

**LINGUISTICA**

SERIES C

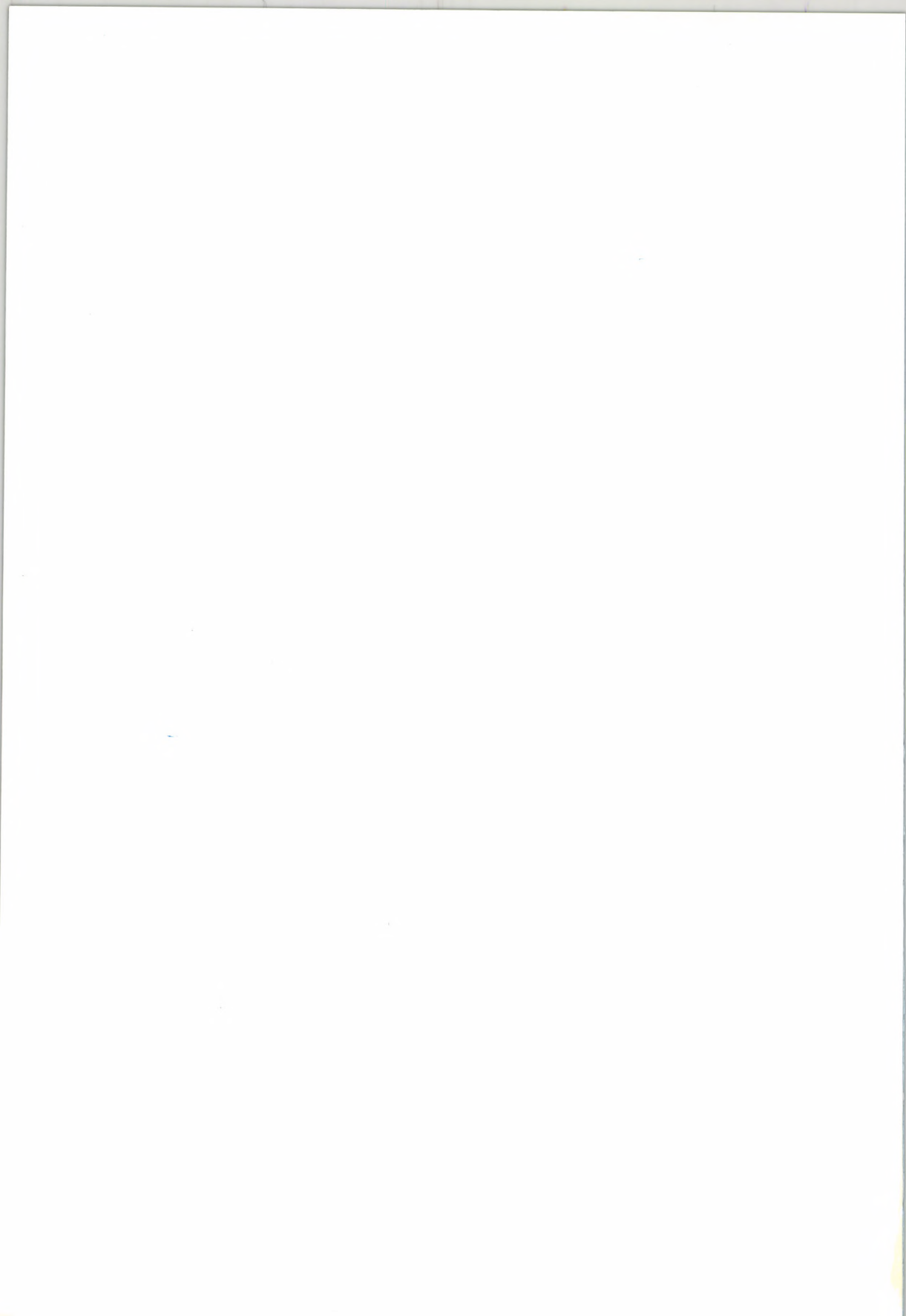
**RELATIONES 11.**

**PAPP FERENC**

**A DEBRECENI THÉSZAURUSZ**

A MAGYAR TUDOMÁNYOS AKADÉMIA NYELVTUDOMÁNYI INTÉZETE  
INSTITUTUM LINGUISTICUM ACADEMIAE SCIENTIARUM HUNGARICAE

2000

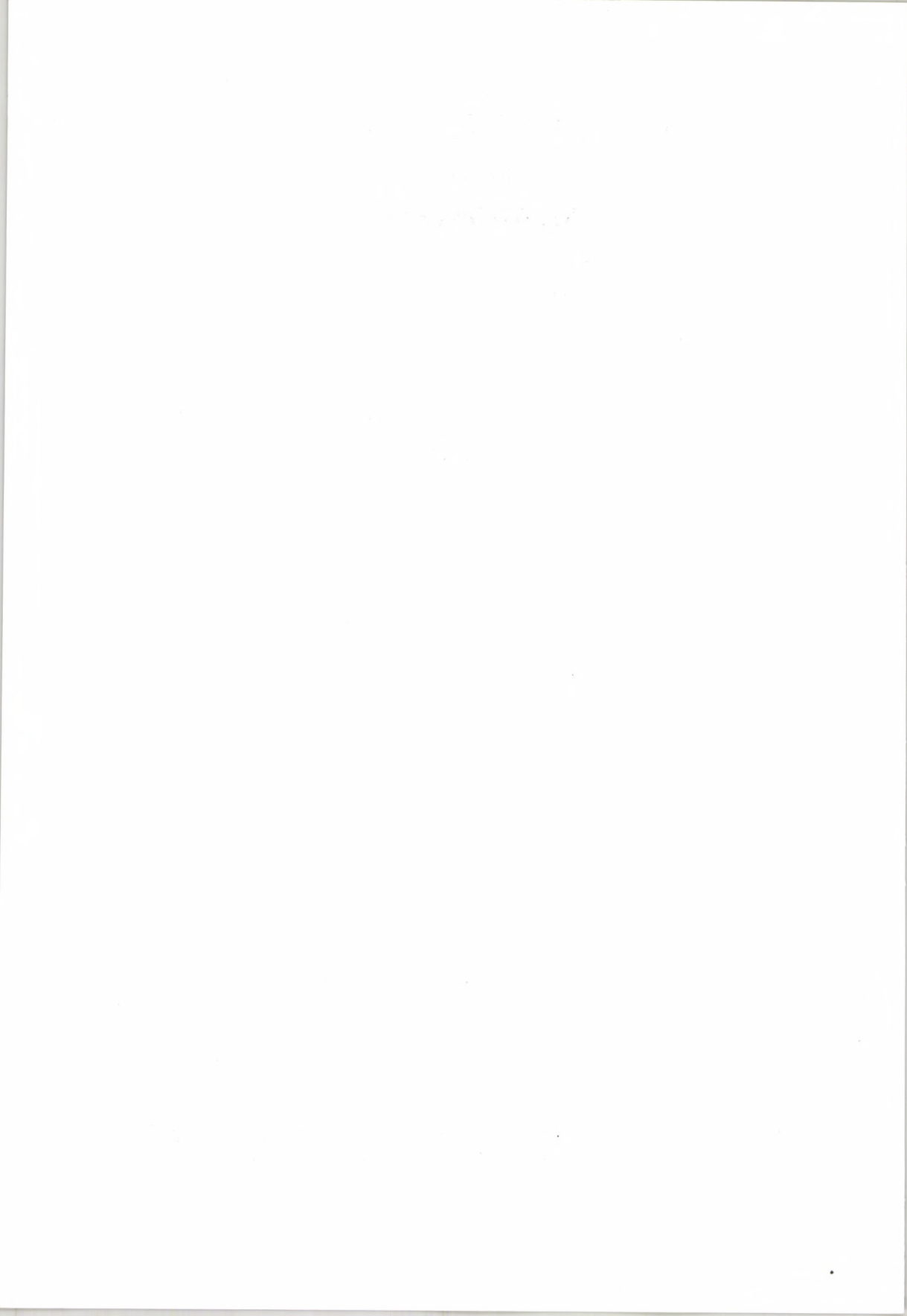


LINGUISTICA  
SERIES C  
RELATIONES 11.

---

---

**A DEBRECENI THÉSZÁURUSZ**



**LINGUISTICA**

SERIES C

RELATIONES 11.

**PAPP FERENC**

**A DEBRECENI THÉSZAURUSZ**

A MAGYAR TUDOMÁNYOS AKADÉMIA NYELVTUDOMÁNYI INTÉZ  
INSTITUTUM LINGUISTICUM ACADEMIAE SCIENTIARUM HUNGAR

2000

ISBN 963 9074 18 7  
ISSN 0866-4196

© 2000. Az MTA Nyelvtudományi Intézete

Felelős kiadó: Kiefer Ferenc

Budapest, 2000.

## Története

0. Ha a témát első közelítésben úgy fogalmazzuk meg, mint a matematikai és számítógépes nyelvészet megindulását hazánkban, egyrészt e korábbi terminusokat a mai alakra kell hoznunk, másrészt rá kell mutatnunk töobgyökerűségekre, a külföldi és a hazai előzményekre, amelyek e történetbe tehát tulajdonképpen nem tartoznak bele. Hanem részint segítenek a hazai események nemzetközi elhelyezéséhez; részint mint itthoni előtörténet lehetnek érdekesekek.

Boot–Weaver 1946-os, Weaver 1949-es megnyilatkozásait követően 1952-ben tartják az első GF-kongresszust az Egyesült Államokban. Az 1954-es georgetowni kísérlet: az első a GF történetében, oroszról angolra. Más oldalról, de természetesen alapvető fontosságú a modern irányzatok számára Chomsky könyvének 1957-es megjelenése. – A külföldi eredmények közül azért utalok az akkori szovjet eseményekre, mert azok, mint látni fogjuk, döntő jelentőségűek a magyar fejlődésre. Sztalin 1950-es, a hagyományos lingviztikát jogaiba visszaállító, valószínűleg Csikobava által írt újságcikkeit követően, 1952-ben igen nagy példányszámmal (15.000) beindul a *Voproszy jazykoznanija* (VJaz.) c. folyóirat (a példányszám rohamosan csökken: 1962-re az eredeti egyharmada – a fordulat évéig a Language mellett ez a legelterjedtebb külföldi folyóirat hazánkban, persze a legolcsóbb is), majd Sztalin 1953-as halálát követően a korábbinál némileg jobb helyzetbe kerül a szovjet nyelvtudomány: nyitás észlelhető a „világ” felé. A VJaz. hasábjain már 1955 végén megjelenik az ifjú Berkovnak és társának rövid ismertetése az 1954-es amerikai GF-kísérletről, utalással arra, hogy hasonló munkálatok a SZU-ban is folynak, amiről hamarosan számot fog adni az orgánium. Ez utóbbi megjegyzés nélkül valószínűleg meg sem jelenhetett volna Berkovék néhány lapos közleménye az amerikai eredményről, de ezen kívül még igaz is volt. Melcsuk, a hruscsovi korszak feltűnésében, majd kiüzettségében oly jellegzetes alakja, Kulaginával (az örök paszportábilis pártonkívülivel) együtt még épp megtörülte a szemüvegét, hogy a számítógép mellett a francia–orosz GF-hez fogjanak; épp a franciához – és nem az angolhoz – nem azért, amiért a Hold túlsó oldalát elsőként lefényképezték, vagyis minden gyakorlati, hadászati, gazdasági stb. cél nélkül, merő költészetből: hanem mert véletlenül mind a két megnevezett személy franciául tudott akkor talán a legjobban. Chomsky 1957-es művére a VJaz. hasábjain 1959-ben jelenik meg Paduceva recenziója, melynek egy első változata még az előző év folyamán – tehát néhány hónappal a recenze-

ált mű megjelenése után! – szűkebb körű, de kvázi „szakmaibb” nyilvánosságot kap az Idegen Nyelvi Főiskola GF-egyesülésének sokszorosított bülléten-jében.

A hazai előtörténet Tarnóczyra (1942) és Vértes Editre (1952–1954) megy vissza. Alacsonyabb szintű számlálás gyorsírói s hasonló praktikus céllal már a század elejétől folyt.

1. Antal László 1957. június végén vagy július elején Zebegényben áthozta e sorok írójához a VJaz.-nak tanszékükre frissen megérkezett számát, benne egy rövid híradással a Moszkvában és Leningrádban megindult szellemi mozgolódásról „bizonyos matematikai kutatási módszerek alkalmazásáról a nyelvészetben”<sup>1</sup> tudva, hogy e sorok írója az ősszel a Szovjetunióba készül több hónapos tanulmányútra; buzdítva őt, hogy nézzen szét: mi is történik e téren a két fővárosban. Valószínűleg ezek voltak Magyarországon az első lépések e mondott, alább pontosítandó irányban, kivált, ha hozzátesszük a számítógépek korai alkalmazását is. Mintha Dömölki Bálint és Fónagy Iván számítógépes vizsgálatai egyes magyar költők hangtanát illetően csak 1959-ben vagy 1960-ban indultak volna meg Szegeden, Kalmár László védőszárnyai alatt, alkalmasint bizonyos kijevei (Kaluzsnyin prof.) gondolatokat is átplántálva ezzel.<sup>2</sup>

Antal Lászlónak ekkortájt jelennek meg első cikkei, vagy inkább ismertetései egyes amerikai nyelvészek gondolatairól. Ma, egy jó emberöltővel a pont elején leírtak után, nem sokkal AL oly korai halála után, tudjuk, eltértek a nézeteink

---

<sup>1</sup> На филологическом факультете МГУ с сентября 1956 г. работает семинар посвященный некоторым применениям математических методов исследования в лингвистике. С докладами и сообщениями на заседаниях семинара выступили лингвисты П. С. Кузнецов, И. И. Ревзин [irva Резвин], С. К. Шаумян, В. В. Иванов, Т. Н. Молошная, И. А. Мельчук и математики Р. Л. Добрушин, В. А. Успенский, О. С. Кулагина. Объединение по проблемам машинного перевода работает с декабря 1956 г. в Московском педагогическом ин-те иностр. языков. Вышли из печати № 1, 2, 3 стеклографированного издания „Бюллетеня” Объединения.

В декабре 1956 г. начал также работать межфакультетский семинар по машинному переводу в Ленинградском ун-те. – ВЯ. № 3, стр. 167 (1957)

<sup>2</sup> Közelebb is találhattunk volna ősforrást. Íme:

Az 1950-es évek legelején felkeresett bennünket a Lenin Intézet Orosz nyelvi tanszékén a szellőzőművek egy dolgozója. Minden újra nyitott tanszékvezetőm (én akkor ott tanársegéd vagy aspiráns lehettem), Satyernyikova professzornő (az érdemes klasszika-filológus, Nyikolaj Satyernyikov leánya) elküldött velem: járjak utána a dolognak. „E dolog” pedig nem volt más, mint az akkor a Kossuth tér és a Széchenyi-rakpart sarkán álló IBM Kft. lyukkártyagép-parkjának megtekintése. A szellőzőművi dolgozó ötlete egyszerű volt: a nyelvben igen sok szó van, rögzítsük őket egyenként kártyákra s azokat kombináljuk-rendezgessük. És én visszamentem professzoromhoz, a jámbor orosz nyelvtörténészhez azzal, hogy ez egy újabb laikus számárság. – Azóta nem láttam ezt az embert. Talán utolsó reménye voltunk. Aztán szétszedték munkatársai a Szellőzőművekben, mint javíthatatlan örültet. S hogy miért pont hozzánk jött, a Lenin Intézetbe?



Antal Lászlóval azt illetően, hogy mi ösztönözte őt nyugaton-keleten új ideák keresésére – és mi e sorok íróját. A DT szempontjából úgyis csupán az a lényeges, hogy az 1. lábjegyzetben idézett s AL által áthozott rövid VJaz-közlemény alapján e sorok írója hamarosan Kuznyecov professzor szellemi karjaiban; a Mohováján elég késő esti órában megtartott szemináriumokon, az Idegen Nyelvi Főiskolán Rozencvejg (Viktor/Mordko Juljevics) és Revzin (Iszaak Ioszifovics) körében találta magát. És ezzel megpecsételődött az ő szellemi sorsa, a hazai, számítógéppel támogatott kutatások egy részéé is, ezen belül a DT-é. Antal László útmutatása nélkül fiatal kutatóként valószínűleg elvesztem volna Moszkva szellemi fergetegében, nem is fogtam volna fel, hogy ez valóban fergeteg. Így volt egy biztos, a fentiekben körvonalazott, összehasonlítási pontom.

Az ezt követő évtized, a „hatvanas évek” világszerte a nyelvtudomány „boom”-jának évtizede, amit leginkább a Chomsky-forradalom évtizedeként szoktak feltüntetni. Valóban, az amerikai nyelvtudományi társaság tagjainak létszáma ebben az évtizedben több mint kétszeresére nő (Newmeyer 1986, 44); Franciaországban többszörösére emelkedik a nyelvészeti tanszékekkel rendelkező egyetemek, a nyelvész-professzorok és -oktatók száma (Chevalier 1984, passim), és így tovább. Valószínűleg pusztán véletlen (?), esetleg országonként is változó konkrét politikai-gazdasági okokkal, hogy a mi „féltekénk” csatlakozik e trendhez: a Szovjetunióban is szaporodnak a matematikus-nyelvész tanszékek és szakok – durván szólva Hruscsov 1964. október 14-i leváltásáig (a valóságban természetesen tovább: a fejlődést nem lehetett megállítani). Nálunk 1961 nyár elején lezajlik a nevezetes strukturalista vita, 1962 márc. 8–10 között az a matematikai-nyelvészeti munkaértekezlet (szintén a Roosevelttéren), mely elhatározza, hogy a debreceni KLTE-n működő matematikai-nyelvészeti munkacsoport egyik fő feladata az MTA Számítástechnikai Intézetével karöltve elkészíteni a magyar nyelv „a tergo” szótárát (I. I.OK XIX. 337). 1963-ban indult el az Ált.Nyelv.Tan.sorozat, egyes kötetei általában Telegdi Zsigmond (a II. kötet TZs. és Kalmár László) szerkesztésében (vö. 3.), 1964. október 15–16. között Budapesten az MTA Nyelvtudományi Intézete és a KLTE említett munkacsoportja rendezésében folyik egy értekezlet „A lyukkártyarendszerű adatgyűjtés és adatfeldolgozás kérdései a nyelvtudományban” elnevezéssel, ahol Kelemen József a fenti címmel tartott előadását, e sorok írójának a magyar szókinccs hasonló gépi feldolgozásáról tartott beszámolóját vitatták meg, nemzetközi részvétellel (J.Stindlová, J.Panevová), Kalmár László részvételével. Az értekezleten jelen voltak azok az egyetemi hallgatók, akik a DT kódolásában részt kívántak venni, s akik ma a VégSz. címlapjának versóján szerepelnek, természetesen tanáraikkal, Jakab Lászlóval és Jánoska Sándorral együtt. A kódolás – az ÉrtSz. adta információk sokszorosított adatlapokra történő rögzítése – az értekezlet után azonnal megindult s néhány hónapra rá be is fejeződött az alább részletesebben érintendő ellenőrző műveletekkel egyetemben. (Ezt megelőzően még egy próba készült az egész anyag 2,5 %-án, nem szabályos mintavétellel, I. Nyr. 88. 4. 457–64 [1964].) Az első, a teljes anyagra vonatkozó publikáció 1966. nyarán jelent meg ugyanott (8 –

„idő előtt”: Szende Alinál ezt a kéziratot kicseréltem egy előbb leadott anyaggal, csak hogy ez minél előbb napvilágot láthasson). A gépi feldolgozás a KSH Buday László utcai központjában folyt. Maga az egész DT-munkálat zöme tehát 1964. őszétől 1969. végéig tartott. Ami e „zömön” kívül maradt, az néhány tucat rendezés, mely főleg a főnév problematikáját (= paradigmarendszerét) járta körül, a debreceni SZÜV telepén készült, a KLTE finanszírozásában. (A VégSz. első kiadását angolul Z.B.-nek és E.W.-nak ajánljuk. Oldjuk fel itt e rejtélyt. **Botka Zoltán** a KSH-ban Budapesten, **Woynarovich Elek** a KLTE tudományos rektorhelyettesi pozíciójában tette lehetővé, hogy vagy térítésmentesen dolgozzunk, vagy kapjunk pénzt az egyetemtől erre a munkálatra. Főleg ez utóbbi történt: Woynarovichnak direkt jól jött, hogy a KLTE tudományos pénzalapjából valamit a bölcskarnak is nyújthatott.) Az „utótörténetre”, az egész kártyaanyag egy eredetileg Hell György–Meinhardt von der Ohe kérésére készült másolatának a Műegyetemtől (= az ottani egyetemi számítóközpont nyitott szekrényéből) a SZTAKI-ba kerüléséről l. Kornai 1986. Pótolhatatlan vesztesége az egész DT-munkálatnak, hogy az anyag fokozatosan csökkent: már a nyomtatott VégSz. nem tartalmazta a SzófSz.-információkat, a stílusra és a hosszúságra vonatkozókat (nem fért bele a nyomdai tükörbe), majd hasonló okokból „lekoptak” a homonima-indexek. Így az alább említendő, ezzel kapcsolatos, rendezések csak az MTA Nyelvtudományi Intézetének raktárában található fel, illetőleg itt számolunk be róluk.

A DT anyagából itthon egy nagydoktori (ebből könyv is) és egy kandidátusi, az akkori NDK-ban egy A-doktori disszertáció, több tucat cikk itthon és külföldön, számos kongresszusi előadás készült magyarul, németül, oroszul, franciául, angolul, ukránul.

## Néhány elvi kérdés

Ezen néhány belső, a tárgyat érintő kérdést (2.), továbbá néhány külső, a tárgyat ideológiailag-politikailag megítélő kérdést (3.) értünk.

2. Fentebb, tanulmányunk legelső mondatában, már adtuk két közelítő megnevezését tárgyunknak (matematikai, számítógépes nyelvészet), pontosabban: az akkori kornak megfelelő megnevezését. Mára tisztult-változott a fogalmi helyzet: a DT a számítógép t á m o g a t t a kutatások sorába tartozik. A számítógépes nyelvészet természetes nyelvek használatáról szól számítógépes rendszerekben (Prószéky 1989 alcíme a külső címlapon), a "computational linguistics" fogalmába, az ilyen nevű konferenciák témakörébe nem tartozik bele a nyelvi anyagra (linguistic material) vonatkozó számítástechnikai alkalmazás; ide nyelvészeti kutatások tartoznak és nem a nyelvi anyag (közvetlen) vizsgálata (Karlgrén 1989, 32 – a *linguistic* terminus alapvető kétértelmősége miatt minden ilyen fogalmazás angol nyelven kétértelmű marad). E két hivatkozás nem véletlenül 1989-es: majdnem csak napjainkra (az 1987-es budapesti kongresszusra) alakult az ki világosan és határozottan; még Karlgrén is azért fogalmaz ilyen élesen, hogy a számítógép *támogatta* kutatásokat ismertető előadásokkal kíméljék meg őt mint a programbizottság elnökét az 1990-es helsinki COLING-kongresszusra. De ezt, úgy látszik, 1989-ben még szükségesnek találta leírni – ő, a KVAL-csoport alapítója, tehát a *kvantitatív lingvisztika* egyik (évtizedekkel ezelőtti) apostola.<sup>3</sup> Ennyi nekünk a DT szempontjából elég: ehhez képest külön problémát jelent a matematikai nyelvészetnek (alkalmasint mint a matematika részének) a meghatározása. A gépi fordítás (GF), mint ami olykor elhomályosította ezt a egész matematikai-számítógépes komplexumot és e szavak szinonimájaként is fellépett, most például, Prószéky említett művében a számítógépes nyelvészet egy fejezete.

Egy további, ennél szűkebb elnevezési-fogalmi kérdés magának a munkálatnak (ma talán „projectumnak”) az elnevezése és tartalma. Az említett határozatban ez áll: „...fordított betűrendű (»a tergo«) szótárának elkészítése” (I. OK XIX. 337). Egy (Kiefer Ferencsel közösen) 1963. dec. 20-án aláírt sokszorosított körlevelünk legelső kérdése a nyelvész-kollégákhoz: „Mi legyen a szótár neve?”. Név nélkül szerepel még a fentebb említett, 1964. októberi konferencia téziseiben is. Végül a MNy. hasábjain az 1965. évfolyam 2. számának cikkcímében éppen úgy, mint a szövegben Kiss Lajos leleményeként megjelenik a „szóvégmutato szótár”, megadva annak egyszersmind „VégSz.” rövidítését is. E jó név feltűntével azonban immár egy fogalmi probléma merült fel észrevétlenül: a „VégSz.”

<sup>3</sup> Valamelyik kongresszuson (talán a bonnin) egy fiatal kolléga kérdésére: mi az a négy betű – KVAL – az ő postai címében, Karlgrén azt felelte: értelmetlen betűrövidítés. Péter megtagadta Krisztust (vagy egyszerűen csak sietett?)

elnevezés rátelepedett az egész elektromechanikus feldolgozásra. Ezen tulajdonképpen csak a von der Ohe által a hetvenes évektől alkalmazott DT elnevezés segített (Ohénél eredetileg természetesen – bár ez nem is olyan természetes: Ohe, akinek a feleséges is magyar, kiválóan ismeri nyelvünket – németül, de ez csak helyesírásában-képzőjében különbözik a magyar elnevezéstől).

Az itt említett „körlevél” és „kérdés a kollégákhoz” némileg jellemző ezekre az időkre. Kevésbé utaztunk (talán Kiefert kivéve), de legalább leveleztünk, kérdeztünk itthon és külföldön: látjuk, hogy a két fő fogalom – a DT és annak legnépszerűbb rendezése, a VégSz. – nem tőlünk, törvényes szüleitől kapta a nevét. Ez a közös munka teljesen elengedhetetlennek látszott, amikor egy ennyire új, nagy dologba kezdtünk. A két fő elnevezésen kívül kaptunk természetesen további jó ötleteket: így Gáldi László felhívta a figyelmünket arra, hogy a feldolgozott anyag nem maga a magyar szókincs, ahogy eleinte neveztük, hanem annak lexikográfiai törzsanyaga; Kalmár László s mások agyában született meg a gondolat, hogyan lehet lyukkártya-gépeken standard operációk segítségével atergósítani (nem egyszerű: tessék végiggondolni!); és így tovább. Döntő jelentőségű volt a Varga Dénestől az 1964-es októberi konferencián kapott ötlet: kódoljuk a kártyákon külön oszlopban a képző meglétét-hiányát. Ez szükségtelen: az *à tergo* sorrend úgyis „kihozza” a képzőket. Emlékezetem szerint Pesten végig ezt hajtogattam, s csak a buszon, Debrecenbe haladtunkban jöttem rá arra, hogy ha ez az oszlop nincs, nem tudjuk kiválogatni a tőszavakat! Ez végzetes hiba lett volna. Másrészt, amikor ez az oszlop már megvolt (ti. utólag rávéstük az adatlapok már kész stenciljére), „enni kért”: így itt jelölhettük a korábban is ismert, de külön nem kezelt „képzőszerű nem-képzőket” (*újra, hajrá, hozzá* stb.); a gépektől visszszakapva a legelső teljes a tergo listát Jakab Lászlóval közösen fel kellett fedeznünk a „gyakori idegen szóvéget” (*linóleum, dokumentáció, miniszter* stb.), azt jelölhettük ebben az oszlopban 8-as kóddal. És így tovább: számos hasznos (Varga Dénes esetében: létfontosságúnak bizonyult) ötlet, tanács.

De, visszatekintve az emberöltővel ez előttire, azt kell mondanom, még több volt a rossz, egyenesen káros ötlet, amelyeket viszont – szerencsére – nem fogadtam meg. Inkább csak az irányát, mint a szerzőit jelzem az efféle tanácsoknak. Ilyen volt a nagyban, gépen végzett munka lényegének meg nem értése, – NB.: én sem értettem, nem érthettem, csak érezhettem vagy sejthettem ekkor még! – a túlságos aprólékos igények érvényesítésének felvetése, mintha a számítógép atombombájával zsilletpenge finomságú műveleteket kívántak volna végrehajtani. Vagy: a hagyomány túlságos tisztelete, követni kívánása. A túlzott igények hangoztatása. Ezekkel a tekintélyes tanácsokkal állandóan küzdenem kellett, a sok „nem”, „nem”, „nem” miatt vettem el majdnem Varga Dénes javaslatát is, mint feleslegeset és túlzott igényűt. Másrészt bennem is megvolt ez az apróságokra, mindenféle igényekre való hajlás, ezért az első rendezések nagyon is részletező volta; idegesítően sok helyen kellett bennük utánanézni valamely dolognak. A hagyomány igényeinek érvényesítése egy helyen lett volna végzetes: az elektromechanikus gépeken a magyar betűrend konzekvens megkövetelése.

Ezt (ti. ennek lehetetlenségét) feltehetően elhallgattam az 1964. októberi értekezlet (nyelvész-) részvevői előtt (vagy még magam sem tudtam?).

Egy jelentős hiányosságra kell még rámutatnom a jövő számára nyújtható tanulságként is (NB: az új nagyszótár szinte „magától” e hiányosságokat maximálisan elkerülendően indult). Ez: a kódolt anyagban meglévő igen nagy mennyiségű hiba. Mi az ÉrtSz. minden egyes szavát egymástól függetlenül kétszer kódoltuk; a két független listát összeolvastuk, a hibásnak ítélt locusokat belső műhelyvita eredményeképpen azonnal javítottuk. Nem hogy nem külön „telephelyeken” végeztük a kódolást, hanem együtt, egy vidéki kis város özszezárságában, ahol nap mint nap öszzetalálkoztunk. Minden héten legalább egyszér hivatalosan („tanrend szerint”) is öszszeültünk a kódolás, majd a két adatlap-csomag öszzeolvásása során, hogy a felmerült problémákat megbeszéljük, maximális egységgel járjunk el, és így tovább. Mégis sok hiba maradt az adatlapokon, és a külön készített kontroll-tablók részint elégtelenek voltak, részint csak a más jellegű, teljesen mechanikus hibákat szűrték ki. (Réműlten gondolok a GyakSz.-ra, ahol a fenti feltételek egyike sem volt meg. Ott mi lehet(ett)? Ráadásul a mi kódolási utasításunk mindöszze egy lap – két oldal – könnyen áttekinthető rendszere volt!). Felmerült annak idején még egy további kontroll-operáció: a teljes betűrendes lista újbóli öszzeolvásása közvetlenül az ÉrtSz.-szel, ettől azonban igen hamar elálltunk egy fentebb már érintett ok miatt: a hagyományos magyar betűrend, amelyben természetesen az ÉrtSz. címszavai is állnak, állandó apróbb-nagyobb eltéréseket hozott létre az ÉrtSz.-beli és a gépi sorrend között. Mi el akartuk végezni ezt a projektumot belátható időn belül, több ilyen akadályon szűkségszerűen átugrottunk. Ma már látjuk, hogy ezt az akadályt alkalmasint nem lett volna szabad átugornunk, bár ki tudja.

Eddig csupán egyetlen vonatkozásban nem találtunk úgyszólván egyetlen hibát sem: ez épp a (fordított) betűrend. Úgy látszik, érdemes egy mégoly felemás rendszert létrehozunk (ilyen a VégSz.-ben alkalmazott kétféle betűrend vö. VégSz., 10), a DT-be beadnunk, ha az géphez simuló (gépbarát): ott az eredmény kiváló lesz. A mai gépek persze, a megfelelő szoftverrel, eleve sokkal emberhez simulóbbak, mint amilyenek a múlt századi elektromechanikus gépek voltak, ma már tehát joggal olyan, szinte hibátlan eredményt várhatunk el minden területen, mint amilyenre itt a fordított (pontosabban: „VégSz.-fordított”) betűrend sikerült.<sup>4</sup>

---

<sup>4</sup> Az egyetlen eddig megtalált sorrendi hiba:

<i>írva:</i> koriander	<i>helyesen:</i> leander
szalamander	koriander
leander	szalamander

4. A mai (kivált fiatal) olvasó számára rejtélynek tűnhet fel, mi lehetett itt az ideológiai kifogás, hogyan merülhetett fel még 1959-ben is egyáltalán a gondolata annak, hogy efféle szovjet (!) eredmények egyszerű ismertetése miatt el lehet/kell valakit távolítani az egyetemről.<sup>5</sup> Könnyebb az itt felkiáltójellel kiemelt szóval kezdeni. Arra természetesen hamar rá lehetett jönni, amit először egy ma Franciaországban élő magyar barátom mondott ki: nekünk a szovjet zászló azért kellett, hogy saját hasonló törekvéseinkkel bátrabban előállhassunk.

De miért kellett „bátorság” és „előállás”? A modern nyelvészeti irányzatoknak (akárcsak a modern művészetieknek, tudományosoknak, mint pl. a kibernetika) sokáig bujdokolniuk, üldöztetniük kellett egyszerűen azért, mert modernek. Ennek mélyebb, elvi gyökereire, ha s ahol erre szükség lesz, nyilván rámutat majd a tudomány- és művészetszociológia, itt csak a vizsgált szűk szaktudományos területen mutatjuk meg röviden, mi is volt „a baj” vele. (Az alábbi, 5. látjegyzet ennek a „filozófiai” helyzetnek kvázi a vulgáris, napi-politikai felszíni

---

<sup>5</sup> Természetesen máig is szinte szóról szóra emlékszem a beszélgetésre a R(ektor) és A(djunktusa) között, egy néma tanú (a tanszék akkori adminisztrátora, Krecsmáriné) előtt, 1959. szeptember végén – októberében.

R: Szóval a NyK-beli kéziratodat idegen nyelven is meg kívánod jelentetni az ALH hátsábjain?

A: Igen.

R: Te ezzel egy – a Szovjetunióban – folyó vitához kívánsz hozzászólni.

A: Ó, hát ez tulajdonképpen csak egy ismertetés-fűzér (ti. az új irányzatok képviselőinek munkáit, a Vjaz-beli vitát ismerteti).

R: Igen, de Te mondani akarsz valamit ezekkel az ismertetésekkel.

A: Igen, anélkül meg sem írtam volna az egészet.

R: De a vita még nem dőlt el a Szovjetunióban. S így végezetül kiderülhet, hogy Te helytelen álláspontot képviseltél. És ily módon ártottál a szovjet nyelvtudománynak.

A: Na ne mondd, hogy én tudnék ártani...

R: Egy szó mint száz: megjelenhet idegen nyelven ez a cikked, de csak ha az egyetem falain kívül vagy.

Én (A) nem akartam az egyetem falain kívül kerülni. Ezért a) táviratoztam az ALH szerkesztőjének, tán Fónagy Ivánnak, hogy visszakérem a német (angol?) nyelvű kéziratot; b) a napi postával Pestről épp abban az órában érkezett (!) NyK-korrektúrámat átnéztem s abban a „szovjet strukturalisták”, „szovjet strukturalizmus” kifejezéseket a hozzá n-számban még legközelebb álló „matematikus nyelvészek”, „m. nyelvészet”-re javítottam. Ez e magyar terminus igaz, bár kissé talán keserű, története. (NB: tessék megnézni, ebben a cikkemben a fenti kifejezés valóban ilyen, csak a szovjetekre vonatkozó, szűkített szemantikájú.) Az akkor fejemben felmerült, de (főleg az n-szám miatt?) elvetett terminusok közül emlékszem még a *neolingvisták*-ra. Valamint arra, hogy a klinika szemészetén, ahová aznap bevonultam, kegyetlenül rossz világítás volt (minek a rossz szeműeknek egyáltalán villany?), nevezett tanulmányom pedig petittel jött („a nyelvtudomány története” – ez a rovat mindig, mindenütt petittel jön; így jött persze legutóbb a VJaz-ban is szösszenetem Lomonoszov magyar vonatkozásairól). Így hát volt elég bajom vele a klinika szemészetén. (Ott és akkor kezdődött sclerosis multiplex-em, szerencsére csak vagy három évtized múlva tudtuk meg.)

képét nyújtja.) Ami strukturalista, az idealista nyelvészetben, irodalomban stb. Mert szakít a történetiséggel, szinkrón struktúrákat vizsgál. Mert a formát (amit egyedül kutat) elszakítja a tartalomtól, figyelmen kívül hagyja a jelentést. Marx és Engels becületes XIX. századi tudósok voltak, számukra minden jelenséget annak története magyarázott meg; az őket kommentálók számára az sem volt kétséges, hogy a formát nem szabad elszakítani a tartalomtól, mert az formalizmus. Ha ezek az ötletek nyugatról jöttek – Bloomfield, Chomsky (aki még hozzá zsidó is!) –, az még rosszabb. Ezt az égtáj szerinti igazság-kiosztást Saussure sem úszhatta meg egészen, bár ő Baudouin (ez meg emigráns lett!) és Scserba révén nagyon beette magát a szovjet nyelvtudományba. A formalizmus egyik csúcsa volt a matematizálás: „más tudományok módszereinek alkalmazása a nyelvtudományban” (ez a kifejezés a hetvenes években gyakran előfordul a VJaz. és egyéb hasonló orgánumok hasábjain – nem dicsérő kicsengéssel). A matematizálás egyik csúcsa a gépesítés: ami gépesítve van, az gépies, mechanikus, ezért idealista. Ezekkel a gondolatokkal számolt le a hazai nyelvészet az 1961. nyár eleji strukturalista vitán, Antal László szikrázó programbeszédével: „Milyen legyen az eljövendő magyar strukturális nyelvtan? Én két alapvető követelményt támasztanék vele szemben. Az egyik az, hogy következetesen szinkronikus síkon mozogjon, és ne tegyen egy jöttányi engedményt sem a maga helyén oly értékes történetiségnek. A másik, hogy a formális elemzésen alapuljon, és ne tegyen egy jöttányit engedményt sem a jelentésnek” (I. OK.18.54 [1961] – Ált. Ny. Tan. II. 138. – Antal mindig előre leírta és „betanulta” előadásait, felszólalásait, ezért azok az élő beszéd benyomását keltették.). Égbekiáltó szavak voltak ezek akkor mind a hivatalos marxizmus, mind a hazánkban oly mindenerős hagyományos nyelvészet szempontjából. És Antal hevessége mellett ott volt Telegdi Zsigmond óriási műveltsége, megfontoltsága; ott voltak az ő kiváló kapcsolatai a mindenkori politikai vezetéssel; ott voltak Szépe György tájakat összekötő gondolatai, e sorok írójának szervezőkészsége és aprómunka szeretete. Így tisztogathatta az utat a ma felé az 1961-es értekezlet. Így ülhetett össze az 1962-es „matematikai nyelvészeti” konferencia, ahol immár Kalmár László zsenialitása is méltó helyet kapott. Így született a többször idézett határozat az a tergo szótárról – így vált hivatalosan is tiszta az út a DT számára.

## A DT mai meghatározása

5. Mi is hát a DT? (Szebben szólva: a DT ontológiai státusa.) Ezt a kérdést azért érdemes felvetnünk, mert mindjárt látni fogjuk: eleinte mi magunk sem tudtuk világosan, mit is csinálunk. Ennek megfelelően olykor, reményeinkben, túlbecsültük, olykor meg alábecsültük az általunk végzetek jelentőségét.<sup>6</sup> Szinte még a szavak, a fogalmi keret sem volt meg annak idején a következő bekezdés elején mondandókhöz.

Eszerint a DT az első kísérlet a magyar szókincs lexikográfiai törzsanyagának (az ÉrtSz. némileg – a Szófsz-szal – kiegészített és explicitté tett) adatbankszerű feldolgozására. (NB.: mikor jelent meg az *adatbank* szó és fogalom a magyarban és általában?) Készült Debrecenben a hetvenes évek első felében; legfontosabb eredményeinek első publikálásai ettől kezdődően mintegy egy évtizedig tartottak. Nem a mai értelemben vett számítógépes munka, hanem a számítógéppel alátámasztott hagyományos nyelvészeti kutatások egy korai megnyilvánulása.

A DT eszerint a modern számítógépek nyelvészeti alkalmazásának egy igen korai megnyilvánulása. Talán ezért is, talán másért is (amiből nem zárjuk ki e sorok írójának fentebb egyszer már említett tompaságát), még a VégSz. sajtó alárendezésekor egy súlyos hiba (? – mondjuk így, semlegesebben: sajátosság) született nyomdai megfontolásokból. Nevezetesen: mivel ezt a normális nyomdai tükör nem engedte meg (és mivel az Akadémiai Kiadó egy azóta elhunyt, ezért róla nulla fiat mentio, bűn rossz szerkesztője jutott osztályrészemül), az eredeti DT-hez képest a VégSz.-ből elhagytuk a hosszúságra, a jelentésszámra, az etimológiára és a stílusminősítésre vonatkozó információkat; ezeknek a státusa különféle okokból úgysis vitatott volt. A rendezések ezen oszlopok és kombinációik szerint a Buday László utcában, ill. a debreceni SZÜV-nél azért persze elkészültek, korabeli publikációinkban megjelentek. Ezek után a DT-vel a következő, jó és rossz – ennek megítélését az olvasóra bizzuk – dolgok történtek: a) megjelent a VégSz., az előbb említett oszlopok és a rajtuk elhelyezett információk nélkül; b) Hell György – Meinhardt von der Ohe kérésére ebből (tehát a VégSz.-ből és

---

<sup>6</sup> Egy jellegzetes „alábecsülés” tanúi lehettünk a VégSz. megjelenése előtt, a Nyelvtudományi Főbizottság egy ülésén, ahová véletlenül (a Nyr. ún. „felülbírálat” kapcsán) becsöppentem. A megjelenendő VégSz.-re szánt ívszámot, azzal a hivatkozással, hogy a VégSz. úgysis a Moutonnal közösen fog megjelenni s így ívszáma nem nálunk számos, Kelemen Jóska könyvére javasoltam áttenni. Benkő Lóránd, aki persze már akkor is, ott is elnök volt, és aki előbb konzultált e furcsa, akkor még sehol sem látott műről Kiss Lajossal, zárszavában ezt a javaslatot lesöpörte az asztalról. (A magyar VégSz. tán a második e nemben. Halljatok csudát: az első a román volt.) Bár ki tudja, nem lett volna-e jobb, ha a VégSz. a Moutonnal közösen jelenik meg? Én csak Jóskán akartam segíteni, bár megvolt róla a szakmai véleményem.



nem a kártyákról az azokon érintetlen információval!) egy másolat készült, ez fölkerült Hell Györgyhöz a Műegyetem számológéppontjába, ahol azután az ismert módon megtalálta őket, taxiba rakta és tovább szállította Kornai – ezt nevezzük a továbbiakban BUT-nak; c) a teljes cédulaanyagból készült tételes és gyűjtött tablók bekötve, szépen sorban ott sorakoztak debreceni egyetemi tanári szobámban (NB: ezt hívtuk orosz nyelvi tanszéknek – volt ott sok minden más is, a japán ábécétől egyes magyar írók-költők számítógépes feldolgozásáig, itt volt a főhadiszállása a hetvenes évek elején a debreceni szemiotikai körnek stb.)<sup>7</sup>; d) ez az anyag Pestre helyezésemet követően rövid ELTE általános nyelvészeti tanszéki kitérő után felkerült az MTA Nyelvtudományi Intézetének lomtárába – előbb tehát a Szentháromság-, majd a Színház utcába. Ezt nevezzük tehát BUT-nak; d) az 1986–89-es években, amikor e változások-vándoroltatások nagy része végbement, összesen 22-szer feküdtem különféle kórházakban. A BUT (sajátos szófaji kódokkal, erről most ne essék említés) került 1991 őszén, Kiss Gábor jóvoltából személyi számítógépek memóriájába.

Alább, ahol ez szükséges, utalni fogunk azokra a rendezésekre, melyek csak a DAT-ban vannak meg. (A BUT-rendezések megvannak minden azt óhajtó személyi számítógépén.)

A fentiekből vonjunk le még egy tanulságot. Kezdetben nem tudtuk, hogy a hard copy semmit sem ér (azon kívül, hogy kézbe vehetjük és olvashatjuk – persze ez se semmi, de:), a korszerű adathordozón, számítógép memóriájában őrzött anyag a minden. Csak két példát erre. Az ÉrtSz. címszavainak jelentésszáma<sup>8</sup> és ezek kombinációja különféle egyéb kategóriákkal csak a DET-ben van meg, ezért ma már szinte használhatatlanok, bár természetesen részint olvashatók az alapjukon készült cikkek a korabeli folyóiratok, kongresszusi előadások-tézisek hasáb-

---

<sup>7</sup> Valamikor 1974-ben vagy az előtt előadásom volt a Nyelvtudományi Társaságban Ady Endre költői műveiről (azok fonémastatisztikájáról). Mint mindig, kevesen voltak. A kevésből is kivált egy anyóka, az megvárt végig, s amikor már egyedül maradtunk, még egyszer meggyőződve arról, hogy a debreceni egyetem orosz nyelvészeti tanszékének vezetőjével áll szemben, inkább talán szempillantással, mint szóval buzdított: amíg ilyen orosz tanszékvezetőink lesznek addig nem veszünk el. Akkor értettem csak meg, mit akart mondani. – Nem mindenki gondolkozott így. Már nem tudom, vagy nem akarom tudni ki, erősen hibáztatott azért, hogy nem vagyok igazi orosz nyelvi tanszékvezető, mert lám, a nagydoktorimat nem is orosz témából írtam (az a MFön volt). – Még egy harmadik oldal, már napjainkban, 1995-ben, díszdoktorim átvételekor, hallottam, amint az egyik, újságíróként ott álldogáló egyetemi hallgató kb. ezt mondja: Hát igen, mindennel (még ezzel is – ti. orosz nyelvvel is) sikert lehet elérni, csak makacsul kell csinálni.

<sup>8</sup> Országh László, persze a maga finom-úri modorában, ellenezte a jelentések felvételét a kártyatervbe. Aztán ő örült a legjobban, mikor ezek az érdekes eredmények kijöttek. – A KLTE elméleti-fizikai tanszékének kalkulátornője, aki itt számunkra a szükséges számításokat végezte az akkori hatalmas számológépeken, sirva jött hozzám: nem tudja a fenti görbe elméleti értékeit és az egyes mérési eredményeket külön-külön ábrázolni: összefolynak. (Ha e könnyek nem is voltak őszinték, jól mutatják: a kalkulátornő tudta, mit óhajta a főnök...)

jain, részint megtekinthetők a megfelelő tablók a Nyelvtudományi Intézet padlásán. (És aki még más akar, ugyanebből az anyagból?) A másik példa a szótárak, melyek elektronikus verzióit kivált Pajzs Júlia ismeri és propagálja. Hard copy: egy „közönséges” szótár. Hét vagy tizenkét vagy akárhány kötetes, szolid, mozgathatatlan. Ugyanakkor az elektronikus verziók akármilyen formátumban, ábrákkal vagy anélkül, egészben vagy részben megjelentethetők, azokból kinyerhetők az idézett szerzők, az idézetek, a nyelvtani információk stb. bármely kombinációban.

## Eredmények

6. Az alább felsorolt eredmények érintenek a betűkészlettől fogva egyes felsőbb szinteket is. Tekintettel arra, hogy teljesen a hagyományos keretek között, a hagyományos fogalomrendszerben maradunk, látszólag igen könnyen, egy az egyben alkalmazhatók is – lennének, lettek volna a hazai nyelvtudományban. Ugyanakkor teljesen beváltak Tompa Józsefnek a 25-ben idézett keserű szavai. Az égvilágon semmi sem szüremlett át eredményeinkből a hazai hagyományos magyar nyelvészeti oktatásba. Például alább látni fogjuk, hogy az eddig ismert egy-, két-, háromszófajiságú elemek mellett az ÉrtSz. nyilvántart néhány négy- és öt- és hatszófajiságút is. A legcsekélyebb jelét sem láttam eddig annak, hogy ezt bárki észrevette, egyetemi-főiskolai előadásaiban felhasználta volna (kivéve talán a Grétsy-tanszéken Tóth Etelkát). Vagy: eredményeink koherens feldolgozása eddig leginkább a magyar főnévre vonatkoztak. Ezen belül egy egész, elég terjedelmes fejezet a hagyományos kérdések egyik leghagyományosabbjára – a hangrendre és az illeszkedésre. Sőt kiderült, ez a fejezet a legbővebb anyagra (ti. az ÉrtSz.-ére) támaszkodó tanulmány a magyar nyelvtudomány egész eddigi történetében a kérdésről. A Nyelvművelő Kézikönyv 1985-ben, tehát pontosan tíz évvel a MFőn. megjelenése után napvilágot látott II. kötete a magánhangzó-illeszkedés címszó alatt még mint irodalmat sem említi meg ezt a forrást...

Vajon mi lehet az oka ennek a teljes visszhangtalanságnak, ezen belül annak, hogy még inkább a „modernek” (Kornai, Prószéky...) (meg a nyelvtörténészek) „haraptak rá” erre az anyagra, pedig ők más rendszerben dolgoznak? El lehet képzelni, hogy teljesen európai nyelvészetünk egy ilyen fontos(nak tartott) ága (ti. a nyelvművelés) többé vagy kevésbé hátsó információ (bibliográfia) nélkül dolgozik? (El.) Kell jönnie egy új Lőrinczének, aki immár nem a nagyközönség, hanem a nyelvművelők felé fordul, azokat magasabb elméleti színvonalra emeli, például a nyelvművelés dokumentációjának gépesítésével. Ha kell. – Ez a helyzet ösztönzött arra, hogy ne az eddig elkészült DAT–BUT rendezéseket soroljam fel:

ma már mindenki rugalmasabb, érdeklődésének megfelelő rendezéseket állíthat elő – Kornai említett igyekezetének köszönhetően otthon, saját PC-jén, ha beszerezte modern adathordozón a BUT (az ÉrtSz./DT) anyagát.

7. Magukat a címszavakat (= önálló szócikkkel rendelkező címszó) két kérdés érinti:

a) Hány címszó van az ÉrtSz.-ban? Az ÉrtSz. szerkesztői szerint 58023 (ÉrtSz. VII. 671), szerintünk 58323, tehát pont háromszázzal több. (Ez a „pont” is némileg valószínűsíti, hogy emberi-számlálási hibával van dolgunk: a gép számára a „kerek szám” semmivel sincs kitüntetve.) Amint korábban is írtuk e két számról s a köztük lévő eltérésről (8), az eltérést egy fentebb már említett ok miatt nem tudtuk pontosan (milyen kezdőbetű alatt stb.) kimutatni: az általunk alkalmazott és a hagyományos, az ÉrtSz.-ben is érvényesített betűrend eltérése miatt. Ehhez még annyit tegyünk itt hozzá, hogy a BUT valóban a fenti mennyiségű címszót tartalmazza (ti. 58.323-at). Ám a kártyák romantikus utóélete, a nyitott szekrény, taxin való utaztatás stb. (vö. Kornai i.h.) elvileg elképzelhetővé teszi, hogy a jelenleg hozzáférhető kártyamennyiség ettől eltérő (nyilván: kevesebb – a valóságban néhányval több van, egyelőre nem tudjuk, miért. Ez lesz a BUT, a DT budapesti verziója.)

Magán a DET-en belül két nem túlságosan jelentős különbség is van: (i) A VégSz. sajtó alá készítésekor annak kéziratán (a [DET] 2a jelzésű outputon) attól eltérő változtatásokat eszközöltem, emlékezetem szerint elég nagy számban. A grammatikai kódokkal ellátott AT-lista ugyanis egyes következetlenségeket szinte „magától” kihozott, ez a nyelvtani (morfológiai) szótárak, így az AT-szótárak sajátja a magyarhoz hasonló („nem bantu típusú”) nyelvekben. (ii) Az DET/AT és az ABC rendezések között némi eltérés van (épp az előzők miatt):

b) Mely karakterekkel vannak rögzítve a címszók? A hetvenes évek derekán (nem egyszer azóta is) felháborodva írtam nyelvész, nem nyelvész (informatikai) orgánumban: az, amit nagy (vagy kisebb) szótáraink elejükön-végükön közölnek, mint *teljes* magyar ábécét – nem teljes még az izolált magyar címszavak tekintetében sem. Évtizedek elmúltak – újabb szótáraink ugyan jelentek meg, de továbbra is illegális, tehát fel nem sorolt, betűkkel. Íme, az illegális jelek, melyek címszók rögzítésekor előfordulnak magyar szótárakban, lexikonokban: à (à 5 Ft, à la carte stb.), ë (zárt e), - kötőjel), – (nagykötőjel: *marxista-leninista, osztrák-magyar, marxizmus-leninizmus, marxi-lenini*; több címszó e jellel az ÉrtSz.-ben nincs is), „, „ (szóköz: *illegeti-billegeti magát* címszó az ÉrtSz.-ben: előbb egy kiskötőjel, majd egy szóköz áll benne: *billegeti magát*). A helyesírási-nyelv-művelő szakirodalomban rendszerint felsorolt *q* [kú], *w* [dupla vé], vagy a számítástechnikai-matematikai gyakorlatban talán inkább [vevé], *x* [iksz], *y* neve [ipszilon], *ejtve* [i] (az ÉrtSz. egyetlen címszavában fordul elő: *brandy*) (vö. pl.

HSzab.<sup>9</sup>, 11) szintén fontosak, már csak sajátos fonematikus megfelelésük miatt is. De nézetünk szerint legalább ilyen fontosak a korábban felsorolt „illegális jelek” is. Azt nem engedhetjük meg magunknak, hogy felsoroljuk a „teljes magyar ábécét” és az ott felsorolt 40 elemen kívül további ötlet (12,5 %) fogunk még élni csak az izolált címszavak rögzítése során is. Szövegek rögzítésekor beugranak még további jelek: a pont, a vessző, a pont-pont-pont („...” – ez különösen kellemetlen, mert először úgy néz ki, mintha pont lenne, utána két üres mondat, melyeket szintén pont zár), a kevés szóalakban (opcionálisan? – még ez sem biztos: az efféle lazaságok ábécé esetében különösen megengedhetetlenek, egyébként erről a ritka írásjelről még sehol sem olvastam, csak tudom, hogy van) előforduló tapadó félidézőjel: *tudj’ isten, majd’ annyi, hál’ isten, ’60-as évek stb.*, de ez más kérdés: jó lenne, ha csupán egy szótár, enciklopédia, lexikon stb. *címszavai*-nak a betűkkel való rögzítését rendbe tudnánk tenni. Eddig ez sem sikerült. (Tehát e rész címe emiatt is, több alább felsorolandó jelenség miatt is ez lehetne: „Eredmények” és eredmények.)

8. A címszavak hossza szerinti eloszlást a VégSz. közli (588–9). Elektromechanikus gépeken a hagyományosnak megfelelő, egy- és többjegyű betűket kezelni nehéz lett volna, ezért az adatok *n*-ekben (karakterekben) vannak adva. (Későbbi, szövegfeldolgozásaink már elektronikus gépeken folytak, így ott legalább a kétjegyű betűket figyelembe tudtuk venni; ugyanígy lehetséges ez a nagyszótár gépi feldolgozása számára.) Manuálisan bejelöltük az *egész/ség, arc/szín, kuka/csin* vagy *kukac/sín* stb.-féle, kétértelmű stringek kötelező szegmentálását. Ezek szerint a hossz karakterekben durván szólva normális eloszlást mutat (ez nem triviális: magyar *szövegekben* például a szóhossz kétmaximumos, alkalmasint a határozott névelő gyakori előfordulása miatt) 8–9 *n* körüli maximummal, egyenként 14–15 %. A leghosszabb szó *n*-ekben mérve a *keresztényszocializmus* (22 *n*), hagyományos magyar betűkben/fonémákban a *lelkiismeretfurdalás* és a *rabszolga-kereskedelem* (20–20 fonéma). A leghosszabb teljesen motiválatlan töszó a *csimborasszó* (12 *n*, 9 betű). (Az utóbbi „csúcs”-adatok már nem a VégSz. idézett helyéről, hanem a 8:182-ről merítettek.) Készültek annak idején (a 70-es évek elején: ezek a klasszikus DET-tablók) rendezések az *n*-ekben mért szóhossz és a jelentésszám kombinálásával. Eredményül természetesen azt kaptuk, hogy a rö-

---

<sup>9</sup> Iszonyú gépi és emberi erőfeszítéseinkből, ha jól látom, csak egyetlen egyáltalán nem is jellemző apróság került be a helyesírási elméletbe (a gyakorlat eddig is ez volt): „8. A többjegyű betűknek csak az első jegyét írjuk nagybetűvel: Csoma, Dzsingisz...” (Hszab<sup>11</sup> 11). Ez sem reguláris, tudományos úton, hanem, emlékezetem szerint, egy véletlen találkozás során Fábíán Pállal. Akkor ő olyasmit mondott: Persze, csak a gép olyan marha, hogy a szabályzat alapján így írja: SZegeD, CSongráD stb. Én akkor erre az egyszerűség kedvéért (a lift előtt álltunk) rábólintottam. Ám látnunk kell, hogy a következetességre nem a gépnek van szüksége (az csak rá-rá koppint a körmünkre, ha nem vagyunk azok), hanem az elmélet tisztaságának. Az ilyesmit a tegnap élt kollégák közül talán csak Telegdi Zsigmond érezte ilyen élesen, a ma élők közül – Zaliznyák akadémikus érzi...

videbb szavak tömegükben (statisztikailag) kevesebb jelentésűek a szótárban is. (A ritka s mégis rövid *ón* és ehhez hasonlók természetesen semmit sem bizonyítanak, a statisztikába ezek is szépen belesimulnak.)

9. Homonímia. Erről igen röviden ugyan, de a (30)-ban adtunk összefoglaló jelentést: „A címszót követi a szótárban a homonima-index. Kiderült, hogy teljes anyagunk mintegy 1650 homonimát és álhomonimát tartalmazott. Ezek nagyobb része természetesen homonima-pár, ám itt láttuk összegyűjtve a hármas, négyes, ötös csoportokat is (ez utóbbiak: *a, c, költés, az, század, tud*), sőt volt egy hatos homonimabokor is (*haj: haj<sup>1</sup>* fn ‘szörszerű képződmény’, *haj<sup>2</sup>* fn ‘héj’, *haj<sup>3</sup>* fn ‘padlás’, *haj<sup>4</sup>* isz, *haj<sup>5</sup>* msz ‘terelésre használt szó’, *haj<sup>6</sup>* msz ‘hé!’). Együtt szemlélhettük immár e homonimákat ábécérendben és a tergo sorrendben, szófajok szerint csoportosítva és külön a tőszavak közül kiválogatva (arányuk ott lényegesen nagyobb az egész szókincshez viszonyítva) stb. Láttuk, hogy általában rövidebbek, mint a nem-homonimák; láttuk egyes további közös vonásaikat is” (30: 148).

10. Hány többől áll a címszó? Ezt a kérdést aligha értelmes önmagában feltennünk. Így, bár a kombinált kérdések értékelésére később fogunk kitérni, erre a szófajonként (a szófaj-oszloppal kombinálva) válaszolunk. Eszerint az igék 40 %-a egytövű, 56 %-a igekötős összetétel. (Mindkét csoportban az egy képzővel ellátottak vannak döntő többségben, 37, ill. 36 %-kal. A főnevek 34 %-a egytövű, 55 %-a áll két többől [33:25.]

ige	1	2	4	6	7	9	össz-k
0	399	37	0	1.787	3	1	
1	5.265	146	1	5.187	21	217	
2	0	2	0	0	0	1	
3	0	42	0	0	1	0	
9	102	72	0	968	1	0	
össz-ö							

1. ábra

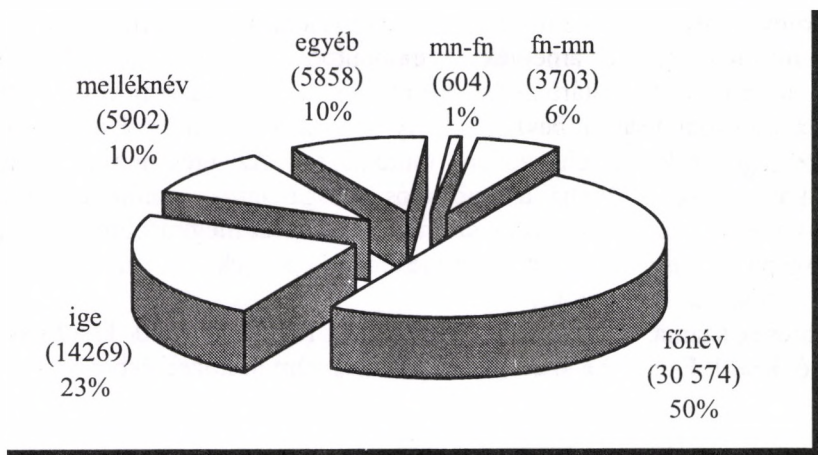
11. Szófaj. A magyar szókincsre, így természetesen az azt tükröző szótárra is jellemző a bizonyos mértékű többszófajúság. Így az 58.328 címszónak 91 %-a egyszófajú, 8 %-a két, 0,2 %-a három, 0,02 %-a négy szófaj valamelyikéhez tarthat, a szövegtől függően (ez utóbbi elemek: határozószó-igekötő-mondatszó-főnév kettő: *újra, vissza*; egy-egy: számnév-melléknév-főnév-névmás: *egy*<sup>1</sup>, melléknév-főnév-határozószó-mondatszó: *jelen*, határozószó-főnév-melléknév-mondatszó: *kontra*, határozószó-igekötő-mondatszó-főnév: *teli*, határozószó-igekötő-névutó-melléknév: *szerte*, határozószó-igekötő-névutó-főnév: *túl*, határozószó-igekötő-kötőszó-mondatszó: *viszont*, vö. 8:185) s maradt 57 címszó (0,1 %) melyet az ÉrtSz. munkatársai szófajilag nem minősítettek (14:161). „Bizonyos mértékű” többszófajságról beszéltem, mert egyes nyelvek szókincsében alkalmassint több (így az angoléban, kivált a gyakori ige-főnév konverzió miatt), másokéban (így az oroszéban) talán kevesebb lenne a többszófajiságú elem), arról nem is beszélve, hogy a szófajhoz tartozás megítélése terén valószínűleg a szótárszerkesztők-szerzők szubjektív megítélése sem elhanyagolható.

A szófaj-kategória rögzítésére három oszlopot tartottunk fenn a kártyatervben. Nem mintha 999 maximális szófaj-számot feltételeztünk volna (ennyi fér el a 3 oszlopon), hanem szókincsünk említett, „bizonyos mértékű” konverzió-hajlamára tekintettel. Ha ugyanis valami ingadozott két érték között, akkor megállapodás szerinti mindkettőt rögzítettük (tőszavak esetében elvileg abban a sorrendben, ahogy az ÉrtSz. közli, egyébként saját nyelvérzékünk szerint). A konverziót is sajátos ingadozásnak számítottuk: ahol valamely lexéma két vagy három szófajiságú volt, ott mind a két vagy három szófajiságot jelöltni kívántuk. Hogy a legtömegesebb kategóriákból hozunk példát: ha a főnév kódja 2, akkor a főnév-melléknév 23, a melléknév-főnév 32, és így tovább. Ha valami négy szófajiságú volt, vagy két ritkább, tehát egyenként két-két oszlopot lefoglaló kóddal rögzített szófajhoz tartozott, akkor mind a három oszlopot „lekilenceztük” a fentebb mondottak értelmében: az *újra* lexéma kódja ebben a három oszlopban 999, mert négy szófajiságú; de 999 pl. az indulatszó (önmagában 010) és a mondatzó (önmagában 011) kettős szófajiságú lexéma, és így tovább. (A teljesség kedvéért megjegyezzük, hogy ha valamely ritkább, tehát önmagában kétjegyű kóddal rögzített szófaj került hármasszófajiságba, az is „lekilenceződött”: az ő két számjegye, plusz a másik két szófajiság egy-egy számjegye már nem fért bele a három oszlopban. A BUT-ban másképpen.)

„Úgy találtuk, összesen 34 szófaji osztály állítható fel úgy, hogy egy-egy osztályba csak azonos szófaj-kombinációkban és azonos fontossági sorrenddel fellépő szavak legyenek. Például: a főnév egy osztály, a melléknév egy másik osztály, a főnév-melléknév egy harmadik, a melléknév-főnév egy negyedik és így tovább” (8:185). Az egész címszóanyagban több, mint a fele (30 574) főnév, negyede ige (14269), melléknév 5902, főnév-melléknév 3703, melléknév-főnév 604 és í.t. (vö. uo.). (Itt jegyzem meg, bár másutt is tehettem volna, hogy ezek és más számok csupán tájékoztató jellegűek: lehet kettővel több vagy tízzel keve-

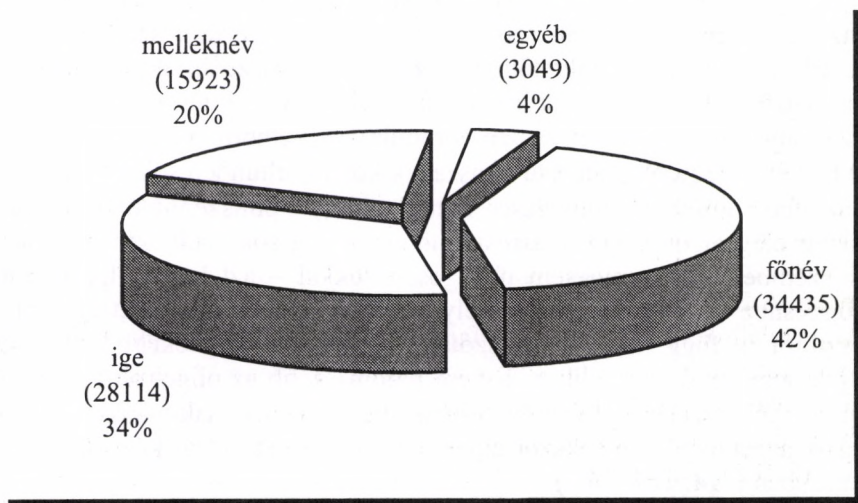
sebb stb. Inkább a grafikonokat tessék elgondolni, a főnevek, igék, fn-mn-ek stb. arányát aszerint elképzelni.) Ugyanitt összehasonlítjuk ezeket az arányokat egy nagy orosz szótár (az Usakov) megfelelő adataival.

### ÉrtSz. (60 000 rekord)



2. ábra

### Usakov (80 000 rekord)



3. ábra

A legfontosabb, amit nem innen tudunk, de több kutató hajlandó állandóan megfeledezni róla: a szótárban uralkodó arányok teljesen másak, mint az élő szövegekben uralkodók. A szótárban például néhány tucat névmás a szövegek jelentős részét teszi ki, ugyanígy egyes segédszók (kötőszók, prepozíciók, posztpozíciók stb.). Ez elég általános lehet. Ugyanígy általános, csak mintegy műfaj-specifikus a szótárban a főnevek viszonylag magas aránya: a szótár, ha nem is lexikon, de elég nagy arányban tartalmazza az előbbieket. A főnév-melléknév és a melléknév-főnév nem elhanyagolható mennyisége a magyarra lehet jellemző (nyilván más nyelvekre is, amelyek e szempontból hasonlítanak rá). Az igék magas aránya az Usakov-szótárban (35 %) részben vagy egészben következhet abból a lexikográfiai hagyományból, amely szerint az egymástól aspektusukban különböző igéket külön címszóként tüntetik fel. Az igék magas aránya az Usakov-következik. Verbális asszociációs vizsgálataink mintha azt mutatták volna, hogy az orosz anyanyelvűek számára, a külföldi megközelítéssel szemben, ezek az igepárok összetartoznak – „egy lexémát képeznek” –, akár a стол–столы, a синий–синяя stb. alakpárok.)

Jelentések száma. A (28) és a VégSz. egyik függeléke (590–1) ad összképet erről a kérdésről. Eszerint a szavankénti jelentésszám jó közelítésben az

$$y = \frac{V}{2^x}$$

összefüggéssel írható le, ahol  $y$  az  $x$  jelentéssel regisztrált szavak száma a  $V$  egyedet tartalmazó szótárban. (Maga az összefüggés a (28)-ban.) A legtöbb jelentéssel regisztrált szavak az ÉrtSz.-ben: 101 – *is*, 65 – *van*, 55 – *úgy*, *hogy*<sup>(2)</sup>, 48 – *csak*, 47 – *jár*, és így tovább. Szófajjal kombinálva e szempontot azt találtuk, hogy az ige a legpoliszémebb szófaj – legalábbis az ÉrtSz. szerkesztőgárdája szerint. Itt ugyanis erősen felvetődött a szerkesztők-összeállítók szubjektív hozzáállása: közülük ki mit tartott külön jelentésnek? egy jelentésen belül csak betűvel jelzett aljelentésnek? vagy éppen homonimának-álhomonimának? (A VégSz. idézett helyén, a legtöbb jelentésű címszavak között feltűnően sok a homonima és kivált az álhomonima. Ez tömegesen mutatja, hogy a poliszémia és a homonímia közti határ elég bizonytalan; másrészt, hogy itt elég sok szubjektívizmussal állhatunk szemben. Hogy mégsem túlságosan sokkal – azt bizonyítja a fentebbi összefüggéshez való jó közelítés. Mely utóbbi ugyan szintén esetleg szótártól-szerkesztőtől is függ: az Usakov-szótárban sokkal „meredekebb” ez a görbe (mondjuk nem kettő az  $x$ -ediken, hanem három), több az oligoszém elem. Vagy ez is pl. az igék kezelésével van összefüggésben: a szinte csak utalószóként fellépő „másik aspektusú” ige sokszor egyjelentésű.) (A két szótár közti különbséget részletesebben l. 14: 167, 169.)

12. Milyen stílusminősítést kapott a fejben a címszó egésze? A kódok – első pillanatra érthetetlen módon – hármasával ugráltak: *argó* – 3, *bizalmas* – 6, *durva*



– 9 és így tovább. Erre az „ugrálásra” az adott módot, hogy még így, hármásával sem léptük túl a 99-et, két oszlop maximumát; ezt a lehetőséget arra használtuk fel, hogy a hárommal osztható számok akkor szerepeltek, ha a fejben csak egy stílusminősítés van (természetesen ez a többség),  $3n+1$  – ha ezen kívül még egy stílusminősítés is szerepelt az ÉrtSz.-ben,  $3n+2$  – ha ezen kívül még kettő vagy több. Az egyes jelentésekhez, szóalakokhoz stb. fűzött stílusminősítést nem vettük figyelembe: mi más vonatkozásban is, ebben is a címszót mint egészet kívántuk jellemezni – tehát minden információnkat a lexikográfiai fejből vettük.

Példaképpen a VégSz. első kiadásában függelékként öt stílusréteget (a másodikban négyet) magunk is közlünk (24:587–8). Az „új” stílusminősítésű elemekről egy munkatársunk külön is publikált egy tanulmányt a *Nyr.* hasábjain (32). További, stílust (is) érintő cikkek: (17), (22).

Tovább 8 oszlop szófajonként részben vagy egészben különböző alaktani információt rögzít:

**14. Főnév:** tőtípus, egyes acc., többes nom, egyes 3. birtokos. Az egész DT-ből ez van a legrészletesebben kidolgozva: (37) (rövidítve MFőn.). Bár ez a feldolgozás nem folyóirat-cikkbe van elbújtatva, mint a többi, hanem önálló kötetbe került, eredményei Amerikától (Vago) Szlovákián át (Sima) a szerzőével szomszédos szobáig (Elekfi: ő a MNyKK. inkriminált „magánhangzó-illeszkedés” címszavának az írója, vagyis e téma s annak kapcsán az egész MFőn. elsikkasztója – abban természetesen nem csupán a jelzett témáról van szó, ami érdekelhetné a kötet(ek) olvasóit...) ugyanolyan ismeretlenek. A kódolás itt is ugyanolyan elvek alapján történt, mint másutt a DT-ben: egytövű szó – ÉrtSz.-kód, többtövű – kollektív debreceni kód; kettőnél több közötti ingadozás, egyéb rendkívüliség: az adott oszloppárban 99 (pl. *bíró–bírá-* tőtípusa 99, ugyanígy *mű–műv-*; *jáger* egyes 3. birt. személyragos alakja is 99, mert az ÉrtSz. szerint négy alak között ingadozik; *-a* v. *-ja* v. *-e* v. *je*, ugyanítt a *kuffer* is 99, mert *-e* v. *-je* v. *ja* (de a *koffer* már elfért, ebben az oszloppárban 42 = *e* v. *ja*) és így tovább.

A MFőn.-et a maga egészében itt újra leközölni, bár kevesen olvasták, nyilvánvaló értelmetlenség lenne. Ezért mindössze három kérdéskör rövid rekapitulációja következik alább – hogy megmutassa egy a DT-hez hasonló gépi feldolgozás erősségét (és gyengeségeit is – l. mindjárt a 14.1-es alpontban:)

14.1. Az egyes 3. birtokos személyragos alak: *lába* de *combja*, *ablaka* de *barackja*, *filmje* de *türelme*, és így tovább. Ezzel foglalkozik a MFőn. leghosszabb fejezete (III., 109–163). A fenti s ahhoz hasonló további megmagyarázhatatlan anomáliák miatt századok óta (ezeket persze nem a gépek, hanem a külön filozmunka fedte fel, a CorpGram.-tól napjainkig) foglalkoztatta a kérdés a magyar grammatikusokat. Azt ugyanis világosan érezzük mind a 15

millióan, hogy határozottan lehetetlen a *baracka* alak (kevésbé lehetséges az *ablakja*), lehetetlen a *türelmje* (kevésbé lehetséges a *filme*), hogy ingadozunk az *üvege-üvegje*, *szappana-szappanja*, *bőre-bőrje* stb. változatok között – hogy mi bújik e határozott, egységes érzéseink mögött, azt kutatták a grammatikusok Sylvester Jánostól (1539) Hámori Antóniáig (1959) és tovább. (Én csakis így tudom értelmezni azt a bizonyos 15 millió magyart.) Az elektromechanikus gépek számos lehetséges változatban élénk tárták az anyagot: hangrend szerint (mert a századok alatt kiderült, hogy a mély hangrendűek – *ceteris paribus*, de mik itt a „*ceteris*”? – jobban vonzódnak a „*j*”-hez, mint a magasak (*padja-vádbeszéde*), ma már azt is tudjuk, hogy a DT-ben szereplő összes mély főnév, főnév-melléknév, melléknév-főnév 59 %-a volt *j-s* vagy *j-s is* (ingadozó), a magasak 85 %-a *j-nélküli* (is); a főnévként használt melléknevek gyakrabban *j-sülnek* (*nagyja, kékje*) mint a tiszta főnevek: a tipikus (később kiderült: az adott korban produktív) valódi főnévképzők kevésbé *j-vonzók* mint a melléknév-, számnév-stb. képzők: *embersége*, de *jobbja*, *tizede*, (ebadta) *nyugtalanja* és így tovább. Az anyagot tehát hangrend, szófaj és végződés, képző, összetettség stb. szerint, azon belül persze *j-j-nélküliség* szerint kellett csoportosítanunk, hogy az eddig felmerült kérdésekre kimerítő (a DT/ÉrtSz. határai között teljes) választ tudjunk adni.

Ugye, az ilyesfajta rendezési kérdéseket az isten is gépekre teremtette, még-hozzá már elektro m e c h a n i k u s berendezésekre is. Úgye, milyen szép lenne, ha erre a kérdésre valóban a gépek feleltek volna meg elsőül, több százados emberi, grammatikusi fejtörés után. Valójában nem (egészen) így történt: két kiváló emberi elme, a Tompa Józsefé és a Károly Sándoré egészen közel jutottak a megoldáshoz, emberi-grammatikusi érzékükre támaszkodva ki is mondták azt – ránk s a gépekre annyi feladat hárult, hogy amit ők megéreztek, ahhoz a pontos számadatokat is hozzuk. (Vö. pl. fentebb a hangrend és a „*j*” kapcsán mondottakat: ők szegények, emberek, még az összes főnevet, főnév-melléknevet, melléknév-főnevet is csak nagy munkával tudták volna kibányászni az ÉrtSz.-ből, nem-hogy azt a fentieknek megfelelően további csoportokra bontani és megszámlálni! Az szárnyaló fantáziájú ember nem is csoportosításra-cédularakosgatásra – és főleg nem számlálásra – való. A következő bekezdésben mi most mégis a saját korunk s fogalmi kereteink szerint s lényegesen kibővítve fogjuk megadni a magyarázatot: ember számára nem nehéz ebből visszaszámlálni a két kortárs-szerző gondolataihoz, megfogalmazásaihoz.

Szerintünk tehát a *j-re* akkor van szükség, ha a pusztá tövet meg kell különböztetni a possesszív tőtől: *gomb-gombja*, hogy ne higgyük pusztá tőnek: *gomba*. A *j-re* nincs szükség, ha produktív képző, változó alakú képző, változó alakú tő, statisztikailag gyakran előforduló tő stb. van a szóvégen (hiszen ezek világosan mutatják: itt a pusztá szóvég). Viszont, megfordítva: kell a *j*, ha szófajilag vagy végződése szerint nem jellemző főnévről van szó, ha a tő nem változik, stb. Fentebb láttunk már szinkron példákat, álljon itt néhány példa a történetből: *uta j* nélkül, ha a tő elszenvedi az *út-ut-* változást, és *út-útja* e változás nélkül; JókK.

*titkában* (mert még nincs *titok–titk* változás a töben), BécsiK *horgia*, *titkjanac* ugyanezért; a korábbi *magzatja* a nyelvújítás körül *magzata* lesz, mert az *at* képző ekkortájt újból produktívá (tehát érezhetővé, a tövéget markánsan jelölővé) vált. A mély hangrendűek között azért kell több *j*, mint a magasak között, mert az *a* pusztá tövég a honfoglalást követően a latin, és kivált a szláv tömeges kölcsönzések miatt hirtelen megszorodott, veszélybe hozva ezzel a birtokos *a* vég felismerhetőségét. Mindez persze feltételezi, hogy a hallgató csak felébe-harmadába hallgat oda a beszélőre, és/vagy nem szívesen nyúl a háttérmemóriában őrzött teljes szótárhoz (ahol persze megvan a *gomb* és a *gomba* pusztá tő is): szüksége van egy közvetlen csengőre, mely figyelmezteti: itt kezd az analízist, minimális tőtárhasználattal.

Melcsuk annak idején (a hetvenes évek elején?) ezt az egész *j*-elméletet elfogadta, csak az *út–uta*, *türelem–türelm*-féle esetekre mondta: a címzett egyszerre csak az *uta*, *türelme* alakokat hallja, honnan tudja, hogy emögött tőváltozás rejtőzik? Most sem tudok jobb választ adni, mint akkor: a beszélő nem csupán fonémarealizációkat, hanem morfofonéma-realizációkat is appercipiál. A *türelem* második *e* morfofonéma-realizációja eszerint nem azonos az *ember* második *e*-jének realizációjával: az elsőt rajta van annak a bélyege is, hogy adott esetben ki fog esni. A változó tő-képző *j*-tlensége igen kemény szabály, régóta észre is vették a grammatikusok. Lehet, hogy ez a magyarázat nem kielégítő, akkor másikat kell keresni – de a nyelvi tény tény marad.

14.2. H a n g r e n d , i l l e s z k e d é s a főnévként (is) minősített címszavak között. Ennek is egy fejezet van szentelve a MFön-ben (IV: 165–86).

14.2.1. Diakrón szempontból azt kellett látnunk, hogy az idő mentén haladva (pl. az egyes, egymást követő etimológiai rétegek főneveit vizsgálva) egy kezdeti (= finnugor főnevek), viszonylag kiegyenlített, kevésbé a veláris felé húzó (*e* rétegben mai alakját tekintve 55 % a mély illeszkedésű) állapot után, a fiatalabb rétegek fokozatosan mélyebb illeszkedésűek (legmélyebbek a latin-görög eredetűek: 92 % s hogy ez az illeszkedés szerinti mélyülés erősebb a hangrend, tehát a tőszókincs magánhangzó-állománya terén, mint az illeszkedés terén volt (a finnugor elemek *i*, *í* nélkül számított magánhangzó-állományának 55 %-a volt mély, a latin-görögök magánhangzó-állományának 92 %-a mély). Ez a mérési eredmény részben összhangban van azzal a más anyagon végzett megfigyeléssel, hogy szókincsünk fokozatosan velarizálódik; másrészt azzal, hogy ez az állomány önmagánál is nagyobb veláris illeszkedést idéz elő. Ez ismét összhangban van azzal, hogy a magyar szövegek fokozatosan velárisabbá válnak, egy korábbi igen palatális csengést megtartva: azelőtt körülbelül kiegyensúlyozott szókincs mellett a szövegek nagy részét kitevő gyakori elemek palatálisak voltak. (Vö. ezt a közismert ténnyel, hogy új ragjaink, igeekötőink, névutóink – talán egyetlen kivétellel (*után*) – mind palatális főnevekből keletkeztek: *be*, *mögé*; *mellett* stb. – persze, mert ezek voltak a leggyakoribbak!)

14.2.2. M a i s z i n t é z i s terén egy szabályt állítottunk fel arra vonatkozóan, hogy az ÉrtSz. összes főnévként (is) minősített egytövíi közül melyek a mély, melyek a magas illeszkedésűek és melyek ingadoznak a két osztály között. Tekintettel éppen arra, hogy ezt a szabályt a kivétellistájával együtt, mint láttuk, alkalmasint az eddigi legnagyobb anyagon alapulva állíthattuk föl és ennek ellenére elég kompakt – mondjuk, nem néhány nyomtatott oldalt betöltő, azt lehetetlen lenne emberi ésszel „betanulni” (erről más szempontból l. még alább, 14.2.3) – valószínűleg a gyakorlatban, így a hazaiak, de kivált az idegenek magyar nyelvoktatásában is elég használható lenne.

Épp e szabályba ugyanakkor a vártnál több hiba csúszott (melyeket főleg von der Ohe disszertációjából, egy fontos ponton Peter Sherwood észrevételéből ismerjük) (vagy másutt is ennyi a hibák aránya, csak éppen ezzel a fejezettel foglalkoztak többet s többen?). Ezért célszerűbbnek látszott a MFön. 167. lapját újra írni (ezen a lapon koncentrálnának a hibák). Íme, az új változat:

(i) Mély illeszkedésűek

- a) a következő 2. tőtípusbeliek: *more, extempore, dözse, andante, parte, forte, áve, csoze, glukóze* és a következő 12. tőtípusbeliek: *guillotine, empire;*
- b) a betűk nevei közül a *h, k, q, y;*
- c) *fiŋg, fi<sup>1</sup>, tik<sup>1</sup>*, valamint a következő, szótári alakjukban egytagú, hosszú í-t tartalmazók: *híd, víg, íj, díj, szíj sík, csík<sup>1</sup>, csík<sup>2</sup>, gyík, nyíl, ín, kín, csin, síp, ír<sup>2</sup>, pír, sír<sup>2</sup>, zsír, nyír<sup>3</sup>*;
- d) *klapec, héj, derék<sup>(1)</sup>, indíték, bolsevik, cél, dauer.*

(ii) Magas illeszkedésűek

- a) a következő 1. tőtípusbeliek: *ábécé, alibi, baltacim, gentleman, oxigén, sztreptomycin, triumvir, aszkézis, matézis, protézis, indonéz, majonéz, szolfézs;*
- b) az i-(b) alatt fel nem sorolt, mássalhangzót jelölő betűk neve az *ly* kivételével.

(iii) Ingadozik a következő 171 lexéma: *jampec, kupec, szanitéc, patécs, szamojéd, honvéd, koszperd, kombiné, agavé, hieroglif, norvég, cajg (cajg, -ot ja; cejg, -et, -je), hokedli, nokedli, pukedli, stampedli, partedli, puszedli, fuszekli, holmi, pofézni, kolibri, harakiri, spagetti, balek, kopek, maszek, derék<sup>(2)</sup>, szik<sup>(2)</sup>, burlesz, obelisz, kábel, rubel, dzsungél, srapnel, fotel, hotel, motel, koktél, bibliofil, szil<sup>1</sup>, modell, flanel, kartell, pasztell, ly, klistély, totem, bohém, pantomim, lóden, pollen, ciklámen, bitumen, szaracén, hidrogén, nitrogén, ómen, fenomén, poén<sup>1</sup>, poén<sup>2</sup>, szatén, rutén, szlovén, arzén, kodein, protein, vatelin, vazelin, aszpirin, abesszin, konzern, kaláber, kandeláber, puccer, valcer, karcser, pucer, káder, pernahajder, leander, szalamander, hipochonder, sóder, púder, koffer, kuffer, jáger, láger, sláger, culáger, tróger, sóher, stájer, fiáker, okker, gukker, bunker, juncker, konstábler, kókler, roller, indiáner. Korner, partner, tapper, stopper, pancser, fuser, zsalugáter, páter, fráter, kráter, bakter, salabakter, karakter, sakter, malter, kóter, gatter, flaszter, majszter, suszter, haver, pulóver,*

*kalugyer, affér, kupcihér, kortes, spektábilis, parainézés, poézés, analízis, paralízis, katalízis, macesz, pajesz, samesz, notesz, soltész, amarillisz, bronchitisz, kadet<sup>2</sup>, sólet, kadét, ankét, oszét, bakelit, barhent, patent, advent, abszint, szint, makett, zsakett, bankett, parkett, balett, omlett, amulett, rulett, kvartett, klozett, pozitív, objektív, kollektív, kurzív, konzerv, szingaléz, trapéz, burizs.*

(iv) A fennmaradó (tehát az itt (i)–(iii) alatt fel nem sorolt) lexémák illeszkedési osztályát a következő szabály segítségével határozzuk meg:

- a) ha a lexéma utolsó szótagjának magánhangzója *a v. á v. o v. ó v. u v. ú*, akkor a lexéma *m é l y* illeszkedésű;
- b) ha a lexéma utolsó szótagjának magánhangzója *e v. ö v. ő v. ü v. ű v. y*, akkor a lexéma *m a g a s* illeszkedésű;
- c) ha a lexéma utolsó szótagjának magánhangzója *é v. i v. í*, akkor az ez előtti szótag magánhangzóját tekintjük utolsóznak: ha az az a) alatti, akkor a lexéma mély, ha b) alatti, akkor magas, ha c) alatti, az az előtti szótag magánhangzóját tekintjük és folytatjuk a c) pontban leírt ciklust, míg ki nem fogytunk a szótagokból; ha kifogytunk a szótagokból, mielőtt kiugrottunk volna a ciklusból, a lexéma illeszkedése *m a g a s*.

Számos további részletre (és „részletre”: így a zárt *ë-t* külön fonémaként ismerő rendszer, általában: a hangzó és nem a helyesírási rendszer alapján működő algoritmus leírása) a MFőn idézett helyén megtalálható: ott ennek néhányszor tíz lapot szentelhettem, nem egyet. Ezért itt tegyünk csak egyetlen záró megjegyzést, mely a MFőn-ben még nem szerepel(het)(et)t, mert akkor ezt a kérdést nem vitük gépre. (A magyar főnévragozás lényegi része a hatvanas évek derekától gépen volt, a MFőn. 1975-ben jelent meg.) Nevezetesen: az általunk Basic-ben elkészített program a magánhangzókat nem így és nem egészen ebben a sorrendben tartalmazza. Az egyes mély, ill. magas magánhangzók nem „v.”... „v.”-gyal vannak egymástól elválasztva, hanem egyetlen „szóvá” (a magánhangzók között space nélkül) tömörítve. A program azután e szóból szeleteli le egyenként a magánhangzó-jeleket és hasonlítja össze előbb a mély, majd a „közönséges” magas, végül a labiális magas magánhangzó-jelekkel, hogy mély, „egyszerű” magas vagy labiális magas magánhangzóról van-e szó, azt a ciklusszámláló mutatja.

14.2.3. Mi mekkora problémát okozott az elmúlt 1100 év alatt a főnevek téren nyelvünk rendszerének? Erre a kérdésre a mi eszközeinkkel a következő módon és a következőket válaszolhatjuk. (NB: „a mi eszközeinkkel” éppen efféle kérdésekre lehet csak válaszolni, ahol szavak ezrei vagy tízezrei kerülnek együttes szemlélésre; ahol – diakrón kérdés esetében – évszázadok jönnek számításba.) Az ÉrtSz.-ben 35172 (ne vegyük *egészen* komolyan ezt a számot sem!) főnévként (is) fellépő címszót találtunk (megjegyzendő, mielőtt elektronikus gépeken, a megfelelő javítások végrehajtása után, nem ellenőriztük a mennyiségeket, tízezer nagyságrendű eltérések előfordulhatnak; következtetéseinket ezek távolról sem érintik). Ebből 4 %, tehát mintegy másfélezer címszó volt egy vagy több helyen is ingadozó (példákat l. alább). Ha a továbbiakban ezt a másfélezer

mennyiséget vesszük 100-nak, akkor az ingadozások a következőképpen oszlanak meg: mintegy 68 % (az egész névszókincs nem egészen 3 %-a) volt az *üvege-üvegje, barátja-barátai*-féle ingadozásnak alávetve, vagy annak is alávetve, 12 % volt az illeszkedésben (is) ingadozó (*jampece-jampeca*). Az összes ingadozók 30 %-a volt a *brilliánst-brilliánsot, akadémikusok-akadémikusak* stb.-féle előhangzóbeli ingadozás. (Ez összesen 110 % az „is”-ek miatt.) Vagyis nyelvünk rendszerének objektíve a j–0 okozta a fő problémát. Ez eredetileg a nagy tömegű szláv, illetve latin származású a végű főnévi szóalak következtében jelentkezik: nem lehet tudni, hogy a *gomba* szóalak pusztán nominativus-e, vagy egyes 3. személyű birtokos: ezért kellett az utóbbi esetben úgy mondani – még a környező szláv nyelveknek is becsületére váló mássalhangzó-torlódással –, hogy *gombja*. (A honfoglalás utáni nagy a végű kölcsönszóáradatról s annak romboló hatásáról alaktani rendszerünkben több alkalommal írtam az elmúlt évtizedek során, vö. leginkább talán 21, 34.)

Másrészt, a mai beszélők szempontjából: mi okozza nekik a legnagyobb problémát az idézett három főnévragozási tényező közül? Erre a kérdésre meg „a mi eszközeinkkel” úgy válaszolhattunk meg, hogy megnéztük: a három közül melyik kategóriában (minek a kódolásakor) tévedtek a legtöbbször munkatársaink. Nos, nagyságrendekkel többet tévedtek a j–0 kérdésében, mint az összes egyéb vonatkozásokban, jóllehet láttuk, milyen többszörös speciális kontrollok voltak beépítve egész munkálatunkba a tévedések (elkódolások) kiszűrésére. (Az ember nem fegyelemre van teremtve. Ez jó vagy rossz – de ezzel számolnunk kell s lehet: az ember fegyelmeztelenségére bizton számíthatunk. Milyen jó.)

14.3. Hiányos paradigmájú elemek. A MFőn-ben ennek is egy fejezetet szentelek (V: 187–212). Ennek a fejezetnek a fő gondolata, hogy a magyarban – az öt környező indoeurópai nyelvektől elérően – nem két (1. hiányos a szám – sing.t. és pl.t. –, 2. hiányos az esetrendszer), hanem három dimenzió mentén kell a hiányosságot elképzelnünk (az előző kettőhöz ui. hozzájárul még a – birtokos személyragozás hiányos volta – l. itt 6. ábra) s hogy ezt a legjobban/legegyszerűbben egy korábbi automata sémánkon tudjuk elhelyezni (Papp F. 1975, 196 l. itt 7. ábra).

Az alábbiakban, a 8. ábrán, az eredeti számozást megtartjuk. Ahol valamely szám kimaradt, annak oka az 1. esetében, hogy ezek a teljes paradigmájú főnevek, csak a számozás teljessége miatt kerültek ide, a többi esetében – hogy éppen ilyen hiányosságot felmutató elem az ÉrtSz.-ben (ti. annak szerzői szerint) nem volt. Íme: 4. SG T, TOT PERS, TOT CAS. Ezek tehát a klasszikus singulare tantumok. Mintegy 3000 ilyen címszó volt az ÉrtSz.-ben; mi Zaliznyakkal egyetértve kétkedünk e kategória objektív nyelvi-nyelvtani meglétén. – 6. SG T, TOT PERS, OBL T az egyetlen *poncius* (ti. csak függő esetes–tárgyesetes–alakok vannak, bármely személy mind a két számban, de maga a főnév mindig egyesben: *a ponciusát, a ponciusodat, a poncius(o)tokat* stb.). – 7. PL T, TOT PERS, TOT CAS. Ezek lennének a klasszikus plurale tantumok: *függönykarikák, tüdő-*

*hólyagocskák, naturáliák, üzelmek, mézeshetek...* 10. TOT NUM, IMPERS T, TOT CAS: *búcsúfia, napnyugta...*; *úrforma, sajtóhang ...*; 12. Mindössze négy címszó: *ólomláb* (csak így: *ólomláb(ak)(on)*), *közelebbi, rettentő* (belső tárgyként), *hosszú* (ÉrtSz.: „főleg többes számú tárgyraggal” – a „többes” megszorítás miatt a 18-ba? – 13. SG T, IMP T, TOT CAS: *maga*<sup>(1)</sup>, *bá*<sup>1</sup>..., *locspocs, nagysád...*, *hintapalinta, háromféle...*; *rosszseb, haragszomrád...*, *sárgája, eleje* stb. – 14. SG T, IMP T, NOM T: *híre-neve, amice, domine, tréjadolog...* 15. SG T, IMP T, OBL T: *sorba, cigányutcára...*; *önmaga, gléda...* 16. PL T, IMP T, TOT CAS: *jelenvoltak, légutak...*; *májmoha, apokrif...* 18. PL T, IMP T, OBL T: csak két lexéma: *világméreteken, katonáék.* – 19. TOT NUM, PERS T, OBL T: *fia, mása*<sup>(1)</sup>..., *atyjafia, magaféle...*; *homonima...*; *katonaév...* – 22. SG T, PERS T, TOT CAS: *napa, holta...*, *netovább, leendőbeli...*; 23. SG T, PERS T, NOM T: *barátocska* (megszólításként, rendszerint első személyű birtokos személyraggal), *takargatnivaló, hátulütő...* – 24. SG T, PERS T, OBL T: *létére* (*létemre, létedre...*), *inaszakadtáig...* 25. PL T, PERS T, TOT CAS: *elei, jogászévek...* – 27. PL T, PERS T, OBL T: egyetlen címszó: *védőszárny.* – 28. Van végül egy kis egyedszámú csoport, melyet külön kellett ábrázolnunk: nem egy végpont, hanem egy út (él) hiányzik belőlük: *lépte-léptek, párthíve-párthívek* (talán jobban illenék ide az eddig a 19-be besorolt *homonima* is: valaminek a *homonimája*–ezek *homonimák*?) A MFön-ben természetesen megadjuk az egyes csoportok pontos előfordulási számát is, összesen 3000 (ezek a fentebb említett, kérdéses SG T-ok) + 693 lexéma (37,211).

**15.** Melléknév: tőtípus, határozóragos alak, középfok (a tőtípust követő oszloppár üres). Ez a szófaj alaktani szempontból még egyáltalán nem került feldolgozásra a DT-ből tudomásunk szerint<sup>10</sup>. Pedig lenne mit kibányászni itt, pl. a szakirodalomban eddig is érintett *-an/-ul* (*hosszan* de *rosszul*) alakpár eddiginél nagyobb (ti. az egész ÉrtSz.-anyag) korpuszon való vizsgálata, a fenti értelemben vett hiányosak stb. Megjegyzendő, az eddig is érintett főnévmelléknevek stb. főnévként voltak kódolva, (több nem fért a három oszloppárba; az első pár, a tőtípusé, közös volt a névszóknál: *buta-butább* éppen úgy, mint *alma-almát*).

**16.** Az ige elvileg másfajta kódolásban részesült (vö. VégSz. 18–20). A speciális igekódokat a VégSz. számára emlékezetünk szerint Jánoska Sándor dolgozta ki (esetleg e sorok írója, vagy ketten közösen), ő és Jakab László vett részt az eredmények feldolgozásában is.

<sup>10</sup> Azóta, hogy ezeket írtam, egy akadémiai előadásomban magam láttam hozzá e kérdés vizsgálatához: 1994. Meg kell mondanom, hogy például Kálmán Bélához – Tompa Jóska-hoz képest semmi elvileg újat nem találtam a nagyobb anyag átnézése, annak ábrákon való rögzítése stb. után sem. (Elvileg új pl. a BMT.)

**17. K é p z ő.** Ennek az oszlopnak regényes (ti. a legkésőbb felvett) sorsáról a történeti részben már szóltam. Azon az ott említett sajátosságon (hogy ti. itt ugrott ki a „8”-as „képző”: gyakori idegen eredetű szóvég, melyet a magyarban nem lehetett még képzőnek tekinteni) és olyan trivialiságokon kívül, hogy az igék jelentős része (igekötős v. anélküli, de) képzős, önmagában keveset lehet erről az információról mondani. Annál fontosabb az összetettség-oszloppal kombinálva: így állnak elő a különféle tőszó-csoportok s azon belül különálló kombinációk, vagy a teljes anyagot ismétlők. E két oszlopnak a jelentésszámmal való kombinálásáról is szó lesz az alábbiakban.

**18. Etimológia.** Részben a tőszókincsen belül érdekes, részint egyéb oszlopokkal kombinálva, illetőleg elektronikus gépeken folytatva az elemzést (azok segítségével az egyes fonémákig lemenve).

(19. A kártyán itt következett a hossz, melyről fentebb szóltunk: ezt az információt az elektromechanikus berendezés automatikusan állította elő és rögzítette a kártyára.)



## Egyes elemi információk kombinációja

20. A legfontosabb kombináció alkalmasint az 1-es összetettségű, 0 képzős szavak tömbje („és ami körülötte van”: a 9-es összetettségű [„nem világos, összetett-e”], 8-as és 9-es képzőjű) ezek a szigorúan vett tőszavak [Ö = 1, K = 0] és ennek a magnak a perifériája. Korábban talán fel sem merült ez a pedig jogos kérdés: hány tőszó is van szókincsünkben (a francia, az angol stb. szókincsben)? A gépkorszak előtt ilyen kérdés fel sem merülhetett: a kutató ódzkodott efféle kérdést feltenni, előre látva, mennyi mechanikus rendező-számláló munkával fog az járni az ő számára, abban mennyi hibát fog véteni, és í.t. Mi először a debreceni nyelvészkongresszuson, 1966. augusztusában számoltunk be a kérdésről – természetesen csak a magyarra vonatkozóan, akkor is sóvárogva, most is, hogy ezt összevethessük más nyelvek hasonló eredményeivel. Ekkor derült ki, hogy nem is olyan egyszerű ez a kérdés: „a tőszavak négy alcsoportját kaptuk meg...: a) vitathatatlan tőszavak (1 tő, 0 képző): *lát, fű, kék* stb.)” (6000 címszó, a 10000-es széles értelemben vett tőszó 60 %-a), „b) gyakori idegen előtagot tartalmazók” (9-es tő, 0 képző): *mikroszkóp, bibliotéka* stb. (206 címszó); „c) gyakori idegen szóvéget tartalmazók” (8-as képző): *pliszé, imitáció, kliens* (2340 címszó); „d) ragos, jeles alakban megmerevedett tőszavak” (9-es tő, 1-es képző): *után, szembe, újra* (1435 címszó) (12,520).

20.1 Tőszókincsünk statisztikai feldolgozása. E téren a legegyszerűbb, elemi adatokat már ebben a közleményünkben megadtuk (12:521). Eszerint az a) típusú (legszigorúbb értelemben vett) tőszavak 43 %-a nincs benne a SzófSz.-ban, 10 %-a finnugor, 9 %-a szláv eredetű, 8 %-a magyar belső keletkezésű. Mindennek felhasználásához itt csak egyetlen példát. Alkalmasint érdemes lenne tudnunk, melyik nyelv mennyire „eredeti”, mennyire van közel tőszókincse az eredeti állományhoz (ami a mi esetünkben a finnugor volt), annak mai alakjában. (A kínai például feltehetően igen közel áll „önmagához”, saját múltjához.)

20.2 Tőszókincsünk fonémaállományának etimológiai rétegenkénti statisztikai vizsgálata elektronikus gépeken volt célszerűen végrehajtható már a legkorábbi időszakokban is.

20.2.1 Egy igen egyszerű kérdés erről a szintről: vajon milyen az egyes rétegek magánhangzó-aránya? Nyilván olyasféle választ vártunk, hogy a finnugor réteg a legmagánhangzósabb és, mondjuk, a szláv a legmássalhangzósabb. Nem ezt kaptuk, vö. 31:19, a 2. táblázattal. Itt csak a dolog summája álljon: a legmagánhangzósabbak lettek a latin-görög eredetű elemek mai magyar, ÉrtSz.-beli alakjukban, őket követik a szlávok (!), majd a törökök, s csak az után a finnugorok. E négy rétegben a magánhangzók aránya rendre 42, 42, 40 és 40 %. Nyel-

vünk rendszere tehát alaposan megmagyarosította a bekerült szláv elemeket, ha „magyarosításon” mássalhangzók elhagyását, magánhangzók betoldását stb. értjük. (Jakobson 1973-ban nálunk jártakor szóban hozzátette ehhez, megtudva róla: amikor a szlávból a kölcsönzés történt, még a magyarot környező szláv nyelvek sem lehettek annyira mássalhangzó-gazdagok, mint ma.) Mai szinkron rendszerünkben ennek nagyon megfelelő, csak ott is némileg meglepő statisztikai képet kaptunk: legmássalhangzósabbak a költői szövegek lettek (ez a meglepő: őket zenélőbbnek, magánhangzósabbaknak képzeltük), bennük a sok ősi szóval (*szív, kéz, láb...* – ősiek, ősiek de igen mássalhangzó-gazdagok!).

Ha jól emlékszem, Bárczi tanár úr mondott annak idején valami olyasfélét, hogy a SzófSz. a maga megjelenésével nagy mértékben hozzá fog járulni az etimológiák tömeges elévüléséhez. A DT részben más okból, főleg talán korán-jött volta miatt természetesen szintén korszerűsítésre vár több oldalról is: a SzófSz. helyett, a TESz, részben elektromechanikus (! – lényegében XIX. századi technika) gépek helyett végig elektronika. Ezt mi már természetesen nem érjük meg, mint ahogy Bárczi Géza sem például a TESz-t, de, ha módunk lesz rá, a túlvilágról is érdeklődéssel fogunk betekintgetni az akkori anyagokba.

20.2.2 Az egyes etimológiai rétegeket el tudtuk helyezni egy többdimenziós térben: itt az egyes fonémák rétegenkénti terheltsége (gyakorisága) adta az egyes dimenziók mennyiségi értékeit. Eszerint – például – fonémaállományát tekintve a magyar belső keletkezésű réteghez van az összes többi réteg a legközelebb, másrészt a finnugor eredetűektől a legtávolabb.

20.3 Megint egy egyszerűbb statisztikai összefüggés: a hosszúság és a magánhangzó-gazdagság közötti (a hosszúság és a tőszó típusa, tehát közvetve etimológiai rétege közötti) kapcsolat talán triviális: az a típusú tőszavak a legrövidebbek (vö. 35:37). Talán nem ilyen könnyen belátható azonban, hogy minél hosszabb egy szó, annál magánhangzósabb: a *szinkronciklofazon*, *ribonukleinsav* stb.-félék jobban „zenélnek” e szempontból, mint a *térd*, de akár a *szív* stb.-félék. (Az a)-típusú, tehát vitathatatlan tőszók átlagos hossza 5 fonéma, az össz fonémaállomány 40 %-a magánhangzó, a c)-típusú tőszók átlagos hossza 8 fonéma, a teljes állomány 44 %-a magánhangzó.) Ez az eredmény összhangban van korábbi, meglepőnek tűnhető, megfigyelésünkkel, amely szerint rövid szavaink s az azokból álló költői szövegek mássalhangzósabbak, mint a modern szakszövegek. Másrészt ugyanezt a megállapítást így is megfogalmazhatnánk: nyelvünkben minél hosszabbak s így bonyolultabbak a szavak egy szövegben vagy a szótárban, annál egyszerűbb (rövidebb) szótagokból épülnek fel és megfordítva. Ez áll a szótár egyes részeire is, amit itt közvetlenül mértünk. Ha ui. egy szöveg (a szókincs valamelyik része stb.) magánhangzóban gazdagabb, ha tehát benne több magánhangzó esik ugyanannyi mássalhangzóra, ott a szótagok is szükségszerűen rövidebbek-egyszerűbbek. (Erre s a többi mennyiségi összefüggésre vö. még levezető tagi székfoglaló előadásomat: (38).) Ámde ne menjünk mélyebbre ezen a

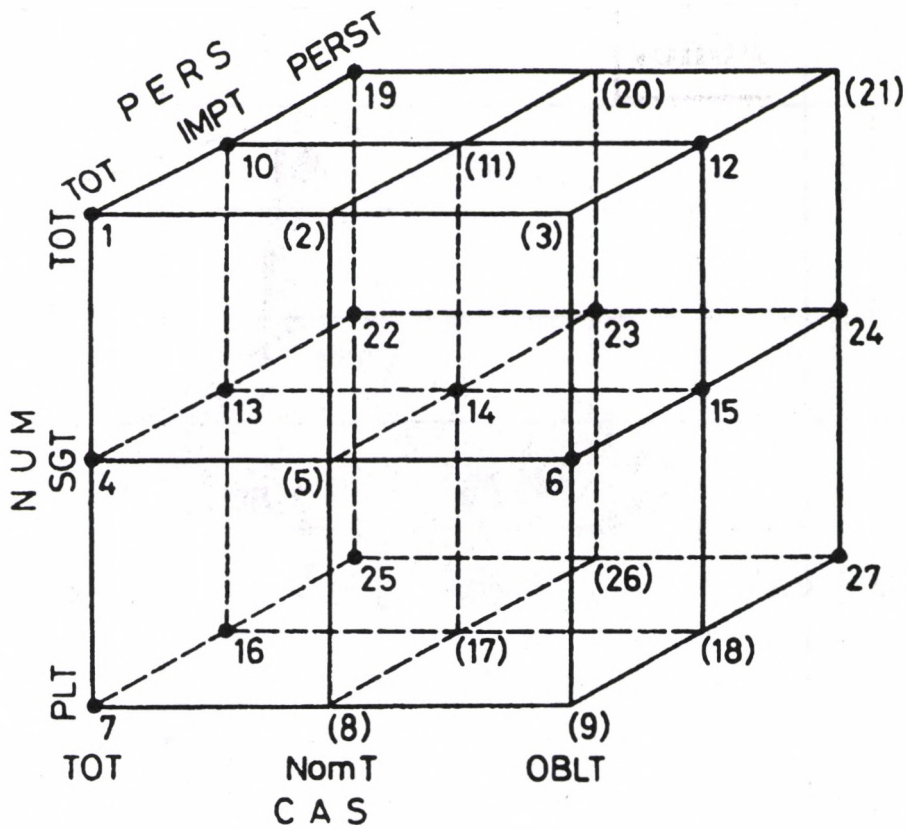
téren: várjuk meg a TESz-t és az elektronikus gépeket! (A valóságban természetesen nem vártam. Erről először az egri konferencián számoltam be.)

## 21. Két ritkább keresztcsoportosítás:

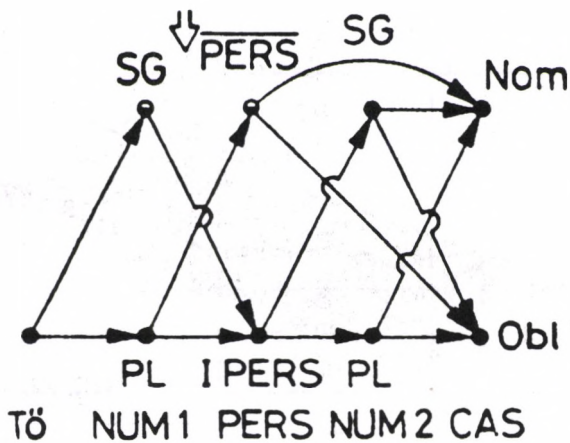
21.1. Az természetesnek tűnik fel, hogy minél bonyolultabb összetételű (összetett, képzős, igekötős stb. és ezeknek különféle kombinációja) minél „motiváltabb” egy lexéma, annál kevesebb jelentése kell, hogy legyen: az egyes morfológiai elemek szűkítik a fő jelentését. Ezt a kérdést járja körül a (33). Itt először szófajonként meg kell állapítani, mi is a fő új-lexéma-képző morfológiai alkotásmód (főneveknél az összetétel, igéknél az igekötős összetétel és a képző). Másrészt meg kellett állapítanunk, hogy az egyes szófajok kvázi természetüknél fogva polyszémebbek (e téren az ige vezet). S azután az alaktani bonyolítás valóban általában csökkentette a többjelentésűséget. É. Kiss Katalin, aki egyébként is elég szigorú hozzám, ezt a kombinációt már túlságosan sok – öncélú? – mennyiségi adatot tartalmazóként értékelte.)

21.2. Vajon a szófaj és az eredet között van-e összefüggés? Van: egyrészt az igék a legősibbek, másrészt minél ősbibb egy etimológiai réteg, annál népesebb származék (összetétel, képzős alakulat stb.) képviseli a szótárban. Vagyis, míg maga a tőszókincsünk például elég „szlávós” (majd’ annyi benne a szláv elem, mint a finnugor), addig a szótárunk egésze lényegesen ősbibb, a származékok ősi elemekből, tövekből lévén nagy arányban összetéve. (Részletesen, a pontos számszerű adatokkal vö. minderre (38).)





6. ábra



7. ábra

PERS NUM	TOT PERS			IMPERS T			PERS T			Σ/CAS			Σ/NUM
	TOT CAS	Nom T	Obl T	TOT CAS	Nom T	Obl T	TOT CAS	Nom T	Obl T	Σ/TOTCAS	Σ/Nom T	Σ/Obl T	
TOT NUM	① /	② /	③ /	⑩ 40	⑪ /	⑫ 4 ÓLOMLÁB	⑲ 40	⑳ /	㉑ /	80	0	4	84
.SG T	④ (kb.3000)	⑤ /	⑥ 1 PONCIUS	⑬ 170	⑭ 20	⑮ 150	㉒ 90	㉓ 10	㉔ 30	260	30	181	471
PL T	⑦ 30	⑧ /	⑨ /	⑰ 100	⑱ /	⑲ 2 VILÁG- MÉRE- TEKBEN	㉕ 5 ELEI	㉖ /	㉗ 1 VÉDŐ- SZÁRNY	135	0	3	138
Σ/PERS/ CAS	30		1	310	20	156	135	10	31	475	30	188	
Σ/PERS	31			486			176						693

## A debreceni Thesaurus irodalma 1963–1983, 1994

### A) A DT irodalma

1963

– a belső címlapon, valójában 1964. márc. a 2. lapon:

Mathematical and Applied Linguistics at the University of Debrecen –213–218  
[214] 0

1964

О готовящемся обратном словаре венгерского языка. – *CL* III. 205–11 1

A magyar szókincs lyukkártya-rendszerű feldolgoása. – A Magyar Tudományos Akadémia Nyelvtudományi Intézete és a Kossuth Lajos Tudományegyetem Matematikai és Alkalmazott Nyelvészeti Munkacsoportja. A LYUKKÁRTYARENDSZERŰ ADATGYŰJTÉS ÉS ADATFELDOLGOZÁS KÉRDÉSEI A NYELVTUDOMÁNYI MUNKÁBAN munkaértekezlet előadás-kivonatai. Budapest, 1964. október 15–16. 8–11 2

A magyar szókincs lyukkártya-rendszerű feldolgoása. – *Nyr.* 88. 4. 457–64 3

1965

A Magyar Szóvégmutato Szótár és a vele kapcsolatos munkálatok. – *MNy* LXI. 2. 187–200 4

Le vocabulaire du hongrois contemporain sur cartes perforées. – *Cahiers de Lexicographie* 7. II. 103–17 5

1966

A magyar szókincs törzsanyagának lyukkártyarendszerű feldolgoása. – *Információ-elektronika*. I. 1. 52–56 6

Обзор таблиц венгерского словаря, полученных на перфокартных машинах. – *CL* VI. 158–62 7

Megszólaltak a gépek. – *Nyr.* 90. 2. 181–8 8

Kettős és hármas szófajiságú szavaink. – *Nyr.* 90. 4. 404–8 9

1967

Jánoska Sándor, A magyar ige automatikus toldalékolásának egy modellje. – *Nyelvt.Ért* 58. 564–8 10

О некоторых количественных характеристиках словарного состава языка. – *Slavica* VII. 51–8 11

A magyar szókincs gépi feldolgozásának egyes eredményei és további problémái. – *Nyelvt.Ért* 58. 518–22 12

37

A Magyar Szóvégmutato Szótár és az alaktani ingadozások vizsgálata [Jánoska Sándorral közösen.]. <i>Mny.</i> LXIII. 2. 138–48	13
Die Bearbeitung des ungarischen Wortschatzes auf Lochkartenmaschinen. – <i>ALH.</i> XVII.1–2. 141–72	14
Néhány fontosabb adat szókinsünk szófaj és ezen belül hosszúság és jelentésszám szerinti megoszlásáról. – <i>Nyelvfeldolgozás és dokumentáció.</i> Szerk. Szépe György. 104–16	15
Tőigéink. – <i>Nyr.</i> 91. 1. 45–52	16
A <i>buli</i> és társai az Értelmező Szótárban. – <i>Nyr.</i> 91. 3. 346–53	17
О некоторых количественных характеристиках словарного состава языка. – <i>Slavica</i> VII. 51–8	18
Машинная обработка словарей.– <i>Конференция по машинному переводу.</i> Тезисы докладов. Ереван, 1967. – 115–117	19
1968	
Сzófaj és végződés. – <i>Nyr.</i> 92. 1. 23–37	20
Amíg a gépi listából olvasható tanulmány lesz. – <i>Nyr.</i> 92. 3. 321–33	21
Hangutánzó és hangfestő szavaink az Értelmező Szótárban. – <i>MNy.</i> LXIV. 1. 32–8	22
Adatok tőszókinsünk etimológiai rétegeinek végződés szerinti eloszlásához. – <i>NyK.</i> LXX. 1. 200–9	23
1969	
<i>A magyar nyelv szóvégmutato szótára.</i> [VégSz.] Szerk. Papp Ferenc. Budapest. Akadémiai Kiadó. 594 pp.	24
Gépi eredményeink felhasználása szótárírásunkban. – <i>Nyr.</i> 93. 3. 376–81	25
Az Értelmező Szótár legtöbb jelentésű szavairól. – <i>Nyr.</i> 93.1. 94–7	26
О машинной обработке одноязычных словарей. – <i>Научно-техническая информация.</i> Серия 2. Информационные процессы и системы. 3. 20–9 [Москва]	27
1970	
Sur quelques lois statistiques du vocabulaire. – <i>Actes du X<sup>e</sup> Congrès International des linguistes.</i> Vol. IV. Bucarest 1970. 471–5	28
Деякі кількісні характеристики одного угорського словника.– <i>Питання структурно? лексикологі?.</i> Киев, 1970. 167–78?	29
1971	
A magyar nyelvre irányuló számítógépes vizsgálatok a Kossuth Lajos Tudományegyetemen. – <i>MNnyj.</i> XVII. 147–58	30



Szókincsünk -at/-et végű névszóiról. (Jakab Lászlóval közösen.) <i>Mny.</i> LXVII. 1. 56–57	31
<b>Hunyadi László</b> , Az Értelmező Szótár „új” stílusminősítésű szavairól. – <i>Nyr.</i> 95. 3. 281–7	32
1972	
A lexémák alaktani szerkezete, szófaja és jelentésgazdagsága. <i>MNy.</i> LXVIII. 1. 23–7	33
1973	
О некоторых общих чертах славянских заимствований в венгерском языке. <i>StSl.</i> XIX. 1–3. 225–34	34
Tőszókincsünk etimológiai rétegeinek hangstatisztikája. <i>NyK.</i> LXXV. 1. 3–40	35
1974	
A lexémák szófaja, alaktani szerkezete és eredete. – <i>MNy.</i> LXX. 1. 55–68	36
1975	
<i>A magyar főnév paradigmatis rendszere</i> [Leírás és automatikus szintézis]. [MFön.] Budapest. Akadémiai Kiadó 329 pp.	37
1978	
Magyar és orosz nyelvű szövegek számítógépes vizsgálata. – <i>MTA I. OK.</i> 30. 4. 365–91	38
1983	
Шандор Яношка, Словарные статьи венгерской части многоязычного автоматического словаря. Международный семинар по машинному переводу. <i>Тезисы докладов.</i> ВЦП Москва. 252–3	39
1994	
Számítógépek és nyelvész-gondolatok. <i>MT XXX/X</i> 1994.10. 1202–1208	40

## B) Egyéb

- Берков, В.П., Б.А.Ершов, О попытках машинного перевода. – ВЯ. 1955.6. 145–148
- Bielfeldt H.H., Rückläufiges Wörterbuch der russischen Sprache der Gegenwart. Berlin 1957
- Chevalier, J.G., P. Encrevé, La création des revues dans les années 60. Matériaux pour l'histoire récente de la linguistique en France – *Langue française* 63: 57–102 (Sept. 1984)
- Chomsky, N., *Syntactic Structures*. The Hague, Mouton and Co. 1957
- Karlgren, Hans, The 13<sup>th</sup> International Conference on Computational Linguistics COLING 90. – *The Finite String* 15. Number 3. September 1989. 31–33
- Kornai, András, Szótári adatbázis az akadémiai nagyszámítógépe *Műhelymunkák a nyelvészet és társadalomtudomány köréből*. MTA Nyelvtudományi Intézet, Budapest, 65–279 (1986)
- Newmeyer, F.J., *Linguistic Theory In America*. Orlando – San Diego – etc.: 1986
- Падучева, Е.В., N. Chomsky, Syntactic Structures (рец.) ВЯ. 1959.1. 133–138
- Papp, F., Új irányzatok a szovjet nyelvtudományban. – *NyK*. LXI. 392–410 (1959) [Eredeti címe: *Strukturalista irányzatok...*]
- Papp, Ferenc, *Mathematical Linguistics in the Soviet Union*. The Hague, Mouton and Co., 1966
- Prószéky, Gábor, *Számítógépes nyelvészet*. Budapest. Számítástechnikai-Alkalmazási Vállalat. 1989
- Tarnóczy, Tamás, A magyar magánhangzók eloszlása. – *MNy*. 38. 352–357 (1942); 39. 369–374 (1943)
- Vértes, Edit, Adalékok a magyar nyelv hangtani szerkezetéhez. – *NyK*. 54. 96–140 (1952), 55. 138–180 (1953), 56. 215–266 (1954)



