

## 9.

### Az empirikus koherenciavizsgálat alapjai

#### Bevezetés a korrelációs korpuszelemzésbe<sup>1</sup>

CSÜRY ISTVÁN

#### 1. Bevezetés: megjegyzések a koherenciavizsgálat módszertani adekvátságáról

Ha a koherencia valódi természetéről, működéséről és tényezőiről empirikusan alátámasztott képet akarunk alkotni annak a koncepciónak megfelelően, melyet az *Officina Textologica* 20. kötetében megjelent, *A zabolátlan koherencia* című tanulmány (Csúry 2017) fejt ki, akkor a szövegek vizsgálatához sajátos módszerre van szükségünk. Ez a felfogás ugyanis a koherenciát egy szövegnek (diskurzusnak) az értelmezésével (értelmezéseivel) és az az alapján levont következtetéssel (következtetésekkel) fennálló relációjaként határozza meg. Ennek értelmében lehetetlenség a koherenciáról bármilyen releváns megállapítást tenni kizárólag a szöveg határain belül maradván, hisz az – a konnexitástól és a kohéziótól eltérően – nem immanens szövegsajátosság. Legalább még egy, az elemzés kiindulópontjául szolgáló szövegre reflektáló másik szöveget is vizsgálnunk kell, hogy az előbbiről megállapíthassuk: milyen értelmezési tartomány(ok) viszonylatában ítélnélhető ellentmondásmentesnek, ebben a keretben milyen konklúziókat tesz lehetővé, és melyek azok az elemei/sajátosságai, melyeket ez az interpretáció koherenciaszervezőként vagy koherenciajelzőként alapul vesz.

Ez a vizsgálati mód az elemzett szövegre egy másik szöveg optikáján keresztül tekint, szemben állva a megszokott módszerrel, melyet a (képzett, szakértő) elemző objektívnek tételezett nézőpontjából végzett közvetlen (csupán az önmagában vett tárgyra irányuló) elemzésmódja jelent. A két módszer közötti ellentmondás akkor oldható föl, ha úgy tekintjük, hogy csak „a” szöveg önmagában való koherencia-szemponthú elemzése ugyanúgy két diskurzus relációjából fakad, mint ahogyan azt a mi koherenciafelfogásunk elvárja, csupán egyikük, az elemző által részben virtuálisan konstituált értelmező szöveg látens, illetve egybeesős az elemzéssel mint metadiskurzussal. A szakértői elemzés csak egy (és ugyanúgy szubjektív, bár kétségtelenül kitüntetetten releváns) a lehetséges értelmező diskurzusok közül, ugyanakkor az esetleg ténylegesen (a szakértői elemzést megelőzően) létező értelmező diskurzusok sokasága sem meríti ki feltétlenül az értelmezési lehetőségeket, még a szöveg alkotója által szándékoltakat

---

<sup>1</sup> A kutatás az EFOP-3.6.1-16-2016-00022 DE „Debrecen Venture Catapult Program” című projekt támogatásával valósult meg.

sem. Így kétségtelen, hogy az elemző (nyelvész, szemiológus, a koherencia kutatója) saját értelmező diskurzusának föltétlen létjogosultsága van. Célszerű azonban explicitté tenni a vizsgálati szöveget *értelmező* és a koherencia tényezőit, működésmódját *elemző* diskurzus különbségét, és mindvégig szem előtt tartani, hogy interpretációról, nem pedig a szövegben eleve adott tényekről van szó.

A megszokott, introspektív megközelítést nemcsak azért szükséges a reflektáló/értelmező diskurzusokat is tekintetbe vevő megközelítéssel kiegészíteni, mert az ismertett koherenciafelfogásból ez logikusan következik. Aligha érvelhetünk triviális, „értelmezéstől függetlenül megálló” szövegtartalmakkal: az ilyenek véltekből is fakadhatnak váratlan interpretációk, és vetődhet új fény a koherencia tényezőire. Kissé elnagyolt fogalmazással úgy mondhatnánk, hogy a koherencia ott kezdődik, ahol a kohézió véget ér: az interdiskurzus terében. Emellett hasznos, ha a koherenciára vonatkozó állításokat kvantitatív, korpusz-alapú vizsgálatokkal nyert empirikus adatokkal is alá tudjuk támasztani. Egy szövegnek más szövegek optikáján át történő megvizsgálása sajátos szöveg-együttest feltételez: „központi” szövegek körül gravitáló, esetleg egymáshoz is kapcsolódó (egymásra reflektáló) különböző szövegeket. Nevezzük az ilyen szövegesoportokból álló szöveggyüttest *korrelációs korpusznak*. A jelen tanulmány tehát az előzőben (Csúry 2017) felvázolt projektum soron következő lépésére: ennek az elemének a részletesebb tárgyalására vállalkozik.

## 2. A korrelációs korpusz

A koherenciakutatás számára adekvátnak tekintett korpuszfajta meghatározásához a következő alapkérdéseket kell tisztázni:

- Mit értünk korrelációs korpuszon, különös tekintettel a más, ismert korpuszfajtákhoz való viszonyára, megkülönböztető sajátosságaira?
- Milyen típusai képzelhetők el a korrelációs korpusznak, és mi lehet ennek a tipológiának az alapja?
- Hogyan képzelhető el a korrelációs korpusz elemzése, különös tekintettel az alkalmazható/alkalmazandó annotációtípusokra?

### 2.1. A korrelációs korpusz meghatározása, szerkezete és viszonya más korpuszfajtákhoz

Az elemi korrelációs korpusz minimum két olyan megnyilatkozásból álló diskurzuskomplexum, ahol minimum egy megnyilatkozás a másikkal/többihez képest elsődleges, amennyiben keletkezése időben azét/azokét megelőzi, miközben rá a másik/többi explicite vagy a kontextusból kikövetkeztethető módon akár direkt, akár indirekt formában reflektál.

Az (elemi) korrelációs korpuszt az elemeit és a közöttük lévő relációkat leíró annotációegyüttes teszi teljessé. Ez összességében bonyolult relációs adatbázis képét vetíti elénk.

Tágabb értelemben korrelációs korpusznak nevezhető olyan elemi korrelációs korpuszok mint alkorpuszok együttese, melyek között nem áll fenn ez a függőségi viszony, csupán a rendeltetési céljuk, szerkesztési és annotációs elveik, valamint felhasználásuk kapcsolja össze őket.

Bár a főnti meghatározás bizonyos elemei csak később, a tipológiai fejtegetések fényében nyernek igazán indoklást, már most, a továbblépéshez szükség lehet kommentálásukra.

A *megnyilatkozás* terminus – bennfoglaló volta révén – alkalmasabb a *szöveg* vagy a *diskurzus* terminusoknál, ha a dialógusokat nem akarjuk kizárni a korrelációs korpuszok lehetséges alkotóelemei köréből. Márpedig ez aligha volna célszerű, lévén a dialógus a koherenciát létrehozó reláció elemi közege. Ugyanakkor indokolt a megnyilatkozások (szövegek, diskurzusok), nem pedig a megnyilatkozók többségét kikötni, mivel az öninterpretáció eseteit is figyelembe kell vennünk, tekintettel a koherencia kettős – a befogadói interpretáció mellett a megnyilatkozói szándékban is gyökerező – természetére.

Az időbeliség triviális volta ellenére hasznos definícióelem: rámutat arra, hogy a kiinduló szöveg elsődlegessége nem relevanciabeli különbséget, hanem egy folyamat dinamikájában elfoglalt pozíciót jelent. Észrevesszük emellett, hogy egy-nél több értelmező megnyilatkozás esetén közöttük párhuzamosság és egymásutániság is lehetséges. Az utóbbi esetet az értelmező megnyilatkozások egymásra reflektálásának lehetősége teszi érdekessé: ekkor nemcsak „másodlagosan elsődleges” korpuszelemeket kapunk, melyeknek koherenciáját önmagában is vizsgálhatjuk az őket (is) értelmező újabb megnyilatkozások fényében, hanem a kiinduló megnyilatkozás koherenciája is komplexebb megvilágítást nyerhet.

Valamely megnyilatkozás a koherenciarelációk vizsgálata szempontjából nemcsak akkor vehető figyelembe, ha explicite – annak azonosítást lehetővé tevő bármilyen említésével és esetleg egyes konkrét szövegelemeinek a felidézésével – utal egy bizonyos elsődleges megnyilatkozásra, hanem akkor is, ha ez az értelmező-reflektáló viszony a kontextusból kikövetkeztethető, például dialógusbeli rákövetkezéséből, a vizsgált megnyilatkozások intézményes kommunikációs kereteiből, közös tartalmi elemekből stb.

Végezetül: az indirekt formában való reflektálás lehetősége annyit tesz, hogy nem várjuk el feltétlenül az interpretáló megnyilatkozástól, hogy benne az értelmezés tárgyát vagy forrását képező szövegelemek konkrét (szó szerinti) formában említést nyerjenek.

Az egymással összekapcsolt, illetve kapcsolataik szerint tekintetbe vett korpuszelemek okán indokolt annak a meghatározása, hogy miben áll a korrelációs korpusz sajátossága a hasonló korpuszfajtákkal szemben. A *párhuzamos korpu-*

szok fordítás vagy átdolgozás révén létrejött szövegpárokat tartalmaznak, és mindenekelőtt az e műveletekkel érintett nyelvi-tartalmi jelenségek vizsgálatát szolgálják, ami természetesen a koherenciarelációk szempontjából is érdekes lehet, de önmagában nem elegendő az itt követett koherenciafelfogás érvényre juttatásához. Az *összehasonlító korpusz* a kontrasztív nyelvészeti kutatás vagy a nyelvváltozatok egybevető elemzésének eszköze lehet, az ebből a koherenciára nézve nyerhető következtetések köre is korlátozott. Az olyan *tematikus-funkcionális* alapon összeállított korpusz pedig, amilyen például bizonyos diskurzuselemzési keretek között használatos (mint például egy választási kampány beszédeiből készült korpusz), ugyan tartalmaz(hat) egymásra reflektáló szövegeket, de abban a korpuszépítésnek és -feldolgozásnak nem vezérelve mindaz, amit fontosabb a koherenciavizsgálat szempontjából sarkalatosnak vettünk. Emellett ráadásul a megnyilatkozók közötti (diszkurzív, pragmatikai) együttműködés, melyre pl. egy választási kampány résztvevői nem föltétlenül mutatnak hajlandóságot, fontos feltétele annak, hogy a koherencia működéséről minél teljesebb képet kapjunk.

A leírt céloknak és követelményeknek megfelelő korpusztípus tehát eltér azoktól, melyek a korpusznyelvészetben elterjedten használatosak. Jegyezzük meg: az utóbbi évtizedekben látványosan gazdagodott a koherenciaviszonyok korpuszalapú kutatása és szakirodalma, csakhogy az általunk követett koherenciafelfogástól eltérő keretekben. Tény ugyanakkor, hogy az ott elért eredményekkel nekünk is számot kell vetnünk, sőt azt kell vizsgálnunk, hogyan illeszthetők be a korrelációs korpusz alapú megközelítés kereteibe. Minderre alább, az annotáció kérdéseire érve térünk ki bővebben.

A korrelációs korpusz elemeinek forrása természetesen nem korlátozandó az írásos médiumú diskurzusra, ellenkezőleg: célszerű törekedni a multimodális kommunikáció legváltozatosabb produktumainak a vizsgálatba vonására. Ez persze tovább növeli a feladat komplexitását, különösen az annotáció és az annotálandó jelenségek terén.

## 2.2. A korrelációs korpusz lehetséges típusai: a tipológia alapjai

A korpusznyelvészetből ismert számos paramétert használhatunk a rendszerezéshez, osztályozáshoz e korpuszfajta esetében is, mint a korpuszelemek egy- vagy többnyelvű, írott vagy orális/multimediális (multimodális) volta, „műfaja”, stílári vagy szociolingvisztikai jellemzői stb. A speciális rendeltetéséből fakadó sajátos felépítése azonban egy másik rendezőelvet állít minden más elé, s ez nem más, mint a kiinduló megnyilatkozáshoz kapcsolódó értelmező megnyilatkozások kommunikációs jellemzői, ezek forrása. Lássuk, milyen esetekkel, elméleti lehetőségekkel számolhatunk ezek alapján.

1. *Monologális* interpretáló megnyilatkozások (amelyek – önmagukban – nem a szó szoros értelmében vett párbeszéd vagy társalgás fordulói).

a) *A kiinduló megnyilatkozás létrehozója* (nevezzük a továbbiakban megnyilatkozónak, és jelöljük M-mel) *nem azonos az interpretáló megnyilatkozóséval* (nevezzük a továbbiakban interpretátornak, és jelöljük I-vel). Jól lehet a kiinduló meghatározásaink alapján eredetileg bajos is más helyzetet elképzelni, mégis érdemes, ahogyan mindjárt látni fogjuk, ezt az alapesetet külön típusként kezelni. Itt tehát a kiinduló és az értelmező megnyilatkozások más-más személytől (közlőtől, beszélőtől) származnak, és formailag lehetnek teljes egész szövegek vagy szövegfragmentumok is (azaz az értelmezés tárgya lehet M egy teljes szövege vagy csupán annak egy része, és az értelmezés is alkothatja I teljes szövegét vagy szövegének csupán egy fragmentumát). Verbális műalkotás és annak értelmező elemzése vagy kritikája, egy polémia irományai, egy per iratai, egy közszereplő megnyilvánulása és az általa kiváltott sajtóbeli reakciók lehetnek többek között a példái az ilyen típusú korrelációs korpusz céljaira alkalmas anyagoknak.

b) *M és I ugyanaz a személy* (vagy entitás, pl. intézményi közlés esetén). Bármilyen meglepőnek tűnhet is első látásra a kiinduló megfontolásaink után ennek a lehetőségnek a megengedése, nem zárhatjuk ki az öninterpretáció eseteit sem, hiszen – egyrészt – ezekből más, M-mel nem azonos, létező vagy virtuális I-k létező vagy feltételezett értelmezéseit lehet (direkt vagy indirekt módon) megismerni, másrészt pedig az így elének táruuló perspektíva a koherenciát a beszélői szándék oldaláról világítja meg. Több altípust célszerű ezen belül megkülönböztetni.

i. *Az egyik az öninterpretációt tartalmazó szekvenciáé egyazon megnyilatkozáson (szövegen) belül.*

Belátható, hogy ha ezt a típust korrelációs korpusztípusként, illetve ilyen korpusz elemeként elfogadjuk, látszólag az olyan egyszerű korpuszhoz jutunk vissza, amely egy szövegen belül maradván próbálja reprezentálni a koherenciaviszonyokat. Ez eltérne a kiinduló koncepcióntól. Mégsem erről van azonban szó, mivel a korrelációs megközelítés nem a szokványosan „koherenciarelációknak” nevezett szövegstruktúrák feltérképezését célozza, hanem a polifónia-jelenségeket mutató szekvenciák azonosítását, elemzését, melyekben tehát M a sajátjától megkülönböztetett nézőpontokat is megjelenít, explicite vagy következtethető formában. Ebből a szempontból pedig az említett relációknak vagy struktúráknak csak egy része jön számításba, mint például a kifejtés, az újrafogalmazás vagy az ellentét, a megengedés bizonyos formái. Sajátos átmeneti változatot jelenthetnek az olyan szövegek, melyek többé-kevésbé változatlanul ismétlik meg M egy korábbi

szövegének tartalmait, de reflektálnak egyes – időközben született – hozzájuk kapcsolódó interpretációkra. Ezzel a jelenséggel találkozhatunk például a politikai diskurzusból, egy jelöltnek egy választási kampány során elmondott beszédeit egybevetve.

Számítanunk kell arra, hogy az (ön)interpretáció jelenségei kevésbé nyilvánvaló formákban is tetten érhetők. Egy olyan, vizsgált szempontunkból „ártatlannak” tűnő szövegrelációban, mint egy tulajdonnévvel említett referenshez (pl. *Mr. Gorsky*) kapcsolódó korreferáló elemek anaforaláncolata, nem mindegy, hogy az illetőt a szöveg egy későbbi pontján *az idős úr* vagy *a vénember* formában látjuk/halljuk újra említeni. Ami persze megint csak a kohézió és a koherencia közötti viszony újragondolásához vezet.

- ii. A másik *a kiinduló megnyilatkozástól elkülönülten, később létrehozott megnyilatkozása*, mely az első interpretációja vagy ilyen tartalmú/funkciójú öninterpretáló szekvenciát tartalmaz. Változatos műfajú, kommunikációs funkciójú, terjedelmű, formai, szerkezeti, tartalmi és kontextuális sajátosságokkal jellemezhető megnyilatkozásokról, szövegekről lehet szó. Ebben az esetben M és I azonossága ellenére is megállnak a korrelációs korpusz szerkezetéről mondottak. Az ilyenfajta vizsgálati anyagtól is azt várhatjuk elsősorban, hogy a koherenciát a beszélői szándék oldaláról világítja meg, indirekt módon azonban nagy valószínűséggel lehetőséget ad más I-k koherenciaítéleteinek a megismerésére is. Ebbe a típusba tartoznak például a valamely korábbi megnyilatkozásból és az általa kiváltott visszhangra, kritikára, támadásra az eredeti megnyilatkozó által válaszként közölt/közzétett reakcióból, nyilatkozatból alkotható korpuszok.
2. *Dialogális* interpretáló megnyilatkozások (amelyek önmagukban párbeszéd, társalgás vagy ezek írásbeli megfelelőinek – mint amilyenek az online fórumok, kommentárok – a fordulói, de szélsőséges példaként ide sorolható akár törvények, bírósági eljárások/döntések és kúriai jogegységi határozatok alkotta szövegegyüttesek is). Esetükben M és I különböző voltát szükségszerűnek tekinthetjük azzal, hogy az egyetlen M által dialógusformában megírt fejtegetés csupán felszíni jegyeiben mutat rokonságot ezzel a fajta diskurzussal. Ez utóbbi kikötést kiterjeszhetnénk akár a műalkotásként létrehozott, irodalmi vagy filmbeli dialógusokra is. A tapasztalat azt sugallja, hogy ezeket kellő körültekintéssel, afféle dialógusszimulációként – nem kizárólagos vizsgálati anyagként ugyan, de – felhasználhatjuk korrelációs korpusz céljaira.

Meg kell jegyeznünk, hogy a leírt monológális/dialogális, illetve „egyszerplős”/„többszereplős” típusok egymásba ágyazódása sem zárható ki, azaz például egy olyan dialógusforduló megjelenése, mely tartalmazza M egy kiinduló megnyilatkozása mellett az arra irányuló saját interpretációja közlését

is. Ha ezzel a jelenséggel találkozunk, ez a dialógus dinamikájának fényébe helyezve rávilágít az I-megnyilatkozások (s ezáltal a koherencia) kettős természetére. Ezek ugyanis nemcsak „hátra hatók”, azaz egy lezártnak tekintett közlés(elem) utólagos feldolgozásának a termékei, hanem „előre hatók” is lehetnek, vagyis előre ki akarhatják jelölni a létrejövőben lévő közlés(elem) lehetséges feldolgozásának az útját. A koherenciának az a felfogása, amely belső szövegtulajdonságnak tekinti, a fogalmat voltaképpen ez utóbbira korlátozza. Az általunk követett dinamikus koherenciafelfogás az előbbivel nem egyszerűen szemben áll, hanem annak elemeit újraértelmezve beépíti.

A dialóguskorpusz használatát két további megfontolás is befolyásolhatja. Először is: a párbeszédes jellegből adódóan olyan megnyilatkozásláncolatokkal találkozhatunk, melyekben az I-megnyilatkozások maguk is elsődleges, interpretált megnyilatkozásokká válnak egy rájuk következő beszédforduló viszonylatában, ami a tanulmányozni kívánt jelenségek sűrített jelenlétéhez vezethet. Másodsor: a dialógusok önmagukban nem biztos, hogy lehetővé teszik a koherenciajelenségek és létrejöttük nyelvi eszközei teljes változatoságának a feltárását.

Végül hadd emlékeztessünk arra is, hogy az Officina Textologica-program korábbi munkálataiban már megjelent a dialógális korrelációs korpusz előképe egy olyan tanulmányban (Csúry 2005a), mely a korreferencia-láncolatok dialógusbeli mintázatait vizsgálva ragadja meg a vizsgált anyag koherencia-jellemzőit, s ezen át egyfajta beszélői stratégiátípust.

Várható, hogy korrelációs korpusz építésének céljaira sok esetben nem szövegegészek, hanem csupán szövegfragmentumok felhasználása indokolt, amennyiben az interpretációt megvalósító megnyilatkozás és annak a kiinduló megnyilatkozásbeli tárgya is lehet szűkebb körű szövegegység egy ezeket magában foglaló szövegegészen belül. Ez felértékeli a célnak megfelelő válogatás és előfeldolgozás szerepét a korrelációs korpusz építésében. Tegyük hozzá: a célra irányuló válogatást az I megnyilatkozásában tapasztalt attitűd, illetve a megnyilatkozások vagy a kommunikációs folyamat eredménye szerint ugyancsak szükségesnek ítéelhetjük, például aszerint, hogy az interakciós jelentéskonstruálás vagy a félreértések pragmatikája-e egy adott vizsgálat tárgya.

### 2.3. A korrelációs korpusz annotációja és elemzése

Fogas kérdésnek tűnhet, hogy a megcélzott eredményt: a koherencia-jelenségeknek az interdiszkurzív térben való megragadását milyen annotációs eszközökkel érhetjük el. Közelebről a következő részkérdések vetődnek fel:

1. Mit (milyen jelenségeket, elemeket, tulajdonságokat) annotáljunk?
2. Mivel (milyen eszközzel, milyen rendszerben) annotáljunk?

3. Hogyan (milyen sorrendben, milyen stratégia szerint) annotáljunk?
4. Mi végett annotáljunk, azaz milyen kérdésekre keressük a választ a korrelációs korpusz annotációinak segítségével?
5. A rendelkezésre álló annotált korpuszokból melyek és milyen mértékben lehetnek használhatók az itt vázolt sajátos célokra?

Mindezen kérdésekre nem lehet célunk e tanulmány keretében kimerítő választ adni, hisz sok tekintetben éppen további – szükséges – kutatási irányokra mutatnak rá. Amellett, hogy az annotáció terén számos létező eredményre támaszkodhatunk, úgy tekinthetjük, hogy a tanulmány következő fő szakaszában megfogalmazandó általános válaszok az empirikus munka kiinduló hipotézisei lehetnek. Ezt – a korrelációs korpusz koncepcióját bemutató – szakaszt a fenti 3. és 4. (a *hogyan* és a *mi végett* firtató) kérdés megválaszolásával zárhatjuk.

A korrelációs korpusz (kiváltképp, ha multimediális anyagot tartalmaz) több annotációs réteggel tehető teljessé, melyek egyformán elkészítendőek külön-külön a kiinduló és az értelmező korpuszrésztek tekintetében is. Nemcsak töbréteggű, hanem több menetben végezhető, rekurzív lépéseket is tartalmazó, ciklikus – azaz még az annotációs sémának akár menet közben történő módosítását is megengedő<sup>2</sup> – annotációs eljárásról van azonban szó, ugyanis meg kell állapítani a korpuszrésztek, illetve a bennük annotált elemek egymás közötti viszonyát is. Ez részint az egymáshoz interdiszkurzív módon kapcsolódó annotátumok összekapcsolását, az annotációk összhangba hozását implikálja, részint pedig akár azt is, hogy az értelmező korpuszrésztek elemzése eredményeképpen esetleg vissza kell térni a kiinduló korpuszrész elemzéséhez, annak bizonyos annotációs rétegeit kiegészítve vagy módosítva.

Az elemzés eredményeként a korpuszból kinyerhető – a koherencia létrejötte és tényezői tekintetében releváns – információk a következő fő területekre összpontosulhatnak:

- M és I megnyilatkozásainak referenciális átfedése (a kiinduló megnyilatkozás referenseinek megtartása, elhagyása, módosítása, kiegészítése az interpretáló megnyilatkozásban, különös tekintettel a korreferáló elemek esetleges jelentésmódosító hatására);
- M és I megnyilatkozásai diszkurzív kereteinek<sup>3</sup> átfedése (a kiinduló megnyilatkozásban azonosítható térbeli/időbeli, tematikus, tulajdonításon, valamint értékeléseken alapuló keretek tükröződése vagy módosulása, helyettük/mellettük újak megjelenése az interpretáló megnyilatkozásban);
- M és I megnyilatkozásai logikai, argumentatív és következtetési relációinak átfedése (a kiinduló megnyilatkozás relációinak tekintetbe vétele, figyelmen kívül hagyása, módosítása, új relációk bevezetése);

<sup>2</sup> ANALEC (ld. Landragin et al. 2012)

<sup>3</sup> Charolles 1997

- az M és I megnyilatkozásaiban megfogalmazott és/vagy azok által implikált/megengedett konklúziók átfedése (azonossága, kompatibilitása, eltérései).  
Mindezek (kvantitatíve is releváns) feltárásától végső soron nem egyedi esetek, hanem esettípusok, összefüggések jobb megismerését, rendszerezését remélhetjük.

### 3. Alapvető annotációs/elemzési feladatok a gyakorlatban

#### 3.1. Tartalmi kérdések

Az elmondottakkal összhangban aszerint kell tehát az annotációs rétegeket meghatározni, hogy milyen (nyelvi) elemektől, struktúráktól, sajátosságoktól várható, hogy koherenciaitételek forrásaiként működnek.

Kulcsszerepe van a koherencia tekintetében a megnyilatkozások lexikai anyagának, és persze nemcsak önmagukban a szavaknak, hanem szemantikai mezőiknek, konnotációiknak, a hozzájuk kapcsolódó – általuk az egyes megnyilatkozókban és értelmező szubjektumokban különbözőképpen aktivált – enciklopedikus/kontextuális ismereteknek. Ezt azonban annotáció formájában bajos volna megragadni. Amellett tehát, hogy a szavak szintjén érdemes lehet (automatikus) morfo-lexikai és szintaktikai elemzést futtatni (többek között a szóalakok egyértelműsítése és lemmatizálása céljából), az látszik indokoltnak, hogy az M- és az I-korpuszelemekről készüljön egybevető lexikai statisztikai kimutatás, mely a tartalmas szavak körén belül kiemeli a közös tendenciát mutatókat csakúgy, mint a pozitív és a negatív értelemben vett kulcsszavakat, és feltárja a szavak jellemző kombinációit, az ismétlődő mintázatokat is. Ez a korpuszvizsgálat eredményeinek elemzésekor jó szolgálatot tesz a kiinduló szöveg és interpretációja közötti tartalmi konvergencia vagy divergencia kimutatásában. Az így kapott adatok nemcsak optimalizált bemenetül szolgálhatnak azoknak a kvalitatív vizsgálatoknak, melyek a kommunikátumok fogalmi verbalizált („értelmi”) jelentésének<sup>4</sup> feltárására irányulnak, hanem a vizsgálati objektum megkettőzött-ségének köszönhetően e jelentés létrehozásának mikéntjéről is árulkodhatnak.

A következő réteg a referáló kifejezéseknek és a korreferenciális láncolatoknak, a korreferencia-viszonyok hálózatának a feltárása. Az így nyert adatok – a főntebb mondottakhoz hasonlóan – a Petőfi S. által referenciális jelentésként megkülönböztetett jelentéssík<sup>5</sup> többdimenziós, dinamikus vizsgálatát szolgálhatják. (A korpusznyelvészeti eljárások ráadásul technikailag, illetve az átláthatóság és a kezelhetőség szempontjából hatékonyabb eszközhöz juttatják a szemiotikai textológiai elemzést is.) A feladatnak – technikai értelemben – egyszerűbb részét

---

<sup>4</sup> Követve Petőfi S. János fogalomrendszerét, aki a szövegnek három jelentéssíkját (az értelmi mellett a referenciális és kommunikatív jelentését) különbözteti meg (2004:103).

<sup>5</sup> Petőfi 2004:103.

jelenti az M- és az I-korpuszelemeken belül létrehozni ezt az annotációt, amit ki kell egészíteni a kettő közötti kapcsolatok annotációjával. Ez utóbbi azonban egy szélesebb perspektívájú megközelítést igényel, amennyiben az I-korpuszelemek kifejezései nemcsak olyan M-korpuszelembeli egységekre utalhatnak, melyek annak korreferencia-hálóiban jelennek meg, hanem különböző strukturális elemekre, szerveződési szintjeire vagy belső relációira is. Az I-korpuszelemek annotációját tehát az M-korpuszelemek (első körös) annotációjának célszerűen meg kell előznie. Ami a korreferenciarelációk annotációjának minőségét, részletességét illeti, fontos elvárás, hogy megjelenjenek benne azok a referenciális módosulások, melyek egy korreferencia-láncon annak kiindulópontjához, az eredeti antecedenshez képest végbemennek, és hogy tükrözze az egyes anaforikus kifejezések hordozta pragmatikai orientációt, esetleges értékelő mozzanatokot is. Módszertani és szoros értelemben vett technikai szempontból is kihívást jelenthet emellett az olyan implicit/kikövetkeztethető tartalmak kezelése, melyekre a korpuszelemek kifejezései valamilyen módon utalhatnak, amelyek jelentésviszonyok kulcsszereplői lehetnek, s amelyeket a Petőfi-féle szemiotikai textológiai keretben a szöveg ún. kiegészített organizációtípusa szintjén azonosíthatunk. Ezek forrása nemcsak valamilyen általános tudáskeretben lehet keresendő, hanem a megnyilatkozók kontextusának specifikus elemei sorában is, mint amilyenek pl. a más (korábbi) megnyilatkozásaikra/interakcióikra, személyiségükre és cselekedeteikre, viselt dolgaikra, ismeret-, tapasztalat- és hiedelemvilágukra vonatkozó ismeretek.

Külön annotációt kíván – nyilván csak komplexebb, terjedelmesebb korpuszelemek esetében – a szövegszerkezet tömbösödése, hierarchikus struktúrája, gondoljunk akár a Petőfi S.-féle fogalmi-terminológiai hagyomány makroegységeire, akár a Charolles-féle keretekre. A fentebb mondottakkal összhangban ez nemcsak a szóban forgó tömböknek és egymáshoz képest elfoglalt hierarchikus vagy mellérendelt helyzetüknek, hanem sajátos jelentésképző funkciójuknak, mint például a bennük foglalt propozicionális tartalmak érvényességére irányuló specifikációnak vagy valamely minősítés ezekre való kiterjesztésének/korlátozásának az annotációját is jelenti.

Ez utóbbi réteget némileg keresztezve, átfedésben egészítheti ki azoknak az értelemi struktúráknak az annotációja, melyeket jelentésszerkezeteknek nevezhetnénk, másutt pedig diskurzusrelációk vagy koherenciarelációk (*discourse / coherence relations*) az elnevezésük. Ennek s az előbbi annotációs rétegnek az elkülönítése vagy összekapcsolása, illetve tartalmának meghatározása azonban olyan alapvető elméleti megfontolások függvénye, melyek igen eltérő eredményekre vezethetnek. Erre a problémára a későbbiekben térünk vissza.

Szerkezeti és pragmatikai specifikumuk miatt sajátos annotációt kívánnak a dialógusok összetevői (az ún. dialógusaktusok, *dialogue acts*) is. E téren is támaszkodhatunk több elméleti keretre ugyanúgy, mint gyakorlati megoldásokra

(például Roulet et al. 2001 vagy SWBD-DAMSL, Jurafsky et al. 1997), sőt akár nemzetközi szabványt is követhetünk (Dialogue Act Markup Language, DiAML: ISO 24617-2:2020<sup>6</sup>).

Ha egy megnyilatkozás nem pusztán verbális anyagból áll, vagy ha egyszerűen csak követni akarjuk a szövegnek Petőfi S. által multimediális kommunikátumként meghatározott koncepcióját, tekintetbe kell vennünk a vizuális (tipográfiai, grafikai, illusztrációs stb.) elemeit is, melyeknek meghatározó szerepe lehet a koherenciaítéletekben. Orális korpusz esetén az intonációs és egyéb szupraszegmentális fonetikai jelenségekre kell figyelemmel lennünk, ha pedig valódi multimodális korpuszunk van, újabb annotációs rétegeket kíván a tekintet, a mimika, a gesztusok és a testtartás/testmozgások.

Az M- és I-korpuszrészecskék relációinak azonosítása/annotációja az összes felsorolt réteg elemeinek vonatkozásában elvégzendő. Az így nyert adatok (Petőfi S. terminológiájával élve) a *sensus* inferenciális organizációformájára, illetve annak interpretációjára segítenek fényt deríteni. Különösen értékesek azok a megfigyelések, melyeket a Petőfi S. által kevésbé interszjektívként jellemzett *fogalmi nem verbális* és a *nem fogalmi sensusra*<sup>7</sup> nézve tehetünk a segítségükkel. A mondottak értelemszerűen kiterjednek arra a negyedik – asszociatív – jelentéssíkra is, mellyel Dobi (2013:175) javasolja kiegészíteni Petőfi S. hármas rendszerét.

### 3.2. Technikai kérdések

Ezeknek tüzetes áttekintésébe igazán csak egy konkrét korrelációs korpusz-építési projekt megkezdésekor érdemes belefogni. Itt mindenekelőtt azt kell kiemelnünk, hogy egy olyan, méreteiben is jelentősebb annotált korpusz megvalósítása, mely a fent leírt teljes metainformáció-együttest tartalmazza, idő-, munka- és forrásigényes vállalkozás. Konkrét kutatási résztémák céljaira kisebb, szűkebb körű annotációval ellátott korpuszok azonban viszonylag könnyen készíthetők, és az ehhez szükséges eszközök is mind fejlettebbek és elérhetőbbek, az általuk létrehozott kimenet pedig egyre inkább átvihető egyik rendszerből a másikba, vagy evégett átalakítható. Ilyen eszközök például: korreferencia-annotációra pl. a SACR<sup>8</sup>, komplex feladatokra a TXM<sup>9</sup>, ill. a bizonyos mértékig elő-

---

<sup>6</sup> Blunt et al. 2020; a szabvány az Interneten: <https://www.iso.org/standard/76443.html>.

<sup>7</sup> Petőfi 2004: 47.

<sup>8</sup> Oberle 2018; a program elérhető: <https://boberle.com/projects/coreference-annotation-with-sacr/#from-and-to-glozz-and-txm>

<sup>9</sup> Heiden, 2010; a program elérhető: <https://txm.gitpages.humanum.fr/textometrie/files/software/TXM/0.8.0/>.

deiként tekinthető, de önállóan továbbra is elérhető Glozz<sup>10</sup> és ANALEC<sup>11</sup>, vagy a WARP-Text<sup>12</sup>.

Ha viszont a létező, diskurzus-annotációt tartalmazó korpuszokra tekintünk, azt láthatjuk, hogy ezek sajátosságai és a mi speciális elvárásaink számunkra legfeljebb részleges kiaknázhatóságukat eredményezik, vagy zárt voltuktól, vagy annotációs rendszerükből, és leginkább abból eredően, hogy bajos volna belőlük olyasfajta szövegközi viszonyokra vonatkozó információkat kinyerni, mint amilyenekre éppenséggel kíváncsiak vagyunk.

### 3.3. Az elméleti keretekkel kapcsolatos nehézségek

Világos, hogy a koherencia korrelációs korpusz alapú vizsgálata konzisztens elméleti alapokon valósulhat csak meg értékelhető eredménnyel. Ez nem csupán a koherenciáról alkotott koncepcióknak átgondoltságát jelenti, hanem a korpusz-elemzés és annotáció különböző rétegeit megalapozó elméleti kereteket is, melyeknek egységesnek vagy legalábbis egymással (és az általunk követett koherenciafelfogással) kompatibilisnek kell lenniük. Gondoljunk mindenekelőtt a szövegtömbösödés és a szövegbéli szemantikai-pragmatikai relációk annotációjára, s az ennek kapcsán tett megjegyzésünkre. Itt ugyanis egy központi problémához érünk.

Az elmúlt mintegy 20 évben látványos korpusznyelvészeti eredmények születtek a diskurzuselemzés, azon belül is az úgynevezett koherenciarelációk feltérképezése terén, különböző elméleti alapokról kiindulva. A kutatások/fejlesztések két talán legismertebb és legnagyobb hatású ilyen eredménye a Penn Discourse Treebank (PDTB; Prasad et al. 2008), illetve a Rhetorical Structure Theory Discourse Treebank (RST-DT; Carlson és Marcu 2001). Mindegyik teljességre törekszik, a szöveg jelentésszerkezeti egységeinek és relációinak kimerítő rendszerezésére és illusztrálására, más-más elméleti alapállásból, eltérő fogalom- és kategóriarendszerrel, de – elkerülhetetlenül – bizonyos fokú átfedéssel is. Mármost ha fenn akarjuk tartani a szöveg tömbösödésével kapcsolatos elképzeléseinket, s ennek elkülönült rétegben történő annotálását, akkor a másik, jelentésszerkezeti annotációjának céljaira az RST-t kell választanunk (ha a választási lehetőséget e kettőre szűkítjük). Az általunk felvázolt annotáció-együttesben megjelenő *keretekre* ugyanis, köztük a tulajdonításokéira nem terjed ki, míg a PDTB a tulajdonításokat is a koherenciarelációkkal együtt tárgyalja. Tehát ha emezt választanánk a jelentésszerkezetek annotálására, akkor – a redundancia és

<sup>10</sup> Widlöcher–Mathet 2009; a program elérhető: <http://www.glozz.org/>.

<sup>11</sup> Landragin et al. 2012; a program elérhető: <https://www.lattice.cnrs.fr/ressources/logiciels/analec/>.

<sup>12</sup> Kovatchev et al. 2018; a program elérhető: <https://github.com/venelink/WARP>.

az inkonzisztencia elkerülése végett – a javasolt két rétegből (tömbösödés + jelentésszerkezetek) egyet kellene alkotnunk.

Bármilyen elméleti rendszerre és annotációs sémára esik is a választásunk (a szöveg tömbösödésének és jelentésszerkezetének a leírása céljából), a kiválasztott rendszer ökonómiája, ellentmondás-mentessége, leíró ereje, adekvátsága lényegi követelmény, aminek fényében indokoltnak láthatjuk akár saját kategória-rendszer kidolgozását is (a meglévőkből<sup>13</sup> kiindulva, azokat tökéletesítve). Nem véletlen, hogy a legújabb kutatások, jelesen a Sanders és kutatótársai tanulmányában bemutatottak ezeknek a különböző rendszereknek az egyesítését/egységeseítését célozzák (Sanders et al. 2021). Ez utóbbi munka egyébként lefed a vizsgált rendszerek diskurzusrélatió-típusaiból, amit csak lehet, így a PDTB-ből is, viszont éppen az ebben szereplő, általunk inkriminált tulajdonításrelációt explicit módon elhagyja, ami az annotációs rétegek általunk prezentált felosztásának tartható voltát sugallja.

### 4. Záró megjegyzések

Tanulmányunk a tényleges értelmezői ítéletekre építő dinamikus-interaktív koherenciafelfogásnak az empirikus kutatás gyakorlatába történő átültetésével foglalkozik. Először is részleteiben tárgyalja az e célra javasolt, korrelációs-nak nevezett korpusztípus sajátosságait és alfajait, majd pedig feldolgozásának és felhasználásának alapvető kérdéseit veszi sorra, áttekintve az annotációjának megkívánt dimenzióit s az általuk felvetett bizonyos problémákat. Kitér az annotáció menetére és lehetséges eszközeire is.

Mindez egyszerre jelenti a Csúry 2017-ben felvázolt problémák részletesebb vizsgálatát és a további kutatási feladatok kijelölését, pontosítását. Ez nemcsak a koherenciajelenségek tanulmányozásának folytatását szolgálhatja, hanem olyan munkálatok irányába mutat, melyek az Officina Textologica-program keretében vizsgált más szövegjelenségek, különösen a korreferencia problémakörének a feltárása terén is gyümölcsözők lehetnek.

### Irodalomjegyzék

- BUNT, H. et al. 2020. The ISO Standard for Dialogue Act Annotation. In: Calzolari, N. et al. (eds.): *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020). Marseille, 11–16 May 2020*. European Language Resources Association (ELRA). 549–558.
- CHAROLLES, M. 1997. L’encadrement du discours: Univers, champs, domaines et espaces. *Cahier de Recherche Linguistique* Vol. 6. 1–73.

---

<sup>13</sup> Nem megfelelően egy saját korábbi javaslatunkról sem (Csúry 2005b).

- CSÚRY I. 2005a. Gázspray, bilincs és egy füstölgő szenátor. (Komplex szövegösszetevők a dialogális szövegek korreferencia-hálójában.) In: Petőfi S. J.–Szikszainé Nagy I. (szerk.): *Officina Textologica 12. A korreferencialitás poliglott vizsgálatá*. Debrecen: Debreceni Egyetem, Magyar Nyelvtudományi Tanszék. 35–72.
- CSÚRY I. 2005b. *Officina Textologica 13. Kis könyv a konnektorokról*. Debrecen: Kossuth Egyetemi K.
- CSÚRY I. 2017. A zabolátlan koherencia. In: Dobi E.–Andor J. (szerk.): *Officina Textologica 20. Esettanulmányok a szövegkoherenciáról*. Debrecen: Debreceni Egyetemi K. 11–38.
- DOBI E. 2013. A nem tipikus szövegek jelentésreprezentációjának egy kérdéséről. *Magyar Nyelvjárások* 51. évf. 169–177.
- HEIDEN, S. 2010. The TXM Platform: Building Open-Source Textual Analysis Software Compatible with the TEI Encoding Scheme. In: Otaguro, R. et al. (eds.): *24th Pacific Asia Conference on Language, Information and Computation – PACLIC24*. Sendai: Institute for Digital Enhancement of Cognitive Development. 389–398.
- JURAFSKY, D.–SHRIBERG, E.–BIASCA, D. 1997. *Switchboard SWBD-DAMSL shallow-discourse-function annotation (coders manual, draft 13). Technical Report 97-02*. Boulder: University of Colorado, Institute of Cognitive Science.
- KOVATCHEV, V.–MARTÍ, A. M. –SALAMÓ, M. 2018. WARP-Text: a Web-Based Tool for Annotating Relationships between Pairs of Texts. <https://aclanthology.org/C18-2029.pdf> (2022-01-18)
- LANDRAGIN, F.–POIBEAU, T.–VICTORRI, B. 2012. ANALEC: a New Tool for the Dynamic Annotation of Textual Data. European Language Resources Association (ELRA). In: Calzolari, Nicoletta et al. (eds.): *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2012)*. European Language Resources Association (ELRA), 357–362.
- OBERLE, B. 2018. SACR: A Drag-and-Drop Based Tool for Coreference Annotation. In: *Proceedings of the 11th International Conference of Language Resources and Evaluation (LREC 2018)*. European Language Resources Association (ELRA). <https://aclanthology.org/L18-1059.pdf> (2022-01-18)
- PETŐFI S. J. 2004. *A szöveg mint komplex jel. Bevezetés a szemiotikai-textológiai szövegszemléletbe*. Budapest: Akadémiai K.
- PETŐFI S. J. 2009. *Officina Textologica 15. Egy poliglott szövegnyelvészeti-szövegtani kutatóprogram II. (Adalékok a verbális szövegek szövegösszefüggőség-hordozóinak vizsgálatához)*. Debrecen: Debreceni Egyetem, Magyar Nyelvtudományi Tanszék.

- ROULET, E.– FILLIETTAZ, L.–GROBET, A. 2001. *Un modèle et un instrument d'analyse de l'organisation du discours: Avec la collaboration de Marcel Burger*. Bern: P. Lang.
- SANDERS, T. J. M. et al. 2021. Unifying dimensions in coherence relations: How various annotation frameworks are related. *Corpus Linguistics and Linguistic Theory* Vol. 17, No. 1. 1–71.  
<https://doi.org/10.1515/cllt-2016-0078> (2022-01-18)
- WIDLÖCHER, A.–MATHET, Y. 2009. La plate-forme Glozz: environnement d'annotation et d'exploration de corpus. In: *Actes de la 16<sup>e</sup> Conférence sur le Traitement Automatique des Langues Naturelles (TALN'09), session posters*. Senlis, France: ATALA.  
<https://hal.archives-ouvertes.fr/hal-01011969> (2022-01-18)

### **Foundations of Empirical Investigations on Coherence** Introduction to Correlational Corpus Analysis

This paper deals with principles and problems of turning into practice the dynamic and interactive conception of coherence based on interpreters' judgments, presented in (Csúry 2017). The author claims that a special type of corpus, called correlational corpus, needs to be designed for the sake of investigating relevant phenomena. A minimal correlational corpus consists of a pair of two pieces of discourse (either utterances, dialogue turns or whole texts) out of which one expresses an interpretation of the other. After establishing a typology of subtypes of correlational corpora, the author examines their possible annotation layers as well as the annotation procedure and its tools. Issues of using known theoretical and annotation frameworks are also being addressed. This study develops further the claims advanced in (Csúry 2017) and initiates, at the same time, a research program.