

A számítógépek a történettudományban

A hazai számítástechnikai szakemberek és a számítógépes eljárások alkalmazására törekvő történészek között eddig nem volt intézményes kapcsolat. Minden remény megvan azonban arra, hogy a közeledés hamarosan valamilyen szervezett formában is létrejön.* Ezért helyénvalónak látszik annak tisztázása, hogy mit jelent a számítógépek megjelenése a történettudományban. Az alábbi referátum úgy szeretne ehhez hozzájárulni, hogy egy általánosabb megállapításokat tartalmazó eszmefuttatás és két példa segítségével megkísérli néhány probléma jellemzését. E problémák az új helyzetben, a történész és a számítástechnika kapcsolatában szükségképpen merülnek fel. Természetesen csupán szerény kísérletről van szó, az elhangzó megállapítások minden bizonnyal megvitatásra és kiegészítésre szorulnak. A referátum szerzője várostörténész, akinek érdeklődése nem terjed túl a XVIII. századon. E körülmény tehát eleve meghatározza orientációját a számítógép felhasználásával kapcsolatban is, mondanivalója ezért egyéni álláspontot tükröz, mely csak őt kötelezi.

1

Mínt hogy a történelmi kutatómunka a források vizsgálatán alapszik, a történettudomány és a számítástechnika közti kapcsolatok megértése sem képzelhető el a történelmi források mint adottságok értelmezése nélkül. A számítógépes feldolgozásra alkalmas források két nagyobb csoportba oszthatók: *a)* nagy terjedelmű vagy mennyiségű forrásanyagok, melyek *elszórtan* tartalmaznak nagyobb tömegű, tipizálható információt, *b)* olyan forrásanyagok, melyek *összefüggően*, vagy *koherensen* tartalmaznak nagy tömegű, tipizálható információt. Ez utóbbiak esetében a forrásanyag terjedelme másodrendű szempont, helyesebben kizárólag az játszik szerepet, hogy a gépi feldolgozás gazdaságos-e vagy sem. Az első típusba sorolhatjuk például egy-egy időszak, korszak, terület, intézmény oklevélananyagát, elbeszélő forrásait stb., amelyek szövegéből kiválaszthatunk bizonyos adalékokat, avagy valamilyen rendszerező elv vagy statisztikai szempont szerint csoportosíthatjuk őket.

A második típusba sorolható források általában statisztikai jellegűek függetlenül attól, hogy készítőiknek mi volt velük a szándéka. Ezek tehát összeírások, számadások, anyakönyvek, egyéb komplex dokumentumok, amelyeknek az a közös jellemvonásuk, hogy numerikus, vagy könnyen numerikussá alakítható, nagy tömegű adat található bennük. Mindkét típusú forrásról el kell még mondani, hogy számítógépes feldolgozásuk igénye akkor merülhet fel, ha a belőlük kiszűrhető információ-anyag valamilyen értelmezés szerint a teljességet tükrözi,

* Időközben, 1972 októberében lezajlott a Neumann János Számítógéptudományi Társaság, a Magyar Történelmi Társulat és a Magyar Agrártudományi Egyesület kerekasztalkonferenciája a „Számítógépek és a Történettudomány” címmel.

továbbá legalábbis lehetőséget ad ahhoz, hogy egy körülhatárolt jelenséggel, szituációval, folyamattal vagy térbeli együttesel kapcsolatban globális képet, összefüggéseket lehessen alkotni. A tények nyelvére fordítva ez a megszorítás azt jelenti, hogy pl. értelmetlen dolog lenne egy oklevélgyűjtemény alapján helynév indexet készíteni úgy, hogy közben a kigyűjtött helynévanyagot nem sikerül azonosítani, azaz egy-egy helynév egyszerre több rejtett variánssal „fut”. Az is értelmetlen lenne, ha nem a rendelkezésünkre álló, témánk szempontjából teljes vagy legalábbis a teljes reprezentatív mintának tekinthető forrásanyagot dolgoznánk fel. (Természetesen ha a cél, tegyük fel valamilyen egységes szempontú nyelvészeti elemzés, akkor e konkrét esetben más a helyzet, s az index esetleg felhasználható.) Egy további példa: az a tény, hogy Sopronban a XV. századtól a XVII. század végéig rendelkezésünkre áll a város adójegyzékeinek nagy többsége, amelyek egyszerű numerikus adatokat tartalmaznak az adófizetőre kirótt adó nagyságáról, nem teszi még lehetővé, hogy e forrásokat minden további nélkül gépi úton elemezzük. Minden egyes jegyzék esetében ui. elemezni kell az adórendszert, azaz a kivetési kulcsot, továbbá ismernünk kell az ár- és bérvizonyokat, tisztázni kell, hogy az egyes adójegyzékek az egész várost és a teljes lakosságot foglalják-e magukba, stb. Így az adókönyvek adatsora a minden részletre kiterjedő feltáró munka elvégzéséig (amelybe egyébként elvileg szintén bevonható a számítógép) kezelhetetlen információhalmaznak tekinthető csupán, annak ellenére, hogy ezek az információk látszólag jól tipizálhatók. Megítélésem szerint itt a történettudomány és a számítástechnika legnagyobb konfliktusáról van szó, amelyre később még vissza kell térni.

A két fő csoport mellett említést tehetünk egy harmadik típusú forrásról is. Ez olyan nagyértékű mű lehet, amelynek részletes (formális vagy lexikológiai) elemzése a továbbiakban fontos tartalmi ismeretekhez vezethet. Gyakorlatilag egy-egy kompilált szövegrész elkülönítése, vagy például valamilyen ismeretlen írás megfejtése lehet a cél, mint az néhány éve az inka Csilam-Balam kódex esetében történt.¹

A számítógép felhasználására történész részről általában háromféle igény merül fel. *A)* Dokumentációs célokra, *B)* tipológiai elemzés céljaira, *C)* statisztikai számítások készítésére kívánják felhasználni a gép által kínált lehetőségeket. A feladat megfogalmazásakor világosan látszik, hogy az e három fajta közül melyikhez sorolható, az eredmények láttán, de gyakran már az algoritmus készítésekor kiderül azonban, hogy a géppel végzett nagy tömegű munka több célra is felhasználható.

Terjedelmi okokból kizárólag utalásszerűen említek néhány külföldön készült kutatást az előbbieket illusztrálására. *Dokumentációs* munkáknak tekinthetők mindenekelőtt a bibliográfiai jellegűek, de ezek a könyvtárak és a könyvtártudomány feladatkörébe tartoznak. Írott források dokumentációja történt egy úttörő francia kísérlet során. Lotharingia vidékének valamennyi 1120—1130 előtti eredeti oklevelét tartalmi és diplomatikai, paleográfiai szempontból elemezték². Az anyagi műveltség tanulmányozását segíti az ásatási leletek, sőt lelőhelyek számítógépes dokumentációja, amely ma már elterjedt módszernek mondható³.

¹ V. A. *Usztynov*: *Primenenie vücsiszlityelnüh masin v isztoriceszkoj nauke*. Moszkva, 1964. 15—20. l.

² L. *Fossier-Dupont*: *Méthode de traitement sur ordinateur des documents diplomatiques du haut-moyen âge*. „Annales. E. S. C.” 1970 január.

³ D. *Deopik*: *Klassifikacija i sistematiceszkij analiz keramicseszkoj kompleksza*

Tipológiai kutatás volt a közelmúltban a Brit Szigeteken található jellegzetes bronzkori szekercék vizsgálata, amelynek során a konvencionális régészeti tipológiai koncepciókat jobban definiált numerikus taxonómiai eljárásoknak vetették alá⁴. Komplexebb tipológiai és dokumentációs módszert dolgoztak ki, ugyancsak a közelmúltban a francia műemléki kataszter felállítása céljából⁵. Tipológiai jellegű az az eljárás, amelyet Borsa Gedeon ír le régi nyomtatványok betűtípusainak azonosítására⁶. Végül a tipológiai kutatásokhoz sorolhatjuk azokat a kísérleteket is, amelyeknek sírfeliratok, kompilált, vagy lexikológiai szempontból vizsgálándó szövegek elemzése a célja.

A *statisztikai elemzések* és a kvantifikáció fogalma a köztudatban összemósodik, ha ezekről a számítógépekkel kapcsolatban esik szó. Mindenekelőtt nyilvánvaló, hogy bonyolultabb matematikai módszerek alkalmazására pl. a tipológiai kutatásokban is sor kerülhet, másrészt olyan történeti statisztikai kutatások is léteznek, amelyek az alapműveletekre szorítkoznak, s meglegszenek a százalékosan kifejezett eredményekkel. Ilyen egyszerű, de a jövő szempontjából véleményem szerint nagy jelentőségű statisztikai módszer a Bertin-Arbellot-féle mechanikus kartográfia, amely a történeti statisztikai információt gép segítségével kartográfiai úton dolgozza fel. A statisztikai táblázatok térképre vitele lehetővé teszi, hogy a kutató vizuális úton ismerjen fel bizonyos korrelációkat⁷. A kvantifikációs módszerek alkalmazása terén a szovjet történettudomány úttörő szerepet játszik. Az egyik legérdekesebb kísérlet (amely egyébként részben saját, alább ismertetendő munkánkban is inspirált) az észtországi parasztságának a városok felé irányuló migrációját és társadalmi helyzetét vizsgálta az 1930-as évekből ismert adatok segítségével. Hat mutató alapján szerzője korrelációs matrixot, majd ennek alapján faktor matrixot készített, s így bizonyította, hogy a férfi lakosság és a magasabb műveltségű rétegek törekedtek leginkább a városokba⁸.

Az eddig elmondottak többek között azt a célt szolgálták, hogy illusztrálják a megállapítást: a történettudományban, és egyes társtudományokban is, a számítógépek alkalmazására irányuló törekvéseket általában a problémakeresés jellemzi. Sehol sem tudunk még összefüggő és huzamosabb ideje tartó munkákról. Szinte minden egyes ilyen természetű kutatás kísérlet, amelynek elsőrendű célja, vagy mellékterméke valamilyen új számítástechnikai eljárás kidolgozása egy adott feladat megoldása céljából. Ezen a tényen egyébként nem kell különösebben csodálkoznunk, hiszen az első, gép segítségével elért eredményekről a hatvanas évek derekán jelentek meg az összefoglaló ismertetések⁹. A legkézenfekvőbb felhasználási lehetőségek a rendszerezés és a statisztikai feladatok elvégzése; a gépi úton nyert eredmények azután a hagyományos történeti módszerek-

poszelenija epohi pozdney bronzü sz. Kirovo. „Drevnosztyi Vosztocsnogo Krüma”. Kiev. 1970; M. Borillo—P. Ihm: A Classification method for archeological Artifacts with incomplete Descriptions and structurally related Features. „Angol—román matematikai és régészeti konferencia.” Mamaia, 1970. szept. 16—24.

⁴ F. R. Hodson: Numerical typology in Archeology. „Angol—román konferencia.”

⁵ Avant-projet de code pour l'analyse des monuments civils. (kiad.) CNRS-CADA. Marseille, 1965.

⁶ Magyar Tudomány. 1971. 7—8. sz.

⁷ G. Arbellot: La cartographie statistique appliquée à l'histoire. Paris. 1970.

⁸ A. Ruuszman: Iz opüta primenenija faktornogo analiza v isztoriceszkom iszszledovanii. „Vesztnik MGÜ. Isztorija.” 1966.

⁹ Usztyinov 1964-ben megjelent könyve mellett ld. még: Computers for the Humanities? Yale University 1965, amely a Yale egyetemen megtartott konferencia anyagát tartalmazza.

kel végzett kutatás rendjébe illeszkednek, azaz annak egy-egy fázisát alkotják. Nem állítható, hogy a számítógép megjelenésével a történész hagyományos helyzete megváltozott volna. A helyzet ma is az, hogy a történész térben és időben lokalizált tényeket tanulmányoz illetve megmagyaráz. Érzékletesebben kifejezve: mintha egy hegycsúcson, netán a fellegek között ülne, s onnan szemlélné a lábai alatt hullámozó múltat.

Gyakorlatilag arról van szó, hogy a számítógépes eljárás alkalmazása előtt a forrás előkészítése részben, vagy teljes egészében továbbra is „manuális” eszközökkel történik. Például egy-egy birtokösszeírás vagy adóstatisztika előkészítése hatalmas nem numerikus és kodifikálhatatlan ismeretanyagot igényel. Amikor e sorok írója egy régi városunk adójegyzéke alapján vagyon és foglalkozási statisztikát akar készíteni, akkor egyebek között tudomásul kell vennie azt a már huzamosabb ideje ismert körülményt, hogy nálunk, a történeti Magyarországon a városi polgárnak egyszerre többféle megélhetési bázisa is lehetett. Ha tehát valakiről a jegyzék azt állítja, hogy *szabó*, emellett még foglalkozhatott borkereskedéssel, földje, állata és szőlője is lehetett. Ezzel szemben egy nagyobb nyugat-európai vagy itáliai városban a foglalkozás-megjelölések az esetek oroszlánrészében valóban a megélhetés kizárólagos forrását jelölik. E példa természetesen nem a hazai viszonyok speciális voltát kívánta kidomborítani. Hasonló nehézségek mindenütt és szinte állandóan felbukkanhatnak. Újabb példák a történészek egyre nagyobb szkepszissel vizsgálják a plébániai anyakönyvek forrásértékét. Igaz ugyan, hogy Nyugat-Európában ilyen anyakönyvek olykor a XV. század óta, hosszú sorozatot alkotva léteznek, vezetésük minősége azonban a plébános személyétől, hajlandóságától függött s az egységes szempontok kezelésükben csak a XVIII. századtól érvényesültek. Hosszú összefüggő sorozat készítésére ritkán alkalmasak. A források feldolgozása tehát számos kritikai szempont érvényesítését, megszorításokat, olykor betömhetetlen „lukakat” is magában foglal, a gépi nyelvre fordítható rendszer mellett.

Nem kisebb horderejű a különféle tudományelméleti álláspontok konfliktusa sem. E rövid referátumban nincsen helyünk ezek kifejtésére, de néhány utalás talán közelebb segíthet a megértésükhöz. A számítógépek alkalmazását a történeti kutatásokban nemcsak a forrásnehézségek determinálják, hanem szemléletbeli megfontolások is.

Legegyszerűbb, ha ismét egy konkrét példából indulunk ki. Marxista történész számára a társadalomtörténet kutatása elképzelhetetlen a tulajdonviszonyok feltárása nélkül. Ennek egyik fontos eszköze a vagyoni viszonyok elemzése, amelyek, — a várostörténetnél maradván — jól ábrázolhatók a különféle vagyonösszeírások vagy adójegyzékek segítségével (a forráskritikai nehézségektől e pillanatban elvonatkoztatunk). Ha ugyanis a vagyonság mértékét illusztráló változó más változókkal is összevethető (így a foglalkozási, topográfiai, termelési, esetleg demográfiai megoszlást jellemző változókkal), akkor lehetőségünk nyílik a vagyonság mértékének statisztikai interpretálására, s ezáltal a társadalmi rétegződés elemzésének szolid, összefüggő bázishoz jutunk¹⁰. Ezt a felfogást a Nyugaton már többször támadták. Az ellenérvek összefoglalása azért tanulságos számunkra, mert azzal, hogy a globális statisztikai vizsgálatok létjogosultságát valamely nézet elveti, egyúttal ezzel a számítástechnika alkalmazásának is korlátokat szab a történeti kutatásokban.

¹⁰ Ismét meg kell jegyezni, hogy várostörténeti példa lebeg előttünk, amelynek tanulságai azonban jóval szélesebb körben is érvényesek.

A társadalmi tagozódás imént jellemzett és „materialistának” bélyegzett elemzését az ellentábor azzal vádolja, hogy a vagyoni szintet nem tekinti minőségnek s e szemléletmód hatására bizonyos elemek indokolatlanul nagyobb hangsúlyt kapnak, mások rovására. A vagyonnak ui. összetétele is éppen olyan fontos, mint a nagysága, nemkülönbön jelentős a termelőeszközökhez való viszony, a tevékenységek szociológiai típusai, az életformák és életvitelek, a foglalkozási, vallási és politikai szervezetek típusai, valamint a kollektív pszichológia. A társadalmi struktúra definíciójának ezek elválaszthatatlan elemei. Maga a társadalmi struktúra kvalitatív fogalom, amelynek elemzése a bírálókat szerint nem zárja ki a statisztikai ábrázolást, minthogy azonban globális struktúráról van szó, amely kis csoportok társadalmi tevékenységének összességét foglalja magában, a gazdasági jellegű mozzanatok nem választhatók el a többi, társadalmi, politikai, pszichológiai mozzanattól, s így ebben az értelemben a statisztikai adatfeltárás jelentősége csekély, inkább dokumentatív szerepet kap. Az érvelés egy további eleme, hogy a társadalmi struktúra megértése a családi, mindenekelőtt a házassági kapcsolatok kutatásával kezdődik, ezen alapul¹¹. Természetes, hogy a számítástechnika alkalmazását történetési feladatok megoldására az ilyen álláspontok határozottan befolyásolják, annál is inkább, mert a gép segítségével végzett munka költséges, ezért magasabb szintű szervek, intézmények döntéseinek a függvénye.

A helyzet világosabb jellemzése érdekében nem térhetünk ki a fentiekben sommásan összefoglalt polémia elől. Ezt annál is kevésbé tehetjük, mert az utóbb ismertetett álláspont mögött a strukturalista antropológiának a történetírással szemben támasztott bizonyos elvi kifogásait véljük felismerni, azaz a konkrét ellentét mélyebb, elvi nézetkülönbség megnyilvánulása, amely a történetész státusát, szemléletét érinti. A polémia részleteiben e szűkreszabott keretek között per sze felesleges elmerülni. A rövidség kedvéért szögezzük le, hogy esetünkben az említett elvi nézetkülönbség két ponton mutatkozik. Az egyik probléma az, jogunk van-e ahhoz, hogy a vizsgálandó struktúrán belül bizonyos mozzanatokat kiemeljünk. Konkrétan a gazdasági mozzanatok kiemeléséről van szó. A másik probléma, hogy a kutatás a maga minden mozzanatra kiterjedő komplex szempontjaival az izolált egyéncsoport (ház-utca; helyiség-vidék) felől haladjon-e a nagyobb egységek felé a globális struktúra megrajzolása során, avagy a szélesebb (mélyebb), illetve hosszabb időközökre érvényes (*longue durée*) összefüggéseket kutassa-e, bizonyos, a történeti mozgás szempontjából elsősorban fontosnak tekintett mozzanatok feltárása alapján. Első látásra az a benyomás, hogy a két álláspont nem is összeegyeztethetetlen, hiszen az eszményi állapot mintha éppen az lenne, ha a két szempont egyformán érvényesülne. Valójában az ellentmondás feloldhatatlan, mert a történetész nem képes feltárni a történeti valóságot a történeti mozgás (ezen belül is annak legfőbb eleme, a gazdasági mozzanat) és a tulajdonviszonyok megértése nélkül. Ezért kerül sor kidomborításukra, ami természetesen nem jelenti azt, hogy mintegy „lehámoznánk” a gazdasági szférát és a tulajdonviszonyok már önmagában is igen komplex problémakörét. A történelmi struktúrák kutatása azonban nem az alkotó elemek mozaikszerű összegezéséből áll, hanem a mozgási folyamat felderítésére irányul, ezért a struktúra definíciója megszilárdulásának és átalakulásának definícióját is magában kell, hogy foglalja. Mindez nem képzelhető el a nagyobb egységek, tágabb összefüggések feltárása nélkül. Mindent

¹¹ Ld. R. Mousniers több tanulmányát, továbbá *J.-Y. Tirat*: Problème de méthode en histoire sociale. *Revue d'histoire moderne et contemporaine*, 1963. 10. köt.

összevetve egy dinamikus és egy statikus szemlélet összeütközéséről van szó, amely egyébként néhány évvel ezelőtt sokat foglalkoztatta a történész közvéleményt¹².

A marxista történettudomány így nem véletlenül, nem a divatnak hódolva, nem is egyszerűen a természettudományok és a szolidabb matematikai bázison operáló társadalomtudományok nyomására fordul a számítástechnika felé. Nem is kizárólag azt várja tőle, hogy adattárolást, csoportosítást és egyéb rutinjellegű feladatokat könnyítsen meg, netán e terheket teljes egészében vállalja át. Amikor a történelem tudományos megismerésére törekszik, célja a történelmi folyamatok, a törvényszerűségek feltárása. Lenin a törvényszerűség elemzésével kapcsolatban hangsúlyozta, hogy az „nem mutatkozhat meg másképp, mint átlagos társadalmi, tömeges törvényszerűségben, szembeállítva az egyik, vagy másik irányban történő egyedi elhajlásokkal”. (Lenin művei. 21. köt. 50. l.) A tudományos eszközökkel végzett marxista történelmi kutatás tehát a valóság statisztikai tényeken alapuló, azaz szociológiai értelemben általánosított feltárására törekszik, s a tudományos felépítésben az empirikus tények az előbbieket mellett reprezentatív szerepet kapnak¹³. Úgy vélem így válik érthetővé, miért jutott olyan kiemelkedő szerep a szovjet történettudománynak a kvantifikációs technikák, valamint a számítástechnika alkalmazása területén, s hogy miért fordult nálunk is a figyelem elsősorban a történelmi-statisztikai feladatok gépre vitele felé. A számítástechnikát olyan eszköznek tekintjük, amely a tudományos megismerésben is előrevivő szerepet fog játszani. Ez a megállapítás úgy értendő, hogy a történelmi kutatás segédmunkáit megkönnyítő eljárások mellett a számítástechnika már most is kínál olyan eljárásokat, amelyek részben a munka extenzitásának egyszerű növelésével, részben más módon a rendelkezésünkre álló információhalmaz összehasonlíthatatlanul nagyobb hányadának az eddigieknél sokkal átfogóbb és alaposabb feldolgozását, elemzését teszik lehetővé. Így megfigyeléseink mennyisége és exaktsága még abban az esetben is nő, ha óvatosan mindössze azt tételezzük fel, hogy a számítástechnikai eljárások ezidőszert a történettudományban „manuális” úton kidolgozott módszerek gépi alkalmazását szolgálják. (Az igazság az, hogy már napjainkban is találkozhatunk *par excellence* gépi módszerekkel, amelyek korábban, a számítógépek megjelenése előtt elképzelhetetlenek voltak. Ilyen például az imént említett kartográfiai illetve faktor analitikus módszer.) Térjünk vissza azonban megegyezően a strukturalista történelem koncepció egyik tételéhez. Lévi-Strauss a *Pensée Sauvage*ban egy helyen arról ír, hogy a történetírásnak, mint minden megismerésnek, kódot kell használnia tárgya elemzésére még akkor is, ha tárgyát folyamatos valóságnak képzeli el. Ezzel kapcsolatban mutat rá arra az ellentmondásra, hogy „ha (a történész) folytonosnak képzeli el tárgyát, lehetetlenné válik, mert a folytonosság leírása végtelen regressziót igényel; ahhoz, hogy a történetírás ténylegesen létezhesen, el kell szigetelnie, 'kvantálnia' kell az eseményeket; ezzel azonban megszünteti az idődimenzió kivételes szerepét, mert minden kvantált esemény úgy tekinthető, mint a priori lehetséges események halmazából történő véletlen választás eredménye (A kiemelés tőlem! — G. Gy.)¹⁴ E megállapítás azért figyelemre méltó, mert elméleti indoklását adja az imént körvonalazott marxista történelmi statisztikai koncepció elvetésének. Ezzel kapcsolatban ismét hangsúlyozni szeretném, hogy a szemléletbeli, koncepcionális, mondhatni világnézeti jellegű megállapítások és megfontolások a „tisztá”

¹² Vö. Világtörténet, 1965. 7. sz., Valóság, 1967. 11. sz.

¹³ E témával kapcsolatban vö. M. A. Barg cikkét a Világtörténet idézett számában.

¹⁴ Az idézetet a Valóság említett számában közölt szemelvény is tartalmazza.

tudományos törekvések mellett napjaink egyre átfogóbb tudományszervezésében szükségszerűen ugyancsak egyre nagyobb jelentőséget nyernek. Szükségszerűnek kell tehát tekintenünk azt a fejleményt is, hogy Franciaországban a marxista történettudomány szemléletével számos rokon vonást mutató ún. „Annales” iskola lett kezdettől fogva egyrészt a számítástechnika, másrészt a statisztikai módszerek alkalmazásának centruma, továbbá hogy a Lévi-Strauss féle strukturalista antropológiát bíráló, vagy azzal ellenséges hatásának mindenestre jóval kevésbé kitett angolszász országokban ilyen alapokról kiinduló megfontolásokkal aligha kellett megküzdeni.

A strukturalista állásponhoz képest úgy látszik az ellenkező véglelet képviseli az ún. „kvantitatív” történelem koncepciója. Erről azért kell szólnunk, mert egyrészt létrejött összefügg a számítástechnika megjelenésével, másrészt mert szigorú elvi bírálatot gyakorol a közkeletű történeti statisztika felett. A „kvantitatív” történelem elvét a közgazdász Marczewski fejtette ki, s az első eredmények már napvilágot is láttak¹⁵. Marczewski szerint a statisztikai módszerek alkalmazása a hagyományos gazdaságtörténeti kutatásokban a kvantitatív szempontok illúzióját keltik. Nem képesek azonban a Marczewski által hagyományosnak nevezett módszerek arra, hogy az információs anyagban mutatkozó részeket betömjék, s a kiválasztott adatok szubjektív karakterét csak részlegesen küszöbölik ki, azaz a történész szintézis alkotó képességét csak csekély mértékben növelik. A statisztikai módszereket alkalmazó gazdaságtörténet végeredményben nem „kvantitatív” mindaddig, amíg kiindulása, azaz a felhasználandó adatok kiválasztása nem kvantitatív módszerekkel történik, s amíg a következtetéseket, amelyekhez elvezet, nem alkalmasak összefüggő, kvantitatív kifejezésre. A kvantitatív módszerek előnyét Marczewski abban látja, hogy elmozdítják azt a pontot, amelyről a megfigyelő szelektál ahelyett, hogy a leírandó valóság megfigyelésének pillanatában lépne közbe. E közbelépés a referencia rendszer kiépítése során, folyamatosan valósul meg. Maga a referencia rendszer arra szolgál, hogy megszámolja az ilyen módon koncepcionálisan homogénné tett tényeket, s felépítése minden történetírói törekvéstől függetlenül megtörténhet.

Ez a felfogás egyrészt áttételes módon felelni akar a strukturalista antropológia által támasztott globális igénynek, másrészt, s a saját gondolatmenetünk szempontjából ez a körülmény a fontosabb, tükrözi azt a természettudományok művelői, valamint a matematikai módszerek alkalmazására jobb lehetőségeket kínáló társadalomtudományok (pszichológia, szociológia, ökonometria stb.) művelői által hangoztatott véleményt, hogy a történettudomány korunk lemaradó, archaikus állapotban megrekedt tudományterületei közé tartozik, mert nem képes a matematika nyelvén elsajátítására, s ezzel végleges korlátokat von saját maga és a megismerendő valóság köré.

Tárgyilagosan le szeretném szögezni, hogy a Marczewski által felvázolt koncepciót a „kvantitatív” történetírásról a magam részéről határozottan csábítónak és elgondolkodtatónak vélem. A benne kifejezett igény egybeesik a modern, tehát nem mitoszgyártó, hanem objektív megismerésre törekvő történettudomány azon igényével, amely ki akarja iktatni a szubjektív, legalábbis indulati, érzelmi indítékokkal elhomályosított látásmódot. Sajnos azonban meg kell állapítani, hogy a közgazdászok általában lebecsülik a történeti megismerés egyik legnagyobb akadályát, azt ti., amely bennünket az ún. *proto-statisztikai* kor-

¹⁵ J. Marczewski: Buts et méthodes de l'histoire quantitative. Cahiers Vilfredo Pareto, 1964. 3. sz.

szaktól választ el. Ez az objektív akadály Európa egyes területein különböző időpontokban jelentkezik, nagyjából mégis a XVIII. század vége és a XIX. sz. közepe közé eső időszakról számítható. Ennél korábbi periódusokból csak szórványos és igen heterogén, statisztikai vizsgálatra alkalmas források állnak rendelkezésünkre, amelyeknek éppen e nem összefüggő jellegük miatt, mint már korábban utaltunk rá, csak nagy és bonyolult előkészítő munka után vehetjük hasznát.¹⁶ A Marczewski által kifejtett elv megvalósításának tehát gyakorlatilag, bármely területet vagy országot vegyünk is számításba, mindkét alapfeltétel hiányzik a XVIII—XIX. század előtt. Nem rendelkezünk sem hosszú periódusokra, sem nagyobb területekre érvényes összefüggő, vagy koherens statisztikai forrásokkal. Sőt ezidőszerint úgy látszik, olyan forrástípusokkal sem rendelkezünk, amelyeket hosszú előkészítés után a bennük elszórtan található információk révén alkalmassá lehetne tenni a globális statisztikai elemzésre, hogy ily módon a nehézséget legalábbis áthidalhassuk.

Kérdés azonban, hogy a történész feladatát ilyen módon kell-e megfogalmaznunk. Megítélésem szerint a számítástechnika megjelenése a történettudományban nem fog totális fordulatot hozni. Jóllehet talán túl sokat is foglalkoztunk a történeti-statisztika témájával, nyilvánvaló, hogy a történész nem csupán kvantálható ismeretekre törekszik. Amellett, hogy mind a források, mind a kutandó területek számosak és bonyolultak, a történész állandóan a rokon tudományok, mint a szociológia, etnográfia, közgazdaságtan, régészet, művészettörténet, irodalomtudomány hatása alatt, azok szempontjait is vizsgálva, adaptálva dolgozik. A történeti valóság objektív interpretálására törekedve olyan rekonstrukciót kíván tehát alkotni, amely a lehetőségekhez képest a legsokrétűbb. De a megismerés áttörhetetlen korlátai miatt a rekonstrukció tökéletlen, s az is marad örökre. Éppen ezért lehet újból és újból megkísérelni egy-egy történeti probléma feloldását. A számítástechnika által kínált hatalmas lehetőségek a történettudomány haladása szempontjából azért jelentősek, mert segítségükkel régóta ismert forrásanyagokat újból, tökéletesebben lehet feldolgozni, új, eddig megközelíthetetlennek vélt forrásokat lehet majd bevonni a kutatásba, s így pontosabb, sőt új eredmények fognak születni. Hogy mit, milyen lehetőségeket nyújtott eddig a számítástechnika a történettudománynak, arról már röviden szót ejtettünk. Még ennél is rövidebben, mintegy kitekintésként megkockáztathatunk néhány feltevést arról is, hogyan segítheti a számítástechnika a történész kutató munkáját.

Mindenekelőtt Marczewskivel ellentétben úgy vélem, hogy a számítógép segítségével végzett munka igenis növeli a történész szintézis alkotó képességét. Az imént utaltunk már arra, hogy ott, ahol erre lehetőség kínálkozik, a gép hallatlan mértékben megnöveli a kutatás extenzitását, s ez a pusztán körülmény minőségi változást hozhat még akkor, is ha az extenzitás növekedése nem jár együtt valamilyen hagyományos értelemben vett felfedezéssel. Minőségi változást jelent már az is, ha eddig *implicit*e mennyiségi jellegű megállapítások a jövőben *explicit*e is azok lesznek. A „nagy többség”, „jelentős hányad” és hasonló kifejezéseket mind több esetben a számszerű kifejezés helyettesítheti. Minőségi változással járhat az is, hogy a számítógép a kutatót határozott állásfoglalásra kényszeríti olyan kritikus esetekben, amelyekben eddig megkerülte az állásfoglalást. Végül,

¹⁶ Kivételesen gazdag források maradtak fenn a *proto-statisztikai* korszakból egyes itáliai városok levéltáraiban. Ezek a *catasto*-k olykor már a XI. század óta léteznek, s rendszeresen megújították őket.

hamarosan eljöhethet az az idő, amikor a matematikai és a számítástechnikai eljárások hatására a történettudomány bizonyos területein a szó, a beszéd mellett más, a lényeget tömörebben kifejező eszközök is a rendelkezésünkre fognak állni.

2

A FORCOD rendszer

Megjegyzés: A következőkben csupán vázlatos ismertetést adunk a FORCOD rendszerről. Részletesebb technikai leírások az alábbi kiadványokban találhatóak:

Couturier, Marcel: FORCOD B. (Kiad.) EPHE, Párizs, 1969 (sokszorosított)

Béaur, Gérard: Le langage FORCOD B. Párizs, 1971.

Poublan, Danièle: Système FORCOD. L'exploitation en FORCOD B. Párizs, 1971 (soksz.).

Couturier, Marcel: Le langage FORCOD C. Principes généraux (Párizs, 1971) (soksz.).

*

1. Általános tudnivalók

A FORCOD szó a *FOR*malisation és a *COD*age automatique szavakból kialakított összevonás, amely tehát az automatikus formalizálás és kódolás kifejezéseket rövidíti.

A FORCOD rendszer kialakítása, a kísérletek és a rendszeres felhasználás Marcel Couturier vezetésével folyik az *École Pratique des Hautes Etudes* Történeti Kutatási Központjában (Párizs). A kiinduláskor az volt a cél, hogy az elsősorban társadalomtörténeti kutatásokban nagy sorozatokban felhasználásra kerülő, azonos típusú források feldolgozását megkönnyítsék, illetve a kutatás megoldhatatlan mennyiségi feladatait számítógépre bízassák. A rendszer alapelve, hogy felhasználójának nem kell ismernie azokat a technikai megkötéseket, amelyek a számítógépek felhasználásával járnak. A gép a kutató szempontjából mesterséges értelmi blokk, amely egy természetes nyelven, a FORCOD nyelven „beszél”. A FORCOD nyelv szabályai a francia nyelvén alapulnak, s megegyeznek számos indoeurópai nyelv szabályaival. (Couturier közlése szerint elvileg még annak sincs akadályja, hogy a FORCOD a magyarhoz igazodjék.) Egyaránt ezt a nyelvet kell felhasználni az adatok összegyűjtésére, a gépnek szóló „ismeretek” közlésére és az utasítások megfogalmazására.

Első, A variációja 1966 óta működött, ezt ma már nem használják. 1967-ben készült el a B változat, amely bináris kódot használ. Működtetéséhez olyan gépek jöhetnek számításba, amelyek 24 bites szavakkal operálnak. A gyakorlati kutatómunka (tehát az alkalmazás) ma ezzel a rendszerrel folyik. A FORCOD B nyelvet kiegészíti a FORTAB B1 nyelv, amely a programozásnál használatos. A legújabb, kísérleti állapotban levő rendszer, a FORCOD C 32 bites szavakkal operáló gépekre készül, s a B változat logikus továbbfejlesztésének tekinthető. A FORCOD rendszer tehát a speciális FORCOD nyelvre épül, amely egyrészt géphez, másrészt feladattípushoz igazodik.

2. Felhasználási lehetőségek

A továbbiakban a ma működő FORCOD B adottságainak ismertetésére szorítkozunk.

Statisztikai számításoknál a jelenségek osztályonkénti összeszámlálását végzi el. Az osztályokat egyetlen jellemzővel avagy összetettebb módon, ÉS illetve VAGY operátorokkal is meg lehet határozni. — Összeget és százalékot számít, akár részleges akár általános formában a táblázat különböző osztályaiban nyert számok alapján. — Lehetővé teszi az adatok különböző szinten végzendő szűrését: bemenet a táblázatok csoportjába, bemenet a táblázatba, majd bemenet a táblázat négy lehetséges dimenziójának bármelyikébe (e dimenziók mindegyik bemeneténél diszperzió vagy szűrés). — Minden táblázat oszloponkénti vagy soronkénti átlaga és mediánja.

Automatikus kartográfiai célkitűzés esetében a felhasználónak tudomásul kell vennie, hogy a rendszer különleges térképeket készít. Az automatikus kartográfiai program ui. nem felületeket, hanem pontokat vesz figyelembe, tehát például egy-egy megye csak egy-egy pont lehet, kivéve persze, ha ennél kisebb alosztályok vannak. A végeredmény a felhasználó választása szerint számokban vagy a szürke szín árnyalataiban fog jelentkezni.

Dokumentációs feladatok megfogalmazásakor a program egy cédulaanyag tételek szerinti csoportosítását teszi lehetővé.

3. A műveletek lefolyása (lásd az organigrammot a 39. oldalon !)

A munka első része az adatok numerikus kóddá alakításából áll, s e folyamat addig tart, míg szalagon elhelyezkedő kódolt és normalizált adatok sora alakul ki. A kódolást egy első elemzés előzi meg, amelynek során kialakul az ún. „meghatározók” szalagja.

A felhasználó a munka során négy ízben lép közbe:

— az adatok összegyűjtése során, ami egyszerűen magnetofon szalagra olvasás útján is történhet. Ilyenkor azonban előfordulhat, hogy koncepcionális előkészítő munkára van szükség az információ rendszerezése és egyszerűsítése céljából.

— az ún. „meghatározók” (*définiteurs*) listájának összeállítására. Ezek olyan szavak, amelyek az információ különböző aspektusait határozzák meg.

— végeleges kódolás, vagyis a gép által *ad hoc* adott kódok ésszerű rendezése.

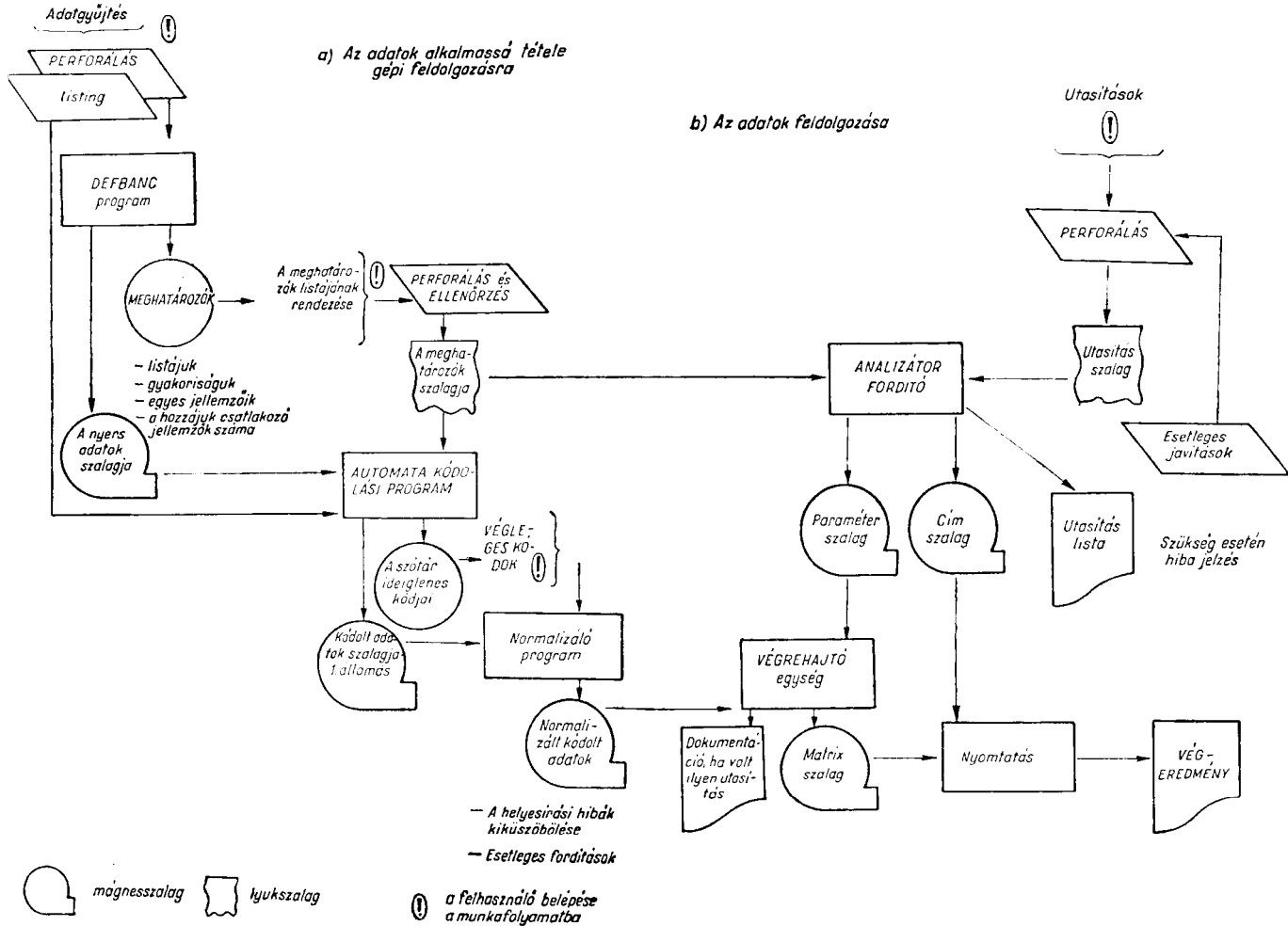
— a program elkészítése a FORCOD-hoz kapcsolódó FORTAB nyelv szabályai szerint.

4. A FORCOD B felhasználásának gyakorlati példája

Az EPHE támogatásával folyó történeti kutatások közül az alábbiakban illusztrációként összefoglaljuk az egyik munka célját és fő fázisait.

A társadalomtörténeti témán D. Letouzey dolgozik. Címe: „*Bicêtre* kastélya: általános kórház és lakói a XVIII. század közepén.” A források közül a legfontosabb az az anyakönyv, amelyet a kórházban vezettek. Ebben kerültek feljegyzésre minden beteg felvételével, személyével és „bentlakásának” megszűné-

A FORCOD organigrammja



sével kapcsolatos tudnivalók. Az egyes betegekre vonatkozó információk formája, természete és száma alig változik. Íme egy bejegyzés magyar fordításban:

<p>Meghalt e házban 1730. február 23-án</p>	<p>1729. december 13 Jacques Foucault, 30 éves, <i>Jarnacque</i>-ból <i>Saintonge</i>-ban, belépett a Főügyész Úr 12-én kelt levelével, hogy nemi betegséggel kezeltessék.</p>
---	--

Az adatok összegyűjtése magnetofon szalagra olvasás által történt. Az olvasás során az információ bizonyos egyszerűsítése és részleges kódolása is lezajlott. Ezért a felhasználó minden betegfelvétel, azaz minden *articulus* esetében előre kimondta, hogy max. 14 információt fog, mint a legfontosabbakat, megtartani. Minden információ meghatározóból (*définiteur*) és a meghatározót leíró jellemzőből (*descripteur*) áll. Az egyszerűség kedvéért rögtön idézzük is a meghatározók listáját. Valójában persze ez a lista csak az adatgyűjtés végén zárul le, minthogy bizonyos pontosabb meghatározások, vagy egy-egy új tulajdonság az utolsó pillanatban is felbukkanhatnak. (Ez egyébként az organigrammon is világosan látszik, a meghatározók listájának elkészítése ui. az adatgyűjtés szakasza után következik.) A 14 meghatározó listája a következő:

HO vagy FM	Férfi vagy Nő, tehát az egyén neme
JOUR	A felvétel napja
MOIS	A felvétel hónapja (Minthogy az év bizonyos számú személynél azonos, ezt nem kell minden esetben feltüntetni)
DIVISION	A kórház azon osztálya vagy szekciója, amelyben a kérdéses személy tartózkodott (ezek száma 1-től 14-ig)
PRENOM	a személy keresztnéve
NOM	családi neve
NIV	helye a társadalmi-foglalkozási hierarchiában
MT	mestersége
EC	családi állapota
AGE	életkora
DLE	lakhelyének egyházmegyéje
MOTIF	felvételének indoka
AUTOR	a hatóság, amely felvételre a kórházba (kódszámuk 1-től 12-ig)
DUREE	a benttartózkodás időtartama napokban kifejezve
DV	egyéb. E meghatározó jelenléte lehetővé teszi, hogy adott esetben az információ valamelyik sajátos elemét fel lehessen használni.

Az információ ilyen módon való „kezelése” az adatokat bizonyos rendszerbe csoportosítva juttatja a gépbe. Az egyik *articulus* tehát például a következő formát öltötte: HO, JOUR 8, MOIS 3, DIVISION 7, PRÉNOM PIERRE, NOM MACRET, NIV COMPAGNON, MT MENUISIER, EC CELIBATAIRE, AGE 61, DLE AMIENS, MOTIF VUE BASSE, AUTOR 7, DUREE 30, DV MORT. Nem szükséges, hogy az információk valamennyi fajtája egy-egy bejegyzésben előforduljon, s ezt a beolvasás során nem is kell külön jelezni. Az sem szükséges, hogy a meghatározó-jellemző párok mindig azonos sorrendben következzenek.

Az adatok összegyűjtése után a felhasználó összeállítja a meghatározók listáját.

A gép véletlenszerűen adja a kódszámokat. Az ésszerűbb felhasználás és a könnyebb kezelés céljából a felhasználó ezeket értelemszerű kódokkal cserélheti fel. Így pl. elhatározhatja, hogy a mesterségek kódjai 20 és 99 közé fognak esni, ezen belül is a mezőgazdasági jellegűek 89-től 99-ig kapják majd kódjelüket. Hasonlóképpen csoportosíthatja a betegségeket stb.

A munka programja ebben a konkrét esetben egyszerűen alakult, mert a felhasználó csupán táblázatok összeállítását igényelte.

B		A
nemek	0 — 15, 15—30, 30—45, 45—60, egyéb	életkor —————>
férfi		
nő		
egyéb		

Megjegyzendő, hogy a rendszer mindig számol egy-egy kiegészítő oszloppal és sorral, hogy ide sorolhassa a meghatározhatatlan egyedeket. Ilyesmi akkor fordul elő, ha a táblázat kerete túl szűk, vagy amikor a szükséges információk hiányoznak az *artículus*-ban. Ebben a táblázatban az A és a B feltétel kapcsolódó követelményeinek megfelelő egyének összeszámlálása történik. Ezután összeg-, százalék- és átlagszámítások következhetnek soronként és oszloponként, továbbá a végösszeg kiszámítása.

A felhasználó „hierarchikus” táblázatokot kívánt készíttetni. E „hierarchia” a feltételek egyre pontosabb megfogalmazásával áll elő.

1. feltétel szint: Csak a férfiakat számolja össze!

2a) feltétel szint: Csak a nős férfiakat számolja össze!

2b) feltétel szint: Csak a 0—15 év közti férfiakat számolja össze, akiknek nincs mesterségük!

3. feltétel szint: A mesterség nélküli 0—15 éves férfiak csoportosítása származási helyük szerint!

4. feltétel szint: milyen betegségben szenvedtek?

	Párizs	vidék	idegen országok
bőrbetegségek			
szembetegségek			

A feltételek e négy szintjéhez kapcsolódik egy további finomítási lehetőség is: az adatok ventilációja minden oszlopon illetve soron belül.

5. Néhány megjegyzés a FORCOD rendszerrel kapcsolatban

A FORCOD rendszer megszületését azoknak a sajátos forráskörülményeknek köszönheti, amelyek Franciaországra és más nyugat-európai országokra jellemzőek. Gyakori eset ui., hogy egy-egy levéltárban hatalmas, folyóméterekben mérhető mennyiségben találhatók tömegszerű adatokat tartalmazó források (házassági szerződések, anyakönyvek, összeírások, egyéb regiszterek). A statisztikai számításokon kívül (ezek az alaplátványokra szorítkoznak) a FORCOD másféle adatfeldolgozásra is alkalmas: kartográfiai ábrázolások, családfák, kis (házakra, utcákra, birtokokra, személyekre, családokra stb.) vonatkozó monográfiák állíthatók össze.

A rendszert megalkotói a hatékonyság növelése céljából bizonyos géptípusokhoz kötötték. Sajnos az MTA birtokában levő CDC 3300-as típusú gépre a fejlettebb FORCOD C rendszer nem alkalmazható.

A FORCOD rendszer tanulmányozása mindemellett megítélésem szerint azért hasznos, mert a történeti feladatok megoldása céljából született, tehát ezek felé „orientáltak”, az adatok gépre vitelét nagymértékben megkönnyíti, s a számítástechnikai kérdésekben járatlan kutató számára is könnyen érthető, nagyobb erőfeszítés nélkül is elsajátítható. Kifejlesztése révén azonban ma már a szűk történeti jellegű feladatok mellett gyakorlati ügyviteli stb. jellegűek megoldására is alkalmazható.

3

Kísérlet történeti-statisztikai adatokat tartalmazó kontingencia táblázatok számítógépes elemzésére

Megjegyzés: az alábbiakban e sorok írója készülő nagyobb tanulmányának egyik részmunkájáról számol be. Az ismertető eljárás H. Jankó Katalin, az Egyetemi Számítóközpont tudományos munkatársa adaptálta a konkrét anyagra, s ugyancsak ő készítette a programot is. Az egész tanulmány történeti-statisztikai számításai a „Marx Károly” Közgazdasági Egyetem Gazdaságtörténeti Tanszékének támogatásával készültek az Egyetemi Számítóközpontban. Az eljárás elméleti alapjai Ruuszman idézett tanulmányában (8), valamint Yule, G.E.—Kendall, M.G.: Bevezetés a statisztika elméletébe. Budapest, 1964. 71—77. l. található.

*

1. A források jellege

A városi társadalom tagozódásának történeti kutatásához nélkülözhetetlen alapot adnak a tömegszerű információkat tartalmazó források. Ezek statisztikai elemzése révén olyan kiinduló helyzethez juthatunk, amelyben már összefoglaló képünk van egy-egy fontos tendenciáról, például az ár- és bérvizonyok alakulásáról viszonylag hosszabb időn keresztül, vagy — bizonyos időpontban — áttekintésünk van az adott városi társadalom horizontális tagozódásáról. Ez utóbbi helyzet fényképfelvételhez hasonlítható, amelyet kedvező esetben kisebb vagy nagyobb időközökben meg is lehet ismétetni.

Kassa város társadalmi tagozódásának vizsgálatához ilyen pillanatfelvétel készítésére alkalmas forrás két jegyzék a XVI. század derekáról. E jegyzékekben házról-házra haladva írták össze a városlakókat, s bizonyítható, hogy ily módon egy térbelileg jól körülhatárolható területen az összeírás valamilyen formában szinte mindenkit érintett. A jegyzékekkel kapcsolatos legfontosabb tudnivalókat a túldoldali táblázat foglalja össze.

Három észrevételt kell tennünk a fenti táblázattal kapcsolatban. 1. A későbbi jegyzék csonka, ezért a biztosabb összehasonlítás kedvéért számításainkból most elhagyjuk az 1554-es jegyzéknek a város 4. negyedére vonatkozó részét. Így ténylegesen csak 11 utca és 370 háztartásfő kerül 1554-ből megfigyelésre. 2. A '''-gal jelölt cellákban olyan adattípusok szerepelnek, amelyek a forrásban nem szerepeltek, hanem bizonyos kritériumok alapján több osztály összevonásával keletkeztek. Hasonló az utcák negyedenként történt összevonása, ez a városi

A jegyzék neve:	1554. év	1557. év
	népesség összeírás	adójegyzék
az összeírtak száma:	457	461
negyedek száma:	1—4	1—3
utcák száma:	1—14	1—11
foglalkozások száma:	0—53	0—58
foglalkozási ágak száma:	1—10*	1—10*
háztartások létszáma:	1—24 fő	.
adóösszeg adófizetőként:	.	11-féle
adókategóriák száma:	.	1—8*

fiscus-tól származik. 3. Egyes adattípusoknál külön osztályokat nyitottunk a definiálhatatlan egyedek számára.

2. A Pearson-féle együttható

Közkeletű eljárás szerint az egyes adattípusokon belül a vizsgálandó sokaságot tulajdonságai szerint szokás csoportosítani, s a statisztikai vizsgálat ezután abból áll, hogy a tulajdonságok alapján osztályokra bontott adattípusokat kontingencia táblázatokban összevetjük. Ezek a numerikus, vagy az oszloponként illetve soronként számolt százalékos megoszlást tükrözhetik. A kontingencia táblázat általános képlete tehát a következő:

TULAJDONSÁG	A1	A2	—	—	As—1	As	ÖSSZEG
B1	(A1B1)	(A2B1)	—	—	(As—1B1)	(AsB1)	B1
B2	(A1B2)	(A2B2)	—	—	(As—1B2)	(AsB2)	B2
—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	—
Bt	(A1Bt)	(A2Bt)	—	—	(As—1Bt)	(AsBt)	Bt
	(A1)	(A2)	—	—	(As—1)	(As)	N

AmBn osztály gyakoriságát az *m*-edik oszlop és az *n*-edik sor által meghatározott cella tartalmazza; az összesítő adatok az oszlopok alján illetve a sorok végén az elsőrendű gyakoriságokat, azaz az Am-ek és a Bn-ek értékét adják. A jobb alsó sarokban levő összeg a megfigyelések teljes számát adja (N).

A szokásos eljárás ezután (amelyet természetesen mi sem nélkülözhetünk) az, hogy a táblázatot különbözőképpen 2×2 -es alakra redukáljuk, s így nyílik lehetőség arra, hogy bármelyik vagy több A és bármelyik vagy több B között vizsgáljuk a kapcsolatot. Ha ez a tárgyalási mód helyesen valósul meg, a tulajdonságok kapcsolatáról elvileg minden információt ki tudunk bontani. A gyakorlat azonban azt mutatja, hogy komoly nehézségekbe ütközünk, ha a táblázatnak akár csak egyik oldala is túlságosan hosszú. Elméletileg pedig további nehézség, hogy nem csupán az egyes A-k és az egyes B-k közötti kapcsolat érdekel bennünket, hanem arra a kérdésre is feleletet várunk, vajon függenek-e az A-k *összességükben* a B-ktől, s ha igen szoros-e ez a kapcsolat vagy laza.

A történeti-statisztikai elemzések e sarkalatos elméleti problémájának megoldására javasolta Ruuszman idézett tanulmányában a *Pearson-féle* „négyzetes kontingencia” együttható alkalmazását. Meg kell jegyezni, hogy ő az együtt-

ható kiszámítását csupán a faktor analízishez vezető eljárás első lépésének tekintik. Megítélésem szerint azonban a Pearson-féle együttható kiszámításán túl a XVIII. századnál régebbi magyar várostörténeti forrásanyag az adattípusok viszonylag csekély száma miatt nem indokolja hasonló, bonyolultabb matematikai eljárások alkalmazását.

A „négyzetes kontingencia” együttható lényegében olyan összefüggés jellemzésére szolgál, amelyben az egymástól függő ismérvek (adattípusok) jelenléte között nincsen teljes összhang, másszóval az összefüggés nem funkcionális, hanem korrelációs jellegű. Azt mutatja meg, hogy az egyik adattípus, mint tulajdonságok összessége milyen mértékben függ a másik adattípustól mint tulajdonságok összességétől. Gyakorlatilag tehát a kapcsolat erősségét vagy gyengeségét tükrözi.

Az együttható kiszámítása a

$$C = \sqrt{\frac{S - N}{S}}$$

képlet alapján történt. Ebben C a Pearson-féle „négyzetes kontingencia” együttható, N a táblázat jobb alsó sarkában levő összeg.

S definíciója:

$$S = \sum \frac{(AmBn)^2}{(AmBn)_o}$$

amiből

$$(AmBn)_o = \frac{(Am)(Bn)}{N}$$

A Pearson-féle együttható használatának komoly hátránya azonban, hogy bár elvileg az 1 határérték felé halad, ezt az értéket gyakorlatilag sohasem, elvileg végtelen számú osztály esetében veszi fel. Yule-Kendall könyvében szimmetrikus táblázatokra érvényes határértékeket közöl, amelyeket adott táblázat méret esetén az együttható sohasem haladhat meg. Így például 2×2 -es táblázat esetében a C értéke maximálisan 0,707, s csak 10×10 -es táblázatnál jut el 1 közelébe (0,949). E fogyatékoság kiküszöbölésére azonban Ruuszman tanulmányának függelékében (90. l.) egy minden szempontból kielégítő korrekciót javasol, amelynek segítségével a C értékek homogénné tehetők. Módszerét alkalmazva történt meg tehát a fenti úton kiszámított C értékek korrigálása.

3. Eredmények és értékelés

A két kassai forrás alapján a következő „négyzetes kontingencia” együtthatókat számítottuk ki:

Kassa, 1554

1. negyed	2. utca	3. fogl. ág	4. fogl.	5. család létsz.
1. <input type="checkbox"/>	0,997	0,439	0,668	0,466
	2. <input type="checkbox"/>	0,616	0,845	0,677
		3. <input type="checkbox"/>	0,999	0,587
			4. <input type="checkbox"/>	0,799
				5. <input type="checkbox"/>

Kassa, 1557

1. negyed	2. utca	3. fogl. ág	4. fogl.	5. adókat.	6. adóössz.
1. □	0,96	0,348	0,567	0,253	0,256
	2. □	0,566	0,814	0,622	0,584
		3. □	0,99	0,469	0,499
			4. □	0,517	0,727
				5. □	0,98
					6. □

Mielőtt rámutatnánk néhány, a táblázatok elemzéséből leszűrhető megfigyelésre, hangsúlyozni kell, hogy az olvasó egy kísérlet első eredményeit látja maga előtt. A kísérletre számítógép nélkül sohasem kerülhetett volna sor, hiszen a fent leírt számítások annyi munkát igényelnek, hogy „kézi” eszközökkel való elvégzésük szinte reménytelen feladat lenne.

Ami a mondott megfigyeléseket illeti, úgy véljük a referátum keretében elegendő, ha ezekkel kapcsolatban négy fontosabb tanulságot emelünk ki.

1. A két különböző évből származó jegyzék típusában is eltért egymástól, így a bennük szereplő adatok forrásértékét hosszú és körültekintő vizsgálat alapján lehet elbírálni. Ehhez fontos adalékok jelent a két táblázat összehasonlítása a foglalkozási és a topográfiai tagozódás mutatói alapján. Az 1554-ben készült lélekszámösszeírás valamennyi számbavehető *C* együtthatója szorosabb összefüggést jelez, mint az 1557-es adólajstrom alapján kiszámított megfelelő együtthatók. Azaz igazolják a más eszközökkel is alátámasztható felismerést, miszerint az 1554-es összeírás következetesebben törekedett az egyes összeírtak foglalkozásának meghatározására. Ugyanakkor a *C*-k összevetéséből az is kiderül, hogy nem koncepcionális eltérés mutatkozik a két jegyzék között. Inkább árnyalatnyi különbségről beszélhetünk, amely feltevésünk szerint azzal magyarázható, hogy a korábbi összeírás védelmi célokból készült, s nagyobb pontosságra azért törekedett, mert a foglalkozások szerinti megkülönböztetés révén a céhek, egyesületek (Kereskedő Társaság!) erejét is fel kívánta mérni. Kassán, mint általában minden középkori városban, az egyes falszakaszokat, bástyákat a különböző korporációk védték ostrom esetén. (A családnév alapján történő foglalkozás meghatározás általános szakmai kérdéseiről a referátum keretében természetesen nem beszélhetünk.)

2. A *C* értékek nagyságát vizsgálva arra a megállapításra jutunk, hogy 1,0-hoz legközelebb valamennyiük közül a foglalkozási megoszlás és az utcánkénti megoszlás összefüggését reprezentáló értékek állnak.¹⁷ E megállapítás mindkét vizsgált évre érvényes. Az is kiderül továbbá, hogy, bár kisebb mértékben, de a családlétszám is függött a foglalkozástól. Erősnek tekinthetjük a foglalkozás és a kirótt adó közti összefüggést.

E számok mindenekelőtt meghatározzák a tudományos diszkurzió menetét, amelynek a foglalkozások szerinti megoszlásból kell majd kiindulnia. Ha emlékezetünkbe idézzük a referátum első részében elhangzott megállapítást a magyarországi városi polgárság sajátos megélhetési viszonyairól, akkor most kitűnő példáját láthatjuk annak, hogy a statisztikai eredmények milyen komplex interpretációt igényelnek a történésztől. Annyi már most is megállapítható, hogy

¹⁷ Az osztályok összevonása útján keletkezett adattípusok (*negyed, foglalkozási ág, adó-kategória*) összefüggését a velük identikus, de bővebben tagolt adattípusokkal szintén közöltük a táblázatban. Ez természetesen értelmetlen összevetés, afféle tükörkép, amely a Ruuszman által javasolt korrekciós eljárás helyességét igazolja.

Kassán egyes foglalkozások (ezek elsősorban kézműves jellegűek) határozott topográfiai rendszerben szerveződtek. Tulajdonképpen bizonyos vízigenyes iparokról van szó (mészáros, kovács), amelyekhez más foglalkozások (kerékgyártó) csatlakoztak. Külön utcájuk volt a lakatosoknak is.

A tudományos diszkurzióhoz tartozik az is, hogy a C együttthatók két táblázata egyszerre tükrözi valamennyi adattípus valamennyi kapcsolatának erősségi fokát. A táblázat alapján akár grafikus model is készíthető, amely a molekulák felépítését ábrázoló rajzokhoz hasonlít.

3. Történész részről az előbbiekkal szemben azt a kapitális ellenvetést lehet felhozni, hogy számításainkat végül is az adott forrás determinálja. Valóban például az „ismeretlen” foglalkozásúak részaránya (ezeket 0-val jelöltük és számításainkban is figyelembe vettük) 1554-ben 42% , 1557-ben 46% . A foglalkozások túlnyomó része kézműves jellegű. A vizsgálandó sokaság foglalkozás szerinti megoszlását feltüntető táblázat szélső oszlopába tehát tekintélyes számú definiálhatatlan egyed kerül, s ez az oszlop tovább nem bontható. Minthogy pedig olyan, a lakosság egészét felölelő forrás, amely egy időpontban készült, s másféle szempontok alapján csoportosít, nem áll rendelkezésünkre, az „ismeretlen” foglalkozásúak csoportja csak más forrásokban elszórt kiegészítő adatok segítségével elemezhető. Távrolról sem remélhető azonban, hogy segítségével megnyugtató végeredményre jussunk. Továbbá igaz ugyan, hogy a foglalkozási, topográfiai, vagyoni és más tagozódások többé-kevésbé rekonstruálhatók statisztikai eszközökkel, a tagozódás többi eleme azonban homályban marad.

Az ellenvetés arra hívja fel a figyelmet, hogy a rendelkezésünkre álló statisztikai források elemzése révén csupán a társadalmi valóság egy részletéről kapunk felvilágosítást, amely nem is tekinthető pontosnak. S bár így alapvetően fontos információk birtokába jutunk, a tény mégis az, hogy az egykorú forrás határozza meg statisztikai munkánk menetét. Kiegészítésül azért mindehhez hozzá kell tenni, hogy fennmaradt városi összeírásaink a XVIII. századig bezárólag nagyjából a kassaiakéhoz hasonló módszerrel készültek, továbbá, hogy általában a XVI. század közepéig a bennük előforduló családnevek túlnyomó részükben nem öröklődő nevek voltak, hanem a tényleges foglalkozást vagy a származás földrajzi helyét stb. jelölték. Mindent összevetve a szóbanforgó forrástípus tág tűrési határok között a történeti Magyarország városainak társadalmi tagozódását tükrözi, de statisztikai eszközökkel e tagozódásnak csak bizonyos aspektusai szűrhetők ki belőle. Ezek az aspektusok lehetnek tehát adott esetben a statisztikai elemzés tárgyai, s csupán ezen aspektusok összefüggéseit vizsgálhatjuk az ímént leírt módszerrel is.

4. Ha a „négyzetes kontingencia” együttthatókat a továbbiakban más városok hasonló forrásai alapján is kiszámítjuk, fontos tudományos eszközhöz juthatunk. Kellő számú példa segítségével elvileg talán meghatározható lesz, hogy bizonyos erősségű kapcsolat esetén más kapcsolat vagy kapcsolatok milyen erősségűek lehetnek. Így lehetőség nyílnék bizonyos kapcsolatok szimulálására, ami további fontos előrelépést jelenthet a tudományos megismerésben (következtetések hiányzó adattípusokra, a forrás hitelességének, egyedi sajátosságainak elemzése stb.). Úgy véljük azonban, hogy erre a mostaninál alkalmasabb lenne egy olyan eljárás kidolgozása, amely nem a numerikus, hanem az oszloponként és soronként számolt százalékos megoszlásból indulna ki.

Talán nem érdektelen ebből a szempontból, ha röviden bemutatjuk egy kezdetlegesebb körülmények között végzett első kísérlet eredményeit. A C együttthatók ennek során egy másik város, Brassó, 1475. évi adójegyzékéből kerültek

kiszámításra. Sajnos a vagyoni és a foglalkozási tagozódás adattípusait csak a nagyobb sokaságot magukban foglaló *foglalkozási ágak* illetve *adó kategóriák* alapján elemeztük. (Zárójelben a megfelelő kassai adatok 1557-ből)

Brassó, 1475

1. utca	2. fogl. ág	3. adókat.
1. □	0,5001 (0,566)	0,590 (0,622)
	2. □	0,304 (0,469)
		3. □

Az összehasonlítás számos lényeges, de nem ide tartozó problémát vet fel. Ezek a két város eltérő jellegével, a mintegy 80 év különbséggel és a forrásokban tükröződő topográfiai realitások eltéréseivel kapcsolatosak. Valójában a különbségek a magyar város evolúció szélesebb spektrumába állítva másodlagos jelentőségűek. Megítélésem szerint feltűnőnek minősíthetjük azt a tényt, hogy a kassai és a brassói mutatók között nincsen lényeges eltérés. Közepes vagy gyenge összefüggéseket reprezentálnak, s a kassaiak mindhárom esetben valamivel magasabbak. Mindkét város adójegyzékénél azt tapasztaljuk, hogy a kapcsolatok erőssége ugyanabban az irányban növekszik; a növekedés iránya az 1—3., 1—2., 2—3. viszonylat. Ez azért érdekes megfigyelés, mert a foglalkozási ágak illetve az adó kategóriák megállapításánál egyforma elveket alkalmaztunk, tehát kiindulási alapunk elvileg hasonló volt.

A foglalkozások meghatározása szempontjából a kassai forrás pontosabb, mert a brassói kizárólag a kézműveseket említi, az „ismeretlen” foglalkozású adófizetők mintegy 66 $\frac{0}{100}$ -ot tesznek ki. Ez a körülmény elsősorban a vagyoni és a foglalkozási viszonyok összehasonlításakor érezteti hatását. A topográfiai tagozódás foglalkozási áganként viszont még így is szemléletesnek látszik. Az adó kategóriák, tehát a vagyoni viszonyok és a topográfiai viszonyok összehasonlítása pedig azt mutatja, hogy a két városban nagyjából hasonló volt a helyzet ebből a szempontból. Valóban, az egyes utcák között mindkét városban jelentős különbségek voltak, rangjuk volt, amit az is bizonyít, hogy bizonyos városi tisztségviselők szinte törvényszerűen egy-két rangosabb utcában, vagy az ún. „Ring”-en laktak.

Kísérletünk rövid leírása során természetesen nem szólhattunk kimerítően az eljárás valamennyi kiaknázzható eredményéről. Ez a szakmai elemzésre tartozik, amely összefoglalóan tárgyalja a kassai viszonyokat a XVI. század derekán. A kísérlet során arra keresünk választ, hogy hagyományos kontingenciátáblázatok előállításán túlmenően a számítógép kínál-e olyan lehetőséget, amely továbblépést jelent a történeti-statisztikai elemzésben. Nem állíthatjuk, hogy a kísérlet lezárult volna, de azt sem, hogy negatív eredményt hozott.