

Korpusznyelvészet és a tankönyvek

Tóth Etelka

1. A magát korpusznyelvésztként definiáló tudományágat a nyelvészet kutatási területei között a legfiatalabbak egyikeként szokás említeni. Ennek a kijelentésnek azonban inkább csak a paradigma megnevezésében, illetve az utóbbi évtizedekben alkalmazott vizsgálati módszereket és eszköztárat tekintve van létjogosultsága.

A latin eredetű *korpusz* szó egyik – számunkra fontos – alapjelentése a Bakos-féle *Idegen szavak és kifejezések kéziszótára* (1994) szerint: 'a vizsgálat során feldolgozott forrásanyag, az adatok összessége'. A Pusztai Ferenc szerkesztette *Magyar értelmező kéziszótár* (2003) továbblép ennél a meghatározásnál, és a 'valamely kérdéskörre vonatkozó írások, források összessége' értelmezés mellett megjelenik az 'adattár, rendezett adathalmaz' jelentés is. Tolcsvai Nagy Gábor 2007-ben megjelent *Idegen szavak és kifejezések szótára* jóval átfogóbb értelmezést közöl, amely kitér a *korpusznyelvészetre* is (megjegyezzük, hogy a 2000-ben kiadott *Nyelvi fogalmak kisszótárában*, amelynek társszerzője Tolcsvai Nagy Gábor, még nem szerepel ez a nyelvészeti diszciplína a címszavak között):

korpusz [lat 'test'] 4. *ny, ir* meghatározott módszerrel és előismeretekkel összegyűjtött nyelvi v. irodalmi adatmennyiség, amely a tudományos kutatás v. vizsgálat alapja 5. *ny* számítógépre vitt és elemzőprogramokkal előzetesen feldolgozott, további kutatásokra (pl. gyakorisági vizsgálatra) alkalmas, különböző szövegtípusokból gyűjtött szövegmenyiség

korpusznyelvészet ny 1. különböző szövegtípusokból gyűjtött szövegmenyiség, nyelvi adatbázis számítógépre vitele és tárolása előzetes elemző programokkal feldolgozott állapotban, további kutatási célokból 2. számítógépes nyelvi, szöveges adatbázissal dolgozó nyelvtudományi irányzat

E három különböző szótárból idézett példa – ha távoli megközelítésben is, de – hűen tükrözi a tudományág gyors fejlődését, tartalmi gazdagodását. A *korpusznyelvészet* műszo Jan Aarts és Willem Meijs nevéhez fűződik (vö. Szirmai 2005, 17), akiknek a szerkesztésében ezzel a címmel – *Corpus linguistics: new studies in the analysis and exploitation of computer corpora* – 1984-ben jelent meg egy tanulmánygyűjtemény (W1). Ebben az értelemben tekinthető tehát fiatalnak a diszciplína megnevezése. Ami pedig a módszert és az eszköztárat, valamint ezzel párhuzamosan a tartalmi megújulást illeti, az a számítógépek nyelvészeti kutatásokban való megjelenéséhez (1950-es évek vége) köthető, amely napjainkban egy valóban új tudományágban, a nyelvtechnológiában ér el mind nagyobb perspektívát kínáló eredményeket.

2. A korpusz-, pontosabban a nyelvi tényeken/adatokon alapuló nyelvelírások kezdete az ókorig vezethető vissza. Gondoljunk csak a strukturalizmus egyik korai előfutárának tartott óind Pāṇini (i. e. 4. sz.) munkásságára, amelynek jelentősége (többek között a saussure-i *langue-parole* kettősség gyökerei) a 19. századi történeti-összehasonlító nyelvészet eredményeinek tükrében, illetve valamivel később, az amerikai deskriptív nyelvészeti irányzat megalapítójának, Leonard Bloomfieldnek köszönhetően vált ismertté (Máté 1997, 42). A mai értelemben vett – ám az indulásnál számítógéppel még nem támogatott – korpuszkutatások első, nemzetközileg kevésbé ismert állomása a gépi fordítás problémáival foglalkozó, szerb Đorđe Kostić nevéhez fűződik, aki az 1950-es évek közepén kezdte meg 11 millió szövegszót tartalmazó, különböző szempontok szerint kódolt korpuszának létrehozását (e korpusz számítógépes feldolgozása azonban csak 1996-ban kezdődött el). A másik, nemzetközileg széles körökben ismert kutatás a Quirk nevéhez fűződő *Survey of English* (elérhető: <http://www.ucl.ac.uk/english-usage/archives/seu-biblio.htm>) projekt (1959) volt, amely nagy mennyiségű (200 mintából álló, egyenként 5000 szavas), írott és beszélt nyelvi szövegből kiindulva kívánta leírni a művelt brit lakosság nyelvtani szerkezeit és szóhasználatát (vö. Szirmai 2005, 60–61). Ez utóbbi projekt külön érdeme, hogy az aktuális nyelvhasználat strukturális leírásával lényegében kísérletet tett a *parole* és a *langue* szembenállásának a feloldására is. Az adatok feldolgozása kezdetben kézi erővel történt, az elektronikus változat elkészítése több mint 25 évet vett igénybe.

Az első elektronikus korpusz, amely a korpusznyelvészet térnyeréséhez vezetett, az Amerikai Egyesült Államokban 1961-ben publikált szövegekre építő, valamivel több, mint egymillió szövegszót tartalmazó – Nelson Francis és Henry Kučera vezetésével készült – Brown University Standard Corpus of Present-Day American English, rövid nevén *Brown Corpus* (W2, W3) volt, amely a szerkezete, a mintavétel nagysága és tartalma miatt a későbbekben több más korpusz (például: LOB; vö. W4) mintájául is szolgált. Az ún. nagy korpuszok nyitánya 1980-ra tehető. Ekkor indult a Birminghami Egyetem és a Collins Publishers Kiadó együttműködésének köszönhetően a John Sinclair által vezetett projekt, amelynek eredményeképpen megszületett a *COBUILD* (*Collins Birmingham University International Language Database*) korpusz, majd belőle az első korpusz alapú szótár, amelyben a szójelentések sorrendjét az előfordulásuk gyakorisága határozta meg. A *COBUILD* továbbépítése eredményezte az első, szóanyagában folyamatosan bővülő *Bank of English* (*BoE*) korpuszt, amelynek jelenlegi szóanyaga 57 millió szó (W5). Nem sokkal később három kiadó (Oxford University Press, Longman Group – UK, Chambers), két egyetem (az oxfordi és a lancasteri), valamint egy könyvtár (British Library) támogatásával kezdődött el a *British National Corpus* (*BNC*) építése, amelynek modern brit angol írott és beszélt nyelvi szövegei 100 millió szövegszót tartalmaznak (W6).

A korpuszok bemutatását sokáig lehetne folytatni, hiszen ma már minden egyes, nyelvét kodifikáló nemzet törekszik a saját nyelvi korpuszának a kiépítésére. Ehhez azonban feltétlenül hozzá kell tennünk azt is, hogy a magyar kutatók a világ élvonalában voltak és vannak e tudományterület kutatásában. Voltak, hiszen Papp Ferenc és munkatársai 1969-ben a *debreceni thésaurusszal* létrehozta egy olyan számítógépen tárolt, grammatikai adatbázist, amely kiváló alapja volt az agglutináló magyar nyelv strukturális leírásának (nem véletlenül lett ez a forrás a későbbi helyesírás-elemző programok egyik pillére). A jelentésről pedig gondoljunk csak a Magyar Tudományos Akadémia Nyelvtudományi Intézetének Nyelvtechnológiai Kutatócsoportjára, amely többek között a mai magyar írott köznyelv korpuszát, a *Magyar Nemzeti Szövegtár* gondozza (W7). Hasonlóképpen kiemelkedő a magyar irodalmi és köznyelv nagyszótárának korpusza, a *Magyar Történeti Korpusz* (W8), amely a 18 kötetesre tervezett, világviszonylatban is különlegesnek számító munkának, *A magyar nyelv nagyszótárának* (W9) az alapja. E történeti jellegű értelmező szótár a szavak jelentéstartalmát nemcsak a mai nyelvállapotnak megfelelően írja le, hanem mind címszóállományában, mind a szótározott jelentésekben és a jelentéseket illusztráló példákban túllép a szinkronia határain, és az elmúlt két évszázad szókincséből merítve a történeti dimenziót is megjeleníti.

Fontos fejlesztés a *Kárpát-medencei magyar nyelvi korpusz* is, amelynek kialakítása a Magyar Tudományos Akadémia Etnikai-nemzeti Kisebbségkutató Intézetének, valamint a Nyelvtudományi Intézet Nyelvtechnológiai Osztályának szervezésében indult 2005-ben. A korpusz sajátossága, hogy a nyelv különböző – sajtó, szépirodalom, tudományos próza, hivatalos nyelv és személyes közlések – regiszteriből merítve szövegeit, a következő megoszlásban tartalmazza a határon túli magyar nyelvváltozatokat: a szlovákiai magyar rész 4 millió, a romániai 6 millió, a kárpátaljai 3 millió, a vajdasági 2 millió szövegszó (W10).

3. A példák után következzen a korpusznyelvészeti megközelítésben a *korpusz* meghatározása: „megadott szempontok szerint válogatott, géppel olvasható, természetes nyelvi megnyilatkozások gyűjteménye, melyet automatikusan vagy fél automatikusan lehet feldolgozni és/vagy elemezni” (Balaskó 2011, 167). Tágabb értelemben a korpuszt tekinthetjük adatbázisnak is, hiszen olyan módon rendszerezett adatok gyűjteménye, amely lehetővé teszi az adatok gyors elérését, kezelését és megújítását (Szirmai 2005, 20).

E meghatározásokból következik, hogy a reprezentatív mintavétel érdekében az adatgyűjtést a kizűzött célokra megfelelően kell előzetesen megtervezni és kivitelezni. Az adatok automatikus vagy félautomatikus feldolgozásába tartozik például az, hogy a lemmatizáló program elvégzi a szövegszavak szótövesítését (pl. *nyelvek, nyelvet, nyelvnek* → *nyelv*), amely a morfológiailag sokszínű magyar nyelv esetében különösen fontos, hiszen az adatbázisban való keresés során a jelentés meghatározásához a felhasználó a szótári alak

előfordulásait szeretné megkapni különböző szövegek környezetben. Ha az elektronikusan tárolt szövegeket a korpusz építése során további információkkal (jelekkel) látjuk el, akkor a keresési lehetőségek is árnyaltabbá válnak, újabb szempontokkal gazdagodnak. Az ún. korpuszannotáció (olyan plusz információk és jelek, amelyeket az eredeti szöveg nem tartalmazott, de a korpusz készítése vagy feldolgozása során meghatározott céllal kerültek be a szövegbe) leggyakoribb eseteinek tartjuk a szófaji címkézést (*tagging*) és a szintaktikai kapcsolatok meghatározását (*parsing*), de e kategóriák mellett (és tőlük függetlenül) további elemzési kategóriákat építhetünk be annotációs elemként az adatbázisba.

A nyelvészeti szempontból hasznos keresés/lekérdezés eredménye annak függvénye, hogy mire készítjük fel az elemzőprogramot. A keresés eredménye lehet kvantitatív és/vagy kvalitatív. A legáltalánosabb és leggyakoribb outputok a következők: meghatározhatjuk az adatbázis nagyságát, a szavak hosszúságát, készíthetünk szójegyzékeket alfabetikus vagy *a tergo* (szóvégek szerinti) sorrendben, kilistázhatjuk a szövegszók, illetve a szótári szók előfordulását növekvő vagy csökkenő gyakoriság szerint, vizsgálhatjuk egyes karakterek vagy karakteregyüttesek, illetve egyes szavak vagy szókapcsolatok szövegek környezetbeli előfordulást (konkordanciák), kereshetünk szófajokra, szófajcsoportokra.

A vizsgálati szempontok annak függvényében is formálódhatnak, hogy a mintavétel módja szerint milyen korpusszal vagy korpuszokkal dolgozunk, és azokat mire akarjuk felhasználni (vö. Szirmai 2005, 32–37). Az ún. statikus korpuszok, amelyek a nyelvi tényeket mintegy pillanatfelvételnélként ábrázolják, kiválóan alkalmasak adott nyelvállapot feltérképezésére, a nyelvi szerkezetek leírására és a jelentésstruktúrák meghatározására. Ha pedig ugyanazon témában több – azonos elvek szerint annotált – statikus korpuszt helyezünk egymás mellé, akkor kiderül, hogy az adott téma milyen variációkban jelenik meg a vizsgált szövegekben.

A továbbiakban arra mutatunk be példát, hogy a korpusznyelvészet kínálta lehetőségek mennyiben segíthetnek kibontani az azonos ismeretkörben írott tankönyvek szövegeinek azonosságait és különbségeit. A bemutatásra kerülő makrometszet egy nagyobb volumenű, intelligens keresési lehetőségeket biztosító tankönyvadatbázis kiépítésének szükségességére hívja fel a figyelmet.

4. Ahogy arra már utaltunk, a korpuszok építésében fontos szerepe van a tervezésnek. Mintapéldánk a tervezési folyamat első lépéseit, a makroszint kiépítésének irányait szemlélteti.

A vizsgálat alapkritériuma az a tény, hogy a zárt rendszert képező kiinduló adatoknak – esetünkben a rendelkezésünkre álló tankönyvek ugyanazon témát feldolgozó szövegegységeinek – a témaazonosságból következően alkalmasnak kell lenniük az elemzésre, amely a későbbiekben meghatározott szempontok szerint történik. Jelen esetben a mintaadatbázis forrásanyagát három különböző, 9. osztályosoknak írott *magyar nyelv* tankönyv (Antalné–Raátz 2005,

Balázs–Benkes 2006, Misad–Simon–Szabó Mihály 2009) szövegei szolgáltatják. Ahhoz, hogy ezek a forrásszövegek valóban összevethetőek legyenek, mindegyik előtt meg kellett találnunk köztük a tematikus kapcsolódási pontokat. Ezért elkészítettük, és párhuzamba állítottuk az egyes tankönyvek témahálóját, amelyet a mellékletben közölt táblázat tartalmaz (a háló kidolgozása hallgatóimmal, Horváth Krisztinával és Varga Nórával történt korpusznyelvészeti szeminárium keretében).

A háló kibontásánál azt a címstruktúrát vettük alapul, amelyet a tankönyvek belső címhierarchiája (főcím = 1, alcím = 2, másodrendű alcím = 3) követ. Amint azonban a mellékletben látható hálóból kiderül, hogy a tankönyvekben megtalálható ismeretkörök csak részben mutatnak egymással kapcsolódási pontokat, és ezek a kapcsolódási pontok gyakran különböző szinteken jelennek meg. Ez a tény, bár úgy vélem, komoly elgondolkodásra készíthet minden tankönyvszerzőt és pedagógust, számunkra elsősorban azt jelenti, hogy a kapcsolatokat feltérképezését ki kell terjesztenünk a felmenő és a lemenő évfolyamokra egyaránt. Könnyen előfordulhat ugyanis, hogy a téma időben eltolódva, más évfolyamon jelenik meg, mint ahogy ez például nálunk is megfigyelhető a *retorika* esetében (a szlovákiai tankönyv a 9. évfolyamon, a magyarországi tankönyvek a 10. évfolyamon tárgyalják).

A táblázat színezett háttérrel jelölt elemei a közvetlen megfeleltethetőséget emelik ki. Látható az is, hogy a Magyarországon kiadott tankönyvek között több a kapcsolódási pont, de a tematikus címek egyúttal arra is utalnak, hogy a kapcsolat mélysége és hatóköre eltérő (így van ez például a *helyesírás* esetében). Ha pedig a *szöveg* témakört nézzük, akkor kiderül, hogy azzal a Szlovákiában kiadott tankönyv közel sem foglalkozik olyan mélységben, mint az anyaországbeliek. Ellenben megjelennek benne olyan témák (*kétnyelvűség, nyelvi emberi jogok*), amelyek a Magyarországon publikáltakban nem szerepelnek.

A megfeleltethető témakörök (például *kommunikáció*, illetve *helyesírás*) szövegeit a továbbiakban elektronikusan rögzítjük, és kvantitatív, illetve kvalitatív mutatókkal – egymással is összevetve – jellemezzük. A terjedelmi mutatók meghatározásán túl többek között megállapítjuk a szövegek kohézióját létrehozó kulcsszavakat és kifejezéseket, megvizsgáljuk, hogy ezek milyen környezetben, milyen összefüggésekben, milyen nyelvi példákkal igazolva jelennek meg.

Feltételezésünk szerint a módszer és az alkalmazott eszköztár alkalmas arra, hogy feltárja a tankönyvszövegek tartalmi elemeinek erősségeit és gyengeségeit, és ezáltal hozzájárulhat ahhoz, hogy a jövőben elméleti és módszertani szempontból egyaránt egységesebb tananyagok születhessenek.

Irodalom

- Antalné dr. Szabó Ágnes – dr. Raátz Judit; 2005: *Magyar nyelv és kommunikáció. Tankönyv a 9–10. évfolyam számára*. Budapest: Nemzeti Tankönyvkiadó.
- Balaskó Mária, 2011: Korpusznyelvészet. In *Nyelvészetről mindenkinek*. Szerk. Balázs Géza. Budapest: Inter, 167–170. p.
- Balázs Géza – Benkes Zsuzsa, 2006: *Magyar nyelv a gimnáziumok és a szakközépiskolák 9. évfolyama számára*.⁴ Budapest: Nemzeti Tankönyvkiadó.
- Kugler Nóra – Tolcsvai Nagy Gábor, 2000: *Nyelvi fogalmak kisszótára*. Korona Kiadó, Budapest.
- Máté Jakab, 1997: *A 19. századi nyelvtudomány rövid története I*. Budapest: Nemzeti Tankönyvkiadó.
- Misad Katalin – Simon Szabolcs – Szabó Mihály Gizella, 2009: *Magyar nyelv a gimnázium és a szakközépiskolák 1. osztálya számára*. [elektronikus kézirat Bratislava: Slovenské pedagogické nakladateľstvo – Mladé letá, spol. s r. o.
- Papp Ferenc, 2000: A debreceni thésaurusz. *Lingustica series C Relationes 11*. Budapest: MTA Nyelvtudományi Intézete.
- Pusztai Ferenc, 2003: *Magyar értelmező kéziszótár*. Budapest: Akadémia Kiadó.
- Szirmai Mónika, 2005: *Bevezetés a korpusznyelvészetbe*. Budapest: Tinta Könyvkiadó. [letölthető: http://www.korpusz.com/Monika/A_szerzorol.html]
- Tolcsvai Nagy Gábor, 2007: *Idegen szavak szótára*. Budapest: Osiris.

Internetes hivatkozások

- W1: http://books.google.com/books?id=WSvo6Lyt8ulC&hl=hu&source=gbs_similarbooks_s&cad=1
- W2: <http://khnt.aksis.uib.no/icame/manuals/brown>
- W3: <http://www.helsinki.fi/varieng/CoRD/corpora/BROWN>
- W4: <http://khnt.hit.uib.no/icame/manuals/lobman>
- W5: <http://www.mycobuild.com/about-collins-corpus.aspx>
- W6: <http://www.natcorp.ox.ac.uk>
- W7: <http://corpus.nytud.hu/mnsz>
- W8: <http://www.nytud.hu/hhc>
- W9: <http://www.nytud.hu/publ/nszt/index.html>
- W10: http://ht.nytud.hu/index.php?option=com_content&task=view&id=15&Itemid=33

Melléklet: A vizsgált tankönyvek témahálója

A jelölések feloldása:

- a: Balázs Géza – Benkes Zsuzsa 2006 *Magyar nyelv a gimnáziumok és a szakközépiskolák 9. évfolyama számára.*⁴
- b: Misad Katalin – Simon Szabolcs – Szabó Mihály Gizella 2009 *Magyar nyelv a gimnázium és a szakközépiskolák 1. osztálya számára.*
- c: Antalné dr. Szabó Ágnes – dr. Raácz Judit 2005 *Magyar nyelv és kommunikáció. Tankönyv a 9–10. évfolyam számára.*

1, 2, 3: A belső címhierarchia egyes szintjei

Korpusová lingvistika a učebnice

(Zhrnutie)

Štúdiá, vychádzajúc z definície korpusu, podáva prehľad najvýznamnejších etáp zrodu korpusovej lingvistiky. Pomocou zahraničných a domácich príkladov načrtáva teoretické a metodologické otázky disciplíny, jej súbor prostriedkov, ktorý – aj vďaka počítačovej technike – poskytuje možnosti na čoraz hlbšie a rozvetvujúce sa analýzy. V našom prípade súbor prostriedkov môže byť vhodný na vybudovanie takého paralelného korpusu, ktorého statickú databázu tvoria texty učebníc objavujúce sa v totožných predmetoch a napísaných rôznymi autormi. Tematická totožnosť poskytuje možnosť na obsahovú analýzu, na skúmanie vnútornej štruktúry textu, na odhalenie variability odborného jazyka, a to so zvláštnym zreteľom na používanie jazyka v materinskej krajine, ako aj za hranicami Maďarska. Analýzou vzorkových textov sa upozorňuje aj na to, že vzájomná prepojenosť medzi jednotlivými sériami učebníc v rámci daného učebného predmetu je v nejednom prípade spochybniteľná. Vzájomná nadväznosť medzi sériami učebníc by však bola veľmi potrebná tak z hľadiska pedagogického, ako aj metodického.

Corpus Linguistics and Textbook

(Summary)

The study reviews the most important stages of the development of corpus linguistics starting from the definition of corpus. Based on Hungarian and international examples, it presents the theoretical and methodological questions and tools of the discipline, which provide deeper and deeper and more and more varied ways of analysis thanks to computers. In our case it may be possible to build a parallel corpus whose static database is comprised of textbooks of the same subject written by different authors. Having the same subject allows for analysis of content, the examination of the inner structure of the text, the revealing of terminological variance, with special attention to the use of the language of the mother country and abroad. The analysis of samples also calls attention to the fact that in many cases transferability between textbook sets within a given subject is questionable, whereas it would be very important from both pedagogical and methodological aspects.

A belső címhierarchia egyes szintjei

a1	a2	a3	b1	b2	b3	c1	c2
magyar nyelven szóban, írásban (ismétlés)							
a kommunikáció tényezői és funkciói			információ és kommunikáció			mi a kommunikáció?	
a tömegkommunikáció	médiaműfajok			a kommunikáció			a tömegkommunikáció
	a hír				álatti kommunikáció	a médiaműfajok	
	a tudósítás			új kommunikációs létművek	embertől kommunikáció		a sajtóműfajok a rádiós műfajok
	a riport						a televíziós műfajok
	az interjú						
	a cikk						
	a kritika és az esszé						
	a jegyzet						
	a glossza						
	a sajtó nyelve és műfajai						
	a rádió nyelve és műfajai						
	a televíziós nyelv, műfajai						
a nyelvi szintek grammatikája				a nyelv mint jelrendszer		jelek és jelendszerek	
	nyelv, grammatika, nyelvi szintek				a nyelv szó jelentése	jel, jelentés, szójelentés	
		jel-, jelrendszer, nyelvi jel			nyelvi és nem nyelvi jelek	a hangalak és a jelentés kapcsolata	
		a nyelvtan modelleje			jel, jelítő és jelölt		

a1	a2	a3	b1	b2	b3	c1	c2
	a jelentéssel nem bíró nyelvi elemek (fonémák és szótagok) szintje				a jeltypusok	a beszéd és a nyelv	
		a beszédhangok meghatározása és felosztása			a nyelv mint sajátos jelrendszer	a hangok és hangtövények	
		a hang és a jelentés viszonya a nyelvben			nyelv és beszéd		a beszédhangok képzése
		a hangok előfordulásának szabályszerűségei a szavakban					a magánhangzók képzésének jellemzői
		a szótagok meghatározása és felosztása					a mássalhangzók képzésének jellemzői
	a szóelemek (morfémák) szintje					a szóelemek	
		a szóelemek (morfémák) meghatározása és osztályai					a szóelemek típusai
		a tömorfémák formai és jelentéstani tulajdonságai				a szóalkotás módjai	
		a toldalékmorfémák formai és jelentéstani tulajdonságai					a szóösszetétel
		szóelemkapcsolatok (morfémakapcsolatok) formai és jelentéstani tulajdonságai					a szóképzés
							a ritkább szóalkotási módok

a1	a2	a3	b1	b2	b3	c1	c2
		a mondat meghatározása és a mondatok osztályai					a mondatok fajtái a kommunikációs tartalom és a beszélő szándéka szerint
		az egyszerű mondatok formai és jelentéstanai tulajdonságai					a mondatok fajtái szerkezetük szerint
		az összetett mondatok formai és jelentéstanai tulajdonságai					
a szöveg fogalma, szintjei és szerkezete			a szöveg			mi a szöveg?	
	a szöveg fogalma					a szöveg és a kommunikáció	
		a mondat és a szöveg elhatárolása				a szöveg szerkezete	
		szövegmondat – szöveg					a szöveg szerkezeti egységei
		szövegmeghatározások					a szöveg szerkezettypusai
		antológia					
	a szöveg szintjei	a szöveg szintjeiről általában				a szöveg jelentése	a szöveg témája
		a szöveg pragmatikai szintje					jelentésbeli kapcsolatok a szövegben
		a szöveg nyelvi-jelentéstanai szintje					
		a szöveg stílusjegyei szintje				a szöveg grammatikája	
	a szöveg szerkezete						

a1	a2	a3	b1	b2	b3	c1	c2
	a szöveg megszoalaltatása, a szövegfonetikai eszközök	a szövegszerkezetéről általában					
a magyar helyesírás alapelvei			az írásbeliség kialakulása	írásszerek és írástíjak a magyar írásbeliség		helyesírásunk rendszere helyesírásunk alapelvei	a kiejtés szerinti írásmód a szövelemzés szerinti írásmód a hagyományos írásmód az egyszerűsítő írásmód
					magyar írásbeliség latin környezetben művelődés nemzeti nyelven a magyar helyesírás		
					szabályzatok, kézikönyvek	helyesírásunk szabályai	
					helyesírásunk alapelvei		a különírás és az egybeírás a kezdőbetű az ítélen szavak helyesírása az elválasztás az írásjelek a rövidítések és a mozaikszók helyesírása a számok helyesírása
			a kétnyelvűség				

a1	a2	a3	b1	b2	b3	c1	c2
				a beszédfrázisok			
				a (szóbeli) beszéd hagyományos részei			
				az érvelés fajtái			
					az érv felépítése		
					az érvek elrendezése		
					a kapcsolatteremtés és -fenntartás nyelvi és nyelven kívüli eszközeit		