

Lyukkártyarendszerű adatgyűjtés és feldolgozás a nyelvtudományban¹

1. *Lyukkártyának* nevezzük azt a meghatározott méretű, vízszintes és függőleges sorokkal tagolt papírkarton-kártyát, amelyet arra alkalmas készülékkel vagy géppel meghatározott helyeken kilyukasztanak, s így olyan számjegyeket, számjegysorokat, betűket, betűsorokat rögzítenek rajta, amelyek révén e kártya alkalmas gépek segítségével sokféleképpen felhasználható. A lyukkártyát számjegy- és betűírásra, ill. leolvasásra, másolatok készítésére, kulcsszámok szerinti vagy betűrendes rendezésre, nyomtatásra egyaránt használják.

1.1. A lyukkártya használata a XIX. század elején kezdődött. Első gyakorlati alkalmazásaként J. M. JACQUARDnak (1752—1834), lyukkártyavezérlésű szövőgépét szokás tekinteni. Ezt 1805-ben mutatták be Párizsban. 1837-ben már vezérlésre, tárolásra és összegezésre alkalmas gépet is szerkesztettek. 1839-ben CH. BABBAGE (1792—1871) angol matematikus a szövőgép vezérlését számológépre alkalmazta.

A mai értelemben vett lyukkártyarendszert az 1880-as amerikai népszámlálás munkájában szerzett tapasztalatok alapján H. HOLLERITH 1885-ben dolgozta ki J. POWERS technikus közreműködésével. Ugyanekkor szerkesztette meg a lyukkártyarendszerhez való statisztikai és rendszerező géppark őseit. Ezt 1890-ben tökéletesítette. A tökéletesített „géppark” az alábbi egységekből állt: 1. kézi lyukasztó; 2. táblázó gép: a) leolvasó részzel, b) számoló részzel, c) osztályozó részzel. Az 1890-i népszámlálásban alkalmazták először a Hollerith-féle lyukkártya-gépeket. Hazánkban 20 évvel később használtak első ízben lyukkártya-gépeket.

A lyukkártya-gépeknek nyelvészeti célú adatgyűjtésre és adatfeldolgozásra való alkalmazása jóformán csak a legutóbbi két évtized folyamán bontakozott ki, főleg az olasz ROBERTO BUSA és a francia BERNARD QUEMADA munkássága nyomán. Az utóbbi Besançonban megszervezte a francia szókincs gépi tanulmányozásának első központját (Centre d'Étude du Vocabulaire Français), itt adtak ki első ízben rendszeresen olyan kiadványokat, amelyek a nyelvészeti célú gépi adatfeldolgozás elvi és gyakorlati kérdéseivel foglalkoznak (Cahiers de Lexicologie I—VII. [1959—65]; Bulletin d'Information du Laboratoire d'Analyse Lexicologique I—VII. [1960—64]).

1.2. A lyukkártyáknak többféle típusa van forgalomban. Vannak 21, 27, 40, 80, 90 oszlopú lyukkártyák. Egy-egy oszlopban 0-tól 9-ig terjedő

¹ A Magyar Tudományos Akadémia Nyelvtudományi Intézete és a debreceni Kossuth Lajos Tudományegyetem Matematikai és Alkalmazott Nyelvészeti Csoportja közös munkaértekezletén 1964. október 15-én elhangzott előadás. Néhány részlettel kiegészítve.

10—10 számjegyet, s ezen kívül a 0 felett még 2 sort találunk: A számjegyek részint kulcsszámok alkotására, részint betűk, jelek jelölésében egyik tényező — hármas lyukasztás esetén két tényező — meghatározására, kivételesen egy-egy betű jelölésére szolgálnak. A 0 feletti két sor lyukasztásai részint vezérlésre, részint a betűk vagy jelek másik tényezőjének meghatározására valók.

2. A lyukkártyáknak három típusa terjedt el szélesebb körben: a HOLLERITH, a BULL és POWERS kód szerint szerkesztett lyukkártya. Ezek közül is különösen az első kettő van elterjedve. Mai fejlettebb változatában már mind a HOLLERITH, mind a BULL rendszerű lyukkártyákon hármas lyukasztást is alkalmaznak, s így 39 különféle betűnek, 10 jelnek, 10 számjegynek s néhány tipográfiai jelnek megfelelő lyukasztással dolgoznak.

2.1. Mivel nálunk főleg HOLLERITH-rendszerű lyukkártyák használatosak, az alábbiakban elsősorban ezekre leszek tekintettel.

A HOLLERITH-rendszerű nálunk használt gépeken a betűk többségét két szám lyukasztásával képezzük:

	1	2	3	4	5	6	7	8	9
R (12)	A	Á	B	C	D	E	É	F	H
X (11)	J	K	L	M	N	O	Ö	P	R
O	/	T	U	Ü	V	W	X	Y	Z

Egy lyukasztás szükséges a G (6), az I (1) és az S (8) betűhöz.

E szerint az egyes betűk lyukasztása a lyukkártyán a 345. lapon látható képet mutatja:

3. Ha ebben a rendszerben kell leírniunk, illetőleg lyukkártyára lyukasztanunk a magyar szöveget, akkor a *cs, dz, gy, ly, ny, sz, ty, zs* hangoikat két-két, a *dzs (ž)*-t három betű jelével jelölhetjük éppen úgy, mint irodalmi nyelvünk szokásos helyesírásában. Ezek a betűk kiírásakor (felíratásakor) a szokásos alakban jelentkeznek.

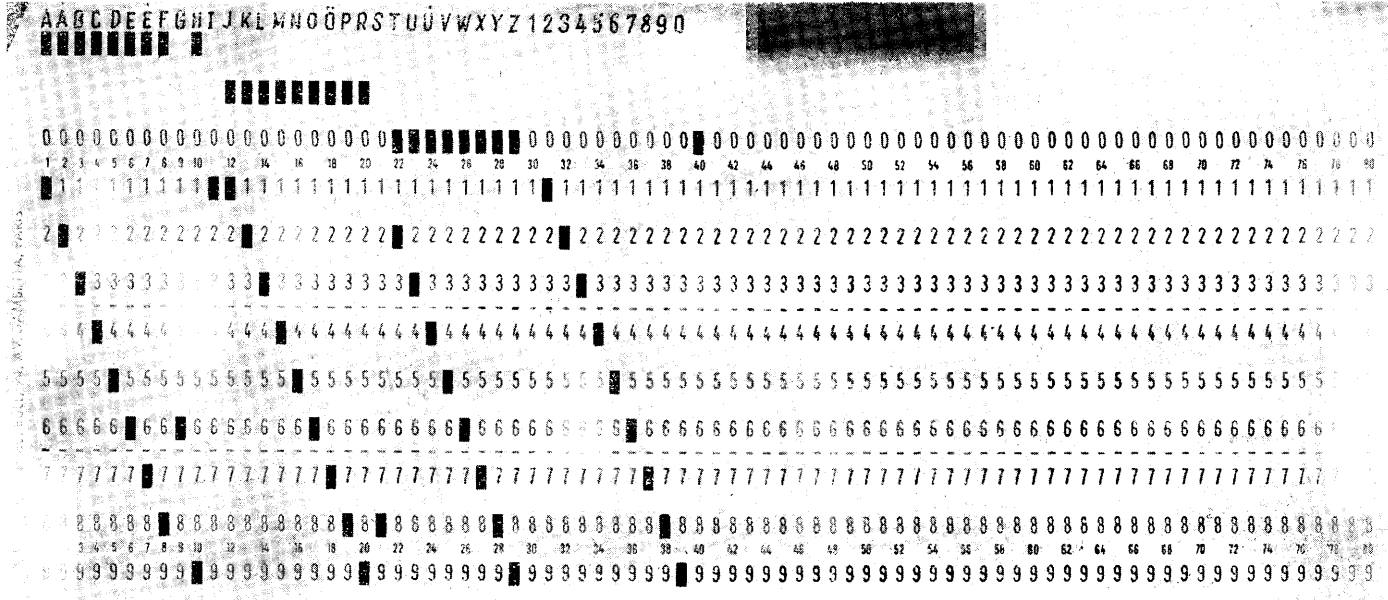
3.1. Nehezebb a helyzetünk az interpunkció, az idézőjelek, a hosszú *í, ó, ő, ú, ű* kódolásakor, valamint a tipográfiai különbségek érzékeltetésében.

Ezeknek a sajátságoknak kódolására ebben a tervezetben egyelőre számjegyeket és számjegyek kombinációit használtam annak a megfontolásnak alapján, hogy irodalmi, sajtónyelvi szövegekben, magán- és közéleti írásos érintkezésünkben a betűk és számjegyek általában nem érintkeznek közvetlenül egymással: vagy kötőjel, vagy pont vagy térköz van a betű és a számjegy, illetőleg a számjegy és a betű közt. Ha tehát a lyukkártyának betűlyukasztásra szánt mezéjén betű és számjegy közvetlenül egymás szomszédságában áll, ez a helyzet arra figyelmeztet, hogy a számjegyet nem számértékben, hanem meghatározott jelértékben kell vennünk. E jelértékben vett számjegyek közt azonban nem szerepelhetnek az 1., 6., 8. számjegyek, mert ezek az I, G, S betűk jeleként is használatosak, jel-értékben való használatuk tehát zavarokat idézhetne elő.

3.2. A továbbiakban abból indultam ki, hogy a 2 számjegy és a kettőspont összekapcsolása könnyen megjegyezhető, tehát a kettőspont jelét a megelőző szó utolsó betűjéhez szorosan csatlakozó 2 számjegyet használtam. Mivel pedig eléggé elterjedt

	A	Á	B	C	D	E	É	F	G	H	I	J	K	L	M	N	O	Ö	P	R	/	S	T	U	Ü	V	W	X	Y	Z	
R (12)																															
X (11)																															
0																															
1																															
2																															
3																															
4																															
5																															
6																															
7																															
8																															
9																															

Az egyes betűk lyukasztásának elhelyezése a nálunk használatos lyukkártyagépeken



Egy lyukkártya képe, természetes nagyságban, a nálunk elterjedt betűlyukasztási rendszerrel és a számjegyek lyukasztásával, a lyukkártya felső szélén a lyukasztásnak megfelelő betűvel, illetőleg számjeggyel

átírási rendszer szerint a kettőspont a magánhangzó hosszúságának jele, a hosszú *i, ó, ő, ú, ü* jelül szintén a rövid *i, o, ő, u, ü* betűhöz szorosan csatlakozó 2 számjegyet használtam: $i2 = \acute{i}$, $o2 = \acute{o}$, $ő2 = \acute{ő}$, $u2 = \acute{u}$, $ü2 = \acute{ü}$. Így egyetlen eljárással hat jel hiányát hidalhatjuk át. (Ez az eljárás csak abban az esetben okoz zavart, ha rövid *i, o, u, ü*, vagy hosszú *i, ó, ú, ü* után kettőspont következik, ez azonban rendkívül ritka; ilyenkor rövid magánhangzó után hézag + 2 számjegyet lyukasztunk; hosszú *i, ó, ő, ú, ü* értékben pedig a kettőspontot a szó utolsó betűjéhez (i, o, ő, u, ü-höz) tapadó 22 számjeggyel helyettesítjük: ebből az első 2 a magánhangzó hosszúságát, a második a mondat-tagoló kettőspontot jelöli.

3.3. Hasonló módon jelölhetjük a többi interpunkciós jelet a következő számjegyekkel:

.	= 3
...	= 333
,	= 4 (vesszőként a szó utolsó betűje után; fél idézőjelként az idézet első betűje előtt és utolsó betűje után)
;	= 34
?	= 5
!	= 7
?!	= 57
!?	= 75
-	= 23 (kötőjel és elválasztójel); (sorvégi elválasztójel: 2323, ha kiírása szükséges)
-(nagy-kötőjel és gondolatjel)	= 24; bekezdés élén 5924
„ (idézőjel)	= 44 (az idézet első betűje előtt és utolsó betűje után)

3.4. A nagy kezdőbetűt a betű elé tett 9 jelöli, a verzális szedésű szót, szórészt vagy szövegrészt pedig az első betű elé és az utolsó után tett 52–52 kettős számjegy; így a ritkított szedést 53–53, a félkövéret 54–54, a kövéret 55–55.

A bekezdés jele: 59. Ilyenkor a nagy kezdőbetű nem kap külön jelet; a bekezdés és gondolatjelé, párbeszédben 5924.

3.5. A kerek zárójel () jele: kezdő 27, záró 72, a szögletes []: kezdő 29, záró 92.

3.6. Az írásjeleket és tipográfiai jeleket helyettesítő számjegyek után egy betűnyi hézag hagyandó, — kivéve a kezdő idézőjelet (4,44), a kezdő zárójelet (27, 29), a kezdő gondolatjelet (24), a nagy kezdőbetű jelét (9) és a tipografizálás kezdő jelét.

3.7. A kísérletek idejére megfelelőnek tartom ezt az átírási rendszert, mert a rendelkezésre álló szám-, illetőleg lyukasztás-kombinációs rendszerben egyértelműen átírhatjuk vele modern irodalmi szövegeinket, a helyesírási és a tipográfiai sajátosságokat is pontosan tükrözve. Sőt némi átalakítással és kiegészítéssel a kódexirodalom vagy a nyelvjárási szó- és szövegművelek feldolgozására is alkalmazható a Hollerith-rendszernek ez a nyelvészeti szempontból kezdetleges változata.

3.8. Kollektív munkák nagyobb létszámú szerkesztőségének egyes tagjai azonban bizonyára mindvégig idegenkedni fognak efféle szükségmegoldásoktól. A kísérletek utáni időre, a nagy kollektív munkák cél-gyűjtésének idejére feltétlenül kívánatosnak tartom olyan géppark létesítését, amely legalább 64 jeltől álló betű—írásjel—számjegy jelelemet tartalmaz, és változtatás nélkül alkalmas a magyar helyesírás és a főbb tipográfiai különbségek hű visszaadására. Ilyen géppark ma már mind a Hollerith-, mind a Bull-rendszerű lyukkártyákhoz beszerezhető.

Ilyen géppark egységei mind bemenetkor, mind kimenetkor a következő betűket, számjegyeket és jeleket (összesen 64 elemet) tartalmazhatnak:

a) betűk: a, á, b, c, d, e, é, ë, f, g, h, i, í, j, k, l, m, n, o, ó, ö, ő, p, q, r, s, t, u, ú, ü, ű, v, w, x, y, z (36); esetleg a régi nyomtatványokra való tekintettel ezeken kívül még: f, ő, ű vagy: f, β, e, ebben az utóbbi esetben diakritikus jelként (összesen 39);

b) számjegyek: 1, 2, 3, 4, 5, 6, 7, 8, 9, 0 (összesen 10);

c) jelek: . , ; : ? ! ' - / () []

* [a betű előtt, a verzális jeleként]

◇ [a tipografiaiilag kiemelt szó, szóréssz, kifejezés, hosszabb szövegrész első betűje előtt és utolsó betűje után] (összesen 15).

A HOLLERITH-rendszerű lyukkártya egyik korszerűbb változata hármas lyukasztást is tartalmaz

Ilyen modernebb gépparknak megteremtésében kívánatos lesz áttérni a nagybetűs írásról a kisbetűs írásra. Nemcsak azért, mert megszoktuk, hanem azért is, mert a kisbetűk nagyobb mértékben különböznek egymástól, mint a nagybetűk és így könnyebben olvashatók.

3.9. A fényképező eljárások újabb fejlődése lehetőséget ad olyan közvetítő megoldásra is, amely szerint a szövegnek a lyukkártya egyik részére való fényképezése (pl. xerox-másolata) adja a szót vagy a vizsgálat tárgyát képező egyéb nyelvi jelenséget, megfelelő szöveggörnyezettel együtt; a lyukkártya másik részének számoszlopai az anyagrendezést biztosítják.²

4. A lyukkártyák lyukasztása magában is többféle rendezést tesz lehetővé, így pl. a betűrend szerinti anyagrendezést balról jobbra (nyelvünkben és sok más nyelvben ez a szokásos betűrend) és — megfelelő eljárással — jobbról balra („a tergo” betűrend, megfordított betűrend, a szövegről a szó eleje felé haladó betűrend, képes kifejezéssel élve: „rák” betűrend); a szókezdetek, szövegek, szóbelseji hangkapcsolatok szerinti rendezést.

4.1. De a lyukkártya számoszlopai, megfelelő mezőkre osztva rendkívül sokféle egyéb szempontból való rendezhetőséget biztosítanak a tervszerűen előkészített szövegelemzés számára.

4.1.1. Annak érzékeltetésére, hogy fő tárgyunk, a folyamatos szövegből való szó- és morfémagyűjtés és feldolgozás hogyan viszonylik a lyukkártyákon végezhető egyéb gyűjtésekhez és feldolgozásokhoz, bemutatom a lyukkártya-típus kódolásának lehetséges módját az 1. és 2. oszlopon: az 1. oszlopon aszerint, hogy milyen forrásból származó anyagot viszünk lyukkártyára, a 2. oszlopon aszerint, hogy milyen célt szolgál a lyukkártyatípus.

4.1.1.1. Az első oszlop lehetséges kódolása: 1. Beszélt nyelvből (hangszalagról, hanglemezről) származó nyelvi anyag. 2. Szótárból való nyelvi anyag. 3. Nyelvtanból vett anyag. 4. Kézírással szövegből vett adat. 5. Nyomatott szövegből vett adat. 6. Nyelvtudományi részlettanulmányból származó anyag. 7. Bibliográfiai műből (szó-, név-, tárgymutatókból) származó anyag.

4.1.1.2. A második oszlop lehetséges kódolása: 1. Szövegrész-kártya (az alapul vett főszövegről). 2. Szövegváltozat-kártya. 3. Szóalakkártya. 4. Cím-szó-kártya. 5. Cím-szó-változat-kártya (szótári utaló kártya). 6. Morféma-kártya. 7. Jelentés-kártya. 8. Mondattani (szintaktikai) kártya. 9. Hangtani (fonetikai, fonológiai) kártya. [0. Bibliográfiába szánt kártya.]

² A nyelvészeti kutatások mechanizálásának és automatizálásának kérdéseiről Prágában 1966. jún. 9-én tartott nemzetközi értekezleten A. DURO, az olasz nyelvtörténeti szótár szerkesztőségnek igazgatója mutatott be ilyen kombinált lyukkártya-tervet: ezen összesen 24 oszlopnyi helyet biztosítanak részint a címszó v. szóalak (lemma) lyukasztására (15 oszlop), részint a legfőbb rendezői szempontokéra (9 oszlop); a lyukkártya fennmaradó részére xerox-eljárással másolják a szövegrészletet a szerző nevével, a mű címével, keletkezésének és kiadásának évszámával, lapszámával, a szövegrész sorszámával. E DURO-féle kombinált lyukkártya egyik példányának kicsinyített fényképmásolata a 349. lapon látható.

1441. DATI Leonardo. Scena,
in Certame coronario (1952)

v.

- 35 a virtù e lode! Oh! merto, per qual ne sarai
grato a' celicoli, quasi d' officina Minerva,
udite, o studiosi: èmmi l' ascoudita cosa
qual cercate voi, parmi. scoperta, patente,
meglio che ancora nullo movimento riveli.
- 40 Tra li celesii, del Nilo e Pallade figlia,
una Dea escelle, che formosissima vince,
o non men che Venus, tanta sua forma decora;
passeggia il cielo, mo sopra dove l' arduo fende
Eridano, mo donde al cielo le Pleiadi sotto
- 45 cinsero, mo donde gli Dei la sguardano tutti
vestita e nitida, distinta in mille colori
dell' iris succinta, il che suo lembo ritesse
di gemme e d' oro lustro non men che l' Orion
e circuntesta è d' ogni mirabile fatto.
- 50 Questa (suo ufficio) manifesta l' aurea porta
dell' Oceano a quelli alipedi che 'l putto superbo

4.1.1.3. Ennek megfelelően tehát a folyamatos szövegből gyűjtött szövegkártya típuszáma 41 (kézírásos szövegből) vagy 51 (nyomtatott szövegből); az ilyen eredetű szóalakkártya típuszáma 43 vagy 53, a címszókárttyáé 44 vagy 54, a szótári célú utalókártyáé 45 vagy 55, a morfémakártyáé 46 vagy 56, a jelentéskártyáé 47 vagy 57, a mondattani kártyáé 48 vagy 58, a hangtani kártyáé 49 vagy 59 (a forrás természete szerint).

4.2. A folyamatos szövegből való szó- és morfémagyűjtés (valamint a szövegelemzéssel kapcsolatos egyéb gyűjtések) alapja az úgynevezett *s z ö v e g k á r t y a*. Ennek első 16 oszlopán azok a lyukasztások szerepelhetnek, amelyek részint a folyamatban levő munka megkülönböztetésére valók (pl.: vállalati szám, a lyukkártyatípus száma), részint a forrásjelzést, az évszámot tartalmazzák. A 17–80 oszlop a szövegrész lyukasztására és a szótári rendezés kódszámaira van fenntartva. A szövegrész egyes kártyái sorszámot kapnak, s annak alapján a szükséges szövegelőzménnyel vagy szövegfolytatással az egyes részletek gépi rendezéssel kiegészíthetők. A szóalakkártyák szótári rendezésének kódszámai (74–80 oszlop) a címszó sorszámát és esetleg a segédkártya sorszámát jelölik.

4.3. Elkészítettem a folyamatos szövegből való szó- és morfémagyűjtés és -rendezés alapjául szolgáló *s z ó l a k k á r t y a* ideiglenes általános lyukkártya-tervét.

Ez a terv tartalmazza a lyukkártya számoszlopainak olyan elv szerinti mezőkre osztását, hogy a szövegnek a lyukkártyán elemzett szóalakját lehetőleg sokszempontú lexikológiai-morfológiai vizsgálatnak vethessük alá, figyelembe véve a szó vagy a szóalak esetleges stilisztikai kötöttségét, szakmai nyelvi használatát, a mondatbeli meghatározó tényezőkből is annyit, amennyi a szó szófaji értékeléséhez, a szóalak jelentésének pontosabb meghatározásához vagy a morféma szabatosabb meghatározásához támpontul szolgál.

4.3.1. A kísérletek idejére a következő szempontok szerinti lyukkártyamezők megkülönböztetését javaslom:

- | | |
|---|--|
| 1— 2 Munkaszám (vállalati v. a lyukkártya-típus száma) | } Esetleg: 3—6 szerző és összegyűjtött műveinek kötettszáma; 7—15 egyéb rövidítés a mondat, a sor sorozatával, az évszámmal. |
| 3— 6 Szerző, mű.
7 Kötet. | |
| 8—10 Lapszám. | |
| 11— 5 A lyukkártya azonossági száma (a fel-
dolgozásra kerülő művön belül). | } |
| 16 Kiejtés: ha a szöveg utal rá; szótagok száma. | |
| 17— 8 Szóalak hossza: betűszám. | |
| 19 Homonimák sorszám. | |
| 20—39 Szóalak. | |
| 40— 2 Szófaji minősítés (a szófaji alosztály
megjelölésével). | |
| 43— 7 Fogalomkörü jelzés; jelentés-elemzés. | |
| 48 Igekötő; előképző, előjel. | |
| 49—50 Szóismétlés; ikerítés, összetétel. | |
| 51— 3 Szótő. | |
| 54 „Kötőhangzó” („tővéghangzó”, „elő-
hangzó”) (képzők előtt). | |
| 55— 8 Képző; képzőbokor, képzősor, képző-
szerű szóvég. | |
| 59 A kódolásból kihagyott többszörösen
képzett származék (külön lyukkártyá-
kon kerül elemzésre a képzők szem-
pontjából). | |
| — (Csak névszók, egyes partikulák
kártyáján: 60—68) — | — (Csak igék kártyáján:
60—68) — |
| 60 „Kötőhangzó” (a névszói jelek előtt). | 60 „Kötőhangzó” (módjel
előtt). |
| 61— 2 Névszói jelek
(-é, -k, -i, -ik, -b v. -bb) és kapcsola-
taik. | 61—2 Módjelek. |
| 63 „Kötőhangzó” (személyrag előtt). | 63 „Kötőhangzó” (időjel
előtt). |
| 64— 5 Személyrag (+ -é; -éi). | 64 Időjel. |
| 66 „Kötőhangzó” (esetrag előtt). | 65 „Kötőhangzó” (személy-
rag előtt). |
| 67— 8 Esetrag. | 66—8 Személyrag. |
| 69—70 A műre vonatkozó főbb adatok. | |
| 71— 2 A szövegkörnyezetre vonatkozó főbb
adatok. | |
| 73 A szóra vonatkozó főbb adatok. | |
| 74— 5 Stílusrövidítés, földrajzi, társadalmi
kötöttség. | |
| 76— 7 Szakmai minősítés, szaknyelvi haszná-
latra utalás. | |

- 78 Etimológia.
 79 Mondatrész-érték; szintagmatag-érték.
 80 Mondatjelentéstani mozzanat.

4.3.2. A 60–68. oszlopok párhuzamosítása azon az elven alapul, hogy a rendezésben a szófaji értékek szerinti rendezés megelőzi a jelek és a ragos alakok rendezését, tehát ugyanazok az oszlopok mind névszói, mind igei jelek, ragok és előhangzók (tővéghangzók, kötőhangzók) kódolására felhasználhatók (vö. 4.19.).

4.3.3. Sőt ugyanezen az alapon esetleg a 48–59 oszlop (igekötős, előjeles, ismétléses, ikerített, összetett, képzett szók oszlopa) is párhuzamosítható.

4.4. Pl. ennek a mondatnak: „*Csibe felduzzasztotta a száját . . .*” (Móricz Zsigmond: *Csibe*. Elb. VI. 1936/55. 508) egyes szavai a következő oszlopokon szerepelnek:

1. *Csibe*: a 16. oszlopon mint két szótagú, a 17–18.-on mint 5 n hosszúságú, a 20–39.-en mint betűlyukasztásos szóalak, a 40–2.-en mint a főnév, illetőleg a tulajdonnév egyik alfaja: személynek nem keresztnévi (személynévi) eredetű neve, a 43–7.-en mint leány (neve), az 51–3.-on mint hangutánzó-állathívogató tő (*csib-*), az 55–8.-on mint mondat-szóból -e képzővel képzett névszó, a 78.-on mint a magyar nyelv külön életében keletkezett hangutánzó tőből képzett származék-szó (név), a 79.-en mint alany, a 80.-on mint érzelmileg színezett mondat szava (neve);
2. *felduzzasztotta* szava szerepel a 16. oszlopon mint 5 szótagú, a 17–18.-on mint 15 n hosszúságú, a 20–39.-en mint betűlyukasztásos szóalak; a 40–2.-en mint ige, megjelölve az alosztályt is (műveltető), a 43–7.-en mint emberi cselekvés, magatartás kifejezője, a 48.-on mint igekötővel alakult származék, az 51–3. oszlopon mint hangfestő szótő (*duzz-*), az 54.-en mint képző előhangzójának v. tő véghangzójának v. kötőhangzójának (*a*) példája, az 55–8. oszlopon mint az -sz műveltető képző példája; a 62.-en mint kötőhangzójának v. időjel előhangzójának v. toldalék-véghangzójának (*o*) példája, a 64.-en mint a -tt múlt idő jel példája, a 66–68. oszlopon mint tárgyas ragozású egyes sz. 3. személyű -a személyrag példája; a 78.-on mint a magyar nyelv külön életében keletkezett hangulatfestő szó; a 79.-en mint állítmány; a 80.-on mint érzelmileg színezett mondat szava;
3. az *a* névelő a 16. oszlopon mint 1 szótagú, a 17–18.-on mint 1 n hosszúságú szóalak, a 19.-en mint homonima sor egyik tagja, a 20–39. oszlopok közül az elsón mint betűlyukasztásos szóalak szerepel; a 40–2. oszlop megfelelő lyukasztású helyén mint személyragos főnév névelője; a 78.-on mint a magyar nyelv külön életében keletkezett szó; a 80.-on mint érzelmileg színezett mondat szava;
4. a *száját* szóalak a 16. oszlopon mint 2 szótagú, a 17–18.-on mint 6 n hosszúságú, a 20–39.-en mint betűlyukasztásos szóalak szerepel; a 40–2.-en mint főnév egyik alfaja (egyedi köznévi), a 43–7.-en mint emberi (és állati) testrész neve, az 50–3.-on mint tőváltozat (*száj-*), a 65.-en mint személyrag változata (*-á-*), a 67–8.-on mint (*-t*) ragos alak, a 78.-on mint finnugor eredetű szó, a 79.-en mint tárgy (mondattárgy), a 80.-on mint érzelmileg színezett mondat szava (szóalakja).

4.5. A bemutatott példán is láthatjuk, hogy a szó- és morfémagyűjtést és rendszerezést érdemes lesz összekapcsolni; a sok szempontú szövegelemzés mind a lexikográfia és lexikológia, mind a morfológia számára rendkívül gazdag

és használható anyagot ígér, s a kettős feldolgozás sok szempontból támogatja egymást.

4.6. A szóalakkártya első 15 oszlopa a lyukkártya azonosítására és erősen rövidített forrásjelzésre való. Ebben a keretben kellene biztosítani a helyet a feldolgozásra kerülő mű évszámának is, esetleg csak 3 jegyű számmal (az ezres elhagyásával), amint a franciák teszik. A nagymérvű rövidítés ugyanis nem ad elég kronológiai fogódzót, s a rövidítésnek minden adatnál való visszakeresése az évszám megállapítása céljából túlságosan megterhelné a gépi adatrendezést vagy az adatok felhasználóját.

4.7. A 16. oszlop kiejtés-rovatának (a 0 számjegyek) beállítására azért gondolok, mert kivételesen irodalmi szövegekben is történik megjegyzés vagy utalás a kiejtésre (főleg idegenszerű, nyelvjárási vagy beszédhibás megnyilatkozások esetén). Azonkívül ezzel az oszloppal egységes keretbe foglalhatjuk a szótelemzést, a szótári elemzést (l. alább 8.), a nyelvtani elemzést (l. alább 9.) és nyelvtudományi részlettanulmányok (l. alább 11.) eredményeinek idevágó részleteit.

4.7.1. Azokról a lyukkártyákról, amelyek kiejtésjelölésre utalnak, a 39. oszlopig másolat készítenőd, s a kiejtés ezeken külön kód szerint elemezhető.

4.7.2. A 16. oszlop 1—8 számjegye a szóalak szótagjainak számát jelöli; „9” a kilenc szótagú és kilencnél több szótagú szavak együttes jelölésére szolgál. Ezek száma nem lesz olyan nagy, hogy elemzésükre egy további oszlopot igénybe kellene vennünk.

4.7.3. Hogy a szótaghosszúság elemzésére elég egy oszlop (1—9 számjegy lyukasztásával), s a „9” lyukasztás a 9 és 9-nél több szótagú szavak jelölésére gyakorlatilag alkalmas, arra meggyőző bizonyítékot találunk NEMES ZOLTÁN Szóstatisztika (1939/41: 328—353) c. művében: a 0—8 (helyesebben 1—8) szótagú szavak százalékos aránya a 9 és 9-nél több szótagú szavakhoz viszonyítva 99,94 : 0,06. Ez más szóval azt jelenti, hogy 10 milliónyi szóból valószínűleg 6000 szót kell külön vizsgálni a szótaghosszúság szempontjából, ha a 9 és a 9-nél több szótagú szavak megoszlási arányát is vizsgálni akarjuk.

4.7.4. A 16. oszlop tehát egyrészt megbízható statisztikai adatokat biztosít a szöveg szóalakjainak szótagokban mért hosszúságáról, a szótagok számáról, másrészt nyersanyagot nyújt a szótagok minőségi elemzéséhez, adalékokat a kiejtés vizsgálatához.

4.8. A 17—18. oszlop (Szóalakhosszúság: n-ekben) részint a szóalak írásképeinek — ezen belül áttételesen — a szóalak egyik fontos jellemzőjének tanulmányozásához kíván elvszerűen megvizsgált anyagot biztosítani, másrészt segédeszközül is szolgál azokra az esetekre, amelyekben a szóalak több mint 20 n-nyi hosszúságú, tehát a gépi anyagrendezésben kiegészítő kártyák alkalmazását teszi szükségessé (pl. *rabszolgakereskedelemmel*: 24 n hosszúságú szóalak).

4.8.1. Célszerűnek látszik a szóhosszúságnak n-ekben való megjelölése. Tehát pl.

1 : 1 n hosszúságú szóalak, pl. *a* (mutató névmás, névelő, indulatszó),
e (mutató névmás), *ó*, *ő*;

2 : 2 n hosszúságú szóalak, pl. *én*, *őt*, *te*, *-e* [kötőjel és kérdő *e*];

3 : 3 n hosszúságú szóalak, pl. *azt*, *fel*, *haj*, *háj*, *így*, *oly*;

4 : 4 n hosszúságú szóalak, pl. *alma*, *atya*, *biró*, *gömb*, *öreg*;

5 : 5 n hosszúságú szóalak, pl. *agyar*, *ágyat*, *bátor*, *korom*, *ujjam*;

6 : 6 n hosszúságú szóalak, pl. *asztal, kormot, növény, rántsd*;
 20 : 20 n hosszúságú szóalak, pl. *igazságtalanságairól, műkorcsolyázó-baj-
 nok*.

4.8.2. Elképzelhető olyan kódolás is, amely az n-ekben mért szóalakhosszúságnak és a hangokban mért szóalakhosszúságnak kombinációján alapul. Pl.

- 1 : 1 n hosszúságú, 1 hangból álló szóalak, pl. *a, e, ó, ő, s*;
- 2 : 2 n hosszúságú, 1 hangból álló szóalak, pl. *sz*;
- 3 : 2 n hosszúságú 2 hangból álló szóalak, pl. *én, őt, te*;
- 4 : 3 n hosszúságú 2 hangból álló szóalak, pl. *ész, így, szó, szb*;
- 5 : 3 n hosszúságú 3 hangból álló szóalak, pl. *azt, fel, haj*;
- 6 : 4 n hosszúságú 3 hangból álló szóalak, pl. *begy, gyár, zsír*;
- 7 : 4 n hosszúságú 4 hangból álló szóalak, pl. *alma, haja, őzek*;

Kérdés azonban, efféle kódolás nem jelent-e annyi idővesztést és munkaerő felhasználást, hogy nem érdemes a kombinációk figyelembevételére, mert az n-ekben mért hosszúság szerint rendezett szóalakok gépi rendezésével rövidebb idő alatt elvégezhető az n-ekben mért azonos hosszúságú szóalakok hangtani elemzése.

4.8.3. Még kevésbé lenne célszerű a szóalakhosszúságnak olyan kódolása, amely a kísérleti időszak alatt bevezetendő szükség-jelölésből adódó hosszúságnövekedést is figyelembe venné. Pl. így:

- 1 : 1 n hosszúságú, 1 hangból álló, 1 pozíciós szóalak, pl. *a, e, s*;
- 2 : 1 n hosszúságú, 1 hangból álló, 2 pozíciós szóalak, pl. *o2, ő2*;
- 3 : 2 n hosszúságú, 1 hangból álló, 2 pozíciós szóalak, pl. *sz*;
- 4 : 2 n hosszúságú, 2 hangból álló, 2 pozíciós szóalak, pl. *az, én*;
- 5 : 2 n hosszúságú, 2 hangból álló, 3 pozíciós szóalak, pl. *őt2 ('őt')*;

4.8.4. Hogy 20 n-nél több helyet nem kell fenntartani a szóalakok elemzésére, arra következtethetünk NEMES (i. m. 328, 353) adatainak közvetett tanulságából. Az átlagos szótaghossz nyelvünkben eddigi eredményeink szerint ugyanis 3 alatt van (kb. 2,5 n), NEMES adatai szerint pedig a hétnél több szótagú szavak száma újságnyelvi szövegben nem több mint 0,28 %, a nyolcnál több szótagúaké pedig csak 0,06% (i. h.), azaz pl. 10 000 000 adatból mindössze 28 000 hétnél több szótagú s csak 6000 nyolcnál több szótagú szóalakot várhatunk, a 99,72%—99,94%-ra (9 972 000, ill. 9 994 000 adatra) elég lesz a 20 n-nyi terjedelem, a 0,28%-nyi, illetőleg 0,06%-nyi vagy ennél is kevesebb mennyiségű szóanyag másodkartyák bevonásával elemezhető.

4.9. A 19. oszlop a homonimáknak sorszám szerinti megkülönböztetésére szolgál. A homonimák jelöléséhez később külön segédletet fogok kidolgozni, amely lehetőség szerint áthidalja a különféle szótárak közti ellentmondásokat, s amely az etimológiai és a szinkronikus összefüggéseken alapuló homonimafelfogást az esetek jelentős részében összeegyezteti.

A homonimák sorrendjének kérdésében főleg gyakorlati szempontból azt javaslom, hogy ne gyakoriság alapján, hanem elsősorban szófaji különbség és szófaji alosztályok alapján foglaljuk sorozatba a homonimákat, természetesen tekintetbe véve a több szófaji értékű szavakat is. Tehát pl.

- 1. *ad* ige, 2. *ad* (latinosan szerkesztett határozó viszonyító szava, a ragtalan szó előtt); 1. *áll* ige, 2. *áll* főnév; 1. *akar* ige, 2. *akar* 'akár' kötőszó; (1.) *az* névmás; (2.) *az* a névelő; 1. *bukó* melléknév és főnév, 2. *bukó* melléknév; (1.) *egy* számnév, melléknév, főnév és névmás, (2.) *egy* határozószó, (3.) *egy* névelő; (1.) *fokos* főnév, (2.) *fokos* melléknév; (1.) *föl* főnév, (2.) *föl*

határozószó; 1. *hány* ige, 2. *hány* névmás; 1. *hív* ige, 2. *hív* melléknév és főnév; (1.) *hogy* határozószó; (2.) *hogy* kötőszó; 1. *hők* 'hüvelyk' főnév, 2. *hők* mondatszó; 1. *kenet* ige, 2. *kenet* főnév; 1. *ki* névmás, 2. *ki* határozószó és igekötő; (1.) *mind* névmás, határozószó és főnév, (2.) *mind* kötőszó; (1.) *mint* határozószó, (2.) *mint* kötőszó; 1. *ó 'óv'* ige, 2. *ó* főnév, 3. *ó* melléknév, 4. *ó* indulatszó; (1.) *otthon* főnév, (2.) *otthon* határozószó; 1. *ör 'öröl'* ige, 2. *ör* főnév; (1.) *pulya* 'gyermek' főnév, (2.) *pulya* 'elpuhult' melléknév és főnév; 1. *szív* ige, 2. *szív* főnév; 1. *tejszín* főnév (tejnek gazdag zsírtartalmú része), 2. *tejszín* főnév és melléknév (színnév); 1. *vált* ige, 2. *vált* melléknévi igenév; 1. *zsupp* 'zsúp' főnév, 2. *zsupp* mondatszó.

Ezek legnagyobb részében, javaslatomra, az ÉrtSz. is a szófaji érték alapján foglalja sorozatba a homonimákat.

A kódolónak mindenestre könnyebb állást foglalni a sorrend kérdésében a szófajoknak meglehetősen hagyományos sorrendje alapján, mint akár gyakoriság, akár történeti sorrend, akár etimológiai hovatarozás vagy egyéb szempont alapján.

Azonos szófaji értékű homonima lényegesen ritkább nyelvünkben. Ezek sorrendjét a szónak általános nyelvbéli beágyazottsága alapján igyekszem megállapítani és kódba foglalni. Pl. 1. *avar* főnév (lehullott lomb az erdőben), 2. *avar* (nép); 1. *rost* főnév (növényi, állati), 2. *rost* 'rostély'; 1. *szín* 'Farbe', 2. *szín* (épület), 3. *szín* (színpad; színmű része, színpadi előadás).

4.10. A 20—39. oszlop a betűkkel írt, illetőleg betűket jelölő számkombinációval lyukasztott szóalakot tartalmazza úgy, ahogyan a szövegben előfordul.

A szóalak betű- és hangstatisztikai elemzéséhez, a szóalak hosszúságának elemzéséhez, a morfémaelemzéshez erre az alakra van szükség, azért célszerűbb a szöveg áttételekor a szóalakot tekinteni az elemzés egységének, s azokban az esetekben, amelyekben a szóalak (vagy annak eleje) nem egyezik a szótári címszóval, a lyukkártyáról másolatot lehet készíteni s a másolaton a szótári címszót lyukasztani. Erre különösen akkor van szükség, ha az igekötős ige igekötője az ige mögé kerül vagy előtte áll ugyan, de nem közvetlenül, hanem úgy, hogy közte és igéje közt egy másik szó is van.

4.11. A 40—42. oszlop (Szófaji minősítés) az elemzett szöveg szófaji kategóriáin és azok statisztikai megoszlásán kívül feleletet ad a szófaji alkategóriák (pl. a névmási alosztályok, az igenevek alosztályai) nyelvi és statisztikai megoszlására, valamint az egyes szófajok belső tagozódására vonatkozólag. Pl. az „ige” kategóriáján belül:

101 létege: létezés és hiánya (a *van* ige paradigmatis alakjai a „van” és „vannak” nélkül); *nincs* (~*nincsen*; *nincsenek*), *sincs* (~*sincsen*; *sincsenek*);

102 *van*, *vannak* (jelentő módú jelen időbeli egysz. és tbsz. 3. sz.) (a nominális állítmányú mondatra való tekintettel célszerű külön gyűjteni);

103 állapotban levést, maradást jelentő ige, tn (*alszik*, *áll*, *ül*, *él* stb.; *betegeskedik* stb.);

104 ugyanaz, ts (*átalussza* az éjszakát; *állja* a sarat stb.).

4.11.1. Az állandósult szókapcsolatokról készíthető másolattal esetleg megjelölhetjük az egyes szófajokra jellemző szókapcsolatokat. Pl. 010 igei (értékű) szókapcsolat, mint *mozgásba hoz* 'elindít; megmozgat'; 020 főnévi (értékű) szókapcsolat, mint *vérehulló fecskefű*.

Az effélék ugyanis sok tekintetben úgy viselkednek, mint a megfelelő szófajba tartozó szók, de másrészt különbség is van köztük, mind a beszélők

nyelvérzékéhez való grammatikai viszony szempontjából, mind a szókapcsolat stilisztikai értékét tekintve.

4.11.2. Azonkívül a szófaji rendezés azt a célt is szolgálja, hogy a további elemzés számára a más-más szempontokat kívánó fogalomszókat, formaszókat és mondatzókat elválassza egymástól, és így teret biztosítson mind-egyik típusú szóanyag sajátos szempontjainak és problematikájának. Így pl. az igék és névszók elválasztásával lehetővé teszi, hogy 9 oszlopon az igeragozást is, a névszóragozást is sajátos paradigmikus viszonyainak megfelelően elemezzük tovább anélkül, hogy ennek érdekében kétszer 9 oszlopot kellene felhasználnunk. (Vö. 4.3.2. és 4.3.3.)

Ugyanígy a névelőnek a többi szófajtól való elválasztásával lehetővé válik, hogy a névelő teljes grammatikai-stilisztikai problematikáját ugyanazon a lyukkártyán elemezhesük anélkül, hogy ezzel több oszloposort elvonnánk a többi szófaji kategóriába tartozó szó elemzésétől.

4.12. A 43—7. oszlop (fogalomköri jelzés; jelentés-elemzés) azokat a kulcsszámokat foglalja magában, amelyek az elemzett szöveg fontosabb fogalomköri csoportjait olyan rendszerben jelölik, amely nem törekszik ugyan több ezer kategória felsorakoztatására, de a felölelt fogalomkörök a részletes fogalomköri elemzés kereteiben úgy helyezkednek el, hogy nagyobb mennyiségű anyag esetén a fogalomköri hálózat tovább finomítható legyen a gépi elemzés számára.

Ezek a kulcsszámok nemcsak a fogalomköri kategóriákba sorakozó szavak mennyiségi elemzésére, az egyes fogalomkörök, fogalmi csoportok gyakoriságának vizsgálatára alkalmasak, hanem az egy és több fogalomkörbe tartozó szavak arányának tisztázására, valamint az egyes szavak jelentéshálózatában levő fogalomköri csomópontok meghatározására, s a szeména-összefüggések statisztikai elemzésére is.

4.12.1 A 43—7. oszlop adatainak, a szókárttyák szóhasználati adatainak (stílusminősítésnek, szakmai minősítésnek, állandósult szókapcsolatokra utalásnak), valamint a szöveggártyáknak segítségével gépi úton nagymértékben előkészíthetjük mind az egyes szavak, mind a szókapcsolatok jelentésének elemzését.

4.13. A 48. oszlop (Igekötő; előképző, előjel) olyan összefüggések adatainak gyűjtésére és statisztikai elemzésére alkalmas, mint az igekötős és igekötő nélküli igék aránya, jelentéstani összefüggése; az elöl és hátul álló igekötők aránya; a közvetlen és közvetett sorrendi hely előfordulása és aránya; az ismételt (pl. *be-benéz, el-elmegy, elő-elővesz, meg-megretten*), az összetett igekötős igék (pl. *ki-bejár, össze-visszazagyvál; el-bémegy*) adatai, az igekötős ige és igekötős igenév, főnévi, melléknévi származék aránya, összefüggése (pl. *elad: eladni; eladó, eladott, eladandó; eladva, eladván; eladás, eladatlan, eladható, eladhatatlan*); a felsőfokú és túlzófokú alakok aránya (pl. *legszebb, legeslegszebb; legtöbb, legkevesebb; legtovább, legeslegkorábban; legjava, legvége; legelőül, leghátul*); a felső- és túlzófokú melléknévi igenevek — igekötő nélküliek és igekötősek — aránya (pl. *legbántóbb, legműveltebb, legállandóbb; legelveteműltebb, legmegrögzöttebb*); igekötős igéből képzett felső- és túlzófokú melléknévek (pl. *legfeltartóztathatatlannabb, legmegvesztegethetetlenebb*) adatai és gyakorisága; az igekötős és nem igekötős utótagú összetételek adatai és gyakorisága (*napfelkelte — napkelte*).

4.14. A 49—50. oszlop a szóismétlés, az ikerszó és az összetett szók anyagát gyűjti össze.

Ezek az oszlopokon kódoljuk:

4.14.1. egyrészt az *alig-alig*; *egy-egy* (egy-et-egy-et); *kiki* (kinek-kinek), *mármár*, *mennék-mennék*, *néhanéha* típusú szóismétléses szavakat, szóalakokat;

4.14.2. másrészt a *csigabiga*, *tarkabarka*, *ingó-bingó*, *icipici*; *diribdarab*, *gírbegörbe* ~ *gírbegurba*, *mendemonda*, *ireg-forog*, *izeg-mozog* (izgett-mozgott) típusú ikerszókat.

4.14.3. Ugyanitt lyukasztjuk az összetételek különféle típusait:

4.14.3.1. a tapadásos vagy szeretlen összetételeket, összevont szavakat (mint *ábécé*, *ámde*, *azután* ~ *aztán*, *dejszen*, *délelőtt*, *délután*, *egymás*, *egyszer-egy*, *mégsem*, *mintegy*, *nahát*, *sohasem*, *úgyhogy*, *úgymint*);

4.14.3.2. a szűkebb értelemben vett vagy szerves összetételeket: a) a mellérendelő összetételeket, (mint *ágbog*, *ágas-bogas*, *két-három*; *tíz-húsz*; *száz-kétszáz*; *jön-megy*, *jövés-menés*, *kisebb-nagyobb*; *imitt-amott*, *imígy-amígy*; *hetet-havat*, *szórén-szálán*);

b) az alárendelő összetételeket, mint: az alanyos összetételeket (*mennydörög*, *magvaváló*, *porlepte*, *szárnyaszegett*; *szívszorongva*);

a tárgyias összetételeket (*egyért*; *jótevő*; *kárvallott*; *szófogadatlan*; *vállvetve*).

a határozós összetételeket (*aláír*; *alattvaló*; *túlerős*; *általút*, *ellenfél*, *tejbekása*; *egyedüllét*, *észrevétel*, *szemrehányás*; *tizenegy*; *százszorszép*; *vizbő*);

a jelzős összetételeket (*aprópénz*, *melegágy*; *forgószél*, *hálófülke*, *írótoll*; *ezeremester*, *háromszög*, *sokszög*; *egyhangú*, *féleszű*; *búzaliszt*; *gőzfürdő*; *gyöngyvirág*; *aranyárga*; *mozsárágyú*, *jégverem*; *borizú*);

a birtokosjelzős összetételeket (*ablakkeret*; *napsugár*; *vásárfia*); elkülönítve az alanyi birtokosjelzős összetételeket (*égzengés*, *érszűkület*; *napkelte* ~ *napfelkelte*, *napnyugta*) és a tárgyi birtokosjelzős összetételeket (*kézfogás*; *magaviselet*; *tanúságtétel*);

4.14.3.3. a szintaktikailag többféleképpen magyarázható összetételeket, mint *nyúl-gát* MNy. LV (1959) 80—7; — a más-más értelemben más-más típusba sorolhatókat, mint *ablaküveg* 1. 'ablakkeretbe foglalt üvegtábla' (birtokos összetétel), 2. 'ablaküvegezésre alkalmas üveg' (más célra szolgáló üvegfajtákkal szemben) (minőségjelzős összetétel), 3. 'szemüvegre tett vagy monoklinak használt sima vagy igen kis dioptriájú üveg' (hasonlító jelzős összetétel);

4.14.3.4. a szintaktikai elemezhetőségüket veszített, de a tagok önállóságát megőrző összetételeket, pl. *másodmagával* 'vele együtt ketten', *sokadmagával*; *harmadfél*, *negyedfél* stb.

Elhelyezhetők a különféle átmeneti típusok is, így:

4.14.4. Az ismétléses előtagú összetételek, pl. *két-kétezer* (vö. a *be-benéz* típusú igekötős származékokkal 4.13.).

4.14.5. Az ismétléses utótagú összetételek, mint *huszonöt-öt* '25-25' a régi nyelvben.

4.14.6. A csonka előtagú kétszeres összetételek vagy „mellérendelő” előtagokból álló „alárendelő” összetételek vagy „alárendelő összetételek”-ből álló olyan „mellérendelő összetételek”, amelyekben az azonos utótag csak egyszer (a mellérendelő összetétel utótagján) van kitéve, pl. *két-háromszáz* „200-300”, *tíz-húszezer* „10 000—20 000” (mint a *ki-bejár* típusú igekötős származékok).

4.14.7. Az ikerszók és összetételek közti átmenetek, pl. *cseng-bong*, *csetlik-botlik*; *dúl-fúl*; *immel-ámmal*; *szedett-vedett*; *tipeg-topog*; *tutyi-mutyi*; *cica-*

maca (MNy. LX (1954) 275, némileg más felfogással); *aligon-alig*; *egyed-egyedül*; *köröslén-körül*; *végestelen-végig*.

4.14.8. Az ikerszók és egyszerű szók közti átmenetek, pl. *cécó*, *gézengúz*, *hórihorgas*.

4.14.9. Az összetételek és egyszerű szók közti átmenetek, pl. *felebarát*, *felenyáj*, *harminc*; *tétova*.

4.14.10. Az összetételek és a képzett származékok közti átmenetek, pl. *királyné* 'királynő', *mosóné* 'mosónő'.

4.14.11. Az összetételek és előképzős származékok közti átmenetek, pl. *néha*, *néhol*, *némely(ik)*; *némi* (R. *néki* is); *akárhány*, *akárhol*; *bárki*, *bárhol*.

4.14.12. Az összetételek és szókapcsolatok közti átmenetek, pl. *tűzön-vízen át*; *cserbenhagy* (de *ránca szed*).

4.15. Az 51—3. oszlopon nemcsak az egyes tőváltakozás-típusokat lehet elemezni, hanem az egyes tőváltozatokat is; mind a változattípushoz tartozó egyes szavak megfelelő tőváltozatainak előfordulása (gyakorisága) szempontjából, mind a változattípusba tartozó összes esetek gyakorisága szempontjából; anyagot nyújt továbbá ez a lyukkártyamező az egyes tőváltozatoknak a tőváltakozásban játszott szerepe tanulmányozásához, a tő és a toldalék kapcsolódása módozatainak elemzéséhez.

A ragos alakokban jelentkező abszolút szótó változatai szófajok szerint (igetövek, névszótövek, egyéb tövek), azokon belül tőváltakozástípus szerint, majd a nevező alakkal egyező és a függő tőváltozatok sorrendjében kódolva különféle összefüggések egyidejű elemzését teszi lehetővé, pl. változatlan és változó tövek szövegbeli gyakoriságának aránya, a nevező alakú és a függő tőváltozatok aránya szavanként, szócsoportonként.

A képzett származékok töelemzése, jóval bonyolultabb a többszörös továbbképzés, a képzőhöz járuló jelek és ragok különféle kapcsolódási lehetősége miatt.

Az 51—3. oszlopon a képzett származékoknak lehetőleg abszolút tövét, illetőleg abszolút tőváltozatait elemezzük, pl. a *jóságos*-féle származékok *jórésztét*, a *javadalmazásbeli*-féle származékok *jav-* részét. A képző után járuló képző, jel, rag tőproblémáit (pl. *szomorká-s*, *könyvecské-k*, *else-je*) az 55—8. oszlopból álló lyukkártyamezőn alkalmasabban lehet elemezni. A két lyukkártyamező tőanyaga gépi úton összerendezhető.

A csak képzett származékban jelentkező abszolút tő változatai, mint a becéző származékok tőváltozatai, szintén beleférnek az 51—3. oszlop mezejébe, és sajátos természetüknek megfelelően elválaszthatók a többi tő-típustól.

4.16. Az 54., 60., 63., 66. „Kötőhangzó” („tővéghangzó”, „előhangzó”) jelzésű oszlopok elválasztása egyrészt azon az elvi megfontoláson alapul, hogy az adatgyűjtés és adatrendezés szempontjából nem kell különbséget tennünk tővéghangzó és előhangzó közt, sem kötőhangzó és a másik kettő közt, mert az elvszerűen rendezett anyag megkönnyíti a szétválasztást, ha elméleti megfontolások azt kívánják; másrészt annak megfontolásán, hogy e vitatott szerepű hangzók nemcsak az előhangzós elmélet szerint nem elemezhetők egy oszlopon, hanem a tővéghangzós-kötőhangzós elmélet szerint sem: a *bot-o-t*: *bot-o-k-a-t*, *bot-o-m-a-t*, a *szomorú-n* — *szomorú-a-n*: *szomorú-bb* — *szomorú-a-bb*: *szomorú-bb-a-n* (~ ? *szomorú-a-bb-a-n*); a *hall-a-na*: *hall-o-tt*: *hall-o-k*, *hall-o-m*-féle alakok alaktani összefüggése világosan mutatja eljárásunk cél-szerű voltát.

Azonkívül ezeknek a sajtóságos hangzóknak külön oszlopokon való lyukasztása azzal az előnnyel is jár, hogy a gépi rendezés során automatikusan egymás közelébe kerülnek a paradigmaticus szempontból szorosabban összehasonlítható esetek; minden nehézség nélkül lyukaszthatók a más-más kronológiai síkba tartozó változatok (pl. *nézzünk ~ nézzönk*), az esetleges nyelvjárási változatok (pl. *várjanak ~ várjonak*); könnyen összerendezhetők a „kötőhangzó” („tőhangzó”, „előhangzó”) nélküli változatok és ilyen hangzós változatok (pl. *száll-t — száll-o-tt; száll-na — száll-a-na*): egyszerűen a tő és a szóban forgó hangzó nélküli toldalék kulcsszáma szerint rendezhetjük az anyagot, s a gép automatikusan kivetíti a keresett változatok bizonyító és szemléltető anyagát. Végül ezzel a módszerrel olyan változatok lyukasztását is biztosítjuk, amelyekre esetleg az előzetesen elemzett anyag hiányos volta miatt nem is gondoltunk (pl. van-e *mindenütt típusú változat?).

4.17. Az 55—8. oszlop (Képző, képzőbokor, képzősor, képzőszerű szóvégi) kódolásának elveit később fogom véglegesen megszovegezni; akkor, amikor a nagyrészt együtt levő anyagnak kulcsszámok szerinti rendezését és kódolását befejeztem. Annyit már az eddigi anyag alapján is megállapíthatónak tartok, hogy a képzett származékokban a szinkronikus és a diakronikus szempont elvszerű alkalmazása sok nehézségbe ütközik, de erre az elválasztásra mégis szükség van, ha a szóalakok nyelvbéli életére, stilisztikai különbségeire, nyelvi rendszerbeli szerepére tekintettel akarunk lenni. A szépirodalom művészi archaizmusaira, a népies szereplők nyelvi sajtóságaira, valamint irodalmi nyelvünknek a népnyelvből gyakran merítő eljárására való tekintettel célszerűnek látszik a képzők és képzőszerű elemek fontosabb népi és régi nyelvi változatait is felvenni a jegyzékbe.

Másrészt gondolhatunk a nagyon gyakori képzőegyütteseknek egy kulcsszámon való elemzésére, hogy az anyag felesleges szaporítását elkerüljük. Tehát pl. ha *-ítás, -ítés*-féle származékokat több kötetnyi szövegben külön elemezzük az *-ít* képző és külön az *-ás, -és* képző szempontjából, az az anyagot feleslegesen szaporítja. Ha pl. az *-ítás* képzőcsoport anyaga együtt van, a dokumentáció szempontjából szükség szerint gépi úton könnyen berendezhető akár az *-ít*, akár az *-ás* képző anyagába; ha pedig csak a gyakorisági adatokra van szükség, az *-ít* és az *-ás* képző szempontjából nem is kell berendezni az anyag egészét.

A képzők és képzősorok fontosabb alakváltozatai külön kulcsszámot kapnak azért, hogy a képzőváltozatoknak (pl. *-ság ~ -ség ~ -síg; -dül ~ -dül ~ -dul ~ -dül; -ít ~ -it ~ -ét; -dogál ~ -dégél ~ -dögél*) egymáshoz való aránya is tanulmányozható legyen. Vannak ugyanis olyan képzőink, amelyeknek alakváltozatai nagyjából egyensúlyban vannak egymással, vannak olyanok, amelyeknek egyik vagy másik változata rendkívül ritka, vagy csak archaizmusként, vagy csak népies stílusrétegben fordul elő mai nyelvünkben.

A képzőnek az alapszó, illetőleg a tő szófaji típusától a származék szófaji értékétől való függését lehetőleg elvszerűen vizsgáljuk. A deverbális és a denominális típus mellett a departikuláris (határozószó, névutóból, mondatszóból képzett) származékokat is következetesen megkülönböztetjük, külön megjelölve a gyakoribb képzők esetén a mondatszói vagy mondatszószerű hangutánzó, valamint a hangfestő tövet, mert bizonyos képzők használata jellemzőbb ezekre a tövekre, mint a tisztán igei vagy névszói tövekre. A gyakoribb képzők esetén a denominális képzésen belül is megkülönböztetjük a fő-

névi, a melléknévi, a számnévi, a névmási tövekhez járuló képzőhasználatot ahol ez a megkülönböztetés keresztülvihető. Ahol nem, ott marad az általánosabb „denominális” megjelölés. Tekintettel leszünk a kódolás előkészítésében a szófajilag elemezhetetlen (*bakafántos*), a bizonytalan, vitatott vagy vitatható szófajú (*csúfondáros*), illetőleg többféle szófaji értékű tőhöz járuló képzőhasználatra (az utóbbira példa: *nyugi! nyug-* [a „nyugalom!” vagy „nyugodj!” töve] + *-i*).

A képzett származék szófaji értékének figyelembevételekor tekintettel leszünk arra, hogy itt messzemenő árnyalásra nincs szükség, mert a származék szófaji értékét a 40—2. oszlop szabatosan meghatározza. Az egyes képzők „szófaj-determináló” tendenciáját tehát a képző kulcsszáma(i) és a szófaji érték kulcsszáma alapján statisztikailag is meghatározhatjuk.

4.18. Az 59. oszlopot arra tartjuk fenn, hogy a különösen bonyolult képzőegyütteseket, a komplexebb szóalkotásmódokat a többi anyagról leválaszthassuk.

(A szóalkotásnak egyéb módjai, mint népetimológia, szóvegyülés, elvonas, amelyeknek nincsenek alaktanilag az ismétléshez, ikerítéshez, „összetétel”-hez, szóképzéshez hasonló jellemzőik, az etimológiai részben kapnak kódszámot.)

4.19. A 60—68. oszlopon, a jelek és ragok elemzése során külön-külön sorozatban szerepelnek a névszóragozás és az igeragozás jelei és ragjai (4.3.2.). Ebből semmi zavar nem származik, mert a kétféle ragozási rendszer belső összefüggései lényegesen mások, s még ahol a kétféle rendszerhez tartozó alaktani elemek hang(betű)alakjukban egybeesnek is (*-m* személyrag, *-nk*: *-unk*, *-ünk* személyrag; *-i* többesjel: *-i* személyrag), jelentésük, funkciójuk meghatározása csak a ragozási rendszeren belül történhetik, és gépi elemzésük a szófaji érték figyelembevételével történik, ez pedig minden félreértés lehetőségét kiküszöböli.

4.20. A névszókra a 61—2., 64—5. oszlopon technikai okokból kapcsoljuk az *-é* (*-éi*) birtokos-birtokjelet (és többes *-i* jellel való kapcsolatát) a többi névszói jelhez, illetőleg a személyragokhoz, hogy ne kelljen az *-é* elemzésére külön oszlopokat igénybe venni. A szabatos kódolás és a részletek összefoglalásának megkönnyítésére zárójelben közöljük az utalószámokat. Ugyanezen a két lyukkártyamezőn foglaljuk össze az *-i* birtok-többesjel használatát; az egyszerű *-i*-nek a többi jellel és a személyragokkal való kombinációját utaló számokkal foglaljuk egységbe.

A *-k* többesjel és a névszói jelek kapcsolatát szintén a 61—2. oszlop kulcsszámai alapján elemezzük, az összefüggéseket utaló számokkal foglaljuk egységbe.

4.21. Az igéken a mód- és időjelek kódolásában (61—2., 64. oszlop) a feltételes mód jelének *-na*, *-ne*, *-ná*, *-né*, *-nó*, *-nő* változatai, az elbeszélő múltnak ezzel párhuzamosan *-a*, *-e*, *-á*, *-é*, *-ó*, *-ő* változatai szerepelnek külön kulcsszámmal; így a személyragoknak amúgy is bonyolultabb rendszere könnyebben kódolható.

4.22. A névszói (birtokos) „személyragok” v. „személyjelek” (64—5. oszlop, egyes esetekben 63—5. oszlop) elemzését összekapcsoljuk a személyrag (személyjel) után következő *-é* jel és *-éi* jelkapcsolat elemzésével.

4.23. Az igei személyragok elemzésében tekintettel kívánunk lenni a személyragok változataira s a mód és időbeli alakokban való jelentkezésükre is. A rendszertani összefüggések elemzését utaló számokkal könnyítjük meg.

4.24. A 69—70. oszlopon (a műre vonatkozó adatok) olyan, főleg műfaji szempontokat érdemes rögzíteni, amelyek a nyelvi adatok értékelésében eligazító értékűek. Pl. társadalmi regény v. novella; történeti regény v. novella; társadalmi (középfajú) dráma v. vígjáték vagy bohózat, vagy tragédia; történelmi (középfajú) dráma vagy vígjáték vagy bohózat vagy tragédia; drámai költemény; társadalmi tárgyú elbeszélő költemény; történeti tárgyú elbeszélő költemény; eposz; lírai vers (pl. dal, óda, himnusz, rapszódia, elégia, epigramma), költői leírás; elmélkedő próza; vitairat; szónoki beszéd; tudomány-népszerűsítő mű; tudományos mű (értekezés); hírlapi cikk; napihír; apróhirdetés; levél.

4.25. A 71—2. oszlop (a szöveggörnyezetre vonatkozó főbb adatok) szempontjai között ilyenfélékre gondolok: elbeszélő szövegrészben; elmélkedő szövegrészben; párbeszédes szövegrészben; kérdésre adott feleletben; állandósult szókapcsolatban; viszonyító szókapcsolatban, pl. többtagú kötőszókban, mellékmondatra utaló szókapcsolatban; vezényszóban v. egyéb mondatzó szókapcsolatban; szitkozódásban v. káromkodásban; köszönésben; nyelvi klisében; szólásmódban; szóláshasonlatban; közmondásban; szállóigében; a beszélő társadalmi helyzete a műben, esetleg — ha jellemző — életkora.

4.25.1. A szereplők közül csak azokat érdemes külön kulcsszámmal jelölni, akiknek szavait, vagy gondolatait — párbeszédbeli helyzettel vagy egyéb módon meghatározott — idézetként tartalmazza a szöveg, s a szóhasználat jellemző rájuk. Ilyenkor a szereplő társadalmi helyzetére utaljon a kulcsszám. Pl. munkás; iparos; paraszt; kereskedő; értelmiségi (bíró, gyógyszerész, mérnök, orvos, tanár, tanító, ügyvéd stb.); katona; (közkatona, altiszt, katonatiszt); arisztokrata; tanuló; gyermek; ifjú; serdülő leány; háziasszony és hasonló.

4.26. A 73. oszlopon (a szóra vonatkozó főbb adatok) olyan szempontokat kódolhatunk, mint tipográfiailag kiemelt szó; (1) dőltbetűs, (2) ritkított szedésű, (3) félkövér, (4) kövér betűs, (5) csupa nagybetűs, (6) idézőjellel kiemelt szó; — vagy: (7) alliteráló szó; (8) rímelő szó; (9) a szöveg más helyén minősített vagy értelmezett szó.

4.27. A 74—5. oszlop (stílusminőség, földrajzi, társadalmi kötöttség) kulcsszámait elsősorban az ÉrtSz. minősítései alapján állíthatjuk be a lyukkártyatervbe, esetleg néhány kiegészítéssel pl. „... szereplő szava”; „az író szava .. szereplője gondolataként”; „képes használatban”; „idézetben”, „indirekt beszédben”; „korábbi szövegváltozatban”, „első szövegváltozatban”, „későbbi szövegváltozatban”.

Csak azt a stílusminősítést kell jelölni a szóalak kartonján, amelyet a közelebbi vagy távolabbi szöveggörnyezet támogat. A kódolónak olyan egyéni stílusminősítése, amelyet a szöveg nem támogat, ne kerüljön a lyukkártyára. (Az esetleg külön, speciális vizsgálatként lyukkártyákon elemezhető, hogy valamely szöveg egyes szavai stilisztikai értékének milyen szórásképe van különböző egyének nyelvérzékében, de ez az általános lyukkártya keretében nem oldható meg.)

4.28. A 76—7. oszlop (Szakmai minőség, szaknyelvi használatra utalás) alapja szintén csak a szöveg lehet. Erre tehát szaknyelvi szövegek feldolgozása-kor kerülhet sor és olyan szépirodalmi és egyéb szövegek egyes részletei esetén, amelyek szaknyelvi szóként használják a lyukkártyán szereplő szót, illetőleg szóalakat. A szaknyelvi hovatartozás meghatározására szolgáló megnevezéseket és kulcsszámokat külön melléklet tartalmazza.

4.29. A 78. oszlop (Etimológia) szókinsünk fontosabb etimológiai rétegeinek gyakorisági elemzésére szolgál. Az etimológia meghatározására egyelőre Bárczi Szófejtő Szótára használható, a vizsgálat későbbi szakaszán esetleg már az előkészületben levő etimológiai szótár. Az etimológiai kódolás keretét és szempontjait külön melléklet tartalmazza.

4.30. A 79. oszlop (Mondatrész-érték, szintagmatag-érték) csak azokat az eseteket tartalmazza, amelyek a szófaji érték meghatározása vagy a jelentés elemzése szempontjából hasznosak lehetnek. Ilyenek pl. (1) állítmány; (2) alany, tárgy; (3) „főnévi” természetű határozó (-*ban*, -*ba*, -*ból* stb.); (4) „melléknévi” természetű határozó (*bátran*, *betegén*); (5) általános természetű határozó (-*nek*, *hisz*, *lát*, *tart*); (6) (tulajdonképpeni) jelző; (7) birtokosjelző; (8) utójelző (hátravetett jelző); (9) értelmező. A nem névszói természetű határozókat nem kell külön kódolni ezen az oszlopon, mert ezek a szófajjelölésben szerepelnek (pl. határozószó, hat. igenév).

4.31. A 80. oszlopon ilyenféle mondatjelentéstani kérdések rögzítése látszik célszerűnek: (1) higgadt közlésben; (2) érzelmileg színezett közlésben; (3) óhajító mondatban; (4) kérésben; (5) szigorú parancsban; (6) higgadt kérdésben; (7) nyugodt feleletben; (8) érzelmileg színezett kérdésben; (9) érzelmileg színezett feleletben.

Irodalom

BUSA, R., Mechanisierung der philologischen Analyse (Nachr. Dok. 3 [1952], 14–9).
Uő., Entwicklung der Mechanisierung der sprachlichen Analyse. (Nachr. Dok. 4 [1953], 202–4).

Lexicologie et Lexicographie Française et Romanes. Orientations et exigences actuelles. Strassbourg 12–16 Novembre 1957 (Colloques internationaux du Centre National de la Recherche Scientifique). Paris [1960].

QUEMADA, B., Mécanisation dans les recherches lexicologiques (CL. 1 [1959] 7–46).
Uő., La technique des inventaires mécanographiques (Lexicologie et Lexicographie [1960], 53–63).

Uő., L'inventaire lexicographique en vue d'un Thesaurus National (CL. 3 [1962], 119–31).

GREIMAS, A. J., Les problèmes de la description mécanographique (CL. 1 [1959], 47–75).

IMBS, P., Au seuil de la Lexicographie. (CL. 2 [1960], 3–17).

SCHODEL, H. – REIMANN, R., Observations sur l'usage contemporaine du mot „style”. (BILAL V [1961], 43–53)

TOLLENAERE, F. DE, Automatisering in de Lexicologie. Automatisering en Taalkunde [1961], 24–34).

Uő., La documentation lexicographique et ses propres besoins (CL. 3 [1962] 101–15).
Uő., Nieuwe Wegen in de Lexicologie. Amsterdam, 1963.

HÜBNER, G., Die Ex[z]erption des dritten Bandes von Goethes Farbenlehre mit Hilfe von Lochkarten und Maschinen (CL. 3 [1962], 99–100).

PHAL, A., Les groupes de mots et les problèmes qu'ils posent dans la prédiction de textes scientifiques destinés à l'analyse mécanographique. (CL. 4 [1964], 45–60)

KELEMEN J., Gépi adatgyűjtés és adatfeldolgozás a lexikográfia szolgálatában. (Szótártani Tanulmányok [szerk.: Ország László] Bp. [1966], 29–54): MNyTK. 114. sz.

4.32. Mind a szövegekártyáknak, mind a szó-, illetve szóalakkártyáknak a fentiekben vázoltnál összevontabb típusára lesz szükség l e x i k o g r á f i a i c é l ű g y ű j t é s b e n. A szóalak ugyanis gyakran nem azonos a szótári címszóval: pl. *fát*, *fák*, *fákat*, *fára* stb. címszava: *fa*; *álma*, *álmot*, stb. címszava: *álm*; *szava*, *szavát* stb. címszava: *szó*. Még fontosabb a külön lexikográfiai kártya („címszó-kártya”) a *menj be* vagy *be sem ment* típusú igekötős igealakok vagy a *legrégibb* (címszó: *régi*), *leghátulsó* (címszó: *hátulsó*), *legtovább* (címszó: *tovább*)-féle felsőfokú melléknevek és határozószók esetén. Ezek lexikográfiai helyének megállapításához célszerű legalább hét oszlopot (pl. a 74–80. osz-

lopot) üresen hagyni, hogy megfelelő sorszám lyukasztása és a címszó alá sorolása számára helyet biztosítsunk: 6 oszlopot a címszó sorszámának, egyet a segédkártáknak.

4.32.1. Ilyen megfontolás alapján a szóalak kártya fenti tervéből a szótári célú gyűjtésben elhagyható lenne pl. a 74—5 (stílusminősítés), 76—7 (szakmai minősítés), 78 (etimológia), 79 (mondatrész-érték), 80 (mondatjelentéstani mozzanat): összesen 7 oszlop. Ezek helyett egyik oszlopon (pl. az újabb lyukkártyatípus 73. oszlopán ilyen figyelmeztető lyukasztás lehetne: a szövegben stílusminősítés, a stilisztikai érték meghatározására valló mozzanat van; a szövegben szakmai minősítés vagy ilyenre utaló megjegyzés van. Ebben az esetben az efféle megjegyzéssel ellátott szóalakkártyákról másodkártya készülne a 51—72. oszlopok üresen hagyásával, s az elmaradt szempont kódolását a másod-kártyák tartalmazzák. A mondatrész-értékre és a mondatjelentésre vonatkozó lyukasztást a szótári gyűjtésben a szövegekártyamásolatok pótolják. Az etimológiai kódolás a címszókártákra kerül. (vö. 4.32.2.4).

4.32.1.1. A szótári szóalakkártya 1—50., 55—8., 74—80. oszlopa megegyezhet a szóalakkártya eddig részletezett típusával.

4.32.1.2. Az 51—4. oszlop azokat az adatokat tartalmazhatja, amelyek a szóalakkártya szótári besorolása szempontjából fontosak: Pl. 51—2. oszlop: [A szóalakkártya és a hozzátartozó idézet helye a szócikkben: a] .. [számú jelentéshez tartozik].

53. oszlop: A szövegbeli szóalak minősítése a címszóhoz viszonyítva I. Pl. (1) ejtészváltozatként értékelhető alakváltozat; (2) írásváltozat;

54. oszlop: A szövegbeli szóalak minősítése a címszóhoz viszonyítva II. Pl. (1) alaktani szerkezete a címszóéval egyezik; (2) esetragos alak; (3) „személyragos” (esetragtalan) alak; (4) „személyragos” — esetragos alak; (5) „személyragos” — jeles alak; (6) jeles (ragtalan) alak; (7) jeles — esetragos alak; (8) jeles — „személyragos” — esetragos alak.

4.32.1.3. Az 59—73. oszlop megegyezhet a címszókártya megfelelő oszlopaival:

59—62. Stílusminősítés [a szóalakra, a szövegrészben való stilisztikai értékre vonatkozólag].

63—5. Szakmai minősítés [a szövegrészben való használatra vonatkozólag.]

66—71. Etimológia (ahol statisztikai elemzés szempontjából fontos).

72—73. Utaló kártyák kódszáma.

4.32.1.4. A 74—80. oszlop a szövegekártyákhoz hasonlóan a címszó kódszámát és a segédkártya számát tartalmazza.

4.32.2. A szóalakkártyákat összefoglaló címszó kártya a szótári gyűjtésben a szótári szó lexikográfiai elemzéséhez nyújt kellő teret a 3—19., 43—80. oszlopokon. (Az 1—2., 20—42. oszlop szempontjai azonosak lehetnek a szóalakkártya szempontjaival, de a 40—42. oszlop a szó összes szófaji értékeit tartalmazza.)

4.32.2.1. A 3—15. oszlop pl. ilyenféle lexikográfiai szempontok lyukasztására használható: első előfordulás (szövegben, szótárban: a címszóval egyező alakban és közzsóként; egyéb alakban, tulajdonnévként vagy tulajdonnév részeként); utolsó előfordulás (szövegben, szótárban).

4.32.2.2. A címszókártya 16—17. oszlopa a homonimák árnyaltabb elemzésére használható. Pl. (1) etimológiai értelemben vett homonimák; (2)

végső fokon azonos eredetű de más-más nyelvből átvett szók, mint homonimák; (3) divergens jelentésváltozással keletkezett, szinkronikus összefüggések, ill. különbségek alapján megkülönböztetett homonimák.

4.32.2.3. A 18–19 oszlop a címszónak *n*-ekben meghatározott hosszúságát jelöli.

4.32.2.4. A címszókártya 43–73. oszlopa például a következő szempontok szerint használható fel:

43–52. Fogalomköri jelzés, jelentéselemzés.

53. A címszó alakja I.: 1. címszóalak 2. ejtéváltozatként értékelhető alakváltozat; 3. írásváltozat.

54. A címszó alakja II.: 1. Ø jel – Ø személyrag Ø esetrag; 2. Ø jel – Ø személyrag – esetrag; 3. Ø jel – személyrag – esetrag; 4. jel – Ø személyrag Ø esetrag; 5. jel – Ø személyrag – esetrag; 6. jel + személyrag + esetrag.

55–8.: A címszó (alakváltozat stb.) összes előfordulásának száma.

59–62. Stílusminősítés [a címszóra vonatkozólag].

63–5. Szakmai minősítés [a címszóra vonatkozólag].

66–71. Etimológia.

72–73. Utaló kártyák kódszáma.

4.32.2.5. A 74–80. oszlop a címszó sorszámát tartalmazza, s a szövegkártya és a szótári szóalak-kártya gépi besorolását biztosítja.

4.33. A szóalak- és morfémakártya, szótári szóalak-kártya és a címszó-kártya fenti általános tervéhez összehasonlításul közlöm a QUEMADA-féle „s z ó k á r t y a” beosztását:

- 1 oszlop: A kártyatípus száma: az egyes kártyák átcsoportosítására és a különféle műveletek utáni ellenőrzésre szolgál.
- 2–16 oszlop: A szerző, a szöveg adatai, forrásjelzés, kötet, oldal verssor v. prózai sor, magyarázó értékű egyéb szöveggörnyezet (lekszisz), a szó sorszáma a versben v. prózai sorban (15 oszlop).
- 17–36 oszlop: A szó a szöveg helyesírásával (20 oszlop).
- 37–40 oszlop: Nyelvtani adatok (pl. nem, szám, tranzitív v. intranzitív stb. (4 oszlop).
- 41–49 oszlop: A jelentés kódszámai, a fogalomkörök (jelentésmezők) szerint (9 oszlop).
- 50–52 oszlop: A szöveg keletkezése (3 oszlop).
- 53–56 oszlop: Adatok a szerzőről (kor, nemzedék, hivatás, vallás stb.) (4 oszlop).
- 57–61 oszlop: Adatok a műről (próza, költészet, irodalmi (szépprózai) szöveg, technika stb.) (5 oszlop).
- 62–66 oszlop: Adatok a szöveggörnyezetről (idézet, elbeszélés, felelet, a beszélő szociális helyzete, szövegváltozat, javítás stb.) (5 oszlop).
- 66–71 oszlop: Adatok a szóról (dőltbetűs, nagybetűs; idézőjel; változat, a szó helye a verssorban v. a mondatban, rím) (5 oszlop).
- 72–80 oszlop: Numerikus kódjelek (a szó matrikulaszáma az általános jegyzékben, amely lehetővé teszi a helyesírási változatok, a ragozási alakok szerinti átcsoportosítást).

Irodalom:

- QUEMADA, B., CL 1 (1959), 25–7; BILAL I (1960), 13–4.
TOLLENAERE, F. DE, Nieuwe Wegen in de Lexicologie 1963. 22–3.
Uó., BSI 6 (1963), 36–7.

5. A szintaktikai célú szöveggyűjtés egészen más problémákat vet fel, mint a szó- és morfémagyűjtés. Bár az utóbbi a ragok és egyes szófajok (határozószó, névutó, kötőszó, mondatsszó) és egyes állandósult szókapcsolattípusok gyűjtésével és elemzésével kétségtelenül támogatja a szintaktikai kutatásokat is, ez az anyag azonban feltétlenül kiegészítendő sajátos mondattani, szószerkezetani lyukkártyaterv alapján végzendő gyűjtéssel. A mondattan és a szószerkezetan tárgyának zömét a szószerkezetek, szószerkezethalmazok, tagmondatok, egymással kapcsolatban levő mondatok alkotják: vizsgálatukra tehát olyan lyukkártyatervre és olyan hosszabb szövegrészekre van szükség, amelyek alkalmasak a szószerkezetek, tagmondatok, mondatok, mondatkapcsolatok bonyolultabb összefüggéseinek elemzésére.

5.1. Természetesen nem alkalmasak a szó- és morfémagyűjtés céljából előállított lyukkártyák a mondatfonetikai sajátságok elemzésére sem; még akkor sem, ha az azonossági számok alapján több lyukkártyát összekapcsolnánk. Ezek sajátos, mondattani — szintaktikai elvű kódolást kívánnak, ezeket tehát az ilyen lyukkártyákon kell elemezni, s ennek az elemzésnek az összefüggő hangsorok és szünetek vizsgálatán, a beszédíramon és változásain, a hangfekvésbeli és hangmenetbeli (hanglejtésbeli), valamint a hangsúlyozásbeli változatokon kell alapulnia.

5.2. A mondat jelentésének elemzése külön kódolást kíván, mert a mondat gyakran bonyolult jelentésviszonyok együtteseként jelentkezik, amelyben a valóság tükröződése, a beszélőnek a valóság jelenségeivel szemben való állásfoglalása és a mondatbeli közléshez való intellektuális-érzelmi-akarati viszonyulása is kifejeződik.

5.3. A mondat szerkezeti elemzésének lyukkártyatervében a mondat egészének szerkezeti jellemzése mellett is valószínűleg kétirányú kódolást kell elvégeznünk: egyrészt a mondat lineárisan tagolódó kisebb egységeit („tagjait”) és ezek egymásutánját kell megkülönböztetnünk (pl. előkészítő rész — főrész — toldás; közrefogó rész első fele — közrefogott [közbeékelt] rész — közrefogó rész második fele), esetleg e kisebb részek („tagok”) kisebb részeit és ezek egymásutánját is (pl. közvetve és közvetlenül megelőző rész — központi rész — közvetlenül és közvetve követő rész); másrészt az „egymás mellé rendelt”, illetőleg a „föle- és alárendelt” részek („tagok”) egymáshoz való viszonyát kell meghatározni a lineáris elhelyezkedéstől függetlenül és a lineáris elhelyezkedéssel kapcsolatban.

5.4. Külön tisztázandó, a mondatfonetikai, mondatjelentéstani és mondat-alaktani, szerkezetani kérdések elemezhetők-e a szövegrészletet kísérő egyetlen lyukkártyán vagy pedig — s ez a valószínűbb — a mondat komplex elemzéséhez több lyukkártyára van szükség.

6.1. Bizonyos hangtani vizsgálatok céljára számos vonatkozásban a szó- és morfémagyűjtés céljára készült lyukkártyák is felhasználhatók, mindig figyelembe véve természetesen helyesírásunk hangjelölésének következetlenségeit. A kiejtésen, az etimológiai összefüggésen és a hagyományon alapuló, egyszerűsítéseket is tartalmazó helyesírás természetesen sokban nem a valóságos hangtani viszonyokat tükrözi, de megfelelő előkészítéssel összeállíthatók olyan hangtani szempontok, amelyeket helyesírásunk is meglehetősen következetesen jelöl, így pl. a magánhangzók időtartambeli különbségeit különösen *a-á, e-é, o-ó, ő-ő* esetében (az *i-í, u-ú, ü-ű* hangok jelölésében aránylag gyakori az ingadozás); a zöngés és zöngétlen mássalhangzók nem hasonlóságot. A betűstatisztika, a betűkapcsolatok tipizálása

és statisztikája tehát bizonyos hangtani következtetések levonására is alkalmas.

6.2. Elemezhető a szó- és morfémagyűjtés anyagán is a magyar szóalakoknak szótagtípusokból való felépülése, valamint a különféle típusú szótagok gyakorisága, a különféle szótagokból álló szavak gyakorisága stb., figyelembe véve egyes szótagtípusok grafémikus és fonémikus képének tipizálható eltérését (pl. *önts: öncs; tartsd: tardzd; folyt: foit*).

6.3. De természetesen nem végezhetünk el minden hangtani elemzést a szó- és morfémagyűjtés szempontjából előkészített lyukkártyákon. Finomabb hangtani árnyalatok elemzéséhez sajátos szempontok szerint megtervezett lyukkártyákra lesz szükség, de a szóalakok lyukkártyáinak szövegrésze (20—39. oszlop) ilyen vizsgálatok jelentős részére is alkalmazható és egyszerű gépi másolással olcsón átvihető újabb lyukkártyákra, amelyeknek 16—19. oszlopa elég tág teret biztosít árnyalt fonetikai és fonológiai szempontú elemzések programozására.

6.4. Végezhető olyan hangtani elemzés is a szó- és morfémagyűjtés szempontjából előkészített lyukkártyákon, amelyeken változatlanul hagyjuk a 20—39. oszlop lyukasztását akkor, ha a szóalak (nagyolt) fonetikai átírása egyezik az irodalmi helyesírással, s a 20—39. oszlop üresen hagyásával készítünk másolatokat azokról a szóalakokról, amelyeknek nagyolt fonetikai átírása különbözik az irodalmi helyesírástól, s ezeket a szóalakokat nagyolt fonetikai átírással lyukasztjuk a 20—39. oszlopon, s utána az egész anyagot gépi rendezéssel elemezzük hangtani szempontok szerint.

6.5. A hangtörténeti kutatások számára a szó- és morféma-gyűjtés igen jó nyersanyagot szolgáltat, ha e gyűjtést a nyelvemlékes kor egészére kiterjeszthetjük.

6.6. Az eddigi hangtörténeti részletkutatások szintézise lyukkártyákon jól előkészíthető a változás egyes mozzanatainak kettős kódolásával: külön oszlopon ill. mezőn kódolhatjuk a hangtörténeti előzményt és a változás eredményét. Ezzel biztosíthatjuk mind a progresszív (deszcendens), mind a regresszív (aszccendens) szemléletű szintézist. Ugyancsak külön oszlopon kódolható a feltett hangváltozás bizonyított volta (külön-külön lyukkártyákon a bizonyító anyag részleteivel) vagy nem bizonyított, feltételes volta. Ismét más oszlopon vagy oszlopokon lyukasztatók a hangváltozás feltételei, feltehető oka vagy okai, valamint a hangtörténeti analógiák; a hangváltozás kora; esetleges egyéb szempontok.

7. A mai helyesírás részletkérdéseinek lyukkártyarendszerű vizsgálatától elsősorban a Helyesírási Tanácsadó Szótár kiegészítése terén, e Szótár anyagának a Helyesírási Szabályzat rendelkezéseivel való összehasonlítása terén várhatunk hasznos eredményeket.

7.1. A helyesírástörténet kérdéseinek tanulmányozásában más-más lyukkártyatervet kell kidolgozni egyrészt a szórványemlékek, a kódexek és egyéb kéziratok helyesírásának vizsgálatára, másrészt a nyomtatványok helyesírásának tanulmányozására. Az előbbi csoportban ugyanis sokkal bonyolultabb betűkódolást kell kidolgozni, s ebben olyan „betű-változat”-okra is tekintettel kell lenni (pl. az *ly* hang jelölésére: *ly, li; l* mellett *li, l' l', 'l*; az *ny* hang jeleként: *ny, ni, ng, gn, n* mellett *n, ñ, n''; ñi*), amelyeknek gépi jelölésében előreláthatólag nagy szerepe lesz a diakritikus jelek alkalmazásának s a betűk és számjegyek kombinációjával való azonosításának. A nyomtatványok hangjelölésében

egységesebb betűtípusokat találunk, de ezek is felvetnek megoldandó azonosítási kérdéseket (pl. az *ó, ú, ő', ú', ő, ü* megkülönböztetésének kérdését).

8. Kész szótárak lyukkártyarendszerű elemzése számos szempontból történhetik. Az elemzés szempontjait a feldolgozásra kerülő szótár természetéhez kell szabni. Tehát más tervet kell kidolgozni a kétnyelvű szótárakra, mint az egynyelvűekre; mást a dokumentáló célú történeti szótárakra, mint a jelenkori elemző szótárakra; mást a szótári rendű frazeológiai gyűjteményekre, mint a szinonima-szótárra; mást a konkordanciákra, mint az értelmező szótárakra.

8.1. De bizonyos lexikográfiai szempontokat mindenfajta szótár lyukkártyatervének tartalmaznia kell. A helyesen szervezett szótárelemzéshez tehát szükség van olyan általános lyukkártyatervre, amely az általános szótárírói gyakorlatban érvényesülő és a különféle szótártípusokban jelentkező közös vonások részére kijelöli a lyukkártya általános kereteit és fenntart bizonyos oszlopokat speciális szótártípusok elemzésére.

8.2. A szótárak általános lyukkártyatervében teret kell tehát biztosítani a forrásjelzésen és évszámjelzésen kívül a mindenféle szótárban szereplő címszóalaknak (betűjelek lyukasztásával); a több szótártípusban felbukkanó címszóalakváltozatoknak (főcímszó, utaló címszó, egyéb alakváltozat, amely címszóként nem jelentkezik); a grammatikai szerelés általános keretének; a jelentéstagolás keretének (jelentések száma, értelmezés módja, jelentések sorrendje); a szemléltető anyag tipizálásának; a sok szótártípusban szereplő bokrosításra utaló számoknak (pl. értelmezés nélküli származék a betűrend figyelmen kívül hagyásával a szócikk belsejében vagy végén, szemléltető anyaggal vagy anélkül; értelmezett származék a szócikken belül, a betűrend figyelmen kívül hagyásával; értelmezett összetétel a szócikken belül; értelmezés nélkül közölt összetétel a szócikken.)

8.3. Szintén több szótártípusban vannak etimológiai megjegyzések, nem etimológiai szótárban esetleg csak bizonyos típusú szavakra vonatkozólag, (pl. idegen szavak; nemzetközi vándorszavak).

8.4. Speciális szempontú elemzésre tarthatnak számot pl. a két- és többnyelvű szótárak, az idegen szavak szótárak, a szakszótárak, a tájszótárak, a nyelvtörténeti anyagot tartalmazó szótárak (mint Nyelvtörténeti Szótár, az Oklevélszótár, a Nyelvújítási Szótár), az etimológiai szótárak, a szinonima-szótárak, a stíluszótárak, a nyelvművelő célú szótárak és szójegyzékek.

8.5. A szótárak elemzésében külön figyelmet érdemel a nem-irodalmi szemléltetőanyag. A jobb szótárakban ugyanis tetemes élőnyelvi anyag található. Ezeknek szó- és morfémaanyagát, frazeológiai részleteit tehát érdemes lesz külön is feldolgozni.

8.6. Ha a szótárak címszóanyagát morféma-gyűjtésre is fel akarjuk használni, erre a célra valószínűleg másod-kártyákat kell igénybe venni, mert a morfémagyűjtés szempontjai túl sok teret elvonnának a sajátosan szótári problémák elemzésétől.

8.7. A szótárak elemzésének általános lyukkártyatervét szintén elkészítettem, de véglegesítése előtt néhány szótártípuson szeretném kipróbálni. Ennek ismertetésére később kerül sor.

8.8. Előkészületben van néhány speciális szótár elemzésének lyukkártyaterve, pl. Czuczor—Fogarasi szótáré, a Szófejtő Szótáré, az Oklevélszótáré, a Nyelvtörténeti Szótáré, a Nyelvújítási Szótáré, a Magyar Tájszótáré. [Ezek

elemzésének sajátos szempontjait vázlatosan tárgyalom alább idézett tanulmányomban.]

Irodalom:

- QUEMADA, B., L'exploration mécanique des dictionnaires bilingues: Index français-flamand du Vocabulaire de N.[oël] de Boerlaimont. BILAL IV. [1961], 9–49.
 Uő., L'inventaire des dictionnaires bilingues à propos du Dictionnaire néerlandais-français de N. de Berlaimont (1536). CL 2 [1960], 67–78.
 ŠTINDLOVÁ, J., Sur le classement des mots et sur ce qu'on appelle „Dictionnaire inverse”. CL 2 [1960], 79–86.
 MATER, E., Herstellung eines rückläufigen Wörterbuchs. FLP 4 [1962], 122–5.
 Problems in Lexicography. Edited by F. W. Houzeholder and S. Saporta. (Part IV. International Journal of American Linguistics, Volume 28, Number 2, April) 1962.
 TOLLENAERE, F. DE, Nieuwe Wegen in de Lexicologie. 1963, 108–9. (Moderne Lexicografie), 117–22.
 KELEMEN J., Gépi adatgyűjtés és adatfeldolgozás a lexikográfia szolgálatában. Szótár-tani Tanulmányok (szerkesztette Országh László) 1966: MNyTK. 114. sz. 39, 50–4.

9. A nyelvtanoknak a lyukkártyák segítségével való elemzésehez össze kell állítani a nyelvtani hagyomány nyelvtani fogalmait, nyelvtani szabályait és azok figyelembe vételével kell elkészíteni a lyukkártya-tervet. A nyelvtanok anyagának elemzése kiterjedhet a nyelvtani fogalmakra, azok rendszerezésére, az egyes nyelvtani szabályokra, azoknak egymással való kapcsolatba hozására, a nyelvtanban tárgyalt morfémák, szavak szó szerkezetek, mondatok, nyelvtani tárgyalására, a nyelvtani rendszer tárgyalásának részleteire és egészére, a nyelvtani szemléltető anyag elemzésére.

Irodalom:

- MITTERAND, H., Observation sur la description mécanographique des Formes grammaticales. BILAL I [1960], I–XV (17–31).
 BONNARD, H., Remarques pour la partie grammaticale du fichier. BILAL II [1960], 9–16.
 Uő., Proposition pour le codage des caractères grammaticaux. BILAL III [1960], 1–25.
 GOUGENHEIM, G., Tableaux de classement d'éléments grammaticaux et des constructions. BILAL III [1960], 26–35.
 GREIMAS, A. J., Remarques sur la description mécanographique des formes grammaticales. BILAL II [1960], I–XXV [17–41].
 BIERWISCH, M., Die Verwendung von Lochkarten bei der grammatischen Analyse. Mitteilungsblatt für Mitarbeiter der Deutschen Akademie der Wissenschaften zu Berlin 6 [1960], 255–8; FLP 4 [1962], 94–102.

10. A jelentéstani kutatásokat a lyukkártya-rendszerű szervezett szó- és morfémagyűjtés pusztá nyersanyagával is jelentős mértékben támogatja, mert a nagytömegű anyag részint a lyukkártyán (szövegkártyán) levő szövegkörnyezetével, részint — ahol szükséges — gépi úton könnyen kiegészíthető szélesebb szövegkörnyezettel az eddigi kutatásnál biztosabb alapot nyújt a jelentéstani elemzésekhez.

10.1. Az értelmező és történeti anyaggal dokumentáló szótárak lyukkártya-rendszerű elemzése közvetlenül, a szótári jelentések tipizálásával (pl. fogalomköri tipizálással, szinonimacsoportok egybeállításával, egyes szavak, szócsoportok jelentéshálózatának, a szócsoporton belüli jelentésváltozatoknak és analóg jelentésviszonyoknak meghatározásával) is támogatja a rendszeres jelentéstani kutatások előkészítését.

10.2. Végezhetünk sajátos szempontú jelentéstani kutatásokat is lyukkártyák felhasználásával. Kereshetjük pl., melyek azok az elemi jelentésmozzanatok, amelyek a bonyolultabb jelentésekből levezethetők, kielemezhetők, s amelyekből a bonyolultabb jelentések felépíthetők. Kereshetjük a nyelv-

rendszerben egyidejűleg élő többjelentésűségnek, a nyelvi jelhez tartozó jelentéseknek és jelentéscsoportoknak nyelvbeli beágyazottságát, rétegződését, használati kötöttségét. Vizsgálhatjuk a jelentésviszonyoknak, pl. a rokonértelműségnek, az ellentétes értelműségnek, az összefoglaló-általánosító és részletező-egyedítő jelentések viszonyának különféle fokozatait és módozatait, a részjelenségek arányait.

10.3. Alkalmazható a lyukkártya a jelentésváltozások tipizálásának vizsgálatára is: mind a régiebb és az újabb jelentés közti logikai viszony meghatározására, mind a jelentésváltozások feltételeinek, feltehető okainak vizsgálatára, ha a kódolást kellő körütekintéssel készítjük elő és végezzük.

11. A nyelv tudományi részlettanulmányok eredményeinek lyukkártyákon való elemzése a szótári-nyelvtani elemzéshez kapcsolódik, s részint azok kiegészítésére használható, részint egy átfogó nyelv-tudományi bibliográfiához ad részleteket.

12. A stilisztikai kutatásokban is többféleképpen használhatók a lyukkártyák, és részben már hasznosították is őket.

12.1. Különösen hasznos szolgálatot tehet a lyukkártyákkal végzett elemzés az egyes írók szókincsének, állandósult szókapcsolatainak összehasonlításában. Mivel az írók által használt szókincs rendkívül sok szóegyedet tartalmaz, az író kedvelt szavának kiválasztása és más írók kedvelt szavaival való összehasonlítása emberi erővel végzett elemzés esetén rendkívül lassan halad, éppen azért a legtöbb stíluskutatót ilyen munkától visszariasztó feladat. A frazeológiában a kisebb terjedelem mellett a változatokban való gazdagság jelent nehézséget kézi erővel való feldolgozásban. Gépi adatgyűjtésre támaszkodó statisztikai összehasonlítás viszont sokkal rövidebb idő alatt és sokkal pontosabban elvégezhető, mind a szókincsre, mind a frazeológiai anyagra vonatkozólag.

12.2. A stíluskutatás azonban nemcsak ilyen természetű eredményeket várhat a gépi adatfeldolgozástól. Hasznos adalékokat nyújthat a stilisztikai eszközök összegyűjtése és leltározása terén; a korstílus meghatározó jegyeinek összeválogatásában; az egyes műfajok hagyományos stíuselemeinek összegyűjtésében és történeti dokumentálásában; az egyes szavak, szókapcsolatok, fordulatok stilisztikai értékelésének tisztázásában, a stílusrétegek, stílusnemek tudományos vizsgálatában stb.

Irodalom:

PIGNAULT, J., Le vocabulaire d'Arnolphe dans l'École des Femmes de Molière. BILAL V [1961], 1–29. (Betűrendes szójegyzék Arnolphe fogalmi szavairól [mots pleins] i. h. 17–29.)

FINGERHUT, K. H., Recherches sur le vocabulaire de la Critique de Jazz en 1960–1961. BILAL V [1961], 30–42. (Etimológiai szócsoporthoz uo. 41–2.)

ANTAL LÁSZLÓ, Kritika III. [1965], 42–3 (Matematika és esztétika c. cikkében, a gépi-matematikai stíluselemzés néhány kérdése).

13. Verstan kutatások terén is történtek kísérletek a gépi feldolgozásra.

13.1. Verses szövegek elemzésében elsősorban az ütemek, verslábak, sorok tipizálásának és változatainak, a sorok tagolódásának és kapcsolódásának, a strófászerkezetek és rímek (rímtípusok) gyakoriságának vizsgálatától várhatunk új eredményeket.

13.2. A rímshótárak összeállításában különösen nagy szerepe lehet a gépi elemzésnek. Itt nemcsak az *a tergo* shótárak jelentőségére gondolok, amelyek kétségtelenül tetemes nyersanyagot tartalmaznak a rímshótárak számára, hanem a költők rímelő gyakorlatának gépi statisztikai elemzésére is.

Irodalom:

SEBEOK, TH. A. and ZEPS, V. J., Concordances and thesaurus of Cheremis poetic language. *Janua linguarum, Studia memoriae Nicolai van Wijk dedicata, Series maior VIII.* 's Gravenhage 1961.
KELEMEN J., i. h. 44.

14. A szövegváltozatoknak gépi úton való elemzése, a lyukkártyáknak a szövegvizsgálóban való hasznosítása szintén járható útnak bizonyult. Ebben a munkában az egyes változatok megfelelő részletei külön-külön lyukkártyákra kerülnek, a változat értékelését megkönnyítő megjegyzések kulcsszámaival. Pl. „legrégibb kézirat(ban)”, „újabb kézirat(ban)”, „kétségtelen másolási hibákat tartalmazó kézirat(ban)”, „bizonytalan korú kéziratban”, „nyomtatott szövegváltozat(ban) az... évből” stb. Más csoportokon egyes változatok egyes szövegrészeihez vagy szóalakjaihoz tartozó kulcsszámok, pl. „a szerzőnél szokatlan alak”, „más nyelvjárásra valló alak”, „utólagos beszúrás” (a sor fölött, vagy a lapszálon, vagy a sor alatt [a szerzőtől v. kortárstól v. későbbi kéztől]), „áthúzott szó”, „áthúzott szó után irt helyesbítés a szerzőtől”, „másolási hiba”, „sajtóhiba” stb.; „más olvasata is van”, „más olvasat is lehetséges”, „más olvasat szerint” ... stb.; „a szó eleje levágva (legyalulva)”, „a szó vége levágva (legyalulva)”; „a szó belsejében az [mp] betűk helyén szakadás” stb.; „olvashatatlanul áthúzott szó, hét betű”; „a [...] betűk áthúzva”, „a [...] betűk utólag beírva (a szerzőtől v. idegen kéztől)” stb. Változatokat egyesítő lyukkártyák tartalmazzák a változat sorszámát, a szöveg azon sorának sorszámával, amelyben a változat jelentkezik; a kéziratokat jelölő számokat vagy betűket a változat után, pl. 11. [váltakozás] (52. [sorban]) iam nunc CGHMSW [változata:] nunc iam T stb.

Irodalom:

FROGER, DOM. J., La critique textuelle et la méthode des groupes fautifs. CL. III. [1961/2.] 207–24.
GEISS, HEINZ, FLP. 4. [1962], 110.
MATER, ERICH, FLP 4. [1962], 209–10 (FUCKS vizsgálatairól).

15. A lyukkártyáknak ismeretlen jelrendszerek megfejtésében való alkalmazása részint azon a tényen alapul, hogy a szó- és morféma-készletnek a szövegben való eloszlása bizonyos törvényszerűséget mutat: az egyes szavak és morféma-k más-más gyakorisági számmal jellemezhetők; részint azon, hogy bizonyos típusú szóalakok elsősorban meghatározott típusú szóalakok szomszédságában szoktak előfordulni. Régóta ismert tény például, hogy a nyelvtani fogalmakat jelentő szavak a szöveg leggyakoribb szavai közé tartoznak, tehát bizonyos hosszúságú szövegben észrevehetően különválnak egyéb szavaktól. Azt is megfigyelések igazolják, hogy bizonyos szóalkakvégződések a nyelvek nagy részében meghatározott mondattani szerepre utalnak, mások a jelölt tárgyi jelentésről árulnak el egyet-mást. Ilyen szempontok tervszerű kódolásával, a szóképek gyakorisága és a sejthető végződések gyakorisága alapján támpontokat ad a gépi adatfeldolgozás ismer-

retlen írás megfejtéséhez, ismeretlen nyelvű nyelvemlékek nyelvének vizsgálatához.

Irodalom:

JEVREINOV, JE. V. – KOSAREV, F. G. – USTINOV, V. A., Primenenije elektronnyh vyčislitel'nyh mašin v issledovanii pis'mennosti drevnich majja. Novosibirsk, 1961.

GEISS, H., Aufgaben und Perspektiven der Arbeitsgruppe „Minoische Texte“ des Instituts für griechisch-römische Altertumskunde (Mitteilungsblatt für die Mitarbeiter der Deutschen Akademie der Wissenschaften zu Berlin. 6 [1960], 250–4.)

U6.: Untersuchungen an Texten mit unbekannter Schrift und Sprache mit Lochkarten. FLP. 4. [1962], 103–13.

16. A lyukkártyáknak bibliográfiák összeállítására való felhasználása szintén az utóbbi időben fejlődött ki.

16.1. A bibliográfia anyagát a decimális rendszer szerint megtervezve, az egyes egységeket sorszámokkal jelölik. A szerző nevét, a címet, a megjelenés helyét, ahol szükséges, az érintett tárgyköröket egy-egy lyukkártyamezőre kerülő betűérték nélküli számkombinációk lyukasztása vagy betűlyukasztás rögzíti a lyukkártyán. Az évszámot a megfelelő számjegyek lyukasztásával viszik rá a lyukkártyára. Több szerző esetén, gyűjteményes munkák feldolgozásában, hosszabb címek és alcímek esetén több lyukkártya összekapcsolása módot ad terjedelmesebb címleírás felvételére is.

Az így előkészített lyukkártyák rövid idő alatt elrendezhetők a kívánt szempontok szerint: tárgyköri csoportok, szerzők neve, időrend, megjelenési hely szerint. A rendezés szempontjai a célnak megfelelően bővíthetők. A leírt művek sorszáma alapján és az érintett tárgykörök kulcsszámai alapján mind az utalásoknak, mind a csoportok alkotásának tág teret biztosítanak a lyukkártyák.

Ilyen bibliográfiát adott ki pl. SCHEELE, M., Die Lochkartenverfahren in Forschung und Dokumentation mit besonderer Berücksichtigung der Biologie. Stuttgart 1959 (1954) c. művében (vö. FLP. 4. 1962. 85–7). A lyukkártyáknak bibliográfiai célú felhasználásáról RUZSICZKY ÉVA beszámolója is megemlékezik: Általános nyelvészeti tanulmányok II. 1962/4. 255–6.

16.2. Természetesen hasznosíthatók a lyukkártyák különféle mutatók (szómutatók, név- és tárgymutatók), r e s z - b i b l i o g r á f i á k a n y a g á n a k r e n d e z é s é r e. Így pl. a nyelvészeti folyóiratok évenkénti szó-, név- és tárgymutatóit, több kötetre vonatkozó „kalauzait”, „mutatóit”, évkönyvek, gyűjteményes munkák kötetében levő szó- és névjegyzékeket leggyorsabban lyukkártyákon egyesíthetjük és rendezhetjük. A rendezés többféle elv szerint történhetik. Pl. 1. folyamatos betűrendben, tipográfiai különbségekkel: a) szerző neve: verzális v. kapitálchen; b) tárgymutató: antiqua; c) szómutató: kurzív; d) régi eredeti írású nevek esetén: antiqua ritkített; régi eredeti írású közszavak esetén: kurzív ritkített; f) problematikus kategóriájú anyag esetén (kevés ilyen lesz!) jellel megkülönböztetett antiqua 2. A nyelvészeti mutatókban jól bevált hármas csoportosítás szerint: 2.1. Tartalomjegyzék: Név (álnév, jel): cím; 2.2. Tárgymutató (tulajdonnevek; fogalmi körülírások; földrajzi, társadalmi rétegződés, történelmi egységek stb. 2.3. Szómutató; szókapcsolatok mutatója. Az utóbbi újabb két oszlopon tovább bontható nyelvek szerint.

17. Röviden szeretnék kitérni arra a kérdésre, mit várhatunk a lyukkártyák alkalmazásától a nyelvtudományban.

17.1. Megfelelő méretű anyagi beruházás mellett hatalmas méretű, rendezett és újabb szempontok szerint gépi úton rövid idő alatt újrendezhető dokumentációs anyagot.

17.1.1. Ilyen anyagot várhatunk a szókincs különféle szempontú elemzéséhez, különféle szempontú szójegyzékekhez, szótárakhoz, pl. írásműszó-jegyzékek (indexek) részére (szövegkörnyezet nélkül); a szavakat szűkebb vagy tágabb szövegkörnyezetükkel bemutató írásműkonkordanciák részére; jelentések szerint rendezett anyagú írásmű-szótárak részére; írói konkordanciák részére; írói szótárak részére; különféle részleges szakszójegyzékek és szak-szótárak részére, a *tergo* szójegyzékek (szótár) részére; egyéb speciális szótárak, mint pl. rím��tár részére; különféle típusú áttekintő (összefoglaló) szótárak részére; így általános műszaki szótár részére; mai általános szótár (könyvelvi és irodalmi nyelvi szótár) részére; egyes korszakok szókincsét felölelő szótárak részére; — az akadémiai nagyszótár részére; — végül mindezek anyaga esetleg egyesíthető illetőleg bedolgozható egy nagy thesaurus-szerű összefoglaló gyűjteménybe esetleg ennek alapján kidolgozott szótárba.

Hogy ezek közül egyik, a thesaurus jellegű nagyszótár milyen arányokban való munkát jelenthet, arra idézem Bernard Quemada becslését a francia nyelvtörténeti szótár tervezett adattömegére vonatkozólag. Összesen mintegy 250 millió adatot kívánnak összegyűjteni a következő korszakolással:

XIII. sz.	20 000 000
XIV—XV. sz.	40 000 000
XVI. sz.	40 000 000
XVII—XVIII. sz.	60 000 000
XIX—XX. sz.	90,000 000
Összesen:	250 000 000

ebből 120 000 000 „teljes” v. „lexikális” szó, 130 000 000 — tehát több mint fele — „eszköz”-szó, grammatikális szó (CL. III. [1961/2] 121).³

17.1.2. Hatalmas méretű dokumentációs anyagot várhatunk a gépi adatfeldolgozástól egyes speciális morfémátárak és az összefoglaló morfémátár részére.

17.1.3. A szövegek lyukkártyarendszerű feldolgozása az eddigi vizsgálatoknál nagyobb arányú és rendezettebb hangtani archívum alapanyagául szolgálhat, és komoly segítséget jelent a tudományos leíró hangtani és hangtörténeti összefoglalások, valamint a helyesírástörténeti kutatások számára.

17.1.4. Megvethetjük a gépi adatfeldolgozással a magyar mondatnasiszintaktikai adattár alapját. Ez az anyag feltétlenül hasznos és értékes kiegészítése lesz az eddigi mondatnasiszintaktikai kutatásoknak.

17.1.5. Értékes stilisztikai adattár,

17.1.6. verstani adattár és

17.1.7. nyelvtudományi bibliográfia egészíti ki a gépi adatfeldolgozás szótári-nyelvtani anyagát.

17.2. Módszerben nemcsak azt várhatjuk a gépi adatfeldolgozás nyelvtudományi alkalmazásától hogy hozzászoktat bennünket a nyelvtatisztikai

³ Ugyancsak nagyméretű, thesaurus-jellegű nyelvtörténeti szótári gyűjtés tervéről számolt be az 1966-i prágai értekezleten A. Duro, az olasz nyelvtörténeti szótár szerkesztőségének igazgatója. A XIII—XX. sz.-ból kb. 100 000 000 lyukkártyát kívánnak összegyűjteni a szóalakok bőséges szövegkörnyezetének xerox-másolatával.

módszerű vizsgálatokhoz, hanem azt is, hogy a kódolás révén az eddiginél nagyobb gondot fordítunk a szempontok és szempontcsoportok következetes alkalmazására.

A gépi adatfeldolgozás előkészítése és a nagy tömegű adat nyelvtudományi értékelése a módszeres fogások és eljárások terén új lehetőségeket tár fel.

Több lyukkártyamező párhuzamos gépi elemzése olyan komplex összefüggéseket is felvillant, amelyeket eddig vagy nem, vagy kevesebb szempontból tettünk vizsgálat tárgyává.

17.3. A nyelvvizsgálat alapelvei szempontjából a gépi adatfeldolgozás gyakorlata új oldalról világítja meg a nyelvi jelenségek alogikus természetét. A gépek logikusan rendezett számain, logikus felépítésű számrendszerén elemzett alogikus nyelvi anyag jobban tudatosítja a nyelvnek logikus rendszerbe nem kényszeríthető alogikus elemeit és jelenségeit, de egyszersem azt is igazolja, hogy megfelelő kódolási elvek alkalmazásával kisebb térre szorítható a rendszerezhetetlennek hitt nyelvi jelenségek száma.

A nyelv rendszerszerűségének logikus felépítésű számrendszer segítségével való elemzése új oldalról világítja meg a nyelv rendszerszerűségének elvét. A szókinccs mint rendszerszerűen kapcsolódó elemekből álló halmaz, valamint a morfémakészlet, főleg a módosító (toldalék)-morfémák készlete, maga többirányú rendszerszerű kötöttségével a lyukkártya párhuzamos mezőin a valószínűsághoz hívebben elemezhető, mint az egyes elemi részletek mozaikszerű vizsgálatával.

A hangtanban főleg a hangkapcsolatok vizsgálatától várhatunk új elvi megállapításokat.

Nagy tömegű kronológiailag rendezett anyag biztosításával a gépi adatfeldolgozástól a nyelvtörténeti kutatás is új lendületet kaphat.

17.4. A gépi adatfeldolgozás nyomán nyilván egész sor olyan mű készül el évek távlatában, amelyek hatalmas dokumentációs anyagra támaszkodva szabatosabban fogalmazzák meg nyelvi törvényszerűségeket, változási tendenciákat, árnyaltabban fejtegetnek nyelvi jelenségesoportokat, mint a gépi adatfeldolgozás nélkül készült művek.

A legközelebb várható művek közé tartozik nyelvünk *a tergo* szójegyzéke (szótára).

Szándékunkban van a kísérletek után egy hosszabb írásművet vagy egy írói életművet gépi adatfeldolgozás segítségével monográfiaszerűen feldolgozni.

Ha megfelelő anyagi támogatást kapunk, néhány éven belül elkészítünk egy magyar gyakorisági szótárt, amely írók, műfajok és szövegtípusok szerinti gyakorisági szójegyzékeket is tartalmazni fog.

Megfelelő anyagi és személyi támogatás mellett elkészítünk egy hasonlóan tagolt gyakorisági morfémátárat.

18. A gépi adatfeldolgozás jelenlegi súlyponti kérdését abban látom, hogy a már megkezdett munkálatokat befejezzük, illetőleg abba a szakaszba juttassuk, hogy a folyamatos munka biztosítva legyen. Tehát segíteni kell az *a tergo* szójegyzék elkészítését; anyagi támogatást, megfelelő gépeket kell biztosítani a szövegekből történő lyukkártyarendszerű szó- és morfémagyűjtés számára.

Kívánatosnak tartom a lyukkártyarendszerű és egyéb gépi adatfeldolgozás szervezetszerű biztosítását, lehetőleg a Magyar Tudományos Akadémia Nyelvtudományi Intézetének keretén belül.

Biztosítani kell azt a minimális személyi keretet, amellyel a folyamatos lyukasztás (lyukszalagé és lyukkártyaé), a lyukasztás ellenőrzése, a rendezés, válogatás, a jegyzékek (szó-, morféma- és egyéb jegyzékek) készítése és a feliratozás fennakadás nélkül végezhető.

A munka sikeres végzéséhez biztosítani kell olyan helyiségeket (kb. 4–6 helyiséget), ahol a géppark elhelyezhető és olyan temperált hőmérsékletű és meghatározott nedvességtartalmú további helyiséget, ahol a lyukkártyák tárolhatók. A helyiségek kiválasztásakor figyelembe kell venni azt a körülményt, hogy a gépek működése meglehetősen zajjal jár, ezeknek közvetlen közelében hangszigetelés nélkül tehát nagy elmélyedést kívánó munka nem végezhető zavartalanul.

Meg kell tenni a hivatalos lépéseket a magyar helyesíráshoz alkalmazott géppark megszerzésére.

19. Bizonyára felmerült már a hallgatóság körében az a gondolat, hogy azt is meg kellene nézni, vajon a lyukkártyák alkalmazása milyen előnyök és hátrányok között jelent és milyen nehézségeket rejt magában. Mert az a tény, hogy valamely munka gépi úton végezhető, hogy használhatóságát eredményes kísérletek igazolják, még nem jelenti feltétlenül azt, hogy gazdaságilag vagy a munka szervezése szempontjából olyan mérvű előnyöket jelent, amelyek megokolják a gépek beszerzését és üzemeltetését.

A lyukkártyarendszerű adatfeldolgozás előnyeit ma már számos országban folyó munka igazolja. Ezek közül az alábbiakat emelem ki:

19.1. A dokumentációra szolgáló szöveget csak egyszer kell leírni és ellenőrizni, egyes részleteit akárhány példányban kívánjuk felhasználni. A másolás gépi úton, hibátlanabban, gyorsabban, olcsóbban végezhető, mint írógéppel történő szövegggyűjtéssel, nem is szólva a kézírással történő gyűjtésről.

A másolatoknak gépi úton való gyors előállítására lehetővé teszi, hogy akár nagyobb anyagrészen is egyidejűleg párhuzamosan dolgozhassék két vagy több kutató, anélkül, hogy egymást zavarnák a munkában. Erre sor kerülhet például nagyobb létszámú szerkesztőség munkájának összehangolásában.

19.2. Az első ellenőrzést más gépirónó végzi, mint a lyukasztást, ugyanolyan billentyűzetű gépen, mint a lyukasztógépé, de nem új lyukakat a lyukkártyába, hanem a gépirás helyességét ellenőrzi azzal, hogy ha a lyukasztásnak megfelelő billenyűket üti le, a kártya zavartalanul továbbmegy a gépben, ha más billentyűt üt le, a gép azonnal megáll. Az ellenőrzésnek ez a módja azon a lélektani megfigyelésen alapul, hogy ugyanazt a munkafolyamatot végző két személy nem szokta ugyanazokat a hibákat elkövetni. Ez nem teljes biztosíték ugyan a hibák kiküszöbölésére, de az eddigi beszámolókból nem látszik annak nyoma, hogy az ellenőrzésnek ez a módja ne lenne elég hatásos.

Azzal, hogy az első ellenőrzést gépirónó végzi, az adatellenőrzés olcsóbb lesz. A filológiai ellenőrzést végző tudományos segédmunkatárs gyorsabban végezheti munkáját, mert kevesebb javítanivaló marad a szövegben.

Azzal, hogy a gépelt szöveget csak egyszer kell filológiailag ellenőrizni, — mert az ellenőrzött szövegrészt a gép hibátlanul sokszorosítja, gépirásos szövegeknél legalább annyiszor, ahányszor gépelik, s hiba esetén az egyszeri gépelés példányait végig kell javítani, — a gépi adatfeldolgozás ellen-

őrző munkarészlete gyorsabb is, olcsóbb is, mint az írógépen indigóval sokszorosított adatoké, nem is szólva a kézírással készített másolatok ellenőrzésének idő- és költségtöbbletéről.

A xerox-másolással összekapcsolt lyukkártyákon pedig a szövegekörnyezetet egyáltalán nem kell filológiailag ellenőrizni, csak a címszó vagy a címszóalak (lemma) lyukasztását és feliratozását.

19.3. Általában nagy jelentőségűnek tartom a lyukkártyák alkalmazásában a r é s z l e g e s g é p i m á s o l á s t. Ez módot ad a lyukasztott szövegrészek szinte számtalan szempontú feldolgozására, azzal, hogy a szóalakok, vagy szövegrészeket nyelvi anyagát megfelelő kódolás alkalmazásával újabb és újabb szempontú elemzésnek vethetjük alá.

A részleges másolás a szempontok gazdagításának lehetővé tételével az ismert tények és okok összefüggéseinek olyan mérvű r e n d s z e r e s k i a k n á z á s á r a ad alapot, amilyenre eddig nem is gondolhattunk.

19.4. De nemcsak a szöveg leírása, illetőleg lyukasztása, ellenőrzése és a szövegrészek sokszorosítása terén vannak kétségtelen előnyei a gépi adatfeldolgozásnak, hanem az a d a t o k r e n d e z é s e terén is. Ez pedig sokkal fontosabb az eddigieknél, mert a gépi adatrendezés összehasonlíthatatlanul gyorsabb és lényegesen olcsóbb a kézi erővel végzett adatrendezésnél. Ebben a vonatkozásban a lyukkártyarendszerű adatfeldolgozás a fényképezés szokásos módszereivel előállított cédulákkal s a szöveg példányainak felvágott anyaggyűjteményével szemben is szembetűnő előnyöket mutat.

A lyukkártyarendszerű gépi adatrendezés lehetőséget nyújt arra, hogy e g y i d e j ű l e g t ö b b f é l e r e n d e z é s i szempont érvényesüljön, hogy egymást keresztező szempontok érvényesítésével, rövid idő alatt, kisebb csoportokra bontsuk a rendezésre váró anyagot.

19.5. Az ilyen adatfeldolgozás lehetővé teszi a megvizsgált szöveg szavainak s z ó j e g y z é k b e f o g l a l á s á t, akár betűrendben, akár meghatározott csoportok szerint. Az ilyen szójegyzékek rövid idő alatt elkészíthetők. Egy német kutató, G. HÜBNER szerint pl. egyetlen IBM. 1410. jelzésű géppel több munkatárs sok hónapi munkája néhány nap alatt elvégezhető. (CL. III. 1961/2. 100).

19.6. Lehetőséget nyújt a gépi adatfeldolgozás a szövegnek vagy szövegeknek sokoldalú statisztikai vizsgálatára, így pl. az egy műben, egy író összes műveiben való szógyakoriság megállapítására (kétféle rendszerben: a szavak betűrendjében, a gyakorisági mutatóval; a szavak előfordulási számának fogyó rendjében), kronológiailag vagy műfajilag vagy egyéb szempont szerint rendezve is; az egy műben vagy egy író összes műveiben levő összes morfémák, fontosabb morfémakapcsolatok gyakoriságának megállapítására.

19.7. A lyukkártyarendszerű adatfeldolgozás kellő beruházás mellett lehetőséget biztosít olyan k o m p l e x f e l a d a t o k megoldására, mint egyes művek vagy egyes írói életművek teljes szókincsére és teljes morfémakészletére támaszkodó beható elemzése, esetleg hasonló jellegű mondattani-szintaktikai és hangtani elemzéssel kiegészítve, a szótári és nyelvtani statisztikai elemzések tanulságainak egymásra vetítése. Ilyen komplex feladatok kézi erővel szinte el sem végezhetők.

19.8. Gépi adatfeldolgozással hitelesen rávetíthető pl. egy-egy s z ó t á r í r ó saját műveinek s z ó k i n c s e a szótáríró s z ó t á r á b a n vagy szótáraiban feldolgozott szókészletre, akár gyakorisági szempontok figyelembe

vételével is (pl. Szenczi Molnár Albert, Baróti Szabó Dávid, Czuczor Gergely, Fogarasi János).

19.9. Gépi adatfeldolgozással lehetőség nyílik olyan jellegű vizsgálatokra is, hogy egy-egy nyelvtaníró nyelvteni rendszerezése hogyan viszonylik ahhoz a nyelvteni képhez, amely szöveges műveinek nyelvhasználatában jelentkezik (pl. Szenczi Molnár Albert, Verseggy Ferenc, Révai Miklós, Fogarasi János).

19.10. Nagyméretű kollektív munkák (pl. tudományos igényű nagyszótárak, nagy nyelvtanok) készítésében nemcsak az anyag összegyűjtésében és az összegyűjtött anyag rendezésében hasznosíthatók jól a lyukkártyarendszerű gépek, hanem a mű szerkesztése közben is, pl. a megoldandó kérdésekkel rokon jelenségek összegyűjtésére, a már elkészült részletekben alkalmazott megoldási módok nyilvántartására és rendszerbe foglalására; szótári munkában, ezenkívül a címszókiválasztás előkészítésére, illetőleg elvszerűbbé tételére stb.

A gépi adatrendezésnek szerkesztés közben való hasznosítását megkezdték pl. a Mai Német Nyelv Szótára (Wörterbuch der Deutschen Gegenwärtssprache) szerkesztőségében.

Irodalom:

- KLAPPENBACH, R., L'emploi des Cartes perforées dans le dictionnaire de [la] Langue Allemande Contemporaine. CL II. [1960/1.], 87–92.
 BLUMRICH, M. – TELLENBACH, E., Die Anwendung von Lochkarten im Wörterbuch der Deutschen Gegenwartssprache. FLP 4. [1962], 88–93.
 RUZSICZKY É., Az elektromechanikus adatfeldolgozó gép nyelv tudományi alkalmazásának néhány kérdéséről. 5. A lyukkártyamódszer alkalmazása a lexikográfiában. Ált. nyelvészeti tanulmányok. II. [1962/4.], 258–60.
 KELEMEN J., i. h. 47–50.

Kezdeményezésük megmutatta, hogy az út járható: a szótár szerkesztése közben szerzett tapasztalatok, megoldási módok, analóg esetek gépi úton gyorsan és megbízhatóan összegyűjthetők, és a szerkesztés további folyamán jól hasznosíthatók. De helyes kezdeményezésük tovább fejleszthető. Ők ugyanis megálltak a gépesítésben a lyukkártyák bizonyos szempontok szerint való rendezésénél, az elrendezett lyukkártyák adatait már gépirónók gépelik külön jegyzékbe, s ez lassítja a szerkesztés egyes munkameneteit. Ha numerikus rendszerű géppark helyett alfanumerikus gépparkon végezzük el a szerkesztés közben szükségessé váló rendezői munkákat, akkor a jegyzékbe foglalást is táblázó gépen végeztethetjük el, s ez gyorsítja mind a rendezés további műveleteit, mind a szükséges jegyzékek elkészítését, akár több példányban való elkészítését is.

19.11. Mindezek alapján kétségtelenül megállapíthatónak tartom, hogy a lyukkártyarendszerű adatfeldolgozás egyik legfőbb előnye az alkotó tudós szellemi energiájának nagymértékben való felszabadítása az idegkimerítő mechanikus munkák egész sorának nyúge alól, s ezzel alkotó erejének a tulajdonképpeni tudományos kérdésekre való összpontosításában egyik legmegbízhatóbb támogatója. A gép sohasem lesz ursa az alkotó tudósnak, csak sokoldalúan felhasználható engedelmes szolgálja.

Jól tudjuk, hogy az igazi tudós anyaggyűjtés közben is gondolkodik, érleli gondolatait, de az is kétségtelen, hogy az anyaggyűjtéssel és rendezéssel járó megfeszített figyelés, az ezzel járó elfásulás, kimerülés gátolja is gondo-

latainak kibontakozását, megindult gondolatmenetének végiggondolását. A gépi adatfeldolgozás tehát ebben a vonatkozásban is jótékony hatással lesz az alkotó tudásra.

20. A lyukkártyák alkalmazásának természetesen vannak bizonyos korlátai, fogyatékoságai és a velük való munkát nehezítő körülményei is.

20.1. A helyes kódolás, különösen a sokszempontú szöveggyűjtés helyes kódolása a gépek működésének bizonyos fokú ismeretét is megköveteli a kódolótól. Vannak ugyanis olyan számsoregyüttesek, amelyeket a gép nem tud megkülönböztetni. Ennek alapkövetelménye, hogy ugyanannak a lyukkártyának minden mezőjén csak egyetlen szám vagy számkombináció lyukasztható. Ez tehát korlátokat szab a lyukkártya-mező kihasználásának és a különféle mezők kombinálásának. Ezért van szükség például a névszói személyragozás elemzéséhez két, a szótóváltozatok elemzéséhez három számszlopra, a képzők és változataik elemzéséhez, úgy látszik, legalább négy számszlopra. Ezért van szükség a lexikológiai-lexikográfiai vagy a morfológiai elemzéshez legtöbb esetben legalább két lyukkártyára: egyre a megfelelő szövegkörnyezet miatt, egyre a lexikológiai-lexikográfiai-morfológiai elemzés miatt. A kettős kartonokat viszont azonossági számmal kell ellátni, s ez lyukasztásban is, adatrendezésben is idővesztéssel jár.

20.2. A lyukkártyák használhatósága a kód helyes kidolgozásának függvénye. Bizonyos vonatkozásban azt mondhatjuk, a lyukkártya nem ad új információt, csak azt adja vissza, amit előre kidolgozott terv szerint beleviszünk. Ez természetesen igaz a belevitt részletek számára és minőségére vonatkoztatva. De nem helyes ez az állásfoglalás abból a szempontból, hogy új információt a lyukkártyára rávitt információk kombinációja és statisztikai elemzése is ad. A kombinációkat tartalmazó lyukkártyák összerendezése és megszámlálása pedig összehasonlíthatatlanul gyorsabb és pontosabb a gépi adatfeldolgozás zavartalan biztosítása esetén, mint emberi erővel végzett rendezés esetén.

20.3. A lyukkártyák anyagának, a kartonnak esetleges gyengébb minősége érezteti hatását az anyag feldolgozásában: a kártyák törése, szakadása késlelteti a munkát és szaporítja a hibalehetőségeket. Ez persze kiküszöbölhető jó minőségű lyukkártyáknak nagyobb mennyiségben való biztosításával.

20.4. Bér munka esetén a lyukasztó és feldolgozó üzem egyébmunkáinak függvényeként a tervbe vett nyelvészeti kutatás folyamatosága akadozik. Azonkívül, ha a lyukasztás és a feldolgozás távollevő üzemben történik, aligha ellenőrizhető az anyag teljessége, az esetleg sérült kártyák újra lyukasztásának hibátlan volta. — Ezen úgy lehet legegyszerűbben segíteni, ha az Akadémia saját gépparkot rendez be.

20.5. A lyukkártyák tárolására temperált hőmérsékletű és a levegő megfelelő nedvességtartalmát biztosító helyiségre van szükség. Ez az adottság elhelyezési nehézségeket jelenthet.

20.6. A gépek zaja zavarja a nagy elmélyülést kívánó kutató munkát. A gépek zajának szigetelése viszont csak jelentős költséggel oldható meg.

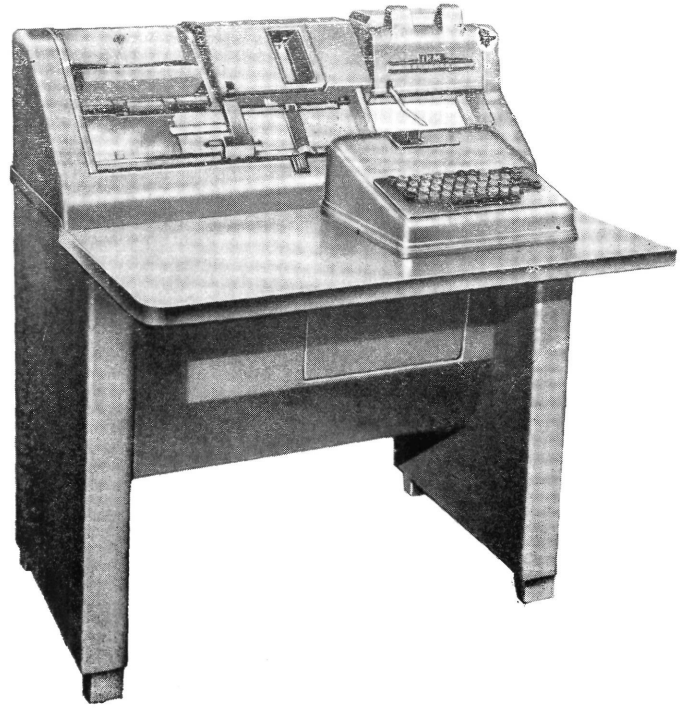
20.7. Legfőbb kifogás a gépesítés bevezetése ellen rendszerint az, hogy tetemes összegű beruházással valósítható meg. Ez természetesen helytálló ellenvetés, ha kis méretű munkák végzésére gondolunk, de



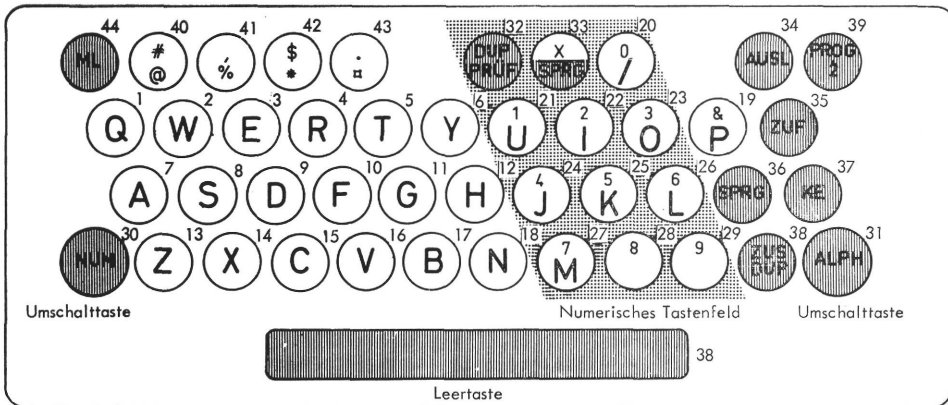
1. ábra.



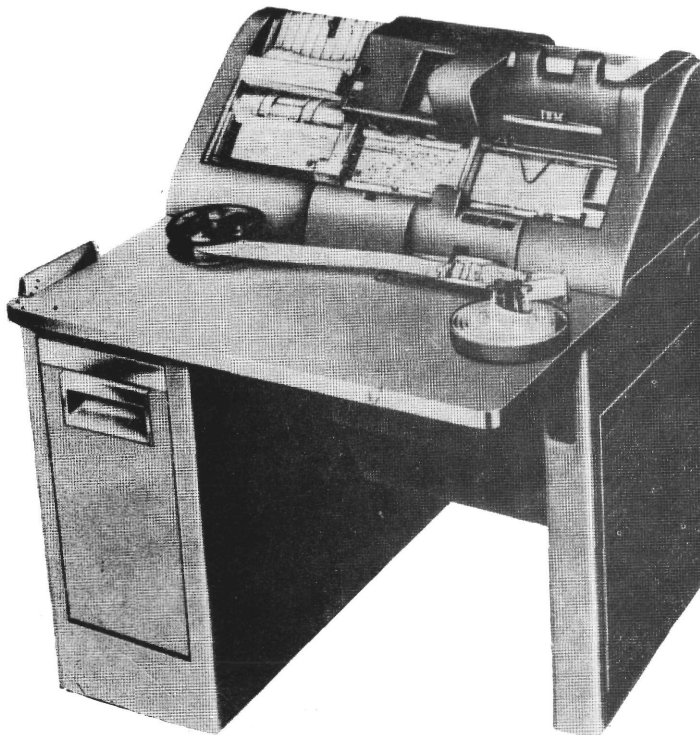
2. ábra. IBM kártyalyukasztó



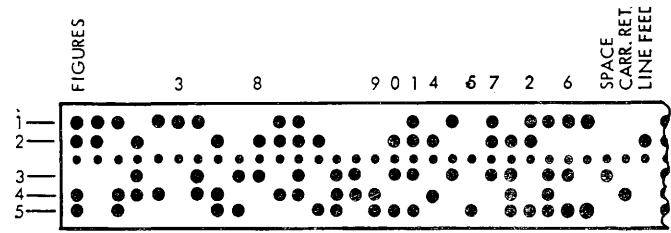
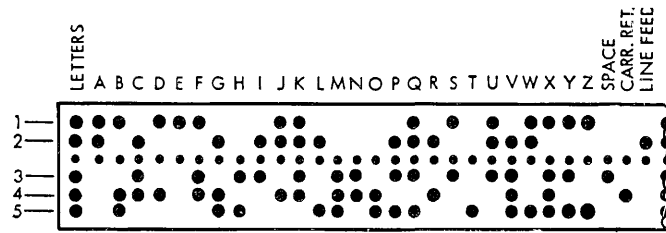
3. ábra. IBM 56 típusú lyukasztásellenőrző
(kártyavizsgáló kontrollgép)



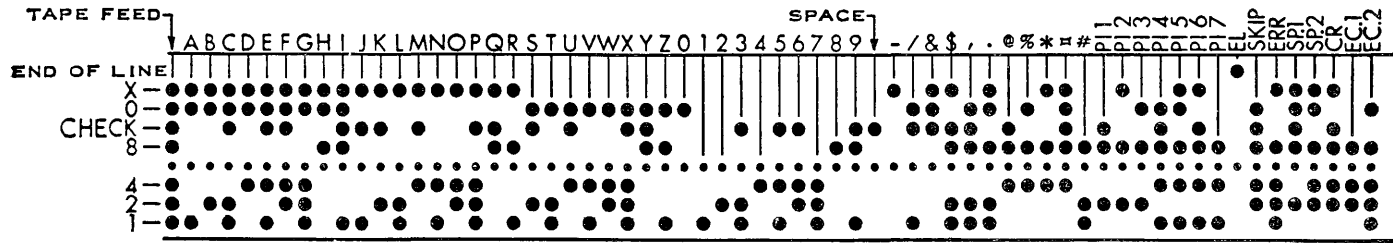
4. ábra. A lyukasztó- és az 56 típusú lyukasztásellenőrző alfanumerikus billentyűzete



5. ábra. Az IBM. 46—47 típusú szalagból kártyát lyukasztó gép (konverter),
5 és 8 csatornás lyukszalagra



6/a ábra. Az 5 csatornás lyukszalag betű- és jelkódja



6/b ábra. A 8 csatornás lyukszalag betű- és jelkódja

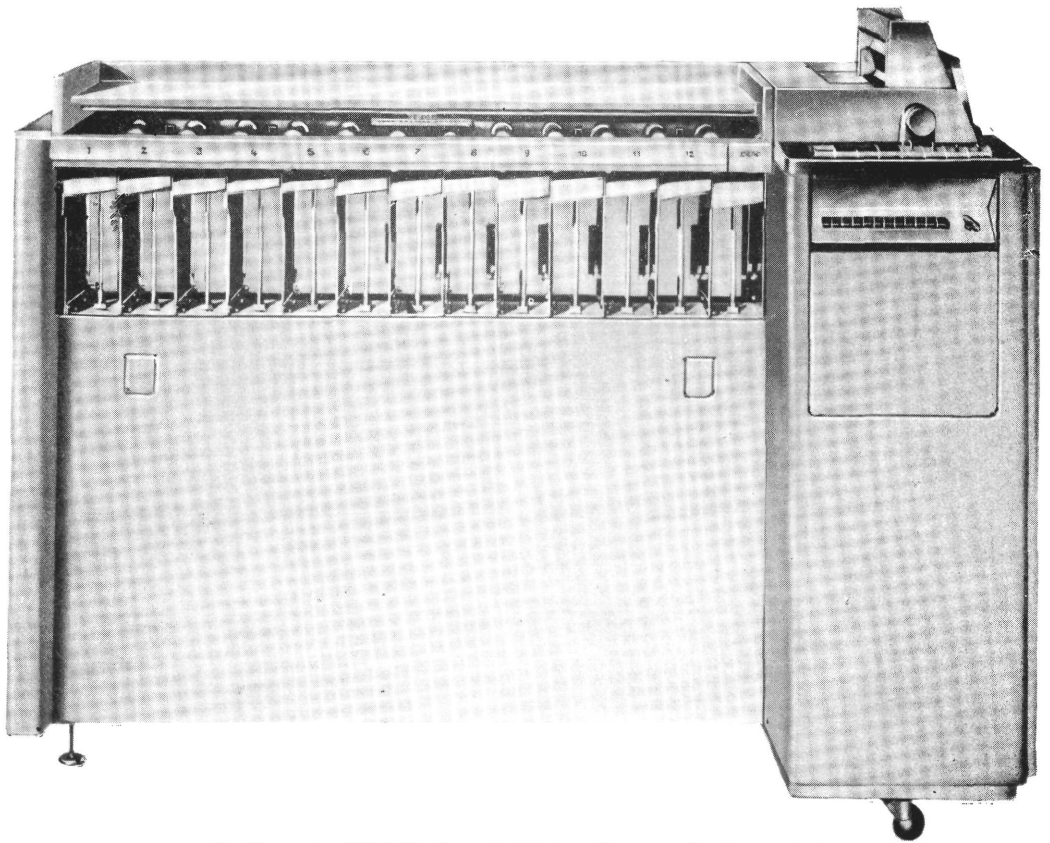
CHARACTER		CARD CODE	BCD CODE (Core Storage)						
Report	Program		C	B	A	8	4	2	1
b		No Punches	C						
.		12-3-8		B	A	8		2	1
□)	12-4-8	C	B	A	8	4		
		12-5-8		B	A	8	4		1
<		12-6-8		B	A	8	4	2	
≠		12-7-8	C	B	A	8	4	2	1
&	+	12	C	B	A				
\$		11-3-8	C	B		8		2	1
*		11-4-8		B		8	4		
}		11-5-8	C	B		8	4		1
;		11-6-8	C	B		8	4	2	
△		11-7-8		B		8	4	2	1
-		11		B					
/		0-1	C		A				1
,		0-3-8	C		A	8		2	1
%	(0-4-8			A	8	4		
~		0-5-8	C		A	8	4		1
\		0-6-8	C		A	8	4	2	
#		0-7-8			A	8	4	2	1
6		2-8			A				
#	=	3-8				8		2	1
@		4-8	C			8	4		
:		5-8				8	4		1
>		6-8				8	4	2	
√		7-8	C			8	4	2	1
?		12-0	C	B	A	8		2	
A		12-1		B	A				1
B		12-2		B	A			2	
C		12-3	C	B	A			2	1
D		12-4		B	A		4		
E		12-5	C	B	A		4		1
F		12-6	C	B	A		4	2	
G		12-7		B	A		4	2	1
H		12-8		B	A	8			
I		12-9	C	B	A	8			1
!		11-0		B		8		2	
J		11-1	C	B					1
K		11-2	C	B				2	
L		11-3		B				2	1
M		11-4	C	B			4		
N		11-5		B			4		1
O		11-6		B			4	2	
P		11-7	C	B			4	2	1
Q		11-8	C	B		8			
R		11-9		B		8			1
≠		0-2-8			A	8		2	
S		0-2	C		A			2	
T		0-3			A			2	1
U		0-4	C		A		4		
V		0-5			A		4		1
W		0-6			A		4	2	
X		0-7	C		A		4	2	1
Y		0-8	C		A	8			
Z		0-9			A	8			1
#	0	0	C			8		2	
1	1	1							1
2	2	2						2	
3	3	3	C					2	1
4	4	4					4		
5	5	5	C				4		1
6*	6	6	C				4	2	
7	7	7					4	2	1
8	8	8				8			
9	9	9	C			8			1

COLLATING SEQUENCE

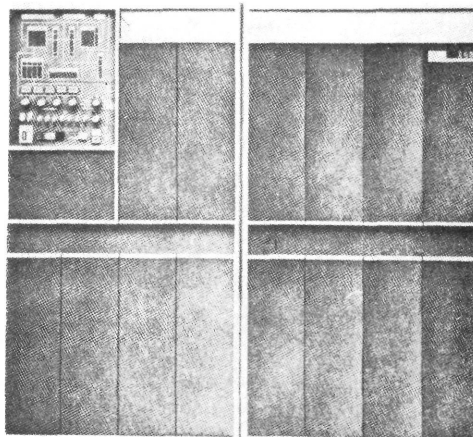
Low →

High →

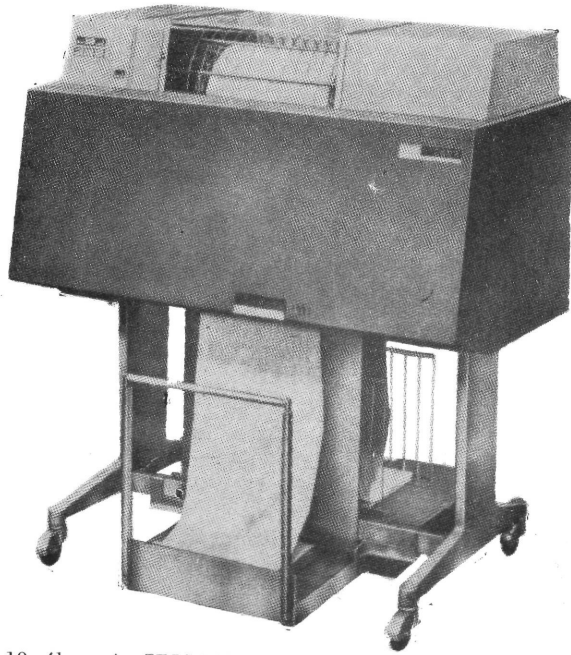
7. ábra. Szabvány BCD betű- és jelkód 69 egységre (karakterre)



8. ábra. Az IBM 83 típusú elektronikus rendező (szorter)



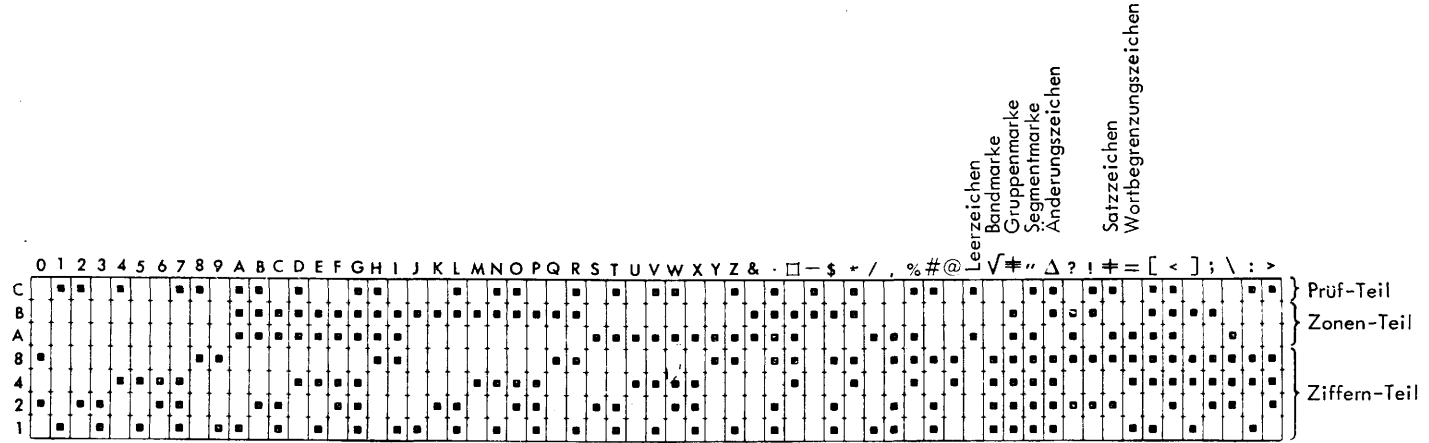
9. ábra. Az IBM 1401 típusú központi egység



10. ábra. Az IBM 1403 típusú kiíró (nyomó, printer)



11. ábra. Az IBM 557 típusú feliratozó (interpreter)



12. ábra. A 7 csatornás mágnesszalagkód

nem helytálló, ha a jövőben elvégzendő nagy méretű feladatok megoldása szempontjából nézzük és sokévi teljesítmény egészéhez viszonyítjuk a beruházás összegét.

20.8. Másik súlyos kifogás a gépek ellen az, hogy azok működtetéséhez sok emberre lesz szükség. De ez az ellenvetés is veszít jelentőségéből, ha ahhoz a nagyméretű és sokrétű feladathoz viszonyítjuk, amit az ilyen jellegű kutatásoktól várunk és joggal várhatunk. A vázlatosan érintett sokféle munka egyes részeinek gépi végzésével anyagilag is megtérül az a befektetés, amit a géppark ára és üzemeltetése s a gépparkot foglalkoztató személyzet fizetése összességében jelent. Pedig a pénzbeli megtakarításnál jóval többet jelent, hogy nagyméretű munkák belátható időn belül elkészülnek és tervszerűen gyűjtött és megbízhatóan rendezett állapotban állnak a műveket megalkotó kutatók rendelkezésére.

20.9. Végül kétségtelen, hogy a nálunk használt lyukkártyagépek betűkészlete, írása nem kielégítő, de nem nyugodhatunk bele abba, hogy a kísérleti szakaszon túl, végleges adatfeldolgozásra a technika mai fejlettségi fokán ne olyan géppark álljon az Akadémia rendelkezésére, amely alkalmas a mai magyar helyesírás szokásos betűivel és jeleivel való szövegírásra.

21. Engedjék meg, hogy röviden áttekintsem, mi történt eddig a gépesítés terén nálunk.

21.1. Az első lépés néhány idevágó tárgyú ismertetés volt (SZÉPE GY.: I. OK. 19 [1962]. 364–6; RUZSICZKY É. idézett referátuma: Általános nyelvészeti tanulmányok II. [1962/4.] 253–61.).

21.2. A Magyar Tudományos Akadémia I. Osztályának megkeresésére Intézetünk igazgatósága 1963-ban engem bízott meg a nyelvészeti kutatások részleges gépesítésének előkészítésével. Még ez év májusában előkészítettem egy kisebb szójegyzéket az ÉrtSz. címszóanyagából, s kijelöltem egyes szavakat folyamatos szövegből, hogy a lyukkártyáknak a szótári elemző munkában és valamely szöveg szavainak összegyűjtésében való használhatóságáról meggyőződhessünk. A kísérlet eredménye pozitív volt.

Ennek alapján 1963-ban és 1964-ben több javaslatot készítettem ebben a tárgyban; ezeknek egy részét Intézetünk dolgozóival szűkebb körű megbeszélésen vitattuk meg: ezen elsősorban a szótári tervmunkák vezetői vettek részt. Egyik javaslatot 1963. július 8-án az Akadémia I. Osztálya is megvitatta és támogatásáról biztosította a tervezett munkálatokat.

1963. szeptemberében összeállítottam az ideiglenes lyukkártyatervet a folyamatos szövegekből való szó- és morfémagyűjtéshez, valamint a szótárak elemzéséhez. (Ennek egyik másolatát tájékoztatóul közöltem Papp Ferencsel is.)

Ezzel párhuzamosan elkészítettem a folyamatos szöveg átírásának, lyukkártyára vitelének kódját a HOLLERITH-rendszerű lyukkártyáknak nálunk használatos betűjelölésére való tekintettel. Ennek másolatát szintén közöltem Papp Ferencsel és debreceni munkatársaival.

Előkészítettem Móricz Zsigmond egyik Csibe-novellájának (Csibe. 1936/55. ÖM. Elb. VI, 506–16) lyukkártyákra vitelét (1963. IX. 29). A kódolás szempontjait Dénes Józseffel és a Statisztikai Hivatal egyik szakemberével, Karsai Istvánnal megvitattam és a Statisztikai Hivatal gépparkjának adottságaihoz alkalmaztam (1964. május). De az anyag lyukkártyákra vitelére mindeddig nem kerülhetett sor. Részint anyagi okok miatt, részint a gépek kapacitásának más munkák részére való igénybevétele miatt.

1963-ban megkezdtem a részletes lyukkártyaterv kidolgozását. Ebből elkészült a szófajok és a szófaji alosztályok jelölésének terve; az egyes szófajok részletes elemzésének tervei közül az ige, a főnév, a melléknév, a számnév, a névmás és a névelő és a mondatszók kódolási terve; előkészületben van az igevevek, a határozószók, a névutók és a kötőszók kódolási terve.

Kidolgoztam a szótövek elemzésének részletes tervei közül a névszótövek kódolásának tervét (másolatban közöltem a debreceni munkaközösség vezetőjével), majd az igetövek kódolásának részletes tervét is.

Befejezéshez közel van a szóismétlés, az ikerítés, a szóösszevonás és a szóösszetétel, továbbá a képzők, képzőbokrok, képzősorok és képzőszerű szóvégek és fontosabb változataik kódolási terve. Előkészületben van a névszóragozás és az igeragozás részletes kódolásának terve.

Úgy gondolom, ennek az évnek végére vagy a jövő év első negyedében teljesen elkészül a folyamatos szövegekből való szó- és morfémagyűjtés részletes terve, s a jövő év második felében kipróbálhatjuk a részletes lyukkártyaterv összes részleteinek a gépi aatfeldolgozásban való érvényesítését.

Közben, 1964. júniusában a magyar–német kulturális egyezmény keretében Berlinben helyszínen tanulmányozhattam a hasonló jellegű munkákat: megtekinthettem a gépi adatrendezésnek a R. Klappenbach vezette szótári munkában való felhasználását; megismerkedhettem E. Maternek lyukszalagon és lyukkártyákon végzett munkájával, e munkák szervezési kérdéseivel. Berlinből visszatérőben Prágában pedig betekintést nyertem J. ŠTINDLOVÁ gépi adatfeldolgozó munkájába.

Jelenlegi beszámolómm elkészítésében berlini és prágai személyes tapasztalataimat is felhasználtam.

21.3. A részletes lyukkártyaterv kidolgozásának előkészítésében részt vett RUZSICZKY ÉVA: a képzők első jegyzékét ő állította össze SIMONYI Tüzetes Magyar Nyelvtana (1895), Laczkó Géza: A játszi szóképzés (1907), D. BARTHA KATALIN: Magyar Történeti Alaktan. Szóképzés (1959) és az akadémiai leíró nyelvtannak (I. 1961) a képzőkről szóló fejezete alapján. Ugyancsak ő vállalta a névszóragozás elemzésének előkészítését.

21.4. Az előkészületek kibernetikai vonatkozásaira Dénes Józseftől kaptam értékes felvilágosításokat.

21.5. Ezekkel az előkészületekkel csaknem egyidőben, Papp Ferenc vezetésével a debreceni Kossuth Lajos Tudományegyetem Matematikai és Alkalmazott Nyelvészeti Csoportja hozzákezdett a magyar *a tergo* szójegyzék előkészítéséhez; majd az eredeti tervtől eltérően kiegészítették az *a tergo* szójegyzék tervét az ÉrtSz. címszóanyagának szóalaktani, etimológiai és stilisztikai szempontú elemzésével.

22. Meg vagyok győződve arról, hogy az itt érintett részletkutatásokon kívül számos egyéb területen is bevezethető lesz a gépi adatfeldolgozás. De természetesen helytelen lenne azt várnunk, hogy a gépesítés bevezetésével azonnal gyökeresen megváltozik a nyelvtudományi kutatómunka; hogy néhány hónap, esetleg egy-két év forradalmi változásokat eredményez a nyelvtudományban. Helytelen lenne úgy felfognunk a gépesítés kérdését, hogy annak bevezetésével a régi gyűjtő- és rendező munka feleslegessé válik. A régi és az új munkamódszer még sokáig egymás mellett, egymást kiegészítve fog szerepelni; a nyelvész egyéni kutatásához ezután is fog egyéni gyűjtőmunkát végezni, ritka nyelvi jelenségek elemzésében ezután sem fognak gépi adatrendezés végeztetni; s ha a gépi adatrendezés során kiderül, hogy egy-egy jelenségcso-

portra nagy tömegű nyelvi adatból is kevés példa akad, ezeket bizonyára továbbra is gépek igénybevétele nélkül maga rendezi tovább a kutató. De másrészt azt is valószínűnek tartom, hogy nagy tömegű anyaggyűjtést és adatrendezést kívánó általános érdekű, nagyméretű vállalkozások megvalósításában egyre nagyobb szerephez kell jutniuk a gépeknek, s megfelelő beruházás esetén egyre nagyobb szerephez is fognak jutni.

Azt is látnunk kell: még hosszú az út a gépesítés nagyszabású várható eredményeinek jelentkezéséig. De kitartó munkával, tervszerű előkészítéssel, kellő anyagi beruházással el fogunk jutni arra a fejlettségi fokra, amikor beérik a ma még csak sejtett, a legközelebbi években elvetni remélt mag gazdag termése.

A gazdagságon nemcsak a statisztikailag értékelt adatok nagy tömegét értem, hanem a munka közben felmerülő új szempontoknak, új gondolatkapcsolatoknak várható gazdagságát is.

Hiszem, hogy ebben a várakozásban, a technikai fejlődés hatalmas méreteinek korában nem kell csalódnunk, nem fogunk csalódnunk.¹

Rövidítések:

- BILAL:** Bulletin d'Information du Laboratoire d'Analyse Lexicologique. Publications du Centre d'Étude du Vocabulaire Français, Besançon, I—III. 1960., IV—V. 1961; VI. 1962., VII. 1964.
- BSI:** Beiträge zur Sprachkunde und Informationsverarbeitung. München, Heft 1—2. 1963., 3—4. 1964., 5—6. 1965.
- CL:** Cahiers de Lexicologie. Publié par B. Quemada 1. Besançon 1959., 2. Paris 1960., 3. 1961., 4., 5. 1964., 6., 7. 1965.
- FLP** Forschung. Lehre. Praxis. Schriften der Gewerkschaft Wissenschaft. Heft 4., Berlin, 1962.

KELEMEN JÓZSEF

¹ A munkaértekezlet óta a debreceni csoport — Papp Ferenc irányításával — elvégezte az Értelmező Szótár önálló szócikkben kidolgozott címszavainak lyukkártya-rendszerű elemzését a következő szempontok szerint: összetétel, homonimák, szófajjelölés, jelentések száma, stílusminősítés, szótípus, ragok, eredet, képző, szóhosszúság n-ekben. A szójegyzék mind balról jobbra, mind jobbról balra haladó betűrendben készült a fenti adatokkal s többféle teljes- és rész-szójegyzék készül belőle.

PAPP FERENC több cikkben számolt be a végzett munkáról: A magyar szókinés lyukkártya-rendszerű feldolgozása. (Nyr. LXXXVIII [1964], 457—64.) — A Magyar Szóvégmutató Szótár és a vele kapcsolatos munkálatok. (MNY. LXI [1965], 187—200.) — Le vocabulaire du Hongrois Contemporain sur cartes perforées. (CL VII [1965], 103—17.)

Magam előkészítettem az értekezlet óta egy irodalmi mű (Madách: Az Ember Tragédiája) lyukkártya-rendszerű elemzésének tervét a Berlieni Német Tudományos Akadémia Gépi Információfeldolgozó Bizottságával együtt végzendő munka számára.

J. ŠTINDLOVÁNAK, a prágai Csehszlovák Tudományos Akadémia Nyelvtudományi Intézete munkatársának előadása (Sur les problèmes de la mécanisation et l'automatisation de quelques recherches linguistiques) és J. PANOVOVÁ előadása (Исследование перфокарточных машин при многостороннем анализе текстов) a Computational Linguistics legközelebbi, V. kötetében jelenik meg franciául, ill. oroszul.

TARTALOM

1. A lyukkártya vázlatos jellemzése, használatosabb típusainak rövid bemutatása	343
2. A Hollerith-rendszerű lyukkártyák ismertetése a nálunk használt változatra való tekintettel	344
3. A szöveg átírásának főbb kérdései a Hollerith-rendszerű lyukkártyáknak nálunk használatos betűrendszere alapján.....	344
4. A lyukkártyáknak szövegből történő szó- és morfémagyűjtésre való alkalmazása; a gyűjtés és feldolgozás általános szempontjait tartalmazó lyukkártyaterv ismertetése	348
5. A szintaktikai célú szöveggyűjtés kérdése; szintaktikai célú lyukkártyaterv	364
6. Hangtani vizsgálatok a szövegnek lyukkártyarendszerű elemzésével..	364
7. A helyesírás és a helyesírástörténet kérdéseinek vizsgálata lyukkártyákkal	365
8. Szótárak elemzése lyukkártyákkal.....	366
9. Nyelvtanok elemzése lyukkártyák segítségével.....	367
10. Jelentéstani kutatások lyukkártyák segítségével.....	367
11. Nyelvtudományi részlettanulmányok eredményeinek nyilvántartása lyukkártyák segítségével	368
12. Stilisztikai kutatások lyukkártyákkal.....	368
13. Verstani kutatások lyukkártyák alkalmazásával.....	368
14. Szövegváltozatok elemzése; a szövegkritika és a lyukkártyák.....	369
15. Lyukkártyák alkalmazása ismeretlen jelrendszerek megfejtésében...	369
16. A lyukkártyák felhasználása a bibliográfiában.....	370
17. Várható eredmények: anyagban, módszerben, elvekben, művekben..	370
18. Súlyponti kérdések, különös tekintettel a munka kezdeti szakaszára..	372
19. A lyukkártyákkal történő adatfeldolgozás előnyei.....	373
20. A lyukkártyák alkalmazásának korlátai.....	376
21. Előkészületek.....	377
22. Kitekintés.....	378