

**6.**  
**A koreferencia kérdései**  
**a számítógépes szövegfeldolgozás szempontjából**  
BODA I. KÁROLY—PORKOLÁB JUDIT

**Bevezetés**

Az alábbi tanulmányban kísérletet teszünk arra, hogy bemutassuk a szövegek *koreferenciális elemzésének*, mint hatékony szöveg-elemzési módszernek egy lehetséges felhasználását, amely lehetővé teszi koreferenciálisan elemzett szövegek számítógépes vizsgálatát. Írásunkban először egy kiválasztott szövegrészlet koreferenciális elemzését végezzük el, és ennek kapcsán rámutatunk azokra a módszertanbeli kérdésekre, amelyek megoldása szükséges ahhoz, hogy a koreferenciális elemzés eredménye egy számítógépes szövegfeldolgozó program bemenete, inputja lehessen. Ezután bemutatjuk az általunk az elemzés során használt számítógépes programok által szolgáltatott konkrét eredményeket, amelyek véleményünk szerint meggyőzően bizonyítják, hogy a koreferenciális elemzés és a számítógép együttes alkalmazása rendkívül hatékony szövegnyelvészeti, ill. *szövegtani eszközt* ad a témával foglalkozó kutatók kezébe. Végül röviden foglalkozunk az eredmények lehetséges elméleti értelmezésével, amely véleményünk szerint elvezethet a koreferenciális elemzés, és rajta keresztül a szövegnyelvészet, ill. szövegtan érintett kognitív vonatkozásainak jobb megértéséhez.

**1. Koreferenciális elemzés**

Az általunk választott szövegrészlet Szent János Apostol Jelenéseinek Könyvéből származik:

Ve: Szent János Apostol Jelenéseinek Könyve. 1:1-5. Bevezetés és Címzés (részlet).

Szent János Apostol Jelenéseinek Könyve

<sup>1</sup>Jézus Krisztus kinyilatkoztatása, amelyet azért adott neki az Isten, hogy szolgáinak megmutasson mindent, aminek csakhamar be kell következnie. Elküldte angyalát, így adta tudtul szolgájának, Jánosnak, <sup>2</sup>aki tanúskodik Isten Igéjéről és Jézus Krisztus tanúságtételéről: mindenről, amit látott. <sup>3</sup>Boldog, aki

olvassa és aki hallgatja ezeket a prófétai szavakat, és meg is tartja, ami bennük meg van írva, mert az idő közel van.

<sup>4</sup>János a hét ázsiai egyháznak. Kegyelem és békesség nektek attól, aki van, aki volt és aki eljő, meg a hét lélektől, aki trónja előtt áll, <sup>5</sup>és Jézus Krisztustól, a hűséges tanútól, a halottak elsőszülöttétől és a föld királyainak fejedelmétől!

(Biblia. Újszövetségi szentírás. - Szent István Társulat, 1996)

#### **Az első szövegmondat koreferenciális elemzése:**

A szöveg koreferenciális elemzése során a szövegmondatokat szétbontottuk kisebb, ún. *k o m m u n i k á c i ó s e g y s é g e k r e*. Az első szövegmondat esetén ez a következőképpen történt:

[K01] = [k01]&[k02]&[k03]

Szent János Apostol Jelenéseinek Könyve <sup>1</sup>Jézus Krisztus kinyilatkoztatása, amelyet azért adott neki az Isten, hogy szolgálainak megmutasson mindent, aminek csakhamar be kell következnie.

Látható, hogy a [K01] indexszel jelölt szövegmondat három kommunikációs egységből tevődik össze, amelyeket a [k01], [k02] és [k03] indexek jelölnek. A továbbiakban mind a szövegmondatok, mind a kommunikációs egységek sorszámozása folyamatosan, növekvő számsorrendben történik. A szövegmondatok és kommunikációs egységek kapcsolatát minden szövegmondat esetén explicite feltüntetjük; ennek során az & operátort használjuk a kommunikációs egységek egy szövegmondatban való előfordulásának jelölésére.

#### **Az első szövegmondat első kommunikációs egységének koreferenciális elemzése:**

[k01] = (=i03)(=i04)[=!i01] B (=i05)[=i02]

// [k01/&vb/&ind]:

Szent^János^Apostol[=i03]^

Jelenései(i03)[=i04]nek^

Könyve(i03)(i04)[=i01]

Jézus^Krisztus[=i05]^

kinyilatkoztatás(i05)[=i02]a,

[i01] Szent János Apostol Jelenéseinek Könyve

[i02] Jézus Krisztus kinyilatkoztatása

[i03] Szent János Apostol

[i04] Szent János Apostol jelenései

[i05] Jézus Krisztus

A koreferenciális elemek, ill. koreferencia-indexek meghatározása, valamint a használt jelölésrendszer nem igényel különösebb kommentárt, mivel fő vonalaiban megegyezik az *Officina Textologica* második kötetében kikristályosodott módszerrel. Az elemzett kommunikációs egység szövegének tördelése pusztán az elemzés során kapott koreferenciális elemek áttekintését segíti. Új viszont, és alapvető változás az eddigiekhez képest a [k01] kommunikációs egység explicit reprezentálása kizárólag a koreferenciális elemzés során bevezetett szimbólumokkal. Ez az index-reprezentáció (Ko/&ind) képezi a szöveg számítógépes feldolgozásának alapját.

Az első kommunikációs egység index-reprezentációjában alkalmazott B operátor a kommunikációs egység, mint mondat állítmányát jelöli, esetünkben ez a teauruszokban is alkalmazott generikus („absztrakt/konkrét”) reláció szemantikai megfelelője, erre utal a B jelölés is (*broader term*). Egy R reláció, mint operátor operandusainak helyzete szemantikai jelentést hordoz, így azok nyilvánvalóan nem felcserélhetőek. Pontosabban: az operandusok felcserélése egy új reláció bevezetését tenné szükségessé. Bár ezt a továbbiakban nem használjuk („helyette” a Co szintagmatikai jellegű relációt alkalmazzuk, ld. később), megjegyezzük, hogy a B reláció „inverze”, az N reláció (*narrower term*) bevezetésével a [k01] kommunikációs egység egy ekvivalens reprezentációjához juthatunk el

// [k01] = (=i05)[=i02] N (=i03)(=i04)[=!i01]

módon. (A // jelekről később lesz szó, szemantikai funkciójuk nincs.)

A B reláció mellett még két (paradigmatikai jellegű) relációt használunk az elemzés során. Az U reláció az operandusok közötti szinonim kapcsolatot jelzi (*use*), míg a B\* reláció arra utal, hogy az operandusok között csak számban van eltérés. Például 'Szent János Apostol' U 'János' ([c02]), ill. 'Jézus Krisztus szolgálja' B\* 'Jézus Krisztus szolgálja' ([c03]).

A szövegben használt // jelek a számítógépes program számára jelzik, hogy az utánuk következő sort hagyja figyelmen kívül — a tanulmány szövegét ugyanis közvetlenül használhatjuk a számítógépes szövegfeldolgozó program inputjaként. Valójában a program a számára hasznos adatokat (sorokat) a [k01], [k02], ... indexekkel, ill. később a [c01], [c02], ... indexekkel kezdődő sorok felismerésével azonosítja. A program ilyen megvalósítása azt a célt szolgálja, hogy a szövegeket elemző kutatók koreferencia elemzést végezve közvetlenül a tanulmány szövegét dolgoztathassák fel a számítógéppel — akár „menet közben” is, pl. ellenőrizve a munka során felmerült észrevételeiket, hipotéziseiket.

## A koreferencia kérdései a számítógépes szövegfeldolgozás szempontjából

Mivel az elemzés során előre nem lehet tudni — legalábbis saját tapasztalataink szerint —, hogy milyen egyéb információk (pl. új koreferencia indexek) bevezetése, „beszúrása” válik majd szükségessé, ezért egy külön segédprogramot fejlesztettünk ki, amely az elemzés szövegében előforduló indexeket előfordulásuk sorrendjében újrászámozza. Emiatt figyelhető meg pl. az, hogy az első kommunikációs egységben a (=i03)(=i04)[=!i01] koreferenciális kifejezés fordul elő, mint a 'Szent János Apostol Jelenéseinek Könyve' koreferenciális elem indexe, és *nem* a „manuális” elemzés során kézenfekvőbbnek tűnő (=i01)(=i02)[=!i03] kifejezés. Az újraindexelő program ugyanis soronként dolgozza fel a szöveget, és először a szögletes zárójelekkel határolt, alapvető *referenciális jelentést* hordozó indexeket azonosítja, és csak ezután az elemek szintagmatikai (az *értelmi jelentésen* alapuló) kapcsolataira utaló, kerek zárójelekkel határolt indexeket.

Itt kell megjegyeznünk, hogy egy formális, de hasznosnak tűnő egyszerűsítéssel éltünk a koreferenciális elemzés során: a kerek zárójellel jelölt koreferencia-indexeket csak a szögletes zárójellel jelölt koreferencia-indexek első (= jellel jelölt) előfordulásakor jelöltük explicite, a további előfordulásokkor nem. (Tehát az első előfordulás után már csak [i01] fog szerepelni a 'Szent János Apostol Jelenéseinek Könyve' koreferenciális elem indexeként, és nem az (i03)(i04)[i01] koreferenciális kifejezés.) Ennek oka, hogy a számítógépes feldolgozó program az első előfordulás után már „tudni fogja”, tehát automatikusan figyelembe veszi a jelölt kapcsolatot, így felesleges azt többször jelölni. A számítógép által generált kapcsolatok pl. az előző példa esetén ugyanis

```
// [g01] = [i03] G [i04]
// [g02] = [i04] C [i03]
// [g03] = [i04] G [i01]
// [g04] = [i01] C [i04],
```

(részlet az ún. G/C relációk listájából)

ahol a G/C relációk *jelzik* az [i03] és [i04], valamint [i04] és [i01] koreferencia-indexek között fennálló relációt (amiből egyébként a reláció tranzitivitása folytán következik az [i03] és [i01] közötti reláció). „Lefordítva”, pl. [g01]-ben a G reláció arra utal, hogy a 'Szent János Apostol jelenései' koreferenciális elemében a 'Szent János Apostol' koreferenciális elem (itt: birtokos jelzőként) előfordul (*generate*), míg a [g02] index ennek a relációnak a „fordítottját”, inverzét jelöli, a C relációval (*cut*). Ezeknek a relációknak a szöveg szövegtani

vizsgálatakor, a szöveg asszociatív kapcsolatainak feltárásakor lesz nagy jelentősége.

Végezetül még egy jelölésről kell szólnunk, amelyet a számítógépes feldolgozás segítésére vezettünk be. Látni fogjuk később, hogy a koreferenciális elemek közötti kapcsolatokat egy speciális adatszerkezetben, ún. koreferenciaindex-hálóban foglaltuk össze. Ebben minden kommunikációs egység pontosan egyszer szerepel, mint a „!” jellel jelölt, ún. *bázis koreferenciaindexnek* megfelelő fogalom (koreferenciális elem) egy *tulajdonsága*. Megjegyezzük, hogy megfigyeléseink szerint ez a jelölés a kommunikációs egység ún. *kommunikatív jelentésével* áll kapcsolatban, ugyanis gyakorlatilag minden esetben a kommunikációs egység szövegben szereplő lineáris elrendezettségét vettük alapul a „!” jellel jelölt bázis koreferenciaindex kiválasztásakor. (A referenciális, értelmi és kommunikatív jelentéssel kapcsolatban lásd PETŐFI 1998)

**Az első szövegmondat második kommunikációs egységének koreferenciális elemzése:**

[k02] = [=i06] [=r01]/t0, kij [!i02] + [i05],[k03]

// [k02/&vb/&ind]:  
amelyet (=a kinyilatkoztatást)[i02]  
azért...  
adott[i06][=r01]  
neki (=Jézus Krisztusnak)[i05]  
az^Isten,[=i06]  
...hogy[k03]

[i06] Isten  
[r01] vki(a) kinyilatkoztatást(t) ad/tesz(R) vkinek(v) vmiért(w)  
t0: Isten kinyilatkoztatásának ideje

Az előző kommunikációs egység elemzéséhez képest a legnagyobb különbség az [r01] *relációs index* megjelenése, amely az elemzett kommunikációs egység, mint mondat állítmányát adó ige (itt: „adott”) szótári alakját tartalmazza azzal a „mintával” együtt, amely meghatározza a kommunikációs egység további elemeinek mondatbeli funkcióját (*verb pattern*, lásd pl. HORNBY 1980). Esetünkben ez a (logikai) alanyt (a), az igét (R), az ige (kötelező) bővítményét (t) (itt: tárgy, de pl. [r06] esetén állandó határozó), a részeshatározót (v) és a célhatározót (w) jelenti. Az állítmányt R-rel jelölve a [k02] kommunikációs index felépítése definíció szerint az

**a R t + v, w**

mintát követi (ahol a „+” és „,” operátoroknak elválasztó funkciójuk van). A relációs indexek meghatározásakor a továbbiakban mindig ezt a mintát fogjuk alapul venni, legfeljebb további mondatrészekkel kiegészítve (pl. x, y, ... esetén: aRt+v,w,x,y,...), ill. hiányzó mondatrészek esetén a „,” operátorral „kitöltve” a minta megfelelő elemét (pl. [r09] esetén az aR. minta, [r15] esetén pedig az aR.+v minta adódik). Érdemes észrevenni, hogy a mintában a koreferencia-indexek mellett a kommunikációs indexek is szerepelhetnek (pl. [k02]-ben w helyén a [k03] index), valamint hogy egyes mondatrészek „halmozottan” is előfordulhatnak, ilyenkor az azonos mondatbeli funkcióval rendelkező koreferencia-indexeket az „&” operátorral kapcsoljuk össze (lásd pl. [k07]-ben). Az „=” operátort a koreferenciális elemzésben megszokott értelemben használjuk a relációs indexek esetében is (tehát azok első előfordulásakor).

Mivel a relációs indexek csak az ige szótári alakját tartalmazzák, ezért szükséges a relációs indexek kiegészítése két további „dimenzióval”: az igeidővel és egy további („egyéb”), a jelentést módosító elemmel, amit rövidítve tüntetünk fel (itt: az igemóddal, t.i. ’kij’ a kijelentő módra utal). Ezeket az információkat a / jel után és vesszővel elválasztva, közvetlenül a relációs index után tüntetjük fel. Az elemzett szöveg időbeli kiterjedését nagyon jól jelzik a lehetséges idődimenziók:

- t0: Isten kinyilatkoztatásának ideje
- t1: Jézus Krisztus kinyilatkoztatásának ideje
- t2: a jelenések ideje
- t3: a Jelenések Könyve megírásának ideje
- t4: a Jelenések Könyvében leírtak megismerésének ideje
- t5: a kinyilatkoztatásban szereplő események bekövetkezésének ideje

Jellemző a Jelenések Könyvének szemantikai zártságára, hogy a fenti idődimenziók mindegyike előfordul (explicite v. implicite) az elemzett szövegrészletben!

**Az első szövegmondat harmadik kommunikációs egységének koreferenciális elemzése:**

[k03] = [!i05] [=r02]/t1, felsz [=i07] + (i05)[=i08]

// [k03/&vb/&ind]:  
(Jézus Krisztus)[i05]  
szolgái(i05)[=i08]nak  
megmutasson[i05][=r02]  
mindent (=minden eseményt),^aminek^csakhamar^be^kell^következnie.[=i07]

[i08] Jézus Krisztus szolgálja  
[i07] minden esemény, aminek csakhamar be kell következnie  
[r02] vki(a) megmutat(R) vmit(t) vkinek(v)  
t1: Jézus Krisztus kinyilatkoztatásának ideje

**Kommentár:**

[c01] = [!i07] [=!r03]/t5, kell.

('i07' mindazokat az eseményeket jelenti, amik hamarosan be kell, hogy következzenek – más szóval biztosan bekövetkeznek)

[r03] vmi(a) bekövetkezik/megtörténik(R)  
t5: a kinyilatkoztatásban szereplő események bekövetkezésének ideje („hamarosan”)

Maga az elemzés nem tartalmaz új elemet az előzőekhez képest. Új viszont egy [c01] indexszel jelölt ún. *kommentár egység* megjelenése. Ismeretes, hogy a szövegek szemantikai tartalmának értelmezéséhez nem elegendő a szövegek vizsgálata pusztán szövegnyelvészeti eszközökkel. A teljes jelentéstartalom feltárásához szövegtani eszközökre is szükség van, amelyek képesek a *világra vonatkozó* ismeretek kezelésére is (lásd PETŐFI 1998). Ezeket az ismereteket építjük be a koreferenciális elemzés eszköztárába a kommentárok segítségével, amelyeket a számítógépes feldolgozás érdekében ugyanúgy formalizálunk, mint a kommunikációs egységeket (Co/&ind reprezentáció). Valójában közöttük csak a forrásban van különbség: a kommunikációs egységek szigorúan az elemzett szöveg részét kell, hogy képezzék, míg a kommentár egységek az elemző „világra vonatkozó” háttérismeretein alapulnak (amely magában foglalja az elemző általános, köznapi (kognitív) sémáit, nyelvi, irodalmi és egyéb lexikális ismereteit, az elemzett szövegrészletet tartalmazó teljes művet, annak keletkezési körülményeit, a mű mások által végzett elemzéseit, stb.).

Új elem a [c01] kommentár egység elemzésekor a „!” operátor használata a relációs indexben. Ez azt jelöli, hogy a [c01] kommentár az [i07] koreferenciális elem *definitív tulajdonsága* (azaz: olyan tulajdonság a koreferenciaindex-hálóban, amely a koreferenciális elemre mindig teljesül).

Ezzel az első szövegmondat elemzése kapcsán minden olyan lényeges elemet áttekintettünk, amely szükséges a további szövegmondatok koreferenciális elemzésének megértéséhez.

**A második szövegmondat koreferenciális elemzése:**

[K02] = [k04]&[k05]&[k06]&[k07]

Elküldte angyalát, így adta tudtul szolgájának, Jánosnak, <sup>2</sup>aki tanúskodik Isten Igéjéről és Jézus Krisztus tanúságtételéről: mindenről, amit látott.

**A második szövegmondat első kommunikációs egységének elemzése:**

[k04] = [!i05] [=r04]/t1,kij (i05)[=i09]

// [k04/&vb/&ind]:  
(Jézus Krisztus)[i05]  
elküldte[i05][=r04]  
angyalá(i05)[=i09])t,

[i09] Jézus Krisztus angyala  
[r04] vki(a) elküld(R) vkit(t) vkihez(v)

**A második szövegmondat második kommunikációs egységének elemzése:**

[k05] = [i05] [=r05]/t1,kij [i02] + [=i10],[!i09]

// [k05/&vb/&ind]:  
így (=angyala[i09] révén)  
adta^tudtul[i05][=r05]  
(Jézus Krisztus)[i05]  
(a kinyilatkoztatást)[i02]  
...  
János[=i10]nak,

[i10] János  
[r05] vki(a) tudtul ad/tudat(R) vmit(t) vkinek(v) vhogyan(w)

**Kommentár:**

[c02] = [!i03] U [i10]  
(János Szent János Apostolt jelenti)



**A második szövegmondat harmadik kommunikációs egységének elemzése:**

[k06] = [!i10] B (i05)[=i11]

// [k06/&vb/&ind]:  
szolgájá(i05)[=i11]nak,  
János[i10]nak,

[i11] Jézus Krisztus szolgája

**Kommentár:**

[c03] = [!i11] B\* [i08]

(Jézus Krisztus szolgája Jézus Krisztus szolgálói közé tartozik)

**A második szövegmondat negyedik kommunikációs egységének elemzése:**

[k07] = [!i10] [=r06]/t3,kij (i06)[=i12]&(i05)[=i13]&[=i14]

// [k07/&vb/&ind]:  
aki (=János)[i10]  
tanúskodik[i10][=r06]  
Isten[i06]^Igéjé(i06)[=i12]ről  
és Jézus^Krisztus[i05]^tanúságtételé(i05)[=i13]ről:  
mindenről (=minden dologról),^amit^(János[i10])^látott[=i14].  
[i12] Isten Igéje  
[i13] Jézus Krisztus tanúságtétele  
[i14] minden dolog, amit János látott  
[r06] vki(a) tanúskodik(R) vmiről(t)  
t3: a Jelenések Könyve megírásának ideje

**Kommentár:**

[c04] = [i10] [=!r07]/t2,kij [!i14]

('i14' mindazokat a dolgokat jelenti, amiket János látott)  
[r07] vki(a) lát(R) vmit(t)  
t2: a jelenések ideje

[c05] = [!i10] [=r07]/t2,kij [i07]

## A koreferencia kérdései a számítógépes szövegfeldolgozás szempontjából

(János látott mindent, aminek csakhamar be kell következnie)

[c06] = [!i07] B [i14]

(minden esemény, aminek csakhamar be kell következnie, része annak, amit János látott)

[c07] = [!i04] U [i14]

(János jelenései mindazok a dolgok, amiket János látott)

[c08] = [!i10] [=r08]/t3,kij [i14] + [i01]

(János leírta a Jelenések Könyvében mindazt, amit látott)

[r08] vki(a) ír(R) vmit(t) vkinek(v) vhol(w)

### **A harmadik szövegmondat koreferenciális elemzése:**

[K03] = [k08]&[k09]

<sup>3</sup>Boldog, aki olvassa és aki hallgatja ezeket a prófétai szavakat, és meg is tartja, ami bennük meg van írva, mert az idő közel van.

### **A harmadik szövegmondat első kommunikációs egységének koreferenciális elemzése:**

[k08] = [=!i15] [=r09]/t4,kij. + [k09]

// [k08/&vb/&ind]:

Boldog[i15][=r09],

aki[=i15]^olvassa^és^aki^hallgatja^

ezeket^a^(Jelenések Könyvé[i01]ben levő)^prófétai^szavakat(i01)[=i16],^

és^meg^is^tartja,^

ami[=i17]^bennük[i16]^meg^van^írva,

mert[k09]

[i15] (az az) ember, aki olvassa és aki hallgatja a Jelenések Könyvében levő prófétai szavakat és meg is tartja, ami bennük meg van írva

[i16] (a) Jelenések Könyvében levő prófétai szavak

[i17] (azok a) dolgok, amik a Jelenések Könyvében levő prófétai szavakban meg vannak írva

[r09] vki(a) boldog/áldott(R) vmiért(v)

t4: a Jelenések Könyvében leírtak megismerésének ideje

**Kommentár:**

[c09] = [!i15] [=!r10]/t4,kij (i01)[=i16]

('i15' olyan embert jelent, aki olvassa a Jelenések Könyvében levő prófétai szavakat)

[r10] vki(a) olvas(R) vmit(t)

[c10] = [!i15] [=!r11]/t4,kij [i16]

('i15' olyan embert jelent, aki hallgatja a Jelenések Könyvében levő prófétai szavakat)

[r11] vki(a) hallgat(R) vmit(t)

[c11] = [!i15] [=!r12]/t4,kij [=i17]

('i15' olyan embert jelent, aki megtartja azokat a dolgokat, amik a Jelenések Könyvében levő prófétai szavakban meg vannak írva)

[r12] vki(a) megtart/megszívlel(R) vmit(t)

[c12] = [!i17] [=!r13]/t3,kij. + [i16]

('i17' (mind)azokat a dolgokat jelenti, amik meg vannak írva a Jelenések Könyvében levő prófétai szavakban)

[r13] vmi(a) meg/le van írva(R) vhol(v)

[c13] = [!i16] [!r13]/t3,kij. + [i01]

(a Jelenések Könyvében levő prófétai szavak meg vannak írva a Jelenések Könyvében)

Megjegyzés: az 'aRtvw' formális mintával leírt 'r08' reláció megfelel az 'a1R1v1' formális mintával leírt 'r13' relációnak az 'a1=t', 'R1=R' és 'v1=w' helyettesítések mellett, miközben 'a' határozatlan; ennek megfelelően

### A koreferencia kérdései a számítógépes szövegfeldolgozás szempontjából

[c14] = .[!r08]/t3,kij [!i17] + [i16]

másrészt

[c15] = .[!r08]/t3,kij [!i16] + [i01]

tehát

[c16] = .[!r08]/t3,kij [!i17] + [i01]

(azok a dolgok, amik meg vannak írva a Jelenések Könyvében levő prófétai szavakban, meg vannak írva a Jelenések Könyvében)

ezt összevetve [c08]-cal, adódik

[c17] = [!i14] B [i17]

(mindaz, amit János látott, része azoknak a dolgoknak, amik meg vannak írva a Jelenések Könyvében levő prófétai szavakban)

#### **A harmadik szövegmondat második kommunikációs egységének koreferenciális elemzése:**

[k09] = [=!i18] [=r14].

// [k09/&vb/&ind]:

az^idő[=i18]^(amikor a megjövendölt események[i07] bekövetkeznek)

közel van[i18][=r14].

[i18] (az) idő, amikor a megjövendölt események bekövetkeznek

[r14] vmi(a) (időben) közel van(R)

#### **Kommentár:**

[c18] = [i07] [=!r15]/t5,kij. + [!i18]

('i18' azt az időt jelenti, amikor azok az események, amiknek a jövendölés szerint csakhamar be kell következnie, bekövetkeznek)

[r15] vmi(a) bekövetkezik(R) vmikor(v)

Megjegyzés: a t5 idősík [i18]-nak felel meg

**A negyedik szövegmondat elemzése:**

[K04] = [k10]

<sup>4</sup>János a hét ázsiai egyháznak.

**A negyedik szövegmondat első kommunikációs egységének koreferenciális elemzése:**

[k10] = [!i10] [r08]/t3,kij [k11] + [=i19]

// [k10/&vb/&ind]:

János[i10]

(írja)[i10][r08]

(a következőket)[k11]

a^hét^ázsiai^egyház[=i19]nak.

[i19] (a) hét ázsiai egyház

**Az ötödik szövegmondat koreferenciális elemzése:**

[K05] = [k11]&[k12]&[k13]&[k14]&[k15]

Kegyelem és békesség nektek attól, aki van, aki volt és aki eljő, meg a hét lélektől, aki trónja előtt áll, <sup>5</sup>és Jézus Krisztustól, a hűséges tanútól, a halottak elsősülöttétől és a föld királyainak fejedelmétől!

**Az ötödik szövegmondat első kommunikációs egységének koreferenciális elemzése:**

[k11] = .[=r16]/t4,felsz. + [!i19],[i06]&[=i20]&[i05]

// [k11/&vb/&ind]:

Kegyelem és békesség (legyen)[=r16]

nektek (=a hét ázsiai egyháznak)[i19]

at[i06]től (=Istentől),

...

meg a^hét^lélek[=i20]től,

^aki (=a hét lélek)[i20]^(Isten[i06])^trónja(i06)[=i21]^előtt^áll,

és Jézus Krisztus[i05]től,

[i20] (a) hét lélek, aki Isten trónja előtt áll

[i21] Isten trónja

[r16] kegyelem és békesség legyen(a-r) vkinek(v) vkitől(w)

**Kommentár:**

[c19] = [!i20] [=!r17]/t4,kij. (i06)[=i21]

(az 'i20' indexszel azonosított hét lélek Isten trónja előtt áll)

[r17] vki(a) áll(R) vhol(v)

**Az ötödik szövegmondat második kommunikációs egységének koreferenciális elemzése:**

[k12] = [!i06] [=!r18]/t3,kij.

// [k12/&vb/&ind]:

(Isten)[i06]

aki (=Isten) van(i06)[=r18],

[r18] Isten(a) van(R)

**Az ötödik szövegmondat harmadik kommunikációs egységének koreferenciális elemzése:**

[k13] = [!i06] [=!r19]/t0,kij.

// [k13/&vb/&ind]:

(Isten)[i06]

aki (=Isten) volt(i06)[=r19] és

[r19] Isten(a) volt(R)

**Az ötödik szövegmondat negyedik kommunikációs egységének koreferenciális elemzése:**

[k14] = [!i06] [=!r20]/t5,kij.

// [k14/&vb/&ind]:

(Isten)[i06]

aki (=Isten) eljő(i06)[=r20]

[r20] Isten(a) eljön(R)

**Az ötödik szövegmondat ötödik kommunikációs egységének koreferenciális elemzése:**

[k15] = [!i05] U [=i22]

// [k15/&vb/&ind]:  
(Jézus Krisztus[i05]től)  
a hűséges^tanú[=i22]től,

[i22] a hűséges tanú

**Az ötödik szövegmondat hatodik kommunikációs egységének koreferenciális elemzése:**

[k16] = [!i05] U (=i24)[=i23]

// [k16/&vb/&ind]:  
(Jézus Krisztus[i05]től)  
a halottak[=i24]^elsősülötté(i24)[=i23]től

[i23] a halottak elsősülötte

[i24] a halottak

**Az ötödik szövegmondat hetedik kommunikációs egységének koreferenciális elemzése:**

[k17] = [!i05] U (=i26)(=i27)[=i25]

// [k17/&vb/&ind]:  
és (Jézus Krisztus[i05]től)  
a föld[=i26]^királyai(i26)[=i27]nak^fejedelmé(i27)[=i25]től!

[i25] a föld királyainak fejedelme

[i26] a föld

[i27] a föld királyai

## 2. A számítógépes szövegfeldolgozó program által szolgáltatott eredmények

Az általunk kifejlesztett szövegfeldolgozó programok, ill. az általuk szolgáltatott fájlok listája:

communit.lst	711	1-12-99	0:11	Ko/&ind reprezentáció
genrels.lst	1029	1-12-99	0:11	G/C relációk listája
komatrix.lst	26176	1-12-99	0:11	kapcsolati mátrixok listája
koref.exe	29600	1-11-99	23:05	szövegfeldolgozó program
koref.ind	8791	1-11-99	21:50	koreferenciális elemzés szövege
koref.pas	27612	1-11-99	23:05	szövegfeldolgozó program forráskódja
korefind.lst	1054	1-12-99	0:11	koreferencia-indexek jegyzéke
reindex.exe	7712	12-30-98	22:16	újraindexelő program
reindex.pas	4697	12-30-98	22:16	újraindexelő program forráskódja
relatind.lst	816	1-12-99	0:11	relációs indexek jegyzéke
suppunit.lst	691	1-12-99	0:11	Co/&ind reprezentáció
szemant.ind	4900	1-12-99	0:11	koreferenciaindex-háló
textsnt.lst	153	1-12-99	0:11	szövegmondatok index-reprezentációja

(A programok Turbo Pascal 7.0-ban készültek.)

A koreferencia-indexek és relációs indexek jegyzéke, a Ko/&ind és Co/&ind reprezentáció és a szövegmondatok index-reprezentációja manuálisan, kézzel is elkészíthető, és nem igényel különösebb kommentárt. (Más a helyzet új indexek beszúrásakor, tehát utólagos korrekció, ill. bővítés esetén — ilyenkor már hatalmas segítség a számítógépes program által kínált automatikus újraindexelés.) A számítógép által generált listák a mellékletben megtalálhatóak.

A G/C relációk listája, valamint a koreferenciaindex-hálóban a bázis koreferencia-indexek osztályképző, paradigmaticai (B, B\*, U) relációi fogalomcsoportok, ún. kulcsszó-listák automatikus kialakítását teszik lehetővé, amit pl. irodalmi művek komplex elemzésekor, így a szerzők által kidolgozott *kapcsolatelemzési módszer* során alkalmazhatunk (lásd BODA 1999).

A számítógép igazi „erejét” a koreferenciaindex-háló és az indexek kapcsolatát, összefüggését jellemző ún. kapcsolati mátrixok generálásakor mutatja meg; ezeket gyakorlatilag nem, vagy csak nagyon nagy munkával lehetne manuálisan létrehozni. Először a koreferenciaindex-háló felépítését tekintjük át, majd az ebből létrehozható kapcsolati mátrixokat, ill. ezek lehetséges alkalmazását vizsgáljuk meg.

A koreferenciaindex-háló egységei, „szócikkei” az elemzett szöveg koreferencia-indexeinek kapcsolatrendszerét és tulajdonságait írják le az elvégzett



koreferenciális elemzés eredményei alapján. Vizsgáljuk meg a generált koreferenciaindex-háló néhány kiemelt szócikkét:

[i01] Szent János Apostol Jelenéseinek Könyve

OSZTÁLY

[k01] = (=i03)(=i04)[=!i01] B (=i05)[=i02]

KAPCSOLAT

[g04] = [i01] C [i04]

[g24] = [i01] G [i16]

[g29] = [i01] Gd [i17]

DEFINÍCIÓ

TULAJDONSÁG

[i02] Jézus Krisztus kinyilatkoztatása

OSZTÁLY

KAPCSOLAT

[g06] = [i02] C [i05]

[g07] = [i02] Co [i01]

DEFINÍCIÓ

TULAJDONSÁG

[k02] = [=i06] [=r01]/t0,kij [!i02] + [i05],[k03]

[i03] Szent János Apostol

OSZTÁLY

[c02] = [!i03] U [i10]

KAPCSOLAT

[g01] = [i03] G [i04]

DEFINÍCIÓ

TULAJDONSÁG

.....

[i05] Jézus Krisztus

OSZTÁLY

[k15] = [!i05] U [=i22]

[k16] = [!i05] U (=i24)[=i23]

[k17] = [!i05] U (=i26)(=i27)[=i25]

KAPCSOLAT

## A koreferencia kérdései a számítógépes szövegfeldolgozás szempontjából

---

[g05] = [i05] G [i02]

[g08] = [i05] G [i08]

[g10] = [i05] G [i09]

[g13] = [i05] G [i11]

[g19] = [i05] G [i13]

DEFINÍCIÓ

TULAJDONSÁG

[k03] = [!i05] [=r02]/t1,felsz [=i07] + (i05)[=i08]

[k04] = [!i05] [=r04]/t1,kij (i05)[=i09]

[i06] Isten

OSZTÁLY

KAPCSOLAT

[g17] = [i06] G [i12]

[g32] = [i06] G [i21]

DEFINÍCIÓ

[k12] = [!i06] [=!r18]/t3,kij.

[k13] = [!i06] [=!r19]/t0,kij.

[k14] = [!i06] [=!r20]/t5,kij.

TULAJDONSÁG

.....

[i10] János

OSZTÁLY

[k06] = [!i10] B (i05)[=i11]

KAPCSOLAT

[g12] = [i10] Co [i03]

[g21] = [i10] Gd [i14]

DEFINÍCIÓ

TULAJDONSÁG

[k07] = [!i10] [=r06]/t3,kij (i06)[=i12]&(i05)[=i13]&[=i14]

[k10] = [!i10] [r08]/t3,kij [k11] + [=i19]

[c05] = [!i10] [=r07]/t2,kij [i07]

[c08] = [!i10] [=r08]/t3,kij [i14] + [i01]

Látható, hogy minden szócikk egy megadott koreferencia-index, mint *címszó* négy alapvető jellemzőjét tartalmazza. Az OSZTÁLY jellemző az adott koreferencia-index paradigmaticai (U, B, B\*) relációit tartalmazza, és (referenciális)

jelentésük alapján egy osztály-hierarchiát definiál a koreferenciaindex-háló címszavai között. (Ez egyebek között az egyes fogalmak „alternatíváit” szolgáltatja, ami pl. a mondandó tárgya (referens) megkülönböztetését teszi lehetővé — lásd EYSENCK, 1997. 354. o.)

A KAPCSOLAT jellemző egy megadott koreferencia-index (címszó) alapvető szintagmatikai kapcsolatait adja meg. Három reláció-típust használunk:  
– a G/C relációkat (korábban már volt szó róluk);  
– a Co relációt, amely a B reláció „inverzének” tekinthető, és a generikus fogalom szócikkének KAPCSOLAT jellemzőjében jelenik meg; például

// [k01] = (=i03)(=i04)[=!i01] B (=i05)[=i02]

az [i01] koreferencia-index szócikkének OSZTÁLY jellemzőjében szerepel, a belőle kapott

// [g07] = [i02] Co [i01]

reláció viszont már az [i02] koreferencia-index KAPCSOLAT jellemzőjeként jelenik meg.

– és a Gd relációt, amely a DEFINÍCIÓ jellemzőhöz besorolt (definitív) tulajdonságok „invertálását” valósítja meg, azaz lehetővé teszi, hogy a definitív tulajdonságban megvalósuló szintagmatikai relációk ne csak a kijelölt bázis koreferencia-index szócikkében, hanem a tulajdonságban szereplő összes koreferencia-index szócikkében is megjelenjenek. Vegyük például a [c04] kommentár egységet:

// [c04] = [i10] [=!r07]/t2,kij [!i14]

('i14' mindazokat a dolgokat jelenti, amiket János látott)

a kommentár egység bázis koreferencia-indexe [i14], ezért a [c04] kommentár egység az [i14] szócikkének DEFINÍCIÓ jellemzőjeként szerepel a koreferenciaindex-hálóban; azonban a

// [g21] = [i10] Gd [i14]

## A koreferencia kérdései a számítógépes szövegfeldolgozás szempontjából

reláció alapján a (definitív) tulajdonságban kifejezett szintagmatikai reláció [i10] szócikkének KAPCSOLAT jellemzőjében is megjelenik.

A KAPCSOLAT jellemző a kommunikációs egységek szintagmatikai szerkezete által létrejött *értelmi jelentéssel* áll szoros kapcsolatban, és lehetővé teszi az elemzett szöveg asszociatív szerkezetének értelmezését, amiről részletesebben a kapcsolati mátrixok vizsgálatakor szólunk.

A DEFINÍCIÓ jellemző azokat a kommunikációs, ill. kommentár egységeket tartalmazza, amelyek az egység *kommunikatív jelentése* alapján kijelölt bázis koreferencia-index (értelmi) jelentése szempontjából meghatározó, definitív jellegűek. Másképp fogalmazva a definitív tulajdonságokat a bennük szereplő bázis koreferencia-index minden szövegbeli előfordulásakor implicite beleértjük az értelmezett szövegbe.

A TULAJDONSÁG jellemző azokat a kommunikációs, ill. kommentár egységeket tartalmazza, amelyek az elemzett szövegben előfordulnak, de nem definitív jellegűek; azaz ezeket a tulajdonságokat a bennük szereplő bázis koreferencia-index egy adott szövegbeli előfordulásakor explicite fel kell tüntetnünk, ha az adott kontextusban érvényesek.

A kapcsolati mátrixok közül először tekintsük a legalapvetőbb, a manuális koreferenciális elemzésben is használt koreferenciaindex-kommunikációs egység mátrixot. A mátrix soraiban a koreferencia-indexek sorszámai, oszlopaiban a kommunikációs egységek indexeinek sorszámai szerepelnek. A mátrix celláiban o-val jelöltük a bázis koreferencia indexeket, és x-vel a kommunikációs egységekben előforduló további koreferencia-indexeket. (ld. az 1. ábrát!)

**KOREFERENCIAINDEX-KOMMUNIKÁCIÓS EGYSÉG MÁTRIX/1**

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
1	o																	1
2	x	o			x													3
3																		0
4																		0
5		x	o	o	x						x				o	o	o	8
6		x									x	o	o	o				5
7			x															1
8			x															1
9				x	o													2
10					x	o	o			o								4
11						x												1
12							x											1
13								x										1
14								x										1
15									o									1
16																		0
17																		0
18									o									1
19										x	o							2
20											x							1
21																		0
22															x			1
23																x		1
24																		0
25																	x	1
26																		0
27																		0
		2	3	3	2	4	2	4	1	1	2	4	1	1	1	2	2	2

1. ábra

(A mátrix utolsó oszlopa a koreferencia-indexek előfordulási gyakoriságát, utolsó sora pedig a kommunikációs egységekben szereplő koreferencia-indexek számát tünteti fel.)

Említettük, hogy a kapcsolati mátrixok az elemzett szöveg asszociatív szerkezetének feltárásában játszanak alapvető szerepet. Ehhez először definiálnunk kell, mit értünk az adott modellben asszociáció alatt:

– asszociáció két kommunikációs vagy kommentár egység között lehetséges; jelöljük ezt a továbbiakban  $A(oa, x_1, x_2, \dots) \rightarrow B(ob, y_1, y_2, \dots)$ -vel, ahol  $oa$  és  $ob$  az A, ill. B kommunikációs vagy kommentár egységek bázis koreferencia-indexét,  $x_1, x_2, \dots$  és  $y_1, y_2, \dots$  pedig az A, ill. B-ben előforduló további koreferencia-indexeket jelöli. Jelöljük továbbá egy adott X koreferencia-index OSZTÁLY jellemzőjében szereplő, vele paradigmaticai relációban levő koreferencia-indexek halmazát OSZTÁLY(X)-vel, és az X koreferencia-index

## A koreferencia kérdései a számítógépes szövegfeldolgozás szempontjából

KAPCSOLAT jellemzőjében szereplő, vele szintagmatikai relációban levő koreferencia-indexek halmazát KAPCSOLAT(X)-vel.

– az asszociáció három típusát különböztetjük meg:

1. típusú (egyszerű vagy közvetlen) asszociáció:  $A(oa, x_1, x_2, \dots) \rightarrow B(ob, y_1, y_2, \dots)$ , ahol  $ob$  eleme az  $\{oa, x_1, x_2, \dots\}$  halmaznak.
2. típusú (kiterjesztett) asszociáció:  $A(oa, x_1, x_2, \dots) \rightarrow B(ob, y_1, y_2, \dots)$ , ahol  $ob$  eleme az OSZTÁLY( $oa$ ) vagy KAPCSOLAT( $oa$ ) halmaznak.
3. típusú (bővített) asszociáció:  $A(oa, x_1, x_2, \dots) \rightarrow B(ob, y_1, y_2, \dots)$ , ahol  $ob$  eleme a OSZTÁLY( $x$ ) vagy KAPCSOLAT( $x$ ) halmaznak valamely  $x = x_1, x_2, \dots$  koreferencia-indexre.

Mielőtt ezek előfordulását vizsgálnánk, fogalmazzunk meg egy hipotézist, amely az általunk alkalmazott textológiai modell alapját képezi:

Egy szöveget akkor tartunk összefüggőnek, ha a benne szereplő kommunikációs egységek egymással közvetlen, vagy az elemzés során bevezetett kommentár egységeken keresztül közvetett asszociatív kapcsolatban állnak. Másként megfogalmazva: egy intuitíven összefüggőnek tekintett szöveg elemzése akkor tekinthető teljesnek, ha a szöveget kommentár egységekkel kiegészítve a fenti értelemben összefüggőnek bizonyul.

Látható, hogy a bemutatott koreferenciaindex-kommunikációs egység mátrix a legegyszerűbb, egyszerű asszociáción alapuló kapcsolatokat tünteti fel. Kettős kötőjellel (--) jelölve ezeket, a következő mátrixhoz jutunk: (ld. a 2. ábrát!)

**KOREFERENCIAINDEX-KOMMUNIKÁCIÓS EGYSÉG MÁTRIX**

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17		
1	o																	1	
2	x--o				x													3	
3																		0	
4																		0	
5		x--o--o		x							x--.....--o--o--o							8	
6		x									x--o--o--o							5	
7			x															1	
8			x															1	
9				x--o														2	
10					x--o--o--.....--o													4	
11						x												1	
12							x											1	
13								x										1	
14									x									1	
15										o								1	
16																		0	
17																		0	
18											o							1	
19												x--o						2	
20													x					1	
21																		0	
22																x		1	
23																	x	1	
24																		0	
25																		x	1
26																			0
27																			0
		2	3	3	2	4	2	4	1	1	2	4	1	1	1	2	2	2	

2. ábra

Figyeljük meg, ahogy kirajzolódtak az elemzett szövegrészlet összetartó kapcsolatok! Valójában a szöveg elemzése „majdnem” teljes, pusztán [k08] és [k09] kommunikációs egységek nincsenek asszociatív kapcsolatban a többi kommunikációs egységgel. Mielőtt megmutatnánk, hogy egy, a kiterjesztett és bővített asszociációkat, valamint a kommentár egységeket is tartalmazó kapcsolati mátrixszal a szöveg összefüggővé tehető, kövessük végig az elemzés folyamatát részletesen, hogy a mátrix által kimutatott összefüggések érthetőek legyenek.

Célunk tehát egy asszociációs lánc felépítése, a [k07] kommunikációs egységből kiindulva, és eljutva a szövegből egyenlőre „kilógó” [k08] és [k09] kommunikációs egységekhez.

Az asszociációs lánc például a következő lehet:

$$[k07] = [!i10] [=r06]/t3,kij (i06)[=i12]&(i05)[=i13]&[=i14]$$

## A koreferencia kérdései a számítógépes szövegfeldolgozás szempontjából

aki (=János)[i10]  
tanúskodik[i03][=r06]  
Isten[i06]^Igéjé(i06)[=i12]ről  
és Jézus^Krisztus[i05]^tanúságtételé(i05)[=i13]ről:  
mindenről (=minden dologról),^amit^(János[i10])^látott[=i14].

1. típusú (közvetlen) asszociáció: [i10]/[k07] -> [i10]/[c08]  
[c08] = [!i10] [=r08]/t3,kij [i14] + [i01]  
(János leírta a Jelenések Könyvében mindazt, amit látott)

3. típusú (bővített) asszociáció: [i01]/[c08] -> [i16]/[c13]  
[c13] = [!i16] [!r13]/t3,kij. + [i01]  
(a Jelenések Könyvében levő prófétai szavak meg vannak írva a Jelenések Könyvében)

2. típusú (kiterjesztett) asszociáció: [i16]/[c13] -> [i15]/[c09],[c10]  
[c09] = [!i15] [=!r10]/t4,kij (i01)[=i16]  
[c10] = [!i15] [=!r11]/t4,kij [i16]  
(‘i15’ olyan embert jelent, aki olvassa és hallgatja a Jelenések Könyvében levő prófétai szavakat)

1. típusú (közvetlen) asszociáció: [i15]/[c09],[c10] -> [i15]/[k08]  
[k08] = [=!i15] [=r09]/t4,kij. + [k09]  
Boldog[i15][=r09],  
aki[=i15]^olvassa^és^aki^hallgatja^  
ezeket^a^(Jelenések Könyvé[i01]ben levő)^prófétai^szavakat(i01)[=i16],^  
és^meg^is^tartja,^  
ami[=i17]^bennük[i16]^meg^van^írva,  
mert[k09]

1. típusú (közvetlen) asszociáció: [i15]/[k08] -> [i15]/[c11]  
[c11] = [!i15] [=!r12]/t4,kij [=i17]  
(‘i15’ olyan embert jelent, aki megtartja azokat a dolgokat, amik a Jelenések Könyvében levő prófétai szavakban meg vannak írva)

3. típusú (bővített) asszociáció: [i17]/[c11] -> [i14]/[c17]  
[c17] = [!i14] B [i17]  
(mindaz, amit János látott, része azoknak a dolgoknak, amik meg vannak írva a Jelenések Könyvében levő prófétai szavakban)

2. típusú (kiterjesztett) asszociáció: [i14]/[c17] -> [i07]/[c06]  
[c06] = [!i07] B [i14]



(minden esemény, aminek csakhamar be kell következnie, része annak, amit János látott)

2. típusú (kiterjesztett) asszociáció: [i07]/[c06] -> [i18]/[c18]

[c18] = [i07] [=!r15]/t5,kij. + [!i18]

('i18' azt az időt jelenti, amikor azok az események, amiknek a jövődőlés szerint csakhamar be kell következnie, bekövetkeznek)

1. típusú (közvetlen) asszociáció: [i18]/[c18] -> [i18]/[k09]

[k09] = [=!i18] [=r14].

az^idő[=i18]^(amikor a megjövendölt események[i07] bekövetkeznek) közel van[i18][=r14].

Nézzük ezek után, hogyan szemléltethető mindez egy kapcsolati mátrix segítségével. A minimális asszociációs mátrixban csak az itt felépített asszociációs láncnak megfelelő kapcsolatokat tüntettük fel, míg a teljes asszociációs mátrixban a teljes elemzés „kapcsolati térképe” megjelenik. A mátrixok első sorában a k, ill. c betűk a kommunikációs, ill. kommentár egységekre utalnak (tehát pl. 17c a [c17] kommentár egységet jelenti). A cellákban alkalmazott betűk jelentése:

o: kommunikációs egység bázis koreferencia-indexe

x: kommunikációs egység (nem bázis) koreferencia-indexe

O: kommentár egység bázis koreferencia-indexe

X: kommentár egység (nem bázis) koreferencia-indexe

k: kommunikációs egység bázis koreferencia-indexével kapcsolatban levő koreferencia-index (kiterjesztett asszociációs pont)

b: kommunikációs egység (nem bázis) koreferencia-indexével kapcsolatban levő koreferencia-index (bővített asszociációs pont)

K: kommentár egység bázis koreferencia-indexével kapcsolatban levő koreferencia-index (kiterjesztett asszociációs pont)

B: kommentár egység (nem bázis) koreferencia-indexével kapcsolatban levő koreferencia-index (bővített asszociációs pont)

Az áttekinthetőség kedvéért kötőjelekkel (-) feltüntettük a legfontosabb asszociatív kapcsolatokat mindkét mátrixban. A feltüntetett mátrixokban a [k09] és [k10] kommentár egységeket elválasztó vonal az elemzett szövegrészlet két bekezdését (Bevezetés: Jel 1:1-3, és Címzés: Jel 1:4-5) választja el. (ld. a 3. és 4. ábrákat!)

MINIMÁLIS ASSZOCIÁCIÓS MÁTRIX

	1k	2k	3k	4k	5k	6k	7k	8c	13c	9c	10c	8k	11c	17c	6c	18c	9k	10k	11k	12k	13k	14k	15k	16k	17k		
1		o	k			b	k	k	X	X	B	B			B	B			k								
2		x	-	o	k	k	x								B	B				b				k	k	k	
3						b	k	k	K										k								
4		k						b	B	B													K	B			
5		b	x	-	o	-	o	x	b	b										b	x	-	o	-	o	-	o
6		b	x			b		b												b	x	-	o	-	o	-	o
7			b	x	k	b	k	k	K											k	b			k	k	k	
8			b	x	k	b	b													k	b			k	k	k	
9			b	k	x	-	o													k	b			k	k	k	
10				b	x	-	o	-	o	-	O									k	b			k	k	k	
11			b	k	k	b	x	k	K											k	b			k	k	k	
12			b		b	k	x	K												k	b	k	k	k			
13			b	k	k	b	k	x	K											k	b			k	k	k	
14				b		b	k	x	X											k	b			k	k	k	
15																				k	b			k	k	k	
16		k																		k	b			k	k	k	
17		k																		k	b			k	k	k	
18				b																k	b			k	k	k	
19					b		k	k	K											k	b			k	k	k	
20																				k	b			k	k	k	
21			b																	k	b	k	k				
22			b	k	k	b														k	b			x	k	k	
23			b	k	k	b														k	b			k	x	k	
24																				k	b			b			
25			b	k	k	b														k	b			k	k	x	
26																				k	b						
27																				k	b					b	
		7	14	12	11	17	11	13	13	6	4	4	3	5	8	6	3	2		12	15	3	3	3	10	11	11

3. ábra

**A koreferencia kérdései a számítógépes szövegfeldolgozás szempontjából**

**TELJES ASSZOCIÁCIÓS MÁTRIX**

	1k	2k	3k	1c	4k	5k	2c	6k	3c	7k	8c	7c	13c	9c	10c	8k	11c	12c	14c	15c	16c	17c	4c	5c	6c	18c	9k	10k	11k	12k	13k	14k	19c	15k	16k	17k				
1	o	k			b	B	k		k	X	K	X	B	B		B	K	K	X	X	B	B	K															20		
2	x	-	o	k		k	x				B		B								B	B																13		
3					b	-	O	k		k	K	K											B	K														9		
4	k					K			b	B	-	O	B								B	B	K	K		B												11		
5	b	x	-	o	-	-	-	o	x		b	K	b																									13		
6	b	x				b				b																													10	
7		b	x	-	O	k	b	B	k														K	-	K	-	X	-	O	X	k							21		
8		b	x			k	b		b	X																													10	
9		b	k			x	-	o																															8	
10					b	x	-	X	-	o	-	K	-	o	-	O	B						K	X	-	O	B												13	
11		b	k		k	b	B	x	-	O	k	K												B	K														16	
12		b			b	B	k		x	K														B	K														13	
13		b	k		k	b	B	k		x	K													B	K														15	
14			b	K		b	B	k		x	X	X																											18	
15																																								10
16	k																																						12	
17	k									b	B	B	K	K	K	k	X	-	O	-	O	K	X	-	O	K	B												16	
18			b	K																																			6	
19						b	B	k		k	K																													9
20																																								3
21		b																																					6	
22		b	k		k	b																																	8	
23		b	k		k	b																																	8	
24																																							1	
25		b	k		k	b																																	8	
26																																								0
27																																								1
	7	14	12	3	11	17	10	11	4	13	13	7	6	4	4	3	5	5	5	6	7	8	11	10	6	3	2	12	15	3	3	3	3	10	11	11				

**4. ábra**

### 3. Elméleti vonatkozások

Végezetül néhány megjegyzést szeretnénk fűzni a leírtak elméleti háttérének, ill. lehetséges kapcsolódási pontjainak felvázolása céljából.

A kommunikációs egységek koreferenciális elemzésével kapott index-reprezentációt (Ko&ind) pusztán (vagy legalábbis túlnyomórészt) szövegnyelvészeti eszközökkel kaptuk, így ezek együttese tekinthető az elemzett szöveg egy lehetséges szövegnyelvészeti modelljének. Ezt kiegészítve formalizált, tehát koreferenciális elemzésnek alávetett kommentárokkal (Co&ind reprezentáció), lényegében külső, a *világra vonatkozó* ismeretekkel egészítettük ki a modellt. Ez tehát az elemzett szöveg egy lehetséges szövegtani modelljének tekinthető.

A koreferenciaindex-háló adatszerkezete leginkább a szakértői rendszerekben alkalmazott *keretekhez* (frame-ek, lásd pl. BORGULYA 1995) hasonlít. Tartalmilag az egynyelvű vagy értelmező szótárakhoz (ill. lexikonokhoz, tezauruszokhoz, stb.) áll közel, és ilyen értelemben a kognitív pszichológia szemantikai háló(zat) fogalmának (lásd pl. EYSENCK 1997) egy lehetséges reprezentációjának is tekinthető.

#### *Irodalomjegyzék*

- Boda I. Károly—Porkoláb Judit: Nyelvi megformálások értelmezése számítógépes programmal (A Jelenések Könyve néhány részlete alapján). *Szemiotikai Szövegtan 12.* (1999). (Megjelenés alatt.)
- Borgulya I.: *Szakértői rendszerek, technikák és alkalmazások.* ComputerBooks, 1995.
- Eysenck, M. W.—Keane, M.T.: *Kognitív pszichológia. Hallgatói kézikönyv.* Nemzeti Tankönyvkiadó, 1997.
- Hornby, A. S.—Cowie, A. P. ed.: *Oxford Advanced Learner's Dictionary of Current English.* Oxford University Press, 1980.
- Petőfi S. János: *Egy poliglott szövegnyelvészeti-szövegtani kutatóprogram. (Officina Textologica I.)* (1998).

## Melléklet

### 1. A koreferenciaindexek jegyzéke

- [i01] Szent János Apostol Jelenéseinek Könyve
- [i02] Jézus Krisztus kinyilatkoztatása
- [i03] Szent János Apostol
- [i04] Szent János Apostol jelenései
- [i05] Jézus Krisztus
- [i06] Isten
- [i07] minden esemény, aminek csakhamar be kell következnie
- [i08] Jézus Krisztus szolgálói
- [i09] Jézus Krisztus anyyala
- [i10] János
- [i11] Jézus Krisztus szolgálója
- [i12] Isten Igéje
- [i13] Jézus Krisztus tanúságtétele
- [i14] minden dolog, amit János látott
- [i15] (az az) ember, aki olvassa és aki hallgatja a Jelenések Könyvében levő prófétai szavakat és meg is tartja, ami bennük meg van írva
- [i16] (a) Jelenések Könyvében levő prófétai szavak
- [i17] (azok a) dolgok, amik a Jelenések Könyvében levő prófétai szavakban meg vannak írva
- [i18] (az) idő, amikor a megjövendölt események bekövetkeznek
- [i19] (a) hét ázsiai egyház
- [i20] (a) hét lélek, aki Isten trónja előtt áll
- [i21] Isten trónja
- [i22] a hűséges tanú
- [i23] a halottak elsőszülötte
- [i24] a halottak
- [i25] a föld királyainak fejedelme
- [i26] a föld
- [i27] a föld királyai

### 2. A relációs indexek jegyzéke

- [r01] vki(a) kinyilatkoztatást(t) ad/tesz(R) vkinek(v) vmiért(w)
- [r02] vki(a) megmutat(R) vmit(t) vkinek(v)
- [r03] vmi(a) bekövetkezik/megtörténik(R)
- [r04] vki(a) elküld(R) vkit(t) vkihez(v)
- [r05] vki(a) tudtul ad/tudat(R) vmit(t) vkinek(v) vhogyan(w)
- [r06] vki(a) tanúskodik(R) vmiről(t)
- [r07] vki(a) lát(R) vmit(t)

- [r08] vki(a) ír(R) vmit(t) vkinek(v) vhol(w)  
 [r09] vki(a) boldog/áldott(R)  
 [r10] vki(a) olvas(R) vmit(t)  
 [r11] vki(a) hallgat(R) vmit(t)  
 [r12] vki(a) megtart/megszívlel(R) vmit(t)  
 [r13] vmi(a) meg/le van írva(R) vhol(v)  
 [r14] vmi(a) (időben) közel van(R)  
 [r15] vmi(a) bekövetkezik(R) vmikor(v)  
 [r16] kegyelem és békesség legyen(a-r) vkinek(v) vkitől(w)  
 [r17] vki(a) áll(R) vhol(v)  
 [r18] Isten(a) van(R)  
 [r19] Isten(a) volt(R)  
 [r20] Isten(a) eljön(R)

### 3. A szövegmondatok index-reprezentációja

- [K01] = [k01]&[k02]&[k03]  
 [K02] = [k04]&[k05]&[k06]&[k07]  
 [K03] = [k08]&[k09]  
 [K04] = [k10]  
 [K05] = [k11]&[k12]&[k13]&[k14]&[k15]

### 4. A kommunikációs egységek index-reprezentációja (Ko/&ind reprezentáció)

- [k01] = (=i03)(=i04)[=!i01] B (=i05)[=i02]  
 [k02] = [=i06] [=r01]/t0,kij [!i02] + [i05],[k03]  
 [k03] = [!i05] [=r02]/t1,felsz [=i07] + (i05)[=i08]  
 [k04] = [!i05] [=r04]/t1,kij (i05)[=i09]  
 [k05] = [i05] [=r05]/t1,kij [i02] + [=i10],[!i09]  
 [k06] = [!i10] B (i05)[=i11]  
 [k07] = [!i10] [=r06]/t3,kij (i06)[=i12]&(i05)[=i13]&[=i14]  
 [k08] = [=!i15] [=r09]/t4,kij. + [k09]  
 [k09] = [=!i18] [=r14].  
 [k10] = [!i10] [r08]/t3,kij [k11] + [=i19]  
 [k11] = .[=r16]/t4,felsz. + [!i19],[i06]&[=i20]&[i05]  
 [k12] = [!i06] [=!r18]/t3,kij.  
 [k13] = [!i06] [=!r19]/t0,kij.  
 [k14] = [!i06] [=!r20]/t5,kij.  
 [k15] = [!i05] U [=i22]  
 [k16] = [!i05] U (=i24)[=i23]  
 [k17] = [!i05] U (=i26)(=i27)[=i25]

**5. A kommentár egységek index-reprezentációja (Co/&ind reprezentáció)**

[c01] = [!i07] [=!r03]/t5,kell.  
[c02] = [!i03] U [i10]  
[c03] = [!i11] B\* [i08]  
[c04] = [i10] [=!r07]/t2,kij [!i14]  
[c05] = [!i10] [=r07]/t2,kij [i07]  
[c06] = [!i07] B [i14]  
[c07] = [!i04] U [i14]  
[c08] = [!i10] [=r08]/t3,kij [i14] + [i01]  
[c09] = [!i15] [=!r10]/t4,kij (i01)[=i16]  
[c10] = [!i15] [=!r11]/t4,kij [i16]  
[c11] = [!i15] [=!r12]/t4,kij [=i17]  
[c12] = [!i17] [=!r13]/t3,kij. + [i16]  
[c13] = [!i16] [!r13]/t3,kij. + [i01]  
[c14] = .[!r08]/t3,kij [!i17] + [i16]  
[c15] = .[!r08]/t3,kij [!i16] + [i01]  
[c16] = .[!r08]/t3,kij [!i17] + [i01]  
[c17] = [!i14] B [i17]  
[c18] = [i07] [=!r15]/t5,kij. + [!i18]  
[c19] = [!i20] [=!r17]/t4,kij. (i06)[=i21]