

D 5334

KANDIDÁTUSI ERTEKEZÉS

Horváth Tibor

I S M E R E T S Z E R V E Z É S É S S Z I N T A X I S

Kandidátusi értekezés

OSZK

Országos Széchényi Könyvtár

Herváth Tibor

Budapest, 1987

TARTALOM

BEVEZETŐ	I
A HAGYOMÁNY	1
1. A PROBLÉMA	1
2. KLASSZIKUS RENDSZEREK	6
2.1 A tudáspiramis	6
2.2 Új strukturák	17
2.3 Facettás rendszerek	27
2.4 Gyakorlati célú mai rendszerek	41
A JELEN	53
3. ALAPKÉRDÉSEK	53
3.1 Kvantálás	53
3.2 A heurisztikus információ	60
3.3 Az osztályozás "követő volta"	64
3.4 Specifikusság	67
3.5 Relációmegőrző transzformáció	71
4. AUTOMATIKUS OSZTÁLYOZÁS	75
4.1 Az automatikus osztályozás természete és problémái	75
4.2 Az eljárás technikai leírása	83
4.3 Induló paraméterek	90
4.4 Az ismérvek elemzése	93
4.5 További problémák	96
5. A SZINTAXIS	98
5.1 Egy eszme születése	98
5.2 A relációkról	102
5.3 Intermezzo: a tézauruszok	107
5.4 Szintaxisok	112
5.5 Magyar alkalmazás	131
ÖSSZEGZÉS	139
JEGYZETEK	150

B E V E Z E T Ő

Értekezésem célja az osztályozáselmélet egy fontos gondolatának kifejtése és igazolása: az indextételnek nemcsak az a feladata, hogy az osztályzandó objektumok változatos ismérveit megadja, hanem az is, hogy az indexelő kifejezéseket olyan magasabb egységben szintetizálja, amely az eredeti közlemény gondolatának, mondandójának tömör reprodukálását is eléri. Oly módon, ahogyan a természetes nyelv szavait a grammatika mondatokká kapcsolja össze. Az osztályozó/indexelő eljárások legtöbbje az általánosan elterjedt és elfogadott modell értelmében az objektumokhoz a kifejezések egy részhalmazát rendeli. Ebből az eljárásból hiányzik a szintaxis, ugyanis a bevezetett relációk érvénye a kifejezéseket tartalmazó szótárra korlátozódik, a relációk ennek folytán nem képesek olyan szintaktikai funkciók betöltésére, amelyek révén a szerző által az eredeti szövegben létrehozott fogalmi kapcsolatok az indextételben megőrizhetők.

Ennek megfelelően az értekezés erőteljesen az osztályozás szintaktikai problémáira koncentrálna. A fenti gondolattal egy lépéssel tovább jut az eddigi osztályozási elméleteken.

A disszertáció nem akarja a különböző rendszereket leírni és ismertetni. Mindegyikből csak annyit idéz fel, amennyi a fenti gondolat kibontásához szükséges. Kivétel a PRECIS,

II.

mivel ennek hazai meghonosítását és a magyar nyelv sajátosságaihoz illesztését az Országos Pedagógiai Információs Rendszer valósította meg a PRECIS egyetlen magyarországi alkalmazásaként. Ezzel szemben az értekezés be kívánja mutatni ennek a gondolatnak érlelődését, megjelenésének szükségyszerűségét a tudományok szerkezetváltozásainak függvényében. Másfelől annak következményeként, ami az osztályozás feltételeinek fejlődéséből, pl. az adatfeldolgozó technikák megjelenéséből eredt, lehetőséget teremtve mind az osztályozás új elméleteinek kimunkálásához, mind a magas műveletszámú osztályozás bevezetéséhez.

Bár visszanyulik az osztályozás kezdeteihez, a disszertáció nem kívánja az előzmények történetét sem felvázolni. Más kérdés, hogy a különböző osztályozáselméleti iskolák logikai vagy nyelvi megfontolásaikban hordozott, az értekezés mondandóját támogató eszmék, elvek történeti korszakokhoz is köthetők.

Az értekezés számára fontosak jegyzetei. Köztük számos lényegi megjegyzés található, utalván a téma elágazásaira. A szövegbe építve azonban megtörték volna a gondolatmenet ívét. Ugyanitt található a közvetlenül felhasznált és idézett irodalom. A disszertációt azonban jóval szélesebb körű szakirodalom támogatja főleg a rokontudományok köréből. Ezek felhasználása nem annyira közvetlen, inkább heurisztikus természetű. Ha pl. egy filozófus arra az eredményre jut,

III.

hogy a tudományrendszereknek világkép helyettesítő feladatai vannak, akkor egy tudományos kérdés kétezer éves erőfeszítéseire nyújt magyarázatot. De a megállapítás nem az informatika vagy könyvtártudomány kereteiben él, ezek számára csak számos probléma tisztázásához nyújt heurisztikát. A szakirodalom ilyen széles körű hivatkozása az amugy is terjedelmes jegyzéket emészthetetlen mennyiségűre növelte volna.

A téma tipikusan olyan, amit interdiszciplinárisnak szokás nevezni. Valóban, az osztályozás és információkereső nyelvek elméleti problémái a filozófia, logika, általános nyelvészet, informatika, könyvtártudomány, eszmetörténet, tudományelmélet, taxonómia, rendszerelmélet, matematika, számítógéptudomány kereszteződésében merülnek fel. A lélektant is említeni lehetne - különösen Farradane kapcsán és őt követően - de a pszichológiai nézőpont a disszertációnak más fordulatot adott volna. Elnézést kell kérnem mindazoktól, akik észreveszik, hogy a felsorolt tudományok mindegyikében nem sikerült azonos jártasságra szert tennem. Annyi bizonyos, hogy az osztályozás az emberi gondolkodás alapformája, egyidős magával a gondolkodással, és véle karöltve fejlődik. Az ember számára az osztályozás kényszer is a világ rendezett megismerésére: úgy kényszerül a klasszifikációra, mint ahogyan zenét szerez, mesét mond, megszámol, behatol a jelenségek mélyére. A könyvtárak és információ rendszerek osztályozási kérdései csupán morzsányiak a probléma egészében. Ebben a keretben magam érzem legjobban az értékezés korlátait. Azt is, hogy ál-

IV.

litásai sokban vitathatók.

Végül nem kötelező gesztusként, hanem igaz őszinteséggel akarok köszönetet mondani Rózsa Györgynek. Ő győzött meg baráti biztatással az értekezés megírására. Hálás vagyok érte.

Budapest, 1987. szeptember 1.

Dr. Horváth Tibor

OSZK
Országos Széchényi Könyvtár

A HAGYOMÁNY

OSZK

Országos Széchényi Könyvtár

1. A P R O B L É M A

Két feladatban szokás megjelölni a könyvtári /beleértve a dokumentációs, információs vagy informatikai/ osztályozás célját. Az első az egészként vagy egységként tekintett tudománynak, tágabban az emberi ismereteknek részeire vagy specieseire való bontása, felosztása. A második az individuumok, tények, eszmék, elméletek besorolása egy átfogóbb fogalmi osztályba, amelybe valamelyik jellemző tartalmi ismervük alapján tartoznak. [1] Szigoruan az utóbbit lehet osztályozásnak nevezni, amennyiben az osztályozás nem más, mint osztályba sorolás. Az első feladat szabatosan a felosztás logikai terminusával fejezhető ki. [2] Ennek alapján az információnak /vagy a szakirodalomnak, mint rögzített információnak/ osztályozási elmélete a taxonómia általános elméletéből vezethető le. Az utóbbinak egy, a könyvtári-informatikai területre szóló alkalmazásaként határozható meg.

Ez az általánosan elterjedt álláspont nem elegendő a teljes probléma felvetéséhez, mert éppen az utolsó két-három évtized eredményei, elméleti megfontolásai és megállapításai esnek ki abból a keretből, amelyet a fenti megközelítés enged meg. Létezik ugyanis egy harmadik feladat, amely problémaként mindig is fennállt, de az adatbázisok kapcsán különös élességgel bukkant felszínre. Ez pedig az, hogy nemcsak a tudás felosztásáról van szó, nemcsak objektumok, in-

dividuumok osztályozásáról, hanem minden olyan megállapítás, ismérv, tulajdonság meghatározásáról és leírásáról, amely valamely objektumról; individuumról, elméletről, eszméről állítható. Ezek az ismérvek, tulajdonságok nevezhetők predikátumoknak.

A harmadikként aposztrofált kérdés nem tárgyalható a klasszikus logikából táplálkozó felosztás-osztályozás elmélet kereteiben. Ha pl. egy égitestről, mondjuk a Merkúr bolygóról van szó, akkor ezt a tradicionális osztályozás beillesztette a "naprendszer", vagy a "bolygók", vagy az "égitestek" osztályában, amely egy előzetes felosztás eredményeként állt rendelkezésre. Ennek az eljárásnak alapján maga a "Merkúr" egyedi név által jelölt osztály lehet az a legszűkebb, leg részletezőbb, legpontosabb osztály, amelybe besorolható. Holott a Merkúrról nagyon sok ismeretünk van: mérete, alakja, anyaga, keletkezése, mozgásának és pályájának adatai, kémiai, fizikái, geológiai, dinamikai erők, Merkúr az emberi kultúrában /Hermész, Hermész-mitológia és kultusz, hermetikus vallások, asztrológiai vonatkozások, Mercurius, Thot/, a Merkúr kutatás története, mai problémái, stb. Merkúr fogalmunk gazdagságát azoknak az állításoknak sokasága adja, amely róla elmondható.

A klasszikus logikán nyugvó osztályozási elméletektől elszakadni igyekvő osztályozási koncepciók megjelenését - és

nyomukban a gyakorlatét - az ötvenes évekre tehetjük. [3] Kibontakozásuk azonban a hatvanas évek eredménye. Az individualizáló "osztályozás" vagy indexelés megjelenése - amely a predikátumok kezelését biztosította - nem tette feleslegessé a klasszikus, felosztás elvű osztályozást sem. Az első két feladat is megmaradt. [4] Nem arról van szó tehát, hogy az indexelő eljárások vagy az információkereső nyelvek felváltották a klasszikus, hagyományos rendszereket, hanem arról, hogy a módosult, vagy új indexelési, osztályozási feladatok új elméletet és gyakorlatot követeltek és hívtak életre, megteremtve az eljárások koegzisztenciáját.

A mai teljes feladatrendszer szolgálatában álló törekvések és elméletek a következő utakat járták, vagy inkább járják.

1. Megmaradtak a hagyományos tudományfelosztáson nyugvó rendszerek, de újak konstruálódtak a régi elveken, keresvén a kor tudományosságát jobban érvényesítő felosztásokat. /Kidolgozásra került a BSO, Dahlberg ontikus rendszere, a legmagasabb szintű rubrikátor./ Párhuzamosan a gyakorlatba korábban bevezetett hagyományos rendszereket minden eddiginél radikálisabb reformmal igyekeztek korszerűsíteni. /ETO./
2. A klasszikus logika alapján álló, de a felosztás-osztályozás merevségét oldani kívánó eljárások szü-

lettek azzal, hogy a kategória elmélet hagyományait elevenítették fel. Eredményeként üdvözölhetjük a facettás vagy fazettás osztályozásokat, vagy a facetták beépítését különböző más osztályozások eszköztárába. Ebben az esetben is még mindig osztályozásról van szó a kifejezés szigorú értelmében.

3. A statisztikai elmélet és indexelő eljárások abban kerestek megoldást, hogy az objektumok és a term-ek halmazán értelmezett leképezéssel létesítettek kapcsolatot az objektumok és a term-ek között. Objektumon bármely fizikai létezőt, élőlényt, dokumentumot, szöveget lehet érteni, a term-ek pedig a predikátumok halmazát jelentik /terms, properties, features/. Ezt az iskolát G. Salton tökéletesítette a klaszterálásnak /clustering/ az információkeresésben megvalósuló bevezetésével.
4. A nyelvészeti iskolák a predikátumokat úgy fogták fel, mint olyan állítmányt, amely valamely objektumról elmondható. Egy indextétel valóban mondat, amelyben a term-ek /tárgyszó, vagy bármilyen osztályozási kifejezés/ szerepét a nyelvtani esetrendszerrel analóg viszonyokat tükröző ún. operátorok tüntetik fel. Ebben az eszmében létrehozott minden rendszert egy szintaxis definiál, amely az osztályozási kifejezések bármely gyűjteményére alkalmazható. /PRECIS./ Ezek a mondatok azonban nemcsak a kifeje-

zések halmazatai, dokumentumokhoz rendelt csoportjai, hanem új minőségek, amennyiben az osztályozási kifejezések szavait olyan mondatokká fogják össze, amelynek jelentése más, mint amit a szavak összessége hordozni képes.

A fentebb tisztán meghatározott és leírt irányzatok mellett számos átmeneti eljárás született. Köztük leglátványosabban az információs tézauruszok elmélete bontakozott ki, és számos tézaurusz gyakorlati alkalmazására is sor került. Csupán félig jogosul ezeket a nyelvészeti elvű eljárások közé sorolni. Többen "szemantikai" osztályozásról beszélnek velük kapcsolatban, nem egészen indokolhatóan. [5] Annyiban igen, hogy hangsúlyukat a lexikára /szókészlet, szótár, osztályozási kifejezések gyűjteménye/ és a tézaurusz relációkra helyezik. Az utóbbiak azonban a deskriptorok inherens kapcsolatait tükrözik és visszacsempészik a klasszikus logikai kapcsolatokat, nem pedig az adott indextételben a deskriptornak szánt funkciót jelölik. Mondhatnók: alkalmazás, felhasználás előtti kapcsolatokat foglalnak magukba, míg a felhasználás szintaxisa rendszerint kidolgozatlan marad. [6]

Ennyi tarkaság láttán 1977-ben H. Borko /UCLA/ tanulmánya felkiáltásként hangzott el: van-e, vagy lehetséges-e egységes elmélet? [7] Borko kérdése aligha vesztette el aktualitását.

2. K L A S S Z I K U S R E N D S Z E R E K

2.1 TUDÁSPIRAMIS

A gyakorlati célú osztályozási rendszerek és a filozófiai tudományfelosztások évezredek szimbiózisát az a feltevés ösztönözte, hogy bármely ismeret elhelyezhető a tudományrendszerek keretében. A könyvtárakban pl. az állomány egységeiről készült nyilvántartások tematikai rendje alakítható ki. Ehhez természetesen olyan könyvtár eszme kívánatos, amely ezeket az intézményeket a tudomány házának tekintette. Az uralkodó könyvtári gondolat egészen a 19. századig ebben az eszmében fogant.

A filozófiai tudományrendszerek a klasszikus logika felosztáselvén nyugodtak. Eszerint felosztani annyit tesz, mint egy egészet valamely elv alapján részeire /pars/ vagy fajtáira /species/ bontani.

Az "egész" már a gondolkodás kezdeteinél a teljes világot, az univerzumot jelentette. A világ legelső felosztásait a mitológiák őrzik. Beszámolnak arról, hogy a kaosz /sötétség/ egy istenség akaratából átalakult különböző világokká. Az univerzum három "világra" oszlik: ahol az istenek laknak, az alvilágra és a "mi világunkra", az emberi lét szintereire. [8] E világokat azonban valakik vagy valamik összekötik. [9] A lét három szinterét aztán további szférákra osztják [10] és benépesítik, megkisérelvén magyarázatot adni

ontológiai kérdésekre a civilizáció kezdeti fokán is. [11]

Örökségünk ebből a kultúrából a világ egészként való szemléletének igényessége.

Ennek a korszaknak akkor lett vége, amikor az emberi tapasztalat és a ráció fokozatosan kiszorította a mítoszokat és létre jött a mai értelemben vett tudomány. Ez Szokratész és az utána következő korszak, a hellenizmus szellemében izmosodó, az absztrahálódó tudomány kora. Arisztotelész részben létrehozza és rendszerezi a logikát, a klaszszikus logikát, benne azokat az elveket, amelyek máig is a rendszerezés alapjául szolgálnak.

A filozófia /azaz a tudomány/ feladata a szubsztancia megismerése. [12] A filozófiának annyi része van, amennyi a szubsztanciának van. [13]

Arisztotelész a tudományok felosztásának egyik máig érvényes alapját határozta meg: a tárgyi felosztás elvét. Amivel a tudomány foglalkozik, annak kell megvizsgálni részait és fajtáit, eszerint kell kialakítani a diszciplinák rendjét. A rendszerezés módszere az azonosságok és különbségek felismerése. [14]

Ettől kezdve a tudományrendszerek története nem más, mint a felosztási alapok megválasztásának története. Valamennyi rendszerezés "lényeges" felosztási szempontokat, különbségeket keresett /differentia essentialis/. Valamennyi azt állítja magáról, hogy "természetes" rendszer, azt értvén ezen, hogy a felosztás tárgyához, a tudományok termé-

szetéhez, ezen át a valósághoz igazodóak.

A módszer valóban objektívnek tűnik. A kérdés csupán az, hogy mi tekinthető "lényeges"-nek, amelyen a felosztás nyugodhat. Ebben pedig mindig a rendszeralkotói szubjektívizmus nyilatkozott meg, és minden eddigi és mai tudományfelosztás megítélésének szempontja is: miért éppen az adott felosztási alap tekinthető lényegesnek? Az arisztotelészi módszernek ez a hiányossága korán ismertté vált. Igazolására álljon itt egy reneszánsz szerző bírálata. "Csak azt kívánom, hogy az összes dolgoknak, amit meg akarnak határozni, [15] minél előbb találják meg az efféle lényegi különbségeit, amelyekről álmodoznak. Akkor csak észrevennék, hogy miféle tudományt is meritenek hallgatóik az ilyen csonka és hiányos meghatározásoktól. És ne varrják nyakunkba szüntelen az ember általuk untig ismételtetett meghatározását, hanem más hasonlókat képezzenek - ha tudnak - a saját lényegi különbségeikkel. Vagy pedig, mert elismerik, hogy számukra megfejthetetlenek és ismeretlenek, várják ki, hogy Arisztotelész - akinek saját bevallásuk szerint a tanításait követik a fentiekben -, az alvilágból visszatérve ezeket megvilágítsa, s azonközben semmit se definiáljanak és a világról egyáltalán ne értekezzenek." [16] /S. Varga Katalin fordítása./

Valóban, a lehetséges felosztási alapok száma végtelen, és minden tudományfelosztás preferál egy felosztási ala-

pot. Objektíve nem dönthető el, hogy melyik a leglényegesebb. Ha egy rendszer preferál egy - tegyük fel, hogy lényeges - felosztási alapot, egyben mellőzi a lehetséges többi, tehát lemond azokról a lehetséges rendszerekről, amelyek a negligált alapokon jöhettek volna létre, megmutatván a tudományoknak más természetű rendjét.

Az a preferáláson nyugvó elv az arisztotelészi logika azon következményei közé tartozik, amelyek évezredekig megoldhatatlanok voltak. Egészen az 1960-as évekig kellett várni, amíg olyan matematikai statisztikai eljárás született, amely lehetővé tette az objektumok természetes csoportjainak - akár hierarchiájának - kialakítását preferált ismérvek nélkül. Ez azonban egy másik fejezet problémája.

Az egyik póluson tehát megjelenik a tárgyi felosztás arisztotelészi elve. A másik pólust a lehetséges felosztási alapok megválasztásához - teljes kifejelettségében - a kibernetika nyújtja, azzal a nevezetes tételével, hogy igen nagy rendszer vagy bonyolult tárgy vizsgálatának eredményei nem függetlenek a megfigyelőtől. [17]

A megfigyelő álláspontja különböző szempontokban jelenik meg felosztási alapként ; feltűnik, mint megismerési fokozat, vizsgálati módszer, uralkodó tudományos paradigma, ideológiai nézőpont.

A kibernetika és a rendszerelmélet csupán egzakt szak-tudományi igazolását adják annak, hogy a megfigyelő néző-pontja is lehet felosztási alap; azzal nyújtják ezt, hogy cáfolják azt a sokáig uralkodó kartéziánus dichotómiát, amely az anyagi világot, mint kívülállót függetlenítette a megismerő tudattól. Ennek ellenére a megfigyelő nézőpontja jóval korábban megjelenik felosztási alapként. Valószínű, hogy Francis Bacon volt az első, aki ezt érvényesítette. Tudományrendszerét három intellektuális képesség, az emlékezés, képzelet és értelem alakítja. [18] Közvetlen elődje talán a spanyol J. Huarte. [19] Megjegyzendő, hogy huszonhét évvel fiatalabb honfitársa, Hobbes ismét visszatér a tudományok tárgyi felosztásának elvére. "Sokféle fajtájuk van, aszerint, hogy milyen tárggyal foglalkoznak, s olyképpen oszt-hatjuk fel őket, ahogyan én a táblázatban tettem." [20] (ld. 11. lap: táblázat)

Kialakult tehát a filozófiai tudományrendszerek két-pólusu világa, amelyben a két felosztási típust a tárgy, illetve a megfigyelői nézőpont adja. Bármelyik típusu felosztásból örökségként a világ egységben való szemlélete maradt meg. Az az igény, hogy a tudományok rendszerével a világ a tudatban egészként legyen reprodukálható. Az összetartó tudomány, a részeket, az egyes diszciplinákat egységbe szervező tudomány a filozófia volt. Bacon mondta: "minden tudomány ága egyazon fa törzsében találkozik". [21] Jellemző például, hogy a 19. század elején munkálkodó Ampère, aki fizikus volt,

A természetes testek járulékaiknak következményei, vagyis *természettfilozófia*

A természetes testek közös járulékaiknak következményei, nevezetesen mennyiség és mozgás

A meghatározatlan mennyiség és mozgás következményei, amelyeket, mivel a filozófia alapelvei és kiindulópontjai, *filozófia priim*-ek nevezünk

[*Filozófia prima*

A meghatározott mennyiség és mozgás következményei

Ábrákkal Számokkal

Matematika

[*Geometria Aritmetika*

A meghatározott mozgás és mennyiség következményei

A testek mozgásának és mennyiségének speciális következményei

A világ nagy részei – miként a Föld és a csillagok – mozgásának és mennyiségének következményei

Kozmográfia

[*Csillagászat Földrajz*

A speciális fajtájú és alakú testek mozgásának következményei

Mechanika Súlytan

[*Gépészet Építészeti Navigáció*

Átmeneti, vagyis hol felbukkanó, hol eltűnő testek tulajdonságainak következményei

Meteorológia

Fizika, vagyis a tulajdonságok következményei

A csillagok tulajdonságainak következményei

A csillagok fényének következményei. Ennek, valamint a Nap mozgásának tudománya
A csillagok befolyásának következményei

*Sziagráfia
Asztrológia*

A csillagok közti teret betöltő testek, mint levegő, éteri anyag tulajdonságainak következményei

Állandó testek tulajdonságainak következményei

Az ásványok, tehát kövek, fémek stb. tulajdonságainak következményei

A Föld érzékeléssel nem rendelkező részeinek következményei

A növények tulajdonságainak következményei

A látás következményei

Optika

A földi testek tulajdonságainak következményei

Az állatok általános tulajdonságainak következményei

A hangok következményei

Zene

Az állatok tulajdonságainak következményei

A többi érzékszerv következményei

Az ember speciális tulajdonságainak következményei

Az emberi érzelmek következményei

Etika

A beszéd következményei

Dicséretben, ócsárlásban stb.
Rábeszélésben
Gondolkodásban
Szerződésben

[*Költészet
Retorika
Logika
Az igazságosság és igazságtalanság tudománya*

A politikai szervezet járulékaiknak következményei – vagyis a *politika* és *államfilozófia*

1. Az állam létezésének következményei a politikai szerv, vagyis az uralkodó jogaira és kötelességeire nézve
2. Ugyanezek következményei az alattvalók kötelességeire és jogaira nézve

akkor fordul a tudományrendszerezés felé, amikor a részlet-témáktól metafizikai, a legáltalánosabb problémákhoz jut el és a lét egészével kel bírokra. /Ampère olyan dichotomikus felosztást dolgozott ki, amely a lényegbe hatolás fokozataira épül, a megfigyeléstől az ok-okozat feltárásáig, így tisztán érvényesíti a megfigyelő nézőpontját. [22]

Bármennyire is sokasodtak a tudományrendszerek, amelyeket immár két nagy családba lehet csoportosítani, mindinkább kiütközött a rendszerezés hiányossága, hogy ti. mindig egyfajta lineáris tudományrendszert lehetett kapni. A szaporodó diszciplinák és a gazdagodó tartalmuk egyre inkább fessegetni kezdték e felosztások kereteit. A tárgyat, a valóságot is sokféleként lehet részeire tagolni. Ugyanigy a megfigyelő nézőpontját is sokféleként lehet megadni. Így mindkét oldalról a tudományrendszerek sokasága építhető fel.

El kellett következnie annak, hogy a tudományrendszereket egyidejűleg több szempontból építik fel. Az ujkantiánus Rickert már kijelentette: a tudományokat tárgyak és módszereik szerint szükséges felosztani. [23] Ugy értette, hogy egyszerre kell figyelembe venni a tárgyat és a módszert, az eredmény pedig kétdimenziós skémában rögzíthető. Rickert nem lépett túl az elődök szempontjain, csupán egyesítette őket. Lehetővé tette árnyaltabb, összetettebb rendszerek felépítését. Ez a "kétdimenziósság" maradt a klasszikus elveken alapuló, felosztás-elvű tudományrendszerek uralkodó formája a mai napig.

A felosztás-elvű tudományrendszerezésnek utolsó nagy erőfeszítéseit - filozófiai rendszerekre gondolva - a szovjet rendszerező iskola, mindenekelőtt Kedrov gigantikus rendszere fémjelzi. [24] Kedrov tárgyi felosztást alkalmazott az anyag mozgásformái alapján. Ezen belül második felosztási elvként az anyag szerveződésének formáit érvényesítette /elemi részek, atomok, molekulák, stb./. Nehézségei valójában a társadalomtudományokban bukkannak elő. Itt ugyanis a mozgásformák bonyolultabbak és kuszábbak, semhogy a természettudományokéhoz hasonló, egyértelmű rendszert lehessen alkotni. Vegyük észre azt is, hogy ha az eddigiekben felvázolt rendszerezési erőfeszítésekben fejlődést is keresünk, akkor ez az iskola visszatérést jelent a legelső, csak a tárgyat felosztási alapkőnek tekintő - ha úgy tetszik, arisztotelészi - elvekhez.

Az egyébként igen gazdag szovjet iskola részletező elemzése Farkas János munkájában található. [25]

A további lépés az volt, hogy a probléma megoldását a felosztási alapok számának növelésével igyekeztek elérni. Hazánkban - egyébként figyelmet érdemlő - kísérlet volt egy négydimenziós felosztás. [26] Bóna Ervin ilyen rendszert részletesen kidolgozott a kémiában. [27] Négy felosztási alapon hat kétdimenziós táblázat születik, öt felosztási alapon már tiz, és így tovább. Kezelésük fizikailag válik lehetetlenné. Elvileg pedig az a hibája, hogy a fejlődésben csak lineáris előrehaladást tételez fel, paradigmaváltást

nem. Kb. úgy, mintha mondjuk a múlt században a közlekedés sebességének növelésére egyre több lovat fogtak volna a szekér elébe. Ez egy ideig valóban növeli a sebességet, de mégiscsak új eszközt kellett feltalálni ahhoz, hogy a gyorsaság fokozódjék.

A gyakorlati célú osztályozáselmélet - a szovjet törekvések zömének kivételével - az ötvenes évektől kezdve egyre erőteljesebben igyekezett elszakadni a filozófiai felosztásoktól. Abban az értelemben, hogy az osztályozási konstrukcióknak nem volt igényük az, hogy igazolásukat valamelyik filozófiai rendszerben keressék. Vannak kivételek is. Felismerődött ugyanis, hogy más és más feladatról van szó. A filozófiai felosztások mindegyike a diszciplinák szintjén abba marad, mert céljuk szerint eddig kell eljutniuk. A gyakorlati osztályozásban pedig innen kell elindulni. A filozófia olyan makrostrukturákat hozott létre, amelyek nem hasznosíthatók a gyakorlati osztályozás kisebb léptékekből felépülő világában. A filozófus számára fontos lehet a tudományok sorrendje, az információ osztályozásában a sorrend abszolút közömbös. A filozófusnak küldetése, hogy valamely eszme alapján rendezze a világot, kitüntetett elvei szerint teremtmeg a részek és az egész összhangját. Thomas Kuhn szavaival: egyparadigmájú rendszerek. Az információ, a szakirodalom ebben nem kényszeríthető, mert sokparadigmájú.

Vegyünk egy példát. Tegyük fel, az egyszerűség kedvéért

csupán négy szintig alkalmazunk felosztást és - bár a lehetséges felosztási alapok száma igen sok - minden hierarchiaszinten csupán négy felosztási alapunk van. Ez azt jelenti, hogy minden elágazás egyben döntési pont is, [28] mert mindenütt a felosztások egyikét lehet választani. Ha mind a négy felosztási alapon mind a négy szinten elvégezzünk az összes felosztást, akkor összesen 256 különböző rendszert kapunk. A valóságban pedig több érdemi felosztási alap lehetséges, és nem elegendő a négy szint sem. A példában adott 256 rendszer mindegyike egy világ, mindegyikéhez tartozik odakivánczó információ.

A hierarchia és a gyakorlati osztályozás kapcsolatára Jonker [29] állapított meg egy összefüggést. Kimutatta, hogy $xy=c$, ahol x a hierarchizálás foka, y az osztályozás hatékonysága, c pedig konstans. Másszóval a két változó egymás ellen hat, egyik a másik rovására növelhető. Ezt erősíti meg az értékelési vizsgálatok is, elsőként a cranfieldi jelentés. [30]

A filozófus dolga, hogy a lehetséges világok közül felmutassa azt, amelyik eszméinek megfelel. Az információért felelős intézmények dolga pedig az, hogy felmutassa a világok bármelyikéhez tartozó tartalmakat.

Aki a tudományokat hierarchiába rendezi, eleve feltételezi, hogy a valóság is ilyen szerkezetű. Erre aligha található igazolás. [31]

Lehetséges, hogy mégis a filozófus feladata nehezebb. Az ő vezérelve az, hogy egységes elméletbe foglalja a világ totalitását, elméletébe építve a tudományokat is. Világképet kíván alkotni. És ez az, ami miatt a filozófia mindig is meg fogja kísérelni a tudományokat oly módon rendszerezni, hogy azok világképébe illeszkedjenek.

Az pedig egészen más kérdés, hogy a filozófiának ez az ambíciója hogyan valósul meg, vagy megvalósulhat-e egyáltalán. Pillanatnyilag nincs egységes elmélet, világkép, még egy tudományon belül sincs. Nincs a világról szintézise a fizikának, az élettudományoknak. A kutatás a részleteknél tart. A filozófus számára a világkép híján annál inkább szükség van a tudományrendszertanra, mert ezek szerint ez az egyetlen eszköz, ami az emberi tudás egészét keretekbe fogja, megteremtve a részeknek és az egésznek összefüggéseit valamely elv alapján. Talán Jaspersnek van igaza. "Nincs világkép, csak a tudományok rendszere van." [32]

2.2 UJ STRUKTURÁK

A második világháborút követően nagy gyorsasággal szaporodtak azok a tudományos eredmények, amelyek nem voltak beilleszthetők a kialakult tudományrendszerek egyikébe sem. Sőt, egész diszciplínák születtek, amelyek szintén nem voltak elhelyezhetők. A jelenséget általában az interdiszciplináris területek megjelenéseként, illetve a tudományok integrálódásaként szokták magyarázni. Okának azt tartották, hogy a tudomány a tárgyaként adódó valóságot egyre részletesebben vizsgálja, a konkrét kutatás egyre kisebb, szűkebb témákat elemez, ezeket azonban teljes komplexitásukban. Pl. szervezhetnek kutatóintézetet, amely évtizedekig egyetlen enzimet kutat, de ezt fizikai, kémiai, biológiai, orvos- és agrártudományi szempontból, segítségül hiva még a matematikát a kísérleti eredmények kiértékeléhez. A specializálódott kutatás a konkrét téma kapcsán nem tudja eldönteni - nincs is egénye rá - hogy a problémát melyik klasszikus tudomány eszközeivel kívánja megoldani, mert mindegyikét felhasználja. Eltűnnek a tudományok hagyományos határai.

A társadalomtudományok hasonló utakon járnak. Egy mítosz kutatásában pl. egyaránt érdekelt az etnológia, vallástörténet, filozófia, lélektan, nyelvészet, antropológia, történettudomány, művészetek, szociológia.

A jelenségnek mélyebb okai vannak. Valóban, a határ-

területi diszciplinák sokasodása jelzi ezt a folyamatot. Velük párhuzamosan azonban olyan új tudományok is kialakulnak, amelyek nemcsak azzal minősíthetők, hogy más tudományokkal átfedésbe kerülnek, hogy a közös tárgyat, jelenséget több tudomány közelíti, és ebből előbb-utóbb önálló arculatú új, a tudományhatárok metszetében álló diszciplína keletkezik. Hanem azzal is, hogy ezek az új "tudományközi" diszciplinák a tudományos modellek létrehozásában a kereszteződő tudományok lehetőségeit meghaladó, eltérő, azokéból közvetlenül nem származtatható új eszközökkel mélyítik el a megismerést.

Az a tény, hogy a tudományok az antik kortól a 20. századig a tudományrendszerek tanulsága szerint tárgyukból határozták meg magukat, felfedi e harmadfélezer éves tudomány természetét. A "tárgy" ugyanis az érzékelhető valóság volt akkor is, ha az érzékelhetőség többszörösére tágult a műszerek segítségével. A tudományoknak ez a felállása lehetővé tett egyfajta absztrakciót, amely egyben a megismerés mélységét határozta meg. A lényeg megismerésében addig lehetett elérni, amennyire a tudományos absztrahálás lehetővé tette.

A második világháború után kialakult tudományok jelentős részének - nem előzmények nélkül - olyan tárgya, vizsgálódási területe van, amely nem az érzékelhető valóság feldarabolásából keletkezett. Tárgyuk olyan entitás, amely már emberi absztrakció eredménye. Ez az entitás nem

köthető az érzékelhető világ egyik részéhez sem, hanem jelen van a valóság teljességében.

Egyik ilyen "tárgy" még a második világháború előtt került definiálásra, mint egy tudomány vizsgálódási területe. Ez a rendszer fogalma volt. A rendszer sokkomponensű entitás, amelyben szükségszerű kölcsönhatások zajlanak le.

A rendszer nem sorolható be a klasszikus tudományok tárgyainak egyikébe sem. Rendszerek a természetben, társadalomban, a gondolkodás szféráiban egyaránt léteznek. Ilyenképpen a rendszerek tudománya, a rendszerelmélet lefedi az összes tudományt, még hozzá "felülről", mert tárgya azokénál elvontabb entitás.

Egy ilyen tudomány létrejöttének természetesen nem lenne semmi értelme, ha nem hozna új lehetőségeket a tradicionális tudományok számára. A rendszertudomány új lehetőségei abban állnak, hogy sok összetevőjű jelenségek egzakt vizsgálatát teszik lehetővé. Ha a rendszerelmélet meg is hozza ezt a gyümölcsöt, akkor valóban határkövé válhat a tudományok fejlődésében. A tudományok egyik követelménye ugyanis az egzaktság. Hagyományos tudományokban ez akkor érhető el, ha kevés komponensű jelenségeket vizsgálnak. Bármely eddig megállapított természeti törvényben egy-két, vagy két-három változó tűnik fel, a törvény egzaktságának ereje rájuk vonatkozik. Einstein híres formulájának, az $E=mc^2$ -nek három komponense van. Ha pedig bonyolultabb, összetett, sok-

komponensű jelenségeket vizsgáltak, akkor le kellett mondani az egzaktságról. Tehát vagy egzaktság, de kevés változóju törvények, vagy sok változó, de egzaktság nélkül. Ezt a választást nevezhetjük a klasszikus tudományok egzaktsági dilemmájának. Ennek értelmében beszélhetünk a társadalomtudományok kevésbé egzakt voltáról. Bertalanffy [33] a rendszerelmélet egyik feladataként éppen ezt fogalmazta meg: sokváltozós problémák egzakt vizsgálatát lehetővé tenni.

Minden diszciplína, tudomány, a tudományok összessége maga is rendszer, amelyet a rendszerelmélet vizsgál. Ennek következtében a maga eszközeivel mutatja meg sok összetevőjük és sok kölcsönhatásaik következtében kialakult struktúrát, anélkül, hogy preferált felosztási alapokon vitázna. A rendszerelmélet vizsgálja a különböző tudományok törvényeinek izomorfiáját, keresvén azt, ami bennük közös. Az azonos törvénystruktúrák elemzésével remél szintézisre jutni. Ezt a nem kis vállalkozást szintén Bertalanffy jelölte ki.

Arra a kérdésre, hogy az új tudománystruktúrák hogyan képzelhetők el, még visszatérünk. Itt csak annyit, hogy nem a diszciplínák, hanem a fogalmak analízise vezet eredményre. Itt csak az volt a fontos, hogy felmutassuk egy új tudománynak olyan természetét, amely a klasszikus tudományok rendszerét átstrukturálja, "felülről", más vetületben behatol az összes tudományba. Ezen keresztül a valóságról alkotott képünk is megváltozik.

A rendszerelmélethez hasonlóan lehet leírni más tudományok átszervező szerepét. A kibernetika szintén rendszereket vizsgál. Vizsgálódása a rendszerek szabályozó, különösen önszabályozó mechanizmusaira vonatkozik. A kommunikáció másik olyan jelenség, ami áthatja a természetet, a társadalmat és a gondolkodást is. Csak a múlt század végén, vagy inkább a századforduló után tudatosodott, hogy az ember olyan kommunikáló lény, akit ebből a szempontból egy új tudománynak kell vizsgálnia. [34] A kommunikációelmélet ugyanazt a hatást fejti ki a már kialakult tudományokra, mint amelyet a rendszerelmélet.

Hasonlóan új fogalom az információ. Az új tudományok egész sora vizsgálja, de központi fogalmaként tartja számon a fent említett három tudomány is. A matematikai információelmélet, informatika, information science - az utóbbi hadd szerepeljen angol nevével, kerülendő a meddő terminológiai vitákat - ugyanazt a tárgyat, az információt vizsgálja más-más oldalról, holott még azt sem tudjuk jóformán, miféle entitás ez, amely - úgy tűnik - meghatározó a természetben, társadalomban és gondolkodásban egyaránt. Kutatásra méltó tárgyként az információt a tudomány későn fedezi fel. Ahogy Fülöp Géza, a kitűnő erdélyi szerző írja, az emberiség úgy van az információval, mint Molière hőse, aki nem tudja, hogy egész életében prózában beszélt. [35]

A szemiotikához sok kétely tapad, talán, mert sok avatatlan művelője akadt. Az általános nyelvészetből, közelebbről a szemantikából önállósult. [36] Tárgya a jel, a jel pedig rokonságban van az információval, mert az ábrázolt információ is jel. A szemiotika mindent jelrendszerekkel modellál az objektív világtól, számos jelsikon át a nyelvig. Így kerül kapcsolatba az ismeretelmélettel, nyelvészettel, logikával, természetesen a rendszerelmélettel és az információ tudományaival. Az emberi magatartás is jelekben valósul meg, a szemiotika így prezentál új szempontokat a néprajz, de még az etológia számára is. A jel is nagyon absztrakt fogalom, a szemiotika is úgy viszonyul a klasszikus tudományokhoz, mint a rendszerelmélet és a többi.

A való világ leképezése a megismerő tudatban a fentiek szerint három síkot képez. A valóság szféráját /tárgy/, a klasszikusan kialakult tudományokét, amelyek ezt a valóságot leképezik, harmadikként a fenti tudományokét, amelyek egy harmadik, absztraktabb síkot formálnak. E síkok három rétegét csupán deklarativé kockáztathatjuk meg leírni, hiszen nem a jelenlegi tanulmány feladata - témája miatt - ezt részletesen kimunkálni. Ez is a "megfigyelő álláspontja".

A fentebb számbavett tudományok jelentős része a matematikából eredt, de mindegyikük összefonódik a matematikával. Ez természetes, hiszen ő a legabsztraktabb tudomány.

Még jellemzőbb, hogy a klasszikus tudományrendszereknek legtöbb baja a matematikával volt. Vagy nem tudták, vagy erőltetetten voltak képesek rendszereikben elhelyezni. A mai matematika rég túlnőtt azon, hogy a mennyiségi viszonyok és térformák tudománya legyen. Leginkább úgy határozható meg, hogy a relációk tudománya. Reláció pedig mindeütt van, ahol tudományos vizsgálódás folyik.

Összefoglalva három pontban határozható meg a fent említett diszciplinák hatása a tudományok rendszerére.

- Tárgyuk másként adódik, nem a tapasztalható valóság valamely része, hanem a tapasztalható valóság valamely már elvont tudati fogalmában /rendszer, jel, információ, stb./ kereshető.
- Nagyfoku absztrakciós szintet képviselnek, ezzel az emberi absztrakció szintjét is magasabbra emelték. Ez a folyamat karöltve halad a tudományok matematizálásával, legalábbis formalizálásával.
- Alkalmasak arra, hogy megállapításaik a teljes valóságra vonatkozzanak, amennyiben mindegyik képessége és lehetősége, hogy valamilyen szempontból átfogó modelleket nyújtson.

A fentebb elmondottakat - nem igazolásként, csak példa gyanánt - anekdotikusan is alá lehet támasztani. A Tudomány /Scientific American/ érdekes interjút közölt a hazánkban járt B.B. Mandelbrottal, akiről az igen érdekes tulajdonságú Mandelbrot halmazokat elnevezték. [37] Az egyik

kérdés úgy szólt, hogyan lehetséges, hogy életének különböző periodusaiban más-más egyetemeken tanított, először nyelvész volt, majd közgazdaságtant tanított, fizikusként a turbolencia jelenségét kutatta, foglalkozott a Nilus és más folyók viselkedésének szabályszerűségeivel /földrajz/, nevét jegyzik a biológiában, eredményei vannak a matematikában /eredetileg matematikus/, jelenleg az IBM-nél kutató. Válaszának lényege az volt, hogy egész életében ugyanazt csinálta. Rátalált egy összefüggésre vagy törvényre, felfedezte a fraktálokat. Nem tehet arról, hogy ezzel a legtávolabbi tudományok problémái magyarázhatók. A válasz annyira frappáns, hogy az itt tárgyalt jelenséget akár Mandelbrot effektusnak lehetne hívni. [38]

Visszatérve arra a problémára, hogy miben áll a kortárs tudomány arculatának az a vonása, amelyet az integrálódás terminusával szokás nevezni, arra az mondható, hogy itt nemcsak integrálódásról van szó. Ha ugyanis integrálna, akkor a szintézis konturjai is megjelennének. Hanem arról is szó van, hogy az új diszciplínák átstrukturálják a hagyományos tudományokat. Ami integrációnak tűnik, az inkább a fentebb harmadik síkba helyezett tudományok hatása, visszatürelése a klasszikus tudományokba. Ha pedig egyszer szintézis lesz, akkor ez nem a hagyományos tudományok határainak elmosódása és az interdiszciplináris kutatások megerősödése útján jöhet létre, hanem Bertalanffy rendszer-

elméleti programja szerint. Jelenleg azonban ez még csak program, eredményei csekélyek. Még az is lehet, hogy Bertalanffy hipotézise nem igazolódik.

A tudományok új szerkezetének szempontjából a humaniorák és természettudományok között emlegetett szakadék is másként értékelhető. Tágabban a két kultúráról folyó vitáról van szó. A szakadék nem létezik a valóságban. Pl. a kibernetikai szabályozás hatásmechanizmusa egyaránt érvényesül az emberi nyelvben, a lelki jelenségek körében, a gazdaságban, hadászatban, az élő szervezetekben, az élettelen természetben is. A szakadék csak abban az emberi tudatban él, amelyik nem képes magasabb absztrakcióval követni az emberi megismerés folyamatait. Személyes kis tudatok korlátait vetítik rá a tudományra. A vitának a zszurnalisztikában van a helye. [39]

Válasz várható arra, hogy mindez - legyen tehát Mandelbrot hatás - miért ismerszik fel ennyire nehezen? A tudat nehézségeire lehet hivatkozni. Az oktatás a klasszikus tudományrendszerek mintái szerint alakítja az iskolai tantárgyakat. A gyermekkorban ért hatások perdöntőek. Az egyetemek fakultásai, tanszékei, akadémiai szervezetek, kutatóintézetek, az egész tudományos intézményrendszer is úgy tagolódik, hogy erősíti a gyermekkori benyomásokat. Végül kialakulnak a tudati sinpályák, amelyekben gondolkodunk. Ezt a megmerevedett gondolkodást nehéz áttörni. Ha tudatosul is egy másfajta tudományszemlélet, megmarad eszmei felismerés-

nek. Mert nehéz és kockázatos is az intézményrendszer átalakítása.

Különösen akkor, ha a klasszikus tudományok hagyományos területein is nélkülözhetetlen kutatás folyik. Sőt, nagyobb volumenű kutatás, mint a harmadik sík diszciplináiban. Mert a tudomány ennyire szerteágazó. [40]

A további kérdés pedig az, hogy mindez hogyan csapódik le a gyakorlatot szolgáló, ujtörekvésű tudományrendszerekben?

2.3 FACETTÁS RENDSZEREK

Az osztályozási eljárások másik hagyományos vonulatának a facettás vagy fazettás rendszerek tekinthetők.

Az Organon első könyve a fogalmi kategóriákkal foglalkozik, elméletüket Arisztotelész a Metafizikában finomította. [41] A kategóriaelmélet a lét strukturáiról szóló tanítás. A kategóriák olyan állítások, amik mondatról függetlenül /kapcsolat nélkül/ mondhatók a szavakról. "Viszont semmi akadály, hogy az egymás alá /illetve fölé/ tartozó nemeknél azonosak legyenek a különbségek. Hiszen a felsőbb nemeket az alájuk tartozókhöz állíthatjuk. Tehát amik az állítmánynak különbségei, azok az alánynak is különbségei." [42] "Kapcsolat nélkül mondott szavak mindegyike vagy szubsztanciát jelöl, vagy mennyiséget, vagy minőséget, vagy viszonyt, vagy helyet, vagy időt, vagy helyzetet, vagy birtoklást /valamivel való ellátottságot/, vagy cselekvést, vagy elszenvedést. ... Önmagában véve, illetve önállóan mondva e szavak egyike sem foglal magában állítást; csak egymáshoz kapcsolásuk révén jön létre állítás." [43] /Rókafalvi Ödön és Szabó Miklós fordítása/ A kategóriák tehát a létezés legáltalánosabb módjait adják, amelyek a létezőkről állíthatók. Céljuk az, hogy a létezőket a legáltalánosabb genusok alá lehessen sorolni. Ebben mutatkozik meg kapcsolatuk a felosztással. A különbség az, hogy a kategóriák nem eredményeznek a felosztáshoz hasonló

hierarchiát, hanem a fogalmak természetét mutatják, hogy mi módon vannak önmagukban. Meghatározzák a létezők legáltalánosabb típusait. Az osztályozáselmélet a fogalmak legáltalánosabb kategóriáit ezért nevezi "face"-nek. Innen ered a "faceted classification" elnevezés is. /Horváth Magda terminológiai javaslata volt ezeket vetületes osztályozásnak nevezni. Nem honosodott meg./

A kategóriaelmélet, tehát a facetta elv kétféleként gyakorol hatást a mai osztályozási gondolkodásra. Léteznek par excellence facettás rendszerek, másfelől előtűnnek egyik építkezési elvként vagy használati szabályzatokként egyébként nem facettás osztályozásokban.

A facetták felfedezése az osztályozás számára Ranganathannak köszönhető. [44] Szellemi elődje nem Arisztotelész, hanem a hindu filozófia világszemlélete. Osztályozási elveit Shera a "hagyomány megtörésé"-nek nevezi. [45] Kategóriái a PMEST: Personality, Matter, Energy, Space, Time kezdőbetűiből.

A Personality - lehet egyediségnek, vagy karakternek fordítani - definiálható a legnehezebben. Azt jelenti, ami elsődlegesen meghatározza a témát. Lehet dolog, tevékenységek vagy dolgok fajtája. Az Energy eltérő dolgok közös faktora, pl. a tanterv a pedagógiában, a nyelvten a, nyelvtudományban. A többi facetta, Anyag, Tér, Idő érthetőbbek.

A kettőspontos osztályozásban - a kettőspont egyik elválasztó jel a facetták közt, innen az elnevezés - már megfigyelhető a facettás osztályozások néhány jellegzetessége.

- Feltételezik, hogy valamennyi osztályzandó téma azonos szerkezetben írható le.
- Sorrendet irnak elő; e sorrendben a facetta funkciót határoz meg, tehát egyfajta szintaktikai feladatot lát el. Az osztályozási kifejezések különböző facettákban jelenhetnek meg. Pl. az arany lehet M, de lehet P is.
- Feltételezik az osztályozási kifejezések valamilyen gyűjteményét /tárgyszavak, deskriptorok jegyzéke/.
- Az osztályozót rákényszerítik a téma elemzésére, mintegy kikérdezik facettáról facettára haladva.
- Pontosítják az osztályozási kifejezések jelentését.

Az ötvenes években a facettás osztályozás burjánzása kezdődött. Számtalan szaktudományi, specializált facettarendszert dolgoztak ki, ám az általános kategóriák letisztult rendszerét Vickery dolgozta ki mindmáig ható elméleti alapvetésében. [46] Vetületeit a legáltalánosabb genusok síkjának nevezi. Rendszere a következő.

Dolgok, szubsztanciák, entitások

Természetesek

Produktumok

Eszközök

Szellemi alkotások

Dolgok részei

Alkotó elemek

Szervek

Dolgok rendszerei

Dolgok attribútumai

Minőségek, tulajdonságok

Szerkezetiek

Numerikus értékek

Folyamatai, viselkedés

Cselekvés elszenvedője, tárgya

Dolgok közötti viszonyok. Együttműködés

Hatások

Ellenhatások

Dolgokra irányuló műveletek

Kísérletiek

Elméletiek

Attribútumok és műveletek tulajdonságai

Attribútumokon és műveleteken végzett műveletek

Hely

Idő

A facetták felsorolásában bizonyos ritmikusság észlelhető, amelyet az egész /pl. dolgok/ és részeik /elemei, tulajdonságai/ ismétlődéséből ered.

A szemantikai tényezők elemzése útján építi fel J.L. Melton saját kategóriáit. Összesen 214 tényezőt vizsgálva végül a következő eredményre jut. [47]

Általános fogalmak

Eszmék

Kapcsolatok

Tulajdonságok

Tevékenységek területei

Hatóerők

Osztályozás

Kapcsolatok

Gazdaságiak

Fizikaiak

Társadalmiak

Állapotok

Cselekvést kiváltó vagy megkövetelő

Fizikai

Folyamatok

Fizikai

Általános

Anyagokkal kapcsolatos /létrehozása, cseréje,
rombolása, stb./

Anyagok

Általában

Specifikusak: szerves, szervetlen

Készítmények

Tárgyak

Általában : természetiek, produktumok

Specifikusak: természetiek, produktumok

A kategóriák általános elméletét lehetett felhasználni arra, hogy az Országos Pedagógiai Információs rendszer egyik segédeszköze létrejöjjön. Tárgyszavaihoz készült egy tematikus elrendezésű rész, amely facettás szerkezetű. Vázát az alábbi adja.

Emberi tényező

Akiket tanítanak

Akik tanítanak

Egyéb személyek /irányítók, érdeklődők, stb./

Személyiségek szellemi, fizikai, lelki tulajdonságai

Oktatási és nevelési folyamatok, módszerek, eszközök, oktatástechnológia

Az oktatás és nevelés tartalma

Az oktatás tartalma /tanterv, tantárgy, ismeretek területei/

Nevelési területek /pl.erkölcsi nevelés/

Intézmények

Iskolatípusok, intézmények, szervezetek. Oktatásügy

irányítási és vezetési kérdések

Pedagógia, mint tudomány. Határterületek

Egyéb /Tér, idő, dokumentumtípusok, tájékoztatás/

A végső sorrend megváltozott /az intézmények és oktatásügy a sorrend élére került, de ennek nincs jelentősége./

Ez a fejezet egy fontos konkluzióhoz kíván eljutni. Ehhez azonban még két facettás rendszert kell szemügyre venni. A konkluzió az, hogy a facettás rendszerek egyfelől a hagyomány rugalmas összegzését eredményezték, másfelől a nyelvészetben gyökeredző elméletek kiindulási alapja is.

A facetták meghatározása valamilyen viszonyt is kifejez, azt, az osztályozási kifejezés miként szerepel. Lehet tulajdonság, cselekvés, eszköz, stb. Valaminek a tulajdonsága, valamire irányuló cselekvés, valamilyen célú eszköz, stb.

1950 és 1955 között Farradane [48] megkísérelt olyan kategóriákat felállítani, amelyek kidolgozásához igen eredeti gondolatot használt fel. Pszichológiai eredményekre építve a fogalmak összekapcsolásában azt a folyamatot kísérte végig, ahogyan azt a gyermek megérti. Feltetelezte, hogy a felnőtt, illetve a tudományos gondolkodás is ugyanezen szakaszokban társít fogalmakat, csak felgyorsított formában. [49]

Az első fázisban csak az együttes előfordulás ténye fedezhető fel, a kapcsolat jellegének, időre vagy térre vonatkozás nélkül. A második szakaszban a kapcsolat idő- és térbeli vonatkozásai átmenetileg, időlegesen /tempo-

rarity/ mutatkoznak, anélkül, hogy az emlékezetben megszilárdulnának. A harmadik, végleges fázisban a kapcsolat állandósul. Ez a három szakasz adja a relációk egyik dimenzióját, amely tehát az asszociáció növekedésével és megszilárdulásával jellemezhető.

A második dimenzióban a kapcsolat érzékelésének /perception/ elmélyülése, világossá válása kerül szintén három szakaszban. Az első lépés itt is csupán az együttes előfordulás érzékelése. A másodikban annak felismerése áll, ami a két fogalomban közös /non distinct/. A harmadikban kialakul a megkülönböztetés képessége /distinct/.

A kapcsolatok megismerési vagy felismerési szakasza-ira épülő kategóriák háromszor három relációtípust eredményeznek, az alábbi táblázat szerint.

	Észlelés	Ideiglenes emlékezet	Állandósult emlékezet
Együttes, egyidejű előfordulás	Együttes előfordulás egyébként kapcsolatban nem álló fogalmak között. Jele: /θ	"Maga" tevékenység, ágens/tevékenység Jele: /*	Asszociáció Jele: /;
Nem különböző /Közös/	Ekvivalencia Jele: /=	Dimenzionálás /Idő, tér, helyzet/ Jele: /+	Tartozék, tartozik v hová tartozás Jele: /(
Különböző	Különbözőség Jele: /)	Cselekvés, vmire visszahatás Jele: /-	Funkcionális függőség, okság Jele: /:

Farradane minden kapcsolatot tovább bont. Pl. a tartozás/tartozik valamihez /appurtenance/ alá sorolja az egész-rész, genus-species kapcsolatot, illetve a fizikai vagy belső sajátosságok relációját. A "saját" vagy "maga" tevékenység /self activity/ a tárgyatlan igék hozdozta cselekvést, a dativuszt és a "valami által" kapcsolatot foglalja magában.

Farradane elméletének gyakorlati felhasználásáról nincs jelzés, de elképzeléseinek hatása igen fontos. Az összehasonlító értékelési vizsgálatokba bevonták. [50] Birálták azért, hogy az "asszociáció" nincs kellőképpen definiálva, a kifejezés jelentése homályos maradt. [51] De alighanem S.D. Neil észrevétele a legtalálób, aki szerint Farradane rendszere inkább a percepció tárgyaival hozható kapcsolatba, mint a fogalmakkal. [52] Ő mutat rá viszont arra is, hogy Farradane operátorai nyelvektől függetlenek, ezért alkalmasak nemzetközi indexelési eljárásokhoz. Ez pedig igen fontos szempont.

A facettás eljárásoknak nemcsak osztályozási, hanem heurisztikus jelentőségét, szerteágazó hasznosságát mutatja Jean-Claude Gardin munkássága. Nevét az osztályozáselméletben a SYNTOL /SYNTagmatic Organization Language/ információkereső nyelv tette ismertté. E helyen azonban inkább a SYNTOL-t megelőző munkásságára van szükség. Ekkor kifejezetten a facetták metodikai alkalmazását dolgozta ki az Institute Français d'Archeologie /Beirut/ számára,

majd a Centre d'Analyse Documentaire pour d'Archeologie-ban /Párizs/, majd az EURATOM-nál. Módszerének ismertetése egyben példa arra is, hogy mit jelent az informatika eljárásainak alkalmazása a szaktudományokban. [53]

Gardin eredményeinek összefoglalása előtt ide kívánczik egy figyelmet érdemlő megjegyzés. Vitathatatlanul facettás megfontolásokon nyugszik Bartók Béla népzenei tipológiája. Bartók persze nem használta ezt a kifejezést. Mivel Bartók elgondolását egy későbbi fejezet tartalmazza, mert nemcsak a facettákat fedezte fel spontán, hanem azt a sokoldalú osztályozási elvet is, amelyet a klaszteranalízis képes a gyakorlatban kivitelezni.

Gardin feladata bronzkori, a Balkántól az Indusig előkerülő leletanyag feldolgozása volt. Három szakaszban elvégezhető elemzést vezetett be. Először a leletek morfológiai leírására került sor. Pl. a facetták a vázák esetében: részei, dimenziók, részek közötti kapcsolatok, arányosság a részek közt, stb.

A morfológiai elemzést és leírást az ornamentika elemzése és leírása követte. Itt kétszer tíz szimbólumot általánosított. Az első csoportba az egyenes vonalú alakzatok kerültek: pont, vonal, Z alak, stb., a másodikba görbe vonalúak, mint spirál, hurok. A szimbólumok elrendezésének műveleti és geometriai rendszerezésére harminc változatot dolgozott ki, hat osztályba sorolva. Az elrendezés pl. lehet szimmetrikus, sugaras, rotáló, keresztező, átfedő, stb. Ezek

meghatározása után a fentiek szintetizálásával un. másodlagos és harmadlagos ornamentikát származtatott.

Ugyanis nem az individuális diszitő elemet, hanem a kompozíció egészét kell látni. Itt egy érdekes elemzési analógia született.

Ahogy a nyelvben az egyes szavak fokozatosan alakulnak át szintagmákká és mondatokká, ugyanugy alakulnak át az ornamentika elemei a teljes jelentést hordozó kompozícióvá. Ez azonban már nem morfológiai elemzés, hanem szintaktikai.

Harmadik elemzési szakaszát ikonográfiának nevezi. Ha összehasonlításra kerül egy balta vagy egy korsó, a kép általában nem magyarázható a leírásban szereplő elemek egyszerű egymás mellé helyezésével. Itt újra egy régi nyelvi analógia támad fel magyarázatként és problémaként, az alany és a tárgy közötti különbségtétel. /Milyen szerepe van egy ornamentikai elemnek?/ Az ember megtámadja az oroszlánt és az embert megtámadja az oroszlán mondatok mutatják a különbséget. Szükség van tehát arra, hogy az elemek szerepét a képen valami megmagyarázza, ahogy az ember és oroszlán szerepét a példamondatban. Miután a leletanyag leírását is egy nyelv segítségével lehet elvégezni, kézenfekvő volt, hogy erre a célra ragozást vezessenek be. Gardin öt esetet különböztetett meg. Ezek: alanyeset /vésett, hordozható fa érme/; tárgyesetet /az ember levágta a fát/; minősítőt /ágot tartó ülő ember/; eszközt /az oroszlánt ággal megölő hős/;

végül lokativuszt /az ágon ülő asszony/.

A cselekvéseket egyszerűen két facettába sorolta: a tárgyat hátrányosan, károsan érinti-e vagy sem, negatív vagy pozitív, vagy ahogy ő mondja, fekete vagy fehér.

A topográfiai kapcsolatokra néhány diagramot vezetett be az egymás melletti előfordulás menete szerint. /Nincs kapcsolat, láncos, ellentétes, stb. Ezeket betűkkel és nyilakkal fejezte ki./

A leletek megnevezését egy szótár tartalmazta. Ezek is csoportosíthatók. Pl. görög érmék, élő, élettelen, természeti, stb. Az emberi lét jegyeit részletesebben kidolgozta, /pl. ruházat, szakáll, hajviselet, testhelyzet, gesztus, stb./.

Módszerét, másokkal együttműködve átvitte szövegelemzésre is. Alkalmazták a mezopotámiai táblák, a puebló indiánok mitikus meséinek és a Koránnak elemzésére. Az utóbbi facettáit érdemes példa gyanánt megismerni.

A Korán elemzése fogalmi elemzés volt, tehát nem kifejezetten szövegelemzés. A fogalomelemzést egy szótár támogatta. A szignifikáns szavakat hat fő facettába sorolták:

1. Ontológia
2. Kronológia
3. Természeti létező
4. Antropológiai létező /biológiai, lélektani, személyek, társadalom, politika, jog, gazdaság, kultúra, eszközök/

5. Filozófiai-vallási /eszmék, kinyilatkoztatás, ritusok/

6. Etika /viszonyok, viselkedés, erény, bűn, stb./

Gardinnek soha nem voltak előre elhatározott kategóriái. A leletek/szövegek elemzése során fokozatosan alakultak ki.

Ebből a környezetből nőtt ki a SYNTOL, amelyet munkacsoportjával az EURATOM-mal kötött szerződés alapján dolgozott ki. A SYNTOL azonban már nem egy facettás rendszer, hanem jól kidolgozott szintaxis. Nem szabad elfeledni, honnan az indulás.

Az ötvenes és hatvanas években úgy tűnt, hogy a facettás osztályozás megoldást nyújt számos elméleti kérdésre is. Ennek a két évtizednek erőfeszítései ma már lanyhultak, mert új paradigmák, a statisztikai és nyelvészeti alapozású eljárások tűntek fel. Ezek visszaszorították a facetták gondolatát. Nem mindörökre, mert ebben az eljárásban maradandó elvek is vannak. Maga az alapgondolat, a kategóriák elmélete harmadfélezer éves. A facetták örökös társul szegődtek az osztályozáshoz. Fel-feltűnnek mai rendszerekben is. Sikerét és maradandóságát annak köszönheti, hogy összeköti az eltérő elvű rendszereket. A facetták néha alig különböznek a felosztástól. A Korán elemzésének facettái származhatnak felosztásból is. Valójában akkor derülhet ki igazán, hogy felosztással vagy fogalmi kategóriákkal van dolgunk, ha a részletek is kibomlanak. A facettás rendszerek nem hoznak létre

hierarchiát. Akkor sem, ha egy-két szinten strukturáltak. Arisztotelész gondolata, hogy általános genusok alá rendelje a létezőket, a felosztás gondolatát sugallja.

Másik oldalon a facetták kapcsolatot teremtettek a fogalmi relációkkal. Az ő kibontásuk relációkat - köztük hierarchikus relációkat is - eredményez.

Azok a nehézségek, amelyekkel a klasszikus tudományrendszerek küszködnek, itt nem merülnek fel. Mert a valóságot nem akarják feldarabolni és sémákba erőltetni, hanem a valóság egy-egy arculatát kívánják felmutatni. Ennyi is elég.

Gardin gondolata pedig - szintaxisára gondolva, tehát arra, hogy az osztályozásban, leírásban szerepet kapott elemek egy teljes egésszé álljanak össze, azok, amik tárgyilag összetartozók - más szavakkal a Landry posztulátum megfogalmazása. /Ld. később./ És ez a követelmény már egy nyelvhez vezet el, amelyet információs nyelvnek szokás nevezni.

A facetták joggal tűnnek fel tehát azokban az új próbálkozásokban is, amelyeket a felosztások utolsó mohikánjainak hívhatunk.

2.4 GYAKORLATI CÉLU MAI RENDSZEREK

Fentebb látható volt, hogy az ismeretek egészét átfogó tudományrendszerek közül a gyakorlati céluak lassan leváltak a filozófiai rendszerekről. Az elválás nem a rendszerek létrehozásának elveire vonatkozik, hanem a célokra. Olyan szkémákat szándékoznak létrehozni, amelyek az információ gyakorlati kezelését biztosítják. Az is különbség, ami e célok pontosításával határozható meg. Ezeknek a rendszereknek nem feladata az információ egyedi leírása, hanem átfogó csoportosítása.

Mindez ideig ugyanis az információ /dokumentumok/ osztályozására bevezetett eljárások - jelesen az ETO is - egyaránt kívánták szolgálni mind a csoportosítást, mind a dokumentumtartalom analizáló kivetítését. A kettős feladat hovatovább nem volt tartható egyazon rendszerrel. Az átfogó csoportosítás céljait szolgáló rendszereket sürgetővé tette az információcsere, hiszen minden adatbázis, vagy számítógépes bibliográfia, referáló orgánus géppel olvasható cserére és géppel olvasható szolgáltatásra törekszik.

Ilyen konstrukció felállításának a szokásos nehézségei vannak: egy konkrét téma több átfogó témába, diszciplínába tartozhat egyidejűleg. Az interdiszciplináris területeknek is ki kell jelölni a helyét, stb.

Az osztályozásnak tehát az informatikában elágazó utjai vannak. Az analizáló osztályozás - a bevezetésben har-

madik feladatként szerepelt - életre hívta az információ-kereső nyelveket. A csoportosítási feladatokra változatlanul a hagyományos módszerű szkémák jöttek létre. Az utóbbiak közül három érdemel figyelmet: a legmagasabb szintű rubrikátor, a BSO és Dahlberg ontikus rendszere. Mindhárom valami módon intézményesített kutatások eredménye.

[54]

Közöttük a rubrikátor a legtradicionálisabb. Tudatosan az, klasszikus felosztást alkalmaz, szempontjaiban nincs újítás. Bontása az első két számjegyig alig több felsorolásnál. Viszont nemcsak a tudományokat, hanem objektumokat, termékeket is besorolhatóvá tesz. Elejére a társadalomtudományok kerültek, ennek élén a marxizmus-leninizmus áll. További fejlesztésre nagyon sok üres számot hagy szabadon, néha egész számtartományokat. Az első szinten pl. harmincöt üres hely maradt. Komplex ismeretek nehezen - vagy inkább sehogyszem - helyezhetők el benne. A tudománycsoportokban azonban visszatérő rubrika az "egyéb" tudományok, amely a komplex problémákat is magába foglalja. Ez a visszatérő "egyéb" úgy is értelmezhető, hogy az élő tudomány a satöbbiben kap helyet. Jogos indulatot váltott ki egyik bírálójából, Ungváry Rudolfból, aki a rubrikátor inkonzekvenciáit, monohierarchikus voltát szubjektív stílusban, de elvszerűen kritizálta. [55]

A rubrikátor további bontása sem szerencsésebb. Példaként a 14. rubrikát /pedagógia/ mutatjuk be.

A 14 01 - 14 09 intervallum a pedagógia általános kérdéseinek és a marxizmus-leninizmus pedagógiai vonatkozásainak, továbbá a neveléstörténetnek tartománya. Ezt követi a közoktatási rendszer /14 15/ csoportja. Majd iskolafokozatonként halad: iskoláskor előtti nevelés, általánosan képző iskolák, szakmunkásképzés, gyógypedagógia, középfoku /szak/ oktatás, felsőoktatás, felnőttoktatás, végül a továbbképzés következik. Majd más vetületű csoportok: családi nevelés, a pedagógia részterületei, oktatástechnika, egyes országok oktatásügye.

A pedagógia részterületei címet viselő csoport /14 43/ tűnhet olyannak, amely kibontja a pedagógiát. Ám ide is hagyományos problémák kerülnek: munkapedagógia, összehasonlító pedagógia, stb.

Bár a rubrikátor gyakorlati alkalmazására még nem került sor, megkíséreltünk néhány intenzíven kutatott témát elhelyezni. Az élő pedagógiára nagy hatást gyakorló J. Piaget egész életműve nem helyezhető el. Leginkább az iskoláskor előtti neveléshez illeszthető lenne, de ez roppant megkérdőjelezhető. Piaget ugyan ezzel a korosztállyal foglalkozott leginkább, de valódi témája az ember intellektuális, érzelmi és morális képességeinek, fogalmainak kialakulása és ezek természetének vizsgálata volt. Pl. az osztályozás képessége is, ami a jelen disszertáció témájához kapcsolódik. Mondhatnók: a kisgyermek kapcsán kutatta mindezt. A gyermekhez azért fordult, mert az eredet, a kialakulás stádium-

mában tisztábban és hatékonyabban vizsgálhatók.

De nemcsak Piaget nem találja meg helyét. Találomra néhány téma: tantervelmélet, kognitív pedagógia /a ma uralkodó pedagógia/, értékelésemélet, mikrotanítás, szociális környezet, a pedagógia és demográfia közös területei, oktatásgazdaságtan, stb.

Bírálatot azonban csak fenntartásokkal lehet mondani. Egyfelől meg kell várni az alkalmazás próbáját. Másrészt ezeknek a rendszereknek kapcsán nem vethető fel az örök kérdés, hogy a felosztás jó-e vagy sem. Csak az vethető fel, hogy a célnak megfelelnek-e. A rubrikátor célja is az információcsere és bizonyos szervezeti funkciók betöltése. Az utóbbi azt jelenti, hogy az adatbázisok közti munkamegosztás a rubrikátorok szerint jön létre, ugyanigy az adatbázisok különböző eszközei, pl. az információs tézauruszok is. Ha pedig ezt az utóbbi feladatot képes betölteni, - miért ne töltené be - akkor az elsőt is /információcsere/ elláthatja.

Ilyen körülmények között a rubrikátornak azt a státuszt kellene vállalnia, ami. El kell határolnia magát attól a szereptől, amelyet ráaggatnak - néha kidolgozóik jóvoltából -, hogy valami módon az egyetemes tudományrendszerek örököse.

A BSO a rubrikátorhoz hasonló feladatokat lát el. Kidolgozó a lehető legnagyobb gonddal vették figyelembe az élő tudomány szerkezetét. Az eredmény egy felosztás, amely

a tudományterületek és az emberi tevékenység produktumainak besorolására való. A kidolgozás módszere azonban nem a felosztás volt, hanem fogalomgyűjtés. A fogalmakat később rendezték laza hierarchiába. Az elnevezés /ordering/ önmagában is szerénységet tükröz. A csoportosítás kialakításának vezérelve a tárgykörök komplexitása volt. A rendszer élére a "metodikai" jellegű tudományok kerültek, amelyeket az összes többi tudomány felhasznál. Pl. logika, statisztika, rendszertan, mérés, informatika, kommunikáció. Ebben a csoportban kapott helyet a filozófia, a tudományelmélet, a matematika és a számítógéptudomány is.

Ezt a blokkot követi a többi tudomány, nagyjából a tárgyukat jellemző bonyolultság szerint. A végén állnak az emberi produktumok /ipari, szellemi, művészeti alkotások/. Az egészet a vallás zárja.

A csoportok kialakításában a facettás osztályozás elveit érvényesítették. (A rubrikátorban is él csoportonként egy felépítési tipusséma: általános kérdések, a tudomány elmélete és módszertana, a tudomány története, a tudomány felosztása. Ez talán ott is kapcsolatba hozható a facettákkal.)

A BSO rugalmasságát azzal is biztosítani kívánták, hogy összetett szakterületek kifejezésére a jelzetek belső és külső összekapcsolását /internal, external combination/ tették lehetővé.

Értékét két szóban lehet összefoglalni: egyszerűség és rugalmasság. Mégsem kapott osztatlan elismerést. Foskett

"Öreg bor új palackban?" címmel méltatta, [56] kérdésében megfogalmazva véleményét is. Dahlberg - akinek része volt a BSO induló fejlesztésében, szempontjait a "BSO-Panel" mindvégig tekintetbe vette - a kétségtelenül leggyengébb részének, a társadalomtudományi blokknak elemzésével mutatta ki a gyengéit. A legnagyobb kritika azonban az, hogy a gyakorlatban nem vezették be. A bírálatok nem a BSO jogosultságát vitatták, hanem továbbfejlesztését javasolták.

Az idézett kritikák mindegyike célt tévesztett. Azt vitatták ugyanis, hogy a rendszer szerkezete mennyire univerzális, illeszkedik-e az élő tudomány képletéhez. A BSO célja pedig sokkal szerényebb: összekötő nyelv /switching language/ kíván lenni az információs hálózatok adatbázisainak eltérő nyelvei között. Ez az összeköttetés pedig sok szálon összefüggő fogalmi - ha rendszer egyetemes, akkor viszonylag generikus fogalmak közötti - strukturák kimunkálásával lehet. Ezt már a 70-es évek legelején nyilvánvalóan igazolták. [57]

Ingetraut Dahlberg módszere annyiban hasonlít a BSO Panel munkatársaiéhoz, hogy ontikus rendszerének felépítését fogalomelemzéssel kezdi. Pontosabban: a tudás elemeit keresi, és elemeiként a fogalmakat, vagy azok reprezentációit találja a fogalmi viszonyokkal. Szerteágazó logikai és ismeretelméleti elemzés után a fogalmakat kategóriákba osztja. Indulásul Arisztotelész kategóriáit tekinti, ezeket átcsoportosítja, majd hosszas levezetés után kilenc fő

területet határoz meg: 1. forma és struktúra, 2. anyag és energia, 3. kozmo- és geo-, 4. bio-, 5. emberi, 6. társadalmi, 7. gazdasági területek, 8. a tudomány és információ, 9. a kultúra szférája. Második dimenzióban a terület növekvő komplexitása szerint halad és szintén kilenc szintet határoz meg. Továbbosztás az un. "szisztematizátor" alapján lehetséges. Ez szintén kilenc: 1. általános és elméleti alapok, 2. objektum vagy objektumok /egységek, komplexumok, organizmusok/, 3. a terület módszertana és technológiája, 4. speciális jelenségek, 5.,6. speciális formák és alakzatok /forms/, 7. kifelé irányuló hatás, 8. alkalmazás, 9. a szakterület környezete /szakemberek, szervezetek, tanítása, dokumentációja, stb./.

Kétdimenziós táblázatában így kilencszer kilenc tudományt kap, amelyeket kétjegyű decimális szám kódol a sor és oszlop helyét jelölve. A "szisztematizátor" szerint minden tudomány tovább bontható. A lélektan pl. így:

- 55 Lélektan
- 551 Általános és teoretikus alapjai
- 552 Észlelés, érzékelés, érzelem, akarat és magasabbrendű mentális folyamatok lélektana
- 553 A lélektan metodikája
- 554 Fejlődéslélektan
- 555 Személyiséglélektan
- 556 Társadalomlélektan
- 557 Parapszichológia

AREAS	1	2	3	4	5	6	7	8	9
1 FORM & STRUCTURE	11 Logic	12 Mathematics	13 Statistics	14 Systemology	15 Organization	16 Measure- ment	17 Control & automation (cybernetics)	18 Standar- dization	19 Testing & monitoring
2 MATTER & ENERGY	21 General physics	22 Matter physics	23 Technical & applied physics	24 Electronics	25 Physical chemistry	26 (Pure) chemistry	27 Applied chemistry	28 Power technology	29 Electrical engineering
3 COSMO & GEO	31 Astro- sciences	32 Space re- search & technology	33 Basic geosciences	34 Atmospheric science & technology	35 Hydrospheric science & technology	36 Geological sciences & mineralogy	37 Mining	38 Metallurgy & materials science	39 Geography
4 BIO	41 Basic biosciences	42 Micro- biology & technology	43 Plant biology & breeding	44 Animal biology & breeding	45 Veterinary sciences	46 Agriculture & horti- culture	47 Forestry & wood science & technology	48 Food science & technology	49 Ecology & environment
5 HUMAN	51 Human biology	52 Theoretical medicine & health	53 Pathology & practical medicine	54 Clinical medicine & therapeutics	55 Psychology	56 Education & pedagogics	57 Occupation, labor, leisure	58 Sports & sports- science	59 Domestic science & engineering
6 SOCIO	61 Sociology	62 State & politics	63 Administra- tion	64 Finances	65 Social wel- fare & relat- ed activities	66 Law	67 Regional & urban planning	68 Military science & technology	69 History
7 ECONOMIC	71 General economy	72 Micro- economics	73 General engineering	74 Mechanical & precision engineering	75 Building	76 Commodity science & technology	77 Vehicle engineering & production	78 Transporta- tion	79 Utilities, services & crafts
8 SCIENCE & INFORMATION	81 Science of science	82 Information sciences	83 Computer science	84 General information	85 Communi- cation science	86 Mass-com- munication	87 Printing & reproduction technology	88 Communi- cation engineering	89 Semiotics
9 CULTURE	91 Language & linguistics	92 Literature & philology	93 Music	94 Arts (fine arts)	95 Theatre	96 Ethnology & archaeology	97 Philosophy	98 General religion	99 Religion of choice, e.g. Christianity

558 Alkalmazott lélektan

559 Didaktika - szervezetek - lélektani információ

Bármennyire is szellemes Dahlberg, bármennyire körültekintő, elmélyült elemzés alapozza meg rendszerét, érezhető, hogy végül a konstrukció lesz urrá az anyagon. Hogyan lehetne másként értékelni azt, hogy mindig mindent kilenc részre lehet tagolni? Elérte tehát ugyanaz a végzet, amely az összes előző egyetemes rendszer haláloka, legalábbis elégtelensége volt: a szerkezet öntörvényűen lenyűgöző, de a valóság nem vesz tudomást róla, mindig szétfeszíti a kereteket.

Dahlberg a gyakorlati használatra is gondol. Rangana-
thanra visszanyulva egy alapformulával /abszolút szintaxis/
kiván leírni:

S	P	C
alany vagy tárgy	predikátum állitmány	komplementum(ok/

A magasabb kidolgozású szintaxissal működő rendszerek-
nél /pl. PRECIS/ látható lesz, hogy ez mennyire kevés. Meg-
nevezhető ugyanis a dolog, tárgy, entitás, amelyről szó van,
leírható cselekvés és folyamat. Az összes többi a gyűjtő fa-
cettába kerül kiegészítésként. Legyen itt a példa. A téma:
számítógépi szoftverek exportjának a nemzeti jövedelmet nö-
velő hatása. Kétféleképpen írható le. Az S ugyanis lehet a
számítógépi export, a P a növekedést serkentő hatás, a töb-
bi a C. De lehet az S a nemzeti jövedelem növekedése, P az
export. Az említett PRECIS-ben a felállítás egyértelmű: a tárgy

a nemzeti jövedelem, a cselekvés a növekedés, az ágens pedig a szoftverek exportja, de mindegyik rész árnyalható a /később tárgyalandó/ un. függő elemek blokkjaival.

A három szakmai ihletésű rendszerről összefoglalva az alábbi állitható.

1. A rubrikátor teljes mértékben hagyományos módon felépített rendszer, módszere a felosztás. Felülről lefelé építkezik. A másik kettő szakit a felosztás tradíciójával és ezt fogalomelemzéssel helyettesíti, mintegy alulról felfelé építkezve. A rubrikátor nem érzékeny az új keletkezésű tudományokra, szerkezetében monohierarchikus.

2. A produktum, maga a rendszer mindhárom esetében a hagyományos rendszerekre hasonlít. A rubrikátorra ez teljes mértékben áll, a BSO biztosítja a legnagyobb rugalmasságot, új tudomány- vagy szakterületek határozhatók meg benne. Az ontikus rendszer konstrukció orientált, önmagában zárt. A fogalomelemzést követően az utóbbi kettő a facettás osztályozást használja - filozófiailag a kategóriaelméletet - a szerkezeti váz kimunkálására. Ennyiben modernebbek.

3. E rendszereket nem szabad szerkezetük gyengeségei vagy erősségei alapján megítélni, mert céljuk nem lehet egyetemes filozófiai tudományrendszer létrehozása - bár egyik sem képes ettől a köteléktől megszabadulni. Sőt, az ontikus rendszer kifejezetten ambicionálja is. Feladatuk nyelvi kommunikációt biztosítani az információs hálózatok változó információs nyelvei között.

Ha ezt az utóbbi funkciót tekintjük, akkor kétséges-sé válik, hogy az összekötő nyelv hierarchikus legyen-e?

Az összeköttetés megvalósulhat a különböző nyelvek generikus fogalmainak síkján úgy, hogy ezek között létesül megfeleltetés. A különböző nyelvek mindegyikében a genusok alatt a saját, specifikus fogalmak halmazata található. Az információ cserében csak az átfogó fogalmak kapnak szerepet. Ez a módszer nagyobb fokú rugalmassággal párosul. Nevezetesen

- a nyelvek ilyen konstrukciójában a felosztás nem diszjunkt, azaz bármely specifikus kifejezés több generikus alá kerülhet; maguk a genusok is lehetnek átfedések;
- bármikor új generikus elem /az alátartozókkal együtt/ beiktatható, törölhető;
- változtatni lehet a generikusság szintjén, tovább általánosíthatók, vagy eltolhatók a specifikusság irányába az igényeknek, vagy új információ rendszerek belépésének megfelelően;
- az összeköttetés egyidejűleg több, nyelvileg különböző rendszerrel is megvalósítható oly módon, hogy a generikus fogalmak nem diszjunkt részhalmazai egyik illetve másik rendszerhez vezetnek. A genusoknál egyszerűen jelölhető, hogy melyik rendszer vagy rendszerek kapcsolatában biztosítanak összeköttetést.

Az Országos Pedagógiai Információs Rendszer /OPIR/ há-

rom nemzetközi rendszerrel létesített - fog létesíteni - kapcsolatot. Az ERIC-kel, a Nemzetközi Nevelésügyi Irodával /BIE, Genf/ és később egy KGST rendszerrel. Az együttműködés nyelvi oldala a fentiek szerint fog alakulni. Egy példával lehet ezt a típusu kapcsolatot jobban megérteni.

A rubrikátorban létezik pl. a családi nevelés szakcsoport, beillesztve és kódolva a megfelelő helyen. Az OPIR tárgyszavai között is megtalálható a családi nevelés generikus tárgyszó. Mögötte áll a rubrikátor kódja: 14 39, mutatva, hogy a kifejezés él a KGST nevelésügyi rendszerében. A rubrikátort azonban nem érdekelheti, hogy a magyar rendszerben a családi nevelés előtt, után, alatt mi szerepel. A csere a fogalmak alapján megy végbe, nem a szerkezetek alapján. A kapcsolat akkor is hathatós, ha a fogalom megszabadul a köré épített konstrukciótól.

A J E L E N

OSZK

Országos Széchényi Könyvtár

3 A Z A L A P K É R D É S E K

Az osztályozáselmélet és információkereső nyelvek két mai iskolájának tárgyalása előtt szükségesnek látszik megfogalmazni az elmélet alapkérdéseit. Ezek jelentős részét a szakirodalom már felvetette, mások exponálására itt kerül sor.

Alapkérdésnek az alábbi problémák tekinthetők.

- Kvantálás
- A heurisztikus információ
- Az osztályozás "követő" volta, precedens nélküli osztályozás
- Specifikusság
- Relációmegőrző transzformáció

A szakirodalomban - főleg a monografikus munkákban bő-
ségesen tárgyalt egyéb kérdések technikai jellegűek. Így pl.
a szinonimia, homonimia problémái is. Ezek elmélete nem az
osztályozás témakörében merül fel. Itt már csak a kezelésük-
re vonatkozó szabályokat, technikát, jelölésüket kell beve-
zetni.

3.1 KVANTÁLÁS

Meghökkenítő ellentmondása a szakirodalom osztályozásá-
nak, hogy elvileg mindig az információ osztályozásáról van
szó, a gyakorlatban pedig egy-egy dokumentum az osztályozás

egysége. Tehát az osztályozandó/indexelendő egység egy közlemény, egy tanulmány, egy monográfia, egy szabadalom, stb. Az osztályozás az információt olyan kvantumokban kapja, amilyenben hordozóik, a dokumentumok spontán, rendszertelenül kötegelik. (Némi iróniával az is mondható, hogy az osztályozás egységeit a könyvkötők és a technikai szerkesztők jelölik ki.) Egy publikáció tematikailag nagyon is változó lehet, tárgyalhat valóban egyetlen kérdést, leggyakrabban azonban - érezhetően - több olyan egység elemezhető benne, amelyek önálló osztályozására szükség lehet. Az élő gyakorlat eleve létrehozza a tároló-kereső rendszerek nem kívánatos egyenetlenségeit.

Ez a probléma nem áll fenn a faktográfiai információknál. A szakirodalom az ő esetükben ugyanis csak az információ forrása, lelőhelye. Ez úgy értendő, hogy az osztályozás tárgya nem a szakirodalom, hanem pl. termék, anyag, szerkezeti képlet, stb. Paramétereit, predikátumait a szakirodalom különböző helyeiről lehet megállapítani és hozzárendelni.

Visszatérve a szakirodalom osztályozására, a kérdés az, hogy a publikációk szövege, pontosabban tartalma felbontható-e elemi részekre, s ha igen, bevezethető-e valamilyen egység? Az osztályozásnak az ismeretek hordozóitól való függetlenítéséről van szó, természetesen csak a kvantálás szempontjából. Ne a dokumentum szabja meg, mi legyen az osztályozás egysége. Az osztályozás oldaláról a dokumentumnak csak az a szerepe, hogy az információ lelőhelyét adja meg. Ez pedig igen

egyszerű jelzeteléssel fejezhető ki.

Felmerülhet a szöveg kisebb egységekre bontása. Ez két okból nem járható. Először azért nem, mert nem biztos, hogy a közlemény több témája szövegrészekben lokalizálható. Másodszor azért nem, mert nincs támpont ahhoz, hogyan végezzük el a részekre osztást.

Az információ egységeivel az információelmélet foglalkozik, kézenfekvő lenne innen kérni segítséget. Ennek egységei azonban egyrészt kvantitatívak, másrészt túl kicsik. Az információelméletnek egyik bővített változata bevezetett másfajta egységeket is. McKay definiálta a logont illetve logontartalmat. [58] A logon az egymástól független paramétereket jelenti, más terminológiával a "szabadságfokkal" írható le. Az időjárásjelentés pl. megadható a hőmérséklet, légnyomás és páratartalom adataival, három független paraméterrel. A független szempontok mindegyikében változó adatok, predikátumok sokasága lehetséges, bennük különböző értékek lehetnek. Pl. a hőmérséklet mondjuk -40°C és $+40^{\circ}\text{C}$ közötti értékeket vehet fel. A logontartalom felfogható a facetták számaként is. Az információelméletnek ez a gondolata levezethető a kategóriák elméletéből is.

A logontartalom - mint a facettáknál látható volt - igen kitűnő eszköz az osztályozásban. Az osztályozási egységeket úgy lehetne ezek segítségével meghatározni, hogy az egységet olyan fogalmak köré szervezzük, amelyekhez az

adott facettarendszerben egy minimális logontartalom hozzárendelhető. Az időjárás fogalma megköveteli, hogy hőmérséklet, légnyomás, páratartalom, csapadék, szél erősség, stb. vetületekből bizonyos számuban a változók értékadatakkal szerepeljenek.

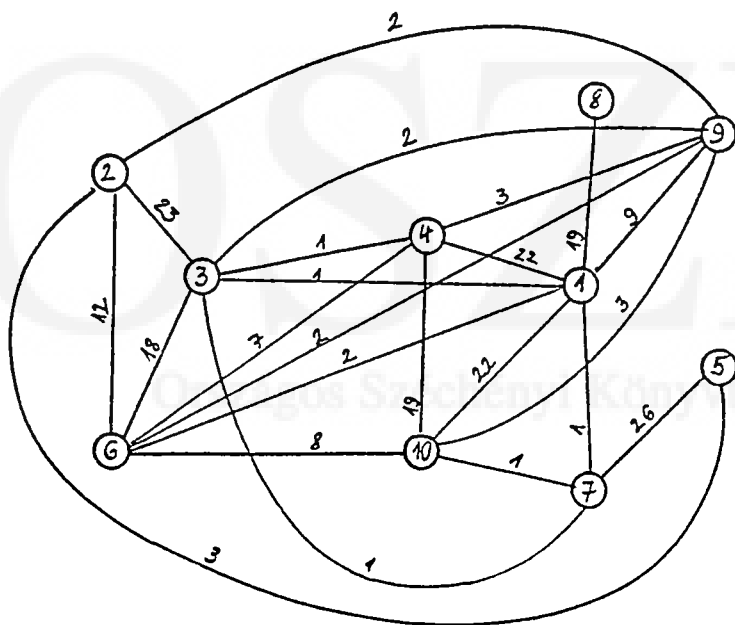
McKay információegysége a gyakorlatban azonban nagyon komplikált, az egységek meghatározása nehezebb lenne, mint maga az indexelés.

Elvi megoldásnak Kunszt György kognitívum fogalma kívánkozik, amely teljesen új, az előbbivel mégis konzisztens egység. A kognitívum "a tudományos megismerés viszonylag egyszerű egységeit: a tudományos téziseket, hipotéziseket, törvényeket, problémákat, a kutatás körülhatárolt, közbülső céljait, témáit, lépéseit, eredményeit, stb." jelenti. [59] Ezt a fogalmat az teszi használhatóvá, hogy a kognitívumok "a tudományos fogalmak jelölésére használt terminusok segítségével írhatók le". A terminusok közt pedig relációk vannak.

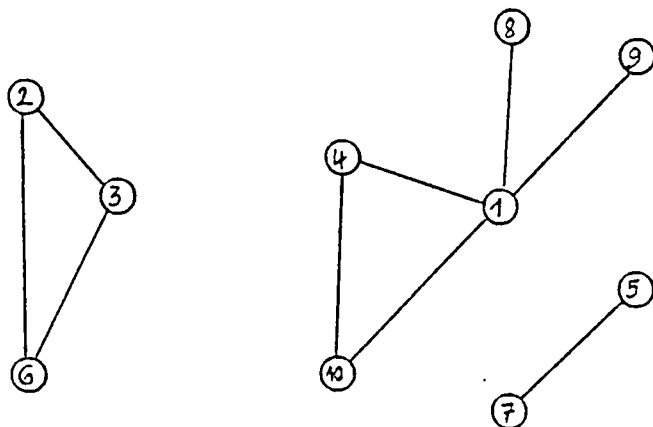
Kunszt ugyan ezeket a fogalmakat a kutatás logikai modellezésére vezette be, de kiválóan alkalmazhatók az itteni problémára is. Maga Kunszt modellként több mátrixot definiál, közöttük kognitívum-term és term-term mátrixot. Fentebb az áll, hogy a terminusok közt relációk vannak. Legyen ez a reláció az együttes előfordulás, a mátrix celláiban pedig az együttes előfordulás eseteinek száma. Illusztrálja ezt egy példa. [60]

*

A mátrixból leolvasható, hogy három kognitívumról, osztályozási egységről van szó. Az elsőt az általános iskola, uttörő mozgalom, zeneoktatás, Zsolnay kísérlet és kommunikáció írja le. A másodikat a gimnázium, fakultáció, nyelvoktatás, a harmadikat pedig a középiskola, munkagépek képviselik. A mátrixhoz tartozó gráf hasonlóképpen ezt illusztrálja. A gráf éleit a gyakoriság súlyozza. A terminusokat az első mátrixban látható kódszámok helyettesítik.



A gráfot a gyenge élek mentén elvágva az alábbi részgráfokat kapjuk.



Két kört találunk a gráfban, ami úgy értelmezhető, hogy a gimnáziumokban a nyelvi fakultációról van szó, az általános iskolákban pedig a Zsolnay-féle kommunikációs kísérletekről.
[61]

Kunszt műve ezt az eljárást nem tartalmazza, de gondolatainak egyik következménye. A mátrixok parciálása a klaszteranalízisben alkalmazott egyik módszer. Átvétele onnan származik.

Az elvi megoldás tehát adott. Gyakorlati bevezetése előzetes szövegstatistikai elemzést feltételez. Igazi értelmét akkor kaphatja meg, ha a teljes szövegtárolás megvalósul.

A vázolt eljárás befolyásolja az indexelés egészét: jelentősen eltolja a specifikus, individualizáló osztályozás irányába.

Valódi individualizáló osztályozást valósít meg.

Országos Széchényi Könyvtár

3.2 A HEURISZTIKUS INFORMÁCIÓ

Pillanatnyilag nem vagyunk képesek az információ heurisztikus leírására és kezelésére. A heurisztikus információ problémája mindenekelőtt a kutatási információ rendszerek kapcsán merül fel. Tehát nem annyira a döntési/vezetői rendszerekben. Összefügg azzal a hipotézissel, hogy a felhasználó az információt tematikusan kéri. Ezt a hipotézist évszázados tapasztalat támasztja alá. Bizonyos, hogy ilyen az igé-

nyek zöme. Felmerülhet azonban az is, hogy a felhasználó azért kér tematikusan, mert tudja, hogy a tájékoztató rendszerek az információt így kezelik, tárolják.

A kutató ezért kutató, mert témájáról minden eddigi eredményt ismer. Ha nem így lenne, akkor nem kutatási, hanem tanulási feladatai lennének.

A valódi kutatói magatartás az információval kapcsolatban a következő. Adva van egy probléma, amelynek megoldásáról felállít egy hipotézist. A problémát a hipotetikus megoldásig a problémamegoldás logikai lépései, menete köti össze, nem is mindig lineárisan. Ha ennek minden láncszeme világos és külön megoldott, akkor a hipotézis is igazolódik. Csakhogy vannak a megoldásban hiányzó, vagy bizonytalan láncszemek, a kutatás éppen ezeknek elosztatására, kidolgozására irányul. A kutató ezekhez keres támogatást, információt: a problémamegoldás logikai menetének hiányaira, igénye heurisztikus természetű. Ezt a heurisztikus információt nem kaphatja meg egy tematikai szerkezetben keresést biztosító rendszertől. Megkaphatná, ha tudná, hogy heurisztikusan hol fordult elő hasonló probléma. Más témák, más tudományok kapcsán.

Egy szakmai példa. Derek de Solla Price törvényszerűséget keresett az információ /dokumentumok/ mennyiségi növekedésében. A növekedés tapasztalati tény volt. A probléma tehát úgy merült fel, hogy ez a növekedés mutat-e törvényszerűséget. Van-e valahol már felismert törvény? Valóban,

ilyenek adódtak a demográfiából /népességnövekedés/, biológiából /sejtszaporodás/, közgazdaságtanból /tőkenövekedés/, stb. Valamennyi statisztikai összefüggést jelentett, így el lehetett jutni a matematikához, Euler általános növekedési formulájához. A probléma ezek után két részletre volt korlátozható: az adatok illeszkednek-e Euler exponenciális görbéjéhez, illetve meg kellett határozni a görbe meredekségét meghatározó együtthatót.

Számos példa hozható arra is, hogy az egyik tudomány heurisztikusan hogyan termékenyíti meg a másikat. Az előző fejezetek néhány esete is idézhető. Farradane-t a lélektan inspirálja. Gardin a diszciplínák egész sorából építkezik, míg esetrendszerét nyelvészeti megfontolások kísérik. A logika, ismeretelmélet pedig az osztályozási rendszerek fundamentumának bizonyultak.

Az osztályozási szakirodalomban a heurisztikus leírás gondolata nem merül fel közvetlenül. Erős a szakmai korlát és szakmai "vakság", a problémát eltakarja az információt matematikailag közelítő rendszerezési paradigma. De van jelzés az igény- és szükségletkutatás oldaláról. Itt is csak mérsékelten. Ezidő szerint Siegfried Balke az egyetlen, aki ezt a gondolatot felvetette ebben az összefüggésben. [62] A használók három típusát, természetét elemzi. A használó először individuum /az egyén mint adatfeldolgozó "rendszer"/, másodszor egy felhasználói csoport tagja /a csoportban megvalósul a személyek közti kommunikáció/, végül problémameg-

oldó szakember, akinek gondolkodásmódját tanulmányozni lehet és kell. Hozzátehető, hogy a kutatócsoport is probléma-megoldó. Balke felveti azt a lehetőséget, hogy a tájékoztató rendszerek felépítésében érvényesíteni lehet heurisztikus szempontokat is.

A tájékoztató rendszerek azonban nem rendelkeznek megfelelő módszerekkel. Helyette pótmegoldások léteznek. Garfield hivatkozási indexe és az erre épülő keresési stratégia előnye éppen abban mutatkozik, hogy feltárja a tudományközi kapcsolatokat. Amennyiben ezek a kapcsolatok hivatkozásokban is megtestesülnek. Pótmegoldás, hiszen az interdiszciplináris kapcsolatok nem feltétlenül heurisztikusak.

A felhasználók nyilatkozataiban azonban fel-felbukkan-
nak olyan momentumok, amelyek erre a kérdésre vonatkoznak. Amikor pl. Wigner beszél "szuper" információ rendszerről, ennek egyik feladataként azt mondja, hogy "a legnagyobb jelentőséget az ellentmondások feltárására helyezi. Nem mereven logikus tehát, hanem heurisztikus. Nem végsőnek látszó, lezárt megismerésre helyezi egyedül a hangsúlyt, hanem ezek feltételeire, érvényességi hatáira, az ellentmondások ki-domborítására, az ellentétek két oldala mellett szóló érvek felsorakoztatására". [63]

Sok minden várható azoktól a kutatásoktól, amelyek a rendszerelmélet fentebb említett Bertalanffy-féle hipotézise alapján a törvények izomorfiajának kimutatására irányulnak.

A jövő kutatásainak egyik területe valószínűleg ez a

téma lesz. Célja egy különleges nyelv kidolgozása lehet, amellyel az információ heurisztikusan leírható. Az információkereső nyelvek mellett, vagy azokba beépítetten az információk leírása ilyen módon is szükséges. A probléma nehézségére azonban álljon itt Wiener egyik megjegyzése. "Ha két tárgynak ugyanaz a technikája és a szellemi tartalma, de teljesen távoleső területekhez tartoznak, akkor még szükség van valakire, akinek az érdeklődése majdnem olyan egyetemes, mint Leibnizé volt." [64]

Akárhogyan is alakul a jövő a heurisztikus leírás kérdésében, egy bizonyos. Azt a rejtélyt soha nem lehet formalizálni, hogy egy gondolkodó miért és hogyan, miért akkor és úgy jutott egy eredeti felfedezésre. Éppen a legnagyobb gondolatok születését fedi homály.

Az információrendszerek beérhetik a formalizálható, metodikai rutinhoz köthető heurisztikus leirással.

3.3 AZ OSZTÁLYOZÁS "KÖVETŐ" VOLTA

A KDSZ referátumait tanulmányozva egy érdekes jelenség keltette fel a figyelmet. A referáló személyek közt akadtak olyanok, akiknek ismertetéseiben soha semmi újdonságot nem lehetett felfedezni, ebből pedig az következett volna, hogy a referált tanulmányokban sincs semmiféle új megállapítás. E referálók ismertetéseiből egy csokorra való összegyűjtése pe-

dig azt mutatta, hogy a kortárs szakirodalom a Sallai-Sebestyén kézikönyv tartalmát ismétli.

Valójában itt egy másfajta folyamat zajlott le. A referáló saját tudását, gyakran szakmai közhelyeit látta bele a tanulmányokba, kivéve belőle az ismert részleteket. Alig van ugyanis olyan tanulmány, amelyik ne ismételne meg valamit az elődök közleményeiből.

A jelenség általánosítható. Sőt, a folyamat erősítésének csaknem intézményes garanciái vannak. Az osztályozás folyamatában rendszerint egy meglévő séma áll rendelkezésre, tükrözve a meglévő tudást, és ebbe kell az analógiák alapján az osztályozandó dokumentumot beilleszteni. Az osztályozandó dokumentumból mindaz, ami benne új, kilóg a sémából. Az osztályozó személy a meglévő szókészlet egységeit illeszti a dokumentum leírásához. Kényszerül rá, hiszen a dokumentumok ujdonságának kifejezésére nincs szótári egysége, nincs osztályozási kifejezése. A tudomány előrehaladása - a tudomány fogalmi megismerés lévén - fogalmi rendszerének örökös változásából, megújításából áll. Néha az egész fogalomrendszer, szerkezetével együtt megváltozik, a tudományos forradalmak, paradimaváltások idején. /Thomas Kuhn kitűnő könyve elevenen írja le ezeket a forradalmakat. [65]/ Nemlétezõ fogalom alá nem lehet besorolni, nemlétezõ kifejezés nem szerepelhet a leírásban. Ismét Wiener idézhető. "Nem lehet besorolni egy könyvet egy ismeretlen címszó alá, csak akkor, ha egy megfelelő személy már felismerte egy címszó

fontosságát arra a speciális könyvre nézve." [66] Az uralkodó osztályozási gyakorlat tehát követő osztályozást produkál. Ezért ellentmondás áll fenn a tudomány folytonos, terminológiájában is megvalósuló változása és az osztályozási rendszerek lehetőségei között. Minél nagyobb az időbeli eltolódás, annál inkább a tudományos újonságok esnek ennek áldozatul. Ha az osztályozó rendszer nagyon régi, elkendőzi az újonságot ahelyett, hogy felszínre hozná: az osztályozás saját feladata ellen fordul.

A kutatói információ rendszereknek a nóvumok megtalálása fontos, nem traktálhat senkit tudományos közhelyekkel. Ami közismert, azt a kutató már tudja. Az újonság tulajdonsága pedig az, hogy precedens nélküli. Akkor osztályozásában is érvényesíteni kell a precedens nélküli osztályozás elvét. Azt kifejezni, besorolni, ami még nem volt.

A tudományos ismeretek és az osztályozás egyidejűségét a kulcsszavas /szövegszavas/ osztályozás biztosíthatja. Bár a kulcsszó elemzés idestova harminc éves /Luhn 1954-ben ebben a témában kezdte kutatásait/, az eszközök még nem eléggé kifinomultak. Az egyidejűség nyereségként jelentkezik, ellen-súlyaként azonban szemantikai és szintaktikai nehézségek lépnek fel. Az utóbbiak kisebb részét az automatikus osztályozás megszüntette, pl. bizonyos relációk statisztikai módszerekkel felismerhetők /ld. később/, de a kérdések egész sora még megoldást kíván.

A mai osztályozási rendszerek olyan arculatát kell ki-

alakítani, amelyben az osztályozási kifejezések a természetes nyelvű szövegek folyamatos elemzéséből származnak, így csökkenthető, akár meg is szüntethető az osztályozás követő jellege. Másfelől ebben a szóanyagban szükséges a szemantikai egyértelműséget biztosítani, kombinálva olyan fajta szintaxissal, mint amellyel fentebb az ikonográfiai paradigmával Gardin biztosította a széthulló elemeknek eszmékké, gondolatokká, üzenetekké, - tehát magasabb egységekké - való összeállítását.

3.4 SPECIFIKUSSÁG

A kulcsszavas osztályozás más kérdést is megold. A visszakereső rendszerek egyik követelménye, hogy az osztályozás/indexelés nyelve ne legyen sem általánosabb, sem specifikusabb, mint a használóé. A használó nyelvében meghatározható terminusok specifikusságának forrása pedig a közlemények nyelve és a keresési témák leírásai.

A specifikusságot más is involválja. Ha az osztályozás egységei a kognitívumok szintjén jelennek meg, nem lenne értelme a generalizáló kifejezéseknek.

A specifikusság kérdése a gyakorlatban igen egyszerű. Nem követeli meg ugyanis, hogy teoretikusan határozzunk meg egy specifikussági szintet, hiszen a természetes nyelv elemzése önként kínálja a kívánt szövegszavakat. Hála a Luhn ál-

tal kezdeményezett eljárásoknak, a módszerek is rendelkezésre állnak, amelyekkel a szövegszavak és szintagmák is meghatározhatók. [67]

Már Luhn és Meadow [68] is tudták, hogy a szignifikáns kifejezések egy meghatározott gyakorisági intervallumban találhatók, amelyet diszciplinákra jellemzően - tudományonként ugyanis bizonyos eltérések mutatkoznak - tapasztalatilag lehet meghatározni. A generikus fogalmak gyakorisága magasabb, a csökkenő gyakoriság - a szignifikáns kifejezések intervallumán belül - összefügg a specifikusság növekedésével. A specifikus szavak gyakorisága általában alacsony. Másfelől a generikus fogalmak eloszlása egyenletesebb, a specifikusaké általában nem egyenletes; eloszlásuk rokonságot mutat a ritka fémek /pl. volfram/ eloszlásával. [69] Ennek hatása van az elemzésre kiválasztott szövegminták meghatározásában.

A specifikusság növekedik akkor is, ha a szövegből egységszavak /uniterm/ helyett szerkezeteket, összetett kifejezéseket határozunk meg.

A kulcsszavaknak ezeket a jellegzetességeit és fontos hatásukat az indexelésre Jonker rendszerezte elsőként és implikálta osztályozási elméletébe. [70] Alapfogalma a terminológiai kontinuum, amely az általánostól a specifikusig terjedő tartomány. Sőt, a természetes, hétköznapi nyelvnek is helyet biztosít. A hétköznapi nyelvet és a szaknyelvet egy "közvetítő" nyelv kapcsolja össze, amely kb. az ujságírás, ismeretterjesztés nyelve. Jonker megállapította még, hogy a

specifikusság növelésével csökken a szavak értelmes kombinációinak száma. Ezt erősíti meg másként Salton. /Az összekapcsolásokra vonatkozóan Jonker másik alapfogalma a kapcsolási kontinuum. A terminológiai és kapcsolási kontinuumra építi osztályozási elméletét./

Salton [71] elemzései még egy érdekességet mutatnak. A generikus kifejezések általában zajosabbak, a specifikusak kevésbé. Itt szemantikai zajra kell gondolni. Ezt erősíti tapasztalatunk is. Pl. az "információ" nagyon zajos, a definíciók özöne ellenére. A "kód" kevésbé zajos, míg, mondjuk a "Shannon-Fano kód" már teljesen egyértelmű. Jonker pedig ezzel erősíthető meg e példával, hogy az "információ" kapcsolási hajlama nyilván magasabb, mint a "kód"-é. A Shannon-Fano kód pedig már csak meghatározott kontextusban található, kapcsolódási lehetőségei alacsonyak.

A specifikus kifejezések meghatározása önmagában még nem biztosítja a hatékony osztályozást. Velük kapcsolatos követelmény az is, hogy az indexelésre választott diszciplína "lefedése" /coverage/ teljes legyen. A specifikusság növelésével ugyanis fennáll annak a veszélye, hogy fehér foltok keletkeznek. Holott nemcsak a lefedés teljessége szükséges, hanem többretegű, nem diszjunkt lefedése. Következik ez a fogalmi kapcsolatok sokoldalúságából és a témák komplex voltából.

A specifikussági probléma palettájához még hozzátartozik, hogy a visszakeresés során pontosság növelő hatása van.

Ezt valamennyi értékelési vizsgálat erősíti, itt elég csak Saltonra hivatkozni. [72]

A szövegszavas osztályozás, eddigi feltételeit együttesen szemlélve, a szótárnagyság duzzadását eredményezi. Az egyetlen igazi veszély ebben áll. Bizonyos méret felett a szóanyag kezelhetetlenné válik és zajosabbá torzul. Az információ rendszerek arra kényszerülnek, hogy lavirozzanak a két ellentmondó követelmény kielégítésében. Osztályozási kifejezéseik legyenek egyfelől elég specifikusak, másfelől a szótárméret kezelhető keretekben maradjon.

A specifikusság problémáját a fenti gondolatmenet homogén témaként mutatta be. Pedig a kulcsszavaknak meghatározásuk utáni sorsa szerint kétféle felhasználása válik szét. Az egyik felhasználásuk arra szolgál, hogy egy szótár lexikai egységeinek birtokába kerülhessen, amely szótár indexelésre való közvetlenül, vagy akár egy információs tézaurusz induló szóanyaga legyen. A szótár karbantartását a rendszeres elemzés biztosítja abból a célból, hogy a "követő" osztályozás zátonyait kikerülje és egyidejű legyen az élő tudomány nyelvével.

A másik felhasználási lehetőség az, hogy minden osztályozási tételt saját szövegének vagy kivonatának elemzése indexeli oly módon, hogy a tételhez rendeli a szövegből meghatározott kulcsszavakat és szintagmákat. Ez az alkalmazás a jövőbe mutat, jelenleg inkább még laboratóriumi gyakorlat. A kutatások azonban igen fontosak, hiszen a jövő

tárolási módja az előrejelzések szerint a teljes szöveg tárolás. Ehhez megvannak a technikai eszközök. A teljes szöveg tárolás is igényel indexelést, mert nélküle a visszakeresés folyamatában mindig a szövegek teljességét kellene átvizsgálni. Egy könyvtárnyi szöveg gyakori és rendszeres vizsgálata megengedhetetlenül költséges lenne.

Az információkereső nyelveknek eddig két tartópillérre derült fény. Az egyiknek az osztályozandó egységek, kvantumok megválasztása, a másiknak a specifikus, az élő szaknyelvvel egybehangzó kulcsszavak bizonyultak. Megvannak tehát az elemek. Bizonyára hiányérzet támad bárkiben, aki látta, hogy ezeket az elemeket értelmes közléssé, összképpé, - Gardint idézve - ikonná, kompozícióvá, mondatná kell összekapcsolni.

A szövegszavas elemzés elemi hiányossága, hogy nem eléggé érzékeny a relációkra. Ez azonban újabb alapprobléma.

3.5 RELÁCIÓMEGŐRZŐ TRANSZFORMÁCIÓ

Gardin következetesen végiggondolt módszerének ismeretében látható volt, hogy az elemzés harmadik fázisában azt kellett megoldania, hogy az individuális elemek hogyan szerveződnek kompozícióvá. Ez már szintaktikai természetű elemzést kívánt meg, s Gardin nem is habozott nyelvtani esetrendszereket átvenni ahhoz, hogy az individuális elemekből monda-

tokat generáljon. Azaz: szavakból mondatok állnak elő. Nyelvet konstruált tehát, és a nyelvnek - formális definíciója értelmében - bizonyos elemeket /szavakat/ és viszonyokat kell tartalmaznia. /A formális nyelvek elméletében a jelentés, mint harmadik komponens nem is definíció értékű./ Ha Gardin problémáját az osztályozandó szövegekre, kognitívumokra, dokumentumokra átvisszük, akkor az osztályozáselmélet legmélyebb, legszigorubb gondolatáig jutunk el. A szövegekben ugyanis a fogalmak bizonyos elrendezésben állnak, rendezettségüket az a viszonyrendszer határozza meg, amelybe a szerző állítja a fogalmakat.

Landry indexelési elméletében az egyik posztulátum erre vonatkozik. [73] A posztulátum azt mondja ki alapelveként, hogy az a transzformáció, amely az eredeti szöveget átviszi reprezentációjába /indextelésbe/, őrizze meg a fogalmak szövegben belüli rendezettségét. Másszóval az a kép, amelyet egy tétel osztályozása nyújt egy szövegről, őrizze meg a fogalmaknak a szövegben egzisztáló viszonyrendszerét. Ez pedig ismét csak szintaktikai probléma.

A gondolat eltérő fogalmazásban máshol is megtalálható. Dagobert Soergel egyik nagyobb lélegzetű tanulmányában a való világnak dokumentációs rendszerekbe való leképezését, mint szigoru matematikus, elvekből vezeti le. [74] Landry posztulátumával rimelő alapelvei: a leképezett rendszer, interpretáció érvényes, ha

/1/ " a való világ minden igaz állítása a formális

rendszer egy teorémájában leképezésre kerül,

és ha

/2/ az olyan állítások, amelyek nem igazak a való világban, nem kerülnek leképezésre a formális rendszer teorémájába." [73]

Tulajdonképpen ennek teljesülése hiányzik a természetes nyelven nyugvó indexelő/osztályozó eljárásokból.

Soergel /2/ elve nem fontoskodás. Itt arról van szó, hogy egy indexelt tételbe belekerülhetnek olyan elemek, amelyek a nyelvi konstrukció következményei és nem az osztályozandó szövegé. /V.ö. 17.sz. jegyzettel/

Az osztályozandó szövegben a fogalmak szervezettsége egyfelől tükrözheti az adott diszciplína általánosan elfogadott vagy ismert tanait. Másfelől individuális, a szerző által létrehozott kapcsolatokat. Minél eredetibb egy közlemény, a szervezettség annál egyénibb, individuálisabb. Az osztályozás pedig a novumokra koncentráló szempontjaival éppen ezt az individuális arculatot hivatott felmutatni. Ugy tűnik tehát, az elv egérfogóba került, hiszen képtelenség annyi fajta kapcsolatot a megőrző transzformáció révén /Soergel nyelven a leképezés következtében/ megtartani, amennyi az újdonságot tartalmazó szövegek, publikációk száma.

A kapcsolatokat tehát tipizálni kell. A tipizálás azonban nem lehet annyira általánosító, hogy a viszonyrendszert a mindenki egyenruhájának mundérjába öltözteti. Ez történt ugyanis a tézaurusz relációk esetén. Sőt, a tézauruszok esetében kontextustól, horribile dictu diszciplináktól füg-

getlenül definiálták őket. /A kérés néhány részlete az 5.3 fejezetben./ De ne is legyen teljesen egyénítő. A kettő között, az általános és az egyedi között kell megtalálni egy harmadikat, mondhatnók, a különösség szintjét.

A teljes egyénítéstől azért is el kell tekintenünk, mert - amint fentebb áll - olyan szervezettség is fennállhat, amely nem a szerző, hanem a diszciplína álláspontjának következménye. Másfelől a megőrző transzformációnak másik feladata is van. Nemcsak a szövegen belüli kapcsolatok átmentése, hanem a különböző szerzők különböző közleményei közötti kapcsolat létrehozása. Ezt nem biztosítja önmagában a kifejezések /tárgyszavak, kulcsszavak/ azonossága, hanem a kifejezések és kapcsolatok együttesen. Folytatva a már megszokott analógiát: "mondatok" között keresünk kapcsolatokat, nem pedig "szavak" között. Kompozíciók között /Gardinre gondolva/ és nem ornamentikai elemek között.

A reláció problémának ez a kulcsa. A megoldást többféle módon keresték - néhányuk, pl. Farradane - már bemutakozott, a többi a későbbiekben látható.

4. A U T O M A T I K U S O S Z T Á L Y O Z Á S

4.1 AZ AUTOMATIKUS OSZTÁLYOZÁS TERMÉSZETE ÉS PROBLÉMÁI

Ha az automatikus osztályozást összehasonlítjuk a hagyományos módszerekkel, szembetűnik természetének két lényeges vonása.

Képes megvalósítani azt a régi óhajt, hogy a csoportok, osztályok kialakításában egyidejűleg érvényesíti az objektumok /dokumentumok/ sok tulajdonságát, valamennyi ismervét. Sőt, kifejezetten akkor virulens, ha sok szempont van jelen. Ha egy dokumentumnak kevés ismérve van - tegyük fel, csak két-három tárgyszóval irták le - az eljárás értelmetlen. Ahogyan egy szimfonikus zenekar is felesleges, ha a zeneművet egy, vagy két-három hangszerre irták.

Ha az automatikus osztályozás sok ismerv jelenlétét kedveli, akkor nyilván olyan indexelési/osztályozási eljárásokhoz kötődhet, amelyek biztosítják a sok szempont meghatározhatóságát, képesek ellátni "nyersanyaggal". A kulcsszavas osztályozás tudja ezt a feltételt teljesíteni akkor, ha az elemzett szöveg nemcsak a címre korlátozódik, hanem hosszabb szövegre, kivonatra, referátumra, akár a teljes szövegre, vagy a teljes szövegből vett mintákra. Alkalmazható akkor is, ha a sok szempont másként biztosítható, akár emberi indexeléssel is.

Másik szembetűnő vonása az, hogy a csoportok kialakítá-

sában nem preferál ismérveket. Előzetesen látható volt, hogy valamennyi hagyományos, felosztás-elvű osztályozás kitüntetett jellemzőkön nyugodott. "A logikában ismertetőjegynek nevezzük azt, aminek alapján valamely osztály tárgyait kiemeljük, és ami lehetővé teszi azt a következtetést, hogy valamely tárgy egy adott osztályhoz tartozik hozzá" - írja J.K. Vojsvillo a fogalomról írt logikai művében. [75] Az eljárás korlátait nagyon hamar felismerték, ahogyan Verancsics Faustus idézetéből kitűnt.

Az automatikus osztályozásnak ez a szelleme, hogy sok ismérvet képes kezelni, nem jelent feltétlenül szembenállást a hagyományos logikai eljárásokkal. Emlékezzünk arra, a tudományfelosztások egyik trendje az volt, hogy az egydimenziós felosztások kétdimenziósakká, majd többdimenziósakká fejlődtek, végül e törekvések a négydimenziós rendszereknél megálltak. Nem elvi okokból, hanem azért, mert a sokdimenziós felosztások kezelése technikailag lehetetlen volt. Ha most elképzeljük, hogy a felosztási szempontokat még tovább növeljük, akkor eljuthatunk egy olyan koordinátarendszerhez, amelynek egyik tengelyén a diszciplinák, a másikon a lehetséges felosztási szempontok - akárhány - ábrázolhatók.

Egy táblázathoz jutunk:

Felosztási alapot Disz- ciplinák	1. felosz- tási alap	2. felosz- tási alap	...	n.felosz- tási alap
1. diszciplinák	1	0		0
2. diszciplinák	0	1		1
⋮				
m. diszciplinák	0	0		0

A táblázatot binárisan lehet kitölteni: az adott felosztási szempont érvényes-e a diszciplinára vagy sem, pl. kísérleti vagy sem, alaptudomány-e vagy sem, stb. Gyakorlatilag közbűös a diszciplinák és felosztási alapok sorrendje.

Bármely oszlop alapján a diszciplinák két csoportra oszthatók, azon belül egy másik oszlop alapján ismét ketté, és így tovább. A felosztás dichotomikus. Különböző oszlopokból indulva és a második, harmadik ... felosztási alapként más oszlop szempontját érvényesítve a lehetséges dichotomikus felosztások variációit kapjuk. Mindegyik ad hoc felosztás, amely átadhatja helyét egy következő rendezettségnek.

[76]

Látható tehát, hogy ez a táblázat, mátrix rokon azzal a reprezentációval, amelyet az információ rendszerek egyik fajta modellálására vezettek be, de azzal is, amelyet az automatikus osztályozás induló mátrixaként kell előállítani.

Ehhez a következőt kell még elvégezni. A diszciplinák helyére kisebb ismeretkvantumokat helyezünk, témát, kognitívumot, egy dokumentumot, tehát egy feldolgozási egységet. A felosztási alapok helyére pedig predikátumokat, termet, tárgyszavakat, kulcsszavakat, stb. A mátrix szintén binárisan tölthető ki:

ismérv Feldolgozási egység	1. ismérv	2. ismérv	...	j. ismérv	...	m. ismérv
1.feldolgozási egység	0	0	1	0	0	1
2. feldolgozási egység	1	0	0	1	1	0
⋮	0	0	1	1	0	0
i.feldolgozási egység	0	0	0	0	0	1
⋮	0	1	0	0	1	0
n.feldolgozási egység	1	0	1	0	0	0

Valójában tehát az automatikus osztályozás alaprepresentációja eszmeileg származtatható a klasszikus felosztások eljárásából, annak mintegy meghosszabbításai. Bár ezeket a gyökereket a klaszteranalízis - az automatikus osztályozás matematikai alapja - nem említi.

Ami pedig azt illeti, hogy az automatikus osztályozás nem preferál ismérveket, korrekten úgy fogalmazható, hogy nem preferál szempontokat, de preferálhat, ha szüksé-

ges. Az automatikus osztályozást "objektív", "természetes" jelzővel szokás illetni azért, mert nem részesít előnybe szempontokat, hanem a dolgokat hagyja összes jellemzőjük, saját természetük szerint csoportosulni.

A klasszikus és automatikus eljárás rokonságát más is mutatja. Kicsit túlzó ugyan az állítás, mégis megkockáztatható, hogy az automatikus osztályozás elveit Bartók Béla fedezte fel, anélkül, hogy a mai szakmai terminológiát /klaszter, hasonlóság, távolság, függvény, stb./ használta volna. Ez a kérdés megér egy kis kitérőt.

Bartók 1923-ban kezdett foglalkozni a népdalok osztályozásával, tipizálásával. [77] Bartókot bizonyára zavarta az a körülmény, hogy a népdalokat egyetlen jegyük alapján osztották be egy fő típusba. Pl. jeles napok, katonadalok, virágének, stb. [78] Holott különböző jegyeik alapján más-más típusba lennének besorolhatók. A népdalelemzés egyéb jegyek kivetítését is megkívánja. Bartók összegyűjtötte az elemzés lehetséges facettáit mind a dallamra, mind a szövegre vonatkozóan: szerkezet, architektorika, szöveg-dallam viszony, hangrend, kádencia, ismétlődések szövegben és dallamban, előadásmód, stb. Minden szemponthoz kidolgozta a lehetséges válaszokat és ezeket számokkal, nagy- és kisbetűkkel, néha görög betűkkel jelölte. Izelitőül két oldal illusztrálja Bartók elképzeléseit.

A témára amerikai tartózkodása idején, 1940-ben visszatért, de végleges jegyzéke nem készült el.

Végleges ritmusjegyzék 1940

A.I. a/
A.I. b/ α)
A.I. b/ β)

A dalok szerkezetüket tekintve:

Kötött, strófikus szerkezetűek
Kötetlen /= nem strófikus/ szerkezetűek



A.
B.
C.

Egyéb.

A dalok a bennük előforduló sorok számát tekintve:

Kétsorosak Háromsorosak Négysorosak

C.IV.

C.III.

C.I.II.

Van-e architektónika a dallamversszakban?

Van = Visszatérő dallamszerkezet
Nincsen = Visszatérés nélküli dallamszerkezet

+

B.

A.
C.I.II.

Alkalmazkodik-e a dal ritmusa a szöveg ritmusához?

Igen = Alkalmazkodó ritmus, pontozott giusto
Nem alkalmazkodó, pontozás nélküli giusto + parlando

+

A.II.
C.I.

A.I.
C.II.

A sorok szótagszám szempontjából:

Egyformák = Izometrikusak
Nem egyformák = Heterometrikusak



A.I.

C.II.

Az egyes sorok szótagszáma:

5 6 7 8 9 10 11 12 13 14 15 16

A dalok előadásmódja:

Tempo giusto
Non giusto = Parlando rubato

+

A.I. b/ A.I. a/

A sorok ritmus szempontjából:

Egyformák = Izoritmikusak
Nem egyformák = Heteroritmikusak

+

A.I. b/ α) A.I. b/ β)

Végleges ritmusjegyzék 1940

A.I. $b/\alpha)$

A.I. $b/\beta)$

A sorok szótagszám szempontjából:

Egyformák = Izometrikusak
Nem egyformák = Heterometrikusak

+

A.I.

C.II.

A sorok ritmus szempontjából:

Egyformák = Izoritmikusak

A.I. $b/\alpha)$

egyfajta ritmus: a

négyszer ugyanaz:

4a

a a a a

Nem egyformák = Heteroritmikusak

A.I. $b/\beta)$

kétfajta ritmus: a b

az egyik háromszor + a másik egyszer: $3a+1b, 1a+3b$

a a b a
a a a b
a b b b
a b a a

az egyik kétszer + a másik kétszer: $2a+2b$

a a b b
a b b a
a b a b
a a a a

/"szlovák szűk" ritmus:/"

háromfajta ritmus: a b c

kétszer az első:

$2a+1b+1c$

a a b c
a b a c
a b c a

kétszer a második:

$1a+2b+1c$

a b b c
a b c b

kétszer a harmadik:

$1a+1b+2c$

a b c c

négyfajta ritmus: a b c d

mindegyik egyszer:

$1a+1b+1c+1d$

a b c d

Ugy értesültem, hogy Kodály ennek egyszerűsített változatával kívánt dolgozni. Azért, mert nem volt ismeretes olyan technikai eszköz, amely Bartók igényes rendszerét kezelni tudta volna. Évtizedeket kellett várni, amíg Bartók szándéka technikailag is megvalósíthatóvá vált. Egy szubjektív megjegyzést. Bartók zsenialitása abban is megmutatkozott, hogy gondolkodásában olyan eljárások születtek meg a kutatási témával való viaskodás és a probléma mély átélése folytán, amely messze előre mutató. Gondolatai ezen a téren máig sincsenek kellően kiaknázva.

Visszatérve az eredeti gondolatmenethez, az automatikus osztályozás nemcsak a klasszikus felosztás-elvű rendszerezések folytatása a fenti módon, hanem hasonlóan rokonítható a facettás módszerekkel. A klaszteranalízisnek a szöveges információ osztályozására, általában kezelésére vonatkozó alkalmazása G. Saltonnak köszönhető. Salton beleütközött az indexelés problémájába, ezért munkatársai között nyelvészek is voltak /Ohio State University/. A tulajdonságokat, attribútumokat a végletekig leegyszerűsített indexelési elméletében típusokba sorolva is elképzelte. [79] A fenti mátrixban tehát elképzelhető, hogy az oszlopokat formáló term-ek sorozata facetták szerinti szakaszokra oszlik. Ez nem zavarja a felsorolás folytonosságát. Sokat segít akkor, ha a mátrixot részeire kell bontani, vagy részgráfokat szükséges kivágni. Salton "típus"-ról beszél. Még Ranganathan PMEST-je is elképzelhető mint típusalkotó szem-

pont. A típusokhoz "érték"-et rendel, amely már konkrét kulcsszó, tárgyszó, deskriptor, amely értéket a típus felvesz. Ez a terminológia pedig ismerős a számítástechnikai szervezésből, közelebbről már az adatbázis szervezés terminológiája.

4.2 AZ ELJÁRÁS RÖVID TECHNIKAI LEÍRÁSA [80]

Jelen gondolatmenetben nem szükséges az automatikus osztályozásnak és matematikai eszközének, a klaszteranalízisnek ismertetése, miután a vonatkozó formulákról, a klaszterálás lényegéről a Könyvtári Figyelő-ben található összefoglalás [81], továbbá doktori disszertációm is ezzel foglalkozik. [82] Csupán az idézett munkákon tulmutató néhány lényeges szempont tárgyalása kívánkozik ide, amely beleillik az osztályozáselmélet itt vázolt gondolatmenetébe. A klaszterálás elvi lényege azonban a gondolatmenet töretlenségéhez kívánatos. A klaszterálási formulák /hasonlósági függvények/ a disszertációból valók, elemzés nélkül az alábbi táblázatban láthatók. A táblázat eredetileg Fritsche-től származik. [83]

A klaszteranalízis /clustering, cluster analysis/ kiindulása az a dokumentum-ismérv, vagy dokumentum-term mátrix, amely fentebb látható. A mátrix bináris, bár súlyozott ismérvek alapján nembinárisan is kitölthető. Minden sorvek-

tora egy dokumentum /kognitívum, osztályozási egység/ reprezentációja. Az a tény, hogy a vektor melyik komponense nem zérus, mutatja, hogy a dokumentumhoz melyik predikátumok, állítások, term-ek, stb. tartoznak. Minden oszlop egy ismértv reprezentációja, nemzérus komponensei mutatják, hogy az ismértvhez melyik dokumentumot rendelték.

A klaszterálás az osztályozásméletben vagy indexelésben két probléma megoldását szolgálja.

1. A dokumentumok csoportosítását, ezt tekintjük szűkebben automatikus osztályozásnak.
2. A kifejezések, term-ek, ismérvek, stb. elemzését, beleértve bizonyos kapcsolatok meghatározását is.

A dokumentumok csoportosítására a sorvektorok olyan számítási műveleteivel kerülhet sor, amelynek lényege a sorvektorok összehasonlítása és amelyek megmutatják vektorok /a dokumentumok vagy osztályozási egységek/ hasonlóságának mértékét. Ez a mérték általában $(0,1)$ intervallumba esik, bár bizonyos formulák intervalluma ettől eltérhet. A teljes hasonlóság mértéke 1, a teljes különbözőségé 0. Egy empirikusan meghatározott küszöbérték feletti hasonlóságot mutató dokumentumok, objektumok alkotnak egy klasztert. A klaszter tehát nem azonos a megszokott osztály fogalommal, hiszen nincs osztályképző ismértv.

Az összehasonlítást különböző formulákkal lehet elvégezni. Tanimoto képlete a legegyszerűbb. [84] Azt mutatja, hogy két dokumentumot jellemző összes ismértv közül mennyi a

Hasonlósági függvények	Intervallum
1. $h/\underline{a}, \underline{b}/ = N/\underline{a} \cap \underline{b}/$	0,n
2. $h/\underline{a}, \underline{b}/ = \frac{1}{n} N/\underline{a} \cap \underline{b}/$	0,1
3. $h/\underline{a}, \underline{b}/ = \frac{1}{n} [N/\underline{a} \cap \underline{b}/ + N/\bar{\underline{a}} \cap \bar{\underline{b}}/]$	0,1
4. $h/\underline{a}, \underline{b}/ = \frac{1}{n} [nN/\underline{a} \cap \underline{b}/ + nN/\bar{\underline{a}} \cap \bar{\underline{b}}/]$	0,n
5. $h/\underline{a}, \underline{b}/ = \frac{N/\underline{a} \cap \underline{b}/}{N/\underline{a}/ + N/\underline{b}/ - N/\underline{a} \cap \underline{b}/}$	
bináris esetben	
$\frac{N/\underline{a} \cap \underline{b}/}{N/\underline{a} \cup \underline{b}/}$	0,1
6. $h/\underline{a}, \underline{b}/ = \frac{N/\underline{a} \cap \underline{b}/}{N/\underline{a}/ + N/\underline{b}/}$	0,1
7. ÁTFED $h/\underline{a}, \underline{b}/ = \frac{N/\underline{a} \cap \underline{b}/}{\min N/\underline{a}/, N/\underline{b}/}$	0,1
8. ASZIM $h/\underline{a}, \underline{b}/ = \frac{N/\underline{a} \cap \underline{b}/}{N/\underline{a}/}$	0,1
9. cos $h/\underline{a}, \underline{b}/ = \frac{N/\underline{a} \cap \underline{b}/}{\sqrt{N/\underline{a} \cap \underline{a}/ \cdot N/\underline{b} \cap \underline{b}/}}$	0,1
10. $h/\underline{a}, \underline{b}/ = \frac{nN/\underline{a} \cap \underline{b}/ - N/\underline{a}/ \cdot N/\underline{b}/}{\sqrt{nN/\underline{a} \cap \underline{a}/ - N/\underline{a}/^2 \cdot nN/\underline{b} \cap \underline{b}/ - N/\underline{b}/^2}} - 1,1$	

közös, azaz a két vektor közös komponenseit vetíti rá a két vektorban szereplő összes nemzérus elemre. Másszóval, méri, hogy a két dokumentumra kiosztott összes ismerv hányada közös. Ha az összes közös, a hasonlóság mértéke 1, és feltehetően ugyanarról a dokumentumról /másik példány, új kiadás/ van szó. Ha egy közös ismervük sincs, a hasonlóság 0, a témák teljesen idegenek.

A másik elterjedt formula az ún. koszinusz képlet. Egy koordinátarendszerben ábrázolt vektornak az X tengelyre eső vetülete a vektor és e tengely által bezárt szög koszinusza. Ha két vektort ábrázolunk, a hajlásszögük koszinuszát az X tengelyre eső vetületek különbsége adja, a koszinusz értéke $(0,1)$ közé esik. Ha a szög kicsi, akkor sok a közös elem, a hasonlóság 1-hez áll közel, ellenkező esetben 0-hoz.

A többi formula a fentiekéhez hasonló megfontolásokon nyugszik.

Tegyük fel ezek után, hogy számítási műveletek sorozatával a hasonlóságot a sorvektorok összes lehetséges párosításában kiszámítottuk. /Ezt a temérdek műveletet gyakorlati fogásokkal töredékére lehet szorítani./ Elvileg fennáll, hogy az összehasonlító műveleteket az összes lehetséges párosításban végezzük. (Valójában ez csak az egyik lehetséges eljárás.) A számítások eredményeként rendelkezésre állnak a páronkénti hasonlósági együtthatók.

Bizonyos formulák annyira finomodtak, hogy a számítási eljárásokba a közös nemzérus elemeket is bevonják. Ennek az a

filozófiája, hogy az "egyikre sem jellemző" is valamilyen hasonlóság. Két embernek pl. lehet közös ismérve, hogy mindkettő délamerikai. De az is valamilyen mértékű hasonlóság, hogy egyik sem európai.

Ha a küszöbérték magas, akkor nagyobb a hasonlóság. Az így képzett klaszterek elemeinek hasonlósága, kohéziója magasabb, de kevesebb a klaszterbe eső dokumentumok száma. Ha alacsony, akkor kisebb a kohézió, de több dokumentum vonható a klaszterbe. Így szabályozható, hogy nagyobb vagy kisebb klasztereket kívánunk-e, alacsonyabb vagy magasabb összetartozást, rokonságot követelve meg.

A klaszterálás következő menetében újabb mátrix, egy dokumentum-dokumentum mátrix készül. Celláiba beírjuk a hasonlósági együtthatókat. Álljon itt egy kis illusztráció mindössze öt dokumentummal, amelyek hasonlósági együtthatói rendelkezésre állnak már.

	1.	2.	3.	4.	5.
1.		2/3	1/5	0	2/3
2.			1/6	1/5	2/4
3.				2/5	2/5
4.					0
5.					

A küszöbérték legyen 0,4. Hagyjuk el a mátrixból a küszöbérték alattiakat, helyükre 0-t, a küszöbérték felettié helyére pedig 1-et írva. Hiszen a továbbiakban már csak arra van szükség, hogy melyik dokumentumok hasonlóak. A példában az 1. dokumentum a 2.-kal, és 5.-kel, a 2. dokumentum az 1.-vel és 5.-kel, a 3. pedig a 4. és 5. dokumentummal.

A további lépés az, hogy a hasonlóságokat figyelve ezt a dokumentum-dokumentum mátrixot olyan részeire bontjuk, amelyek meghatározzák a klasztereket. Erre számos eljárás lehetséges. Egyik eljárás látható a 3.1 fejezetben, ahol az egyik módszert a kognitívumok elkülönítéséhez használtuk. Az eljárás a sor- és oszlopvektorok permutálása volt, ennek következtében a hasonló dokumentumok vektorai "szomszédokká" váltak, így a mátrix /ott a gráf is/ szétvágható volt, szabatosan parciálható.

A fenti gondolatmenet tulságosan leegyszerűsítő. De nem is az eljárás mechanizmusát, hanem a klaszterformálás szempontjainak megmutatását szolgálja. A parciálás ugyanis szerteágazó módszereket takar. Pl. megtehetjük felülről lefelé: kezdetben minden dokumentumot egyetlen szuperklaszter tagjának tekintve, majd ezt daraboljuk egyre kisebb és kisebb klaszterekké. Indulhatunk alulról felfelé nagyjából az alábbi algoritmus szerint.

Vegyük elő a dokumentumot. Tekintsük őt egyelemű klaszternek. Vegyük a második dokumentumot. Kiszámítjuk hasonlóságot. Ha hasonlóak, akkor beosztjuk a klaszterbe, így kételemű klaszterünk van. Ha nem hasonló, a második dokumentumot

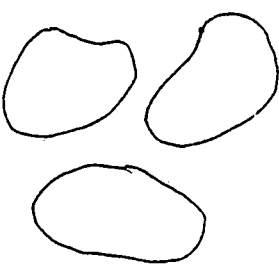
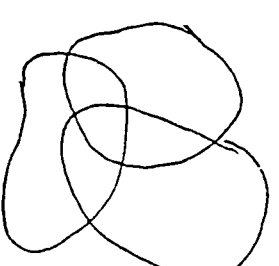
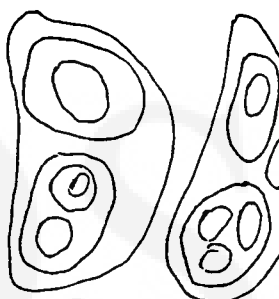
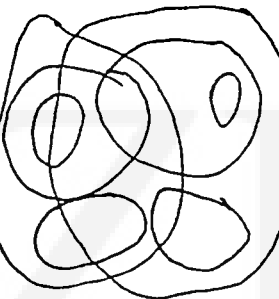
is egyelemű klaszternek tekintjük. Vegyük a harmadik dokumentumot. Összehasonlítjuk az első klaszterrel, ha hasonló, besoroljuk. Ha nem, összehasonlítjuk a második klaszterrel. Ha hasonló, besoroljuk, ha nem, ő lesz a harmadik klaszter. Az eljárás ezzel az algoritmussal folytatható.

A klaszterálást vázoló gondolatmenet abban is sántított, hogy nem mondta meg világosan, hogy egy klaszteren belül mi mihez hasonlít, hiszen az ismertetésben csak a párokra kiszámított együtthatókról volt szó. A klaszteren belüli hasonlóság többféleként értelmezhető. Az eltérő értelmezések már különböző klasztertipusokat eredményeznek. Nevezetesen

- az összes dokumentum egyetlen dokumentumhoz hasonló, ilyenkor "csillag"-ról van szó;
- az első dokumentum hasonlít a másodikhoz, a második a harmadikhoz, a harmadik a negyedikhez és így tovább. Ekkor "lánc" keletkezik;
- minden dokumentum mindegyikhez hasonló /klikk/, ebben igen magas a kohézió;
- a hasonlóság a klaszterre definiált küszöbértéket meghaladja, tehát nem feltétlenül mindegyik hasonlít mindegyikhez, de mindegyik több másokhoz /klamp/.

Ha a dokumentumok csak egy klaszterbe tartoznak, akkor elvágóak /diszjunkt/. Ha több klaszterhez sorolhatók, akkor átfedések. Másfelől a klasztereket is lehet klaszterálni. Kisebb /erős kohézióju/ klaszterek nagyobb egységekben, alacsonyabb hasonlóságot mutató klaszterekben összefoglalhatók.

Ilyenkor rétegzett, ellenkező esetben egyszerű klaszterekről van szó. E típusokat mutatja az alábbi ábra.

	ELVALÓ	ATFEDÉSES
EGYSZERŰ		
RÉTEGZETT		

Látható, hogy a klaszterek hierarchiát is alkothatnak. Meszszire jutottunk a hierarchiák területén is.

4.3 INDULÓ PARAMÉTEREK

Nem egészen technikai probléma, hanem az indexelés elveit érintő kérdés az, ahogyan az induló paramétereket meg lehet választani. Egészen szimplán fogalmazva a probléma úgy szól, hogy a dokumentumok a klaszterálás legelső fázisában, még az indulásnál mihez hasonlítsanak? A következő lehetőségek vannak.

- Nem választunk induló paramétert, hanem az előző fe-

jezetben leírt eljárással a dokumentumok páronkénti összehasonlításával, majd a klasztereket az ennek eredményeként született mátrixon végzett további műveletekkel alakítjuk ki.

- A diszciplína, szakterület ismeretében definiálni lehet kutatási témákat. Ezek leírásának vektorai képezik azt a reprezentációt, amelyhez a dokumentumok hasonlíthatók. A kutatási témákat képviselő vektorok egy klaszter közepének tekinthetők, amely köré a hozzá hasonló dokumentumok összehozhatók. A definiáláshoz fel lehet használni kutatási programokat, terveket. Kutatásnyilvántartások elemzése mutatja, hol vannak a kutatások csomópontjai, milyen aktuális problémák vannak napirenden. Felhasználhatók professzionista kutatási prognózisok, trendek elemzései. Ez a módszer hívható szükséglet orientált paraméteres eljárásnak.
- Induló paraméterek, vektorok állhatnak elő úgy, hogy "tipikus dokumentumok"-at keresünk. Mondván, hogy ezeknek vektorai köré klaszterált dokumentumok tükrözik a tudomány kutatási szerkezetét. Az előbbi eljárás változata. Míg a fentiben a trendek is érvényesülnek, ebben inkább az eredmények testesülnek meg. A fentéhez a tudomány elemzése, az ittenihez a dokumentumok elemzése nyújt támpontokat. Az eljárás hívható eredmény orientált paraméterű módszernek. /Érdekes meg-

jegyezni, hogy a "tipikus dokumentum" módszer a visszakeresésben is alkalmazható. Ha a keresőprofil összeállítása valamilyen okból nem sikerül, gyakran kéri a felhasználót, hogy mondjon tipikus dokumentumot, amelyhez hasonlókat keres. Ennek a megnevezett dokumentumnak vektora lesz a kereső vektor./

- Végül klaszterálhatók a kereső kérdések, felhasználói igények. A kérdéseket ugyanazokkal a kifejezésekkel indexeljük, mint a dokumentumokat. Az igényklaszterek tükrözik a felhasználók tipikus kérdéseit, a klaszterek változásai pedig az igények változásait követik nyomon. Az igényeket reprezentáló vektorok lehetnek a dokumentumok klaszterálásának induló paraméterei, vektorai. Ennek eredményeként a dokumentumokat mintegy felfűzzük, hozzárendeljük a mindenkori igényszerkezethez. Az eljárást ezért igény orientált paraméteres módszernek lehet nevezni. Ennek azonban van két hátránya. Először, nem biztos, hogy az igények hűen tükrözik a szükségleteket. Pl. elmaradott kutatások esetében bizonyosan van különbség a kettő között. Második hátránya az, hogy a kutatások radikális változásakor a teljes állomány újraklaszterálását igényli. Ezt ellessúlyozza az, hogy az igényorientált klaszterek folyamatosan igazodnak a felhasználóhoz. Ezen az elven épül fel Salton dinamikus könyvtári modellje, amelyben a tematikai keresést biztosító intellektuális eszközök változásban vannak, naprakészen igazodnak a fel-

használói igényekhez. [85]

4.4 AZ ISMÉRVEK ELEMZÉSE

A 4.2 fejezet említette, hogy a klaszteranalízis nemcsak dokumentumok csoportosítását, hanem az indexelő kifejezések elemzését is szolgálja. Lehetőség van a kifejezések közötti kapcsolatok némelyikének automatikus meghatározására.

Az elemzés azon nyugszik, hogy a kifejezéseket különböző dokumentumokhoz rendeltük hozzá. A vizsgálatban tehát az eredeti modell /mátrix/ oszlopvektorai vesznek részt.

Elsőként azt kell megfontolni, hogy tapasztalat szerint a generikus kifejezések az indexelés során sokkal aktívabbak. Másszóval több dokumentumhoz tartoznak, mint a specifikusak. Az utóbbiak felhasználási gyakorisága kisebb. Ha tehát összevetjük az oszlopvektorok komponenseit, egyszerűen a komponensek számának meghatározásával, már következtetni lehet arra, hogy generikusabbak-e vagy specifikusak. /E fejezet minden következtetését statisztikai összefüggésként kell érteni./

Vegyük a következő eseteket. Legyen a két vektor \underline{v} és \underline{w} .

$$\underline{v} = (\text{lolololo})$$

$$\underline{w} = (\text{lolololo})$$

A két ismérvvektor minden komponense megegyezik. Az eset szinonimát valószínűsít. Ha a vektorok majdnem minden komponense

egyezik, magas hasonlóság áll fenn közöttük, jelentésük is rokonságban van, akár majdnem szinonimák lehetnek. Legyen két vektorunk a következő.

$$\underline{v} = (1110010)$$

$$\underline{w} = (1010001)$$

A formulák táblázatában a 8. képlet alkalmas a szimmetrikus hasonlóság kiszámítására.

$$h_{v,w} = \frac{N(\underline{v} \cdot \underline{w})}{N \underline{v}} = \frac{2}{4}$$

és

$$h_{w,v} = \frac{N(\underline{v} \cdot \underline{w})}{N \underline{w}} = \frac{2}{3}$$

ahol N a komponensek összegzését jelenti. /Algebrai jelöléssel pl. a

$$h_{v,w} = \frac{\sum_{i=1}^k \underline{v}_i \cdot \underline{w}_i}{\sum_{i=1}^k \underline{v}_i}, \quad k \text{ komponensű vektor}$$

esetén./

Látható, hogy köztük metszetviszony van, továbbá, mivel $h_{v,w} < h_{w,v}$, ezért \underline{v} átfogóbb, generikusabb, tágabb. Az állításból nem feltétlenül következik, hogy közöttük genus-species vagy totum-pars kapcsolat áll fenn, csak valószínű, hogy \underline{v} átfogóbb kifejezés, mint \underline{w} . A kapcsolat mibenlétét nem lehet pontosan meghatározni. (Egy tágabb és egy szűkebb fogalom metszetviszonya.)

Végül meghatározható "beletartozási viszony", amely szintén csak valószínűsíthető. Ez akkor áll fenn, ha \underline{v} vektor minden komponense \underline{w} -ben is megtalálható, de \underline{v} rendelkezik még további komponensekkel. Pl.

$$\underline{v} = (111001)$$

$$\underline{w} = (110000)$$

esetén látható, hogy \underline{w} -t minden olyan esetben használtuk az indexelésben, amikor \underline{v} -t, de \underline{v} -t még további esetekben. Ebből következik, hogy \underline{w} által képviselt téma átfogóbb. \underline{v} olyan témát jelöl, ami valami módon \underline{w} résztémája.

A kifejezésekre vonatkozó eljárások egyike az együttes előfordulást vizsgálja. Az eljárás már bemutatkozott a kognitívumok meghatározásának egyik módjaként. Az együttes előfordulás másért is fontos. Ha felírjuk az együttes előfordulás mátrixát /gráfját/, a gyenge élek mentén eszközölt vágásokkal a kifejezések olyan csoportjait kapjuk meg, amelyek a szakirodalomban együtt szoktak előfordulni, köztük tárgyi kapcsolat van. Így alakítható ki a kifejezések szótárának szakrendje, illetve a témák szerint csoportosított szótár.

Bár az OPIR tárgyszavainak létezik az említett szakcsoportos változata, a tervekben ezzel a módszerrel készült szakcsoportosítás is szerepel. Ez pillanatnyilag azért nem készíthető el, mert a tárgyszavakkal éppen elindult az indexelés, a felhasználás gyakoriságára még nincs elegendő eset.

A klaszteranalízis széleskörű informatikai alkalmazására álljon példaként a tudományometriából ismert együttlidézési klasztertechnika. Bujdosó Ernő szerint célja "komplex halmazban rejlő szerkezet felderítése az egyedek hasonlóságának felismerése révén". Az elemzésre szintén Tanimoto formulája szolgál. Az eljárás az együttlidézett közlemények számát a közlemények külön-külön kapott idézeteinek összegére vetíti. Ha a kettő hányadosa 0, akkor egyetlen klaszter jön létre, amelyben a dokumentumok idézetei egyben együttlidézettek is. Ha a hányados 1, a vizsgált dokumentumok mindegyike egy-egy egyelemű klasztert formál, tehát nincs együttlidézés. Optimális értéket 0 és 1 között nehéz előre meghatározni. Viszont az együttlidézési szám változtatásával a szerkezetbe mélyebb betekintést lehet biztosítani.

Az eljárás hasonlít az un. bibliográfiai társítás /bibliographic coupling/ technikához, amely azon az alapon kívánta a tematikailag hasonló dokumentumokat csoportosítani, hogy azokban az együtthivatkozás küszöbértéke magas.

4.5 TOVÁBBI PROBLÉMÁK

A klaszteranalízis alkalmazása számos osztályozási problémát megold. Nem érzékeny a kifejezések származására. Az ismérvek meghatározhatók szövegből, a dokumentumhoz rendelhetők egy szótárból; egy szótárban szereplő kifejezések felismertethetők a szövegből, eredhet tárgyszavas osztályozási eljárásokból, bármiből. Ha a dokumentumok kívánt reprezentációja előállt, akkor a további műveletek az összetartozó csoportokat automatikusan alakítják ki.

Másik előnye, hogy vonzalmat érez a természetes nyelvhez. Ezáltal olyan kifejezésekkel dolgozik, amelyeket a kutató használ.

Végül sokszempontu, kimerítő, részletező, kellően specifikus osztályozás végezhető segítségével.

Ezzel szemben igen durva megoldásokat tesz lehetővé a kifejezések szemantikai kapcsolataiban. Fogalmi viszonyokat bizonytalanul ismer fel és a kapcsolatoknak csak igen szerény számu típusára következtet. Az előző fejezet ugyan mutatja, hogy az automatikus osztályozás ezen a téren nem tanácstalan teljesen, de messze elmarad a kapcsolatoknak pl. a tézauruszokban szokásos kiépítettségének felismerésében. További kutatásoknak kell az automatikus osztályozás eljárásait a kapcsolatok felismerésében élesíteni.

Hasonló hátránya, hogy nem utal a szintaktikai kapcsolatokra sem. A szintaktikában egyetlen produktumra képes. A szavakat, kifejezéseket egymás mellé helyezi valamilyen sor-

rendi előírás szerint vagy anélkül. Elmarad azoktól a finom megoldásoktól, amelyeket pl. Gardin, Farradane, mások után már természetesen igénylünk. Az utóbbiról nem is vagyunk hajlandók lemondani. A gondosan kidolgozott szintaxisok kényelméről, hatékonyságáról, amelyek már elkápráztattak minket.

A szintaktikai kapcsolatok azonban nem az "idegen test" az automatikus osztályozásban. Nincs kizáró ok, ami megakadályozná az automatikus osztályozásnak és a szintaxisoknak szervezett egységét. Hiszen az automatikus osztályozás szövegszavas osztályozás esetén működhet legjobban. Ha a szövegben szó- és szerkezetelemzés végezhető, akkor szintaktikai is. A szövegben jelen vannak a viszonyjelek, ragok, jelek, képzők.

Az automatikus osztályozás jövője a szakirodalmi tájékoztatásban attól függ, hogy képes-e az automatikus szintaktikai elemzéseket importálni. A kutatási téma tehát az, hogy a természetes nyelv viszonyjelölőiből hogyan lehet következtetni olyan kapcsolatokra, amelyeket az indexelés használ.

5. A S Z I N T A X I S

5.1 EGY ESZME SZÜLETÉSE

Ismét Gardin gondolata - Kolumbus tojása - idézhető. Ha objektumok, eszmék, dokumentumok leírására valamilyen nyelv vehető igénybe, miért ne lehetne ennek a nyelvnek relációit a természetes nyelvből származtatni?

- Az információ rendszereknek feltett kérdések is, a rendszer nyújtotta válaszok is természetes nyelvűek. Miért nem lehet közel álló az is a természetes nyelvhez, ami a kérdés és a válasz között van? Kisebb lenne a minden átalakítást, transzformációt kísérő információ veszteség.

- Már nem vitás, hogy a természetes nyelv produkálja az indexelő kifejezéseket, vagy közvetlenül szövegszavak formájában, vagy közvetve tárgyszójegyzékek, szabályozott szótárak, tézauruszok induló szóanyagaként. Ha a természetes nyelv alkalmas az indexelő terminusok prezentálására, miért nem alkalmas a kapcsolatok meghatározására is?

- A természetes nyelv abban is mintája az osztályozásnak, hogy benne az elemi egységek /szavak/ a gondolatot hordozó mondatokká formálódnak. Vagy sokat citált analógiánkkal fogalmazva: az ornamentikai elemek kompozícióvá állnak össze benne.

- Végül, a tájékoztatás helyzete az információ áradatával szemben teljesen reménytelen automatizálás nélkül. Az au-

tomatizálás pedig szövegelemzésre épül. Ha a szövegek amugy is analizálásra kerülnek, akkor miért nem lehet a szó- és kifejezés elemzés eljárásait összekötni a viszonyrendszer elemzésével?

Ez a négy kérdés három feltételt szab meg.

- Minden szövegelemzés konkrét nyelvhez van kötve. /Orosz, angol, magyar, stb./ Az elemzési eljárások csupán analógiaként vihetők át egyik nyelvről a másikra. Az elemzéseket ennek folytán a tájékoztatás forrásdokumentumainak minden nyelvén el kellene végezni.

- A tudomány nemzetközi, a szövegek különböző nyelvűek. Azokat a relációkat azonban, amelyeket bármilyen eljárás meghatároz, már nem lehet egyik nyelvhez sem kötni. Valamilyen "metanyelvi", nyelvek fölötti, minden nyelvben élő, közös viszonyrendszerre lenne szükség. Kérdés, hogy az általános nyelvészetben megvan-e a felkészültség erre?

- Végül, a kapcsolatoknak nemcsak szemantikai relációknak kell lenniök, amely tisztába teszi a kifejezések jelentését és bizonyos kontextustól független viszonyok meghatározását engedi meg /pl. hogy a ló patás emlős/, hanem szintaktikaiaknak, hogy az adott szövegben érvényes viszonyrendszer transzformációja megőrizze a szövegbéli rendezettségét. Azaz: a csak fogalmaknak a dokumentumokhoz rendelésével végzett osztályozás helyett - ez képes rámutatni a témára - olyan eljárás szülessen, amely a gondolatot, tézist, elméletet, tehát a szöveg mondandóját, állítását képes felmutatni. Az osz-

tályozás számára a természetes nyelvnek fontossága mind lekszikai, mind szintaktikai szempontból a fentiekből ered. Ebből még nem következik, hogy az információt leíró nyelvnek szövegszavasnak kell lennie.

Aligha kell azonban azt kifejtteni, hogy a természetes nyelv viszonyrendszere soha nem képezhető le egy-egyértelműen osztályozási kapcsolatokká. Az elemzés nehézsége éppen abban áll, hogy a kapcsolat több-többértelmű: egy osztályozási viszonynak több grammatikai eset felelhet meg és viszont.

Az első /alighanem egyetlen/ közlemény, amely a magyar nyelv esetrendszere és az osztályozási relációk közt kapcsolatot keresett, 1968-ban látott napvilágot. [86] Ez is csak a magyar határozórendszerrel foglalkozik. Példaként bemutatjuk annak igazolására, hogy a határozók esetrendszere miként alkalmas osztályozási relációk meghatározására. A magyar határozók az alábbiak.

inessivus	/valamiben/
elativus	/valamiből/
illativus	/valamibe/
superessivus	/valamin/
delativus	/valamiről/
sublativus	/valamire/
adessivus	/valaminél/
ablativus	/valamitől/
allativus	/valamihez/
locativus	/valahol/

terminativus	/valameddig/
dativus	/valami részére/
temporalis	/valamikor/
iterativus	/ismétlődés, pl. naponta/
essivus	/valamiül való lét, pl. kutyául/
factivus	/valamivé/
instrumentalis	/valamivel/
comitativus	/valamistül/
causalis	/valami miatt, okból/
finalis	/valamiért, valami célból/
modalis	/valami módon/
formalis	/valamiképpen, valami alakban/
distributivus	/részletekbe vétel, pl. egyenként/
multiplicativus	/valahányszor/

E huszonnégy határozóból öt csoport képezhető, amelyek az osztályozásban relációként értékelhetők:

1. csoport: -ben, valamin, -nél, valahol, -kor;
2. csoport: -ből, -ről, -től, valami miatt;
3. csoport: -be, -re, -hez, valameddig, -nek, valamivé, valami célból;
4. csoport: ismétlődés, valamiül, valami módon, -ként, valahányszor;
5. csoport: -vel, -stül.

Az öt kapcsolatfajta azonban csak kontextusban élhet, nem építhető be tézauruszba, tárgyszógyűjteménybe.

5.2 A RELÁCIÓKRÓL

Mindjárt adódik tehát az információkereső nyelvek és az osztályozási rendszerek relációinak egyik fontos típuspárja. Léteznek olyanok, amelyek az indextételben élnek és a kifejezések szintaktikai viszonyait fejezik ki. Az indextételt egészzé, mondatértékűvé konstruálják. A másik fajta kapcsolat az osztályozási kifejezések szótárába, tézauruszba, tárgyszójegyzékekbe épülnek be. Függetlenek az indextelektől.

E megkülönböztetés megjelenik a SYNTOL-ban, ott a priori a kapcsolat akkor, ha előre tudható és a term definíciójától függ. A reláció aposteriori, ha csak a szöveg vizsgálatából határozható meg. Az előbbi szemantikai kapcsolat, az utóbbi szintaktikai. Pl. a "szék" alárendelt viszonyban áll a "butor"-ral, ez a fogalom definíciójából derül ki: a szék olyan butor, amely... Ez a viszony tehát szemantikai, a priori. A szék azonban lehet pl. artisták kelléke egy mutatványban. Itt a szék az artista kellékek egyike, de ez a kapcsolat nem ered a definícióból, csak kontextusból állapítható meg. Csak az adott szövegben keletkezett viszony, a szerző hozta létre, érvényessége is csak az adott szövegre állítható.

Kezd kibontakozni a kapcsolat két fajtájának természete. Az is, hogy miért kell ezeket így elkülöníteni. Hiába épül be ugyanis egy tézauruszba vagy más szótárba a szék-bu-

tor faj-nem reláció, mert ez aligha segít azon, akit a székekkel végrehajtható artista mutatványok után keres. A keresőkérdések sem az egyedi fogalmak szerint, vagy ezek egyszerű halmazata alapján hangzanak el, hanem a mondatok szintjén, a kérdésben az elemek szintaktikai kapcsolatokkal összeforrasztott kifejezéseké válnak. A példánál maradva: a szék instrumentum, eszköz, a természetes nyelvben eszközhatározó, az osztályozásban instrumentális viszonyban kell megjelenjék. Semmire nem megyünk azzal, hogy a szék butor.

A PRECIS hasonlóan tesz megkülönböztetést. A kapcsolat ott is lehet a priori és aposteriori. A Manuel definíciója szerint az a priori kapcsolatokat a szótár tartalmazza, az aposteriorit pedig a PRECIS indextételei. [87] Austin példája az egér. Definíció szerint az egér rágcsáló. Ez az, amit mondat nélkül tudni lehet róla. Tehát a priori reláció, szótárban a helye. Az egér azonban lehet kísérleti állat, kártevő a mezőgazdaságban, betegségek terjesztője, stb. Ezek pedig ismét csak szövegben létrehozott kapcsolatok. Aposteriori.

Soha nem szabad elfelejteni a relációmegőrző transzformáció posztulátumát!

De honnan származzanak akkor a relációk? Ismét csak kérdezhető, hogy ha az információ leírására szolgáló nyelv elemei a természetes nyelvből valók, akkor eredhetnek-e innen a relációk is?

Derek Austin érdeme, hogy a generatív grammatika mélyesetrendszerét felfedezte erre a célra és inkorporálta az

osztályozáselméletbe. Lépése igen merész volt, mert a nyelvészek maguk is bizonyos szkepszissel fogadják ezt az elméletet. Noam Chomsky sem tekinti mélystruktúra koncepcióját megállapodott elméletnek. [88]

A mélyesetrendszeret Fillmore dolgozta ki ideiglenesen, [89] fenntartva azt, hogy a továbbiakban több mélyesetre lesz szükség. Esetrendszere az alábbi.

1. Agentivus, a cselekvőnek, a cselekvés kezdeményezőjének esete. Pl. János fut az erdőben.
2. Instrumentális, a cselekvésben szükségszerűen részt vevő tárgy vagy erő. János kinyitotta az ajtót. /Ugyanaz, mint a kulcs kinyitotta az ajtót-ban a kulcs/
3. Dativus, az igével megnevezett állapottól vagy a cselekvéstől érintett lény. János azt hitte, győzni fog. Meggyőztük Jánost, hogy győzni fog.
4. Faktitivus, az igei jelentés részeként értelmezett tárgy vagy személy. János házat épített.
5. Locativus, a cselekvés helye, térbeli iránya. Chicago szeles. /Ugyanaz, mint Chicagóban fúj a szél./
6. Objektivus, egy semleges eset, kifejezhető főnévvel, amelynek szerepét az ige által hordozott cselekvésben vagy állapotban maga az ige határozza meg. /János egy dalt énekelt./

Austin a PRECIS számára ennek továbbfejlesztett változatát vezette be. Az esetrendszer így alakul.

1. Szenvedő, pacientivus /patient/

Dolog, amelyen végbemegy a rajta végrehajtott, az ige által megnevezett cselekvésnek megfelelő feltétel- vagy állapotváltozás. A fa magas. János betörte az ablakot.

2. Komplementum [90]

Kiegészíti, részletezi, specifikálja az igét. János egy himnuszot énekelt. János hatvan fontot nyom.

3. Eredmény, faktitivus

Az ige által kifejezett cselekvés eredménye, terméke. János épített egy házat.

4. Kedvezményezett /beneficiary/

Rá irányul a cselekvés, anélkül azonban, hogy helyzete, vagy a feltételek megváltoznának. János elvesztette pénztárcáját.

5. Cselekvő, agentivus /agent/

A cselekvés végrehajtója, ösztönzője. János ujságot olvas.

6. Eszköz, instrumentális

Amit/akit a cselekvő felhasznál a cselekvés végrehajtásához vagy ösztönzéséhez. János egy késsel elvágta a kötelet.

7. Átéltő /experienter/

Képessége az ige által hordozott cselekvés felfo-

gása, de maga nem megy végbe állapotváltozáson.

János szerelembe esett Máriával.

8. Hely, locativus

A térbeli korlátozás esete. Budapesten hideg tél volt.

9. Idő, temporalis

Az időbeli korlátozás esete. Tegnap esett az eső.

Az eseteket generáló igéknek négy fajtája van. Kifejezhetnek

- állapotot [91] (az ut széles, a váza törött)
- folyamatot (a ruha beszakadt)
- "tárgyatlan" cselekvést (a madarak elköltöztek)
- "tárgyas" cselekvést (János betörte az ablakot)

Az esetrendszer "univerzális"-nak mondják, mindenesetre nyelvtől függetlennek. /A néki tulajdonított univerzális grammatika ellen Chomsky tiltakozott, mondván, hogy nincs univerzális grammatika, csak univerzális grammatika elmélet. [92]/

Ha a mélyesetrendszer bármely természetes nyelvben képes megszerkeszteni a mondatot, akkor az osztályozásban is viszont kell látnunk őket. Mielőtt erre a problémára rátérnénk, a relációk miatt egy viharos gyorsasággal elterjedt irányzatot érdemes megvizsgálni.

5.3 INTERMEZZO: A TÉZAURUSZOK

A hatvanas és hetvenes években általánossá vált az a felismerés és meggyőződés, hogy a relációk az osztályozásban elem-értékűek. Ez a rendszerelméleti fogalom azt jelezte, hogy a reláció az osztályozó/indexelő rendszereknek ugyanolyan értékű része, mint a szótári egységek, illetve fogalmak. Valamire ugyanis sikerült megoldást találni. Eladdig az osztályozó/indexelő eljárások, tárgyszórendszerek vagy enumeratív módszereket alkalmaztak az indextételek felírásában. Kétségtelenül a fogalmak rugalmas kombinálhatóságának előnyeivel. Vagy a strukturált, hierarchikus osztályozási rendszerek konstrukciós módszereit, amelyek jobbra monohierarchikusak, de legalábbis monoparadigmájuk voltak. A tézauruszok elkerülték ezt a mindenképpen hátrányokkal járó alternatívát.

Ugy tűnt, a tézauruszok alkalmasak a fogalmi kapcsolatok befogadására. Lexikai egységeik pedig származhattak a természetes nyelvből is.

A tézaurusz relációk azonban kontextus függetlenek, a kapcsolatokból annyit lehet egy tézauruszba beépíteni, amennyit a fogalmakról önmagukban állítani lehet. A már ismert kifejezéssel élve a priori relációkra koncentrálni. Egy tézaurusz nem képes megmondani felhasználásának módját, nem szól semmit arról, hogyan lehet a szöveg kapcsolatait megőrizni. Tehát nem tesz eleget a kapcsolatmegőrző transzformáció követelményének.

Eredményei ellenére a tézauruszok megjelenése azt a fejlődési vonalat szakította meg, amely az osztályozó/indexelő rendszereket egyre erősebben a természetes nyelv szintaxisához kötötte. A tézauruszok kapcsolata a természetes nyelvvel csak abban állt, hogy fogalommeghatározó, fogalomgyűjtő módszerei között előnyben részesítette a természetes nyelv szó és szerkezet gyakorisági vizsgálatát.

A tézauruszépítés maga munkaigényes feladat. Hiszen a kifejezések összegyűjtése és szelektálása után fel kellett ismerni minden deszkriptor kapcsolatát minden olyan más deszkriptorral, amelyhez a bevezetett relációk valamelyike érvényes. Az intellektuális ráfordítás igen magas volt, amit nem mindig lehetett biztosítani. Ennek az lett a következménye, hogy a relációk terén újabb és újabb kompromisszumok születtek. Az igényesség lanyhult, a tézauruszok tendáltak a régi tárgyszórendszerek felé. Végül alig maradt több, mint a szinonima, homonima reláció, genus-species, totum-pars, és egy "rokon" /related term/ kapcsolat. Pedig a lehetőségek száma sokkal bővebb. A fentiekén kívüli kapcsolatokra álljon itt Horváth Tibor - Varga Dénes táblázata. [93]

(ld. 109. oldal)

Számos esetben a relációkat kevésbé mereven kezelték. Pl. a nem-faj, egész-rész kapcsolatot egy elmosódottabb "tágabb-szűkebb" reláció helyettesíti. [94] Voltak törekvések arra, hogy a fogalmak kombinálhatóságát erősítsék. Példa rá a British Standards Institution un. ROOT tézauru-

	Produktum, termék	Melléktermék	Módszer, eljárás	Folyamat, művelet	Feltétel, állapot, befolyásoló tényező	Ellentét	Hatás, ellenhatás	Ok	Cél, rendeltetés	Eszköz	Tulajdonság, minőség	Elmélet	Korrelatív	Hasonlóság
Produktum, termék		1		1					1		1			1
Melléktermék								1			1			
Módszer, eljárás	1			1	1				1	1		1		
Folyamat, művelet	1	1	1		1		1		1	1		1		
Feltétel, állapot befolyásoló tényező				1	1		1	1	1		1			1
Ellentét			1					1			1			
Hatás, ellenhatás				1	1	1		1	1	1		1		1
Ok					1	1	1				1	1		
Cél, rendeltetés	1		1	1		1				1				
Eszköz	1		1	1			1		1					1
Tulajdonság, minőség	1	1			1	1			1					1
Elmélet			1	1	1				1					
Korrelatív													1	
Hasonlóság	1				1	1		1			1			

sza. [95] Kapcsolatrendszere /a jelekkel együtt/ az alábbi.

< tágabb

> szűkebb

~ rokonsági kapcsolat

*< tágabb egy alternatív hierarchiában

*> szűkebb egy alternatív hierarchiában

*~ rokonsági egy alternatív hierarchiában

= szinonima

→ a szinonima inverze, a nyíl után álló deszkriptor használatos a nyíl előtti helyett

+ két fogalom szintézise, amely egy tiltott fogalom helyett áll

**szintetizált fogalom, bizonyos értelemben az előbbi inverze, egy kombináció helyett áll

=** annak szimbóluma, hogy a jelet megelőző fogalmak használatosak az őt követők helyett

... értelmezés /szöveg/

A tézauruszok elégtelensége nem abban áll, hogy ezeket a relációkat a szótárba bevezették, hanem abban, hogy itt rekedtek meg. Szemantikai megoldást kínálnak olyan problémára, amelyik alapvetően szintaktikai természetű. A tézauruszok relációi nem vesznek részt az indextételek felírásában. Az osztályozás/indexelés ezért a végtermékben, a dokumentum image-ében nem tudja a tézauruszrelációkat szerephez juttatni. Ezeknek a relációknak haszna csak abban áll,

hogy az indexelőt segítik a helyes kifejezések kiválasztásában. /A visszakeresésben e relációk kaphatnak szerepet. Ezzel a témával itt nem foglalkozunk./

Maga a tézauruszelmélet nem vonja feltétlenül maga után ezt a következményt. Sem a tézauruszok elszegényítését, sem a szintaxis megligálását. A szintaxisra orientáló nyelvek, pl. a PRECIS kezelni tud tézauruszból származó kifejezéseket is. Pontosan: tudna kezelni.

A tézaurusz kérdést erősen megkeverte a többnyelvűség problémája. Világosan látni kell, hogy a tézaurusz relációk szemantikai kapcsolatokat tükröznek, ezek pedig jobbra nyelvhez kötöttek. A többnyelvű tézaurusz feltételezi a kifejezések mellett a kapcsolatok megfeleltetését is. Mivel veszteség nélkül nem lehetséges, azért a javasolt megoldások formálisak, tehát nem megoldások. [96] A többnyelvű tézaurusz deszkriptorcikkei feltüntethetik a két /több/ nyelv közötti különbségeket. De ez nem szünteti meg a problémát, csak jelzi, hogy létezik. Azzal a jelzéssel, ami nem jár együtt a zajok eltüntetésével, nem kezdhető semmi. Egy példával mindez jobban elmagyarázható. A magyar "középiskola" oroszul "srednaâ škola". De a Szovjetunió tizosztályos iskolarendszerében a magyar középiskolának nincs valódi, csak körülbelüli megfelelője. Az eltérés a tézauruszban jelezhető. A visszakereső rendszer srednaâ škola-t keres akkor is, ha a tézaurusz rámutat az eltérésre. Ha a rendszeret lekérdezzük, az eredmény nem a középiskolásra vonatkozó irodalom lesz, hanem valami nagyon zajos információ.

A tézauruszok a fentiek miatt egy fejlődésben megrekedt kezdeménynek bizonyultak. Mivel nemzetközi rendszerek pártfogása alatt állnak, a tézaurusz-intermezzo olyan közjátéknak bizonyult a hangversenyben, ami az egész előadás helyére küzdötte fel magát, miközben maga az előadás elmaradt. [97]

5.4 SZINTAXISOK

Négy-öt szó halmaza még nem mondat. Akkor sem, ha jelentésüket pontosították egy szótárban és relációikat más lexikai egységekhez meghatározták.

Szükségszerű, hogy az információkereső nyelvek elmélete elvezessen ehhez a felismeréshez és keresse azt a grammatikát, amely az indextételekben az elemeket magasabb egységbe szintetizálja. Ennek a grammatikának azt is biztosítania kell, hogy átmentse az eredeti szövegben élő kapcsolatokat az indextételekbe.

Az információkereső nyelvek lexikájában egy fejlődési sor tapasztalható: a fejlődés állomásait a tárgyszavas osztályozás - információs tézaurusz - szintaktikus nyelvek szakaszai mutatják. Mindegyik fejlődési fázis kimunkál egész sor hasznos módszert, megoldást a rákövetkező számára. A tárgyszavas osztályozás csirájában hordozza azokat a szemantikai megoldásokat, amelyek majd a tézauruszokban teljes pompával bontakoznak ki. [98] A tézauruszok a relációk szerepé-

nek hangsúlyozásával előkészítik a talajt a szintaktikai nyelvek számára.

A szintaktikai információkereső nyelvek is leírnak egy fejlődési sort. Foskett [99] azt a Farradane - SYNTOL - PRECIS fejlődésben látja. Ezen az uton a szintaktikai szabályok formalizálódnak. Mindegyik a természetes nyelv grammatikájára tekint, miközben igyekeznek a konkrét természetes nyelvektől elszakadni. Szintaxisuk mintegy meta-nyelvi, a viszonyrendszert mélyebben és következetesebben kívánják definiálni bármely nyelv nyelvtanánál. Ezért alkalmasak nemzetközi használatra is. Óhatatlanul felmerül tehát az a gondolat, hogy a nemzetközi rendszereknek, jelesen az NTMIR-nek sem az eddigi kísérletei szerint, hanem ezen az uton kellene a "nyelvi kihívás" nehézségeit leküzdenie.

Az 5.2 fejezet leírta a mélyeseteket, bár a nyelvészek is provizórikusnak tekintik. A figyelem mégis feléjük fordult.

A SYNTOL is metanyelvről beszél, amelynek igen egyszerűnek kell lennie azért, hogy képes legyen az inputok /különböző természetes nyelvek/ széles körét befogni. Bevezeti a kételemű szintagmákat /dyadic string/, amelyek két osztályozási kifejezésből és ezek viszonyából állnak. A kapcsolatok négy általánosított nyelvtani esetben jelennek meg. Ezek megfogalmazása még nem árulja el a természetes nyelvi gyökereket, értelmezésük, kibontásuk azonban igen. Minden eset

kimunkálja a maga számára azt, hogy grammatikailag hogyan értelmezhető. A négy eset, reláció az az alábbi.

1. Predikativ. Rendszerint melléknév + főnév alakú.

Pl. broader term /tágabb osztályozási kifejezés/.

2. Asszociativ. Alakja szerint főnév + főnév. Pl. információ, visszakeresés.

3. Egymásra következő /consecutive/. A fogalmak függőségét az a sorrend mutatja, amely az eredeti szövegben található. Pl. háború /okozta/ gazdasági válság.

4. Koordinativ. Feltételezi az összehasonlítás eszméjét, jelenlétét. Pl. kórházi ápolás, otthoni ápolás.

Azonnal szembe ötlük, e négy reláció arra törekszik, hogy a szövegekből felismerhető legyen. Figyeljük meg, alkalmasak arra, hogy segítségükkel létrejöjjön a relációmegőrző transzformáció.

A négy reláció pontos definíciója hiányzik. Részben azért, mert csak kereteket kíván adni. Másfelől a Gardin vezette csoport éppen arra törekedett, hogy a kontextus mutassa világosan a viszony fajtáját. A viszonyok meghatározása nem azt mondja meg pontosan, hogy a reláció mi, hanem inkább azt, hogyan fordul elő a kontextusban.

A SYNTOL-t jó eredménnyel tesztelték lélektani, néprajzi, szociológiai és fiziológiai területen. De sor kerül vizsgálatára kemény tudományokban is, pl. a fizikában. /A

SYNTOL megrendelője az EURATOM./

A SYNTOL Farradane relacionális indexelésétől láthatóan határozott határozott előrelépés. A fejlődés újabb fázisát egy egészen perfekt szintaktikájú rendszer, a PRECIS /PREserved Context Indexing System/ képviseli.

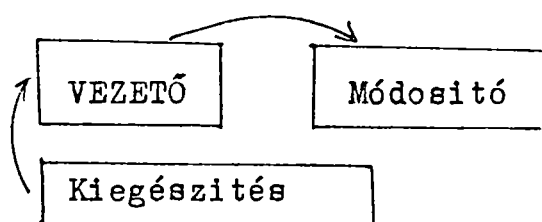
A PRECIS elemzését kicsit részletesebben kívánatos megadni, mivel az OPIR-ban ennek alkalmazására került sor. Másfelől a PRECIS egyszerű leírását a TMT-ben megadtuk [100], az ismertetés túlzott részletezéssel itt nem kívánatos, hanem inkább elemzése. /A közölt táblázat a TMT számára írt összefoglalóból való./

A PRECIS [101] olyan szintaxis, amely különböző típusú osztályozási kifejezésekre alkalmazható. Szabad szövegszavak, tárgyszavak, deskriptorok, egytagú kifejezések és szintagmák esetén egyaránt. Célja az, hogy az osztályozási/indexelési elemekből mondatokat állítson össze, a természetes nyelv mondataihoz hasonló indextételt. Egy PRECIS indextétel annyiban nem hasonlít egy természetes nyelvi mondat-hoz, hogy hiányzik belőle az állítmány. Valójában mégis mondat, mert a "Gigli Verdit énekel" mondat helyett a "Gigli Verdi éneklése" a transzformációs generatív grammatika értelmében ugyanannak a magmondatnak transzformációja. Eszerint egy PRECIS indextétel valódi mondat.

A mondatbeli viszonyokat un. operátorok fejezik ki, amelyeket számmal /0-6/ és betűkkel /f,g,p,q,r,s,t,u/ jelölnek. Különböző technikai fogások jelzésére un. kódok áll-

nak rendelkezésre. Ezeket § + egy-két karakter fejezi ki. E technikai fogások közül néhány : ragok (végződés) beillesztése, több téma szétválasztása, kifejezések helyettesítése másokkal bizonyos feltételek esetén, stb. Az operátorokat és kódokat az alábbi táblázat rendszerezi. (ld. 117. oldal)

Az operátorok és kódok elemzése előtt meg kell ismer-
ni azt, hogy a PRECIS indextételből hogyan állnak elő a
PRECIS mondatok vagy bemeneti tételek. Tegyük fel, hogy egy
feldolgozási egységet A,B,C,D,E - mondjuk - tárgyszavak ír-
nak le. Az operátorokat egyelőre ne vegyük figyelembe. A
tárgyszavak fölé helyezett pipával jelölhető az, hogy a tárgy-
szót besorolási/elérési ismérvnek kívánjuk-e vagy sem. Te-
gyük fel, a példa minden tárgyszava besorolandó. Akkor egy
rotálási szabály e kifejezéseket rendre három pozícióba he-
lyezi úgy, hogy a kifejezések e pozíciókban változtatják egy-
másat. A három hely: vezető /lead/, amely besorolási/elérési
pozíció; tőle jobbra a módosító pozíció áll /qualifier/; a-
latta a kiegészítés /display/ az alábbi módon:



A fenti tárgyszavakból az alábbi bemenetek állnak elő.

A	B	A	C	BA	D	CBA	EDCBA
BCDE	CDE		DE		E		

A kódok és a szerepoperátorok rendszere

Elsődleges kódok

Témakapcsolók	Sx	Koordinált téma első eleme
	Sy	Koordinált téma következő eleme
	Sz	Közös elem
Kifejezőkódok*	Sa	Közös főnév
	Sc	Tulajdonnév
	Sd	Helynév

Másodlagos kódok

Különbségek		
Megelőző különbségek (3 karakter)	S0	1. és 2. karakter: Nem vezető, különírandó
	S1	Nem vezető, egybeírandó
	S2	Vezető, különírandó
	S3	Vezető, egybeírandó
Dátum mint különbség	Sd	3. karakter: 1 – 9, a fókuszról való távolság
Közbevetett különbségek	Sn	Nem vezető közbevetett különbség
	So	Vezető közbevetett különbség
Kapcsolók	Sv, Sw	

Tipográfiai kódok*

Elsődleges operátorok

A magfogalmak környezete	0	Helyek
Magfogalmak	1	Kulcsrendszer: a tárgyas cselekvés tárgya; a tárgyaltalan cselekvés cselekvője (végrehajtója)
	2	Cselekvés, hatás
	3	Tárgyas cselekvés végrehajtója (cselekvő eszköz); a rendszerbe bekerült dolog; tényező
A magon kívüli fogalmak	4	Szempontr mint forma; megközelítési mód
	5	Tanulmányozott, vizsgált eset: pl. tanulmányozott terület, mintasokaság
	6	A dokumentum formája; Használói célcsoport

Másodlagos operátorok

Koordinált fogalmak	f	“Kötött” koordinált fogalom
	g	Koordinált fogalom
Függő elemek	p	Rész; tulajdonság
	q	Kvázigenetikus csoport tagja
	r	Gyűjtőfogalmak, csoportok
Speciális cselekvésosztályok	s	Szerepmeghatározó; irányított tulajdonság
	t	A szerző által létrehozott asszociáció
	u	Kétirányú kölcsönhatás

* Ezeket a kódokat nem tárgyaljuk.

Az ötödik bemeneti lánc az eredeti felírásnak éppen fordított sorrendjét adja. A rotálási szabálynak ez csak egyik esete. A forgatás szükség szerint változtatható, hogy a szavak értelmes sorrendje követhesse a természetes nyelvben megszokottat. Bizonyos operátorsémák a sorrendi változást automatikusan kiváltják.

Egy PRECIS indextétel vázát az (1), (2), és (3) un. magoperátorok hozzák létre. Az indextétel felírása a cselekvés, történés meghatározásával indul, és (2) operátort kap. Olyan tárgyszavak is (2) operátorral lépnek be a tételbe, amelyek cselekvésként értelmezhetők. Pl. nemzetközi kapcsolatok. Vagy hatásként. Pl. fájdalom. A (2) operátor ismétlődhet.

A következő lépésben a cselekvés, hatás tárgyának meghatározása következik, (1) operátorral. Az (1) jelentése: tárgyas cselekvés tárgya. De az (1) operátor jelzi a tárgyatlan cselekvés végrehajtóját is. Mindkettőre példa.

(1) matematika

(1) folyadék

(2) tanítás

(2) áramlás

A (3) operátor jelzi a cselekvés végrehajtóját, ágensét.

E tételszerkezet megfelel a tárgy-állitmány-alany szerkezetű, illetve egy angol szenvedő szerkezetű mondatnak. Pl. "egy népdalt énekelt a kórus"; "the book was written by John". Ezek a példák egy (1), (2), (3) operátorsémát kívánnak.

A forgatási szabály ezt fokozatosan (3), (2), (1) alakuvá változtatja: "a kórus énekelt egy népdalt", illetve

"John wrote the book".

Az (1), (2), (3) operátorséma maga után vonja az un. predikátum transzformációt, a sorrend egyfajta megváltozását. Ez abban áll, hogy ha a (3) kerül vezető pozícióba, akkor a sorrend megőrzi az eredeti felírását: (3), (1), (2) lesz. Tehát "a kórus egy népdalt énekelt" tökéletes sorrendű magyar mondat keletkezik.

Nem állitható azonban, hogy az (1), (2) és (3) operátorok egyszerűen a tárgy, állítmány és az alany mondatrészek megfelelőit jelölik. Az operátorok ugyanis a mélyeseteknek felelnek meg. A magoperátorok több mélyesetet összesítenek megukban.

Az (1) operátor definíció szerint a tárgyas cselekvés tárgya és a tárgyatlan cselekvés végrehajtója, tehát mindezekelőtt szenvedő /patientivus/ esetet jelöl, de ágentivust is tárgyatlan cselekvésnél. A kedvezményezett /beneficiary/ is (1), pl. János elvesztette pénztárcáját mondatban. (1) operátort kíván még a komplementum és faktitívus. Ha meggondoljuk, a komplementum és faktitívus tárgyas cselekvésben annak tárgya, tárgyatlanban az ágens /pl. madár - repülés/ valamiként a feldolgozandó szöveg témáját, tárgyát mutatja.

A (3) operátor funkciója jelölni az ágenst /targyas cselekvésben/, az instrumentálist, /az eszköz a cselekvés végrehajtója pl. János késsel elvágta a zsinórt mondatban a kés/. Az átélő /experienter/ szintén (3). A János szerelembe esett Máriával mondatban János.

Az operátorok tehát a mélyeseteknek nem közvetlen leképezései, hanem összevont megjelenítésükre adnak lehetőséget.

A főoperátoroknak lehetnek ún. függő elemei, azaz bővitményei. A főoperátor és a betűvel jelölt másodlagos operátorok együtt alkotják a függő elemek blokkját. Mondattani analógiával tehát nagyobb szerkezeti egységekké, szintagmákká, frázisokká állnak össze.

Néhány másodlagos operátor részletesebben tárgyalható azért, hogy az összes többi működésére példaként álljon.

A (p) operátor valaminek részét vagy tulajdonságát jelöli. A tézauruszrelációkhoz szokott gondolkodás számára idegennek tűnhet az, hogy ezt a két eltérő kapcsolatot ugyanaz az operátor jelöli. De mondatokban mindkettő birtokos szerkezetben élhet. Pl. az autónak a motorja és az autónak a mérete egyaránt birtokos szerkezetű. Tehát

(l) autó	(l) autó
(p) motor	(p) méret

irandó. A (p) ismétlődhet: a résznek is lehet része, tulajdonsága, tulajdonságnak is tulajdonsága:

(l) könyvtárak	(l) autó
(p) könyvtárosok	(p) motor
(p) pályakép	(p) méret

A (q) operátor kvázigenerikus kapcsolatot fejez ki. Segítségével sok osztályozási finomság oldható meg. A generikus kapcsolat (q) szerint azért kvázi, mert fentebb látható volt, hogy az a priori generikus kapcsolat feltüntetésé-
se a szótár feladata, nem az indextételé. /Bár az indextétel

is feltüntetheti./

(l) artista eszközök (l) kísérleti állatok

(q) szék (q) egér

A (q) operátorral fejezhető ki individuumok nevei, ha az a csoport is fontos, amelynek az individuum tagja.

(l) könyvtárak

(q) Országos Széchényi Könyvtár

Ha egy individuum nem szokásos szerepében jelenik meg, ezt szintén (q) jelzi a következőként.

(l) pedagógusok

(q) Németh László

Homonim kifejezések értelmezése (q)-val:

(l) madarak (l) emelőgépek

(q) daru (q) daru

Végül a (q) funkciók közül egy olyan, amely szerepet kaphat a nemzetközi rendszerek nyelvi nehézségeinek leküzdésében. Ha egy kifejezésnek nincs pontos megfelelője, ezt (q) tünteti fel:

(l) középiskola

(q) srednaâ škola

Ugyanigy pl, ha egy angol nem tudja pontosan visszaadni a magyar "betyár"-t, akkor

(l) outlaw

(q) betyár

alakot ír. Ha meggondoljuk, akkor ez a felírás a visszakeresésben is pontos eredményt nyújt: lekérdezhetők a zsiványok,

utonállók, de ezen belül tovább szelektálhatók a betyárok.

Az (r) gyűjtőkifejezéseket /assembly/ jelez:

(l) elefánt

(l) Henrik, VIII., angol király

(r) csordák

(r) feleségek

A (g) ugyanazon tárgykörön belüli mellérendelést fejez ki.

(l) autó

(p) motor

(g) karosszéria

azt jelzi, hogy a karosszéria a motorral mellérendelten az autó része. A (g) ismételhető. A példában nem (p) ismétlődik, mert

(l) autó

(p) motor

(p) porlasztó

a rész részét jelzi.

Az (f) szintén mellérendelést jelöl akkor, ha a rótálás-kor a mellérendelt elemek mindegyikét együttesen kívánjuk mozgatni, mivel a mellérendelt tárgyszavak elválása értelmi zavart okozna. Ha pl. színészek és rendezők vegyes bizottsága alakul, akkor nem engedhető meg a színészek - vegyes bizottság szerkezet, hiszen kívül vegyes a bizottság, ha csak a színészek alkotják?

Az (u) kétirányú kölcsönhatást fejez ki. Valójában az (1), (2), (1) operátorsémában (2) helyére lépve (1), (u), (1)-et hoz létre. Ebben az esetben a két (1) operátorral jelölt

tárgyszó szerepe felcserélhető: (1) Magyarország (u) nemzetközi kapcsolatok (1) Románia ugyanaz, mint Románia kapcsolatai Magyarországgal. Az operátorsémában (u) megjelenése (2) helyett az indextételből előálló láncok írásmódját változtatja meg.

Az (s) különleges "szerep"-et, "hatás"-t jelöl, amely a megszokottól eltérő.

(1) csecsemők

(2) gondozás

(3) apák

(s) szerepe

Megjegyzendő, hogy az angol PRECIS (s), (3) sorrendű, hiszen a "role of..." ezt kívánja.

A (t) operátor fontos lehet számos tudományos értekezés indexelésekor. Jelentése: összehasonlítva, kifejtve, magyarázva. Pl.

(2) gyógyszeres kezelés

(2) jelentés

(t) összehasonlítás

(t) értelmezés

(2) műtéti kezelés

(3) információ elmélet

A felsorolt operátorok mindegyikénél megfigyelhető, milyen könnyen igazodnak a relációmegőrző transzformáció követelményeihez. A szerző által létrehozott, kontextus függő kapcsolatok nemcsak az (1), (2), (3) megoperátorok sémájában tűnnek elő, ti. hogy ki/mi mit állít, ki mit cselekszik, hanem a kapcsolatok finomabb szövedéke is feltárulhat, amelyet pl. a (t) operátor mutat.

A további számokkal jelölt fő operátorok az indexté-
tel számára fontos információkat jeleznek, de kevésbé illeszt-
hetők be mondatrészenként. /Talán mellékmondatok funkcióját
tölthetik be./

A (4) operátor nézőpontot, álláspontot jelöl: marxis-
ta nézőpontból, lélektani szempontból, szakszervezeti állás-
pont, stb.

Az (5) operátor mintasokaságot vagy tanulmányozott ese-
tet, bizonyítási forrást jelez. Pl.

- | | |
|------------------------|------------------------|
| (2) olvasásvizsgálat | (2) özönviz |
| (5) mintasokaság | (5) bizonyítás forrása |
| (q) ezer falusi olvasó | (q) Biblia |

A (6) operátor dokumentumtipust és felhasználói célcso-
portot fejez ki. Pl.

- (2) családi élet
- (6) animációs film
- (6) gyerekeknek

A kódok szerepe nem lebecsülendő, jóllehet nem reláció-
kat jelölnek. De a relációkat nyilvánvalóbbá teszik, az index-
tétel további árnyalását szolgálják, fokozzák annak rugalmas-
ságát. Esztétikai feladatuk, hogy a mondatok szépek legyenek.
A kifejtés igénye nélkül néhány esetet inkább példaként idé-
zünk.

A \$v és \$w prepozíciók beillesztését szolgálja. A \$v je-
löli azt, hogy ha a \$v-vel ellátott tárgyszó felett álló ki-
fejezés/ek/ kerül/nek/ vezetőbe, akkor az alatta álló kapja a

kód által jelölt prepozíciót. A \$w hasonló, de a felette álló kapja a prepozíciót akkor, ha az alatta álló/k/ a vezető/k/.

(1) Industry

(2) management

(2) control \$v=by \$w=of

(3) personnel

A generált mondatok, bemenetek ezek után az alábbiak. /Egyben arra is legyen példa, hogyan keletkeznek az indextételből a bemeneti láncok./

INDUSTRY

Management. Control by personnel

MANAGEMENT. Industry

Control by personnel

PERSONNEL

Industry. Control of management

Megfigyelhető, hogy az összes jelentés világos és egyértelmű. Az is, hogy ki kontrollál kit. Ti. a management is ellenőrizhetné az alkalmazottakat. Látható a predikátum transzformáció is.

Összetett szerkezetek kezelésére való a S és utána két számjegy. Az első számjegy kifejezi, hogy a szerkezet elemét kívánjuk-e vezetőként, illetve, hogy egybe vagy külön írjuk-e őket. /0=nem vezető, külön irandó. 1=nem vezető, egybe irandó. 2=vezető, külön irandó. 3=vezető, egybe irandó./ A második számjegy a fókuszról való távolságot jelzi. A fókusz a szintagmának az az eleme, amelynek a többi elem determinánsa, PRECIS szóhasználatával élve differenciája, különbsége. A különbség lehet a fókusz differenciája vagy a különbség differenciája. A szerkezetet fordított sorrend-

ben írjuk fel. Pl. a gyorsfagyasztott földi eper összetett szerkezet így alakul:

(1) eper \$21 földi \$21 fagyasztott \$32 gyors
/Minden eleme kerüljön vezetőbe, a földi és a fagyasztott az eperre vonatkozik, a gyors a fagyasztottra. A gyors összeír-
dó a fagyasztottal./

A PRECIS-ben differenciálásnak hívják azt a műveletet, amely az összes többi művelet előtt az összetett szerkezeten elvégzi a rotálást, majd az így létrejött permutációkat egységekként illeszti be az indextételbe.

A PRECIS időtartamot, időt is különbségként kezel, de külön kódja van: \$d. Az idő kifejezhető szóval /\$d középkor/, lehet nyitott /\$d 1945-/. Az idő nem jelenik meg vezetőként.

A \$n illetve \$o is különbség. Zárójeles vagy közbevetett különbségnek nevezzük, tudományos tesztek, mérőskálák, módszerek kifejezésére való. A kód \$n abban az esetben, ha nem kívánjuk vezetőként, egyébként \$o. pl.

(2) vizsgálat \$n IQ teszt

A sub/n↓↑/ helyettesítést tesz lehetővé. Tegyük fel, hogy témánk a K+F tevékenység értékelése. Azt akarjuk, hogy a tétel jelenjék meg a kutatás, a fejlesztés és értékelés alatt is, de ha az értékelés kerül vezető pozícióba, utána K+F álljon. Az indextétel:

(2) kutatás

(2) fejlesztés

sub (2↑) 2 K+F

2 értékelés

A sub(2↑) jelzi, hogy a nyíl irányában látható két kifejezés helyére a K+F irandó akkor, ha az alatta álló lép vezetőbe. A helyettesítés hasonlóan határozható meg lefelé is.

Ha az a cél, hogy bizonyos tárgyszavak ne jelenjenek meg valamelyik láncban, akkor ezt a FN, LN /fölfelé nem, lefelé nem/ jelzi, ha az alatta/fölötte lévő áll vezetőként.

Az \$x, \$y, \$z több téma szétválasztását szolgálja.

Az operátorok és kódok tömör áttekintése után újra folytatni lehet azt a gondolatsort, ami kapcsolatunak mutatta a mélyeset rendszerrel. Nyilvánvaló, hogy \$n és \$o instrumentálist jelöl, az (s) operátor agentivust vagy átélőt /experienecer/. A (t)-nek és (u)-nak ugyanaz a szerepe, mint (2)-nek, generálja az esetrendszert. A függő elemek /(p),(q), (r)/ a főoperátor funkcióját veszik magukra. Ugyanis a főoperátor funkciója vonatkozik az egész függő blokkra.

A mélyeset rendszer és a PRECIS rokonsága riasztóan hathat arra, aki a PRECIS gyakorlati alkalmazására gondol. Ám az osztályozónak, aki a PRECIS-t használja, csak az operátor funkcióját kell ismernie, nem szükséges elmélyednie a mélyesetek tanulmányozásában. Tudnia kell továbbá azt, hogy bizonyos operátor konstellációk hogyan hatnak a bemeneti tételek szerkezetére. Az operátorok eredetét nem kell firtatnia. Az indexelés folyamata úgy zajlik, hogy az indexelő személy megfogalmaz egy természetes nyelvű mondatot, amely az osztályozási egység mondandójának összegzése. A mondat-hoz csak a szótárban /tézauruszban/ felsorolt kifejezéseket

használhatja. Ha ilyen eszköze nincs, mondata állhat szabad szövegszavakból is. A mondatot ezután elemeire bontja. Először a cselekvést, majd ennek tárgyát, majd végrehajtóját határozza meg, tehát a magoperátorokat. Erre a keretre építi a továbbiakat, a kifejezésekhez hozzárendelve a megfelelő operátorokat. Az egész folyamat számítógéppel támogatható.

A PRECIS élő, fejlesztés alatt álló szintaxis, de keretei elég szilárdak ahhoz, hogy gyakorlati bevezetésére gondolni lehessen. Jelenlegi fejlesztése már csak kisebb finomításokra irányul. Viszont igen intenzív a kutatás abban, hogy a PRECIS előnyeit az adatbázisok szervezésében is meg lehessen tartani. Kétségtelen, hogy a tárolási-visszakeresési oldalról a PRECIS kevésbé kidolgozott.

A funkcionáló adatbázisok ugyanis még olyan filozófia alapján születtek - nemcsak a szakirodalmi, hanem más adatbázisok is -, amelyek a Jonker-Salton-Soergel modellre épülnek. Mint láttuk, ebben a modellben az információ rendszerek objektumok és ismérvek halmazából és a közöttük létrehozott leképezésből álltak. /Soergel ezt azzal finomítja, hogy a viszonyokat súlyozza egy "relevancia" mértékkel. Ez kb. azt jelenti, hogy egy ismerv erősebben vagy gyengébben vonatkozik a dokumentumra. [102]/ A modellben tehát dokumentumokhoz az ismérvek egy részhalmaza tartozik, különösebb kapcsolatok nélkül. Vagy fordítva: egy indexelő kifejezéshez a dokumentumok egy részhalmaza. A keresőprofilban a

viSSzakeresés során az ismérvek között a Boole algebra operátorai /és, vagy, negáció/ teremtenek kapcsolatot. Ezek azonban mások, mint a PRECIS relációi, az utóbbiak finomságait nem lehet a szokásos keresőprofilokban érvényesíteni.

Ha pedig a PRECIS relációk nem érvényesíthetők a keresésben, akkor a kurrens kiadványok indexelésének kivételével - a PRECIS-nek magának sincs értelme. Minek az olyan input, amelynek előnyei nem jelentkeznek az outputban? A PRECIS ugyanugy járna, mint a tézaurusz relációk, amelyeket hiába dolgoznak ki körültekintően, ha nincs hozzá az a szintaxis, ami átvinné őket az indextételekbe.

Bár a kutatás ezen a téren a kezdeteknél tart, előzetesen az alábbiak megfontolása szükséges.

A PRECIS indextételekből előállított láncok /mondatok/ tárolhatók, elérésüket a vezető pozícióban álló kifejezés biztosítja.

A keresőprofilok konstruálása során a Boole algebrai operátorok helyett a PRECIS szerepoperátorok adhatják a profil szerkezetét. A logikai operátorokra szükség van ugyan bizonyos részletekben, de nem ők adják a profil vázát.

A PRECIS operátorokkal felírt keresőprofil a mag- illetve főoperátorokra koncentrál, a többi csupán ennek variációiban kap szerepet. A profil írása a (2) operátor, a cselekvés, hatás meghatározásával indul, majd másodikként a/ ha van tárgya, akkor az (1) operátorral jelölt kifejezés határozható meg; b/ ha nincs tárgya, akkor az ágenszt kell leírni. A harmadik lépés a másodiktól függetlenül az ágens le-

írásából áll, ha létezik.

Ez a profil redundáns az indextételek képéhez. A keresés pontosítása további operátorok /része, tulajdonsága, csoportja, stb./ fokozatos beépítésével lehetséges, mindig vigyázva, hogy a profil redundáns volta megmaradjon.

A logikai és, vagy, negáció e profilokban megmarad a következőképpen. Ha a keresésben pl. (1) általános iskola szerepel, kivéve a fővárosiakat, akkor megint (1) általános iskola "de nem" (0) Budapest alak keletkezik. A diszjunkció szerepe a (g) operátorral felsorolt kifejezések összekapcsolását szolgálja, hiszen mellérendelő tárgyszavak felsorolásáról van szó. A (g) ekkor ugyanazt jelöli, mint a "vagy". A konjugálás szerephez jut - példaként - akkor, ha egy operátor ismétlődik, legyen ez a (2) operátor, - tegyük fel, hogy így: (2) oktatás (2) értékelése. A kettő együttes előfordulását tételezzük fel "és" kapcsolattal.

Látható azonban, hogy a profilok elsődleges szerkezetét a PRECIS operátorok határozzák meg, ezt pontosítják és árnyalják a szokásos logikaiak, de az előbbiek meghatározta kereteken belül.

A PRECIS indexelését magában hordozó adatbázis lekérdező nyelve ezért lényegesen bonyolultabb, mint az egyszerűségüknél fogva a felhasználó által is kezelhető jelenlegiek. Lényegesen több /ötszörös/ ugyanis a PRECIS operátorok száma. Valószínű tehát, hogy intelligens profilszerkesztőkre lesz szükség, a lekérdezés pedig kevésbé lesz felhasználó barát.

5.5 MAGYAR ALKALMAZÁS

Az Országos Pedagógiai Információs Rendszer indexelésre a PRECIS-t vezette be. Ehhez a döntéshez azt kellett mérlegelni, hogy az OPIR teljes kiépülésében hálózatot alkot, amelynek tagjai a közoktatás un. háttérintézményei és a husz megyei pedagógiai intézet. Köztük osztott /közös/ feldolgozás valósul meg. Az Országos Pedagógiai Könyvtár és Múzeum szakemberei a PRECIS-t elsajátították két belső, intenzív kurzus révén. De elvárható-e a PRECIS ismerete mondjuk a megyei pedagógiai intézetektől, ahol még a leghagyományosabb könyvtárosi ismeretek sem mindig biztosíthatók. A megyei intézetek információs részlegeiben - idetartoznak könyvtáraik is - nem alakult ki stabil szakértői gárda, a munkatársak pedig gyakran cserélődnek. A PRECIS alkalmazhatósága ezen a kérdésen mulott.

A megoldás a következő. A megyei intézetektől az OPIR nem PRECIS indexelést vár, hanem azt, hogy a feldolgozásra kijelölt dokumentumokról adjanak egy természetes nyelvű mondatot vagy mondatokat. A mondathoz csak az OPIR tárgyszójegyzék kifejezései használhatók.

Ezeket a mondatokat az OPKM munkatársai szedik szét elemeire /tárgyszavakra/, az elemeket ellátják a szükséges operátorokkal és kódokkal. Következő lépésben az így kapott elemek az operátorok sorrendjébe rendezhetők. Ezzel a PRECIS indextétel elkészült. Ezt a folyamatot számítógép támogatja,

az eljárás - tehát a mondat megjelenítése, az operátorok kiosztása, rendezés, stb. - képernyőn zajlik. Az indexelést támogató szoftver Proszéky Gábor munkája.

Ez a megoldás az indexelést annyira megkönnyítő, egyszerű eljárásnak bizonyult, hogy a munkatársak kérésére az OPKM-ben is bevezettük. Az itteni indexelés is egy magyar nyelvű mondat megfogalmazásával indul, amelyből néhány, géppel támogatott lépés után alakul ki a PRECIS indextétel.

E szervezési megfontolás után a PRECIS magyar változatának néhány sajátossága mutatható meg.

A PRECIS eredetileg angol nyelvre készült, de fentebb látható volt, hogy eleve egy "metanyelvi" szintet is tekintetbe vett, mint bármely komoly indexelési/osztályozási eljárás Farradane óta.

A legszembevetőbb különbség az angol és magyar változat között kézenfekvően az, hogy az angol prepozíciók helyét a magyarban ragok vagy más toldalékok foglalják el. Ragozott szerkezeteknek adott esetben a sorrendje is változik. Pl. birtokos jelzős szintagmákban. További feladatként az jelentkezett, hogy a magyarban a toldalék megváltoztathatja a tőalakot is.

A kérdés megoldására elsőként - A PRECIS Manual első kiadása alapján - /ld. 100.sz. jegyzet/ Orbán Éva olyan javaslatot tett, amely erre a célra megőrizte az eredeti \$v és \$w kódokat, de kiegészítette egy < /visszafelé mutató nyíl/, amely ugyanabban a sorban álló tárgyszó toldalékolását jelzi. A "gyermeknek mentése kutyákkal" pl. így:

(1) gyermekek

(2) mentés \$w= e \$v= kal

(3) kutyák

Az ebből keletkező tételek:

GYERMEKEK
Mentés kutyákkal

MENTÉS. Gyermekek
Kutyákkal

KUTYÁK
Gyermekek mentése

A megoldás nem volt kielégítő, pl. nem oldotta meg a bővítettség problémáját, halmozott ragozás esetén nem világos. Több változat után az alábbi megoldás született. A toldalékolást kódolja \$j,. A kód az után a tárgyszó után álljon, amelyik a toldalékot kapja. /Eltérés az angoltól, ahol a kód az alatta/fölötte állónak osztja a prepozíciót, ha a fölötte/alatta álló kerül vezetőbe./ A kódot vesszővel elválasztva sorszámok követik /pl. \$j 1,3,5/, amely azt mondja, hogy hányadik tárgyszó vezető pozíciója esetében kell a ragot a tárgyszóhoz illeszteni. A sorszámokat pontosvessző elválasztó jel követi, majd újabb szám következik, amely arra utasít, hogy a tárgyszóból hány karaktert kell levágni, hogy a toldalék a maradékhoz illeszkedjék. \$j 1,3;1 jelenti pl., hogy az első és a harmadik tárgyszó vezető pozíciója esetén jár a rag úgy, hogy 1 karaktert a tárgyszóból levágunk. Ezt Egyenlőségjel követi, utána a toldalék. A toldalék azonban kiegészül a megváltozott fő végével, tehát azzal, ami a levágott karaktereket helyettesíti. A kutya + kal esetén pl.

a tövégi a megnyulik, á-vá változik. A toldalék tehát -ákkal lesz. Legyen játékos példánk nyelvészeink kedvenc példamondata: "Nádorispán látja Toldit a nagy fával", amelyen a toldalékolást bemutatjuk. PRECIS-ben felírva:

(1) Toldi \$j3,4;0=t

(p) a nagy fa \$j1,3,4;1=ával

(2) látja

(3) nádorispán

/a felíráshoz a (p), "a nagy fa" kíván magyarázatot. Toldinak a nagy fa komplementuma, állapotát mutatja. Így indokolt a (p) operátor./ A felírás szerint "a nagy fa" nem kerül vezetőbe. Toldi tehát t ragot kap a 3. és 4. kifejezés vezető pozíciója esetén, a többől nem kell levágni. A fa az ával toldalékolandó, ha az 1., 3. és 4. kifejezés áll vezetőként, a többől 1 karaktert kell levágni. Az indextételek:

TOLDI

A nagy fával látja nádorispán

LÁTJA. A nagy fával Toldit
Nádorispán

NÁDORISPÁN. Toldit a nagy fával látja

Az utolsó sorhoz magyarázatként: az (1), (2) (3) operátorséma kiváltja a predikátum transzformációt. Egyébként Nádorispán látja a nagy fával Toldit lenne. A predikátum transzformáció megakadályozására adható utasítás - itt nem részletezendő - de Arany János eredeti sorrendje akkor sem alakítható ki. Ugy látszik, hogy a költői nyelv igényeihez a PRECIS még nem eléggé kifinomult. Megbocsájtható hiányosság egy indexelő nyelvénél. A mondatok azonban világos értelműek.

A magyar PRECIS toldalékolása megoldottnak tekinthető. Az az eset is, amikor egy tárgyszó különböző ragokat vesz fel más-más tárgyszó vezető pozíciója esetén. /\$j kétszer, háromszor is kiadható./ Jelenleg egy igazi probléma maradt, amely azonban inkább szépséghibának tekinthető, mert "csunya" bemeneti tételek keletkeznek. Arról van szó, hogy a (3) ágens igen gyakran kívánja meg az "által", "révén" szuffixumot. Ti. valamilyen tevékenység "van végrehajtva" az ágens által. Itt valóban az angol szenvedő szerkezetek hatásáról van szó. "A kocka el van vetve" Caesar által típusu esetek ezek. Az angolban e szerkezetekben rendszerint a by prepozíció áll.

További probléma a birtokos szerkezetek invertált alakjainak felírása. Pl. a munkavédelem ellenőrzése egyszerű, de a rotálás során ellenőrzés, munkavédelem is előáll. Kicsit erőltetett a munkavédelemé alak, vagy munkavédelmet forma.

A fenti két esetet példázza az alábbi. Legyen a téma a szakszervezetek ellenőrzési feladatai a munkavédelemben.

(1) Munkavédelem \$j2;0=é

(2) ellenőrzés \$j1,3;0=e

(3) szakszervezetek \$j1,2;0=által

/a itt spáciumot jelöl/ A bemeneti láncok:

MUNKAVÉDELEM

Ellenőrzése szakszervezetek által

ELLENŐRZÉS. Munkavédelemé

Szakszervezetek által

SZAKSZERVEZETEK. Munkavédelem ellenőrzése

Ha a láncokban a munkavédelem és szakszervezetek nem kap toldalékot, a láncok nem egyértelműek. Azt is jelenthetik, hogy a szakszervezetnél ellenőrzik a munkavédelem betartását.

A láncok esztétikailag kifogásolhatók.

Házi konvenció egyébként, hogy vezető pozícióban nincs toldalékolás.

A PRECIS átalakításának kezdeteinél úgy tűnt, hogy a legtöbb feladat a predikátum transzformációval lesz és a hozzá hasonló un. invertált formával. Mindkettő az indexfogalmak helyes sorrendjének kialakítását szolgálja. Az angol kedvelt szenvedő szerkezeteitől tartottunk, amit az (1)(2)(3) operátorséma is mutat. /Mária van szeretve/szeretett János által./ Csakhogy a rotálási szabály értelmében a láncok visszafelé olvasat szerint alakulnak ki, így a szenvedő szerkezetek az angolban is cselekvővé válnak. /János szereti Máriát./ Megengedhető az (1)(3)(2) séma is, ami megfelel Máriát János szereti mondatnak. Mivel egy un. "üres hely" beiktatásával a program leállítja a predikátum transzformációt vagy az invertálás műveletét, elég rugalmasan lehet a magyar tételek és láncok sorrendjét alakítani. A felvonultatott eszköztár a magyar szórend meghatározásához elegendő, amelyhez az is könynyitésként társul, hogy a magyar nyelv kevésbé érzékeny a szórendre.

Érthető okokból az (s) operátor követi (3) ágenszt, míg az angolban megelőzi. Pl. az "olvasás hatása" a helyes sorrend. Visszafelé olvasva hatása, olvasás lenne, de az (s)

után álló tárgyszó nem kerül vezető pozícióba, a (3) (s) pedig kötötten együtt mozog. Austin megengedi, hogy (s) vezetőbe kerüljön, mondván, hogy valaminek/valakinek a "hatása", "részvétele", "szerepe", "magatartása" fontos lehet pl. egy szociológus számára, aki bárki "részvételét", a "szerep" birtokosait együtt kívánja látni. Ilyen esetekben az (s) az ágens közelebbi szerepét határozza meg, vezetőként pedig mint közvetett, irányított tulajdonság fogható fel. A magyar változat egyelőre nem vezette be. Sőt, megengedi összeírását az ágenssel:

(3) diákok magatartása

A dolgot az is nehezíti, hogy ezek (p) funkciót is betölthetnek:

(1) nők

(p) szerepe

A következőes alkalmazáshoz házi konvenciók szükségesek.

Végül feltétlenül szót érdemel az egyesszám-többszám szabályozása. Az eddigi példákból feltűnhetett, hogy változtatva, elég szabadon jelennek meg mind az egyes, mind a többszámú főnevek. A magyar változat közel áll az angoléhoz. A szabályozás keret jellegű, csupán utmutatásul szolgál. A szabályokat az alábbi táblázat foglalja össze. (ld. 138. oldal)

Tárgyszavak	Példa	Írásmód	Megjegyzés
Megszámlálható fizikai dolgok	kutyák rádió	T	
Nem megszámlálható dolgok, anyagok	épületfa gyémánt	E	De többszám, ha egy osztályhoz tartoznak /pl. közetek/ és a dokumentum ennek speciesével foglalkozik /pl. kvarc
Elvont entitások	hő, idő	E	
Elvont entitások rendszerei	jog, iszlám	E	
Fizikai és elvont entitások rendszerei	általános iskola, egészségügyi szolgálat	E	Ez a típus - (p) operátorral kifejezve - különböző eseteket takar. Pl. (l) iskola (p) tanárok emberi összetevő (l) iskola (p) épületek "hardver" (l) iskola (p) tanterv "szoftver"
Testrészek	tüdő, fogak	E v. T	Ha a szervezetben csak egy fordul elő, akkor E, egyébként T
Egytagu osztályok	Mona Lisa Einstein Egyetemi Könyvtár	E	Az élő egytagu osztályokra a bibliográfiai leírás szabályai érvényesek
Cselekvés, létezés, hatás	osztályozás befolyás	E	Kifejezésük egyesszámu főnevekkel

Ö S S Z E G Z É S

Az értekezés lényegében egyetlen gondolatnak a kifejtése.

Az osztályozó/indexelő eljárásokat két paradigma alapján hívták életre. Az első a hagyományos, logikai értelemben vett osztályozás volt, azaz az osztálybasorolás gondolata. A hozzá kapcsolódó eljárás abban állott, hogy az individuumokat magasabb osztály/ok/ba sorolta. Kérdése az volt, hogy mi mibe tartozik bele.

A második paradigma új eszme jegyében fogant: az osztályozó/indexelő eljárásoknak fel kell mutatnia azokat a jegyeket, ismérveket, amelyek az osztályzandó objektumról állíthatók, leírva azok egyéni arculatát. Erre a célra a tudományrendszerekből származó osztályozási rendszerek alkalmazatlannak bizonyultak, mert a továbbiakban nem az volt a kérdés, hogy az osztályzandó egységek, pl. dokumentumok mibe tartoznak bele, hanem hogy milyen ismérvek hordozói, mi állítható róluk. Erre a célra alakultak ki az információkereső nyelvek, őket megelőzve már a mellérendelő osztályozások. Ennek elméletében /Jonker, Salton, Soergel/ objektumok és ismérvek halmazát, valamint a leképezésüket szolgáló viszonyrendszert tartották modellértékűnek. Eszerint a dokumentumokhoz rendelhetők az ismérvek részhalmaid és viszont, az ismérvekhez a dokumentumok részhalmaid. Az ismérvek részhalma azonban csak az indexelő elemek valamilyen összessége,

semmi több. Még nem szerveződnek össze egésszé. A részek összessége még nem az egész.

A második paradigma gondolatában fogant indexelő rendszerek között a legfejlettek a klaszteranalízis módszereit alkalmazó automatikus osztályozás. Egy ideális rendszer minden jótulajdonsága elmondható róla: nem preferál szubjektív szempontokat, képes kimerítő osztályozásra, a természetes nyelv szövegszavaira és szintagmáira építhet, érzékeny a felhasználóra, automatikusan kezelhető, stb. Egy fontos tulajdonság azonban hiányzik belőle. Pontosan az, ami az osztályozó elemeket egy indextételben egységes arculatúvá, egésszé fogja össze. A szintaxis.

Előnyük még - különösen a tézauruszoké -, hogy a lexikai egységek közötti relációkat elemértékűnek tekintik. De az osztályozási relációkat szótári és szemantikai problémaként kezelik. Egy szótárba, vagy tézauruszba a relációkból az építhető be, ami a fogalmakról a priori megállapítható. E relációk ezért túl általánosak, csak az indexelőt támogatják, és nem mutatják azokat a fogalmi kapcsolatokat, amelyek az indexelendő szövegben fennállnak.

Szükség van tehát egy harmadik paradigmára, harmadik elméletre. Ennek három követelménye van.

Az első az az eljárás, amelyet Gardin a bronzkori archeológiában alkalmazott. Szempontunkból ennek egyik fő problémáját az okozta, hogy a leletekhez megállapított jegek, pl. ornamentikai elemek összessége sem nyújtja a le-

let összképét. Meg kellett oldania azt, hogy az egyedi ismervek hogyan állhatnak össze kompozícióvá. Erre a célra egy nyelv szolgált, amely az individuális jegyeket a természetes nyelvből származtatott esetrendszerrel kapcsolta össze. Az osztályozásban hasonló a probléma: az osztályozási egységeket magasabb egységekké kell szervezni. Ezt nem képes megoldani a második paradigma jegyében született információkereső nyelvek egyike sem. Hiszen relációik szemantikaiak, holott szintaktikaiakra van szükség.

A második követelmény egy Landrytől származó posztulátum. Ez az alapelv azt fogalmazza meg, hogy a fogalmak között a szerző által létrehozott asszociációt nem szabad elveszíteni az indexelés során, hanem át kell menteni. Ez nevezhető relációmegőrző transzformációnak.

A harmadik követelmény az első kettőből ered. Az osztályozandó szövegegységek individuális arculatának felmutatása, továbbá a relációmegőrző transzformáció akkor érhető el, ha a lexikai egységek és a relációk is a természetes nyelvből származnak. Ezt a témát tárgyalja a specifikusságról szóló rész és az egész 5. fejezet. Az erre a célra alkalmas relációk a Farradane relacionális indexelése - SYNTOL - PRECIS fejlődési sort írják le. Az előbbi kettő relációi "metanyelvek", a PRECIS pedig a transzformációs-generatív grammatika Fillmore-tól származó mélyeset rendszerét adaptálja olyan szerepoperátorok meghatározására, amelyek az osztályozási kifejezéseket mondatokká generálják.

A mondatok az osztályozás/indexelés olyan egységei, amelyek nemcsak ismérvek, tematikai jegyek, predikátumok felsorolását eredményezik, hanem alkalmasak az eredeti mű gondolatának visszaadására, tömörítésére, benne jelenhetnek meg eszmék, hipotézisek, elméletek - mindaz, amiből a tudomány valósága felépül.

Mivel az Országos Pedagógiai Információs Rendszer magyar viszonyokra alkalmazta a PRECIS-t, kíváncsú volt ennek részletesebb ismertetése. A PRECIS magyar nyelvhez igazított változatának sajátosságait, megoldásait, eltéréseit angol mintájától az 5.5 fejezet írja le.

Még nem ismeretes a PRECIS-nek viselkedése visszakereső rendszerekben, mert folyamatos tájékoztatásban alkalmazták. Ha nem sikerül adatbázisokhoz kapcsolni, elveszti értelmét. Az értekezés felvázol egy új keresési stratégiát, amely nem a Boole algebra operátoraira, hanem a PRECIS magoperátoraira épül. A keresőprofilok élesíthetők a függő operátorok fokozatos beiktatásával. A logikai és, vagy és negáció e kereten belül él bizonyos szerepoperátor-konstellációkban.

A fentebb vázoltak tehát egyetlen gondolatot hordoznak: hogyan merültek fel a klasszikus tudományrendszerek alkalmazásától kezdődően azok a szükségletek, amelyek újabb és újabb eljárások megszületését eredményezték, míg végül mind az elmélet, mind az osztályozási gyakorlat számos vargabetű bejárása után eljutott az emberi gondolatközlés legősbibbi, legegyszerűbb, pótolhatatlan eszközhöz, a nyelvhez. De

nem szótárakban és szemantikai viszonyokban gondolkodva, hanem szintaxisban. A harmadik paradigma jegyében az osztályozás új korszaka kezdődött.

Az individualizáló osztályozás parancsoló előretörése mellett megmaradtak a hagyományos feladatok is. De a gyakorlati céluak a klasszikus, felosztás-elvű tudományrendszerektől elszakadtak. A tudományrendszerek, bár egyre kevésbé alkalmasak gyakorlati osztályozásra, a filozófia keretében nyerik el értelmüket, ahol mindig is lesz létjogosultságuk, mert világképet pótolnak.

A gyakorlati célokat szolgáló felosztás-elvű tudományrendszerek is szerepet kapnak, mivel az osztályozási funkciók két részre oszlottak. Az egyénítő osztályozás mellett szükségesek az átfogó besorolást biztosító rendszerek is. Az értekezés kritikailag elemzi a rubrikátort, a BSO-t és Dahlberg ontikus rendszerét. Konklúziója az, hogy ezek óhatatlanul és feleslegesen még mindig a filozófiai rendszerek köldökszinórján át táplálkoznak, amit praktikus céljuk, az információ rendszerek közti csere és kommunikáció nem indokol. Az értekezés egy alternatívát ír le, amelynek lényege egy, a genusok síkján létrehozott konkordancia, különbözőbb strukturálás nélkül.

Az értekezés vezérgondolatához számos további részlet csatlakozik.

Köztük említést érdemel az osztályozási egységekről szóló rész. A mai gyakorlat szerint a dokumentumok formál-

ják az osztályozás egységeit, analitikus feldolgozásban egy közlemény, vagy tanulmány képez egységet. Az eljárás az információ esetleges kvantálását eredményezi. Zajossá teszi a tárolási rendszereket, hiszen a tárolási egységként kijelölt információ kvantum önkényes és szubjektív. Az értekezés - elindulva Kunszt György kognitívum fogalmából - az információ természetesebb egységeit eredményező kvantálást javasol. Meghatározására a klaszteranalízis lehetséges eljárásainak egyikét javasolja. Egy olyan gráfból, amelynek csucsain termékek, élein a termékek együttes előfordulásának súlyozása állnak, a gyenge súlyozású élet mentén kivághatók azok a részgráfok, amelyek egy kognitívumot reprezentálnak. A módszer a mátrixok parciálásán nyugszik. Az eljárást kis példa mutatja.

A facettás osztályozásról kimutatja, hogy azok egyfelől a klasszikus megoldások összegzései, feloldván azok merevségét. Annyiban összegzés, hogy Arisztotelesztől eredő módszere - a létezőknek a magasabb genusok alá vonása - rokonítja őket a felosztás elvű rendszerekhez. Másfelől a facettákból erednek a mai osztályozási rendszerek értékelhető relációi, így Gardin-é, Farradane-é is. A facetták számának növelése - hasonlóan a lehetséges felosztási alapok számának növeléséhez - átmenetet biztosít az automatikus osztályozás objektum-ismérv mátrixához, amely modell az utóbbi induló eszköze. Ez a fejezet foglalkozik Bartók Béla osztályozási elképzeléseivel is, megállapítván, hogy a népdaltípológiára vonatkozó törekvései az osztályozáselmélet klasszikusai közt jelöli ki helyét. Bartók facettás típusu megoldásait o-

lyan igénnyel dolgozta ki, amely őt - nem a megoldásokban, hanem a probléma felvetésében - az automatikus osztályozás előfutárai közé emeli.

Az értekezés felveti a heurisztikus információ problémáját. Ez abban áll, hogy az információnak nemcsak tematikus keresését kell biztosítani, hanem - kutatási célokat szolgáló rendszerekben - a problémamegoldás logikai menete szerintit is. A kutató ismeri témáját, információt a problémamegoldás lépéseinek hiányzó láncszeméhez kíván. Ehhez a különböző diszciplínák törvényei, módszerei, elméletei között lévő analógiák felderítése szükséges. Ez a kérdés az osztályozási kutatások új területe és tudományközi jellegű. A megoldásokat a tudományos törvények izomorfijára vonatkozó rendszerelméleti kutatások eredményezhetik Bertalanffy hipotézise szerint.

A dolgozat egyik kitérőnek tűnhető fejezete /2.2/ foglalkozik azzal a folyamattal, amely az élő tudományokat átstrukturálja. A folyamatot egy példa alapján - javaslatként - Mandelbrot hatásnak nevezi. A hagyományban a tudományok szerkezete úgy alakult, a diszciplínák úgy formálódtak, hogy tárgyük a közvetlenül tapasztalható való világ valamely része, szférája volt. Eszerint is, rendszerint tárgyük alapján definiálták magukat. A filozófia mellett - melynek tárgya a lét egésze - kivétel a matematika volt. A matematika tárgya az egész valóság, de csak egy szempontból: a mennyiségi viszonyok és térformák oldaláról. Mai

megfogalmazásban a relációk tudománya. Az elmúlt évtizedekben számos olyan új tudomány született, amelynek természete hasonló a matematikáéhoz, ti. tárgyuk nem a közvetlenül tapasztalható való világ valamely része, hanem a tapasztalható világ valamely absztrakciója. Ilyen tudomány a kibernetika, a rendszerekkel foglalkozó tudományok, szemiotika, az információval kapcsolatos tudományok. Ezek az emberi gondolkodást magasabb absztrakciós szintre emelve átfogják az összes többi, hagyományosan kialakult tudományt, mintegy "felülről", egy magasabb absztrakciós szintről hatolva a tudományok hagyományos rendjébe. Amit általában a tudományok integrálódásának neveznek, az csak részben tulajdonítható a klasszikus tudományok határai elmosódásának és a határterületi kutatások megjelenésének. Inkább köszönhető a fenti fiatal tudományok átrendező hatásának. Ezen a szinten értelmetlen természettudományokról, humaniorákról, társadalomtudományokról, méginkább köztük feszülő szakadékról szólni. Ha létezhet egyszer a tudományoknak valamilyen szintézise, akkor ez az újonnan megjelent, magas absztrakcióju diszciplínák bázisán lehetséges. E fejezet megállapításai erősen hipotetikusak. Felvetései azonban indokoltak annak megértéséhez, amit komplex témaként vagy sokparadigmájú problémaként szokás az osztályozáselméletben aposztrofálni. Az értekezés a "hagyomány" és a "jelen" között húzódó 3. fejezete sok tekintetben ezeket a megállapításokat használja.

Ez a 3. fejezet az osztályozás alapkérdéseit tárgyalja.

Köztük néhányat már az összegzés is felvetett. A dolgozat abban foglal állást, hogy - mellőzve az osztályozás technikai jellegű kérdéseit - az elmélet alapproblémái között az alábbiak a legfontosabbak.

- A kvantálás kérdésköre.

- A heurisztikus információ, mint új kutatási terület.

- Az osztályozás "követő" volta, illetve a precedens nélküli osztályozás problémái. Létezik ugyanis egy paradoxon. A kutatási célokra szervezett információ rendszerekben az indexelés során új eredményeket kell osztályozni, holott az erre alkalmas kifejezések még nem állnak rendelkezésre. Az élő kutatás terminus technikusai megelőzik az indexelő fogalmak meghatározását. Az indexelés visszahuzó lehet a tudományok fejlődésére azzal, hogy meglévő kifejezéseivel igazítja precedens nélküli tudományos eredményekét. Ez a probléma vonja maga után a szövegszavas osztályozás kikerülhetetlenségét.

- A specifikusság kérdése. Az információkereső nyelvek egyénítő indexelési eljárásainak következtében felmerült probléma, amely szintén szövegszavas indexeléssel - vagy szövegszavakon nyugvó tárgyszavak, tézauruszok, szótárak kimunkálásával - oldható meg.

- Relációmegőrző transzformáció.

A dolgozat szelleméből következik, hogy nem állítja szembe a felosztás-elvű, facettás, automatikus, vagy szintaktikára épülő osztályozási rendszereket. Ellenkezőleg: i-

gyekszi kimutatni az érintkező felületeket és azt, hogy a hagyományból a maradandónak tekinthető gondolat hogyan él tovább. Az osztályozó/indexelő eljárások generációváltásait a tudományokban lezajló változások és a technikai feltételek fejlődése váltja ki.

Az osztályozáselmélet mai státuszát az jellemzi, hogy fegyvertárában minden készen áll arra, hogy egy hosszú időre maradandó teóriát produkáljon, amelyben a korszerű osztályozó/indexelő rendszerek változatai formálódhatnak, szolgálva a gyakorlatot.

Ebben a fegyvertárban egyfelől ott található az automatikus osztályozás. Egy kivétellel az elmélet valamennyi problémájára megoldást kínál: kimerítő osztályozás, specifikusság, a természetes nyelvből származó szövegszavak és szintagmák meghatározása és kezelése, az osztályozandó egységek természetes csoportjainak /klaszterek/ kialakítása ismervek preferálása nélkül, kognitívumok meghatározása, a teljes szövegfeldolgozás elterjedésekor /sőt, igazából akkor/ alkalmazható, így jövőbe tekintő, a folyamatok automatizálása.

A fegyvertárban másfelől ott állnak a még éppen lábra álló szintaxisok. Ezek pontosan annak a követelménynek felelnek meg, ami az előzőből hiányzik: a részeket egységgé szintetizálja, a kifejezésekből mondatokat generál. Szempontok, ismervek helyett a gondolat image-ét nyújtja.

A kettőt az köti össze, hogy mindegyikük a természetes nyelvből indul. Az egyik szótárának tökéletesítéséért, a má-

sik a kapcsolatteremtő viszonyrendszerért. Eljárásaik, folyamataik révén nemcsak az érhető el, ami mindig is kíváncsi volt, hogy ti. az osztályozás/indexelés tárgya ne a dokumentum legyen, hanem maga a szöveg. Hanem az indexelést és osztályozást felhasználó barátta teszi. Mert, formalizálva ugyan, de eszköze ugyanaz, mint amiben a kutató, bármely felhasználó gondolkodik: az emberi nyelv.

A fegyvertár egyelőre még két önálló részének összekapcsolása: ez az osztályozáselmélet mai feladata.

J E G Y Z E T E K

1. Legújabb pl. Dunaev, V.V. - Polâkov, O.M.: Metodologičeskie aspekty relâcionnoj teorii klassifikacii. =Naučnaâ Tehničeskââ Informaciâ, Seriâ 2. 1987. 4.sz. 21. l.⁴⁷
2. A szakirodalom jelenlegi szóhasználatában ez a megkülönböztetés elmosódott: "osztályozás"-sal jelölnek felosztást is, kb. abból a megfontolásból, hogy a felosztás eredménye azok az osztályok, amelyekbe az egyedek besorolhatók.
3. Ezen a tényen az sem változtat, hogy a gyakorlatban már korábban számos mellérendelő, tárgyszavas osztályozás alakult ki. A LC subject heading-jei /tárgyszavai/ pl. elég átfogó osztályokat jelölnek. Így ez a fajta indexelés is "alásorolás"-t tesz lehetővé. A radikális elszakadás egyébként Luhn, H.P. KWIC koncepciójának megjelenésére datálható /1954/. Az indexelésnek új közelítését eredményezte. Mai szemmel a KWIC koncepcióban Luhn eredményei nem haladják meg a hagyományos osztályozás teljesítményeit a felértékeltség terén. Viszont automatizálja az indexelést. A KWIC megjelenésének máig ható heurisztikus jelentősége van. Luhn szövegelemző módszerei mindmáig a legkorszerűbbek.
4. Elsősorban az átfogó osztályozási rendszerekben, mint a BSO vagy a rubrikátor.
5. Az osztályozási rendszereknek olyan felosztása, amelyben az egyik típust a "szemantikai" képviseli, az NTMIR-től, nálunk Vajda Erikától ered. Egy oktatási segédletben népszerűsítette. /Utmutató a Bevezetés az osztályozási ismeretek című tantárgy tanuláshoz. Bp. Házi soksz. 1972./ Innen vették mások, jobbra zavarosan. Az elnevezés azért nem szerencsés, mert "szemantikai" minden valamilyen szintig szabályozott információkereső nyelv. Tehát a szövegszavas módszerek kivételével mindegyik. Más kérdés, hogy a szemantikai egyértelműség legigényesebben a tézauruszokban valósul meg.

+/ Az eredeti orosz nyelvű irodalom átírása a szabványnak felel meg. Magyar fordítások esetén a szöveghűség követelményének megfelelően a nevek alakjára a műben található változat mérvadó.

6. A tézauruszok elméletéből ez a hátrány még nem következik. A problémát talán az okozta, hogy a tézauruszok nemzetközi rendszerekben korporálódtak, túl korán szabványosították őket. Így a tézaurusz úgy részben kikerült a tudomány hatásköréből és részben elbürokratizálódott.
 7. Borko, H.: Toward a theory of indexing. 356. l. =Information Processing and Management. Vol. 13. 1977. 355-365. l.
 8. Ezt a hagyományt a teremtesmitoszok őrzik. Itt nincs szükség részletezésükre. A görög mitológiákhoz és analógiáihoz ld. Graves, R.: Görög mitoszok. 1-2. köt. Bp. Európa, 1981.
 9. Ilyen összekötő pl. Hermész, az istenek küldötte. /V.ö. Kerényi K.: Hermész, a lélekvezető. Az élet férfi eredetének mitológemája. Bp. Európa, 1984.
Az összeköttetést biztosíthatja egy szent hegy, oszlop, létra, égigérő fa, stb.
 10. A kozmosznak bonyolult szféráit tudományos igénnyel elsőként Ptolemaiosz /Klaudiosz/ rendszerezi a bolygók alapján emberközpontu kozmogóniájában, még őrizve a mitológiai gyökereket. L. Szabó Á. - Kádár Z.: Antik természettudomány. Bp. Gondolat, 1984. Szabó Árpád itt közli Ptolemaiosz tudományfelosztását is. /49. l./ A mitoszok és tudomány kapcsolatát újszerűen tárgyalja Kákossy László a gnosztikus kódexek alapján. /Fény és káosz. Bp. Gondolat, 1984./
 11. Eszmetörténeti szempontból azzal az érdekes jelenséggel találkozunk, hogy a kezdeti gondolkodásban megjelenik az "egész" fogalma, ami a világegyetem. Számos megfontolásból az is lehetséges lett volna, hogy az akkori élet színterének, egy szűk oikumené-nek részleteiből az "egész" később alakul ki.
A "benépesített" világok kapcsán nem az a kérdés, hogy ezek egy "hamis" tudat termékei-e és ezért a gondolkodásnak ez a szintje "primitív". Hanem az, amit Kerényi írt Hermész kapcsán: "Mi jelent meg a görögöknek Hermészként?" Illetve: "Hogyan jelenhetett meg éppen ez istenként a görögöknek?" /A szerző kiemelése./ Kerényi: i.m. 5-6. l.
- Végül a világmindenségnek a mitológiákban és vallásokban adott felosztása az etnológia vonatkozó szakirodalmának osztályozásához ma is alapot nyújt. Más vonatkozásban is élő hagyományról van szó. Az akkor kialakult szimbólumokat és metaforákat őrzi a néphagyomány, fellelhető a mai gondolkodás szimbólumaiban, megjelennek a művészetek formanyelvében.

12. Arisztotelész: Metafizika, III.

13. Arisztotelész: Met. VI. 1. "Mindezek a tudományok egy bizonyos létezőről és valami körülírt létfajtaról szólnak..." (VI. 1. 1025 b)
A későbbi korszakban tapasztalható merevség azonban itt még nem lelhető fel. "De nem is valószínű, hogy valamennyi szubsztancia egy tudomány körébe tartozzék." (III. 2. 997 a) Arisztotelésznél a tudomány "tárgya" a cselekvés, az alkotás és a tiszta elmélet. Ennek megfelelően három tudomány csoport van. "Az "elmélet" újabb három részre oszlik. (VI. 1. 1026 b)
Az idézetek Halasy-Nagy József fordításából valók. (Arisztotelész: Metafizika. Bp. Dunántul Pécsi Egyetemi Könyvkiadó, 1936.)
14. Ennek a módszernek lenyűgöző eredménye Linné rendszere. Bernal professzor a legnagyobb tudományos teljesítmények egyikének tartja. (Bernal, J.D.: Tudomány és történelem. 421-422. 1. Bp. Gondolat, 1963.)
15. A felosztást a "meghatározás"-on belül tárgyalja, mivel a meghatározás genus alá rendelést jelent, a legmagasabb genus pedig a szubsztancia.
16. Verancsics Faustus: Logica Nova. =Verancsics Faustus Machinae novae és más művei. 201. 1. Bp. Magvető, 1985.
A szerző gunyos indulatát Arisztotelésszel szemben az magyarázza, hogy az erődődő ujplatonizmus szemben állt Arisztotelésszel, akiben a középkori skolasztika látta eszményképét.
17. Ashby, W.R.: Bevezetés a kibernetikába. 137-140. 1. Bp. Akad. K. 1972.
Szovjet kutatók erre szemiotikai magyarázatot adnak. Eszerint a vizsgálat tárgyáról szerzett minden ismeret valamilyen nyelvben, egy jelrendszerben képződik le, több fokozatban, ún. szemiotikai sikokban. A jelrendszerek, így a nyelv is mindig tartalmaz olyan összetevőket, amelyek a nyelvi strukturának következményei és nem a tárgyból származnak. A megfigyelő nincs abban a helyzetben, hogy eldöntse, a jelrendszerben mi származik magából a nyelvi konstrukcióból és mi a tárgyból. Hosszasan elemzi ezt Sztepanov, Ju.Sz.: Szemiotika. 134-161. 1. Bp. Akad.K. 1976.

18. Bacon, F.: Proficiency and advancement of learning. =The works of lord Bacon. Vol. 1. 1-80. 1. London. Reeves and Turner, 1879.

19. Šamurin kitűnő könyvében írja (Šamurin, E.I.: Očerki po istorii bibliotečno-bibliografičeskoj klassifikacii. Tom. I. 135. 1. Moskva. 1955.), hogy harminc évvel Bacon előtt Juan Huarte spanyol orvos és filozófus összesen huszonhét tudományt három fő csoportba sorolt, az emlékezet, értelem és képzelet tudományaiba. Bacon csupán az utóbbi kettő sorrendjét cserélte fel.
Bacon rendszerére épül Diderot felosztása, aki a sorrendet visszacserélte, a rendszert pedig részletesen kibontotta. Bacon amerikai csodálója, W.T. Harris alkalmazta ezt a felosztást könyvtári célokra, sorrendjében ismét Baconhoz igazodva (1870.). Dewey tőle vette át Bacon elveit 1876-ban.

20. Hobbes, Th.: Leviathan. 74. 1. Ford. Vámosi Pál. Bp. Magyar Helikon, 1970.

21. Bacon, F.: i.m. 33. 1. Az idézet pontosan így szól: "Azonban az ismeretek felosztása és elkülönítése nem egy csucsból összefutó egyenesekhez hasonlít, amelyek egy pontban találkoznak, hanem hasonlatosak egy fa ágaihoz, amelyek egy olyan törzsben találkoznak, ami az egésznek és a folytonosságnak dimenzióival és nagyságával rendelkezik, mielőtt még megszűnne ez a folytonosság és ágakra, gallyakra törne szét." Bacon gondolata azért fontos, mert tanúsítja, hogy a hierarchiát nem konstrukciós eszköznek vagy célnak tekintette, hanem a tudomány egységes szemléletének eszközét látta benne.

22. Ampère felosztásában léteznek elemi és magasabbrendű tudományok. Az elemiek adódnak a közvetlen megfigyelésből: autoptikai (autoptique) tudományok; adódhatnak annak felderítéséből, hogy miből áll a tárgy: kriptorisztikus (cryptoristique) tudományok. A magasabbrendűekhez a tárgy változásait kutató tudományok tartoznak: troponomikus (troponomique) tudományok és kriptologikus (cryptologique) tudományok. Az utóbbi az okság szempontjából vizsgálódik. Az így kapott négy tudomány ismét ketté osztható /nyolc tudomány/, majd folytatva dichotomiát egészen 128 tudományig jut el. (Ampère, A.M.: Essai sur la philosophie des sciences, ou exposition analytique d'une classification naturelle de toutes les connaissances humaines. 2. partie. Paris, Bachelier, 1843. Összegző táblázata a 181. lapon.)

23. Rickert, H.: Kulturtudomány és természettudomány. Ford. Posch Árpád. 19., 46. l. Bp. Franklin, 1923.
Ugyanezt az elvet alkalmazta a marxista filozófia alapján egy jobb sorsra érdemes kísérlet. (Nyakó István: Részletek Nyakó István Tudományrendszer c. kéziratából. Bp. OSZK-KMK, 1962.)
24. Kedrov, B.M.: Klassifikaciâ nauk. 1-3. köt. Moskva, Izd. Akad. 1961-1965.
Kedrov szélsőségesen hierarchikus gondolkodású. A tudományok hierarchikus felosztását tartja dialektikusnak, szemben a szerinte metafizikus mellérendelő rendszerekkel. Példázza, hogy mennyi önkény mutatkozott abban, ahogyan a metafizikus vagy dialektikus jelzőt valamire vagy valakire rácímkézték. Farkas János bírálja Kedrovot, konklúziója pedig: "Kedrov kompromittálja a mozgásformákat." (ld. 25. jegyzet, 46. l.) A mai, főleg protestáns teológia kimutatta, hogy a hierarchikus gondolkodás ószövetségi eredetű, ennek világképe a lényeket, dolgokat, szigorú, determinisztikus világképbe rendezi, olyan hierarchiába, amelynek csúcán Jahve áll. Az érdekesség kedvéért lehet idézni nemrég elhunyt pályatársunkat, aki Kedrovval ellentétben a hierarchikus rendszereket tartotta középkori örökségnek, feudálisnak szemben a demokratikus mellérendelő rendszerekkel. Ervelését szintén a marxista filozófiára alapozta. A hierarchikus gondolkodást szintén a középkor jellegzetességének tartja Bernal is. (I.m. 207. l.)
25. Farkas J.: A modern tudomány szerkezete. Bp. Akad. K. 1981.
26. A tudomány néhány elméleti kérdése. Szerk. Bóna E., Farkas J. Bp. Akad. K. 1970.
Négy felosztási alapjuk, vagy ahogy ők nevezik, "metszetük": ontológiai, gnoszeológiai, módszertani és funkcionális. Az ontológiai nagyjából a tárgyi felosztásnak felel meg, a gnoszeológiai a lényegbehatolás fokozatait jelenti.
27. Bóna E.: A kémiai tudományok és kutatási ágak rendszerezési kérdései. Bp. Akad. K. 1971.
28. Horváth T.: A bibliográfiák funkciójáról. 37. l. =Bibliográfiai tanulmányok. Bp. OSZK-KMK, 1978. 11-57. l.
29. Jonker, F.: Indexing theory, indexing methods and search devices. New-York, London. Scarecrow Press, 1964.

30. Cleverdon, C.W.: ASLIB Cranfield research project: report on the testing and analysis of an investigation into the comparative efficiency of indexing systems. Cranfield. College of Aeronautics, 1963.
Magyar összefoglalása: Horváth T.: A második cranfield-i jelentés. =Könyvtári Figyelő, 1968. 5. sz. 1-17. l.
31. Horváth T.: i.m., uott.
32. Jaspers, K.: Bevezetés a filozófiába. 87. l. Bp. Európa, 1987.
33. Bertalanffy, L.: General system theory. A new approach to unity of science. =Human Biology, Vol. 23. 1951. dec. 303-361. l.
Részletek megjelentek a Rendszerelmélet. Válogatott tanulmányok. Bp. Közgazd. és Jogi K., 1969. c. kötetben. A tudományoknak erről a lehetséges közelítéséről kritikus hangvétellel ír Blaumberg, I.V. - Szadovszkij, V.N. - Judin, E.G.: A rendszerkutatás és az általános rendszerelmélet. =Rendszerkutatás. Válogatott tanulmányok. 53-84. l. Bp. Közgazd. és Jogi K. 1973.
34. Cherry, C.: On human communication. Cambridge, (Mass.) MIT Press, 1959.
A kommunikáció és információelmélet kapcsolatára McKay, D.M.: Information, mechanism and meaning. Cambridge, (Mass.) -London, The MIT Press, 1969.
Igen értékesek gondolatai az információ kvantálásáról. A kommunikációelmélet vezető szerzőitől igen jó válogatás: Kommunikáció. Válogatott tanulmányok. 1-2. köt. Összeáll. és szerk. Horányi Özséb. Bp. Közgazd. és Jogi K. 1978.
35. Fülöp G.: Ember és információ. 5. l. Bp. Muzsák, 1983.
36. A szemiotika önmagát Arisztotelésztől származtatja. Ez annyiban jogosult, hogy a nyelv és jelentés problémájának előzményeit hozzá lehet visszavezetni. /Organon, I./
Valódi megalapítója Ferdinand de Saussure, akinél szemiológia néven szerepel. (Bevezetés a nyelvészetbe. 33. l. Bp. Gondolat, 1967.)
Kitűnő összegző monográfia: Sztepanov, Ju.Sz.: i.m.
A klasszikusok írásaiból nyújt válogatást A jel tudománya. Vál. és bev. Horányi Özséb, Szépe György. Bp. Gondolat, 1975.

37. Egyed L.: Szabályszerűség és változatosság. =Tudomány. A Scientific American magyar kiadása. 1978. 4. sz. 41-43. 1.
38. Jellemző azonban Mandelbrot magatartására, hogy mintegy mentegetőzik. Nyilatkozza, hogy amikor könyvet írt, különös gonddal igyekezett az olvasót meggyőzni arról, hogy nem felületes szerzővel van dolga, aki ennyi mindenhez ért.
39. Németh László: "De ahogy történelmi tanulmányaimban előbbre jutottam, azt láttam, hogy ez a szakadék nemcsak én-bennem nincs meg, de nincs meg objektíve sem, az ujkorban legalább, a természet- és szellemtudományok közt. ... Hol van hát, hogy nyílt meg ez a szakadék, amelyről a felszólalók panaszkodtak? Nyilván az emberek képzettségében, s a pedagógia tehetetlensége, ami nyitva tartja." Németh L.: Levél Marx Györgyhöz. =Megmentett gondolatok. 348-349. 1. Bp. Magvető és Szépirodalmi K., 1975.
40. Az értekezés nem szándékozik szembeállítani a kutatások különböző vonulatait. Ezek termékenyítő együttélését kívánja. Elvileg is felmerül egy probléma a magas absztrakcióval kapcsolatban. Ennek tárgyalása itt nem feladat, de röviden csak annyit, hogy magas absztrakciónak van olyan velejárója, hogy nehéz visszajutni a valóság közvetlenebb szféráihoz. Az absztrakció nagyfoku tömörítéssel jár, ennek pedig van olyan foka, ahonnan már nem lehet "dekódolni" pl. az empirikus ismereteket. Wigner Jenő erről: "Nyilvánvaló, hogy van egy határ, amelyen túl a tömörítés - bár önmagában cél lehet - nem hasznos az információ szempontjából." Wigner J.: A tudomány határai = Szimmetriák és reflexiók. - - tudományos esszéi. 206. 1. Bp. Gondolat, 1972.
41. Arisztotelész: Metafizika V.
42. Organon I,3.
43. Uott. I,4.
44. Ranganathan, S.R. - Gopinath, M.A.: Prolegomena to library classification. 3rd ed. Bombay, Asia Publ. House, 1967.

45. Shera, J.H.: Libraries and the organization of knowledge. 101. l. London, Crosby Lockwood and Son, 1965.
46. Vickery, B.C.: On retrieval system theory. London, Butterworths, 1961.
A témával kapcsolatos, az előbbi részben átfedő munkája: Classification and indexing in science and technology. London, Butterworths, 1974. Ebben a művében (146-156. l.) a talajtan területére is kidolgozott egy facettás rendszert. Ezek a facetták: Talajfajták
Talajszerkezetek
Alkotórészek
Tulajdonságok
Természeti folyamatok
Talajműveletek
Laboratóriumi technika
Általános kifejezések
47. Ismerteti Vickery, B.C.: The structure of "semantic coding." A review. =American Documentation, vol. 10. 1959. 234-241. l.
Lásd még: Grolier, E.: A study of general categories applicable to classification and coding in documentation. 117-119. l. Paris, Unesco, 1962.
48. Farradane, J.E.L.: A scientific theory of classification and indexing. =Journal of Documentation, Vol. 6. 1950. No. 1. 83-99. l.
- -: A scientific theory of classification and indexing. =Uott. vol. 8. 1952. No 2. 73-92. l.
- -: Psychology of classification. =Uott. Vol. 11. 1955. No. 4. 187-201. l.
Igen jó összefoglaló található Foskett, A.C.: The subject approach to information. 88-91. l. London, Clive Bingley, 1982. és Coates, E.J.: Subject catalogues: headings and structure. 45-49. l. London, Library Ass., 1960. munkájában.
49. Farradane gondolataihoz hasonló alapról tárgyalja a fogalmi kapcsolatokat Scsedrovickij, G.P., bár elemzése nem osztályozáselméleti szempontú, hanem logikai-rendszerelméleti. (A rendszerkutatás módszertani problémái. =Rendszerkutatás, i.m. 85-155. l.) Scsedrovickij a fogalmi kapcsolatokat úgy osztályozza, ahogyan az a tudományos kutatás folyamatában feltárul. Először csak korreláció áll fenn. /Megfelel Farradane együttes előfordulásának./ Ezután ún. "kapcsolat" jön létre, másszóval kölcsönös függőség, amely a fogalom "magyarázatá"-val jár együtt. Végül kialakul a "viszony". Ez a legszorosabb, és olyan kapcsolatnak definiálja, amelyre érvényes a tranzitivitás. Eredményeit A.A. Zinovjev filozófiai munkáira hivatkozva közli.

50. Keen, E.M. - Digger, J.: Report of an information science index test. Aberystwyth, College of Librarianship Wales, 1972.
Ennek magyar összefoglalása és ismertetése: Horváth T.: Az Aberystwyth-i jelentés. =Könyvtári Figyelő, 1975. 5. sz. 564-571. l.
51. Willets, M.: Investigation of the nature of the relation between terms and thesauri. =Journal of Documentation, vol. 31. 1975. No. 3. 158-184. l.
52. Neil, S.D.: Farradane's relations as a perceptual descriptions. =Journal of Doc. Vol. 31. 1975. 144-157. l.
53. A kutatási jelentések alapján összefoglaló: Grolier, E. i.m. 62-89. és 175-177. l.
54. Meždunarodnoj Centr Naučnoj i Tehničkoj Informacii. Rubrikator meždunarodnoj sistemy naučnoj i tehničkoj informacii. Vtoroe izd. Tom. 1-2. Moskva, 1984.
BSO Broad System of Ordering, Schedule and index. Third revision. Prep. by FID/BSO Pamel: E. Coates, G. Lloyd, D. Simandl. The Hague-Paris, FID, Unesco, 1978.
Dahlberg, I.: Projekt Ordnungssystem der Wissensgebiete. Phase 1: Materialsammlung. Frankfurt, Deutsche Gesellschaft für Dokumentation. 1973.
- -: Grundlagen universaler Wissensordnung. München, Dokumentation Verlag, 1974.
- -: Ontical structures and universal classification. Bangalore, Sarade Ranganathan Endowment, 1978.
55. Ungváry R.: Osztályozáselméleti megjegyzések az NTMIR tárgyköri osztályozási rendszerének - rubrikátorának - szerkezetéről. =Tudományos és Műszaki Tájékoztatás, 1983. 1-2. sz. 22-27. l.
56. Foskett, A.C.: The Broad System of Ordering: old wine into new bottles? =Int. Forum of Information and Documentation, 1979. 3. sz. 7-12. l.
Ugyanitt található Soergel, D. kritikája: The Broad System of Ordering, a critique. 21-24. l.
Dahlberg, I.: The Broad System of Ordering /BSO/ as a basis for an integrated social science thesaurus? =International Classification, 1980. 2. sz. 66-72. l.

57. Samuelson, K.: Coordination of diversified data-bases in information networks through multi-classificatory concept notation in relay-switches and referral directories. = Proc. 2nd seminar on UDC and mechanized information systems held in Frankfurt, 1-5. June, 1970. Copenhagen, Danish Center for Documentation, 1971. 230-243. 1.
58. McKay, I.m.
59. Kunszt Gy.: A tudományos kutatás logikai modellezése és tematikai irányítása. Általános elmélet és építéstudományi alkalmazások. Doktori értekezés. Bp. 1973. Soks. Az idézet: 35. 1.
60. A szöveg minta a magyar oktatásügy aktuális kérdéseire vonatkozó, azonos témájú interjúkból származik. A szöveg-választást indokolja, a kognitívumok meghatározását egy-egy részkérdéssel foglalkozó tanulmányban - kísérletileg kisszámú dokumentum esetében - logikai úton, az elemző személy preconcepciójával is meg lehet határozni, legalábbis befolyásolta volna az eredményt. Ezt elkerülendő került sor összetettebb szöveg meghatározására több interjú alapján. A szövegek végigolvasása az osztályozó személyektől szokásosan nem várható.
61. A szógyakoriság elemzésekor a már klasszikusnak számító Luhn módszert alkalmaztuk. A legnagyobb gyakoriságú szavakat kizártuk /névelők, kötőszók, stb./, továbbá a téma szempontjából túl általános szavakat is, mint a pedagógia, nevelés, iskola, stb. Ezek ugyanis semmitmondóak, mint ahogy egy botanikai rendszernél a "növény" nem hordoz semmi lényegeset. A munkagépek tárgyszó szótári asszisztenciával alakult ki. Az eszterga, marógép, stb. mint kváziszinonimák összevonásra kerültek a munkagépek genus alá. A küszöbértéket önkényesen 10-ben állapítottuk meg, miután alacsonyabb értékekkel kísérleteztünk. Az együttes előfordulást mondaton belül illetve bekezdésen belül értelmeztük. Módszertani kérdéseket itt nem tárgyalhatunk, egy problémát azonban jelezni kell. Meg kell különböztetni az együttes előfordulást és az állandósult szerkezeteket. Állandó szerkezet pl. Zsolnay kísérlet. Együttes előfordulás: Zsolnay kísérlet - kommunikációs készség vagy kommunikáció.
62. Balke, Siegfried: Benutzerprobleme der Dokumentation und Information. = Nachrichten für Dokumentation. Jg. 24., 1973. No. 1. 2-9. 1.

63. Wigner J. - Ákos K.: A tudomány növekedése - kedvező ki-
látások és várható veszélyek. =Magyar Tudomány, 75. köt.
1968. 5. sz. 317. l.
64. Wiener, N.: Információ, nyelv és társadalom. =Válogatott
tanulmányok, 162. l. Bp. Gondolat, 1974.
65. Kuhn, Th. S.: A tudományos forradalmak szerkezete. Bp.
Gondolat, 1984.
66. Wiener, i. m. uott.
67. Horváth T.: Automatikus osztályozás. Bölcsészdoktori disz-
sertáció. Bp. 1979. Kézirat. Ennek első fejezete ezekkel
a módszerekkel foglalkozik, közölve a legfontosabb statisztikai
eloszlásokat is.
68. Meadow, Ch. T.: The analysis of information systems. New-
York, London, Sidney, John Willey, 1967.
69. Shizuo, Fuiwara: Fundamental aspects of information in
modern science. =Int. Forum Information and Documentation.
Vol. 10. 1985. 2.sz. 11-14. l.
70. Jonker. F.: i.m. (29. jegyzet)
71. Salton, G.: A theory of indexing. 5. l. Ed. Society for
Industrial and Applied Mathematics. Philadelphia /Penn./,
Arrowsmith Ltd. 1975.
72. Salton, G.: i.m. 31-36. l.
73. Landry, B.C.: A theory of indexing. Indexing theory as a
model for information storage and retrieval. Ph. D. disser-
tation. Ohio State University, Columbus /Ohio/, 1971.
Landry azonban nem viszi végig saját gondolatát. Számos un.
posztulátumot fogalmaz meg, elmélete azonban megmarad a
Jonker-Salton elmélet kereteiben. Ennek tulajdonítható, hogy
értekezésének alig volt visszhangja.

74. Soergel, D.: Mathematical analysis of documentation systems. An attempt to a theory of classification and research request formulation. =Information Storage and Retrieval. Vol. 3. 129-173. 1. London, Pergamon Press, 1967.
Idézet: 133. 1.
75. Vojsvillo, J.K.: A fogalom. 206. 1. Bp. Gondolat, 1978.
76. A mátrix nem feltétlenül bináris, kitölthető 0, 1, 2, ... értékekkel. A diszciplínák strukturálásakor nem bináris fát, hanem többágut kapunk a mátrix szerint.
77. Köszönettel tartozom a Zenetudományi Intézetnek, hogy Bartók jegyzeteit rendelkezésemre bocsátották.
78. A folklórban máshol is hasonló eljárást követnek. A népmesék Aa-Th /Aarne-Thompson/ tipuskatalógusa a meséket egyetlen típusba sorolja. Pedig a tipizálás Thompson reformja után a motívumok szerint eszközölhető, tehát több elemi egység alapján. Hazánkban ugyanígy Berze Nagy János: Magyar népmesetipusok. 1-2. köt. Pécs, Baranya Megyei Tanács, 1957.
79. Salton, G.: The theory of indexing. Ed. Society for Industrial and Applied Mathematics, Philadelphia /Penn./, Arrow-smith, 1975.
A "típus" legegyszerűbb esete, amikor egy fogalom ekvivalencia osztálya alkotja a típust. Pl. eb, kutya, kutyák, kutyás, stb. (word type), míg az eb, kutya, stb. a szókép (word token). Innen indulva lehet a "típust" kiszélesíteni akár egy facettáig. A típus tagjai már nem ekvivalensek, hanem különböző értékeket vehetnek fel.
80. E rövid összefoglalás az alábbi közleményeken nyugszik.
(Szelektálva, részben kiegészítve a doktori értekezésből.)
- Borko, H.: Automated language processing. New-York-London-Sidney, 1968.
- Borko, H. - Bernick, M.D.: Automatic document classification. Technical memorandum. Santa Monica /Cal./ System Dev. Corp. 1967.
- Brofitt, J.D. - Morgavan, H.L. - Soden, J.V.: On some clustering techniques for information retrieval. Information Storage and retrieval to the National Science Foundation. Scientific report No. ISR-11. Cornell Univ. Ithaca /New-York/, 1966.

Ciganic, M.: Informacné systémy vo vede, technike a ekonomike. Martin, Matica Slovenska, 1969.

Fritsche, M.: Automatic clustering in information retrieval. Commission of the European Communities, Joint Nuclear Research Centre - ISPRA Establishment /Italy/, Scientific Data Processing Centre - CETIS. Luxembourg, 1974.

Jardin, N. - Van Risbergen, C.J.: The use of hierarchic clustering in information retrieval. =Inf. Storage and Retr. 1971. No. 5. 217-240. 1.

Salton, G.: Automatic information organization and retrieval. New-York, St. Louis, San Francisco, stb. 1968.

The SMART System. Experiments in automatic document processing. Ed. G. Salton. Englewood Cliffs, Prentice Hall, 1971.

Az együttidézési klasztertechnika Bujdosó Ernő: Bibliometria és tudománymetria. Bp. OSZK-KMK, MTA Könyvtára, 1986. művéből való. Az idézet a 194. lapon.

81. Horváth T.: Automatikus osztályozás. =Könyvtári Figyelő, 1978. 5.sz. 528-542. 1.
82. Horváth T.: Automatikus osztályozás. Ld. 67. jegyzet.
83. Ld. 80. jegyzetben
84. Tanimoto, T.T.: An elementary mathematical theory of classification and prediction. Scientific report. New-York, IBM, 1958. 10-14. 1.
85. Salton, G.: Proposals for a dynamic library. Ithaca /New-York/, Cornell Univ., Dep. of Computer Science. 1972.
86. Balázs S. - Horváth T.: Kézi lyukkártyák a szakirodalmi tájékoztatásban. 132-134. 1. Bp. OMKDK, 1968. A kezdeményezés visszhangtalan maradt. Talán azért, mert egy manuális technikával foglalkozó könyv szervezési fejezetében található.
87. Austin, D.: PRECIS. A manual of concept analysis and subject indexing. 2nd ed. 37. 1. The British Library Bibliographic Division, 1984.
88. Chomsky, N.: Generatív grammatika. /Beszélgetések Mitson Ronat-val/ 119. 1. Bp. Európa, 1985.

89. Fillmore, J.: The case of case. =Universals in linguistic theory. 1-88. 1. Ed. E. Bach et al. New-York, Holt, Reinhart, 1968.
Ismertetése: Radics Katalin: Ujabb irányzatok a generatív grammatikában. =Nyelvtud. Közl. 77. évf. 1975. 2. sz. 445-479. 1.
"Az utóbbi, a mélyszerkezet fogalma az ellentétes vélemények keresztútjára került, s a generatív grammatika elméleti keretében megszűnt ... Paul Postal azt állította, hogy a szintaktikai mélyszerkezet gondolata mögött az az elképzelés dolgozott, hogy a szintaxis a legkisebb jelentéssel bíró elemeket (morfémák) egymáshoz kapcsolásának szabályait foglalja magában." (448. l.) - írja.
90. Nem azonos Chomsky komplementum fogalmával. Nála ugyanis általában bővítmény.
91. R. Lakoff-ról és P. Postal-ról elnevezett elv, amely azt mondja, hogy a melléknevek az ige egyik fajtáját képezik. Ugyanez a felfogás Fillmore-nál és Austinnál is.
92. Chomsky, N.: i.m. 128. l.
93. Horváth T. - Varga D.: Információs tézauruszok. 51. l. Bp. NPI, 1977.
94. Gyakorlatilag az egész angol nyelvterületen BT=broader term, NT=narrower term használatos.
95. British Standards Institution: The ROOT thesaurus. London, BSI, 1981.
96. Ungvári R. - Orbán É.: A két- és többnyelvű tézauruszok irodalma. Kritikai szemle és annotált bibliográfia. Bp. OSZK-KMK, 1981.
97. A bírálat nem a tézauruszokat, hanem egyfelől a tézauruszok vulgáris felfogását marasztalja el; másfelől azt a hiedelmet, hogy a tézauruszok megoldanak minden osztályozási problémát. A tézauruszok a fejlődés szükségszerű láncszemeként jelentek meg. Furcsa lenne a tézauruszok teljes elmarasztalása, hiszen Varga Dénessel könyvet írtunk a tézauruszokról. (93. jegyzet)

98. A tárgyszavas osztályozások maradandó áttekintését adja Szilágyi Tibor: A tárgyszó és a tárgyszókatalógus. Szemle. Szerk. és ford. -- . Bp. OSZK-KMK, 1968.
99. Foskett, A.C.: i.m. 91. l. (ld. 48. jegyzet)
100. Horváth T. - Orbán É.: PRECIS. =Tudományos és Műszaki Tájékoztatás. 1987. 7. sz. 311-323. l.
Magyar nyelven még az alábbi áll rendelkezésre.
A PRECIS-t elsőként Sárdy Péter említi az indexművekről írt tanulmányában. (Gépi indexek. =Bibliográfiai tanulmányok. Bp. OSZK-KMK, 1978. 180-203.l.) A következő említés Sebestyén György disszertációjában található. (Nemzetközi információs törekvések és ezek akadályai a társadalomtudományi információs rendszerek nyelvi rendszerében. Bp. 1985.)
A Manual első kiadása alapján Orbán Éva készített egy kivonatos ismertetést. Ebben található a PRECIS indexelési eljárásának algoritmusai is. Időközben Austin jelentős változásokat vezetett be, amelyek a Manual 2. kiadásában találhatók. Austin ebben tárgyalja a számítógépes kezelés részleteit is, de az indexelés folyamatával itt nem foglalkozik. A Manual 2. kiadásából Horváth Tibor készített PRECIS 1-24 címmel huszonnégy témára bontott oktatási anyagot az OPIR indexelői számára. /Kézirat./ A kézikönyvek leírását ld. a következő jegyzetben.
101. Az összefoglaló az alábbi közleményeken nyugszik.
Austin, D.: PRECIS. ... Ld. 87. jegyzet.
Austin, D. PRECIS, A manual of concept analysis and subject indexing. London, The Council of the British National Bibliography, 1974.
Austin, D.: PRECIS in a multilingual context. Part 1. PRECIS, an overview. =Libri, Vol. 26. 1976. 1.sz. 137-156. l.
Austin, D. - Sorensen, J.: PRECIS in a multilingual context. Part 2. A linguistic and logical explanation of the syntax. =Libri, Vol. 26. 2.sz. 108-139. l.
Austin, D. - Verdier, V.: String indexing. PRECIS. London, Ontario, The University of Western Ontario, School of Library and Information Science, 1975.
DeHart, F.E. - Glaser, J.: Computer searching on PRECIS. An exploration of measuring comparative retrieval effectiveness. =International Classification, 1984. 1. sz. 3-8. l.
Lambert, G.: PRECIS in multilingual context. Part 4. The application of PRECIS in French. =Libri, vol. 26. 1976. 4. sz. 312-324. l.
PRECIS für die Anwendung in deutschen Bibliotheken Überarbeitete und vereinfachte Form des syntaktischen Indexierungsverfahrens der British Library. Berlin. Deutsches Bibliotheksinstitut, 1984.
Sorensen, J. - Austin, D.: PRECIS in a multilingual context. Part 3. Multilingual experiments, proposed codes and procedures for the Germanic languages. =Libri, vol. 26. 1976. 3. sz. 181-215. l.
102. Soergel, D.: i.m. 134. l.

