

EGY SZÖVEGTÍPUS SZINTAXISÁNAK STATISZTIKAI VIZSGÁLATÁRÓL

B. FEJES KATALIN

0. A szövegtipológia értelmét, azaz pl. az érvelő, a leíró, az elbeszélő stb. szövegek típusonként való megkülönböztetését és jellemzését — nem is kis mértékben — statisztikai (vagy statisztikát is alkalmazó) szövegelemzések indokolják. Az ilyen elemzések irányulhatnak valamely szövegtípus egy-egy már ismert szövegszervező elvének megvalósulási módjára, pl. egyes elbeszélések időszerkezetének, egyes leírások térviszonyainak, egyes érvelések oki-okozati összefüggéseinek stb. eszközhasználati arányaira/aránykülönbségeire. Saját elemzéseim nem egy-egy szövegszervező elvet s eszközeit, hanem a szövegszerveződés grammatikai közegének felső szintjét, a szövegek szintaktikai szerveződését ölelik át. Vizsgálataim ugyanis az elemzett szövegjelenségeknek nem csupán szövegtipológiai jellemzését akarják adni, hanem a szövegalkotók életkori összefüggéseiben, azaz fejlődéslélektani tanulásaival együtt szeretnék feltárni a szövegalkotás szintaktikai vetületét.

1. Mivel e vizsgálatok több részeredménye már olvasható (B. FEJES: 1981., 1985., 1986., 1987., 1988., 1991.), itt elegendőnek vélem csupán a tárgyszavakat felsorolni. Esetünkben a szövegtípus: elbeszélés (narráció), közelebbről fogalmazások az általános iskola 4. osztályában, szerzői 9—10 éves tanulók. A fogalmazások adott témája: *A baleset*. A vizsgálat korpusza: a korosztályt országosan reprezentáló 310 db fogalmazás.

A vizsgálat apparátusát, vagyis a korpuszt alkotó elbeszélések átlagos hosszúságát, terjedelmét, kapcsoltságát, szintmélységét, szintnépességét DEME LÁSZLÓ alapvizsgálataiból vettem át (DEME: 1971.). Jóllehet külföldi műhelyekből is számos hasonló célú kutatás ismeretes, DEME LÁSZLÓ munkája nemcsak egyik legkorábbi, hanem a legalaposabb és a legkövetkezetesebb is.

Vizsgálataim arra a feltevésre épülnek, hogy 9—10 éves tanulók számára már létrehozható olyan közlési helyzet, amelyben — az elbeszélés műfajában — koherens, kohezív és konnex, azaz szövegnek minősíthető fogalmazásokat írnak. Ha ez így van, akkor — az életkorra jellemző mértékben — kölcsönös összefüggés érvényesül a szintaktikai hierarchia eszközrendszerében, abban tehát, ahogyan e 310 db elbeszélés jellemzően folyamatos és tagolt. Elemzésem a fogalmazások mint szövegművek kommunikációs-tematikai tagolódásával, azaz a 'bekezdések' számbavételével kezdődik, majd a bekezdések 'szövegmondatok'-ra (DEME LÁSZLÓ kategóriájával: 'mondategységek'-re), ezeknek 'mondategységek'-re (tagmondatokra és egyszerű mondatokra) történő tagolódásával folytatódik; végül a mondategységek 'mondatfunkciós részek'-re (mondatrészekre) s ezek 'szerkezzettagok'-ra való tagoltságának feldolgozásával fejeződik be.

Mindebben ezen mondatszerkezeti sajátágok kölcsönös összefüggésének feltárása volt célom, ennek ismeretében pedig olyan tankönyvszöveg írása, amely megfelel a

nyelvhasználat életkori sajátosságainak (vö. *Anyanyelvünk*. Tankönyv az általános iskola 2. osztálya számára. Tankönyvkiadó, 1992.).

2. Itt, a jelen dolgozatban a vizsgálat matematikai statisztikai apparátusát mutatom be, elsősorban azok számára, akiknek — hasonló munkáik során — az egyszerű százalékszámítás már nem elegendő; és azoknak, akik egy szövegtípus több sajátosságát együttesen, kölcsönös összefüggéseiben egy gyűjtött (vagy kísérletileg létrehozott) korpuszban akarják vizsgálni.

2.1. A vizsgált komponensek a számítógépre feldolgozott korpuszban 310 db elbeszélő fogalmazás: 3551 db szövegmondat, 6054 db mondategység, 20172 db tartalmas szó.

Az elbeszélések szintaktikai sajátosságainak feltárásához az alábbi komponenseket vettem figyelembe az egyes szerkezeti szintek sorrendjében haladva.

A bekezdések szintjén.* Milyen a bekezdésekre tagoltság; nem hiányzik-e, illetőleg nem indul-e fölöslegesen új bekezdés; mifélek a bekezdések közötti kapcsolásformák (a mellé- és alárendelés fajtái szerint), s ezek jelöltek-e vagy jelöletlenek?

A mondategységek szintjén figyeltem a bekezdést alkotó mondategységek közötti szerves kapcsolódást, a kapcsolásforma jelöltségét-jelöletlenségét, tartalmi-logikai fajtáját. A mondategész szövegbe ágyazottsága szempontjából azt néztem: szerves része-e a mondategész a gondolatmenetnek, vagy éppen fölösleges eleme; esetleg párbeszéd részeként épül-e a szövegbe. A mondategységek szerkezeti tagolódását aszerint is minősítettem, hogy van-e tömbösödés benne (ez természetesen csak a három vagy annál több tagmondatból álló mondategységeknek lehetett szempont), s ha van, milyen tömbösödési típusba tartozik: mellérendelő vagy alárendelő típusú-e.

A mondategységek (tagmondatok) szintjén komponensekként vizsgáltam a tagmondat szintbeli elhelyezkedését (főmondat-e, illetőleg mely szinten álló mellékmondat); a mellékmondatok elhelyezkedését a főmondatukhoz (relatív főmondatukhoz) viszonyítva (előtte, utána, illetőleg közbeékelve állnak-e). Vizsgáltam a mondategységek szerkezeti hiányosságát: okszerű vagy okszerűtlen-e a szerkezeti hiány; melyik mondatrész hiányzik, továbbá, melyeket ismétli a tanuló fölöslegesen. Figyelembe vettem a mondategységek közötti viszonyfajtákat, ezek jelöltségét-jelöletlenségét, valamint azt, hogy megfelelő kötőszót használt-e a tanuló a kapcsolat jelölésére. Vizsgáltam továbbá a mondategységeket: fölöslegesek-e vagy éppen hiányoznak a mondat tagolódása szempontjából. Külön feltüntettem, ha a szegmentum módosító szerepű, tagmondat értékű rész volt. A három vagy több mondategységből álló mondategységekben a megfelelő mondategység kapcsoltságában jelöltem a fő kapcsolat helyét, valamint figyelembe vettem, hogy van-e sajátos jelentésárnyalat a mellékmondatokban.

A mondatrészek szintjén vizsgáltam az egyes mondatrészek minőségét; megkülönböztettem az igei és névszói állítmányt; jelöltem, hogy igei természetű-e az adott mondatrész (főnévi igenévvel és határozói igenévvel kifejezett mondatrészek). Figyelembe vettem a mondatrészek halmozottságát, szerkezetességét, vagyis azt, hogy hány tartalmas szóval fejeződnek ki; a mondategységben az adott mondatrész utalószava-e a hozzá kapcsolódó mellékmondatnak, illetőleg anaforikusan utal-e az előzményekre (a mutató névmás anaforikus használata). A halmozott mondatrészekben néztem a kapcsolat jelöltségét-jelöletlenségét.

* A szint itt nem nyelvészeti terminus technicus; csupán azt jelöli meg, hogy az itt felsorolt komponenseket ebben a tartományban volt célszerű kódolni.

A szerkezettagok szintjén vizsgált komponensek lényegében megegyeznek a mondatrészekével. Különbség abban van, hogy az állítmány ezen a szinten nem jelenhet meg, csak a jelző; s az igei természet jelölésében a melléknévi igenév is helyet kap. A mondategységek szintjéhez hasonlóan jelöltem a szerkezettagok szintbeli fekvését. Ezen a szinten is vizsgáltam a halmozott szerkezettagok közötti viszonyfajtákat, kapcsolódásukat az alaptaghoz (relatív alaptaghoz); s figyelembe vettem, hogy az adott szerkezettag utalószó-e vagy pedig anaforikus elem.

2.2. A mondat szerkezeti sajátosságok körének kibővítése a számítógépben feldolgozott szövegek vizsgálatában

A részletesen feltárt szintaktikai komponensek alapján a számítógépes vizsgálat lehetővé tette, hogy a DEME LÁSZLÓtól számba vett mondat szerkezeti sajátosságokra építve ezeket tovább árnyaljam, részletesebben is vizsgáljam.

A szerkesztettségi mutató árnyalása

A szövegek (fogalmazások) mondategységeire jutó mondategységek átlagos aránya (a szerkesztettségi mutató) nem tükrözi, hogy az adott viszonzyszámban kifejezett arányokon belül milyen eltérő tagolódás van az egyes szövegekben. Vagyis: ha egy tíz mondategységből álló szöveg mondategységeinek száma 18, az átlagos szerkesztettség 1,8. Ez akkor is igaz, ha a 10 mondategységből 9 egy mondategységnyi, azaz egyszerű mondat, és a 9., 10. összetett, és így tovább. Tehát a szöveg belső tagolódási különbségeit csupán az átlagos szerkesztettségi mutatóval nem lehet megfelelően jellemezni. Ezért célszerűnek láttam benne a részesedési viszonyokat is kifejezni. A szerkesztettség belső arányait ki lehet fejezni, ha a szöveg (fogalmazás) összes mondategységéhez viszonyítjuk az egyszerű mondatok (egy mondategységnyi mondategységek) számát, ezzel megkapjuk az egyszerű mondatok mondategység-részesedési mutatóját. Ugyancsak a szövegek szerkesztettségi képét árnyalhatjuk, ha az egyszerű mondatok részesedését mint tagmondat-részesedést vesszük figyelembe, azaz a szöveg (fogalmazás) összes mondategységéhez viszonyítjuk az egyszerű mondatok mint mondategységek számát.

Az információ széttagolódása szempontjából azonban az is fontos lehet, hogy a szöveg összetett mondatai szintaktikailag hogyan épülnek fel. Ezért az általános képet tovább gazdagítja, ha a beszerkesztett tagmondatok arányát is figyelembe vesszük, azaz külön is megnézzük a szöveg összetett mondatainak az átlagos szerkesztettséget: a szöveg összetett mondataira jutó mondategységek átlagos arányát.

A szintagmák átlagos szintmélysége

Vizsgálatomban a szintmélységet nemcsak a mondategységekre vetítve néztem meg, hanem a szintagmákban is. E mélységi mutató elvi megalkotása megegyezik DEME LÁSZLÓ szintmélység-mutatójának kiszámítási módjával. Gondot az okozott, hogy miként tudnám megőrizni, s a képletben is érvényesíteni azt a minőségi különbséget, amely a szintagma abszolút alaptagja mint mondatfunkciós rész és determinánsa(i) között fennáll. Végül úgy döntöttem, hogy a szintjelző számozást nem a mondatfunkciós rész szintjén kezdem, hanem az első determináns megjelenési

szintjén, vagyis az alárendelés első szintjén. Az általános képlet tehát így alakult a szintagmák mélységmutatójában:

$$\frac{Sx1 + S1x2 + S2x3 + \dots + Snxn}{S + S1 + S2 + S3 + \dots + Sn}$$

(Az 1-n a szintek jelzőszáma, S—Sn az adott szinten talált szerkezettagok mennyisége.)

A szintagmák mélységmutatójának kiszámításához természetesen szükség volt szintnépességük vizsgálatára is. Ezért néztem azt is, hogy az összes szerkezettag közül mennyi található az alárendelés első szintjén, illetőleg az alsóbb szinteken, azaz a teljes mennyiségnek hány százaléka kapcsolódik mondatfunkciós részhez, s hány százalék található az alsóbb szinteken.

A mondategységek telítettségi, tagoltsági és zsúfoltsági mutatójának szintenkénti vizsgálata

Ez a vizsgálat azt jelenti, hogy külön vizsgáltam az egyszerű mondatoknak mint mondategységeknek, a beszerkesztett helyzetű, főmondati szintű tagmondatoknak és az alárendelt tagmondatoknak a telítettségét, tagoltságát és zsúfoltságát. Érdekelt, hogy a mondategységekbe való beszerkesztettség tükröződik-e a tagmondatok megszerkesztettségén, s ez miben jelentkezik.

A mondategységek szerkezeti kérdései

Eddigi vizsgálataimban is néztem a mondatrészek zsúfoltságán belül a szerkezetes mondatrészek zsúfoltságát. A számítógépes vizsgálat azt is lehetővé tette, hogy külön is nézzem az igeneves mondatrészek szerkezeti sajátosságait. Kísérletképpen arra is választ kerestem, hogy változik-e a szerkezetes mondatrészek zsúfoltsága, ha e mutatót szintbeli fekvése szerint is figyelembe veszem. Tehát külön vizsgáltam az egyszerű mondatoknak, a főmondati szintű beszerkesztett tagmondatoknak és az alárendelt mondatok mondatrészeinek a zsúfoltságát.

A megoszlási mutató egyszerűsítése

Ha eltekintünk attól a véleményről, hogy a mellérendelő kapcsolást egyszerűbbnek, az alárendelőt pedig bonyolultabb kapcsolásnak tekintjük, akkor egyszerűsíthetjük a megoszlási mutató kiszámítását is. A megoszlási mutató általános képlete így a

következőképpen alakul ki: $M = \frac{Ca}{Ca + Cm}$ (M=a megoszlási mutató; Ca=az alárendelő

kapcsolások száma; Cm=a mellérendelő kapcsolások száma). Az eredmény végül is ugyanaz lesz, csak például nem 1,48 lesz a megoszlási mutató értéke, hanem 0,48. Megfogalmazhatjuk a kapott eredményt úgy is, hogy az összes kapcsolás 48%-a az alárendelő, és 52%-a a mellérendelő.

Egyéb viszonyítási lehetőségek a számítógépes vizsgálat során

A szövegek mondat szerkezeti sajátosságainak gépi vizsgálata lehetővé tette, hogy minden fogalmazást mint önálló szöveget is figyelembe vehessék. Így például a mondatrészek esetében viszonyítási alap lehetett az adott szöveg összes mondatrésze. Ehhez a mennyiséghez viszonyíthattam a szerkezetes mondatrészek, illetőleg az igenes mondatrészek mennyiségét; valamint azt is az összeshez viszonyíthattam, hogy hány mondatrész található az egyszerű mondatokban, a főmondati szintű beszerkesztett tagmondatokban és a mellékmondatokban; végül ugyanezt megnézhettem a szerkezetes mondatrészek esetében is. Így már a részesedési mutatókban kifejezett szerkezeti jellemzőket össze lehetett hasonlítani szövegenként.

Hasonló módon egybevettem az egyes szövegek szerkezeti sajátosságait a szerkezettagok szintjén is.

A szövegek tartalmasszó-mennyiségéből kiindulva is összehasonlítható részesedési mutatókhoz juthattam. Néztem azt, hogy az egyes fogalmazások mint szövegek összes tartalmas szavából mennyi található az egyszerű mondatokban, a főmondati szintű beszerkesztett tagmondatokban, illetve a mellékmondatokban.

A vizsgált összes komponenst figyelembe véve a számítógép a következő jellemzőket adhatta meg az egyes szövegekről:

- az átlagos szerkesztettségét,
- az összetett mondatok átlagos szerkesztettségét,
- az egyszerű mondatok a mondategész-számhoz viszonyított arányát,
- a szövegek bonyolultságát (külön az összetett mondatokat is figyelembe véve),
- a kapcsolások megoszlási arányát,
- a főmondati szintű tagmondatok szintnépességét (az egyszerű mondatokét és a beszerkesztett helyzetű főmondatokét),
- az alsóbb szintek népességét,
- az egyszerű mondatoknak mint mondat egységeknek a népességét,
- a főmondati szintű beszerkesztett mondat egységek népességét,
- a mondat egységek, mondatrészek átlagos telítettségét,
- a mondat egységek és mondat egységek átlagos tagoltságát,
- a mondat egységek és mondatrészek átlagos zsúfoltságát,
- a szerkezetes mondatrészek és az igenévi alaptagú szintagmák átlagos zsúfoltságát,
- a mondat egységek és a szerkezettagok szintbeli fekvése alapján számított átlagos mélységmutatót,
- az egyszerű mondatok, a főmondati szintű beszerkesztett mondat egységek, valamint az alsóbb szintű mondat egységek átlagos telítettségét, tagoltságát, zsúfoltságát,
- az egyszerű mondatok, a főmondati szintű beszerkesztett mondat egységek és a mellékmondatok mondatrészeinek és szerkezetes mondatrészeinek átlagos zsúfoltságát,
- a mondat egységekre és mondat egységekre jutó szerkezettagok átlagos mennyiségét,
- a szerkezetes mondatrészek részesedését a teljes mondatrészmennyiséghez viszonyítva,
- az igenes mondatrészek részesedését a teljes mondatrészmennyiséghez viszonyítva,
- az egyszerű mondatokra, a főmondati szintű beszerkesztett mondat egységekre és az alárendelés első szintjén álló mondat egységekre jutó tartalmas szavak részesedését a teljes tartalmasszó-mennyiséghez viszonyítva,

- az egyszerű mondatokra, a főmondati szintű beszerkesztett mondat egységekre és az alárendelés első szintjén álló mondat egységekre jutó mondatrészek, szerkezetes mondatrészek és szerkezettagok részeseződését a teljes mondatrészmennyiséghez, a teljes szerkezetes mondatrészmennyiséghez és a teljes szerkezettag-mennyiséghez viszonyítva.

2.3. A kutatás módszere

Az egyes szintaktikai komponensek meglétét és kölcsönös összefüggéseit a fogalmazások (s így a gyerekek nyelvhasználata) mennyiségi tulajdonságainak tekintjük; a kutatás módszere tehát magától értetődően statisztikai. Az adatok matematikai statisztikai elemzését a JATE Kalmár László Kibernetikai Laboratórium számítógépén végeztük el. Az anyag számítógépes előkészítése a közösen kialakított kódolási szisztema alapján magára a kutatóra várt, a programtervezést a laboratórium két munkatársa, Vas Zoltán tudományos munkatárs és Tóth Éva programtervező matematikus végezte el. A statisztikai elemzéshez a laboratórium az OSIRIS statisztikai programokat használta fel (Judith Rattenbury — Neal Van Eck: OSIRIS architecture and design. Survey Research Center. Institute for Social Research. The University of Michigan. Ann Arbor, Michigan 1973.) kiegészítve az ilyen típusú adatstruktúrákat kezelő programokkal.

A továbbiakban részletesen ismertetem az anyag előkészítését és a kódolás módját.

2.3.1. Az anyag előkészítése

A szövegek előkészítésében a feltáró vizsgálatból ismert (FEJES: 1981.) feldolgozási módot követtem. Minden egyes fogalmazásról készült egy munkalap. (Ezt a feladatot munkatársaim segítségével végeztem el (Köröcsi Eszter, Mucsányi János, Nagy L. János), hozzá anyagi támogatást kaptam a 6-os főirány keretéből.) Mi került fel egy-egy munkalapra? Az adott szöveg szintaktikai elemzése (lexikai-szemantikai kiegészítésekkel) a DEME-féle lineáris leírással. Ennek bemutatására itt olyan szöveget választottam, amelyben sokféle szintaktikai-lexikai sajátosság lejegyzésére van példa.

Baleset látam

- (1) Ez a baleset, aminek én szemtanúja voltam, november 10-én történt. (2) Egy hűvös, őszi reggelen anyuka elküldött a boltba, hogy vegyek 1 kg kenyeret. (3) Amint elindultam, hirtelen egy fiú száguldott el mellettem versenybiciklivel. (4) Utána kiáltottam, hogy ne menjen olyan gyorsan, mert ebből baleset következhet. (5) De ő rám sem hederített, és a meredeken egyenesen az úttestre hajtott, ahol éppen egy kocsi állt meg. (6) A fiú az autónak a közepében kötött ki. (7) A kocsinak a vezetője, ahogy meglátta a betört üveget és a fiút, akinek már csupa vér volt a keze, és az összetört biciklit, azt hitte, hogy szívbajt kap. (8) Rögtön szaladt hívni a mentőket. (9) A járőrelők mind odarohantak, és nézik a vérben fekvő fiút, aki segítségért kiáltozott. (10) A mentők hamar megérkeztek. (11) Elővették a hordágyat, és ráfektették a fiút. (12) Az orvos megvizsgálta. (13) A keze van neki eltörve, mondta, és beszállt a mentőautóba. (14) A nagy tömeg lassan szétoszlott. (15) De gondolom, sokan elmondták otthon, amit láttak az országúton. (16) Engem is annyira meglepett ez az eset, hogy el is felejtettem megvenni a kenyeret. (17) Szaladtam haza, és elmondtam anyukámnak ezt a háborzongató esetet. (F. K. Harta/4., 4. b osztály, 1980.)

A fogalmazások szintaktikai komponenseinek s a komponensek viszonyainak ábrázolása. (Jelmagyarázat: (F)= főmondat, ill. mondategész értékű egyszerű mondat; (m₁)=első mélységszintű mellékmondat, (jm₁)=jelzői, (jk₁)=kijelölő jelzői stb. mellékmondat. Az egyes mondatrészek és szerkezettagok jelölésére pl.: Á=állítmány, Hh=helyhatározó stb.; a ferde vonal az alárendelő viszonyt jelöli, pl.: (jb₁\Hh), a vízszintes vonal a szinttartó (mellérendelő, ill. közvetett heterogén) viszony jele, pl.: (Hh — Hh), (Háp — Hr). Egy szegmentum közbeékelődöttségét az ék alatti helyzet fejezi ki.)

Az összetett mondatokat kétféleképpen: tagmondati, valamint ezek mondatrészi/szerkezetrészi összefüggéseiben ábrázolom. A fenti elbeszélés számítógépre került komponensei tehát az alábbiak:

- (1) $(F \widehat{jm_1} F; (\widehat{jk_1 \setminus A} \overset{an}{Hr} - A \setminus \overset{\ddot{o}}{A^2} Hi \setminus \overset{\ddot{o}}{A}))$
- (2) $(F/hc_1); \left(((jm_1 + jm_1 \setminus Hi) - A \setminus \overset{\ddot{o}}{A} / Hh) /_{hogy} ([A] \setminus \overset{\ddot{o}}{A} / (jme_1 \setminus T)) \right)$
- (3) $(hi_1 \setminus F); (\underset{amini}{Hi} \setminus \overset{\ddot{o}}{A} / [A] / (Hm - A \setminus \overset{\ddot{o}}{A^2} / Hm - Hm))$
- (4) $(F/hc_1 / hoc_2); ((Hh \setminus \overset{\ddot{o}}{A} / [A]) /_{hogy} (\overset{\ddot{o}}{A^2} / (hfm_1 \setminus Hm) - A) /_{mer} (\overset{an}{Her} - A \setminus \overset{\ddot{o}}{A}))$
- (5) $De (F \text{ és } (F/hh_1)); \left((A - H\overset{\ddot{o}}{a}n \setminus \overset{\ddot{o}}{A^2}) \text{ és } ((Hh - Hm - Hh - [A] \setminus \overset{\ddot{o}}{A}) /_{ahol} (Hh - Hi - A \setminus \overset{\ddot{o}}{A^2})) \right)$
- (6) $(F); (A - (jb_1 \setminus Hh) \setminus \overset{\ddot{o}}{A^2})$
- (7) $(F / \widehat{hi_1} \widehat{jm_2} \widehat{hi_1} \setminus F) / (t_1); \left(((jb_1 \setminus A) /_{ahogy} (Hi - A \setminus \overset{\ddot{o}}{A} / (jm_1 \setminus T) \text{ és } T) /_{akinek} (\overset{ig}{Hr} - Hi - Hfm \setminus \overset{\ddot{o}}{A^2} / A) \text{ és } (jm_1 \setminus T) \setminus (T \setminus \overset{\ddot{o}}{A} / A) /_{hogy} (T - [A] \setminus \overset{\ddot{o}}{A})) \right)$
- (8) $(F); ((Hi - [A] \setminus \overset{\ddot{o}}{A} / (Hc/t_1))$
- (9) $(F \text{ és } (F /_{aki} (jm_1))); ((A - H\overset{\ddot{o}}{a}p \setminus \overset{\ddot{o}}{A} / [Hh]) \text{ és } (A \setminus \overset{\ddot{o}}{A} / hh_2 \setminus jm_1 \setminus T) /_{aki} (A - Hc \setminus \overset{\ddot{o}}{A}))$
- (10) $(F); (A - Hi \setminus \overset{\ddot{o}}{A})$
- (11) $(F \text{ és } F); ((\overset{\ddot{o}}{A} / [A] - T) \text{ és } ([A] \setminus \overset{\ddot{o}}{A} / T - Hh))$
- (12) $(F); (A - [T] \setminus \overset{\ddot{o}}{A})$
- (13) $((t_1 \setminus F) \text{ és } (F)); \left(((A \setminus \overset{\ddot{o}}{A} / H\overset{\ddot{o}}{a}p - Hr) \setminus (\overset{\ddot{o}}{A} / A)) \text{ és } ([A] \setminus \overset{\ddot{o}}{A} / Hh) \right)$
- (14) $(F); ((jm_1 \setminus A) - Hm \setminus \overset{\ddot{o}}{A})$
- (15) \checkmark

(16) $De (F/(t_1 /_{ami} t_2)); \left(([A] \setminus \dot{A}) / \left((H \dot{A} p \setminus \dot{A} / H h - H r) /_{ami} (T - A \setminus \dot{A} / H h) \right) \right)$

(17) $(F /_{hogy} (h f m_1)); \left((T - H f m \setminus \dot{A} / (j k_1 \setminus A)) /_{hogy} \left(([A] \setminus \dot{A}^2 / (T / t_1)) \right) \right)$

(18) $(F \text{ és } F); \left(([A] \setminus \dot{A} / H h) \text{ és } ([A] \setminus \dot{A} / H r - j k_1 \setminus (j m_1 \setminus T)) \right)$

A példaszöveg felszíni szakaszolódása:

$((1) - ((2 - 3) - 4) De ((5 - 6) - (7 - 8) - 9)) - ((10 - 11) - (12 - 14)) \sqrt{([15] De (16 - (17 - 18))}$

(A hiányjel arra utal, hogy a tanuló nem tagolta új bekezdésbe az új témát.)

Az anyag előkészítése során többször okozott gondot bizonyos tartalmi-logikai, szemantikai kérdések eldöntése: jelöletlen mellérendelő viszonyok megítélése, határozók szemantikai besorolása, vonatkozó mellékmondatok kérdése stb., szintaktikailag azonban ezek nem jelentettek problémát: mellérendelő viszony, mondat-funkciós rész, mellékmondat stb.

2.3.2. Az anyag kódolása

A fent leírt módon minden szövegről készült szintaktikai-lexikai séma. Fáradtságos munka volt elkészíteni ezt a lineáris szerkezeti leírást, de a kódoláskor nagy hasznát vettem. A kódolás alapján ugyanis legcélszerűbbnek és legegyszerűbbnek az ún. *fastruktúra* szerinti leírás kínálkozott. Ez azt jelentette, hogy a magasabb szint felől haladtunk a széttagoltság irányába. A csúcspontok felsorolása után az ún. *inorder* bejárást választottuk, vagyis mindig a bal oldalt jártuk végig, s úgy néztük a további elágazásokat. Így végeredményben a fastruktúra linearizálása sok vonatkozásban megegyezett az előkészítés menetével. Az eltérésekről s ezek megoldásáról a későbbiekben még szólok.

A kódoláskor minden szükséges információ kódértéket kapott. Az egyes szerkezeti ismétlések leírására — természetesen — azonos kódértékeket használtam.

2.3.2.1. Az egyes szerkezeti szintek kódértékei:

A bekezdések szintjén: *B

— hiányzik az új bekezdés: V

— a bekezdések közötti tartalmi-logikai viszony fajtái:

kapcsolatos	11
választó	12
következtető utótagú	13
magyarázó utótagú: kifejtő	14
okadó	15
ellentétes: szembeállító	16
megszorító	17
kizáró	18

— a tartalmi-logikai viszony jelöltsége esetén: JM

— fölösleges a bekezdés: FS

A mondatégeszek szintjén: *M

— a mondatégesz nem illeszkedik szervesen a gondolatmenetbe: V

— a mondatégeszek közötti kapcsolás fajtái:

kapcsolatos	11
választó	12
következtető utótagú	13
magyarázó utótagú: kifejtő	14
okadó	15
ellentétes: szembeállító	16
megszorító	17
kizáró	18

— a mellérendelés jelöltsége esetén: JM

— ha a mondatégeszek között nem mellérendelés van:

burkolt alárendelés BA

burkolt fölérendelés BF

— a nem mellérendelés minősége:

állítmányi	20
alanyi	21
tárgyi	22
határozói: hely	23
idő	24
mód	25
állapot	26
állandó	27
jelzői: mennyiség és minőség	28
birtokos	29
értelmezői	30

— ha a nem mellérendelés jelölt: JA

— a mondatégesz párbeszéd része: P

— tömbösödés van a mondatégeszben:

alárendelő típusú mondatégeszek:

$yF/m = A1$	$yF/ym = A5$
$F/xm = A2$	$yF/xm = A6$
$F/ym = A3$	$xF/xm = A7$
$F/zm = A4$	$xF/ym = A8$

mellérendelő típusú mondatégeszek:

$yF-F = M1$	$xF-yF = M5$
$F-yF = M2$	$yF-yF = M6$
$xF-F = M3$	$xF-xF = M7$
$F-xF = M4$	$yF-xF = M8$

— a gondolatmenet szempontjából fölösleges az egész mondat: FS

A mondatégesz szintjén: *T, *T1, *T2, ...

— a mondatégesz jelentésárnyalata:

hasonlító	H
feltételes	F
következményes	K
megengedő	M
feltételes hasonlítás	S

- a mondat egység párbeszéd része: P
 - a mondat egység helyzete a főmondat hoz képest:

utána áll	U
előtte áll	E
közbeékel	B
 - a mondat egység kapcsolása a mondat egészen belüli fő kapcsolás helye: —
 - a mondat egység hiányossága:

okszerű	OK = ...;
okszerűtlen	OT = ...;
 - a mondat egység valamely eleme fölöslegesen ismételt: IS = ...;
 - az alárendelés jelöltsége esetén: JA
 - a mellérendelés jelöltsége esetén: JM
 - a viszony fajtái: közömbös IO

mellérendelés (lásd az előzőekben)
alárendelés (lásd az előzőekben)
 - rossz kötőszó a viszony jelölésére: R
 - fölösleges az egész tagmondat: FS
 - a nyelvi elem módosító szerepű, tagmondat értékű rész: MO
 - a gondolatmenet szempontjából hiányzik egy tagmondat: V
- A mondatrészek szintjén: *R**
- a mondatrész fajtája:

állítmány: igei	LI
névszói	LN
alany	A
tárgy	T
határozó: hely	HH
idő	HI
mód	HM
állapot	HP
állandó	HA
 - a mondatrészt igenév fejezi ki:

főnévi igenév	IF
határozói igenév	IH
 - a mondatrész halmozott: Y

(A halmozás fajtája a mellérendelés kódszámait kapja.)
 - a mondatrész egynél több tartalmas szóból áll: S2, S3 stb.
 - a mondatrész utalószó: UT
 - a mondatrész mutató névmási anaforikus elem: F
- A szerkezettagok szintjén: *S, *S1, *S2, ...**
- a szerkezettag igenévvel fejeződik ki:

melléknévi	I
főnévi	IF
határozói	IH
 - a szerkezettagok között jelölt mellérendelés van: JM
 - a szerkezettagok közötti viszonyfajták: közömbös, mellérendelő, alárendelő.

(Részletesen lásd a mondat egységeknél!)

Jelzői alárendelés esetén a névmási kijelölő jelző kódszáma: 31.

- a szerkezettag egynél több tartalmas szóból áll: S2, S3 stb.
- a szerkezettag utalószó: UT
- a szerkezettag mutató névmási anaforikus elem: F

2.3.2.2. A kódolás legáltalánosabb szabályai:

- a fastruktúra csúcspontjait jelző kódértékek semmiképpen sem hagyhatók el, kivétel a mondategységet jelző *T abban az esetben, ha a mondategész egyetlen mondategységből áll;
- az adott egységre (szövegegység, mondategész, mondategység, mondatrész, szerkezettag) vonatkozó információk zárójelben állnak az egységet jelölő szimbólum (*B, *M, *T, *R, *S) után;
- a zárójel elhagyható, ha nem kerül bele információ;
- a zárójelen belül a kódérték sorrendje tetszőleges, a kódértékeket vesszővel kell elválasztani egymástól.

Álljon itt szemléltetésül a már bemutatott tanulói szöveg kódolása:

*B*M*T*R(A)*S(28,F) *R(HI,S3) *R(LI) *T(B, JA, 28) *R(HA)
 *R(A) *R(LN, S2)

*B(14) *M *T *R(HI) *S(Y,28) *S(11,28) *R(A) *R(LI) *R(HH)
 *T1(JA, 25,U,OK=A) *R(LI) *R(T) *S(28,S2)
 *M(11) *T *R(HI) *R(A) *R(LI,S2) *R(HH) *R(HM)
 *T1(JA,24,E,OK=A) *R(HI) *R(LI)
 *M(11,A2) *T(OK=A) *R(HH) *R(LI) *T1(JA,—25,U,OK=A)
 *R(LI,S2) *R(HM) *S(25) *T2(JA,25,U) *R(HP,F) *R(A) *R(LI)
 *M(JM,16,M4) *T *R(A) *R(HA) *R(LI,S2) *T(JM,—11,OK=A)
 *R(HH) *R(HM) *R(HH) *R(LI) *T1(JA,30,U) *R(HH) *R(HI) *R(A)
 *R(LI,S2)
 *M(11) *R(A) *R(HH) *S(29) *R(LI,S2)
 *M(11,A4) *T *R(A) *S(29) *R(T,UT) *R(LI) *T1(JA,24,B,OK=A)
 *R(HI) *R(LI) *R(T,Y) *S(28,I) *R(T,JM,11) *S(28,I)
 *T2(JA,28,B) *R(HA) *R(HI) *R(HM) *R(LN,S2) *R(A)
 *T1(JA,22,U,10,OK=A) *R(T) *R(LI)
 *M(11,OK=(A) *R(HI) *R(LI) *R(HM,IF) *S(22)
 *M(11,M4) *T(OK=HH) *R(A) *R(HP) *R(LI) *T(JM,—11,OK-A)
 *R(LI) *R(T) *S(28,I) *SL(23) *T1(JA,28,U) *R(A) *R(HM) *R(LI)

*B(11) *M *R(A) *R(HI) *R(LI)
 *M(11) *T(OK=A,HH) *R(LI) *R(T) *T(JM,11,OK=A,HH) *R(LI)
 *R(T)
 *M(11,OK=T) *R(A) *R(LI)
 *M(11,M3) *T(OK=A,HA) *R(LI) *T1(22,E) *R(A) *R(LI) *R(HA)
 *R(HP) *T(JM,—11,OK=A) *R(LI) *R(HH)
 *M(11) *R(A) *R(LI) *R(HH,S2) *R(HH)

*B(V) *M(OK=HH) *R(A) *S(28) *R(HM) *R(LI)
 *M(JM,17,A2) *T(OK=A) *R(LI) *T1(—22,U,OK=A) *R(HP)
 *R(LI) *R(HH) *T2(JA,22,OK=A) *R(T) *R(LI) *R(HH)

*M(15) *T *R(T) *R(HM,UT) *R(LI) *R(A) *S(31,F)
 *T1(JA,25,K,OK=A) *R(LI,S2) *R(T,IF) *S(22,F)
 *M(18) *T(OK=A) *R(LI) *R(HH) *T(JM,11,OK=A) *R(LI)
 *R(HA) *R(T) *S(28,I) *S(31,F,10)

2.3.2.3. Az ábrázolási és kódolási mód közötti főbb különbségek

A fastruktúrára épülő kódolási mód és a DEME LÁSZLÓ által kidolgozott (a szinteződést és tömbösödést egyaránt kifejezni képes) szerkezeti leírás között alapvetően két lényeges eltérés mutatkozott, mégpedig a sorrend és a tömbösödés kódolásában.

2.3.2.3.1. A sorrendben jelentkező kódolási eltérések

Az alaptag és a determináns sorrendje

A magyar nyelvben a szintagmatikus építkezés jobbról balra halad: beszédben-írásban előbb jelenik meg a determináns, mint az alptag. A fastruktúrának megfelelő (függőségi) kódolásban azonban egy szerkezettag (*S) megjelenése szükségszerűen feltételezi az előtte álló mondatrészt (*R). Ez elsősorban azzal a következménnyel jár, hogy a kódolásban eltűnik a főnévi és az igenévi alptagú szintagmák közötti szintaktikai különbség.

Példával szemléltetve:

... <i>el (akartam) menni a kiserdőbe...</i> :	*R(T,IF,S2)	*S(23)
<i>a kiserdőbe (akartam) elmenni:</i>	*R(T,IF)	*S(23)
... <i>egy őszi reggelen...</i> :	*R(HI)	*S(28)

Mivel azonban szórendi, sorrendi kérdéseket (így aktuális tagolást) sem vizsgáltam, nem kellett vállalnom e különbség kódolásának megoldását.

A főmondat és a mellékmondat sorrendje

Helyzetét, beépültségét tekintve a mellékmondat állhat főmondata előtt, mögött és beékelve. A függőségi ábrázolásban azonban előbb kell a főmondatot kódolni, mint a mellékmondatokat. Ezért külön kódértéket kell adni a mellékmondatoknak helyzetük szerint. Például:

Az $m \setminus F$ tagolású mondategész kódolása: *T... *T1(E,...), azaz a T1 szintű tagmondat a főmondata előtt (E,...) áll;

Az $F \setminus m \setminus F$ tagolású mondategész kódolása: *T... *T1(B,...), azaz a T1 szintű tagmondat a főmondatba be van ékelődve (B,...);

Az F/m tagolású mondategész kódolása: *T... *T1(U,...), azaz a T1 szintű tagmondat a főmondat után áll (U,...).

2.3.2.3.2. A tömbösödésben jelentkező kódolási eltérések

A gondolatrészek szorosabb-lazább tartalmi-logikai kapcsolódásait a szerkezet ábrázolásával (zárójelk használatával) ki tudjuk fejezni, de a tömb nem tekinthető olyan szerkezeti egységnek, amely minden mondategységben, illetőleg nagyobb vagy kisebb szerkezeti egységekben minden esetben jelen van. Azaz nem építhető be a fastruktúrába állandó csúcspontnak, ezért ennek a kódolása is gondot okozott.

Tömbösödés a mondatrészek és szerkezettagok szintjén

Mivel a szerkezettagi és a mondatrészi halmozottságon belüli tömbösödés a 10 éves tanulók írásbeli nyelvhasználatában még nem számottevő jelenség, a változók számát nem szaporítottam vele, hanem egyszerűen csak kódoltam, hogy az adott determináns milyen viszonyban áll a fölrendeltjével, és megadtam az előtte álló halmozott taghoz való kapcsolatát. Alábbi példám a szintagmák köréből választottam, a mondatrészek szintjén ugyanis nem találtam jellemző példát. (Csupán néhány esetben szerepelt három mellrendelt mondatrész, de a közöttük levő viszony kapcsolatos volt.)

Tömbösödés a szintagmák szintjén: *bájos, szép, de figyelmetlen kislányt...*

Ábrázolva: $((j_m, +j_m) \leftrightarrow j_m) \setminus T$

A kódolásban: *R(T) *S(28,Y) *S(28,11) *S(28,JM,16)

Gyakoribb eset volt a szintagmák szintjén az a jelenség, hogy a determináns egy, már szerkezetes alaptaghoz kapcsolódott. Részletesen bemutatott szövegünkben is találunk ilyen esetet (az utolsó mondatban): *Szaladtam haza, és elmondtam anyukámnak ezt a hátborzongató esetet.*

Ábrázolva: $j_k \setminus (j_m \setminus T)$ (...ezt a hátborzongató esetet...)

Kódolva: *R(T) *S(28) *S(31,F,10)

Vagyis a kódolásban a szerkezettagok sorrendjével fejeztem ki, hogy melyik determináns kapcsolódott az alaptaghoz előbb. Ez a típus az egymással közömbös viszonyban álló determinánsok esete. Ide soroltam még a következő szintagmákat is: *szomszédunk kétéves kislány...*

Ábrázolva: $j_b \setminus (j_m \setminus A)$

Kódolva: *R(A) *S(28) *S(29,10)

néhány perc múlva a helyszínre érve

Ábrázolva: $(j_m \setminus h_i) \setminus h_h \setminus H_i$

Kódolva: *R(HI, IH) *S(23) *S(24,10) *S1(28)

(Egy megjegyzés. A fenti szintagmák elemzését a hagyományos módon végeztem el. A feldolgozás során azonban egyre jobban megerősödött bennem az a felismerés, hogy ezek a típusok nem azonosak. El kellene különíteni az igenévi alaptagú szintagmákban található közömbös viszonyban álló (KÁROLY SÁNDOR megnevezésével: 'közvetett heterogén viszonyú') szerkezettagokat a névszói alaptagú szintagmákban jelentkezőktől. Szinteződésük alapján ugyan valóban lehetnek egymással közömbös viszonyban a névszói alaptagú szintagma determinánsai is, viszont a kettő között lényeges a különbség abból a szempontból, hogy az igenévi alaptagú szintagmába azonos transzformációs lépéssel kerül a két determináns alsóbb szerkezeti szintre, míg a névszói alaptagú szintagmákban két külön transzformációs lépés eredményeként kerül

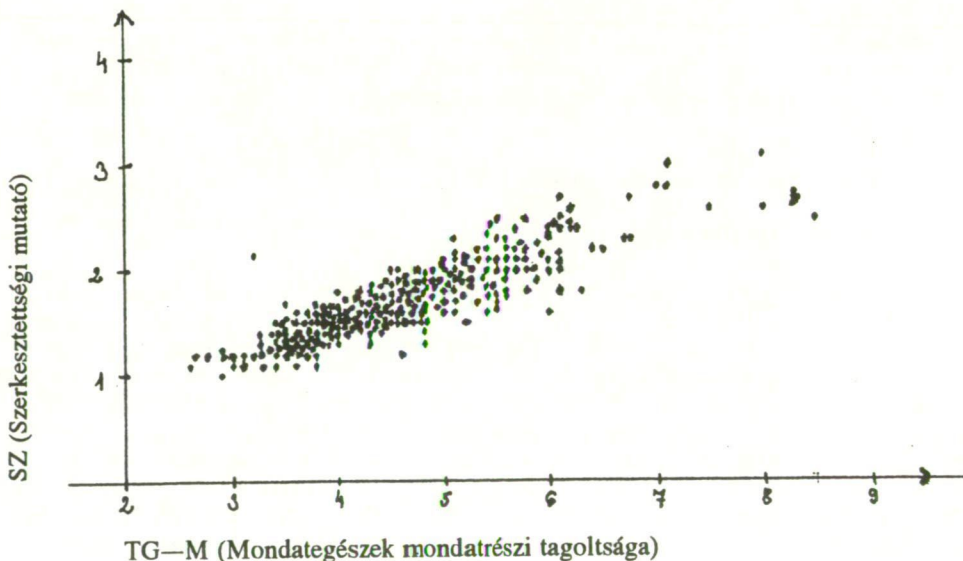
a két determináns (pl. a birtokos jelző és a minőségjelző) szerkezettagi szintre. Elkülönítésüket vizsgálatomban a szófaji háttér kódolásával valósítottam meg.)

Tömbösödés a kettőnél több mondategységre tagolódó mondategységek szintjén

A mondategész szerkezeti vizsgálatában már fontos szempont a tömbösödés. Ebben az életkorban ugyanis a gyerekek már viszonylag sok összetett mondatot használnak. Lényegesek tehát a felvetődő kérdések: Hogyan tömbösödnek a tanulók összetett mondatai? Van-e összefüggés a fő kapcsolás minősége és a tömbösödés típusa között? stb. Mivel azonban a tömbösödést külön nyelvi elem itt sem jelöli, kódoláskor ugyancsak sajátos megoldást kellett választanom. Sorra vettem a várható szerkezeti variációkat, s ezeknek típusonként megkülönböztető kódszámot adtam. Ez a kódérték a mondategészre vonatkozó információként szerepelt: *M(M4). Hogy a fő kapcsolás minőségét is ismerjem, a viszonyfajta kódszáma elé „-” (mínusz) jelet tettem, jelezve, hogy az adott mondatkonstrukcióban itt van a fő kapcsolás. Például a bemutatott szöveg 13. mondategységében: *M(11,M3) *T(OK=A,HA) *R(LI) *T1(22,E) *R(A) *R(LI) *R(HA) *R(HP) *T(JM,-11,OK=A) *R(LI) *R(HH)

3. Az egyes szövegek mondatszerkezeti jellemzőinek összefüggésvizsgálata

Ha az egyes fogalmazások mondatszerkezeti jellemzőit mint változókat tekintjük, akkor vizsgálhatjuk, van-e korreláció a változók között; ha van, akkor milyen irányú (a változók azonos irányban, illetőleg ellentétes irányban változnak-e), továbbá mennyire szoros a változók kapcsolata.

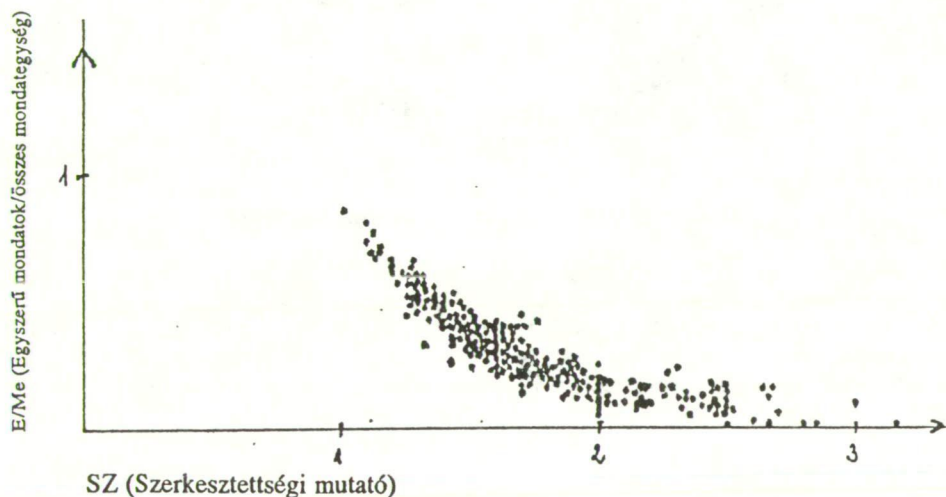


1. ábra

Ez a vizsgálat azt jelenti, hogy minden olyan két mondatszerkezeti jellemzőről, amelyek között korrelálást sejtünk, pontdiagramot készítettünk. Lineáris korrelálás esetén a pontok egyetlen egyenes mentén helyezkednek el, a kapcsolat szorosságát pedig az mutatja, hogy a pontok mennyire szóródnak szét, illetve mennyire tömörülnek az egyenes köré. Az elemzés — statisztikailag számottevő korpusz esetén — számítógép nélkül nem végezhető el, a géptől kapott adatok pontdiagramon való kézi ábrázolása maga is rendkívül sok időt vesz igénybe. (Két-két mondatszerkezeti jellemző pozitív összefüggésének erősségét a 0,00-tól 1,00-ig terjedő értékskálán adjuk meg; az egymással ellentétesen mozgó sajátosságokét pedig a 0,00-tól -1,00-ig. A kapcsolat 0,00-tól 0,3-ig gyenge, 0,31-től 0,6-ig közepes erősségű; 0,61-től 1,00-ig (pozitív és negatív előjellel egyaránt) erős.) Az összefüggésvizsgálat szemléltetése érdekében két pontdiagramot elkészítettem.

Az 1. ábrán a 310 fogalmazás 310 szerkesztettségi mutatójának kapcsolatát láthatjuk a 310 fogalmazás mondategységeinek tagoltságai mutatójával. (Az ábrán nem lehetett mind a 310 pontot elkülönítve feltüntetni, mert az azonos értékek pontjai egybeestek.)

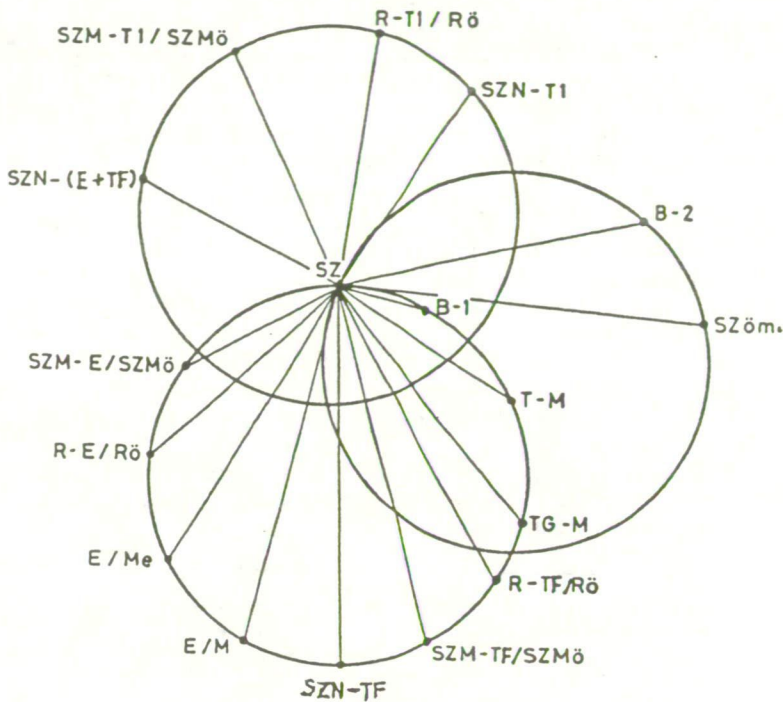
Az 1. ábrán pozitív irányú erős kapcsolatot (0,87) jelez a pontdiagram; jól látszik ez a pontok tömörödéséből, a meghúzható egyenes körüli elhelyezkedésükből: a korreláció tömör elliptikus formát mutat.



2. ábra

A most következő 2. ábrán ugyancsak erős, de ellentétes irányú kapcsolatot szemléltetnek ($-0,87$): a lineáris korrelációs együttható értéke megegyezik a fenti példáéval. Ezért hasonló a pontok elhelyezkedése a meghúzható egyenes körül, de az itt meghúzható egyenes iránya merőleges az 1. ábra egyenesének irányára, hiszen itt ellentétes irányú a kapcsolat.

A fenti pontdiagramokon csupán 2-2 mutató kapcsolata jeleníthető meg. Ha azt akarjuk tudni, hogy pl. a szerkesztettségi mutató még mely további mutatókkal áll korrelációs viszonyban, akkor ezek olyasféle kivetítése látszik célszerűnek, mint pl. amit a 3. ábra mutat.



3. ábra

A szerkesztettségi mutató környezetstruktúrája

Vizsgálataim alapján valószínűnek tartom, hogy — miként a kördiagramról látható — a szövegszerkesztettség átlagos mutatója más szövegtípusokban is hasonlóan informatív, azaz a szintméllységgel, szintnépességgel, telítettséggel stb. kölcsönös összefüggése típusonként jellemző mértékben más és más. Ezek feltárásához, s hasonló jellemzéséhez nagyon sok munka és — meglehetősen drága — gépidő kell. A vizsgálat eredménye azonban különféle metodikájú jelenségvizsgálatokhoz egyértelmű viszonyítási alap lehet.

Irodalomjegyzék

- DEME László:
1971. *Mondatszerkezeti sajátságok gyakorisági vizsgálata*. Akadémiai Kiadó, Budapest.
- B. FEJES Katalin
1981. *Egy korosztály írásbeli nyelvhasználatának alakulása*. Tankönyvkiadó, Budapest.
1985. A szerkesztettségi mutató környezetstruktúrája a korrelációs együtthatók tükrében. *Néprajz és Nyelvtudomány* XXIX—XXX. 37—49.
1986. Lebensalter — Text — Syntax: *Grazer Linguistische Studien* 27. 21—27.
1987. A mondat egység telítettségi mutatójának környezetstruktúrája. *Néprajz és Nyelvtudomány* XXXI—XXXII. 49—61.
1988. A nemzetiségi környezet és a mondat szerkezeti sajátságok néhány összefüggése. In: *A magyar nyelv rétegződése* (szerk. Kiss Jenő és Szűts László). Akadémiai Kiadó, Budapest, 308—311.

1991. A matematikai statisztika térhódítása a nyelvhasználat kutatásában. In: *Tanulmányok a magyar nyelvudomány történetéből* (szerk. KISS Jenő és SZÜTS László). Akadémiai Kiadó, Budapest, 164—167.
- GOPNIK, Myma:
 1986. The Development of connexity in young children. In: *Text Connectedness from Psychological Point of View* (ed. János S. PETÓFI). Hamburg, Buske verlag, 64—95.
 1989. The development of text competence. In: *Text and Discourse Connectedness* (eds. M—E. CONTE, J. S. PETÓFI, E. SÓZER). Amsterdam—Philadelphia, J. Benjamins, 225—244.
- HUNT, K. W.:
 1965. *Grammatical Structures Written at the Three Grade Levels*. NTCE Research Report No 3.
 1964. *Differences in Grammatical Structures Writte at the Three Grade levels, the Structures to be Analyzed by Transformational Methods*. Florida, Tallahassee.
- NAGY Ferenc:
 1972. *Kvantitatív nyelvészet*. Tankönyvkiadó, Budapest.
- PAPP Ferenc:
 1966. On the Depht of Hungarian Sentences: *Linguistics* 25.