

Kvalitatív és kvantitatív szövegelemzés szoftverrel

Juhász Valéria

főiskolai docens

SZTE JGYPK Magyar és Alkalmazott Nyelvészeti Tanszék

juhaszvaleria@jgypk.szte.hu

Kutatási területeim az anyanyelv-pedagógia, az olvasás tanítása és tanulása, az olvasási készség fejlesztése, valamint a médiában és számítógépeken zajló kommunikáció tartalomelemzése. A közelmúltban a MAXQDA tartalomelemző szoftver használati lehetőségeivel és használatának népszerűsítésével is foglalkoztam.

1. A szövegelemző szoftvekről¹⁵

A szövegelemző szoftverek ősei a szószámláló számítógépes programok voltak. Kezdetben a számítógépet csak információ-visszakeresési rutinműveletekre használták a szövegek elemzésekor. A szövegelemző szoftver¹⁶ szövegeket azt sugallhatja, hogy léteznek olyan programok, melyek segítségével automatikusan, igen gyorsan hozzájuthatunk szövegekből származó egyéni vagy társadalmi jelenségeket / folyamatokat jelző, tükröző vagy azt meghatározó, illetve magyarázó fogalmakhoz, viszonyokhoz, jelentésekhez. Ezzel szemben a programok olyan eszközök, melyek segítséget nyújtanak ugyan a kutatóknak a rendszeres, rendezett munkához, de a jelenségek magyarázata, a kódolás az elemzőre marad. A tartalomelemző szoftverek¹⁷ a kódolás folyamatában nyújtják a legtöbb segítséget.

¹⁵ A cikk első megjelenése: Juhász Valéria. 2008. A MAXQDA szövegelemző program. In: Nádor O. (szerk.) *A magyar mint európai és világnyelv. XVIII. Magyar Alkalmazott Nyelvészeti Kongresszus előadásai*. Budapest: MANYE, Balassi Intézet.

¹⁶ Angolul gyakran így található: CAQDAS = Computer Assisted Qualitative Data Analysis Software vagy Computer Assisted Content Analysis.

¹⁷ A tartalomelemzés nem más, mint közlemények, szövegek meghatározott célú elemzése. Azokat az elemeket tárja fel nemcsak szövegekben, hanem képek, filmek közlés jellegű megnyilatkozásaiban, amelyek nincsenek nyíltan kimondva, esetleg az alkotóban sem tudatos közlésként jöttek létre, mégis a dekódolásban akár csak a sejtés szintjén is, de megjelenik a jelentésük. Ez a sorok közötti, sorok mögötti olvasás, értelmezés képessége, amely a kódolás módjában van elrejtve (Antal 1976). A tartalomelemzés módszerének a megismételhetőség kritériumait úgy kell biztosítani, hogy az egyes elemzési egységek kódolása és értelmezése a leírt metodika alapján ismét elvégezhető legyen. A módszeres és objektív eljárás azt jelenti, hogy a vizsgálat folyamán világos és egyértelműen megfogalmazott szabályok alapján dolgozunk. Előre meghatározzuk, milyen elemek, szimbólumok, szavak stb. kerüljenek az egyes kategóriákba. Minden tartalomelemzés két munkafázisra bontható. Az első fázist nevezik durván a kódolás fázisának. Ebben a fázisban a szöveget kódoljuk, tehát szimbólumait, szavait stb. előre megállapított kategóriákhoz

Kódoláson olyan karakterek, szavak, szövegegységek megjelölését értjük, amelyek valamilyen szempontból egy témakörbe tartoznak, van valami közös sajátosságuk, amiben hasonlítanak vagy eltérnek. A közös pont lehet a kód neve.

A forgalomban lévő szoftverek közül a magyar szakirodalomban viszonylag kevésre van utalás a tartalomelemzéssel végzett kutatásokban. Csakúgy, mint az adatokban és eljárásokban, a szoftverekben is alapvetően kétféle programcsomagot kínálnak a fejlesztők, vagy ezek kombinációit: az egyik a kvantitatív kutatásokhoz nyújt segítséget, a másik a kvalitatívhoz.¹⁸ A legrégebbi, ám ma is használt program a szótáralapú The General Inquirer,¹⁹ amit a CHILDES²⁰ követett. Ma ismertebb kvalitatív programcsomag például az ATLAS/ti, Code-A-Text, Computer Assisted Qualitative Data Analysis Software (CAQDAS) Networking Project, The Ethnograph v4.0, Kwalitan 4.0, NUD*IST, MAXQDA, QDA Miner, winMAX.²¹ Hazánkban a kutatási beszámolókból úgy tűnik, hogy az ATLAS/ti és a NUD*IST programokat használják többen. Meg kell említenünk még a készülőben lévő magyar NooJ rendszert, amelyet integrált nyelvelemző környezetnek (INYEK) is neveznek, róla bővebben a következő honlapon olvashatunk: <http://corpus.nytud.hu/nooj>.

Jómagam a MAXQDA-val dolgozom,²² mert a nemzetközi szakirodalomban olvasható, hogy számos területen sikeresen alkalmazzák szociológusok, politoló-

soroljuk. A második, magasabb fázisát az interpretáció szakaszának nevezhetjük. Az első szakaszban nyert mennyiségi eredmények értelmezésére, magyarázatára, a mélyebb rejtett összefüggések feltárására, kikövetkeztetésére kerül sor. Fontos tartalmi mutatóvá válhat valaminek a hiánya is.

¹⁸ A szoftverekről bővebben: <http://academic.csuohio.edu/kneuendorf/content/cpuca/ccap.htm>

¹⁹ A The General Inquirer Stone nevéhez fűződik (1966), új formájában az interneten is megtalálható: <http://www.wjh.harvard.edu/~inquirer/>. A program a Harvard IV-4 szótárt használva kódolja és osztályozza a szöveget olyan értékek megállapításával, mint az Osgood-féle háromdimenziójú szemantikai differenciálskála (a skála az egyes kijelentéseket három dimenzió mentén osztályozza: pozitív-negatív, erős-gyenge, aktív-passzív), kiválogatja az érzellemmel töltött szavakat, megállapítja a kognitív orientációt stb. A program elvégez olyan összegző statisztikákat, mint a szószámlálás vagy a szógyakoriság.

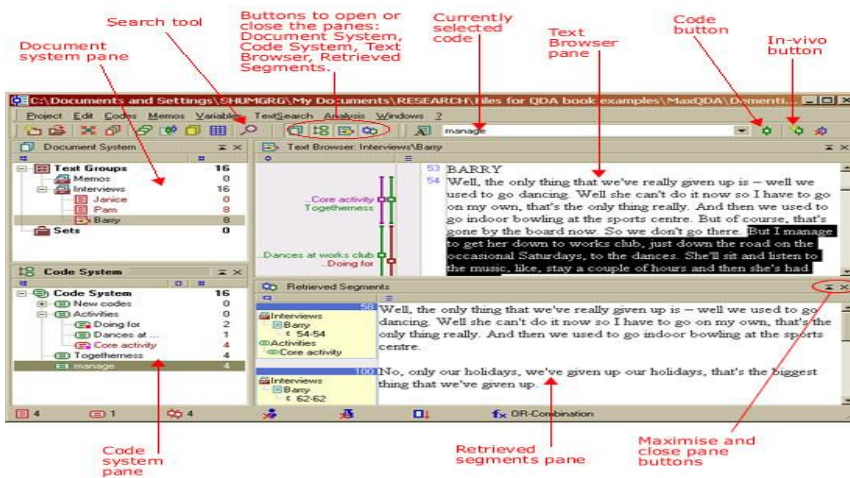
²⁰ CHILDES = Child Language Data Exchange. MacWhinney és Snow alkották meg (1985). Adatbázist, leíró rendszert és szerkesztő, kódoló programot, valamint egy morfológiai, szintaktikai elemző programot tartalmaz. Három (1,6 és 3,6 éves) iráni gyerek beszélgetéseinek elemzése.

²¹ További szoftverek: <http://www.qualitativresearch.uga.edu/QualPage/qda.html>

²² A MAXQDA2 változat két fontos kiegészítést is tartalmaz. Az egyik a MAXDictio alprogram, amely szógyakoriságok és relatív gyakorisági mutatók összeállítására alkalmas, a másik pedig a Visio Tools, amelynek segítségével egyszerű ábrákat készíthetünk az eredmények szemléltetésére.

gusok, pszichológusok, egészségügyi kutatók, antropológusok, piackutatók, közgazdászok stb. Kiemelkedő tulajdonságai közül elsősorban a magas hatékonysági fokát, megbízhatóságát, stabilitását, jól kidolgozott funkcionalitását és felhasználóbarát felületét említik. A program egyszerűen kezelhető és világos struktúrával rendelkezik. A szoftver úgynevezett projektekkel dolgozik. Ezek a munkalapok tartalmazzák a szövegeinket, a kódokat, az emlékeztető megjegyzéseinket, a változóinkat, a kódolt szövegek összesített táblázatát, amelyeket attribútumokként (attributes) tárol a fájlban belül, és amelyekből statisztikai jellemzőkhöz is juthatunk. A különböző funkciók: a szövegben való keresés, a kódolás folyamata és a memóírás lehetősége azonos felületen jelennek meg. A projektet kezelő ablakban legfelül található a menüsor, alatta az eszköztár, amely a gyakran használatos – vagy a gyors hozzáférési lehetőségeket biztosító – gombokat tartalmazza. Alattuk a képernyőn négy ablakot láthatunk egyszerre: a bal felső részen a teljes adatállományt (Document System/Text Groups), a bal alsó részen a kategória- vagy kódrendszert (Code System), a jobb felső ablakban a kiválasztott, vizsgálandó / vizsgált szöveget magát (Text Browser), és végül a bal alsó sarokban látható az alapvető vagy komplex keresési eredményeket tartalmazó ablak. Ez a Retrieved Segments, amelyen nemcsak kialakítani lehet feltevéseinket, de igazolni is. Itt olvashatjuk az azonos kódolású szövegeket, illetve kódok együttes előfordulásából nyert szövegegységeket is, amelyekhez a program analitikai funkciójának segítségével juthatunk.

Az 1. ábrán látható a program felülete.



1. ábra: A MAXQDA felülete, forrás:

http://onlineqda.hud.ac.uk/Step_by_step_software/MAXqda/MAXqda_import_docs.php

2. A szövegek előkészítése, létrehozása és behívása a MAXQDA-ba

A Document System ablakban láthatók az elemzésre előkészített szövegeink, szövegcsoportjaink. Ebbe az ablakba az elemzés elkezdése után is illeszthetünk még újabb szövegeket. Bármilyen szöveget, szövegeket behívhatunk a programba, amely Rich Text Format (.rtf) kiterjesztésű dokumentum. A dokumentumban lehet ábra, Excel-táblázat, PowerPoint-elem, vagy tartalmazhat hyperlinket is, azonban Wordben vagy Office-ban szerkesztett nagyobb táblázatot nem tud kezelni a program, ezeket át kell alakítani szöveges formátumúvá.

Az .rtf kiterjesztésben elmentett szövegcsoportot többféleképpen lehet importálni a projektbe. Vagy úgy, hogy több szöveget egy egységként hívunk be, vagy úgy, hogy sok, viszonylag rövid, előre szerkesztett szöveget külön-külön egységként hív be a program. Szöveget azonban nemcsak importálni lehet a projektbe, hanem magában a projektben is létrehozhatjuk a munkaanyagunkat. Dokumentumainkat rendezhetjük ábécésorrendbe, de egyesével is mozgathatjuk őket, törölhetők, illetve új szövegek illeszthetők a rendszerbe. A Document System ablakocskán belül található egy Text Set nevű lehetőség. Ide úgy rendezhetjük a szövegeinket, hogy a meglévő kódok alapján újabb kis csoportokat hozunk létre. Kijelölhetjük neki az elemzett szövegnek egyszerre több tulajdonságát. Például, hogy csak azokat a szövegeket válogassa ki a program, amelyek hűsvétkor készültek és nők írták (amennyiben ezek a tulajdonságok már kódoltak), és a Text Setbe illesztjük őket, hogy további megfigyeléseket végezzünk rajtuk.

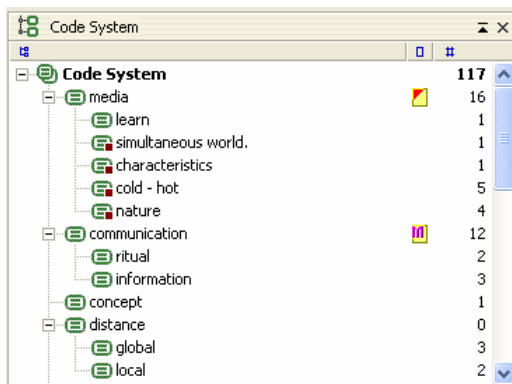
A szövegcsoportokon belül különböző lehetőségek vannak arra, hogy kijelöljünk, aktiváljunk szövegeket, mert azokkal akarunk dolgozni. Az Activate All gomb megnyomásával ki lehet jelölni az összes szöveget, a Ctrl gomb segítségével több szöveget is kijelölhetünk egyesével, valamint a szövegeink általunk megadott változói / attribútumai alapján egyszerűen aktiválhatunk több különböző, azonban egy vagy több változóban megegyező szöveget. (Az attribútumokra később részletesen kitérek.)

3. Kódolás

A MAXQDA egyik fő funkciója a kódolás maga: az, hogy a szöveg bizonyos részeihez, szavaihoz vagy akár csak egyetlen betűhöz kódokat, kategóriákat rendelünk. Ezek a kódok mutatnak rá a szövegben lévő tartalmi, lényegi minták jelenlétére. Ez a tulajdonképpeni kvalitatív elemzés első része. A kvalitatív adatok a kódolás folyamán úgy alakulnak kvantitatívvá, hogy az azonos kódok előfordu-

lásának számát, gyakoriságát őrzi a projekt. Ezek az előfordulások egy olyan mátrixba kerülnek, amelyeket aztán SPSS-be vagy Excelbe exportálhatunk. A végén a kvantitatív adatokat kvalitatív elemzésnek vetjük alá.

A kódok száma korlátlan lehet. Ezek hierarchikusan rendezhetők, és a kódok rendszere fastruktúráként jelenik meg a képernyőn. A kód-, illetve kategória-rendszer tízszintű mélységig, részletességig működik.



2. ábra: A kódrendszer fastruktúrája²³

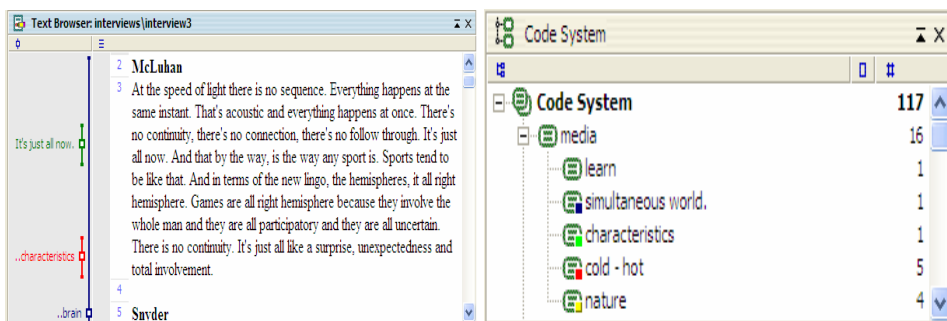
A kód létrehozása olyan, mintha egy üres fiókot alakítanánk ki. Ezeket „felcímkezve” behelyezhetünk olyan dolgokat, amelyek a fiók elnevezésében egyeznek meg. Ha később úgy találjuk, hogy a fiók mégsem a legpontosabb nevet kapta, azt anélkül nevezhetjük át, hogy tartalmában bármiféle változás menne végbe.

A kódolás műveletét háromféleképpen hajthatjuk végre: automatikus rákereséssel (szavakra, szavak kombinációira, együttes előfordulásukra van lehetőség), illetve kézi kódolással kétféleképpen: az adatállományt vizsgálva kijelöljük a kódolandó elemet, majd ezt egy már előre kialakított kódhoz csatoljuk. A kódolás másik módja az in-vivo kódolás. Ez azt jelenti, hogy az adatállományban található elem lesz a kód neve is, azaz az elemet kijelölve, az in-vivo gombot megnyomva a kódot is létrehozuk egyszerre. A program automatikusan legfölülre helyezi az in-vivo kódot, de a kutató ezt később egy áthelyezéssel könnyedén beillesztheti a fastruktúrába. (A kódolásnál lehetőség van kódtörlésre is.)

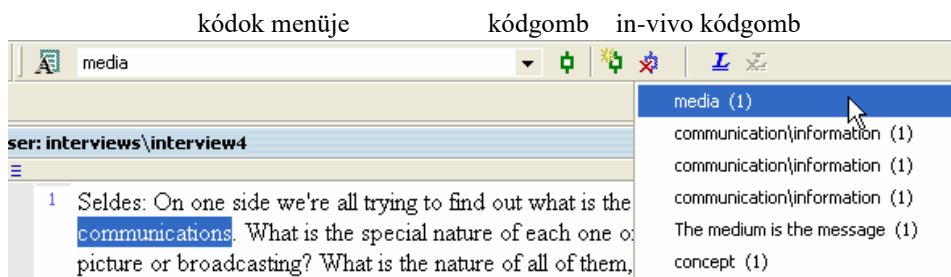
A kódok nemcsak a kategóriarendszerben jelennek meg, hanem az aktivált szöveg mellett is láthatóak, illetve átalakíthatóak attribútumokká, így megjelenik

²³ Az ábrákat a program útmutatójából vettem át.

majd a változók mátrixában. A programban arra is lehetőség van, hogy a kódjeleket különböző színekkel vizuálisan is elkülönítsük. Így olvasás, elemzés közben könnyebben azonosítjuk egy szöveg vagy szövegegység tartalmi vagy egyéb mutatóit. A kódolás során súlyozni is lehet a kódolt részeket aszerint, hogy azok mennyire relevánsak az adott esetben, azaz mennyire jellemzőek a szöveg és a kód összefüggésében.



3. ábra: Színes kód a kódolt szövegrész mellett és kódok a kódfában



4. ábra: A Text Browser ablakának eszköztárában látható kódolási lehetőségek

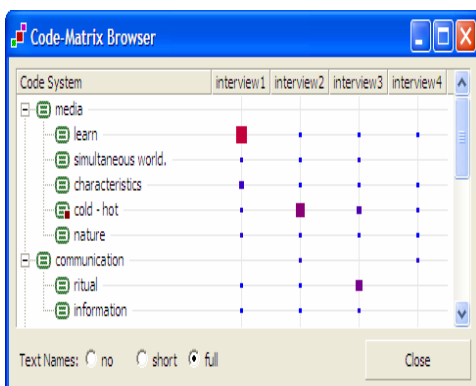
Kiemelkedő lehetősége a programnak, hogy egymást átfedő, egymásba ékelődő kódokat is képes kezelni. A kódfában a kódok nemcsak mozgathatók, hanem egymásba is másolhatók. Ezután viszont már nincs lehetőség az egymásba másolt, azaz egy kódhoz illesztett szövegrészek újabb szétválasztására automatikusan, csak manuálisan. (Ehhez kapcsolódóan megjegyzem, hogy a program egyik legnagyobb hátránya, hogy nincs rajta visszavonó gomb.)

A kódrendszert nemcsak kinyomtatni lehet, hanem egyik adatállományból a másikba exportálni, így – kódrendszerünket más adatállományon alkalmazva – hipotéziseinket igazolhatjuk, vagy újabb eredményekre juthatunk. A kódok előfordulásának gyakoriságát megtekinthetjük a Frequency of Codes alatt, illetve a

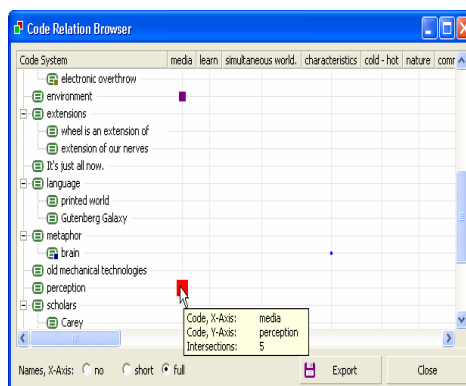
Code-Matrix Browser táblázatában. A Code Relation Browser a kódok közti kapcsolatot, összefüggést vizualizálja, azaz a szövegrészekhez rendelt kódok együttes előfordulását mutatja meg. Ezek a táblázatok segítenek kialakítani és igazolni is hipotéziseket, ezért jól alkalmazhatóak a Glaser és Strauss által leírt megalapozott elmélet (Glaser–Strauss 1967) technikai kivitelezésére.

	A	B	C	D	E	F	G
1	Textname	media	media\learn	media\simultaneous world.	media\characteristics	media\cold - hot	media\nature
2	interviews\interview1	0	11		3	4	3
3	interviews\interview2	0	1		2	1	8
4	interviews\interview3	0	1		3	2	5
5	interviews\interview4	0	1		0	1	2
6							

5. ábra: Kódyakorisági-táblázat



6. ábra: Kódmátrix-böngésző

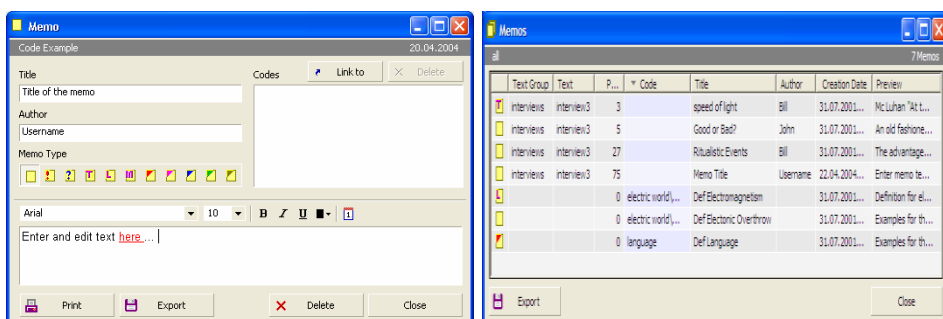


7. ábra: A kódok összefüggését mutató ábra

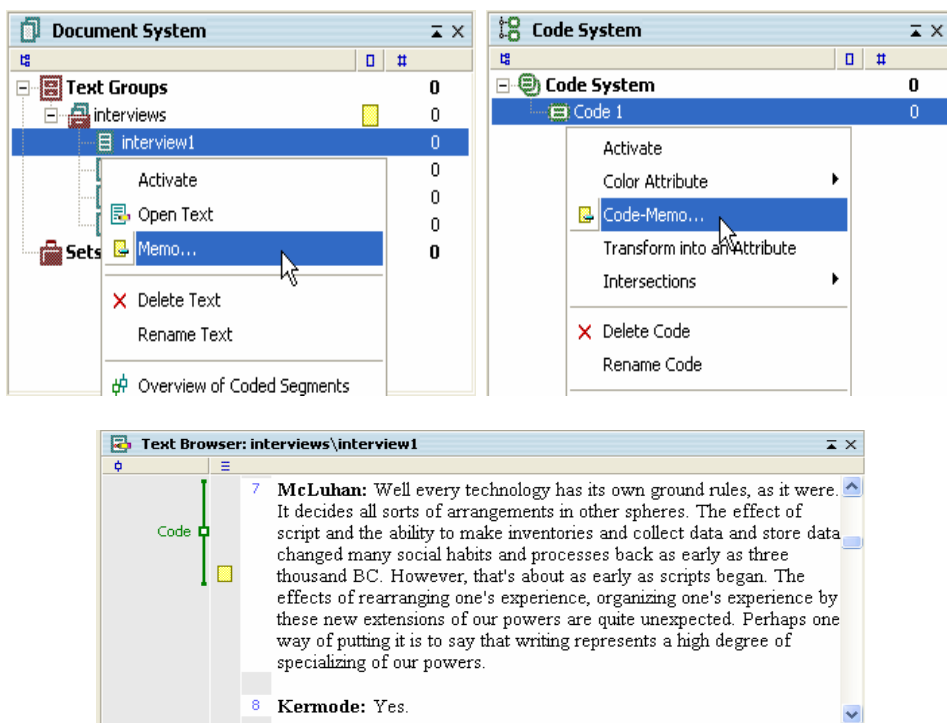
4. Emlékeztetők

A hipotézis kialakításához rendelkezésre állnak az emlékeztetőket tartalmazó úgynevezett memók. Memók mind szöveghez, mind kódokhoz illeszthetők. A „cédulázásnak” ez a lehetősége biztosítja az egyedi esetek, illetve az aktuálisan megfigyelt, később kipróbálható összefüggések azonnali feljegyzését, valamint segít a kategóriarendszerről szóló fogalmaink végső pontosításánál is. A Memo Manager segítségével az összes feljegyzésünket egyszerre is megjeleníthetjük, illetve kereshetünk bennünk, ha elfelejtettük, hogy valamit hova írtunk fel. A memók között különböző típusú ikonok szolgálnak arra, hogy elkülönítsük például azt a megjegyzést, amely a kódra vonatkozik, vagy azt, ami a módszert vagy

akár az elméletet illeti. 10 ilyen különböző memo található a párbeszédablakban. A memókat ki lehet nyomtatni, exportálni lehet másik munkába; a program továbbá tartalmaz egy memorendszert, amelyben rögzítve van, hogy ki írta a feljegyzést, melyik szöveghez, illetve melyik kódhoz kapcsolódik, és itt olvasható a memo neve, a feljegyzés dátuma és a jegyzet első pár szava is. Ez csoportos munka esetében nagyon hasznos.



8. ábra: Memokészítő párbeszédablak és a memorendszer



9. ábra: Szövegmémók, kódmémók és a szövegek melletti mémók

5. Attribútumok

A MAXQDA-ban lehetőség van arra, hogy minden szöveghez jellemzőket, változókat rendeljünk. Változója lehet egy szövegnek a szövegalkotó neve, kora, lakóhelye stb., és a kialakított kódjaink is átalakíthatók ilyen attribútumokká, azaz változókká. Természetesen nemcsak szövegek lehetnek változók, hanem számok is. A változók megadásával kerestethetünk az adatbázisunkban. Például, ha meg akarjuk vizsgálni, hogy a 14 év és 20 év közötti nők milyen értékeket helyeznek előtérbe életükben mindennapjaikról szóló szövegeikben, akkor a változók segítségével csak a kritériumoknak megfelelő szövegek jelenítődnek meg (a náluk idősebbek vagy fiatalabbak szövegei nem).

Ezeket a változókat a program egy négyszög alakú mátrixba rendezi. A mátrix mind a szövegek, mind a számok szerint rendezhető. Egy projektben csak egy ilyen attribútumokat tartalmazó táblázat van. A projekt kezdetén automatikusan megjelenik az attribútummátrixban a szövegcsoport neve, a szöveg neve, a létrehozás dátuma, a kódolt szakaszok száma, a memók száma, a szerző neve és a szöveg mérete byte-okban. Ebbe a táblázatba mi magunk is további elemeket illeszthetünk. Négy attribútumtípussal dolgozhatunk: szöveggel, számmal, dátummal és logikai változóval. (Ha később statisztikai programba akarjuk exportálni a mátrixot, akkor érdemes a numerikus jellemzőket használni.)

Textgroup	Textname	Creation Date	Number of Coded Segments	theory	copy
articles	communication...	25.07.2001 18:24	5	1	yes
articles	McLuhan and s...	25.07.2001 18:15	24	2	no
articles	global hopes	25.07.2001 19:31	12	2	no
documents	literature	25.07.2001 18:17	6	2	yes
documents	new references	25.07.2001 18:11	8	2	yes
interviews	interview1	26.07.2001 00:13	12	2	yes
interviews	interview2	26.07.2001 00:13	7	1	no

textgrou	textname	creation	v3	v4	author	bytes	theory	intervie	year
1 interviews	interview1	26-JUL-01	38	0	username	8664	2	yes	1967
2 interviews	interview2	26-JUL-01	22	0	username	8908	1	no	1969
3 interviews	interview3	26-JUL-01	26	4	username	8427	1	yes	1980
4 interviews	interview4	26-JUL-01	36	0	username	8908	2	no	1969

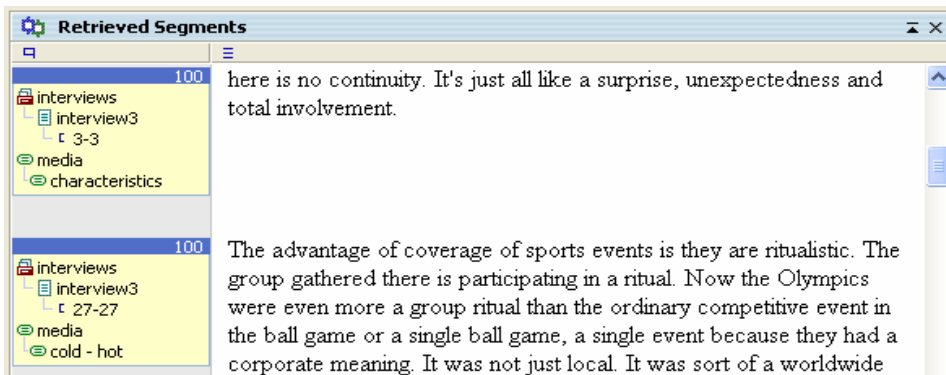
10. ábra: Attribútumtáblázat és SPSS-be illesztett megjelenése

6. Retrieved Coded Segments (RCS)

A képernyő bal alsó ablakában látható az ún. retrieved coded segments. A program itt jeleníti meg a már előzőleg kódolt és most kiválasztott / aktivált szegmenseket. Az újrakikeresésnek két lépése van: először aktiválni kell a szövegeket, majd a kiválasztott kódot vagy kódokat. Ekkor a RCS-ben csak azok a szövegrészek jelennek meg, amelyeket előzőleg aktiváltunk. Az oldal alján lévő sorban látható a kikeresés összefoglalója: hány aktivált szövegünk van, hány aktivált kódunk, ebből hány újrakiválasztott elem lett a kikeresés eredménye.



11. ábra: Állapotjelző csík



12. ábra: A Retrieved Coded Segments ablaka

Hogyha rákattintunk a kikeresett szöveg összefoglaló ikonjára, akkor a Text Browser ablakban megjeleníti azt a szöveget, amelyben a kódolt szövegrész található, hogy további elemzéseknek vessük alá.

A szoftverben tíz analitikai funkcióval tudunk adatokat kinyerni. Az Analysis menü első opciója a logikai kombinációkat tartalmazza. Az aktivált szövegekben az aktivált kódokat előhívhatjuk a „vagy” kombinációval, egymásba illeszkedő, egymást átfedő kódokat kerestethetünk vele, vagy egymás közelében lévőket, illetve az „Only this code” azt jelenti, hogy csak a kijelölt kódokat tartalmazó részeket hívja elő, de a többit ne stb.

7. Zárógondolatok

A MAXQDA olyan csomag, ami lehetőséget nyújt a csapatmunkára is. Ez a tartalomelemzések egyik legfontosabb eleme, hiszen elősegíti és támogatja a kutatói elfogultsággal, prekoncepciókkal szembeni védekezést. A programba bevitt adatbázison különböző gépeken egyszerre lehet kódolni, majd a különböző fájlokat egymásba lehet illeszteni. A MAXQDA alkalmas arra, hogy – tartalomelemzéseik során – széleskörű lehetőségeivel segítse a kutatókat téziseik kialakításában.

A tartalomelemzés módszere egyre elterjedtebb a tudományos kutatások számos területén, mert jelentős eredményeket szolgáltat a társadalmi folyamatok megértéséhez, az összefüggések feltárásához. Bátran és sokrétűen alkalmazható mindenféle nyelvi anyag elemzésére: nyelvtanítási stratégiák javítására, nyelvfejllettségi szint meghatározására, ideológiák feltárására, szövegek összehasonlítására, különböző típusú szövegekben található azonos fogalmak meghatározására, a fogalmak politikai tartalmainak elkülönítésére stb.

Irodalom

- Antal L. 1976. *A tartalomelemzés alapjai*. Budapest: Magvető.
- Glaser, B. G., Strauss, A. L. 1967. *The Discovery of Grounded Theory; Strategies for Qualitative Research*. Chicago: Aldine Pub. Co.