció mutatkozott a tanulók által adott pontszám és a mindennapi életben is használt főnevek és az igék százalék aránya között. Ugyanakkor nagyon magas negatív korreláció volt megállapítható több szövegjellemzővel kapcsolatban is. Negatívan befolyásolta a tanulók értékelését a fizikatankönyvükben is megtalálható főnevek magas százalék aránya ( $r = -0,70$   $p < 0,05$ ), a szaktudományos kifejezések magas százalék aránya ( $r = -0,64$   $p < 0,05$ ), a szavak átlagos ismétlődésének mértéke ( $r = -0,72$   $p < 0,05$ ); a főnevek elvontságának mértéke ( $r = -0,52$   $p < 0,05$ ). Egy ugyanilyen módszerű kutatás (Mikk, 2000), amelyet 7. és 9. osztályos tanulókkal és biológiai témájú szövegek segítségével végeztek, hasonló eredményeket hozott. Ez is megerősítette, hogy a szavak szintjén az elvont főnevek, az ismeretlen szavak, a szakszavak, a hosszú szavak, a túl sok főnév és túl sokszor ismétlődő főnevek gyengítik leginkább a szöveg érdekességét a tanulók számára (Mikk, 2000). A szöveganalízisek értelmezésekor ezen adatokat érdemes az érdekesség szempontjából külön is kiértékelni.

Az új információk mennyisége és a szöveg érdekessége közötti összefüggést egy olyan grafikonnal lehet ábrázolni, amelyik egy ideig meredeken emelkedik, majd egy ponttól kezdve ugyanilyen meredeken zuhanni kezd. Vagyis van az új információk mennyiségének egy olyan optimális mértéke, amelyet érdemes elérni, de amelyet semmiképpen sem szabad meghaladni (Mikk, 2000). A cloze teszttel és a mindennapi nyelvben ritkán használatos szavak mennyiségének mérésével lehet szöveganalízisen keresztül megragadni a tanulók számára új információt jelentő információk mennyiségét: Mindkét módszer azt vizsgálja, hogy mennyire ismerős vagy ismeretlen a szöveg tartalma a tanulók számára.

E tanulmányban igyekeztem bemutatni, hogy a pedagógiai szövegek szókincsének, szóhasználatának analitikus jellegű vizsgálata milyen szempontok szerint, milyen módszerekkel történhet, s melyek azok a nemzetközi és hazai tapasztalatok, amelyek használhatóak lehetnek a hasonló célú kutatásokban. Nem szabad azonban elfelejteni, hogy a pedagógiai szövegek érthetőségével és érdekességével kapcsolatos problémáknak csak egy kis része tárható fel a szavak szintjén. A szöveganalízisen alapuló vizsgálatoknak ki kell terjedni a mondatokra, a mondatközi kapcsolatokra és a szöveg egészére is. E szintek analitikus vizsgálatáról is széleskörű és gazdag tapasztalat áll a kutatók rendelkezésére. Érdemes lenne ezeket is rendszerezett módon megismertetni mindazokkal, akik pedagógiai szövegek készítésével és értékelésével foglalkoznak.

## Irodalom

- Anderson, R., Mason, J. és Shirey, L. (1984): The reading group: An experimental investigation of a labyrinth. *Reading Research Quarterly*, **20**. 1. sz. 6–38.
- B. Fejes Katalin (2002): *A tankönyvszöveg szintaktikai jellemzői*. Juhász Gyula Felsőoktatási Kiadó, Szeged.
- Baker, D. G. és Stokes, W. W. (1968): A simplification of the revised Lorge Readability Formula. *Journal of Educational Research*, **61**. 5–6. sz. 398–400.
- Baker, E. L., Atwood N. K. és Duffy T. M. (1988): Cognitive approaches to assessing the readability of text. In: Davison, A., Green, G. M. (szerk.): *Linguistic Complexity and Text Comprehension. Readability Issues Reconsidered*. Lawrence Erlbaum Associated Publishers, Hillsdale, New Jersey, London. 55–83.
- Barsalou, L. W. (1982): Context-independent and context-dependent information in concepts. *Memory és Cognition*, **10**. 1. sz. 82–93.
- Baumann, M. (1974): *Theoretische Anschätze und erste Ergebnisse der Textuntersuchungen innerhalb der Schulbuchforschung am Pädagogischen Institut Köthen. Informationen zu Schulbuchfragen. Heft 17*. Volk und Wissen Volkseigener Verlag, Berlin. 9–62.
- Begg, J. és Paivo, A. (1969): Concreteness and imaginary in sentence meaning. *Journal of Verbal Learning and Verbal Behavior*, **8**. 6. sz. 821–827.
- Bruce, B. és Rubin, A. (1988): Readability formulas: Matching tool and task. In: Davison, A. Green, G. M. (szerk.): *Linguistic Complexity and Text Comprehension. Readability Issues Reconsidered*. Hillsdale, New Jersey, Lawrence Erlbaum Associated Publishers, London. 5–22.
- Chambliss, J. S. és Calfee, R. C. (1998): *Textbooks for Learning. Nurturing the Childrens Minds*. Blackwell Publishers, Malden.
- Collins, A. és Smith, E. E. (1982): Teaching the process of reading comprehension. In: Detterman, D. K. és Sternberg, R. J. (szerk.): *How and how much can intelligence be increased*. NJ: Ablex, Norwood. 173–185.
- Dárdai Ágnes (2002): *A tankönyvkutatás alapjai*. Dialóg Campus Kiadó, Budapest-Pécs.
- Davey, B. (1987): Relation between word knowledge and comprehension: generalisation across tasks and readers. *Journal of Educational Research*, **80**. 3. sz. 179–183.
- Deme László (1971): *Mondatszerkezeti sajátosságok gyakorisági vizsgálata (Magyar szövegek alapján)*. Akadémia Kiadó, Budapest.
- Elley, W. B. (1969): The assessment of readability by noun frequency counts. *Reading Research Quarterly*, **4**. 411–427.
- Eltis, J. (1992): A readability formula for texts on biology. In: Rimsha, V. (szerk.): *Psychological Problems of Reading. Theses of Papers for the International Scientific Conference*. International Library of Lithuania, Vilnius. 42–44.
- Eye, A., Dixon, R. A. és Krampen, G. (1989): Text recall in adulthood: The roles of text imagery and orienting. *Psychological Research*, **51**. 136–146.
- Flesch, R. (1946): *The art of plain talk*. Harper and Brothers Publishers, New York, London.
- Flesch, R. (1950): Measuring the level of abstraction. *Journal of Applied Psychology*, **34**. sz. 384–390.
- Fry, E. (1977): *Elementary Reading Instruction*. McGraw Hill, New York.
- Geiling, U. (1987): Quantitative Ergebnisse der Untersuchung von verständnisemmenden Stellen in Lehrtexten. *Informationen zu Schulbuchfragen*, **56**. sz. 56–59.
- Gillie, P. J. (1957): A simplified formula for measuring abstraction in writing. *Journal of Applied Psychology*, **41**. 214–217.
- Graf, D. (1989): *Begriffslernen im Biologieunterricht der Sekundarstufe I*. Peter Lang, Frankfurt am Main.

- Groeben, N. (1982): *Leserpsychologie: Text Verständnis – Textverständlichkeit*. Aschendorf, Münster Westfalen. 539.
- Howes, D. H. és Solomon, R. L. (1951): Visual duration threshold as a function of word-probability. *Journal of Experimental Psychology*, **41**. 6. sz. 401–410.
- Julkunen, M. L. (1991): Text types and teaching of concepts in finnish school books. In: Julkunen, M. L. Selander, S. és Ahlberg, M. (szerk.): *Research on texts at school*. Joensuu Yliopiston Kirjasto, Joensuu. 11–34.
- Klare, G. R. (1985): *How to write Readable English*. Hutchinson, London.
- Klein-Braley, Ch. (1985): A cloze-up on the C-test: A study in the construct validation of authentic tests. *Language Teaching*, **2**. 1. sz. 76–104.
- Kojanitz László (2003a): A szakiskola tankönyvek összehasonlító elemzése IV. *Új Pedagógiai Szemle*, 12. sz. 30–41.
- Kojanitz László (2003b): A szakiskola tankönyvek összehasonlító elemzése III. *Új Pedagógiai Szemle*, 11. sz. 77–84.
- Kojanitz László (2003c): A szakiskola tankönyvek összehasonlító elemzése II. *Új Pedagógiai Szemle*, 10. sz. 43–52.
- Kojanitz László (2004a): Lehet-e statisztikai eszközökkel mérni a tankönyvek minőségét? *Iskolakultúra*, 9. sz. 38–56.
- Kojanitz László (2004b): A tankönyvek használhatóságát meghatározó minőségi összetevők elemzése és összehasonlítása. (Kézirat)
- Kojanitz László (2004c): A 3–4. osztályos környezetismeret tankönyvek összehasonlító vizsgálata. (Kézirat)
- Kukemelk, H. és Mikk, J. (1993): The prognosticating effectivity of learning a text in physics. *Glottometrica*, **14**. 82–103.
- Laubig, M., Peters, H. és Weinbrenner, P. (1986): *Methodenprobleme ser Schulbuchanalyse*. Universität, Bielefeld.
- McCann, R., Besner, D. és Davelaar, E. (1988). Word recognition identification: Do word-frequency effects reflect lexical access? *Journal of Experimental Psychology: Human Perception and Performance*, **14**. 693–706.
- McGinnies, E., Comer P. B. és Lacey, O. L. (1952): Visual-recognition thresholds as a function of word frequency. *Journal of Experimental Psychology*, **44**. 8. sz. 65–69.
- Merzyn, G. (1996): A comparison of some linguistic variables in fifteen science texts. In: Welford, G., Osborne, J. és Scott, P. (szerk.): *Research in Science Education in Europe: Current Issues and Themes*. The Falmer Press, London, Washington. 361–369.
- Mikk, J. (1997): Parts of speech in predicting reading comprehension. *Journal of Quantitative Linguistics*, **4**. 1–3. sz. 156–163.
- Mikk, J. (2000): *Textbook: research and writing*. Europäischer Verlag der Wissenschaften, Frankfurt am Main.
- Nagy József (2000): *XXI. század és nevelés*. Osiris Kiadó, Budapest.
- Nagy József (2004): Olvasástanítás: a megoldás stratégiai kérdései. *Iskolakultúra*, 3. sz. 3–26.
- Paivio, A. (1969): Mental imagery in associative learning and memory. *Psychological Review*, **76**. 241–263.
- Pléh Csaba (1998): *A mondatmegértés a magyar nyelvben*. Osiris Kiadó, Budapest.
- Plourde, C. E. és Besne, R. D. (1997): On the locus of the word frequency effect in visual word recognition. *Canadian Journal of Experimental Psychology*, 9. sz. 181–194.
- Selander, S. (1991): Pedagogic Text Analysis. In: Julkunen, M. L., Selander, S. és Ahlberg, M. (szerk.): *Research on texts at school*. Joensuu Yliopiston Kirjasto, Joensuu. 35–40.

Kojanitz László

Thomas, B. Fordham Institution (2004): *The Mad, Mad World of Textbook Adoption*. Thomas B. Fordham Institution, Wasington, D.C.

Wellington, J. J. (1983): A taxonomy of scientific words. *School Science Review*, **64**. 229. sz. 767–773.

## ABSTRACT

LÁSZLÓ KOJANITZ: THE ANALYTICAL STUDY OF EDUCATIONAL TEXTS: VOCABULARY

Within the field of textbook research, textbook analysis constitutes an independent research issue, turning qualitative aspects of textbooks into quantitative data, governed by pre-set rules. This paper gives an overview of the objectives, aspects and methods of the vocabulary analysis of instructional texts as well as the international and Hungarian results that may prove useful when designing such studies. The focus is on analytical research and methods regarding comprehensibility and learnability problems. Also discussed are those measurable text features that former research has proven to have the highest impact on students' reading comprehension and learning. A good instructional text, however, must be interesting as well as comprehensible. Therefore studies examining textbook vocabulary with regard to the potential to rouse and maintain student interest will also be introduced.

Magyar Pedagógia, **104**. Number 4. 429–442. (2004)

Levelezési cím / Address for correspondence: Kojanitz László, Tankönyv és Taneszköz Kutató és Fejlesztő Intézet, H–1055 Budapest, Honvéd u. 22/B