Általános nyelvészeti tanulmányok

II.

Általános nyelvészeti tanulmányok

II.

Szerzők:

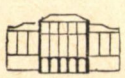
Antal László, Deme László, Dezső László,
Dömölki Bálint, Fónagy Iván, Gábor Mihály,
Gáldi László, Hadas Ferenc, Hajtman Béla,
Hell György, Kalmár László, Kelemen József,
Kiefer Ferenc, Kiss Károly, Kónyi Sándor,
Kozma László, Mikó Pálné, Németh Gyula,
Papp Ferenc, Papp László, Péter Rózsa,
Petőfi S. János, Rényi Alfréd, Ruzsiczky Éva,
Sipőczy Győző, Szalai Sándor, Szende Tamás,
Szépe György, Tamás Lajos, Telegdi Zsigmond,
Vértes Edit

Készült
a Magyar Tudományos Akadémia
Nyelvtudományi Intézetében

Általános nyelvészeti tanulmányok II.

A matematikai nyelvészet és a gépi fordítás kérdései

Szerkesztő
Kalmár László
és
Telegdi Zsigmond



Akadémiai Kiadó Budapest, 1964

A Magyar Tudományos Akadémia
Nyelvtudományi Intézete által

a Magyar Tudományos Akadémia
Elnökségi Kibernetikai Bizottságának
és
a Magyar Tudományos Akadémia
Nyelv- és Irodalomtudományi Osztálya
Általános Nyelvészeti Munkabizottságának
közreműködésével

Budapesten 1962. március 8—9—10-én rendezett munkaértekezlet tudományos anyagának alapján összeállított tanulmánykötet

© *Akadémiai Kiadó. Budapest. 1964*

AK 389 k 6467

Printed in Hungary

A szerkesztők bevezetése

Csupán technikai jellegű bevezetést tartottunk szükségesnek, mivel a tárgykör tartalmi vonatkozásairól a kötetben amúgy is kifejtettük véleményünket. A bevezetés mindössze annyit kíván elmondani: hogyan született meg ez a kötet.

1. A Magyar Tudományos Akadémia Nyelvtudományi Intézete — a Magyar Tudományos Akadémia Elnökségi Kibernetikai Bizottságának és a Magyar Tudományos Akadémia Nyelv- és Irodalomtudományi Osztálya Általános Nyelvészeti Munkabizottságának közreműködésével — 1962. március 8—9—10-én munkaértekezletet tartott „A matematikai nyelvészet és a gépi fordítás kérdései" címmel. Kötetünk ennek a konferenciának az anyagán alapul.

A munkaértekezlet előkészítéséről, rendezéséről és lefolyásáról ebben a kötetben nem lesz szó. Erről tájékoztatnak a konferenciáról írott beszámolók: Szépe György: Magyar Tudomány 7. 521—5 (1962); Petőfi S. János: Magyar Nyelvőr 86. 374—8 (1962); Kiefer Ferenc és Szépe György: A Magyar Tudományos Akadémia Nyelv- és Irodalomtudományi Osztályának Közleményei 19. 321—35 (1962); ez utóbbi helyen megjelentek a konferencián elfogadott határozati javaslatok is (335—9).

2. Ez a kötet a konferencia tudományos anyagát fogja át: az elhangzott előadásokat mint tanulmányokat, az előadásokhoz kapcsolódó vagy helyenként többé-kevésbé önálló hozzászólások tudományos részét. A konferencia formális, protokolláris vonatkozásaitól eltekintünk, egészében mellőzzük a harmadik napi — tudományszervezési jellegű — vitát is.

Az eredeti előadások és hozzászólások sorrendjén a kötetben — apró kivételeket nem tekintve — nem változtatunk. A kötet két részben közli a tanulmányokat és a hozzászólásokat. Mindkét rész írásbeli anyaga a konferencia egy-egy napjának szóbeli anyagán alapul, de korántsem azonos azzal. Ennek az eredeti sorrendnek a lényeges módosítása — esetleges tartalmi előnyei mellett — kihatással lett volna a hozzászólások sorrendjére, a reflektálások visszaadására.

3. A tanulmányok (előadások) egy része és a kisebb közlemények (hozzászólások) túlnyomó többsége magnetofon-szalagról készült áttétel, illetőleg

azoknak írásbeli formába való alapos átdolgozása. Itt mondunk köszönetet Serly Sándornénak (a konferencia lelkiismeretes titkárnőjének), Újváry Ferencnek (aki a hangfelvevő berendezést felszerelte) és Bodor Györgynének (aki a lejegyzés és átgépelés hosszadalmas és fárasztó munkáját szorgalmasan végezte).

4. Két hozzászólást írásban nyújtottak be. Ezeket is közöljük az egyes részek végén. Azért nem soroltuk be őket a többi hozzászólás közé, mivel ezekre a vita már nem lehetett tekintettel.

5. A szerkesztésben arra törekedtünk, hogy a munkaértekezlet anyagának maradandó tudományos részét örökítsük meg. Éppen ezért igyekeztünk a szerzők segítségével az előadásokat műfaji szempontból a tanulmányhoz, a hozzászólásokat pedig a „kisebb közlemény"-hez közelíteni. Ezt kívánja elősegíteni a közlemények erős tagolása, a táblára felrajzolt ábrák közreadása, valamint a néhány szerzőnél nagy számban található jegyzet és hivatkozás is.

Az elhangzás és a megjelenés közti aránylag hosszú időköz miatt néhány szerző bővítette, továbbfejlesztette jegyzet-apparátusát. Hasonlóképpen hozzá kellett járulnunk ahhoz is, hogy egy szerző (Kiefer Ferenc) egy másik szerzővel való részleges átfedés csökkentése miatt módosítsa előadásának címét és tárgyát. J. Soltész Katalinnak „Guiraud statisztikai módszere a szókincs vizsgálatában" című előadása az Általános nyelvészeti tanulmányok I. kötetében jelent meg (263—272).

A konferenciaszerű anyagból meghagytuk az elnöki megnyitót, illetőleg a zárószavakat, mivel ezek a jelentős tudománypolitikai megnyilatkozások dokumentálják a Magyarországon ezzel a kötettel szervezetten színre lépő új nyelvészeti ágazatok jelentőségét, illetőleg távlati fejlődésének néhány problémáját.

6. A kötet az „Általános nyelvészeti tanulmányok" sorozat II. tagja. A sorozat első kötete: Általános nyelvészeti tanulmányok I. (Budapest, 1963) szerkesztő Telegdi Zsigmond, 338 lap.

Németh Gyula

Elnöki megnyitó*

Munkaértekezletünket megnyitom és üdvözlöm az Akadémia Elnökségi Kibernetikai Bizottságának, továbbá az Akadémia I. Osztálya Általános Nyelvészeti Munkabizottságának megjelent tagjait. Most jövünk össze ilyen összetételben nyilvánosan az első alkalommal s minden reményünk megvan arra, hogy ezt az első összejövetelt állandó, zavartalan és eredményes együttműködés követi.

Ez a munkaértekezlet voltaképpen szerves folytatása annak a vitának, melyet egy évvel ezelőtt a Nyelvtudományi Intézet és a Művelődésügyi Minisztérium Nyelvtudományi Munkaközössége rendezett a nyelvtudomány elvi kérdéseiről, illetőleg nyelvtudományi elveink újabb kori átalakulásáról s ezen az alapon a nyelvtudományi kutatás jövő feladatairól. Azok a kitűnő előadások és hozzászólások, amelyek ezen a vitán elhangzottak, nemrégen jelentek meg.[1]

Ez alkalommal nem megyek túl tavalyi elnöki megnyitómban kifejtett elvi álláspontomon. Nem jó az a tudományban, ha az elvi álláspont egy-két év alatt módosul vagy éppen gyökeresen változik: az elvi álláspont megváltozását meg kell előzni evidens ténybeli eredményeknek. Ezek pedig nem jönnek olyan könnyen, mint az új elvi állásfoglalások. Szükségét látom azonban, hogy tavaly jelzett elveim néhány pontját élesebb megvilágításba helyezzem.

Egész életemben a XIX. század folyamán kialakult hagyományos nyelvtudománnyal foglalkoztam, s ezt én már nem tudom megtagadni, legmélyebb meggyőződésem tiltaná meg, hogy megtagadjam. De nem is kell, nem is lehet ezt megtagadni. Ennek a kutatási iránynak a jövőjét egy percig sem féltem. Ami nyelvtények ismeretét, tehát az egyszerű nyelvleírást illeti, azt semmiféle nyelvész, sem az újgrammatikus, sem a strukturalista nem nélkülözheti, s az a nyelvész, amelyik például az új nyelvjárási szótárt vagy a nyelvatlaszt feleslegesnek tartja, maga alatt fűrészeli az ágat. Az elvi elgondolások sohasem lehetnek fontosabbak a tényeknél, amelyekre

* E megnyitót az ülésen erősen megrövidítve olvastam fel, de a fel nem olvasott részeket elmondottam a Kalmár László előadásához fűzött hozzászólásomban. Most — mint ezt az ülésen is jeleztem — a megnyitót eredeti, hosszabb formájában adom ki.

[1] A Magyar Tudományos Akadémia Nyelv- és Irodalomtudományi Osztályának Közleményei 18. 5—174 (1962).

vonatkoznak. Megjegyzem azonban — és ez a megjegyzésem tanácskozási tárgyunk lényegébe vág —, hogy ettől az egyszerű leírástól élesen megkülönböztetem a tudományos strukturalista leírást, amely a legmagasabb tudományos színvonalon mozoghat és alapvető jelentőségű felismeréseket adhat. De nem szabad a két módszert összetéveszteni. Ha én egy nyelvtényt csupán meg akarok ismerni, annak a legegyszerűbb és legkönnyebben megtalálható közlését várom és nagyon rossz néven veszem, ha a grammatikus azt a bonyolult, egyéni strukturális rendszer keretében tálalja fel.

Ami pedig a történeti kutatást, a nyelvi jelenségek keletkezését, fejlődését, egymással, a dolgokkal és a társadalmi élettel való összefüggését illeti, ennek a kutatása egy materialista világnézet képviselői számára olyan elsőrangú kötelesség, mint minden más, a történeti szempontú kutatás számára hozzáférhető tudományterület kutatása.

Az egyszerű történeti érdeklődést ki kell elégíteni. Prágában mondotta egy nyelvész-kollégám, hogy a történeti kutatásokat le kellene állítani, mert a nyelvtudomány módszere még nem alakult ki. És ha valaki megkérdezi tőlem, hogy miféle hasonlóságok vannak a magyar és a finn nyelv között, mi ezeknek a hasonlóságoknak a történeti alapja, hogy hogyan keletkezett a magyarban a *kapa* szó, mi az a *Budapest* név, mi az a *Buda*, mi az a *Pest*, mi a *látok* és *látlak* igealakok egymáshoz való történeti viszonya, mért mondja a Halotti Beszéd ezelőtt 750 évvel, hogy *hadláva χoltát* a helyett, hogy *hallá, hogy meghal*, — azt feleljem, hogy erre a módszer fejletlensége miatt nem válaszolhatok? Világosan látszik a nyelv fejlődése: ne keressem, milyen társadalmi és pszichológiai tényezők vannak e fejlődés mögött?

A dolog úgy van, ahogyan Telegdi Zsigmond ezelőtt egy esztendővel mondotta: „Az új, összehasonlító- történeti, vagy inkább egyszerűen történeti nyelvtudomány rövid idő alatt ragyogó sikereket ért el. Alig egy század határai között feltárta, fő vonásaiban megrajzolta nyelvek és nyelvcsaládok egész sorának fejlődését, megmutatta, hogyan lehetséges egyáltalán a nyelvek mozgását, változását, átalakulását tudományosan vizsgálni, s ezzel új világot, a társadalmi valóság új oldalát hódította meg a a tudománynak."[2] Hozzáteszem, hogy ez a munka csak egyes kivételes területeken van többé-kevésbé — inkább kevésbé — elvégezve. Abbahagyni nem lehet, — ellenkezőleg, fokozott erővel kell folytatni.

A tradicionális nyelvtudománnyal szemben a strukturalizmus ma még gyermekkorát éli. Elvileg kitűnően meg van alapozva, de az alapelvekből levonható konzekvenciák alkalmazásában még tapogat, sokszor kétes értékű próbálkozásai vannak, a forrásban levő borhoz hasonlít, mely sokat ígér, de még nem iható.

A strukturalizmus elvi megalapozásában is vannak tévedések. Ha azt olvassuk — és ezt sokszor olvassuk —, hogy a történeti kutatás nem képes a nyelvet lényegében megvilágítani s ezt a strukturalizmustól várhatjuk, akkor egy súlyos tévedéssel állunk szemben: a keletkezés és fejlődés kutatása

[2] Telegdi Zsigmond, A nyelvtudomány újabb fejlődésének egyes kérdéseiről: A Magyar Tudományos Akadémia Nyelv- és Irodalomtudományi Osztályának Közleményei 18. 11 (1962).

vitán felüli kötelességünk s a strukturalizmus ezt nem pótolja, aminthogy a történeti kutatás sem pótolja a strukturális vizsgálatot. A nyelv vizsgálatának két, l é n y e g é b e n különböző aspektusáról van itt szó, ellenére annak, hogy szorosan egymásba fonódnak és a jövőben még szorosabb lesz a kapcsolatuk. Ugyanez a tévedés jelentkezik a strukturalizmusnak abban a tételében, hogy a történeti emlékekkel nem rendelkező nyelveket csak a strukturalizmus módszereivel lehet kutatni. Lehet, hogy vannak ilyen nyelvek. Ezeknek a strukturalizmus módszereivel való kutatása azonban nem történeti vizsgálatukat pótolja, mint ahogy — megfordítva — a francia nyelv történetének gazdag anyaga nem ment fel bennünket a francia nyelv strukturális kutatásától.

Mindezek ellenére a magam részéről a legszebb reményekkel vagyok a strukturalizmus jövője iránt és — ahogy a törökök mondják — négy szemmel várom a tervezett strukturális magyar nyelvtan megjelenését. Az sem baj, ha a vállalkozás nem sikerül, akár egy-egy részletében, akár egészében.

Azt azonban ki kell jelentenem, hogy a matematikai nyelvészet területére — pedig ez mostani ülésszakunk első tárgya —, csak nagyon csekély távolságig tudom követni az új iskolák híveit. Ez az irányzat nem a nyelvtudományban ért ki. Más a strukturalizmus esete: a strukturalizmus szervesen fejlődött az újgrammatikusok elleni reakcióból, Saussure és a Saussure utáni nyelvészetnek a munkásságából. Ezzel szemben a matematikai nyelvészet nem nyelvészeti, hanem matematikai talajból nőtt ki, a matematika fejlődésében elért új szemléleti formák tették lehetővé. Nem tudom, lehet-e itt történetesen megszerzett matematikai ismeretekkel értékes eredményekre jutni vagy pedig csak az tud jelentős eredményeket elérni, aki az újabb matematikai elméleteket mint problémákat is lényegében látja, tehát matematikus — és nem nyelvész.

Ezt én nem tudom, aminthogy nem tudom azt sem, hogy állanak a többi társadalmi tudományok, például a néprajz, a vallástörténet, a jogtudomány, a történelem, a művelődéstörténet a matematikai vizsgálati módszer alkalmazásában.

Amit most a nyelvészet és a matematikai kutatás viszonyáról mondottam, természetesen a matematikai nyelvészetnek csak a legmagasabb szintjére vonatkozik, s vannak e tudományágnak olyan feladatai, melyeknek végzésébe bármely rendes nyelvész beledolgozhatja magát. Ez egészen más valami. A statisztikai módszer alkalmazása a nyelvtudományban egyszerű kutatási eszköz, melynek lényegében semmi köze a nyelv matematikai szemléletéhez.

Itt meg kell állnom. A gépi fordítás és az információ-kutatások fontosságáról beszéltem tavaly, — most beszéljenek róla részletekben a szakemberek. Ezek nem elméleti, hanem gyakorlati kérdések, s mindent meg kell tennünk, hogy művelésüket elősegítsük.

Szeretném, ha a vita a hasonló természetű kérdések megbeszélésében gyakran mutatkozó túlzásoktól minél inkább mentes lenne, s ha minden kérdést jobbról és balról egyformán higgadt megfontolással vizsgálnánk.

Kalmár László

Matematikai és nyelvi struktúrák

A matematikai nyelvészetnek, ennek a viszonylag új komplex tudományos kutatási irányzatnak, már fejlettsége mai fokán is legalább három különböző, bár egymással szorosan összefonódó ága van kibontakozóban. Ha szabad Németh Gyula elhangzott elnöki megnyitó előadásából átvennem egy gondolatot, azt mondhatom, hogy ezek különböző mértékben megfoghatatlanok, vagy úgy is mondhatni, különböző mértékben megfoghatók a nyelvtudomány szakemberei számára.

1. A matematikai nyelvészetnek az az ága, amelyen végzett munkának legrégebbi tradíciói vannak hazánkban, s amelyen ma is örvendetesen sok eredménnyel folyik nálunk a munka, a nyelvstatisztika. Ez talán a legmegfoghatóbb a nyelvészek számára. Hiszen itt világos, mit kell csinálni: meg kell valamit számlálni. Például azt, hogy valamely fonéma, vagy valamely szó hányszor fordul elő egy hosszabb szövegben, vagy hogy hány egyszótagú, hány kétszótagú, hány háromszótagú szó fordul elő benne. Persze nemcsak egyszerű megszámlálásról, az elemi aritmetikának alkalmazásáról van szó, hanem a megszámlálás eredményének feldolgozásakor sor kerül a valószínűségszámítás törvényeinek alkalmazására is, nem ritkán abban az alakban, hogy felfigyelünk, amikor a megszámlálás eredménye eltér attól, amire a valószínűségszámítás valamelyik törvénye alapján következtethetnénk, és ennek az eltérésnek az okát keressük. Az ok persze az, hogy az adott esetben nem teljesülnek azok a feltételek, amelyek esetén a valószínűségszámítás kérdéses törvénye érvényes, és ez a negatív tény gyakran fontos következtetésre ad alkalmat.

Például ha egy költő valamely hosszabb költeménye alapján megállapítjuk, milyen gyakorisággal használja az egyes fonémákat, akkor a valószínűségszámítás törvényei alapján csak akkor következtethetnénk arra, hogy költőnk más költeményeiben is — bizonyos kisebb statisztikai ingadozásokon belül — ugyanaz lesz az egyes fonémák gyakorisága, ha feltételeznők, hogy a fonémák gyakorisága független attól, milyen költeményről van szó. Minthogy azonban ez a következtetés a tapasztalattal meg nem egyező eredményre vezet, nyilvánvaló, hogy az egyes fonémák gyakorisága ugyanazon költő esetén is függ a költemény tárgyától, a költő hangulatától és további, nyelvészeti kutatás tárgya lehet, hogyan függ

tőle. Ez vezette Fónagy Iván kartársunkat a fonémák fonológiai szerepén túl stílusképző értékük vizsgálatára.

Ide, a nyelvstatisztikai irányzathoz számítom az információelmélet nyelvtudományi alkalmazásait is, amelyek lehetővé teszik például olyan kérdéseknek matematikai módszerekkel való objektív vizsgálatát, mennyire tömörebb, vagy ellenkezőleg, mennyire bővebb beszédű az egyik nyelv, mint a másik, pl. ha egy 20 íves művet németről magyarra fordítunk, hány íves fordításra számíthatunk. Persze itt is függ az eredmény a mű tárgyától, meg a szerzőtől és a fordítótól is, hiszen az egyik ember szűkszavú, a másik meg szószátyár.

A nyelvstatisztika tehát azért viszonylag megfogható a nyelvész számára, mert, bár a matematika különböző módszereit használja, a legelemibbektől, mint az egyszerű megszámlálás, a legmodernebbekig, mint a valószínűségszámítás, a matematikai statisztika, az információelmélet, de eléggé megfogható az, hogy mire használja.

2. A matematikai nyelvészet másik ága, amelyen ugyancsak említésre méltó kezdeti sikereket értek el már hazai nyelvészeink, a gépi fordítás, gépi kivonatkészítés, általában az elektronikus számológépeknek és egyéb matematikai gépeknek nyelvészeti alkalmazásai. Itt már talán sok nyelvtudós számára megfoghatatlannak tűnik az, hogyan lehetséges, hogy gépek — amelyek bár bonyolult emberi alkotások, de az a mozgásforma, amelynek keretében működnek, mechanikai vagy fizikai, tehát primitív mozgásforma az emberi agy működéséhez képest — olyan bonyolult feladatokat képesek elvégezni, mint valamely szövegnek egyik nyelvről a másikra való fordítása, vagy valamely tudományos dolgozat leglényegesebb mondanivalóit összefoglaló kivonat készítése. De itt legalább az a cél, amit el kell akarunk érni a gép segítségével, megfogható a nyelvész számára. S ez már elég ahhoz, hogy a nyelvész megoldhassa e téren a maga feladatát: azt, hogy a kérdéses cél eléréséhez — például az egyik nyelvről a másikra való fordításhoz, dolgozat kivonatának készítéséhez, vagy mondjuk egy konkordancia-szótár összeállításához, vagy másféle adatgyűjtéshez, cédulázáshoz — szükséges teendőket teljes szabatossággal és részletességgel megfogalmazza anélkül, hogy közben valamiféle emberi képességre, mint az intelligenciára, a szöveg megértésére vagy a nyelvérzékre hivatkozna. Az már a matematikus feladata, hogy a nyelvész által ily módon megfogalmazott teendők rendszerét, matematikai szakkifejezéssel: a kérdéses cél elérésére szolgáló *algoritmust*, a gép számára felfogható jelek sorozatává alakítsa, ahogyan mondani szokás, programozza.

3. Ebben az előadásban azonban a matematikai nyelvészet egy harmadik ágáról fogok beszélni, amelyik talán legkevésbé megfogható a nyelvtudós számára, épp ezért a legkevésbé ismert nálunk, bár ifjú nyelvészeink egy része már hozzáfogott, mégpedig nagy energiával, a tanulmányozásához. Ezt az ágat talán úgy lehet jellemezni, hogy a matematikus tapasztalatait igyekszik gyümölcsöztetni a nyelv strukturális vizsgálata számára. Azokról a tapasztalatokról van szó, amelyeket a matematikus a maga tudományának több évezredes fejlődése folya-

mán bonyolult rendszerek szerekezetének vizsgálata során halmozott fel. A matematikai nyelvészetnek ez az ága azért nehezen megfogható a nyelvtudós számára, mert a matematikának azok a fejezetei, amelyekhez az ilyen strukturális vizsgálatok tartoznak, kevéssé ismeretesek a köztudat számára. Kevéssé ismeretesek, mert a középiskolában semmit sem tanítanak róluk. Nemcsak hogy kevéssé ismeretesek, de nem is egyeznek meg a matematikáról való általános elképzeléssel, azzal, amit a *matematika* szó „mennyiségtan" fordítása tükröz: azzal, hogy a matematika állítólag kizárólag mennyiségekkel foglalkozik. Valójában azonban épp a matematika kérdéses fejezetei — elsősorban az absztrakt algebráról, a topológiáról, a halmazelméletről, továbbá a matematika és a logika határterületéről, a matematikai logikáról van szó, amely tárgyánál fogva már a logikához tartozik, s elsősorban módszerei kapcsolják össze a matematikával — nem pusztán mennyiségekből, hanem tetszőleges természetű tárgyakból, elemekből álló rendszerek szerkezetével foglalkoznak és e szerkezeti vizsgálatban kvalitatív szempontok is tekintetbe jönnek.

De azért is nehezen fogható meg a matematikai nyelvészetnek ez az ága a nyelvtudós számára, mert itt nem kész matematikai eredmények alkalmazásáról van szó, mint a nyelvstatisztikában, amely a valószínűségszámítás, a matematikai statisztika, az információelmélet már meglevő tételeit használja fel, nem is már meglevő matematikai gépek alkalmazásáról, mint például a gépi fordítás esetén. Igaz, hogy néha a nyelvstatisztikus is olyan igénnyel lép fel a matematikussal szemben, végezzen például az ő céljaira felhasználható új valószínűségszámítási kutatásokat, s a gépi fordító is megmondja, mivel tudjon többet a legközelebb tervezendő gép az eddigieknél, hogy jobban fel lehessen használni fordításra. A matematika kvalitatív módszereinek a nyelv szerkezetének vizsgálatára való alkalmazása azonban igénybejelentésnél sokkal többet kíván a nyelvésztől. Itt már nem lehet kész matematikai fogalmakat, meglevő matematikai tételeket alkalmazni nyelvtudományi kérdésekre, hanem a nyelvésznek, ha a matematikai nyelvészetnek ezt az ágát meg akarja fogni, el kell sajátítania azokat a matematikai eszközöket, amelyek a matematikus által vizsgált bonyolult rendszerek szerkezetének kutatása során kialakultak, mondhatni, az ilyen kutatáshoz szükséges matematikai gondolkodásmódot, és ezt, természetesen a nyelvész gondolkodásmódjával párosítva, önálló kutatással alkalmaznia kell egészen más jellegű, sok szempontból bonyolultabb szerkezetű rendszernek, a nyelvnek strukturális vizsgálatára. Mert igaz ugyan, hogy az algebrában, vagy a topológiában végtelen sok elemből álló rendszerek szerkezetét is vizsgálja a matematikus, ezzel szemben például a nyelv szavainak halmaza — bár sohasem tekinthető befejezett, lezárt összességnek, mégis — véges számú elemből, szóból áll; azonban más szempontból a nyelv lényegesen bonyolultabb szerkezetű rendszer, mint azok a rendszerek, amelyeket a matematikus valaha is vizsgált; — eltekintve persze attól a néhány újabb vizsgálattól, amelyet matematikusok nyelvészekkel együtt végeztek, és amelyek tárgya épp a nyelv szerkezete.

Ennek az előadásnak az a feladata, hogy a matematikai nyelvészetnek ezt a nehezen megfogható ágát megfoghatóbbá tegye a nyelvészek számára.

Előadásom azonban, és ezt előre meg kell mondanom, sokkal inkább matematikai előadás lesz, mint nyelvészeti, hiszen én matematikus vagyok — közelebbről a matematikai logikával, tehát matematikai módszereknek logikai kérdésekre való alkalmazásával foglalkozom — nem pedig nyelvész, még matematikai nyelvész sem. A nyelvészetben laikus vagyok, éppúgy, mint ahogy kitűnő nyelvész szakemberek laikusok a matematikában, egy-két kivételtől eltekintve, akik ti. a matematikából is, a nyeltudományból is szereztek oklevelet. Ezenkívül azért is inkább a nyelvtudományban felhasználható matematikai módszerekről fogok beszélni, mint ezek nyelvtudományi alkalmazásairól, mert az utóbbiak Kiefer Ferenc kartársnak ezen a munkaértekezleten tartandó előadásának tárgyát alkotják.

4. Ha nem is vagyok nyelvész, úgy gondolom, bevezetőül meg kell mondanom, hogyan látom én, laikus fejjel, azt a kérdést, amiről mostanában oly gyakran folyik a vita nyelvész-körökben, hogy tudniillik a történeti-összehasonlító módszer, vagy strukturális vizsgálat-e a nyelvtudomány igazi módszere. Kétségtelen, hogy a nyelv társadalmi jelenség, amely folytonosan fejlődik, mégpedig, a társadalommal együtt, társadalmi okokból fejlődik, ha fejlődésének a társadalom fejlődésével való kapcsolata nem is olyan egyszerű, mint ahogy ezt Marr tanította. Igaza van annak, aki azt állítja, hogy nem gondolkodik dialektikusan az, aki a nyelv fejlődését nem veszi figyelembe, mert nem alkalmazza az egyetemes mozgás, fejlődés elvét a nyelvre. De ha megvizsgáljuk, hogy mi az, ami a nyelvben a társadalommal együtt fejlődik, rájövünk, hogy nem pusztán az egyes nyelvi jelenségek, hanem a nyelv, mint bonyolult szerkezetű rendszer, egészében fejlődik. A dialektikus gondolkodás nemcsak az egyetemes mozgás, hanem — többek között — az egyetemes összefüggés elvének alkalmazását is megkívánja, tehát az sem gondolkodik dialektikusan, aki nem veszi figyelembe a nyelvi jelenségek közötti bonyolult összefüggéseket, amelyek a nyelvnek mint bonyolult rendszernek a szerkezetét alkotják. Az igazság tehát az, hogy a nyelv a társadalommal együtt fejlődő bonyolult szerkezetű rendszer; ennek az igazságnak egyik fele az, hogy a nyelv a társadalommal együtt fejlődik, a másik fele az, hogy a nyelv bonyolult szerkezetű rendszer. Aki csak az egyik felét látja ennek az igazságnak, az csak féligazságot lát, csak „féldialektikát" alkalmaz, tehát nem gondolkodik dialektikusan, akár az egyik, akár a másik felét látja.

Mármost a történeti-összehasonlító módszer a nyelvnek mint társadalmi jelenségnek fejlődése vizsgálatára szolgáló, jól bevált módszer; a strukturális nyelvvizsgálat módszere pedig a nyelvnek mint bonyolult szerkezetű rendszernek vizsgálatára szolgáló, viszonylag új módszer. A fentiekből világos tehát, hogy mindkét módszer jogosult, mindkettő szükséges, és csak együtt lehetnek alkalmasak arra, hogy a nyelvre vonatkozó teljes igazságot, a nyelvnek, mint bonyolult szerkezetű rendszernek a társadalommal való fejlődése objektív törvényeit feltárják.

Más kérdés az, hogy a nyelvtudományi vizsgálat mely területén lehetnek hasznára a nyelvésznek a matematikus tapasztalatai, amelyeket a matematikai vizsgálat tárgyát alkotó rendszerek szerkezetének tanulmá-

nyozása során felhalmozott. Gyakran hallani azt a véleményt, hogy aki belekóstol a matematikai módszerek alkalmazásába nyelvtudományi kérdésekre, annak örökre hátat kell fordítania a történeti-összehasonlító módszernek, a diakronikus vizsgálatoknak, az végérvényesen eljegyzi magát a strukturalizmussal, a nyelv szinkronikus szerkezeti vizsgálatával. Kétségtelen, hogy — eltekintve a glottokronológia első lépéseitől, amellyel valamely nyelvcsaládhoz tartozó nyelvek keletkezésének időbeli lefolyását egy ősnyelvből annak a képletnek a mintájára próbálja leírni, amely új anyagok keletkezésére vonatkozik valamely radioaktív anyagból, ami mindenképpen csak első durva megközelítést nyújtó modellnek tekinthető, továbbá matematikai módszereknek régi írásrendszerek megfejtésére való alkalmazásától — eddig csak a nyelv szinkrón strukturális vizsgálatára vettek igénybe a nyelvészek matematikai segédeszközöket. Ennek azonban nyilván az az oka, hogy azok a nyelvészek, akik ezeket igénybe vették, a nyelv strukturális vizsgálatát tartották a nyelvtudomány ma legsürgősebb feladatának, mondván, hogy a nyelv szinkrón vizsgálata elmaradt a diakronikus vizsgálatok mögött. Hogy igazuk van-e ezeknek a nyelvészeknek, annak eldöntésére nem vagyok hivatott, ezt döntsék el a nyelvtudomány szakemberei. Annyi azonban bizonyos, hogy a matematikai módszerek természete elvileg nem zárja ki a diakronikus vizsgálatokban való alkalmazásukat sem. Nem zárhatja ki, hiszen a matematika a változó mennyiségektől függő változó mennyiségek vizsgálata, az úgynevezett *matematikai analízis* kialakulásával, mint már Engels is megállapította, a dialektika területére lépett,[1] alkalmassá vált a változások leírására. Hogy más példát ne mondjak, a csillagász, amikor évszázadokra visszamenőleg, vagy éppen előre, leírja a bolygók mozgását, megállapítja a napfogyatkozások idejét, vagy még hosszabb időkre vizsgálja a csillagok fejlődését, szintén diakronikus vizsgálatokat végez matematikai módszerek alkalmazásával.

Ha azonban nem is áll az, hogy a matematikai módszerek természete elvileg kizárja a nyelv diakronikus vizsgálatára való alkalmazásukat, annyi bizonyos, hogy ahhoz, hogy matematikai módszereket diakronikus nyelvészeti kutatásokban alkalmazni lehessen, magának a matematikának is tovább kell fejlődnie. Hiszen eddig kevés kutatás történt a matematika által vizsgált rendszerek struktúrájának valamely változó adattól (matematikai szakkifejezéssel: valamely *paramétertől*), például az időtől való függésére vonatkozólag, s e kutatások is topológiai struktúrákra vonatkoznak, de például algebrai struktúrákkal kapcsolatban ilyen kutatás nem történt. (Például az úgynevezett folytonos csoportok elméletében is az egyes csoportelemeknek, nem pedig az egész csoportnak, mint struktúrának, paraméterektől való függéséről van szó. Ez pedig nagy különbség, hiszen a paraméterek változásával megváltozott csoportelemek is ugyanannak a csoportnak elemei, nem pedig egy megváltozott csoportnak vagy más struktúrájú rendszernek. Ezzel szemben a nyelv fejlődése során az egész nyelv mint rendszer változik.) De ugyanúgy ahhoz, hogy mate-

[1] F. Engels: Hogyan „forradalmasítja" Eugen Dühring úr a tudományt? („Anti-Dühring") (Bp., 1946) 117.

matikai módszereket eredményesen alkalmazhassunk a nyelv szinkronikus, strukturális vizsgálatára, szintén tovább kell fejleszteni a matematikát, minthogy a nyelv másfajta, mint már mondtam, bizonyos szempontból bonyolultabb rendszer, mint azok, amelyeket a matematikusok eddig vizsgáltak. Így a matematikai módszereknek a nyelvtudományban való alkalmazására vonatkozó kutatások nemcsak a nyelvtudomány, hanem a matematika fejlődése szempontjából is hasznosak.

5. Amint mondtam, a matematika ma már nemcsak mennyiségekből, hanem tetszőleges természetű elemekből álló rendszerek szerkezetével foglalkozik. Tagadhatatlan azonban, hogy a matematika, legalábbis annak egyik ága, az *aritmetika*, amelyből később az *algebra* kifejlődött, olyan rendszerek vizsgálatából indult ki, amelyeknek az elemei számok. Ehhez azonban előbb el kellett jutni magukhoz a számokhoz. (Az „előbb" csak logikai értelemben helyes. Valójában a számfogalom egyidejűleg, kölcsönhatásban alakult ki a számok között végezhető műveletek fogalmával. Márpedig a számokból álló rendszerek szerkezetére épp az a legjellemzőbb, hogy elemeiket műveletek kapcsolják össze.) A legegyszerűbb számokhoz, az úgynevezett *természetes számokhoz:* az 1, 2, 3, . . . számokhoz (néha a 0-t is a természetes számok közé értik; ha nem akarjuk közéjük számítani, akkor *pozitív egész számokról* beszélünk) *megszámlálás* útján jutott az ember. Miket számlált meg? Különböző elemekből álló összességeket, matematikai szakkifejezéssel: *halmazokat.* Például a vadász az elejtett vadak halmazát, a juhász a nyáját, tehát az őrizetére bízott juhok halmazát számlálta meg, így állapította meg, hány vadat ejtett el, hány juhot őriz. Ha valamilyen jelentéshez meg kell adni, hány résztvevője volt ennek az értekezletnek, akkor az ebben a teremben jelen levő emberek halmazát kell megszámlálnunk, vagyis megállapítanunk, hány elemből, esetünkben hány emberből áll ez a halmaz. Amikor Tolnai Vilmos felvetette azt a kérdést, hány szó van a magyar nyelvben,[2] ez a kérdés a magyar nyelv szavai halmazának elemei számára vonatkozott. A halmazok eszerint tetszőleges dolgok összességei lehetnek, ezeket a dolgokat a halmaz *elemeinek* nevezzük, a mondott példákban ezek egyszer elejtett vadak, egyszer juhok, egyszer emberek, egyszer szavak. De ugyanúgy lehet a magyar ábécé betűinek, vagy a magyar nyelv fonémáinak halmazáról is beszélni.

Amikor azonban megszámlálunk egy halmazt, vagyis megállapítjuk, hány elemből áll, akkor eltekintünk attól, hogy mik voltak az elemei, juhok, emberek, szavak vagy betűk-e, vagy más természetű dolgok. Eltekintünk, mert csak az érdekel bennünket, hogy hányan vannak, nem az, hogy mik és milyenek. Ez az absztrakciós lépés vezetett a számok, pontosabban, a természetes számok fogalmához. A többi fajta számokhoz másféle absztrakciós lépések vezettek. Hiszen, ha a juhász megszámlálja nyáját, lehet, hogy azt találja, hogy 42 juha van, de azt biztosan nem, hogy 43,6 juha van, azt még kevésbé, hogy —25 juha van.

[2] Tolnai Vilmos, Hány szó van a magyar nyelvben?: Magyar Nyelv 14. 33—4 (1918).

Az olyan fajta számokhoz, mint például 43,6, vagy 0,351, amelyeket *pozitív valós számoknak* nevez a matematika, már nem számlálás, hanem *mérés* útján jutunk el. Például ha megmérem, hány méter hosszú ez az asztal, akkor valamilyen pozitív valós számot kapok mértékszám gyanánt. Megeshetik, hogy a mértékszám természetes szám, például hogy az asztal hossza pontosan 2 méter, épp ezért a természetes számokat is a pozitív valós számok közé számítjuk. Ez azonban rendszerint csak olyankor esik meg, ha az asztalos szándékosan úgy csinálta az asztalt, hogy pontosan 2 méter hosszú legyen. Más esetben azonban rendkívül ritka kivételes eset, hogy valamely hosszúság mértékszáma természetes szám legyen. Általában olyanféle eredményre jut az ember, ha valamely hosszúságot méterrúddal megmér, hogy mondjuk ötször megvan benne a méterrúd hossza, de hatszor már nincs, vagyis a kérdéses hosszúság 5 méter és még valami maradék, amely már kisebb, mint egy méter. Ezt a maradékot a méter valamely részével, például deciméterrel próbáljuk lemérni, és ismét az a ritka kivételes eset, hogy egész számszor, például nyolcszor van meg benne a deciméter — ebben az esetben 5,8 méter a lemérendő hosszúság —, rendszerint megint marad még egy kicsi, amit például centiméterrel próbálunk lemérni és így tovább. Az a ritka kivételes eset, ha a mérés befejeződik, mert nem marad semmi; rendszerint nem fejeződik be, csak abbahagyjuk, amikor már olyan kicsi marad, amit már elhanyagolunk, „kicsire nem nézünk". De ha nem hagynók abba, akkor végtelen tizedes törtet kapnánk pontos mértékszám gyanánt, és csak abban a ritka kivételes esetben, amikor befejeződik a mérés, kapunk véges tizedes törtet. Eszerint a pozitív valós számok (nagy ritkán) véges, vagy (rendszerint) végtelen tizedes törtek alakjában adhatók meg. Az utóbbiak, a végtelen tizedes törtek közül némelyek pontosan megadhatók közönséges tört alakjában is; pl. $0{,}333\ldots = \frac{1}{3}$. A véges tizedes törtek mind megadhatók közönséges tört alakjában is, pl. $3{,}75 = 3 + \frac{75}{100} = 3 + \frac{3}{4} = \frac{15}{4}$. A végtelen tizedes törtek közül azonban csak ritka kivételt alkotnak azok, amelyek közönséges tört alakjában írhatók. Például az, hogy valamely hosszúság pontosan $\frac{47}{73}$ méter, azt jelenti, hogy 47 olyan egyenlő hosszúságú darabból áll, amilyenből 73 tesz ki egy egész métert. Márpedig ha a méterrúdat eltöröm, mindenkinek világos, hogy ritka véletlen, ha úgy törik el, hogy az egyik darabja 47 olyan, egymással egyenlő hosszúságú kisebb darabból áll, amilyenből 73 tenne ki egy métert. Eszerint a természetes számok és a pozitív közönséges törtek, amelyeket közös néven *pozitív racionális számoknak* nevezünk, ritka kivételt alkotnak a pozitív valós számok között; a többieket, amelyek óriási többségben vannak, *pozitív irracionális számoknak* nevezzük. (*Irracionális* itt, legalábbis értelemszerűen, nem 'esztelen'-t jelent, hanem két természetes szám aránya [vagyis hányadosa] alakjában meg nem adható számot; *ratio* ugyanis 'arány'-t is jelent.)

A mérés, amely a pozitív valós számokhoz vezet, szintén absztrakciós lépésekből áll; attól is eltekintünk, mit mérünk (például valamely ruha-

anyag hosszúságát-e), de még attól is, hogy a gyakorlatban mindig abba kell hagynunk valamikor a mérést, akár marad még egy kicsi, akár az a ritka szerencsénk van, hogy nem marad semmi.

A természetes számok összessége is, a pozitív racionális számok összessége is, a pozitív valós számok összessége is példa olyan halmazokra, amelyeknek az elemei számok. Ezek a halmazok azonban abban különböznek például a nyáj juhai, az ábécé betűi, vagy valamely nyelv fonémái, vagy szavai halmazától, hogy elemeik száma már nem természetes szám. Ezt szakkifejezéssel úgy mondjuk, hogy a természetes számok halmaza, a pozitív racionális számok halmaza, a pozitív valós számok halmaza *végtelen halmazok*, míg a nyár juhainak halmaza, az ábécé betűinek halmaza, a nyelv fonémáinak halmaza *véges halmazok*. Véges halmazt alkotnak a nyelv szavai is, habár, mint Vértes Edit helyesen megállapította, és mint már említettük, sohasem tekinthető befejezett, lezárt összességnek, és így nem tudjuk megszámlálni, hány szó van a magyar nyelvben. Vértes Edit ezt szellemesen úgy fejezi ki, hogy a nyelv szavai véges, de megszámlálhatatlan halmazt alkotnak.[3] A *megszámlálhatatlan* szakkifejezést azonban itt nem abban az értelemben használja, mint a matematikus szokta használni.

A halmazelmélet ugyanis végtelen halmazok esetén is felveti azt a kérdést, hány elemből állnak, és választ is ad erre a kérdésre, csak éppen nem természetes szám a válasz, hanem úgynevezett *végtelen számosság*. A halmazelmélet szabatosan definiálja, mikor van két halmaznak ugyanannyi eleme, mikor van az egyiknek több, vagy kevesebb eleme, mint a másiknak. Kiderül, hogy a végtelen halmazok közül azoknak van legkevesebb elemük, amelyeknek ugyanannyi elemük van, mint a természetes számok halmazának. Ezeket nevezzük matematikai szakkifejezéssel *megszámlálhatóan végtelen halmazoknak*. Ezek közé tartozik például a pozitív racionális számok halmaza is. Ez azért meglepő, mert a természetes számok mind közte vannak a pozitív racionális számok között, s rajtuk kívül vannak még — nem is kevesen — más pozitív racionális számok is, tudniillik a törtszámok. Ezt úgy fejezi ki a matematikus, hogy a természetes számok halmaza *valódi része*, *valódi részhalmaza* a pozitív racionális számok halmazának. (A matematikus az egészet is a részek közé sorolja; a *valódi* jelző arra utal, hogy nem az egész halmazról van szó.) A halmazelmélet azonban kideríti, hogy épp ez a meglepő tulajdonságuk különbözteti meg a végtelen halmazokat a végesektől: minden végtelen halmaznak van olyan valódi része, amelynek ugyanannyi eleme van, mint az egész halmaznak, ezzel szemben véges halmaz valódi részhalmazának mindig kevesebb eleme van, mint az egész halmaznak. *Megszámlálható halmaznak* a matematikus a véges halmazokat mind, és még a megszámlálhatóan végtelen halmazokat nevezi; eszerint matematikai értelemben a magyar nyelv szavainak halmaza is megszámlálható, mivel véges. *Megszámlálhatatlan halmazoknak* a matematikus azokat nevezi, amelyeknek több elemük van, mint a természetes számok halmazának. Ilyen megszám-

[3] Vértes Edit, Szótározott szavaink száma és a szókészlet nagysága: Magyar Nyelvőr 79. 329—38 (1955), különösen: 338.

lálhatatlan halmaz pl. a pozitív valós számok halmaza; ez a szabatos jelentése annak, hogy a pozitív racionális számok ritka kivételt alkotnak a pozitív valós számok között, amit megpróbáltunk előbb is plauzibilissé tenni. De még sok másféle megszámlálhatatlan halmaz is van: a halmazelmélet megmutatta, hogy a végtelen számosságok között még nagyobb a változatosság, mint a véges halmazok számosságai, vagyis a természetes számok között.

6. Azonban a végtelen halmazok elméletének eredményei nem fontosak a nyelvész számára, a matematika nyelvtudományi alkalmazásai szempontjából nyugodtan szorítkozhatunk véges halmazokra. Különben is a halmazelmélet csak a legutóbbi századfordulón keletkezett. Addig fel sem igen vetették a matematikusok azt a kérdést, hány elemből áll a természetes számok vagy a pozitív valós számok halmaza, vagy ha felvetették, megelégedtek azzal a negatív válasszal, hogy végtelen sok elemből áll.

Ehelyett a természetes számok halmazát, meg a többi számokból álló végtelen halmazokat, olyan rendszereknek tekintette a matematikus, amelyeknek érdemes a struktúráját vizsgálni. Ezekre a struktúrákra, mint már említettük, az a legjellemzőbb, hogy e rendszerek elemeit, a számokat, *műveletek* kapcsolják össze. Ha két természetes számot, például a 3-at és a 4-et, összeadunk, újabb természetes számot, esetünkben a 7-et kapjuk eredményül. Ezt úgy szokás kifejezni, hogy az összeadás mindig elvégezhető a természetes számok körében, vagy, hogy nem vezet ki a természetes számok halmazából. A szorzás is mindig elvégezhető a természetes számok körében, mert két természetes szám szorzata is megint természetes szám, például $3 \cdot 4 = 12$. A kivonás azonban kivezet a természetes számok halmazából, mert ha egy természetes számból egy másik természetes számot kivonunk, az eredmény csak akkor lesz megint természetes szám, ha nagyobb számból vontunk ki kisebbet (pl. $4-3=1$, $7-4=3$), ellenkező esetben negatív egész számot kapunk (pl. $3-4=-1$, $4-7=-3$), illetőleg ha egy számból önmagát vonjuk ki, akkor 0-t kapunk eredményül. Ha azonban a természetes számokhoz hozzávesszük a *negatív egész számokat* is (és a 0-t is, amennyiben nem számítottuk a természetes számok közé), akkor az *egész számok halmazához* jutunk, amelyek körében nemcsak az összeadás és a szorzás, hanem a kivonás is mindig elvégezhető. Az osztás viszont kivezet az egész számok halmazából is éppúgy, mint a természetes számok halmazából, mert ha egy egész (vagy egy természetes) számot, mondjuk a-t, elosztunk egy másikkal, mondjuk b-vel, az eredmény nem mindig egész (tehát eo ipso nem mindig természetes) szám. Ha mégis az, akkor azt mondjuk, hogy az a szám osztható a b számmal. Például 30 osztható 6-tal, mert $30 : 6 = 5$ egész szám, de 7-tel nem, mert $30 : 7$ nem egész szám.

A pozitív racionális számok körében nemcsak az összeadás és a szorzás, hanem az osztás is mindig elvégezhető, mert két pozitív racionális számnak, akár egész, akár törtszámok, az összege is, a szorzata is, a hányadosa is megint pozitív racionális szám, például $\frac{3}{4}+\frac{5}{6}=\frac{3\cdot 3}{4\cdot 3}+\frac{5\cdot 2}{6\cdot 2}=\frac{9}{12}+$

$+\frac{10}{12}=\frac{19}{12}$, $\frac{3}{4}\cdot\frac{5}{6}=\frac{3\cdot5}{4\cdot6}=\frac{3\cdot5}{4\cdot3\cdot2}=\frac{5}{4\cdot2}=\frac{5}{8}$, $\frac{3}{4}:\frac{5}{6}=\frac{3}{4}\cdot\frac{6}{5}=$
$=\frac{3\cdot6}{4\cdot5}=\frac{3\cdot2\cdot3}{2\cdot2\cdot5}=\frac{3\cdot3}{2\cdot5}=\frac{9}{10}$. A kivonás azonban kivezet a pozitív racionális számok halmazából, mert két pozitív racionális szám különbsége nem mindig pozitív racionális szám, hanem megeshetik, hogy negatív racionális (egész vagy tört) szám, pl. $\frac{3}{4}-\frac{5}{6}=\frac{3\cdot3}{4\cdot3}-\frac{5\cdot2}{6\cdot2}=\frac{9}{12}-$
$-\frac{10}{12}=-\frac{1}{12}$. Az is megeshetik, hogy 0 a különbség, ha tudniillik egy pozitív racionális számot önmagából vonunk ki. De ha a pozitív racionális számokhoz hozzávesszük a negatív racionális számokat is és a 0-t is, a *racionális számok* halmazát kapjuk, amelyek körében nemcsak az összeadás és a szorzás végezhető el mindig, hanem a kivonás is; továbbá az osztás is mindig elvégezhető a 0-sal való osztás kivételével.

Amint e példából látható, nem igaz az, hogy a matematika csupa ,,szabályos'' rendszerrel dolgozik, hogy a matematikában nincsenek kivételek. A matematikában igenis vannak kivételek, ezek a matematikában abban az értelemben erősítik a szabályt, hogy csak akkor lesz ,,erőssé'', szabatossá a szabály, ha megmondjuk, hogy milyen kivételekkel érvényes; ha a kivételeket elhallgatjuk, akkor valójában helytelen szabályt mondunk ki. Az is megesik, hogy végtelen sok kivétel van, mint ahogy a pozitív racionális számok kivételt alkotnak az alól a szabály alól, hogy egy pozitív valós szám, vagyis egy hosszúság mértékszáma általában irracionális szám. Szóval nem áll az, hogy a matematikát nem lehet a nyelvészeti kérdésekre alkalmazni, mert a matematika csak szabályos rendszerekkel foglalkozik, a nyelvben meg sok a kivétel. Igenis, a matematikusnak megvan a maga módszere arra, hogy a kivételeket hogyan vegye számításba, mert bizony számításba is kell vennie.

A pozitív valós számok körében mindig elvégezhető az összeadás, a szorzás és az osztás, de a kivonás nem mindig (ismét, mint a természetes, vagy a pozitív racionális számok esetén is, csak akkor, ha nagyobb számból vonunk ki kisebbet). Ha azonban a pozitív valós számokhoz hozzávesszük a 0-t és a *negatív valós számokat* is, a *valós számok halmazát* kapjuk, amelyek körében szintén mindig elvégezhető a négy alapművelet: az összeadás, a kivonás, a szorzás és az osztás, az utóbbi megint a 0-sal való osztás kivételével.

Azt is láttuk az eddigiekből, hogy a matematika által vizsgált, számokból álló rendszerek szerkezetére nemcsak az jellemző, hogy elemeiket műveletek kapcsolják össze, hanem az is, hogy elemeik között különböző *relációk* állhatnak fenn. Például a ,,nagyobb'', ,,kisebb'', ,,osztható'' szavak ilyen relációkat jelentenek. A matematika e szavak helyett bizonyos jeleket használ e relációk jelölésére; például azt, hogy 7 nagyobb, mint 4, így jelöljük: $7 > 4$, azt, hogy 3 kisebb, mint 4, így jelöljük: $3 < 4$, azt, hogy 30 osztható 6-tal, vagy más szóval, 6 osztója 30-nak, így jelöljük: $6 \mid 30$. Az egyenlőség jele ($=$) is egy relációt jelöl, az egyenlőség-relációt.

De azon kívül, hogy milyen műveletek nem vezetnek ki a számok különböző halmazaiból, és hogy milyen relációk állnak fenn e halmazok bizonyos elemei között, még az is jellemző e halmazoknak mint rendszereknek szerkezetére, hogy e műveleteknek és relációknak milyen *tulajdonságaik* vannak. E tulajdonságok között vannak nagyon egyszerűek, mint pl. az összeadás és a szorzás *kommutatív tulajdonsága*, vagyis az, hogy mindegy, hogy az egyik számhoz adom-e a másikat, vagy a másikhoz az egyiket, pl. $3 + 4$ is 7, $4 + 3$ is 7, és mindegy, hogy az egyik számot szorzom-e meg a másikkal, vagy a másikat az egyikkel, pl. $3 \cdot 4$ is 12, $4 \cdot 3$ is 12. (A kivonás és az osztás nem kommutatív műveletek pl. $3-4$ az -1, viszont $4-3$ az nem -1, hanem 1.) E tulajdonságokat az algebra formanyelve az

$$a + b = b + a,$$
$$a \cdot b = b \cdot a$$

tömör alakban mondja ki. De vannak sokkal bonyolultabb tulajdonságai is a számokon végzett alapműveleteknek. Hogy ezek közül csak egyet említsek, nem is túlságosan bonyolultat, olyat, amivel a nyelvész kartársak is találkoztak már, tekintsük például azt, amelyet az

$$(a + b)^2 = a^2 + 2ab + b^2$$

formula mond ki. Ezt a tulajdonságot szavakban úgy lehetne kimondani, hogy ha két számot össze kellene adnunk, és ami kijön, azt négyzetre emelnünk, vagyis önmagával megszoroznunk, akkor ezt az összetett műveletet nemcsak úgy végezhetjük el, hogy tényleg összeadjuk a két számot és összegüket négyzetre emeljük, hanem akkor is ugyanarra az eredményre jutunk, ha ehelyett az egyik számot megszorozzuk önmagával, a két számot megszorozzuk egymással, majd az eredmény kétszeresét vesszük, s a másik számot is megszorozzuk önmagával, végül e három szorzás eredményét összeadjuk. Például ha 10-hez 7-et kell hozzáadnunk és az eredményt önmagával megszoroznunk, ezt nemcsak így végezhetjük el:

$$10 + 7 = 17,$$
$$17 \cdot 17 = 289,$$

hanem ugyanazt az eredményt, 289-et kapjuk, ha ehelyett a következő műveleteket végezzük:

$$10 \cdot 10 = 100,$$
$$2 \cdot 10 \cdot 7 = 2 \cdot 70 = 140,$$
$$7 \cdot 7 = 49,$$
$$100 + 140 + 49 = 289.$$

Hasonlóan, az algebra minden úgynevezett *azonossága* azt mondja ki, hogy a számok között olyan összefüggés van, hogy ha egy bizonyos műveletsorozatot végzünk rajtuk, akkor mindig ugyanaz a végeredmény jön ki, mint ha egy bizonyos másik műveletsorozatot végzünk ugyanazokon a számokon. (Nem pedig azt, hogy az egyik műveletsorozat helyett a másikat „kell" elvégeznünk, pl. hogy két szám összegét úgy kell négyzetre emelnünk, hogy az egyik számot négyzetre emeljük, a két szám szorzatát megszorozzuk 2-vel, s a másik számot is négyzetre emeljük, végül a kapott három eredményt összeadjuk. Dehogy kell így! Úgy is lehet, hogy a két számot összeadjuk s az összeadás eredményét négyzetre emeljük. A gyakorlatban az dönti el, hogy melyik módon csináljuk, hogy hogyan egyszerűbb, vagy hogy mi a célunk az összetett művelet elvégzésével.) Minden ilyen azonosság újabb részletét fedi fel a számokból álló rendszerek bonyolult szerkezetének.

Hasonló áll a számok közötti relációk tulajdonságaira is: ezek is hozzájárulnak a számokból álló rendszerek szerkezetének felfedéséhez. Például az egyenlőség-relációnak a következő tulajdonságai vannak:

1. $$a = a,$$

vagyis bármely szám egyenlő önmagával. Ezt úgy szokás mondani, hogy az egyenlőség *reflexív* reláció. Reflexív reláció az oszthatóság is (pl. a pozitív egész számok körében), mert bármely pozitív egész szám osztható önmagával ($a \mid a$), hiszen az osztás eredménye, 1, természetes szám. Ezzel szemben pl. a „kisebb" szóval kifejezett reláció (röviden: a „kisebb" reláció) nem reflexív, nem áll az, hogy bármely szám kisebb önmagánál, sőt az áll, hogy egy szám sem kisebb önmagánál, amit úgy mondunk, hogy a „kisebb" reláció *irreflexív*.

2. $$\text{Ha } a = b, \text{ akkor } b = a,$$

vagyis ha egy szám egyenlő egy másikkal, akkor e második szám is egyenlő az elsővel. Ezt úgy szokás mondani, hogy az egyenlőség *szimmetrikus* reláció. Az oszthatóság és a „kisebb" reláció már nem szimmetrikusak: ha $a \mid b$, akkor nem biztos, hogy $b \mid a$ (pl. $2 \mid 4$, de 4 nem osztója 2-nek, mert $4 : 2$ egész szám, de $2 : 4 = \frac{1}{2}$ nem az), és ha $a < b$, akkor nem biztos, hogy $b < a$, sőt, akkor biztos, hogy $b < a$ nem áll:

$$\text{ha } a < b, \text{ akkor } b < a \text{ nem áll,}$$

amit úgy mondunk, hogy a „kisebb" reláció *aszimmetrikus*. Az oszthatóság (bár, mint mondtuk, nem szimmetrikus, de) nem aszimmetrikus reláció: nem áll, hogy ha $a \mid b$, akkor $b \mid a$ nem áll, mert pl. $2 \mid -2$ (ti. $-2 : 2 = -1$ egész szám) és mégis $-2 \mid 2$ (ti. $2 : -2 = -1$ megint csak egész szám).

3. $$\text{Ha } a = b \text{ és } b = c, \text{ akkor } a = c,$$

vagyis ha egy szám egyenlő egy másikkal és e másik egyenlő egy harmadikkal, akkor az első szám is egyenlő a harmadikkal. Ezt úgy szokás mondani, hogy az egyenlőség *tranzitív* reláció. (Mintegy átmegy a-ról b-n keresztül c-re.) Tranzitív reláció az oszthatóság is (ha $a \mid b$ és $b \mid c$, akkor $a \mid c$), meg a „kisebb" reláció is (ha $a < b$ és $b < c$, akkor $a < c$).

A „kisebb" relációnak említett három tulajdonságán

1. $a < a$ nem áll

(irreflexív tulajdonság),

2. ha $a < b$, akkor $b < a$ nem áll

(aszimmetrikus tulajdonság),

3. ha $a < b$, és $b < c$, akkor $a < c$,

(tranzitív tulajdonság) kívül még egy nevezetes tulajdonsága van:

4. ha a és b két különböző szám, akkor vagy $a < b$, vagy $b < a$,

vagy más szóval

4. vagy $a = b$, vagy $a < b$, vagy $b < a$,

amit a „kisebb" reláció *trichotom* tulajdonságának nevezünk. Az oszthatóság nem trichotom reláció, megeshetik, hogy sem $a = b$, sem $a \mid b$, sem $b \mid a$ nem áll, pl. sem $5 = 7$, sem $5 \mid 7$, sem $7 \mid 5$ nem áll, mert sem $5 : 7 = \frac{5}{7}$, sem $7 : 5 = \frac{7}{5}$ nem egész szám.

7. Azonban nemcsak számokkal, hanem másféle dolgokkal is lehet műveleteket végezni, és nemcsak számok, hanem másféle dolgok között is állhatnak fenn relációk. Ezt a nyelvésznek nem kell sokat magyarázni. Például szavakon, pontosabban szóalakokon végzett művelet az összetétel. Ezt a műveletet nem lehet mindig elvégezni a szóalakok körében, mert nem lehet bármely két szóalakot összetenni. Más példa nyelvtani műveletre az egyeztetés (nemben, számban és esetben); ezt pl. a latin nyelvben egy melléknéven és egy főnévalakon lehet elvégezni, a művelet eredménye a melléknévnek a főnévalakkal egyeztetett alakja. Nyelvtani művelet pl. a birtokos esetbe tevés, ez egytagú művelet, mert egy szón, mégpedig pl. bármely főnéven lehet elvégezni, a művelet eredménye a főnév birtokos esete. Más példák egytagú nyelvtani műveletekre valamely szó tövének vagy szótári alakjának képzése.

Relációra példa szavak között a szófajrokonság, vagyis az ugyanazon szófajhoz tartozás; ez, mint könnyen látható, reflexív, szimmetrikus, tranzitív reláció.

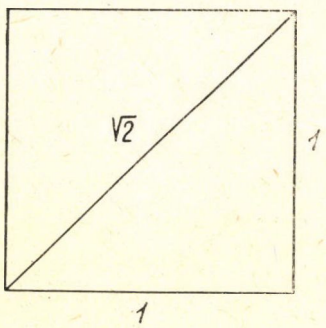
De a matematikus — a legutóbbi matematikai nyelvészeti vizsgálatokig — nem a szavakon, szóalakokon és egyéb nyelvi objektumokon végzett műve-

letek, nem a nyelvi objektumok közötti relációk vizsgálatával folytatta a számokon végzett műveletek és a számok közötti relációk tulajdonságainak kutatását. Hanem előbb olyan objektumokon, elemeken végzett műveletek tulajdonságainak vizsgálatán fejlesztette tovább az algebrai módszereket, amelyeknek valami kapcsolatuk volt a számokkal, és az ilyen elemek közötti relációkat vizsgálta.

Így már az ókori görögök, valószínűleg már az időszámításunk előtti ötödik században működött pythagoreusok, megkülönböztetettek a természetes számok között páros (vagyis 2-vel osztható) és páratlan (vagyis 2-vel nem osztható) számokat. Például 2, 4, 6 páros számok, 1, 3, 5 páratlan számok. (Egyébként a pythagoreusok számon még kizárólag természetes számot értettek, sőt, még az 1-et sem sorolták a számok közé.) Azt is tudták, hogy

$$\text{páros} + \text{páros} = \text{páros},$$

olyan értelemben, hogy bármely páros számhoz bármely páros számot hozzáadva megint páros számot kapunk (pl. $4 + 6 = 10$ megint páros szám). Meg azt is, hogy



1. ábra

$$\text{páros} + \text{páratlan} = \text{páratlan}$$

(vagyis bármely páros számhoz bármely páratlan számot hozzáadva páratlan szám lesz az eredmény, pl. $6 + 3 = 9$ páratlan szám), továbbá, hogy hasonló értelemben

$$\text{páratlan} + \text{páros} = \text{páratlan}$$

és

$$\text{páratlan} + \text{páratlan} = \text{páros}.$$

Lám, a „páros"-sal és a „páratlan"-nal is lehet műveletet, összeadást végezni, pedig ezek nem számok, hanem számok tulajdonságai. De azt is felismerték az ókori görögök, hogy szorzást is lehet ezeken a tulajdonságokon végezni, tudnillik, hogy hasonló értelemben

$$\text{páros} \cdot \text{páros} = \text{páros},$$

$$\text{páros} \cdot \text{páratlan} = \text{páros},$$

$$\text{páratlan} \cdot \text{páros} = \text{páros},$$

$$\text{páratlan} \cdot \text{páratlan} = \text{páratlan}.$$

Az utóbbi tényt arra is felhasználták, hogy bebizonyítsák: ha egy (természetes) szám négyzete páros, akkor az a szám maga is páros: tudniillik, ha páratlan volna, akkor négyzete, vagyis önmagával való szorzata, két páratlan szám szorzata, tehát páratlan szám volna, nem pedig páros. Ennek felhasználásával bizonyították azután be, hogy a hosszegységnek

nincs olyan törtrésze, amely pontosan egész számszor volna meg annak a négyzetnek az átlójában, amelynek oldala a hosszegység (l. az 1. ábrát), vagy, ahogy ma mondjuk, az átló hosszának mértékszáma (amelyről a Pythagoras-tétel alapján, vagy anélkül is, könnyen be lehet látni, hogy $\sqrt{2}$) irracionális szám.

A ,,páros''-sal és a ,,páratlan''-nal kivonást is lehet végezni:

$$\text{páros} - \text{páros} = \text{páros},$$

abban az értelemben, hogy ha tetszőleges páros számból kivonunk egy tetszőleges olyan páros számot, amelyet lehet úgy, hogy az eredmény természetes szám legyen (vagyis nálánál kisebb páros számot), akkor az eredmény páros; hasonló értelemben

$$\text{páros} - \text{páratlan} = \text{páratlan},$$

$$\text{páratlan} - \text{páros} = \text{páratlan},$$

$$\text{páratlan} - \text{páratlan} = \text{páros}.$$

Ha ezeket összehasonlítjuk a ,,páros'' és ,,páratlan'' összeadására vonatkozó eredményekkel, a ,,páros'' és ,,páratlan'' tulajdonságokból álló rendszer meglepő tulajdonságához jutunk: ebben a rendszerben a kivonás mindig ugyanarra az eredményre vezet, mint az összeadás:

$$\text{páros} - \text{páros} = \text{páros} + \text{páros} = \text{páros},$$

$$\text{páratlan} - \text{páros} = \text{páratlan} + \text{páros} = \text{páratlan},$$

$$\text{páros} - \text{páratlan} = \text{páros} + \text{páratlan} = \text{páratlan},$$

$$\text{páratlan} - \text{páratlan} = \text{páratlan} + \text{páratlan} = \text{páros}.$$

Az első két egyenlőség azt is mutatja, hogy ha ,,páros''-t akár ,,páros''-hoz, akár ,,páratlan''-hoz adunk hozzá, vagy vonunk ki belőle, ugyanazt kapjuk (a ,,páros'' és a ,,páratlan'' közül) eredményül, mint amihez hozzáadtuk, vagy mint amiből kivontuk. Ebből a szempontból tehát a ,,páros'' úgy viselkedik ebben a rendszerben, mint a 0 a számok között: az sem változtat azon a számon, amelyhez hozzáadjuk vagy amelyből kivonjuk ($a + 0 = a - 0 = a$). Ezt úgy szokás kifejezni, hogy a ,,páros'' és ,,páratlan'' tulajdonságokból álló rendszernek a ,,páros'' a 0-*eleme*.

Ezek után nem meglepő, hogy ,,páros''-sal nem lehet osztani, mint ahogy a számok körében a 0-val sem. Valóban, páros : páros lehet páros is, páratlan is (pl. $8 : 4 = 2$ páros, $12 : 4 = 3$ páratlan), hasonlóan, mint ahogy 0 a 0-ban akármilyen számszor megvan; viszont páratlan : páros sem páros, sem páratlan nem lehet, mert páratlan szám nem lehet osztható páros számmal, hasonlóan, mint ahogy például 1 : 0-nak egyáltalában nincs értelme. Viszont ,,páratlan''-nal lehet osztani:

$$\text{páros} : \text{páratlan} = \text{páros},$$

azaz, ha valamely páros szám osztható valamely páratlan számmal, akkor az eredmény mindig páros szám, pl. 12 : 3 = 4 páros szám, és hasonló értelemben

páratlan : páratlan = páratlan.

Így tehát a „páros" és „páratlan" tulajdonságokból álló rendszerben mind a négy alapműveletet el lehet végezni, a rendszer 0-elemével, a „páros"-sal való osztás kivételével. Igaz, itt a „kivétel" a két lehetséges eset felét jelenti, mégis meglepő, hogy ennek a rendszernek ilyen szerkezeti tulajdonsága van (mint pl. a számok közül a racionális, vagy a valós számok halmazának), nemcsak azért, mert e rendszer elemei a természetes (nem pedig a racionális, vagy a valós) számok bizonyos tulajdonságai, hanem azért is, mert e rendszer véges számú, tudniillik két elemből áll.

8. Az ókori görög matematikusok ezen felismerései a semmiféle gyakorlati alkalmazásra számot nem tartó *θεωρία*, szemlélődő gondolkodás eredményei voltak. Csak sokkal később, az újkor elején, találkozott a matematikus a matematika alkalmazásai által felvetett olyan kérdésekkel, amelyek szintén műveletek végzését kívánták számoktól különböző objektumokon. Ilyen kérdések elsősorban a mechanikával kapcsolatban léptek fel.

Az ókorban a mozgás még többnyire csak az ember mozgását jelentette. Például, amikor Zénón azt állította, hogy nincs mozgás, vitatkozó partnere, szinópéi Diogenész, úgy cáfolta meg ezt, hogy elkezdett sétálni. Az újkorban azonban a mozgás már a gépek alkatrészeinek mozgását is jelentette. A gazdaságos termelés szempontjából fontossá vált, hogy a gépeket a célnak megfelelően szerkesszék meg és használják. Ehhez pedig tudni kellett, hogyan mozognak a testek különböző erők hatására. A mozgás törvényeinek kutatása Galileivel kezdődött; Newton már pontosan megfogalmazta ezeket a törvényeket. Ezek alkalmazása szükségessé tette, hogy olyanféle mennyiségekkel is műveleteket végezzünk, mint a sebesség, a gyorsulás, az erő. Ha például egy testre egyidejűleg két erő hat, például ha innét is húzzák, onnét is, akkor ezek együttes hatása ugyanaz, mintha egy harmadik erő hatna helyettük; ezt a harmadik erőt nevezzük az egyidejűleg ható két erő *összegének* (vagy *eredőjének*). Ennek a harmadik erőnek azonban nem egyszerűen úgy kapjuk meg a nagyságát, hogy a két erő nagyságát összeadjuk; hanem az, hogy hogyan kapjuk meg, attól is függ, milyen volt a két erő iránya. Az erőnek a hatása (ami a mozgó test sebességének megváltoztatásában áll, amit úgy szokás mondani, hogy az erő gyorsulást ad a mozgó testnek) maga is függ az erő irányától, nemcsak a nagyságától. Eszerint az erőt nem lehet egyetlen mérőszámmal, a nagyságával megadni, hanem ehhez az erő irányának megadása is hozzátartozik. Az olyan mennyiségeket, amelyek megadásához a mennyiség nagyságán kívül irányát is meg kell adni, *vektoroknak* nevezzük. A sebesség, a gyorsulás is példa vektorokra.

Másféle mechanikai kérdések viszont a vektorok *szorzására* vezettek. Kétféle szorzásról szokás beszélni. Az egyik fajta szorzás kivezet a vektorok köréből, mert ha két vektoron végrehajtjuk, az eredmény nem vektor lesz, hanem szám. Ezt a fajta szorzást *skaláris szorzás*nak nevezzük;

ugyanis azokat a mennyiségeket, amelyeket egyetlen számadattal megadhatunk, *skalároknak* szokás nevezni. A vektorok összeadásának is, skaláris szorzásának is, megvan a kommutatív tulajdonsága. Ezzel szemben a vektorok másik fajta szorzásának, az úgynevezett *vektoriális szorzásnak*, amely nem vezet ki a vektorok köréből, mert ha két vektoron végrehajtjuk, eredménye megint vektor lesz, már nincs meg a kommutatív tulajdonsága. Ha ugyanis a két vektort, amelyeknek vektoriális szorzatát képezzük, felcseréljük, akkor vektoriális szorzatuknak ellenkezőre fordul az iránya, mindenesetre más vektor lesz belőle.

A vektorokat, ha nem is egy, de három számadat meghatározza: az egyik a vektor nagysága, a másik kettő, két szög mérőszáma, a vektor irányának meghatározására szolgál. Ezt a három számadatot célszerűbb szimmetrikusan választani, például úgy, hogy a vektort a tér három olyan egyenesére, úgynevezett *koordináta-tengelyekre* vetítjük, amelyek közül bármely kettő merőleges egymásra, és e vetületek, az úgynevezett *vektorkomponensek* nagyságát adjuk meg. Ez a három adat függ a koordináta-tengelyek választásától; a vektor fogalmához, azon kívül, hogy három számadat határozza meg, az is hozzátartozik, hogy hogyan függ tőlük.

A rugalmas testekben alakjuk megváltoztatásakor fellépő feszültségek leírása olyan mennyiségekhez vezetett, amelyeket már nem három, hanem kilenc számadat határoz meg és ezek az adatok is határozott módon függnek a koordináta-tengelyek választásától. A koordináta-tengelyek választásától ezen a módon függő kilenc számadattal meghatározható mennyiségeket *tenzoroknak* nevezzük. A rugalmas testek mechanikája szükségessé tette műveletek végzését tenzorokon is.

A vektort és a tenzort eleinte mechanikai fogalmaknak tekintették; később azonban kiderült, hogy a matematikusnak is jól használható segédeszközei, például a görbe vonalak és felületek geometriájában. Így e fogalmak lassanként a matematikában is polgárjogot nyertek. Sőt, kiderült, hogy bizonyos általánosításaik, az úgynevezett *többdimenziós vektorok*, a *többdimenziós tenzorok* és a *magasabbrendű tenzorok* is hasznos segédeszközök nemcsak geometriai, hanem algebrai vizsgálatokban is; ez arra vezetett, hogy a matematikus ezekkel is végzett műveleteket. Más algebrai kérdések, pl. a több ismeretlenes elsőfokú egyenletrendszerek vizsgálata, ismét másfajta olyan objektumokon végzett műveletekhez vezettek, amelyeket több számadat határoz meg. Ezek az objektumok a *mátrixok;* nevüket onnan kapták, hogy az őket meghatározó számadatok úgy sorakoznak egymás mellé és alá, mint a mátrikulában, az anyakönyvben a bejegyzések. Ezek, és végtelen sok dimenziós általánosításaik, amelyeket meghatározó számadatok már végtelen sok sorban sorakoznak egymás alá és minden sorban végtelen sok számadat sorakozik egymás mellé, később az atomfizika törvényei leírásának hathatós segédeszközeivé váltak.

Az a minőségi változás a matematika fejlődésében, amelyet már említettem, amikor a matematika alkalmassá vált a változás és az összefüggés leírására, tudniillik a matematikai analízis kialakulása, szintén ahhoz vezetett, hogy a matematikus újfajta objektumokkal kezdett műveleteket végezni. Ezeket az objektumokat *függvényeknek* nevezzük. Ezek sem mennyiségek, sőt, alapvetően különböznek a mennyiségektől. Minthogy e

tekintetben sok a félreértés, nem utolsósorban a középiskolai oktatás hibájából, erről valamivel részletesebben kell beszélnem. Téves definíció az, hogy függvény olyan mennyiség, amely más mennyiségtől vagy mennyiségektől függ, vagy hogy a függvény változó mennyiség, vagy hogy olyan változó mennyiség, amely más változó mennyiségtől vagy mennyiségektől függ. A valóságban a függvény nem mennyiség, hanem *függésmód:* az a mód, ahogy valamely mennyiség egy vagy több mennyiségtől függ. A függvény fogalmához erős absztrakció vezetett. A matematikus akkor jutott ehhez a fogalomhoz, amikor eltekintett attól, hogy mennyi az értéke annak, ami függ (egy vagy több) valamitől, attól is, hogy mennyi az értéke annak vagy azoknak, amiktől függ (ezektől változó mennyiségek esetén el kellett tekintenie, hiszen ezek az értékek a változás során mindig mások lettek), és csak azt nézte, hogyan függ tőlük. Egy függvény tehát egészen más dolog, mint egy szám, nem arra a kérdőszóra felel, hogy *hány, mennyi,* hanem arra, hogy *hogyan függ (valamitől* vagy *valamiktől).* Így, amikor a matematikus függvényeken végez műveleteket, ismét más objektumokon végzi ezeket, mint számokon.

Egyébként a halmazelmélet lényegesen általánosította a függvény fogalmát is: nem kívánja meg sem attól, ami mástól, vagy másoktól függ (az úgynevezett *függő változótól*), hogy mennyiség legyen, sem attól, vagy azoktól, amitől vagy amiktől függ (az úgynevezett *függő változóktól*), hogy mennyiségek legyenek, hanem olyan függvényeket is vizsgál, amelyeknél ezek tetszőleges dolgok, tetszőleges halmazelemek. Függvényen eszerint ma már bármiféle olyan módot értünk, ahogyan valami valamitől vagy valamiktől függ. Például az is egy függvény, ahogyan egy főnév birtokos esete magától a főnévtől függ.

A függvény és a művelet fogalma között szoros kapcsolat van: bármely műveletet meghatároz az a függvény, amely megadja, hogy a művelet eredménye hogyan függ azoktól az objektumoktól, amelyeken a műveletet végrehajtjuk (az ún. *operanduszoktól*). Így, minthogy nemcsak számokon, mennyiségeken lehet műveleteket végezni, hanem másféle objektumokon is, a műveletet ily módon meghatározó függvény független változói nemcsak számok, mennyiségek, hanem másféle objektumok is lehetnek; és minthogy a művelet eredménye nemcsak szám, mennyiség lehet, hanem másféle objektum is, a műveletet meghatározó függvény függő változója is tetszőleges objektum lehet.

A pythagoreusoknak a számok párosságára és páratlanságára, a ,,páros'' és ,,páratlan'' tulajdonságokon végezhető műveletekre vonatkozó megállapításai sem maradtak általánosítás nélkül. Páros számok, mint mondtuk, azok, amelyek 2-vel oszthatók; páratlanok viszont azok, amelyekből, ha 2-vel elosztjuk őket, 1 marad. (Ha páros számú ember áll sorba párosával, akkor mindegyik megtalálja a párját; ha páratlan számú ember, akkor egyikük pár nélkül marad.) A 2 helyett más egész számot is választhatunk *modulusnak*, pl. 3-at, 4-et, 5-öt; de akkor nem kétfélének (párosnak és páratlannak), hanem többfélének bizonyulnak az egész számok. Például a 3 modulus szempontjából háromféle egész szám van: olyan, amelyik maradék nélkül osztható 3-mal; olyan, amelyikből, ha elosztjuk 3-mal, 1 marad, és olyan, amelyikből 2 marad, ha elosztjuk 3-mal. Például 12

osztható 3-mal, 16-ból, ha elosztjuk 3-mal, ami 5-ször van meg benne, 1 marad; 20-ból pedig, amelyben 6-szor van meg a 3, 2 marad. Ezt úgy szokták mondani, hogy a 3-ra mint modulusra nézve háromféle *maradékosztály* van: a 3-mal osztható számok osztálya (vagyis halmaza), a 3-mal osztva 1 maradékot hagyó számok osztálya, és a 3-mal osztva 2 maradékot hagyó számok osztálya. A 4-re mint modulusra nézve négyféle, az 5-re mint modulusra nézve ötféle maradékosztály van. És e maradékosztályokkal is lehet műveleteket végezni: akármilyen modulus esetén össze lehet adni, ki lehet vonni, szorozni lehet őket, és ha a modulus úgynevezett prímszám, vagyis olyan szám, amely nem osztható más természetes számmal, mint 1-gyel és önmagával (mint pl. a 2, 3, 5), akkor a maradékosztályokból álló rendszer 0-elemének (a modulussal osztható számok osztályának) kivételével az osztást is mindig el lehet végezni a maradékosztályok körében.

A matematikus a számok oszthatóságának vizsgálata során sem maradt a természetes, vagy az egész számok körében, hanem kiterjesztette ezeket a vizsgálatokat a valós (és a komplex) számok közül bizonyosokra, az ún. *algebrai egész számokra*, amelyek sok szempontból hasonlóan viselkednek, mint a közönséges (vagy más néven racionális) egész számok. Az algebrai egész számokból álló bizonyos halmazokon belül végzett oszthatósági vizsgálatok során a modulus szerepét nem mindig algebrai egész szám vette át, hanem néha másféle, *ideálnak* nevezett objektum. És az oszthatósági viszonyok felderítése megkívánta, hogy ideálokkal, és az ideálmodulusokhoz tartozó maradékosztályokkal is végezzünk műveleteket.

Ismét másféle objektumokkal végzett műveletekre vezettek a magasabbfokú egyenletek megoldására vonatkozó vizsgálatok. A másodfokú egyenlet megoldását tanítják a középiskolában. Az $ax^2 + bx + c = 0$ másodfokú egyenlet megoldásait az

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

képlet szolgáltatja; ebből az is látszik, hogy a másodfokú egyenlet megoldásához a négy alapműveleten kívül elég gyökvonást végezni. Hasonló, csak sokkal bonyolultabb képlet van a harmadfokú és a negyedfokú egyenlet megoldására is; ezeket is meg lehet oldani a négy alapművelet és a (négyzet- és köb-)gyökvonás segítségével. De az ötöd- és magasabbfokú egyenletek közül már nem mindet lehet e műveletek segítségével megoldani, hanem megoldásuk még más műveleteket is igényel. Annak a kérdésnek vizsgálata, hogy melyeket lehet mégis, véges halmazok elemei *permutációin* végzett műveletek tulajdonságainak vizsgálatát kívánta. (Permutáció felcserélésmódot jelent, vagyis azt, hogy hogyan cseréljük fel egy véges halmaz elemeit.)

9. Ilyenféle és még más előzmények után, amelyek ismertetése nagyon messze vezetne, vagy pontosabban azok során, fokozatosan alakult ki az úgynevezett *absztrakt algebra*. Ez egy még további absztrakciótól kapta a nevét: attól, hogy a matematikus attól is eltekintett, vajon mennyiségeken, vagy másféle objektumokon végez-e műveleteket, egyáltalában

attól, hogy miféle objektumokon végzi azokat. Csak arra volt tekintettel, hogy milyen műveleteket lehet azon objektumok körében elvégezni, amelyeken a műveleteket végezzük, és hogy milyen tulajdonságaik vannak ezeknek a műveleteknek. Ezekkel jellemezhető azon objektumok rendszerének szerkezete, amelyeken a műveleteket végezzük.

Az absztrakt algebra eleinte speciális ilyen szerkezetek, a matematikában valóban használt szakkifejezéssel: *struktúrák* vizsgálatával kezdett foglalkozni. E struktúrákat furcsa nevekkel látta el: *test*, *gyűrű*, *csoport*, *ferde test*, *félgyűrű*, *félcsoport* stb. Ebben is megmutatkozott a matematikusnak az a szokása, amit a nyelvész bizonyára rossz szokásnak tart, hogy ha valamely új fogalmat alkot és annak nevet kell adnia, nem gondolkodik sokat, hogy jó szakkifejezést találjon rá, hanem az első „kezeügyébe" eső szót használja szakkifejezés gyanánt. E nevek tehát eredetileg bizonyos, tetszőleges elemekből álló rendszereknek valamilyen szempontból közös szerkezetét jelentik; azonban, jelentésváltozás folytán, szokták magukat a kérdéses szerkezetű rendszereket is *testeknek*, *gyűrűknek*, *csoportoknak* stb. (és általánosan *struktúráknak*) nevezni.

Például *test* olyan rendszert jelent (illetőleg az olyan rendszerek közös szerkezetét), amelynek a — tetszőleges természetű — elemeinek körében mindig el lehet végezni négy műveletet: az összeadást, kivonást, szorzást és osztást, az utóbbit a rendszer 0-elemével való osztás kivételével; e műveletek tetszőleges módon lehetnek definiálva, azzal a megszorítással, hogy meglegyenek mindazok a tulajdonságaik, amelyeket az algebrának a számok körében érvényes azonosságai fejeznek ki. (Az, hogy a testnek van 0-eleme, mégpedig csak egy, ebből a megszorításból, felhasználva azt is, hogy elemei körében a kivonás mindig elvégezhető, már következik.) Például testet alkotnak a racionális számok is, a valós számok is; de pl. bármely olyan modulushoz tartozó maradékosztályok is, amely prímszám, így többek között a „páros" és „páratlan" tulajdonságok is. Az utóbbiak példát szolgáltatnak véges számú elemből álló, röviden: véges testekre.

Egy testet tehát úgy lehet megadni, hogy megmondjuk, milyen elemekből álljon, és azt is, hogyan értelmezzük ezen elemek között az összeadást, a kivonást, a szorzást és az osztást. Az így megadott rendszerről persze még meg kell mutatni, hogy elemeire érvényesek az algebra összes azonosságai. Ez utóbbit szerencsére megkönnyíti az a tény, hogy elég megmutatni, hogy az alább felsorolandó azonosságok, az úgynevezett *testaxiómák* érvényesek, mert ebből már következik, hogy mindazok az azonosságok érvényesek, amelyek a számok körében teljesülnek. A testaxiómák a következők:

(1) $$a + b = b + a$$

(az összeadás *kommutatív* tulajdonsága),

(2) $$(a + b) + c = a + (b + c)$$

(az összeadás *asszociatív* tulajdonsága),

(3) $$ab = ba$$

(a szorzás kommutatív tulajdonsága),

$$(ab)c = a(bc) \tag{4}$$

(a szorzás asszociatív tulajdonsága),

$$(a + b)\, c = ac + bc \tag{5}$$

(a szorzás *disztributív* tulajdonsága az összeadásra nézve),

$$a + (b - a) = b \tag{6}$$

(a kivonás *definíciós azonossága*), végül

$$a \cdot (b : a) = b \tag{7}$$

(ezen eset kivételével, amikor a a test 0-eleme; (7) az osztás definíciós azonossága). Ha tehát valamely tetszőleges természetű elemekből álló rendszer elemeire érvényesek az (1)—(7) azonosságok, akkor pl. az

$$(a + b)^2 = a^2 + 2ab + b^2$$

azonosság is érvényes. (Itt $2ab$ $ab + ab$-t jelent, mert nem biztos, hogy a 2 szám eleme a rendszernek, tehát, hogy lehet vele a rendszer elemein végzett szorzás definíciója értelmében szorozni; és pl. a^2 $a \cdot a$-t jelent.) Sőt, ehhez a (6) és (7) azonosságok érvényessége nem is kell, még az sem, hogy a kivonás és szorzás elvégezhető legyen a rendszer elemei körében, elég az összeadás és a szorzás elvégezhetősége és az (1)—(5) azonosságok érvénye. (Sőt, a (4) azonosság érvénye sem kell.)

A (6) és (7) azonosságokat azért nevezzük a kivonás, illetőleg az osztás definíciós azonosságának, mert módot adnak arra, hogy segítségükkel tetszőleges test elemei kivonását és osztását definiáljuk az összeadás, illetőleg a szorzás segítségével: $b - a$ a testnek az az eleme, amelyet a-hoz hozzáadva b-t kapjuk; $b : a$ pedig a testnek az az eleme, amellyel a-t megszorozva b-t kapjuk. Ahhoz azonban, hogy ezeknek a definícióknak legyen értelmük, ki kell kötnünk, hogy mindig van ilyen elem (kivéve a második definícióban azt az esetet, amikor a a test 0-eleme; ekkor ugyanis vagy nincs olyan elem, amellyel a-t megszorozva b-t kapjuk, vagy, ha ti. b is a test 0-eleme, minden elem ilyen); ebből már és a többi testaxiómából következik, hogy csak egy ilyen elem van. Eszerint a test fogalmát így is definiálhatjuk: testen tetszőleges természetű elemek olyan rendszerét értjük, amelyek körében két műveletet: az összeadást és a szorzást mindig el lehet végezni; e műveletek tetszőleges módon lehetnek definiálva azzal a megszorítással, hogy meglegyenek az (1)—(5) axiómákkal kifejezett tulajdonságaik, továbbá, hogy bármely a és b elemhez legyen olyan elem, amelyet a-hoz adva b-t kapjuk (ezt röviden úgy mondjuk: az összeadás *invertálható* legyen), továbbá, bármely, a rendszer 0-elemétől (amelynek létezése az

összeadás invertálhatóságából következik) különböző a elemhez és bármely b elemhez legyen olyan elem, amellyel a-t megszorozva b-t kapjuk (a szorzás invertálhatósága).

Mint mondtuk, az egész számok körében mindig elvégezhető az összeadás, a kivonás és a szorzás, de az osztás nem mindig, vagyis a szorzás már nem invertálható. Az ilyen szerkezetű rendszert (vagy az ilyen szerkezetű rendszerek közös szerkezetét) *kommutatív gyűrűnek* nevezzük. A kommutatív gyűrű pontos definíciója: olyan, tetszőleges elemekből álló rendszer, amelyek körében két műveletet: az összeadást és a szorzást mindig el lehet végezni; e műveletek tetszőleges módon lehetnek definiálva, azzal a megszorítással, hogy teljesítsék az (1)—(5) axiómákat és hogy az összeadás invertálható legyen. E kikötéseknek következménye, hogy (ha a kivonást ismét úgy definiáljuk, hogy $b - a$ a kommutatív gyűrűnek az az eleme, amelyet a-hoz adva b-t kapjuk, röviden, ha a kivonáson az összeadás *inverz* műveletét értjük) a kommutatív gyűrű elemeire érvényesek mindazok, az egész számok körében érvényes azonosságok, amelyekben csak összeadás, kivonás, szorzás szerepel, osztás nem.

Azt is mondtuk, hogy a matematikus találkozott olyan objektumokkal is, amelyek körében definiált, szorzásnak nevezett műveletnek nincs meg a kommutatív tulajdonsága. (Az asszociatív és a disztributív tulajdonsága azonban megvan; ez készteti a matematikust arra, hogy e műveletet mégis szorzásnak nevezze.) Az olyan rendszereknek, amelyeknek az elemei között ilyen nem-kommutatív szorzás van definiálva, már egészen más szerkezetük van, mint a testeknek és a kommutatív gyűrűknek, így többek között az $(a + b)^2 = a^2 + 2ab + b^2$ azonosság sem mindig érvényes az elemeikre. Ilyen, más szerkezetű (nem-kommutatív szorzással „felszerelt") rendszerek a *ferde testek* és a *gyűrűk*; ezeknek a definíciója annyiban különbözik a testek, illetőleg a kommutatív gyűrűk definíciójától, hogy az (1)—(5) axiómák helyett csak az (1), (2), (4) és (5) axiómák teljesülését kívánjuk meg, emellett az összeadás invertálhatóságát is, továbbá a ferde testek esetén az osztás invertálhatóságát abban az alakban, hogy ha a a rendszer olyan eleme, amely nem a rendszer 0-eleme, akkor a rendszer bármely b eleméhez van olyan x eleme is, hogy $ax = b$, meg olyan y eleme is, hogy $ya = b$. Ez az x és y elem általában különböző: a ferde testekben kétféle inverz művelete is van a szorzásnak, hasonlóan, mint ahogy a hatványozásnak is kétféle inverz művelete van a valós számok körében: a gyökvonás és a logaritmálás. (Ennek oka éppen az, hogy a hatványozásnak nincs meg a kommutatív tulajdonsága, hasonlóan, mint ahogy a ferde testekben a szorzásnak sincs.)

A természetes számok körében az összeadás sem invertálható: a kivonás sem végezhető el mindig. Az olyan rendszereket, amelyeknek a (tetszőleges természetű) elemei körében két műveletet, az összeadást és a szorzást mindig el lehet végezni és e két művelet úgy van definiálva, hogy az (1)—(5) axiómák érvényesek rájuk (egyébként tetszőlegesen), *kommutatív félgyűrűnek* nevezzük; ha pedig csak az (1), (2), (4) és (5) axiómák érvényét kívánjuk meg, *félgyűrűnek*.

A matematikus olyan rendszerekkel is találkozott a matematika fejlődése során, amelyek körében csak egy művelet definiálható természetes

módon; így pl. permutációk szorzásáról beszélhetünk, de összeadásukról nem szokás beszélni. Ez a körülmény megint másféle struktúrák vizsgálatára vezetett. Ilyenek pl. a *csoportok* és a *félcsoportok*. Félcsoportnak nevezzük az olyan rendszert (vagy az ilyen rendszerek közös szerkezetét), amelynek a (tetszőleges természetű) elemei körében egyféle műveletet, amelyet rendszerint szorzásnak szokás nevezni (de néha összeadásnak is nevezik), mindig el lehet végezni, s e műveletnek megvan az asszociatív tulajdonsága (egyébként tetszőlegesen lehet definiálva). Ha még emellett megkívánjuk a csoport-művelet invertálhatóságát is hasonló értelemben, mint a ferde testek esetén, de anélkül a kikötés nélkül, hogy a ne legyen a rendszer 0-eleme (0-elemről nem is lehet beszélni, ha összeadás nincs definiálva), akkor a rendszert (illetőleg szerkezetét) csoportnak nevezzük; ha emellett még a kommutatív tulajdonságát is megkívánjuk, akkor *kommutatív csoportról*, vagy (Abel norvég matematikusról) *Abel-féle csoportról* beszélünk.

Ismét másfajta struktúrák vizsgálatára vezetett az a körülmény, hogy, mint már említettük, a számokon nemcsak műveleteket lehet végezni, hanem bizonyos relációk is fennállnak közöttük, amelyeknek szintén megvannak bizonyos tulajdonságaik, és hogy a matematikus másféle elemből álló rendszerekkel is találkozott, amelyeknek az elemei között hasonló tulajdonságú relációk állnak fenn. Ilyenek az úgynevezett *elrendezett struktúrák :* az *elrendezett testek*, *elrendezett gyűrűk* stb. Például elrendezett testen olyan rendszert értünk, amelynek az elemei körében kétféle művelet: az összeadás és a szorzás, továbbá egy reláció: a „kisebb" reláció, definiálva van, még pedig oly módon, hogy egyrészt teljesüljenek rájuk az (1)—(5) axiómák, másrészt az összeadás és a szorzás invertálható legyen, harmadrészt a „kisebb" relációnak (amelyet a „$<$" jellel jelölünk) meglegyen az irreflexív, az aszimmetrikus, a tranzitív és a trichotom tulajdonsága (lásd a 23. lapon), továbbá még a következő két, az összeadással és a szorzással kapcsolatos, ún. *monoton* tulajdonsága:

1. ha $a < b$, akkor bármely c elemre $a + c < b + c$;
2. ha $a < b$ és $0 < c$ (ahol 0 a rendszer 0-eleme), akkor $ac < bc$.

Egyébként azonban az elrendezett test elemei tetszőleges természetűek lehetnek és a rajtuk végzett összeadás és szorzás definíciója, valamint a „kisebb" reláció definíciója is tetszőleges lehet. (A „kisebb" reláció definíciója abban áll, hogy megmondjuk, milyen esetben áll fenn és milyen esetben nem áll fenn az $a < b$ reláció.)

10. Csak mutatóba soroltam fel néhány, az absztrakt algebra által részletes szerkezeti elemzés alá vont speciális struktúrafajta definícióját. E struktúrafajták esetén a rendszer elemei már nem szükségképpen számok, hanem tetszőleges természetű elemek lehetnek; a rajtuk végzett műveletek és a közöttük fennálló relációk definíciója is tetszőleges lehet, csak azt kívánjuk meg, hogy bizonyos axiómáknak eleget tegyenek. Ezeken az axiómákon azonban még meglátszik, hogy a matematikus a számokból álló rendszerek szerkezeti vizsgálatából indult ki eredetileg. Ugyanis ezek

az axiómák mind olyan tulajdonságait fejezik ki az összeadásnak és a szorzásnak, illetőleg a „kisebb" relációnak, amelyek a számok összeadásának és szorzásának, illetőleg a számok közötti „kisebb" relációnak is tulajdonságai. Épp ezért nevezzük ezeket a műveleteket összeadásnak és szorzásnak, ezt a relációt „kisebb" relációnak, és ezért jelöljük őket ugyanazokkal a jelekkel, mint a számok esetén. Ugyanis a számokon végzett műveletekkel és a számok között fennálló relációkkal közös tulajdonságok lehetővé teszik, hogy a struktúra elemeit a számok általánosításának, a rajtuk végzett műveleteket a számokon végzett műveletek általánosításának, a közöttük fennálló relációkat pedig a számok között fennálló relációk általánosításának tekintsük. Márpedig a matematikusnak az a szokása, hogy ha egy fogalmat általánosít, az így keletkezett új fogalmat, legalábbis egy darabig, gyakran ugyanazon a néven nevezi, és ugyanúgy jelöli, mint azt a fogalmat, amelynek általánosítása útján keletkezett.

A további fejlődés azonban az úgynevezett *univerzális algebrához*, a struktúrák és struktúrafajták általános elméletéhez vezetett, amely már nem tételezi fel, hogy a számokon végzett műveletekhez bármiben is hasonló műveletek, a számok közötti relációkhoz bármiben is hasonló relációk jellemezzék a struktúrát. A fejlődés e pontján tehát már nemcsak azok az elemek váltak tetszőlegessé, amelyeken műveleteket végzünk (ez már előbb, az absztrakt algebra létrejöttével, megtörtént), hanem azoknak a műveleteknek és relációknak a definíciója is, amelyeket ezeken az elemeken végzünk, illetőleg amelyeknek ez elemek közötti fennállását vagy fenn nem állását vizsgáljuk.

Az univerzális algebra *struktúrán* tetszőleges olyan rendszert ért, amelynek — tetszőleges természetű — elemei körében, tetszőleges módon, műveletek és relációk vannak definiálva. E műveletek és relációk száma is tetszőleges lehet (esetleg 0 is: megeshetik, hogy csak műveletek vannak definiálva, relációk nem, vagy az is, hogy csak relációk vannak definiálva, műveletek nem; ha egyik sincs definiálva, a struktúra fogalma a halmaz fogalmába megy át), meg az is tetszőleges lehet, hogy egy-egy művelet, vagy reláció hánytagú. Tehát nemcsak egytagú és kéttagú, hanem háromtagú (három elemen végzett), négytagú stb. műveletekről is lehet szó; hasonlóan, egytagú, kéttagú és többtagú (több elem közötti) relációkról is. (A matematikus egytagú reláción tulajdonságot ért; ilyen pl. a valós számok körében valamely szám pozitív volta.) Persze az is lehet, hogy a relációk közül pl. az egyik egytagú, a másik két- vagy háromtagú stb.

A *struktúra* szakkifejezés már említett jelentésváltozása folytán eszerint a matematikus magát a rendszert érti struktúrán, vagyis egy halmazt, elemei körében definiált, mindig elvégezhető műveletekkel, valamint relációkkal együtt. Az ilyen rendszerek, „struktúrák" szerkezetét ennélfogva kénytelen más szakkifejezéssel megnevezni. E célra a matematikus a *struktúrafajta* szakkifejezést használja. *Struktúrafajtán* azon „struktúrák" közös szerkezetét érti, amelyek ugyanazon, a műveletekre és relációkra vonatkozó feltételeknek, úgynevezett *struktúra-axiómáknak* tesznek eleget. Eszerint a struktúra-axiómák minden egyes rendszere másmás struktúrafajtát jellemez. Ilyen struktúrafajták pl. a test, a ferde test, a gyűrű, a félgyűrű, a csoport, a félcsoport stb., de persze még sok más-

féle is van, olyan is, amelyet a „klasszikus" absztrakt algebra nem vont be vizsgálata körébe. (Az absztrakt algebrát szokták a ma már idejét múlt *modern algebra* néven is emlegetni a kizárólag csak számokból, vagy hasonló objektumokból álló rendszerek szerkezetét kutató *hagyományos algebrával* szemben; de ha ezt az elnevezést használnók, akkor az univerzális algebrát már a *„hipermodern algebra"* körébe kellene sorolnunk.)

Maguk a struktúra-axiómák szintén tetszőleges alakúak lehetnek. Kifejezhetik valamely struktúra definíciójában szereplő műveletek és relációk általános tulajdonságait, mint az azonosságok; kifejezhetik a struktúra bizonyos tulajdonságú elemének létezését, mint pl. valamely művelet invertálhatósága. Lehetnek feltételes ítéletek is, mint pl. a monoton tulajdonságokat kifejező axiómák, és még sok másféle alakúak is. Hogy milyen az általános alakjuk, azt csak a matematikai logika tisztázta teljesen.

Azonban a „hipermodern" algebra fejlődése az univerzális algebránál sem állt meg. Az univerzális algebra értelmében vett struktúrákat egyetlen halmaz és bizonyos számú, annak elemei körében értelmezett művelet és reláció határozza meg. Az algebra további fejlődése, részben belső szükségleteiből, részben a gyakorlati alkalmazások céljára, olyan általánosabb rendszerek szerkezeti vizsgálatához vezetett, amelyek meghatározásához már több halmaz megadása szükséges (e halmazok elemein végezhető műveletekkel és esetleg a közöttük fennálló relációkkal együtt). Így pl. az *automaták* legújabban keletkezett algebrai elmélete az automatákat olyan rendszereknek tekinti, amelyeket három halmaz és két művelet megadásával lehet meghatározni. Az első halmaz az automata összes lehetséges (úgynevezett *belső*) *állapotainak* halmaza; a második halmaz az automatának adható jelek (biomorf kifejezéssel: *ingerek*) halmaza, a harmadik pedig az automata által kiadható jelek *(válaszok)* halmaza. (Ez utóbbi két halmazt az automata *bemenő*, illetőleg *kimenő ábécéjének* szokás nevezni. Ez a szokatlan elnevezés onnan ered, hogy az automatának rendszerint nem egyetlen jelet szokás adni, hanem jelek egy egész sorozatát és azt figyeljük, milyen (kimenő) jelsorozattal válaszol rá. Az ilyen jelsorozatokat a bemenő, illetőleg a kimenő ábécé elemeiből, *betűiből* alkotott *szavaknak* szokás nevezni. Egyébként az *ábécé*, *betű*, *szó* szakkifejezések már előbb polgárjogot nyertek az absztrakt algebrában, pl. az úgynevezett *szabad félcsoportok* elméletében, valamint a matematikai logika egyik fejezetében, az *algoritmusok elméletében*, amely maga is a szabad félcsoportok elméletére épült.) Az automatát meghatározó két művelet közül az első, az úgynevezett *átmenet-függvény*, kéttagú, az automata belső állapotai halmazának és bemenő ábécéjének egy-egy tetszőleges a, illetőleg b elemén lehet elvégezni, a művelet eredménye pedig ismét az automata egy belső állapota, tudniillik az, amelybe az automata átmegy, ha akkor, amikor az a belső állapotában van, a b ingert adjuk neki. A második művelet, az úgynevezett *kimenő függvény*, szintén kéttagú, szintén az automata belső állapotai halmazának és bemenő ábécéjének egy-egy tetszőleges a, illetőleg b elemén lehet elvégezni, azonban a művelet eredménye az automata kimenő ábécéjének egy eleme, tudniillik az, amelyet az automata válaszként kiad, ha akkor, amikor az a belső állapotban van, a b ingert adjuk neki. (Furcsán hat talán, hogy e *műveleteket* átmenet-*függvénynek*, illetőleg

kimenő *függvénynek* nevezik; azonban gondoljunk arra a már említett kapcsolatra a művelet és a függvény fogalma között, amely abban áll, hogy bármely műveletet meg lehet adni azzal a függvénnyel, függésmóddal, ahogyan a művelet eredménye az operandusoktól függ.)

E példából nemcsak azt látjuk, hogy az algebra legújabb, akármennyire is absztrakt fogalmai és módszerei gyakorlati kérdések (pl. automaták összeállításának kérdése egyszerűbb automatákból) megoldására is alkalmazhatók, hanem azt is, hogy az algebra, fejlődése során, mindinkább közeledett olyanféle rendszerek szerkezeti elemzéséhez, mint amilyen a nyelv. Hiszen világos, hogy a nyelv olyan bonyolult rendszer, amelyet nem lehet egyetlen halmaz segítségével megadni. Így pl. Kulagina szovjet matematikai nyelvész is — aki pedig csak egy szempontból, a szófaj- és szintagmafaj-elmélet szempontjából vizsgálja a nyelvet — lényegében két halmaz és egy (egytagú) művelet segítségével jellemzi a nyelvet.[4] Az egyik halmaz, Ξ, a nyelv szókészlete, szavainak, pontosabban: *szóalakjainak halmaza;* a másik halmaz, Θ, a nyelv grammatikailag helyesen képezett (ahogy Kulagina mondja: kitüntetett) *mondatainak halmaza ;* végül a nyelvet a Kulagina-féle elméletben jellemző egytagú művelet olyan, hogy Ξ elemein, vagyis szóalakokon lehet elvégezni, s a művelet eredménye a megfelelő szó *összes alakjainak halmaza*, ahogy újabban mondani szokás, *paradigmája*. Ha x valamely szóalak (pl. *házat*), akkor a Γ művelet eredményét, ha x-en elvégezzük, Kulagina $\Gamma(x)$-szel jelöli és x *környezetének* nevezi; példánk esetében Γ (*házat*) a *ház* szó összes alakjainak (*ház, házat, háznak, házba, házban, házból, házra, házon, házról, . . . , házak, házakat, házaknak* stb.) halmaza (Kulagina a Ξ „szóalakkészlet", a Θ „mondathalmaz" és a Γ „környezetfogalom" által jellemzett nyelvet így jelöli: $\Xi(\Gamma, \Theta)$.) Kulagina elmélete eszerint a morfológiát a Γ művelettel adottnak tekinti, nem megy bele az elemzésébe. De adottnak tekinti a Θ halmazzal a szintaxist is, hiszen az határozza meg, melyek a (grammatikailag helyesen képezett) mondatok. (A továbbiakban mondaton mindig grammatikailag helyesen képezett mondatot értünk.) A fonetikába és fonológiába pedig bele sem megy, hiszen nem a beszélt, hanem csak az írott nyelvre gondol; de még azt sem veszi figyelembe elméletében, hogy az írott nyelvben a szavak betűk (véges) sorozatai, hasonlóan, mint ahogy a mondatok szavak véges sorozatai (ez utóbbit figyelembe veszi). Ugyancsak figyelmen kívül hagyja a jelentéstant, illetőleg azt, hogy a szavaknak is, a mondatoknak is jelentésük van. Elgondolhatjuk, hogy ha mindezeket és általában mindazt, ami a nyelvhez hozzátartozik, figyelembe vennők, a nyelv milyen bonyolult definíciójához jutnánk a Kulagina-féle definíció helyett. Ez világosan mutatja, hogy a nyelv sokkal bonyolultabb rendszer, mint azok a rendszerek, amelyek szerkezetét az algebra valaha is vizsgálta.

11. A nyelvnek az algebra által vizsgált struktúrafajtákkal és még bonyolultabb rendszerek szerkezetével való összehasonlítása azonban nem egészen

[4] O. Sz. Kulagina, Об одном способе определения грамматических понятий на базе теории множеств: Проблемы кибернетики 1. 203—214 (1958). különösen: 203.

korrekt, mert a nyelv nemcsak bonyolultabb, hanem még matematikai szempontból is más természetű rendszer, mint azok, amelyek szerkezetét az algebra vizsgálja. Az algebrai struktúrák esetén ugyanis kikötjük, hogy a struktúra-műveletek ne vezessenek ki a struktúra definíciójában szereplő halmazból (az univerzális algebra szakkifejezésével: a struktúra *tartójából*); de még pl. az automaták algebrai elméletében is kikötjük mindkét, az automatát jellemző műveletről, az átmenet-függvényről és a kimenő függvényről, hogy eredménye mindig az automatát jellemző halmazok egyikének, az automata belső állapotai halmazának, illetőleg a kimenő ábécéjének eleme legyen. Ezzel szemben már a Kulagina-féle elmélet nyelv-definíciójában is olyan Γ műveletről van szó, amelynek eredménye sem a Ξ, sem a Θ halmaznak nem eleme (sem szóalak, sem mondat), hanem a Ξ halmaznak részhalmaza (szóalakok halmaza). Ebből a szempontból a Kulagina-féle elmélet értelmében vett nyelv inkább a *topológiai struktúrákhoz* hasonlít, mint az algebraiakhoz. (Még egy lényeges eltérés adódik abból, hogy az algebrában a rendszert jellemző halmazok, ha egynél több van belőlük, mint pl. az automaták algebrai elméletében az automata belső állapotainak halmaza, bemenő ábécéje és kimenő ábécéje, egymástól függetlenül adhatók meg, csak a műveletek kapcsolják őket össze, ezzel szemben a Kulagina-féle értelemben vett nyelv esetén Ξ és Θ nem egymástól függetlenül megadható halmazok, hanem Θ-nak a Ξ elemeiből, tudniillik szóalakokból képezett összes lehetséges véges sorozatok halmaza részhalmazának kell lennie.) Ezért is hasznos lehet a nyelvész számára, ha a matematikusnak azokkal a tapasztalataival is megismerkedhetik, amelyeket az a topológiai rendszerek szerkezeti vizsgálata során szerzett.

A geometria fejlődését, amelyet megtett, míg a topológiához, sőt, az *általános topológiához* jutott, még kevésbé részletesen ismertetem, mint az algebra fejlődését. (Az algebra fejlődésének igazán részletes ismertetéséhez azoknak a módszereknek a vázolása is hozzátartoznék, amelyeket az algebra a különböző fajta struktúrák, vagy pl. az automaták szerkezeti vizsgálata során használ; ez azonban nagyon messze vezetne.)

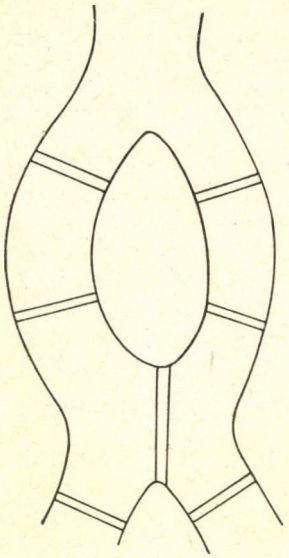
A geometria, mint tudjuk, *síkbeli* vagy *térbeli alakzatok*, *idomok* (például egyenesek, szögek, háromszögek, sokszögek, körök és egyéb görbe vonalak, síkok, síklapú és görbelapú testek, görbe felületek) vizsgálatából indult ki. Ezek voltaképpen pontokból álló halmazok, röviden: *ponthalmazok*. Csakhogy e ponthalmazok geometriai szerkezetére már nem az jellemző, ami az algebrai struktúrákra, tudniillik, hogy műveleteket lehet az elemeik körében végezni, hanem az, amit a *folytonosság* szakkifejezés — nem egészen szabatos — használatával lehet kifejezni. Ez abban áll, hogy egyik pont többé vagy kevésbé megközelíthet egy másik pontot, szűkebb vagy tágabb *környezetébe* eshetik.

Azt, hogy egy pont egy másik pontnak milyen szűk, vagy milyen tág környezetébe esik, a geometria eleinte a két pont távolságával mérte. A távolság már szám (mértékszám); ennélfogva a távolságokra mindazt lehet alkalmazni, amit az algebra a számokról megállapított. Így a mérés, mégpedig a távolságok mérésén kívül a szögmérés, a terület- és felszínmérés, a köbtartalommérés lehetővé teszi, hogy az algebrát a geometriai problémák megoldása során segédeszközként felhasználjuk. Ezt leginkább

az *analitikus geometria* más néven: *koordinátageometria* mutatja, amely a geometria problémáit algebrai problémákra vezeti vissza.

Azonban a matematikusok gyakran úgy érezték, hogy a méréssel, a geometriai viszonyok elemzésének számolásra való visszavezetésével elsikkad az igazi geometriai lényeg. Kerestek hát tiszta, nem mérő és nem számoló geometriai módszereket a geometriai idomok szerkezeti elemzésére.

Az ember hasonlít az árnyékára, annyira, hogy arról fel is lehet ismerni. Pedig az a „hasonlóság" nem hasonlóság olyan értelemben, mint ahogy két háromszög hasonlóságáról beszélünk. Nemcsak a méretei mások az árnyéknak, mint az ember megfelelő méretei, de az alakja is más. Mégis kell lennie valami közös geometriai tulajdonságuknak, különben nem hasonlítanának egymásra; de ez már nem a méretek megegyezésében, vagy, mint a hasonló háromszögek esetében, arányosságában megnyilvánuló hasonlóság. Még inkább így van ez valamely síkidom és árnyéka, vetülete között. Az olyanfajta geometriai tulajdonságok vizsgálatából, amelyek a síkidom és vetülete közös tulajdonságai, született meg a *projektív geometria*, amelyről azt gondolták, hogy a *metrikus geometriával* szemben ez az igazi tiszta geometria. Azonban kiderült, hogy vetítéskor változatlanul maradó tulajdonságokat is ki lehet fejezni a mértékszámok segítségével, így a projektív geometria csak annyiban maradt tiszta geometria, amennyiben ezt a tényt figyelmen kívül hagyta.



2. ábra

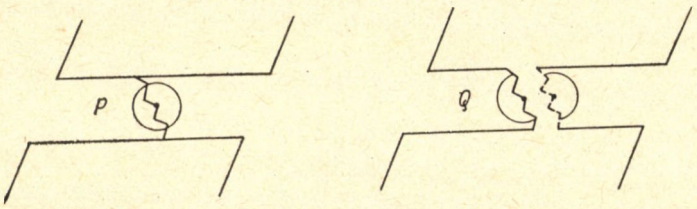
Ma a *topológiát* tartjuk a geometriai lényeget tükröző tiszta geometriának. Ez még általánosabb, a méréstől független geometriai tulajdonságok vizsgálatával foglalkozik. Olyanfélékkel, amelyek pl. a körnek közös tulajdonságai bármely zárt görbe vonallal, vagy az egyenes szakasznak bármely nyitott vonallal.

Topológiai kérdés pl. a königsbergi hidak híres problémája: át lehet-e menni Königsbergben (a mai Kalinyingrádban) a hét hídon, amelyek térképét az ábra mutatja, úgy, hogy mindegyik hídon átmenjünk, de mindegyiken csak egyszer.

A térképet fejből rajzoltam; bizonyára nem mutatja hűen a méreteket. De ez nem baj: a méretek nem lényegesek a felvetett probléma szempontjából. Képzeljük el, hogy a térkép szárazföldi részét és a hidakat tésztaszerű, gyúrható anyagból készítjük el, és hogy ezt a tészta-térképet deformáljuk. Eközben egyes távolságok, pl. a hidak hosszai, megnyúlhatnak, mások pedig megrövidülhetnek; egyenes vonalak, pl. a hidak középvonalai, elgörbülhetnek, görbe vonalak viszont kiegyenesedhetnek, vagy másképpen görbülhetnek. Az így deformált térkép még mindig hasonlít az eredetihez többek között annyiban, hogy ha az eredeti térképen be lehet

járni a hét hidat úgy, hogy minden hídon átmenjünk egyszer, de csak egyszer, akkor a deformált térképen is lehet s ha az eredetin nem lehet, akkor a deformált térképen sem lehet. Feltéve persze, hogy egyik hidat sem szakítjuk szét, és új összeköttetéseket sem létesítünk, például úgy, hogy a folyó két partjáról, a kis sziget felett, elkezdünk nyújtani egy-egy tésztacsikot egymás felé és ezeket összetapasztjuk úgy, hogy új híd keletkezzék.

Mit jelent ez a megszorítás pontosan? Képzeljük el, hogy az eredeti térképen megjelölünk egy pontot, amely pl. egy épületet, vagy az egyik hídfőt ábrázolja. A megfelelő pontot (amelyik ugyanazt az épületet vagy hídfőt ábrázolja), persze a deformált térképen is meg kell tudnunk találni. Sem eltűnnie nem szabad, sem szétválnia két vagy több ponttá, vagyis az eredeti térkép minden pontjának egy egyértelműen meghatározott *képs pontnak* kell megfelelnie a deformált térkép pontjai közül. Ezt úgy szoká-



3. ábra

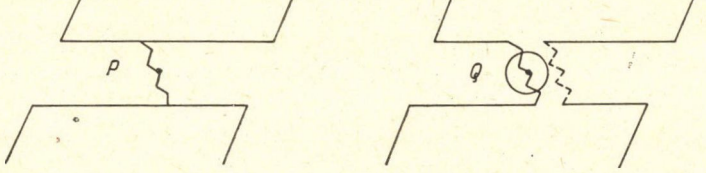
kifejezni, hogy a deformációnak az eredeti térkép pontjai halmazát *egyértelműen* kell leképeznie a deformált térkép pontjainak halmazára. De fordítva, új pontoknak sem szabad keletkezniük a deformált térképen, olyanoknak, amelyek nem felelnek meg az eredeti térkép egy pontjának sem, sem pedig az nincs megengedve, hogy az eredeti térkép különböző pontjainak a deformált térkép egybeeső pontjai feleljenek meg. Más szóval, az is kell, hogy a deformált térkép minden pontja is az eredeti térkép egy egyértelműen meghatározott pontjának legyen a képpontja. Ezt viszont, az előbbi követelménnyel együtt, amely a leképezés egyértelműségét kívánta meg, úgy szokás szaknyelven mondani, hogy a deformációnak az eredeti térkép pontjai halmazát *kölcsönösen egyértelműen* kell leképeznie a deformált térkép pontjainak halmazára.

Ez a megszorítás azonban csak látszólag zárja ki pl. a hidak szétszakítását, vagy új hidak létesítését. Igaz ugyan, hogy ha egy híd valamely, rajta keresztben rajzolt vonal mentén szétszakad, akkor e *szakadási vonal* pontjainak a szakadás útján keletkezett mindkét hídroncs szélén felel meg egy-egy pont (lásd az ábrát). Ha azonban csak az egyik, pl. a baloldali hídroncs szélét számítjuk hozzá a képpontok halmazához, a jobboldaliét nem, ezzel helyreállítottuk a leképezés kölcsönös egyértelműségét, de a szakadást nem szüntettük meg.

Miben mutatkozik meg a leképezésnek ez a *szakadásossága?* Tekintsük a szakadási vonal egy P pontját; ennek tehát, ha a jobboldali hídroncs szélét nem számítjuk hozzá a képpontok halmazához, csak a baloldali hídroncs szélén felel meg egy Q képpont. Ha azonban a P pont egy környezetét tekintjük (olyan kicsit, hogy egészen a hídon legyen), akkor az

ehhez a környezethez tartozó pontok képpontjai a deformált térképen csak részben esnek a Q képpont „félkörnyezetébe", részben a jobboldali hídroncs széléről elhagyott megfelelő pont félkörnyezetébe esnek. Ennélfogva, ha a Q képpontnak tekintjük olyan kis környezetét, amely nem ér el a jobboldali hídroncsig, akkor a P pontnak nem lehet olyan környezetét találni, hogy a hozzá tartozó pontok képpontjai mind a Q pont tekintett környezetébe essenek (mert egy részük mindenképpen a jobboldali hídroncsra esik). Ha tehát az eredeti térkép bármely P pontjáról és a deformált térképen neki megfelelő Q képponttól megkívánjuk azt, hogy a Q képpont bármely (még oly kis) V környezetéhez legyen az eredeti P pontnak olyan (tudniillik annyira kis) U környezete, hogy az U környezethez tartozó bármely pont képpontja a V környezethez tartozzék, akkor ezzel kizártuk a szakadás lehetőségét (nemcsak azt, hogy valamely hidat szétszakítsunk, hanem azt is, hogy a tésztát nyújtás közben valahol beszakítsuk, pl. a folyó valamelyik partján fekvő szárazföldet ábrázoló részén). Az olyan leképezést, amely ezt a feltételt teljesíti, *folytonosnak* nevezzük.

E feltétellel azonban csak a tészta elszakítását zártuk ki, új összeköttetések létesítését (tésztadarabok összeragasztása útján) nem. Azonban most már könnyen megtalálhatjuk azt a feltételt is, amely kizárja az összeragasztást. Ha ugyanis összeragasztás útján új összeköttetést létesítettünk és vissza akarjuk csinálni a deformációt, helyre akarjuk állítani az eredeti térképet, akkor ezt az összeköttetést ismét szét kell szakítanunk. Ennélfogva az új összeköttetés létesítését úgy zárhatjuk ki, ha megkívánjuk, hogy ez a „visszacsinálás", szakkifejezéssel: a deformáció által létesített leképezés *inverz leképezése* (vagyis az a leképezés, amely a deformált ábra minden egyes Q pontjának az eredeti ábrán azt a P pontot felelteti meg, amelynek a deformáció által létesített leképezés éppen a Q pontot feleltette meg képpont gyanánt), szintén folytonos legyen. Tehát a szakadást is, az új összeköttetések létezését is kizárjuk, ha megkívánjuk, hogy a deformáció olyan leképezést létesítsen, amely maga is, inverz leképezése is folytonos legyen, amit rövidebben úgy szokás mondani, hogy *mindkét irányban folytonos (bikontinuus)* leképezést létesítsen.

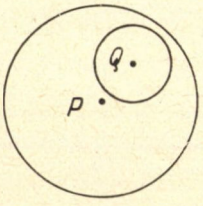


4. ábra

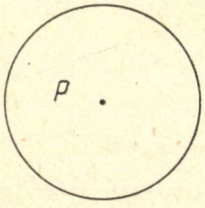
Eszerint, ha az eredeti térképen be lehet járni a hét hidat úgy, hogy minden hídon átmenjünk egyszer, de csak egyszer, akkor minden belőle kölcsönösen egyértelmű és mindkét irányban folytonos leképezéssel keletkező térképen is lehet, s ha az eredeti térképen nem lehet, akkor a belőle ilyen leképezéssel keletkező térképen sem lehet. A topológia általában a geometriai idomoknak azokat a tulajdonságait vizsgálja, amelyek a belő-

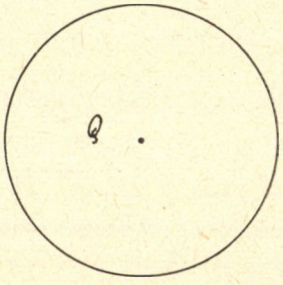
lük kölcsönösen egyértelmű és mindkét irányban folytonos leképezéssel keletkező idomokkal közös tulajdonságaik. Ezek az igazi, tiszta geometriai tulajdonságok.

12. Igen ám, de a leképezések folytonosságának definíciójához megint csak kellett a környezet fogalma, arról pedig még mindig nem mondtuk meg, mi az. Az ábrákon kis körlapnak rajzoltam a környezeteket, az azonban metrikus fogalom: azon pontok halmaza, amelyeknek *távolsága* a kör középpontjától (amelynek a környezetéről van szó) kisebb, mint a kör sugara. Azonban szerencsére éppúgy definiálhattuk volna a leképezések folytonosságát négyzet alakú, vagy téglalap alakú környezetekkel is. Általában csak az a fontos, hogy olyan ponthalmazokat értsünk a pontok



5. ábra





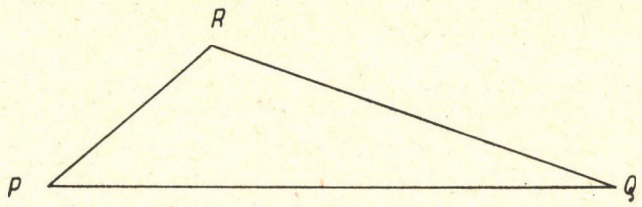
6. ábra

környezetén, amelyek eleget tesznek mindazoknak a feltételeknek, amelyekre a folytonos leképezések elméletének felépítéséhez szükség van. Ezeket a feltételeket Hausdorff német matematikus fogalmazta meg, azért *Hausdorff-féle axiómáknak* nevezzük azokat. A Hausdorff-féle axiómák a következők:

1. Minden pont eleme bármely környezetének.
2. Egy pont bármely két környezetéhez van olyan harmadik környezete, amely mindkettőnek része.
3. Ha egy Q pont eleme egy P pont valamely U környezetének, akkor Q-nak van olyan környezete, amely része U-nak. (Lásd az 5. ábrát, amelyen a környezeteket köröknek rajzoltam.)
4. Bármely két különböző P és Q pontnak van egy-egy olyan környezete, amelynek nincs közös eleme (lásd a 6. ábrát).

Az algebrához hasonló okokból a geometria sem állt meg az eredeti értelemben vett pontokból álló halmazok vizsgálatánál. Már mondtuk, hogy az analitikus geometria a geometria problémáit algebrai problémákra vezeti vissza. Ez úgy történik, hogy a pontokat síkbeli probléma esetén a koordinátáikból álló számpárokkal, térbeli probléma esetén pedig a koordinátáikból álló számhármasokkal pótoljuk. Sokszor azonban célszerű fordítva, az algebrai problémákat geometriailag szemléltetni, vagyis a valós számokból álló számpárokat, számhármasokat pontokkal pótolni. Ennek mintájára aztán később a valós számokból álló számnégyeseket, számötösöket stb. is kezdték *pontoknak*, az összes ilyen számnégyesek, számötösök

stb. halmazát pedig *térnek (négydimenziós, ötdimenziós* stb. *térnek)* nevezni és a geometriai fogalmakat az ilyen átvitt értelemben vett terekre is általánosítani. Később pedig olyan *függvényterek* elmélete is kialakult, amelynek elemei („pontjai") függvények. Így lassanként teljesen mellékessé vált, milyen elemekből áll a tér, vagy milyen elemekből állnak az egyéb, a geometria által vizsgált idomok, vagyis a geometria is tetszőleges természetű elemekből álló halmazokat kezdett vizsgálni. Csak az volt a fontos, hogy e halmazok elemeinek, amelyeket továbbra is *pontoknak* nevezünk, vagy *távolsága* legyen valahogyan definiálva (de úgy, hogy a távolságnak a közönséges térben meglevő tulajdonságai meglegyenek továbbra is), hogy a metrikus geometria fogalmait lehessen e pontokra is definiálni, vagy



7. ábra

környezeteik legyenek valahogyan definiálva (de úgy, hogy a Hausdorff-féle axiómák érvényesek legyenek rájuk), hogya topológia fogalmait ki lehessen terjeszteni e pontokra is. Így alakult ki a *metrikus tér*, illetőleg a *Hausdorff-féle tér* fogalma.

Metrikus téren olyan halmazt értünk, amelynek — tetszőleges természetű — elemei körében definiálva van egy távolságfogalom (két pont távolsága), mégpedig tetszőleges módon lehet definiálva azzal a megszorítással, hogy a következő tulajdonságai, amelyeket *távolság-axiómának* nevezünk, meglegyenek.

1. Bármely pont távolsága önmagától 0, bármely más ponttól pozitív szám.

2. Két pont távolsága nem változik, ha a két pontot felcseréljük.

3. Ha P, Q, R három tetszőleges pont, akkor a P és Q pontok távolsága nem lehet nagyobb, mint a P és R pontok távolságának, továbbá az R és Q pontok távolságának összege (lásd az ábrát, amelyből azt is megértjük, miért nevezzük ezt a tulajdonságot *háromszög-egyenlőtlenségnek*).

Itt a *távolság* olyan művelet, amelyet két ponton lehet elvégezni, de eredménye — eltérően az algebra műveleteitől — nem a halmaz (a metrikus tér) eleme, vagyis nem pont, hanem valós szám (még pedig pozitív, vagy 0).

Hausdorff-féle téren pedig olyan T halmazt értünk, amelynek — tetszőleges természetű — elemei körében definiálva van egy környezetfogalom úgy, hogy a környezetek a T részhalmazai legyenek; mégpedig tetszőleges módon lehet definiálva azzal a megszorítással, hogy meglegyenek a Hausdorff-féle axiómák által kifejezett tulajdonságai.

Milyen természetű fogalom a környezet fogalma? Minden környezet a T Hausdorff-féle tér részhalmaza; ha tehát pusztán abban állna a kör-

nyezet-fogalom, hogy a T halmaz minden pontjához meg legyen adva annak egy-egy környezete, akkor a környezet-fogalmat olyan egytagú műveletnek tekinthetnők, amelyet pontokon, vagyis a T halmaz elemein lehet elvégezni, de amelynek eredménye (a pont környezete) a T halmaznak nem eleme, hanem részhalmaza, hasonlóan, mint ahogy a Kulagina-féle elmélet nyelv-definíciójában is olyan Γ műveletről volt szó, amelynek eredménye, ha a szóalakok Ξ halmazának valamely elemén végezzük el, a Ξ halmaz részhalmaza. Ez a hasonlóság az oka annak, hogy Kulagina $\Gamma(x)$-et az x szóalak *környezetének* nevezi.

Azonban a környezet-fogalom azt jelenti, hogy a T halmaz minden pontjához meg van adva *összes környezeteinek halmaza*. Így a környezet-fogalom olyan egytagú művelet, amelyet pontokon, vagyis a T halmaz elemein lehet elvégezni, de amelynek eredménye egy olyan halmaz, amelynek elemei (tudniillik azon pont környezetei, amelyen a műveletet elvégeztük) a T halmaz részhalmazai. A topológia környezet-fogalma tehát bonyolultabb fogalom, mint a Kulagina-féle elméletben szereplő Γ művelet.

Vannak olyan topológiai vizsgálatok is, amelyekhez a 4. Hausdorff-féle axióma (az úgynevezett *szétválasztási axióma*) helyett enyhébb feltétel is elegendő, amely vizsgálatokat tehát a Hausdorff-féle tereknél általánosabb rendszerekben is el lehet végezni. Ha a 4. Hausdorff-féle axióma helyett azt kötjük ki, hogy bármely két különböző pont közül legalább *az egyiknek* legyen olyan környezete, amelynek a másik pont nem eleme, akkor az úgynevezett T_0*-tér* definíciójához jutunk, ha pedig azt kötjük ki, hogy bármely két különböző pont *mindegyikének* legyen olyan környezete, amelynek a másik pont nem eleme, akkor az ún. T_1*-tér* definíciójához jutunk (amely lényegében azonos a topologikus tér Riesz Frigyes által bevezetett fogalmával). A Hausdorff-féle tereket pedig szokás T_2-tereknek is nevezni.

A topologikus tereket jellemző axiómákat szokás a környezetfogalom helyett másképp, pl. a *nyitott*, vagy a *zárt ponthalmaz* fogalma segítségével is kifejezni. Bizonyára idegenszerűnek hatott, amikor azt mondtam, hogy a két hídroncs közül csak az egyiknek a szélét számítsuk hozzá a térképhez. Ennek az oka az, hogy amikor olyanféle valóságos alakzatokról beszélünk, mint pl. egy hídroncs, akkor ezekhez mindig hozzászámítjuk a „szélüknek" minden egyes pontját, szakkifejezéssel: minden egyes *határpontjukat*. A környezet fogalmával így lehet valamely ponthalmaz határpontját definiálni: a (topologikus) tér olyan pontja, amelynek bármely környezetében van a térnek olyan pontja is, amely a kérdéses ponthalmazhoz tartozik, meg olyan pontja is a térnek, amelyik nem tartozik a kérdéses ponthalmazhoz. Az olyan ponthalmazokat, amelyek minden határpontjukat tartalmazzák, zárt ponthalmaznak nevezzük; az olyan ponthalmazokat pedig, amelyek egy határpontjukat sem tartalmazzák, nyitott ponthalmaznak. Például ha egy körlaphoz az őt határoló körvonalat teljes egészében hozzászámítjuk, zárt ponthalmazhoz jutunk (a topológiai tér szerepét itt a sík játssza), ha pedig egy pontját sem számítjuk hozzá, nyitott ponthalmazhoz jutunk.

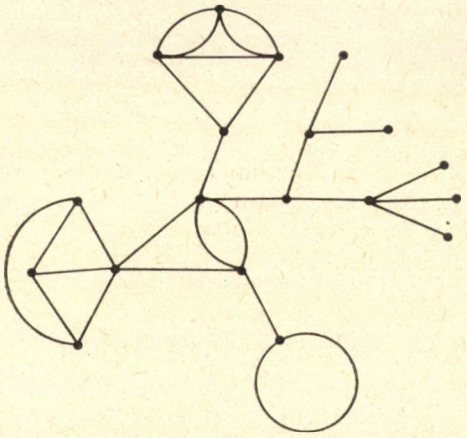
A topologikus tereket jellemző axiómáknak a környezetfogalom helyett a nyitott, vagy a zárt ponthalmaz fogalma segítségével való megfogalmazása

a térfogalom további általánosításaira vezetett. Ezeknek ismertetése messze vezetne, ezért attól el kell tekintenem. Ugyancsak el kell tekintenem attól is, hogy azokat a módszereket vázoljam, amelyeket a topológia a különböző fajta topológiai terek és azok részhalmazai szerkezetének vizsgálata során használ. Megemlítem azonban, hogy a topológiai terek között, éppúgy, mint az algebrai struktúrák között is, vannak olyanok, amelyek véges számú elemből (pontból) állnak. Ezek esetében a „folytonosság" csak nagyon erősen átvitt értelemben jelent folytonosságot. Mégis a topológiai terek közé számítják ezeket is, mert szerkezetük (és részhalmazaik szerkezete) felderítésére hasonló módszereket lehet alkalmazni, mint az „igazi" topologikus terek szerkezetének vizsgálatára, amelyek körében a „folytonosság" valóban folytonosságot jelent. A topológus persze az ilyen véges, úgynevezett *diszkrét* topológiai terek szerkezetére vonatkozó vizsgálatokat az igazi, vérbeli topológiai vizsgálatok melléktermékének tekinti; azonban nincs kizárva, hogy a topológiának ezek a megvetett mellékeredményei épp a nyelvtudományban találnak fontos alkalmazásra. Ezt valószínűvé teszi az, hogy a nyelv vizsgálatában szintén rendszerint véges halmazokról van szó, viszont a nyelv mint rendszer, szerkezeti szempontból inkább az ilyen véges topológiai terekhez hasonlít, mint az algebrai struktúrákhoz.

Egyébként egy és ugyanaz a halmaz el lehet látva, ahogy a matematikus mondja, fel lehet szerelve egyidejűleg algebrai, tehát a halmaz elemei körében elvégezhető műveletek és a halmaz elemei közötti relációk segítségével jellemezhető szerkezettel is, meg topológiai, tehát a halmaz elemei környezetei, vagy a halmaz nyitott, vagy zárt részhalmazai segítségével jellemezhető struktúrával is. Például a valós számok halmaza algebrai szempontból elrendezett test, amelynek elemei között el lehet végezni a négy alapműveletet (a 0-val való osztás kivételével) és amelynek elemei nagyság szerint el vannak rendezve; ugyanakkor topológiai szempontból (egydimenziós) topológiai térnek tekinthető, ha egy x valós szám környezetein az olyan *nyitott intervallumokat* értjük, amelyekben x benne van, vagyis nagyság szerint olyan a és b valós számok között fekvő valós számok halmazát, amelyekre $a < x$ és $x < b$ áll. A kétféle: algebrai és topológiai struktúra között persze szoros kapcsolat áll fenn; pl. ha minden valós számhoz ugyanazt a valós számot hozzáadjuk, vagy mindegyiket ugyanazzal a 0-tól különböző valós számmal megszorozzuk, a valós számokból álló topológiai térnek önmagára való kölcsönösen egyértelmű, mindkét irányban folytonos leképezéséhez jutunk. Ez azért lehet tanulságos a nyelvész számára, mert a nyelv is olyan rendszer, amely egyidejűleg többféle struktúrával fel van szerelve (legalább négyfélével: fonetikai-fonológiai, morfológiai, szintaktikus és jelentéstani szerkezettel), amelyek szintén szoros összefüggésben vannak egymással.

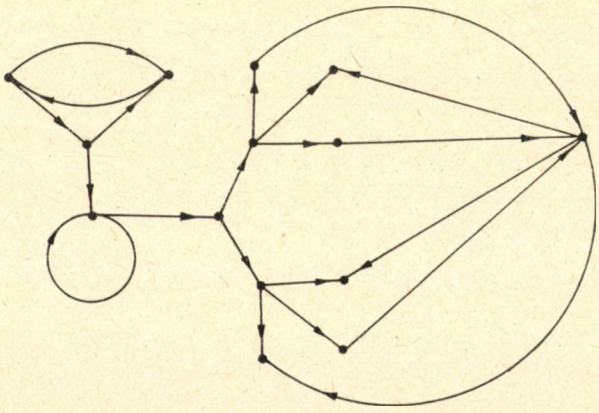
13. A topológiával kapcsolatban meg kell emlékeznem még az úgynevezett *gráfelméletről. Gráfnak* az olyan alakzatot nevezzük, amely úgy keletkezik, hogy pontokat vonalakkal összekötünk (lásd a 8. ábrát). A pontokat, amelyeket összekötöttünk, a gráf *szögpontjainak* nevezzük, a vonalakat, amelyekkel ezeket összekötöttük, a gráf *éleinek*. A gráf szögpontjait a szemlé-

letesség kedvéért a sík, vagy a tér pontjainak képzelhetjük, elvileg azonban lehetnek bármiféle átvitt értelemben vett tér, úgynevezett *absztrakt tér* pontjai is, véges számban, vagy végtelen sokan. A gráf éleit az egyik végpontjuktól a másik felé mutató (nyíllal jelölt) iránnyal is elláthatjuk, ekkor úgynevezett *irányított gráfot* kapunk (lásd a 9. ábrát). Két szögpont egynél több éllel is össze lehet kötve; egy szögpont önmagával is össze lehet kötve a gráf egy vagy több élével.



8. ábra

Gráfokat a matematika legkülönbözőbb ágaiban alkalmazunk bonyolult rendszerek elemei közötti kapcsolatok szemléltetésére, áttekinthetőbbé tételére. Ezt a matematikán kívül, más területen is megtehetjük. Például a magyar nyelvnek az Értelmező Szótárban tükröződő *jelentéstani szerkezetét* szemléltethetjük, ha az Értelmező Szótár minden egyes címszavának megfeleltetünk egy pontot, s mindazon szavaknak megfeleltetett pontokból, amelyeknek valamely alakja előfordul a kérdéses címszóhoz tartozó cikkben, élet irányítunk a kérdéses címszónak megfeleltetett pont felé. Az így keletkező irányított gráf olyan bonyolult, szövevényes lenne, hogy a valóságban meg sem tudnók rajzolni; nos, ilyen bonyolult, szövevényes a magyar nyelv jelentéstani struktúrája. E gráf *ciklusai*, vagyis ugyanazon irányban egymáshoz csatlakozó élei alkotta önmagába visszatérő pályái,



9. ábra

az Értelmező Szótár definícióiban fellelhető circulus vitiosusokat szemléltetik. Ezek nem hibái az Értelmező Szótárnak, hanem a nyelv jelentéstani szerkezetének természetéből folynak; ugyanis nagyon egyszerű gráfelméleti meggondolás mutatja (felhasználva a szavak halmazának véges voltát), hogy ilyen ciklusoknak szükségképpen lenniük kell. Annak vizsgálata, hogy milyen módon lehet a gráfot ciklus-nélkülivé tenni azáltal, hogy bizonyos szögpontjaiba befutó éleit elhagyjuk (könnyű látni, hogy ezt nagyon sokféle módon meg lehet tenni), arra a kérdésre ad választ, mely magyar szavak jelentését kell egy magyarul nem tudó embernek megtanulnia (természetesen a magyar nyelvre vonatkozó morfológiai, szintaktikus stb. ismeretek megszerzésén kívül), hogy az Értelmező Szótár segítségével bármely más magyar szó jelentését is megtalálhassa (ez összefügg a magyar nyelv *bazifikálásának* kérdésével). Persze „kézi erővel" reménytelen volna ilyen vizsgálatba belefogni, hiszen már a gráf megrajzolása is emberfeletti feladat volna; elektronikus számológép segítségével azonban nem okozna leküzdhetetlen nehézséget ilyen vizsgálat elvégzése.

Szándékosan választottam e példát a jelentéstan köréből; ugyanis egyúttal azt az elterjedt téves véleményt is meg szeretném cáfolni, hogy matematikai módszereket nem lehet szemantikus kérdésekre alkalmazni. Bizonyára onnan ered ez a vélemény, hogy ha pl. egy fordítási algoritmust valamely gép számára akarunk megfogalmazni, nem hivatkozhatunk jelentéstani fogalmakra, hiszen a gép nem érti meg a jelentését annak a szövegnek, amit egyik nyelvről a másikra fordít (a gép ugyanis semmit sem ért meg). Hasonló oka van annak, hogy a matematikus haladásnak tekinti azt (pl. a matematikai logikában), hogy egy szemantikus fogalmat sikerül morfológiai, szintaktikus és ezekhez hasonló *formális* (tehát a gép számára is felfoghatóvá tehető) fogalmakra visszavezetni.

A gráfelmélet a gráfok szerkezetére vonatkozó általános törvényszerűségeket vizsgálja; ezek felvilágosítást adhatnak a gráfokkal szemléltetett rendszerek szerkezetének felderítéséhez is. A gráfelmélet annyiban tartozik a topológiához, hogy a gráf szerkezete szempontjából mellékes, milyen távolságra rajzoltuk egymástól a pontjait és hogy milyen alakú vonalaknak rajzoltuk az éleit. Egyébként a gráf fogalma a legegyszerűbb speciális esete az ún. *kombinatórius* (vagy *algebrai*) *topológia* egy fontos fogalmának, a *komplexusnak*, amelyben a két-két pontot összekötő éleken kívül több pontot összekötő általánosításaik, az ún. *szimplexek* is szerepelhetnek. (A komplexusok olyan rendszerek vizsgálatára alkalmazhatók, amelyeknek több mint két eleme közötti kapcsolatok is fontosak.) Másrészt azonban a gráfelméletben alkalmazott módszerek eltérnek a topológia más fejezeteiben alkalmazott módszerektől, és inkább *kombinatorius jellegűek*, ezért a gráfelméletet szokás a matematika önálló, a topológiától független fejezetének is tekinteni. E tekintetben azonban nem egységes a matematikusok véleménye; azok, akik a gráfelméletet a topológiához számítják, arra hivatkoznak, hogy a kombinatórius topológia is használ kombinatórius módszereket, hiszen a nevét is innét kapta.

14. Különösen jellemző a kombinatórius módszerek alkalmazása (és nagyfokú továbbfejlesztése) a matematika egyik legfiatalabb, a századforduló

körül keletkezett ágára, a *halmazelméletre*, hiszen a kombinatórika tulajdonképpen a véges halmazok elmélete egyik fejezetének tekinthető.

A halmazelmélettől való idegenkedésben nagy része van annak a hiedelemnek — amelynek terjesztésében a matematikusok is hibásak —, hogy a halmazelmélet állítólag a matematika egyik legelvontabb fejezete. Valójában a halmazelmélet alapvető fogalma, a *halmaz* fogalma, konkrétabb, mint a számfogalom, hiszen már a természetes szám fogalma is úgy keletkezett, hogy eltekintettünk, milyen elemekből állnak a halmazok, csak arra voltunk figyelemmel, hány elemük van.

A halmazelmélet kialakulása jó példa a matematikai absztrakció természetének megvilágítására is. Sokan azért idegenkednek az absztrakciótól, különösen annak a matematikában használatos válfajától, vagy azért gondolják azt, hogy az absztrakció útján kapott matematikai fogalmakat nem helyes az olyan konkrét kérdésekkel foglalkozó tudományban alkalmazni, mint a nyelvtudomány, mert attól félnek, hogy az absztrakcióval végérvényesen, jóvátehetetlenül eltekintünk valamitől, ami valamilyen szempontból lényeges lehet. Valójában a matematikus, akármilyen merészen alkalmazza is az absztrakció módszerét, sohasem felejti el, mitől tekintett el az absztrakció során, s így azt sem, hogy az absztrakció eredményét csak bizonyos feltételek teljesülése esetén alkalmazhatja, amelyek annak biztosítása végett szükségesek, hogy attól a valamitől, amitől eltekintett, valóban el lehessen tekinteni annak a kérdésnek szempontjából, amit éppen vizsgál. Ha ezek a feltételek nem teljesülnek, egyszerűen „visszacsinálja" az absztrakciót. Így a matematikai absztrakció sohasem jóvátehetetlen, végérvényes lépés. Lehet, hogy évszázadok telnek el, míg a matematikus visszatalál absztrakciójától a konkrét (vagy konkrétebb) fogalmakhoz, mint ahogy csak a múlt század vége felé talált vissza a számoktól a halmazokhoz, de amikor aktuális lesz, visszatalál.

Bizonyos, hogy a halmazelméletnek vannak ismét nagyon absztrakt fejezetei, pl. a *végtelen számosságok elmélete* (amelyhez úgy jutott a halmazelmélet, hogy végtelen halmazokra vonatkozóan is felvetette azt a kérdést, hány elemük van, tehát megint csak eltekintett attól, mik az elemeik). Azonban a halmazelméletnek azok a fogalmai és módszerei, amelyek a nyelvész számára tanulságosak, nem tartoznak e nagyon absztrakt fejezetekbe.

A halmazelmélet a halmazok, főként pedig az olyan halmazok szerkezetének vizsgálatára, amelyeknek az elemei megint halmazok, vagyis a halmazok összességei szerkezetének vizsgálatára, többek között a halmazok közötti olyan kapcsolatokat is figyelembe veszi, amelyek abban állnak, hogy a halmazokon is lehet olyan *műveleteket* végezni, amelyeknek eredménye ismét halmaz. Ilyen művelet pl. a halmazok összeadása.

Az összeadásról, iskolai tanulmányaink során, megszoktuk, hogy olyan művelet, amelyet számokon végzünk. Valójában azonban már a természetes számok összeadása is absztrakció útján keletkezett a (véges) halmazok összeadásából. Amikor két juhász megegyezett abban, hogy együtt őrzik juhaikat, összeadták, egyesítették nyájukat, tehát magukat az őrizetükre bízott juhok halmazát, nem pedig juhaik számát adták össze. Persze valószínű, hogy a két nyáj egyesítése után megszámlálták, hány juhot őriznek most együtt. De a számok összeadására csak akkor jutott

az ember, amikor rájött, hogy a két (közös elem nélküli) halmaz egyesítése útján kapott új halmaz (pl. esetünkben az egyesített nyáj) elemeinek száma csak attól függ, hány eleme volt az egyesítés előtt a két halmaznak, attól nem, milyen elemekből álltak ezek a halmazok, s így e két halmaz elemeinek számán végzett művelet segítségével is meg lehet határozni. Ha olyan volna a világ, hogy egy négy almából álló halmaznak és egy három almából álló halmaznak egyesítése útján ugyan hét almából álló halmazt kapunk, de egy négy körtéből álló halmaz és egy három körtéből álló halmaz egyesítése útján már nyolc körtéből álló halmazt, akkor sohasem jutottunk volna arra a gondolatra, hogy a 4 és a 3 számokat adjuk össze, de a négy vagy három almából, körtéből álló halmazokat akkor is összeadhatnók.

Két halmaz *összege*, vagy *egyesítése*, azon elemek összessége, amelyek vagy az egyik, vagy a másik halmaznak elemei (vagy esetleg mind a kettőnek). Például az ebben a teremben levő nők halmazának, és az ebben a teremben levő férfiak halmazának összege az ebben a teremben levő emberek halmaza. Vagy az ebben a teremben levő olyan emberek halmazának, akiknek nyelvtudományból van valamilyen oklevelük, és az ebben a teremben levő olyan emberek halmazának, akiknek matematikából van oklevelük, az összege az ebben a teremben levő emberek halmaza, elhagyva belőle az olyan vendégeinket, akiknek sem nyelvtudományból, sem matematikából nincs oklevelük. Ebben az esetben a két halmaznak, amelyet összeadtunk, van néhány közös eleme: az a néhány itt jelenlevő kartárs, akinek nyelvtudományból is, meg matematikából is van oklevele. Ezek a közös elemek is csak éppoly elemei a két halmaz összegének, mint azok, amelyek az összeadandó halmazok közül csak az egyiknek elemei: halmaz „többszörös" eleméről nem szokás beszélni. Olyankor persze, amikor az összeadandó halmazoknak van közös elemük, nem kaphatjuk meg úgy az összeadás eredményeként adódó halmaz elemeinek a számát, hogy a két összeadandó halmaz elemeinek számát összeadjuk, hanem még ki kell vonni az elemeik számának összegéből a közös elemeik számát.

A halmazok összeadását — egyébként éppúgy, mint a számok, vagy az algebrai struktúrák másféle elemei összeadását is — ki lehet terjeszteni kettőnél több összeadandó esetére is. Több halmaz összege is azon elemek összessége, amelyek az összeadandó halmazok közül valamelyiknek, esetleg többnek is, elemei. Például valamely nyelv főnevei halmazának, melléknevei halmazának, névmásai halmazának és számnevei halmazának összege a kérdéses nyelv névszóinak halmaza; a nyelv magánhangzói halmazának, mássalhangzói halmazának és félhangzói halmazának összege a nyelv fonémáinak halmaza.

Ha egy halmaz más halmazok összege, azt szokás mondani, hogy e halmazt *felosztottuk* e „más halmazokra". Többnyire azt is meg szokás kívánni a felosztástól, hogy azon halmazok között, amelyekre felosztottuk a halmazt, ne legyenek olyanok, amelyeknek van közös elemük. Például a névszók halmazát felosztathatjuk a főnevek halmazára, a melléknevek halmazára, a névmások halmazára és a számnevek halmazára.

Másik példa halmazok közötti műveletre a halmazok *metszetének* képzése. Két sík metszetén (ha ugyan nem párhuzamosak) azt az egyenest

szokás érteni, amely a két sík közös pontjaiból áll. Hasonlóan, általában két (vagy több) halmaz metszetén e halmazok közös elemeinek összességét szokás érteni, vagyis azon elemek összességét, amelyek e halmazok mindegyikének elemei. Például a magyar nyelv mássalhangzóval kezdődő szavai halmazának és a magyar főnevek halmazának metszete a mássalhangzóval kezdődő magyar főnevek halmaza.

Szokás két halmaz kivonásáról is beszélni. Két halmaznak, A-nak és B-nek, *különbségén* azon elemek halmazát szokás érteni, amelyek az A halmaznak elemei, de a B halmaznak nem. Például ha a magyar főnevek halmazából kivonjuk a magánhangzóval kezdődő magyar szavak halmazát, a mássalhangzóval kezdődő magyar főnevek halmazát kapjuk a két halmaz különbsége gyanánt (ha a *j*-t is mássalhangzónak tekintjük). Mint ez a példa is mutatja, a B halmaznak nem kell feltétlenül az A halmaz részének lennie. Ha része, akkor az A és B halmazok különbségét szokás a B halmaz *komplementerének* is nevezni az A halmazra nézve.

A halmazok összességei szerkezetének vizsgálata szempontjából nemcsak a halmazokon végezhető műveletek fontosak, hanem a halmazok között fennálló *relációk* is. Ilyen reláció két halmaz között pl. az, amely abban áll, hogy az egyik halmaz része a másiknak, vagy az, amely abban áll, hogy az egyik halmazt kölcsönösen egyértelműen le lehet képezni a másikra. Véges halmazok esetén ez utóbbi reláció akkor és csak akkor áll fenn két halmaz között, ha ugyanannyi elemük van; épp e tény ad módot arra, hogy a halmaz elemei számának fogalmát végtelen halmazokra is kiterjesszük.

Természetesen a halmazokon végzett műveletek és a halmazok közötti relációk *tulajdonságai* is fontosak a halmazok összességei szerkezetének vizsgálata szempontjából. A halmazok összeadásának vannak a számok összeadásától eltérő tulajdonságai is, pl. az, hogy ha egy halmazt önmagához hozzáadjuk, visszakapjuk magát a halmazt (míg egy számnak önmagával való összege e szám kétszerese). Emiatt a halmazok körében másféle műveleti jeleket szokás használni, mint a számok körében, pl. az A és B halmazok összegét így jelöljük: $A \cup B$ (bár szokásos az $A + B$ jelölés is), az A és B halmazok metszetét pedig így: $A \cap B$ (bár szokásos, a metszetnek a számok szorzatához hasonló tulajdonságai miatt, az AB jelölés is). A $\cup$ jel kiejtése: kupa (angolul: *cup*), a $\cap$ jelé: kapu (az angol szaknyelvben: *cap*).

A halmazelmélet tehát új fajta struktúrákhoz vezetett s már ezzel is gazdagította az algebrát, amely általánosabb vizsgálat alá vette az ilyen struktúrafajtákat. Sokkal fontosabb azonban ennél, hogy a halmazelmélet új *módszerekkel* termékenyítette meg az algebrát és a topológiát, sőt, a matematika többi ágait is. A modern matematika lépten-nyomon használ halmazelméleti fogalmakat és módszereket, egy lépést sem tud ezek nélkül tenni. Ma már szinte el sem tudjuk képzelni, hogyan tudott a matematika évezredekig meglenni halmazelméleti fogalmak nélkül. Valójában addig sem volt meg nélkülük, csak nem ismerte fel a közöst e fogalmakban és így a matematika minden fejezetében másképp nevezte őket, illetőleg ott fellépő speciális esetüket; és a halmazelmélet ezekből, a matematika különböző fejezeteiben másképp és másképp nevezett fogalmakból absztrahálta

a maga fogalmait. Az univerzális algebra és az általános topológia pedig természetesen csak a halmazelmélet létrejötte után alakult ki.

Annyi bizonyos, hogy halmazelméleti fogalmak nélkül, vagy azok helyett mindig más és más elnevezésű, sokszor speciálisabb fogalmakat használva, sokkal nehézkesebben fejezte ki a matematika a maga mondanivalóit, mint ma. Én nem akartam e nehézkes kifejezésmóddal is megnehezíteni azt, hogy a nyelvész kartársak megértsék ezt az előadást; ezért pl. a halmaz fogalmát kezdettől fogva használtam, olyan matematikai fogalmak megvilágításához is, amelyek jóval azelőtt keletkeztek, hogy a halmazelmélet tudatossá tette a halmaz fogalmát.

Sokszor viszont *rendszerről* beszéltem halmaz helyett, anélkül, hogy pontosan megmondtam volna, mit értek rendszeren. Ezt olyankor tettem, amikor egy (vagy több) halmazon kívül még másvalamiről is volt szó, pl. a halmaz elemein elvégezhető műveletekről, vagy a halmaz elemei környezetének fogalmáról stb. Halmazelméleti szabatossággal rendszer helyett a halmazból és a műveletekből, vagy a halmazból és a környezetfogalomból stb. álló *halmaz*-ról kellett volna beszélnem, vagy még szabatosabban az ezekből álló *rendezett halmaz*-ról, minthogy pl. a műveletek megadásának sorrendje is fontos, ti. ez határozza meg, hogy pl. melyik játssza az összeadás, melyik a szorzás szerepét. (A rendezett halmaz fogalmához az is hozzátartozik, milyen sorrendben tekintjük az elemeit. Ezt azzal az úgynevezett *rendezési relációval* adhatjuk meg, amely akkor áll fenn a halmaz két eleme között, ha az egyik megelőzi a másikat. Ez a rendezési reláció tetszőlegesen lehet megadva, azzal a megszorítással, hogy irreflexív, aszimmetrikus, tranzitív és trichotom reláció legyen. Természetesen más-más ilyen tulajdonságú rendezési relációhoz más-más rendezett halmaz tartozik. Ezeknek a rendezett halmazoknak ugyanazok az elemeik, csak más-más sorrendben tekintjük őket. Jó példa rendezett halmazra a szótárban levő szavak halmaza, ábécé-sorrendben, más néven lexikografikus sorrendben tekintve őket. Ebben az esetben a rendezési reláció definíciója a következő. Két különböző szó közül az jön előbb, amelynek az ábécében előbb jön az a betűje, amely eltér a másik szó (elölről számítva) ugyanannyiadik betűjétől; ha pedig a rövidebbik szó mindegyik betűje megegyezik a hosszabbik szó ugyanannyiadik betűjével, akkor a rövidebbik szó jön előbb.)

Így pl. a test definíciója halmazelméleti szabatossággal a következőképpen szólna. Testen értünk olyan rendezett hármast (vagyis három elemből álló rendezett halmazt), amelynek első eleme, K, egy tetszőleges természetű elemekből álló halmaz, második és harmadik eleme pedig két, e halmaz elemei körében mindig elvégezhető kéttagú művelet, amelyeket a $+$ és a $\cdot$ műveleti jellel jelölünk, s amelyek tetszőleges módon lehetnek definiálva azzal a megszorítással, hogy meglegyenek a testaxiómákkal kifejezett tulajdonságaik (beleértve e műveletek invertálhatóságát, a második művelet, tehát a rendezett hármas harmadik eleme esetén a már többször említett kivétellel, amely a 0-elemmel való osztásnak felel meg). Vagy pl. az automata halmazelméleti szabatosságú algebrai definíciója: olyan rendezett ötös (vagyis öt elemből álló rendezett halmaz), amelynek első három eleme, A, X és Y, három halmaz, negyedik és ötödik eleme két

olyan δ és λ kéttagú művelet, amely az A és az X halmaz egy-egy tetszőleges elemén elvégezhető s eredménye a δ művelet esetén az A halmaznak, a λ művelet esetén pedig az Y halmaznak eleme. Vagy pl. a Hausdorff-féle tér hasonló stílusban megfogalmazott definíciója: olyan rendezett pár (vagyis két elemből álló rendezett halmaz), amelynek első eleme tetszőleges T halmaz, második eleme pedig egy, a T halmaz elemei körében definiált olyan környezetfogalom (vagyis olyan, a T halmaz elemein elvégezhető egytagú művelet, amelynek eredménye mindig egy olyan halmaz, amelynek elemei a T halmaz részhalmazai), amely eleget tesz a Hausdorff-féle axiómáknak. (E definíciókat nem a nyelvész olvasó elrettentése végett, hanem azért fogalmaztam meg halmazelméleti szabatossággal, hogy ha majd hasonló stílusban megfogalmazott definícióba ütközik valamely matematikai nyelvészeti dolgozatban, ne tűnjék neki az túlságosan idegenszerűnek.)

A halmazelmélet nyújtotta új, szabatosabb kifejezésmódnál — amit elvégre olyan mondanivalók kifejezésére használ a matematikus, amelyeket addig is ki tudott valahogyan fejezni, ha valamivel pongyolábban is, de sokszor érthetőbben — sokkal fontosabbak azok az új módszerek, amelyekkel a halmazelmélet gazdagította a matematika különböző fejezeteit. E módszerek ismertetése azonban nem lehetséges ennek az előadásnak keretében; mindössze a legegyszerűbb ilyen módszerek közül ismertetek néhány olyat, amelyet nyelvészeti példákkal is lehet illusztrálni.

Az egyik ilyen halmazelméleti módszer valamely halmaznak felosztására vonatkozik páronként közös elem nélküli (úgy is szokták mondani, hogy páronként *egymástól idegen*) részhalmazaira. Felosztással már a hagyományos formális logika is foglalkozott. A formális logika tanítása szerint a felosztáshoz mindig előbb meg kell adni a felosztás *szempontját* (más néven: *alapját*). Például a nyelv szavainak főnevekre, melléknevekre, igékre stb. (a halmazelmélet kifejezésmódján: szavak halmazának a főnevek, melléknevek, igék stb. halmazára) való felosztása esetén a felosztás szempontja az, hogy a szó milyen *szófajhoz* tartozik. Igen ám, de mik a szófajok? Amikor a nyelvtudomány azt a kérdést veti fel, hogyan lehet szabatosan definiálni a szófajokat, akkor éppen az a kérdés, hogyan történik a szavak felosztása szófajokra, erre a kérdésre pedig ugyancsak semmitmondó az a válasz, hogy aszerint történik, hogy melyik szó milyen szófajhoz tartozik.

Hasonló nehézségek a matematikában is előfordulnak, ha ki akarjuk elégíteni a formális logika említett követelményét. A felosztás szempontját a matematikában is gyakran csak utólag lehet kielemezni, sokszor fáradságos munkával, magából a felosztásból. Ezért a matematikus gyakran túlteszi magát a formális logika említett követelményén és más, halmazelméleti módszereket alkalmaz valamely felosztás megadására. A leggyakrabban alkalmazott ilyen módszer abban áll, hogy megadunk egy, a felosztandó halmaz elemei körében értelmezett relációt avégett, hogy a felosztandó halmaz két elemét akkor és csak akkor soroljuk ugyanabba a részhalmazba, ha ez a reláció fennáll közöttük. Ez azonban csak akkor lehetséges, ha a kérdéses reláció reflexív (mert a halmaz bármely a elemének ugyanabba a részhalmazba kell tartoznia, mint az a elemnek, s így az a elem és önmaga között fenn kell állnia a kérdéses relációnak), szim-

metrikus (mert ha a halmaz valamely *a* eleme ugyanazon részhalmazba tartozik, mint a *b* eleme, akkor a *b* elem is ugyanabba a részhalmazba tartozik, mint az *a* elem, s így ha az *a* és a *b* elemek között fennáll a kérdéses reláció, akkor a *b* és az *a* elemek között is fenn kell állnia), és tranzitív (mert ha a halmaz *a* eleme ugyanabba a részhalmazba tartozik, mint a *b* eleme, ez pedig ugyanabba a részhalmazba, mint a halmaz *c* eleme, akkor az *a* elem ugyanabba a részhalmazba tartozik, mint a *c* elem, s így ha az *a* és a *b* elemek között is, a *b* és a *c* elemek között is fennáll a kérdéses reláció, akkor az *a* és a *c* elemek között is fenn kell állnia). Az olyan relációkat, amelyeknek az említett három tulajdonságuk (reflexivitás, szimmetria, tranzivitás) megvan, *ekvivalenciarelációknak* szokás nevezni. A halmazelméletben megmutatják, hogy valamely halmaz elemei körében értelmezett bármely ekvivalenciarelációhoz van olyan felosztása a halmaznak páronként egymástól idegen részhalmazaira, amely felosztás során a halmaz két eleme akkor és csak akkor kerül ugyanabba a részhalmazba, ha fennáll közöttük a kérdéses ekvivalenciareláció. A felosztás szempontja helyett ezért gyakran ezt az ekvivalenciarelációt adjuk meg a matematikában. Azokat a részhalmazokat, amelyekre az így adódó felosztás felosztja az adott halmazt, az ehhez az ekvivalenciarelációhoz tartozó *absztrakcióosztályoknak* szokás nevezni.

Például a Kulagina-féle szófajelméletben (és már korábbi nyelvszerkezet-vizsgálatokban is) fontos szerepet játszik a szóalakok körében következőképpen definiált ekvivalenciareláció: akkor és csak akkor mondjuk, hogy ez az ekvivalenciareláció fennáll két szóalak között (rövidebben: akkor és csak akkor mondjuk, hogy két szóalak *ekvivalens* ennek az ekvivalenciarelációnak értelmében), ha bármely mondatból, amelyben a két szóalak valamelyike előfordul, ismét (grammatikailag helyes) mondatot kapunk, ha e szóalakot (szabatosabban, amennyiben az több példányban is előfordul a kérdéses mondatban, annak bármelyik példányát) a kérdéses két szóalak közül a másikkal pótoljuk. Például a magyar nyelvben a *házat* szóalak ekvivalens a *széket* szóalakkal, mert bármely magyar mondatból ugyancsak magyar mondat lesz, ha e szóalakok egyikét, amennyiben előfordul a mondatban, a másikkal pótoljuk. Így pl. a *vettem egy házat* mondatból a *vettem egy széket* mondat lesz, ha a *házat* szót a *széket* szóval pótoljuk, az *a házat lakásra használjuk* mondatból pedig a következő mondat: *a széket lakásra használjuk;* ez grammatikailag helyesen képezett mondat, ha értelmetlen is. A *házat cseréltem, házat adtam érte* mondatból pedig a *széket cseréltem, házat adtam érte,* vagy a *házat cseréltem, széket adtam érte* mondat lesz, aszerint, hogy a *házat* szóalak melyik példányát pótoljuk a *széket* szóalakkal.

Könnyű megmutatni, hogy az így definiált reláció valóban ekvivalenciareláció (vagyis reflexív, szimmetrikus és tranzitív; a szimmetriát azzal értük el, hogy azt kívántuk, hogy a két szóalak, *bármelyikét* a másikkal pótolva mondatból ismét mondatot kapjunk). Ennél fogva fel lehet osztani a szóalakok halmazát az ehhez az ekvivalenciarelációhoz tartozó absztrakcióosztályokra. Könnyű látni, hogy egy-egy absztrakcióosztályba az ugyanazon szófajhoz tartozó szavak ugyanazon nyelvtani alakjai tartoznak, pl. az egyikbe az alanyesetben álló főnevek, a másikba a tárgyesetben álló főnevek, ismét másikba a feltételes mód jelen idő többes szám második

személyében álló igék stb. Azonban külön absztrakcióosztályt alkotnak azok a szóalakok, amelyek ugyanazon, esetleg különböző szófajokhoz tartozó szavak ugyanazon, esetleg különböző nyelvtani alakjai lehetnek. Például külön absztrakcióosztályba tartoznak az olyanféle szóalakok, mint a *falat*, *szelet* stb., amelyek mindegyike lehet alanyesetben álló főnév is, tárgyesetben álló főnév is, sőt, jelentő mód jelen idő egyes szám harmadik személyben álló (műveltető) ige is, másik absztrakcióosztályba az olyanféle szóalakok, mint az *ég*, *fal*, *fagy*, *lép*, *tűz* stb., amelyek mindegyike alanyesetben álló főnév is, jelentő mód jelen idő egyes szám harmadik személyben álló ige is lehet. (A *ház* és a *néz* szó már nem ebbe, hanem másik absztrakció osztályba tartoznak, mert pl. a következő két mondat közül: *az ég kék*, *a tűz nagyon ég*, az elsőből ugyan grammatikailag helyes mondat lesz, ha az *ég* szóalakot a *ház* szóalakkal pótoljuk: *a ház kék*, de a másodikból már nem: *a tűz nagyon ház* már nem helyes mondat: viszont épp a másodikból lesz helyes mondat, ha az *ég* szóalakot a *néz* szóalakkal pótoljuk: *a tűz nagyon néz*, de az elsőből nem: a *néz kék* nem helyes mondat.)

Ez a felosztás „*szóalakfajokra*", *a ragozási* (pontosabban: szóalakképzési) *homonimia* figyelembevételével, csak látszólag erőltetett: a gépi fordítás alkalmával a gépnek az ugyanabba az absztrakcióosztályba sorolt szóalakokkal ugyanazt kell csinálnia (pl. ha egy szóalak egy főnév s egy ige valamely alakjának szerepét is játszhatja, meg kell állapítania, hogy a lefordítandó mondatban melyik szerepben áll). További elemzés még más eltéréseket is felfedhet a közelfekvő értelemben vett szóalakfajokra való felosztástól, ilyen eltérést okoz pl. a magyar nyelvben az alanyi és a tárgyi ragozás; de be lehet látni, hogy pl. a gépi fordítás szempontjából mindig értelmes eltérésekről van szó.

15. Az *absztrakcióosztály* kifejezés onnan ered, hogy ezt az eljárást a matematikában leggyakrabban valamilyen absztrakciós lépéssel kapcsolatban szokás alkalmazni. Ennek az absztrakciós lépésnek az általános szkémája a következő. Adva van egy halmaz. Ennek a halmaznak az elemeit vizsgáljuk, azonban el akarunk tekinteni valamilyen tulajdonságaiktól. Más szóval, a halmaz két elemét nem tekintjük különbözőnek, ha (legfeljebb) csak a kérdéses tulajdonságaikban térnek el egymástól, egyébként megegyeznek. Ilyenkor azzal az ekvivalenciarelációval alkalmazzuk a fenti felosztási eljárást, amely akkor áll fenn a halmaz két eleme között, ha a kérdéses tulajdonságaiktól eltekintve megegyeznek egymással. Könnyű látni, hogy ez valóban ekvivalenciareláció, s így a hozzá tartozó absztrakcióosztályokra való felosztás valóban lehetséges. Egy-egy absztrakcióosztály a felosztandó halmaz olyan elemeiből áll, amelyek a kérdéses tulajdonságaiktól eltekintve megegyeznek egymással, és minden ilyen elemet tartalmaz. Ennélfogva ezek az absztrakcióosztályok épp az absztrakciós lépéssel keletkező absztrakt fogalmak halmazelméleti realizációinak tekinthetők.

Ez a halmazelméleti realizáció is igen fontos, a matematikában léptennyomon használt halmazelméleti módszer. Ez tehát abban áll, hogy az absztrakt fogalmakat az alájuk tartozó konkrét speciális esetek halmazával realizáljuk. Így realizálja a matematikus a háromszögek alakját az összes, kérdéses alakú háromszögek halmazával (tehát a háromszögek hasonló-

ságához mint ekvivalenciarelációhoz tartozó absztrakcióosztályokkal), vagy az egyenes irányát az összes, kérdéses irányú egyenesek halmazával (tehát az egyenesek párhuzamosságához mint ekvivalenciarelációhoz tartozó absztrakcióosztályokkal). Ennek a halmazelméleti realizációs módnak az az előnye, hogy az absztrakt fogalmak helyett, amelyekkel eleinte nehéz volna dolgozni, halmazelméleti realizációjukkal dolgozhatunk, amelyeket már jól tudunk kezelni a halmazelmélet módszereivel.

Ezt a fontos halmazelméleti módszert is megpróbálom néhány nyelvészeti példával megvilágítani. A *fonéma* fogalma úgy keletkezik a beszédben használatos fizikai hang fogalmából, hogy e hang bizonyos, a mondanivaló jelentése szempontjából lényegtelen tulajdonságaitól eltekintünk. Ezt az absztrakciós lépést a halmazelméleti realizáció módszerével úgy végezhetjük el, hogy minden fonémát az összes olyan hangok halmazával realizáljuk, amelyek a beszédben a kérdéses fonémát képviselhetik. Vagy, ha mindezeket a hangokat a kérdéses fonéma *variánsainak* nevezzük (kissé eltérően a fonémavariánsnak a fonológiában használatos fogalmától, hiszen pl. a hang különböző magasságú változatait a fonológus nem szokta különböző fonémavariánsoknak tekinteni), és a matematikustól eltanuljuk azt a (veszélytelen) pongyolaságot, hogy az absztrakt fogalmakat sokszor egyenesen *azonosítja* halmazelméleti realizációjukkal, azt is mondhatjuk: a fonéma bizonyos hangoknak, tudniilllik a kérdéses fonéma variánsainak halmaza. (Ilyen azonosítással volt dolgunk akkor is, amikor a „páros" és „páratlan" tulajdonságoknak 2 helyett más modulusra való általánosítása esetén a megfelelő tulajdonságok helyett a maradékosztályokról beszélhetünk.)

Persze az, hogy mely hangok tartoznak egy fonémához, attól is függ, melyik nyelvről van szó. Például a japán nyelvben a magyar [l] és [r] ugyanannak a fonémának variánsai, viszont az [l] és az [l'], amelyek a szláv nyelvekben különböző fonémák, a magyarban ugyanannak a fonémának variánsai. Így hát különböző nyelvekhez más és más olyan, a beszédhangok közötti ekvivalenciareláció tartozik, amelynek absztrakcióosztályai a kérdéses nyelv fonémái.

Ugyanígy a *szó* (lexéma értelmében használt) fogalma is absztrakcióval keletkezik a *szóalak* konkrétabb fogalmából. (A szóalak fogalma annyiban konkrétabb, hogy a mondat szóalakokból, nem pedig lexémákból áll.) A halmazelméleti realizáció módszerét alkalmazva, a szavakat szóalakok halmazaival realizálhatjuk (vagy az említett pongyolaságot megengedve, azonosíthatjuk), tudniillik a szó valamennyi nyelvtani alakjának halmazával. Így pl. a *ház* szót a *ház, házat, háznak, házba, házban, házból, házra, házon, házról* stb. szóalakokból, a *lát* szót a *látok, látsz, lát, látunk, láttok, látnak, látom, látod, látja, látjuk, látjátok, látják, láttam* stb. szóalakokból álló halmaz realizálja.

Amikor a *ház* vagy a *lát* szóról beszélünk, akkor is olyan módszert alkalmazunk, amelyet a matematikus is lépten-nyomon használ. Ez a módszer abban áll, hogy ha páronként egymástól idegen halmazokról van szó (pl. valamely halmaz felosztásában szereplő részhalmazokról), akkor ezeket a halmazokat úgy adjuk meg, hogy egy-egy elemüket megnevezzük. Az e célból megnevezett elemet a halmaz *reprezentáns elemének* szokás nevezni.

Ezt a halmaz elemei közül tetszőlegesen választhatjuk. Ilyen önkényesen választott reprezentáns elem pl. valamely szó *szótári alakja*. Amikor a szótár a *lát* ige német megfelelőjeként a *sehen* igét nevezi meg, ezzel nem azt akarja mondani, hogy a *lát* igealaknak a *sehen* igealak a megfelelője, hanem itt a *lát* igealak a *látok, látsz, lát* stb. igealakokból álló halmaznak, a *sehen* igealak pedig a *sehe, siehst, sieht* stb. igealakokból álló halmaznak önkényesen választott reprezentáns eleme.

Ha a halmazok, amelyekről szó van, nem idegenek egymástól páronként, akkor a reprezentáns-elemek módszere nem mindig alkalmazható jól, pl. az „*ég* szó" nem egyértelmű kifejezés, jelentheti az *ég* főnevet is (amelyet lehetne *eget* szónak is nevezni), az *ég* igét is (amelyet lehetne *égni* szónak is nevezni). A zárójelben adott példák a reprezentáns-elemek módszerének olyan módosítására utalnak, amikor a szóban forgó halmazokból olyan reprezentáns elemeket választunk, amelyek e halmazok közül csak az egyiknek elemei. De e módosított alakban is csak akkor alkalmazható a reprezentáns elemek módszere, ha a szóban forgó halmazok mindegyikének van olyan eleme, amely a többinek nem eleme.

Persze van más mód is a halmazok megadására. Például az, hogy névvel látjuk el őket s nevüket adjuk meg. Például ilyen, a szófajok megadására szolgáló nevek: *főnév, melléknév, ige* stb. (Persze a szófajok is halmazok; pl. a *főnév* szófaj az összes főnevek halmaza. Amikor összességről, osztályról, fajtáról, fajról és hasonlókról beszélünk, mindig tulajdonképpen halmazról van szó.) Vagy megadhatunk halmazokat úgy is, hogy a szóban forgó halmazok mindegyikének valamilyen módon más-más elemet feleltetünk meg (nem szükségképpen a kérdéses halmaz elemét) és a megadandó halmaznak megfeleltetett elemet nevezzük meg. Ezt tesszük pl. a Karinthy Frigyes-féle „mit csinál a kicsoda a micsodában?" játék esetén: valamely irodalmi idézetben szereplő szóalakok mindegyikéről úgy adjuk meg, hogy melyik szóalakfajhoz tartozik, hogy a megfelelő kérdőszó-alakot nevezzük meg. Ezzel már bizonyos szempontból megadtuk az idézet szerkezetét, tudniillik azt, milyen szóalakfajhoz tartozó szóalakokból, milyen sorrendben épül fel. Maga a játék azon alapul, hogy e „szóalakfaj-szerkezet" annyira jellemző az idézetre, hogy nagyon gyakran ki tudjuk találni annak alapján, hogy melyik idézetről van szó.

Még könnyebben ki lehet találni egy idézetet, ha a benne szereplő szóalakok mindegyikéről azt adjuk meg, hogy melyik szónak alakja, anélkül, hogy megadnók, melyik nyelvtani alakjáról van szó. Így is az idézet szerkezetét adtuk meg bizonyos, de az előbbitől különböző szempontból; az így megadott szerkezetet az idézet *szószerkezetének* nevezhetjük. Ki ne találná ki, melyik idézet szószerkezete a következő: *haza rendületlen lesz hív óh magyar*, pedig a *hív* szót nem is adtuk meg egyértelműen (nem mondtuk meg, hogy a *hív* főnévről, nem pedig a *hív* igéről van-e szó).

Valamely irodalmi idézet, vagy általában valamely mondat szóalakfaj-szerkezete és szószerkezete két példa arra, amit Kulagina általában a mondat *szerkezetének* nevez. Ha a nyelv szóalakjainak halmazát tetszőleges módon felosztjuk páronként egymástól idegen részhalmazokra, akkor valamely mondatnak ehhez a felosztásához tartozó szerkezetét úgy kapjuk meg, hogy a mondatban szereplő szóalakok mindegyikéről sorra megadjuk,

mely részhalmazhoz tartozik. Ha a felosztás a szóalakfajokra való felosztás, akkor a szóalakfaj-szerkezethez, ha pedig az a felosztás, amelynél egy és ugyanazon szó különböző alakjai kerülnek egy részhalmazba, akkor a szószerkezethez jutunk. (Attól a körülménytől most tekintsünk el, hogy ezek a részhalmazok már nem idegenek egymástól páronként, mert különböző szavaknak lehetnek homoním alakjaik.)

A szóalakok halmazának minden egyes felosztása módot ad egy ekvivalenciareláció definíciójára e halmaz azon részhalmazai körében, amelyekre felosztottuk a szóalakok halmazát: két ilyen részhalmazt ekvivalensnek nevezünk, ha bármely olyan mondatnak a kérdéses felosztásához tartozó szerkezetéből, amelyben a két részhalmaz egyike szerepel (vagyis a mondatban szereplő valamely szóalak e két részhalmaz egyikéhez tartozik), ismét valamely mondatnak ugyene felosztáshoz tartozó szerkezetét kapjuk, ha ezt a részhalmazt a két részhalmaz közül a másikkal pótoljuk. Más szóval: ha minden ilyen mondathoz van olyan másik, ugyancsak grammatikailag helyes mondat, amelyben ugyanannyi szóalakpéldány fordul elő, mint az eredeti mondatban; és amelyben minden egyes szóalak ugyanahhoz a részhalmazhoz tartozik, mint az eredeti mondatban szereplő ugyanannyiadik szóalak, kivéve egyetlen szóalak egyetlen példányát, tudniillik azt, amely e másik mondatban ugyanannyiadik helyen áll, mint az eredeti mondatban szereplő azon szóalaknak valamelyik (esetleg egyetlen) példánya, amely a két részhalmaz egyikéhez tartozik; s a másik mondatban e helyen álló szóalak a két részhalmaz közül a másikhoz tartozik.

Az ehhez a relációhoz (amelyről meg lehet mutatni, hogy valóban ekvivalenciareláció) tartozó absztrakcióosztályok olyan halmazok, amelyeknek az elemei olyan részhalmazok, amelyekre a szóalakok halmazát felosztottuk. Ha e részhalmazokat minden egyes absztrakcióosztályon belül egyesítjük, a szóalakok halmazának újabb, kevesebb számú és bővebb részhalmazokra való felosztását kapjuk (kivéve azt az esetet, ha minden egyes absztrakcióosztálynak csak egyetlen eleme van). Az eredetileg tekintett felosztásból ily módon leszármaztatott újabb felosztást Kulagina az eredetileg tekintett felosztás *deriváltjának* nevezi. A derivált fogalma segítségével Kulagina kétféle olyan fogalmat is definiál, amelyek logikai terjedelem szempontjából nagyjából megegyeznek a szófaj hagyományos fogalmával. Kulaginánál azonban a szófajok olyan halmazok, amelyeknek nem a kérdéses szófajhoz tartozó szavak az elemei, hanem e szavak összes nyelvtani alakjai. Azt is megvizsgálja Kulagina, hogy milyen tulajdonságú nyelvek esetén vezet ez a kétféle definíció ugyanazokhoz a szófajokhoz. Végül kiterjeszti az elméletet az úgynevezett *konfigurációkra* is (amelyek lényegében a szintagmafajoknak felelnek meg), oly módon, hogy figyelembe veszi egy-egy szóalaknak több szóalakkal való olyan pótlási lehetőségeit is vagy viszont, amelyek mondatokat mindig ismét mondatokba visznek át. (Például a *rét* szóalak ilyen értelemben pótolható a *kopár szik sarja* szintagmával; e pótlás pl. az *ég a rét* mondatot az *ég a kopár szik sarja* mondatba viszi át. Hasonlóan, az *ég* igealak is pótolható az *ég a napmelegtől* szintagmával; e pótlás az utóbbi mondatot az *ég a napmelegtől a kopár szik sarja* mondatba viszi át. E pótlás lehetősége nyilván azzal függ össze, hogy a *kopár szik sarja* főnévi, az *ég a napmelegtől* pedig igei szintagma.)

A szófaj fogalmának egy nagyjából helyes definíciójához úgy juthatunk el, hogy a szóalakok halmazának a szavakra való felosztásából indulunk ki (minden szót nyelvtani alakjainak halmazával azonosítva). Igaz, ez a felosztás (a szóalakképzési homonímiák miatt) nem páronként idegen szóalakokra való felosztás (pl. az *égni* igének és az *eget* főnévnek közös szóalakja az *ég* is, az *égnek* is); mégis tartozik e felosztáshoz is a fentihez hasonlóan definiálható ekvivalenciareláció. Ennek az ekvivalenciarelációnak a definíciója: két szót ekvivalensnek nevezünk, ha bármely olyan mondat szószerkezetéből, amelyben e két szó valamelyikének valamely nyelvtani alakja szerepel, ismét valamely (grammatikailag helyes) mondat szószerkezetét kapjuk, ha az eredeti mondat szószerkezetében e szó bármelyik példányát (amennyiben több példányban is előfordul) a két szó közül a másikkal pótoljuk. Az ehhez az ekvivalenciarelációhoz tartozó absztrakcióosztályok — a hagyományos szófajfogalomtól való kisebb eltérésekkel — épp a szófajok. (Hasonló okból, mint ahogyan fentebb a szóalakfajokhoz jutottunk, amikor két szóalakot ekvivalensnek tekintettünk, ha bármelyikük valamelyik példányát a másikkal pótolva olyan mondatban, amelyben legalább egy példányban előfordul, ismét grammatikailag helyes mondathoz jutunk, s az ehhez az ekvivalenciarelációhoz tartozó absztrakcióosztályokat tekintettük. Csakhogy most eleve eltekintettünk attól, hogy a szavaknak melyik nyelvtani alakjáról van szó, hiszen a mondatok helyett, amelyek szóalakokból állnak, szószerkezetükkel dolgoztunk, amely a megfelelő szavakból áll. Ezért jutottunk most a szóalakfajok helyett a szófajokhoz.)

16. Még azokról a matematikai módszerekről kell megemlékeznem, amelyeket a *matematikai logika* alkalmaz logikai kérdések megoldására. A matematikai logika keletkezése Leibniz munkásságához fűződik. Leibniz, mint tudjuk, kiváló matematikus is volt, kora filozófusai között is előkelő helyet töltött be. Így gyakran volt alkalma matematikus kollégáival is vitakozni, meg filozófus kollégáival is. Azt tapasztalta, matematikus kollégáival mindig megértették egymást, a vita során mindig kiderült, kinek volt igaza, ha másképp nem, kiszámították, s az eredményt az is mindig elismerte, aki előzőleg mást állított a vitában. Ha viszont filozófus kollégáival vitatkozott, se vége, se hossza nem volt a vitának, semmi eredményre nem vezetett. Vizsgálni kezdte, miért van így.

Eközben, mai szemmel nézve naivul, figyelmen kívül hagyta azokat a tényezőket, az egyéni érdekektől az osztályérdekekig, amelyek az emberek különböző filozófiai nézeteit motiválják, és pusztán csak *nyelvi kérdést* látott e jelenségben. Úgy gondolta, a filozófusok azért nem értik meg egymást, mert a köznapi nyelvet használják a vita során, az pedig tele van többértelmű szavakkal. Ezeken egyikük mást ért, mint a másik, főleg, ha olyan filozófiai szakkifejezésről van szó, mint pl. *szubsztancia*, vagy esetleg semmit sem ért rajta a másik, mint pl. akkor, ha ő *monasz*-okról beszél. Ezzel szemben a matematikusnak van egy sokkal szabatosabb jelrendszere, a matematika „formulanyelve", amely már nem ad lehetőséget félreértésre. Így aztán Leibniz abban látta minden eredménytelen filozófiai vita kiküszöbölésének lehetőségét, hogy a matematika formula-

nyelvéhez hasonló, de nemcsak a matematikai fogalmak, hanem mindenféle fogalom kifejezésére alkalmas egyetemes fogalomnyelvet, *characteristica universalis*-t kell szerkeszteni, egy olyan algoritmussal, *calculus ratiocinator*-ral, egyetemben, amelynek segítségével mindenféle vita esetén el lehet dönteni, mintegy ki lehet számítani, kinek van igaza.

Ma mosolygunk Leibniznek ezeken az utópista gondolatain. Arra azonban jók voltak, hogy első lépésként hozzáfogjon a matematika formulanyelvének tökéletesítéséhez. A matematika formulanyelve ugyanis nem önálló jelrendszer, még kevésbé volt az Leibniz idejében. Csak aki sohasem látott még matematikai szakcikket, képzeli azt, hogy a matematikus mindent formulákkal fejez ki. A matematikai dolgozatokban éppúgy van a szerző anyanyelvén, vagy más nyelven írt szöveg is a képletek között, mint ahogy a dalnak is van szövege a kotta alatt. Sőt, a matematikai képletek még szervesebben illeszkednek bele a nyelvi szövegbe, mint ahogy a dal kottája kapcsolódik a szövegéhez. Hiszen pl. egy algebrai kifejezés lehet a mondat alanya, vagy tárgya, vagy más része (pl.: „$a + b$-t úgy szorozhatjuk meg c-vel, hogy ac-hez bc-t hozzáadjuk"), formula illeszkedhetik mellékmondatként a nyelvi mondatba (pl.: „ha ezt az egyenletet megoldjuk, azt kapjuk, hogy $x = \frac{a+b}{2}$"). A matematikai dolgozatok nyelvi szövegrésze nagyon is lényeges dolgokat mond, sokszor már egy-egy, vagy néhány szava is. Egészen mást jelent az, ha két egyenletet az *és* szóval köti össze a matematikus, pl. amikor, egyenletrendszer megoldása során, x és y olyan értékeit keresi, amelyekre

$$2x + 3y = 18$$

és

$$4x + 5y = 32,$$

mint az, amikor két egyenlet közé a *tehát* szót tesszük, mert arról van szó, hogy az egyik egyenletből következik a másik, pl.

$$2x + 3y = 18,$$

tehát

$$4x + 6y = 36.$$

(A fenti egyenletrendszer megoldása során erre is hivatkozhatunk; s akkor az egyenletrendszer második egyenletével összevetve máris látjuk, hogy y csak 4 lehet, mert 36 ennyivel több, mint 32.)

Leibniz tehát azzal kezdte, hogy az olyanféle logikai fogalmakat jelölő szavak számára, mint *és*, *tehát*, vezetett be a matematikaiakhoz hasonló jelöléseket. Úgy gondolta, ezzel fontos lépést tett a *characteristica universalis* megalkotása felé, hiszen e fogalmak nemcsak matematikai, hanem másféle, többek között filozófiai meggondolásokban is fontos szerepet játszanak. Ezt a Leibniz által megkezdett munkát csak a legutóbbi századforduló körül fejezték be Peano olasz, továbbá Whitehead és Russell angol matematikai logikusok, amennyiben a logikai jelölések olyan rendszerét vezették be, amelyekkel, a matematikai jelölések mellett, koruk matematikájának minden

definícióját, tételét, és e tételek bizonyítását is, fel lehet írni anélkül, hogy az olasz, angol, vagy valamiféle más természetes nyelv szavait használnók. E leibnizi cél elérése, a matematika úgynevezett *teljes formalizálása* azonban nem vezetett a matematikai meggondolások érthetőbbé, vagy áttekinthetőbbé válásához; ebből a szempontból ma is célszerűbbnek látszik, matematikai logikai dolgozatokban is, csak addig menni a formalizálásban, amíg az növeli az áttekinthetőséget, a többit pedig szavakkal fejezni ki.

Leibniz azonban a *calculus ratiocinator* megalkotására is gondolt. Megfigyelte, hogy annak az, ítéleteken végezhető, műveletnek, amelyet az ítéletek közé tett *és* szóval szokás kifejezni (s amelyre e szó helyett jelölést kívánt bevezetni) hasonló tulajdonságai vannak, mint a számok összeadásának, s így hasonlóan lehet vele „számolni". Például mindegy, hogy x és y olyan értékeit keressük-e, amelyekre

$$2x + 3y = 18$$

és

$$4x + 5y = 32,$$

vagy olyanokat, amelyekre

$$4x + 5y = 32$$

és

$$2x + 3y = 18;$$

ugyanez igaz más ítéleteknek a kérdéses művelettel való összekapcsolására is, más szóval, e műveletnek megvan a kommutatív tulajdonsága. Leibniz emiatt a számokon végzett műveletek jeleit használta a logikai műveletek jele gyanánt is. Ettől a szokástól később eltértek, nemcsak azért, mert kiderült, hogy az ítéleteken (és, ami Leibniz számára még fontosabb volt, a fogalmakon) végzett műveleteknek vannak a számokon végzett műveletektől eltérő tulajdonságaik is (pl. mindegy, azt mondom-e, x és y elégítsék ki a

$$2x + 3y = 18$$

és

$$2x + 3y = 18$$

egyenleteket, vagy csak egyszer mondom-e elégítse ki a

$$2x + 3y = 18$$

egyenletet), hanem azért is, mert épp a matematika formulanyelve egyértelműségének megőrzése végett célszerű, ha különböző fogalmakat, pl. az ítéletek összekapcsolását az *és* szóval és a számok összeadását, másképp jelöljük, még akkor is, ha vannak hasonló tulajdonságaik.

A logikai műveletek jelölésrendszere még máig sem alakult ki véglegesen, különböző szerzők más és más jelöléseket használnak; egy ilyen elterjedt jelölésrendszer az ítéleteknek az *és* szóval való összekapcsolását (úgy-

nevezett *konjunkcióját*) az ítéletek, vagy az azokat kifejező formulák közé tett „$\wedge$" jellel, az ítéleteknek a (megengedő értelemben használt, vagyis a két ítélet együttes igazságát sem kizáró) *vagy* szóval való összekapcsolását (úgynevezett *diszjunkcióját* vagy *alternatíváját*) a „$\vee$" jellel, a *ha* és *akkor* szókkal való összekapcsolásukat (úgynevezett *implikációjukat*) az ugyancsak az ítéletek, vagy a megfelelő formulák közé, tehát az *akkor* szó helyére tett „$\rightarrow$" jellel, a matematikában nagyon gyakran szokásos *akkor és csak akkor, ha* szókkal való összekapcsolásukat (úgynevezett *ekvivalenciájukat*) a közéjük tett „$\leftrightarrow$" jellel, ítélet tagadását (úgynevezett *negációját*) pedig az ítélet, vagy a megfelelő formula elé tett „$\neg$" jellel jelöli. Például ahelyett, hogy

$$2x + 3y = 18$$

és

$$4x + 5y = 32,$$

röviden azt írhatjuk, hogy $2x + 3y = 18 \wedge 4x + 5y = 32$, ahelyett, hogy ha

$$2x + 3y = 18,$$

akkor

$$4x + 6y = 36,$$

azt írhatjuk, hogy $2x + 3y = 18 \rightarrow 4x + 6y = 36$, és pl. azt, hogy ha

$$ax^2 + bx + c = 0$$

és a nem egyenlő 0-val, akkor vagy

$$x = \frac{-b + \sqrt{b^2 - 4ac}}{2a}$$

vagy

$$x = \frac{-b - \sqrt{b^2 - 4ac}}{2a}$$

így írhatjuk tömören:

$$(ax^2 + bx + c = 0 \wedge \neg\, a = 0) \rightarrow \left(x = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \vee \right.$$

$$\left. \vee\, x = \frac{-b - \sqrt{b^2 - 4ac}}{2a} \right).$$

Ha még azt is ki akarjuk fejezni, hogy ez a, b, c és x minden értékére áll, a még egy zárójelbe zárt fenti formula elé az $\forall a\, \forall b\, \forall c\, \forall x$ jeleket is tehetjük ($\forall$ a köznapi nyelvben a *minden* szóval és az a, b, c, illetőleg után tett *-ra, -re* raggal kifejezett úgynevezett *általános kvantor*). Azt pedig, hogy za a, b és c együtthatók minden értékére, ha a nem egyenlő 0-val, van olyan

x szám, amelyre teljesül az $ax^2 + bx + c = 0$ másodfokú egyenlet, az $\exists$ jelölést (ún. *exisztenciális kvantort*, jelentése: 'van oly ..., amelyre') is alkalmazva így írhatjuk:

$$\forall a \, \forall b \, \forall c \, (\neg a = 0 \rightarrow \exists x \, (ax^2 + bx + c = 0)).$$

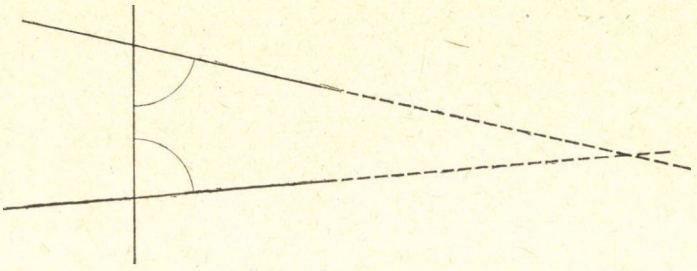
A matematikai logika tárgya hosszú ideig a logikai műveletek és a számokon végzett műveletek tulajdonságai közötti megegyezések és eltérések vizsgálata volt. Ezek a vizsgálatok főleg az algebrát fejlesztették. Az első olyan algebrai struktúrafajtára, amely más jellegű, mint a számokból álló struktúrák, a matematikai logika szolgáltatott példát Boole angol matematikai logikus vizsgálataival; az ilyenfajta struktúrákat ma is *Boole-féle algebráknak* nevezik. (Példát szolgáltat rájuk az ítéletek halmaza, a konjunkció, a diszjunkció és a negáció műveletével „felszerelve".)

16. Mind a matematika teljes formalizálására irányuló törekvések, mind a logika szolgáltatta fogalmakból, ítéletekből, osztályokból stb. álló rendszerek szerkezeti vizsgálata előkészítette a talajt azoknak a mélyebb logikai kérdéseknek a vizsgálatára, melyekkel a matematikus a múlt század első felétől kezdve a maga tudományszakjának fejlesztése során találkozott. Ilyen kérdés pl. az *axiomarendszerek ellentmondástalanságának* kérdése. Ezzel a kérdéssel először a *nem-euklideszi geometria* létrejöttével találkozott a matematikus.

Az ókori görögök matematikája főleg abba különbözött az egyiptomiak, babiloniaiak és egyéb ókori keleti népek matematikájától, hogy, míg ezek megelégedtek a (sokszor csak megközelítőleg érvényes) matematikai tételek és eljárások kimondásával, sőt, legtöbbször még ehelyett is csak példákkal való illusztrálásával, a görögök igyekeztek matematikai eredményeiket deduktív módon *bebizonyítani* s ezzel biztosítani, hogy azok szigorúan érvényesek legyenek. Eközben persze rájöttek, hogy minden matematikai bizonyítás hivatkozik más matematikai tényekre, ezek bizonyítása ismét másokra, stb. Minthogy azonban ez újabb és újabb tények bizonyítását nem folytathatjuk a végtelenségig (ez a *végtelen regresszus lehetetlensége* később Arisztotelész által megfogalmazott elvének speciális esete), a matematikai tételek bizonyításának rendszere szükségképpen bizonyítás nélkül elfogadott tényekből indult ki. Ezeket a bizonyítás nélkül elfogadott tényeket *axiómáknak*, néha *posztulátumoknak* nevezték. Hogy mikor alakult ki a görög matematika deduktív rendszere, hogy mikortól fogva alkalmazzák az *axiomatikus módszert*, különösen pedig hogy kialakulásában mi volt a szerepe az ókori görög matematika és filozófia kölcsönhatásának, arról még ma is folyik a vita a matematika történetének szakemberei között. Annyi azonban bizonyos, hogy a matematikának az axiomatikus módszerrel felépített deduktív rendszere az alexandriai korban már kialakult, mert Eukleidész Elemek (*Στοιχεῖα*) című műve viszonylag tökéletes alakban hagyta ránk az ókori görögök geometriájának axiomatikusan felépített rendszerét.

Eukleidész axiómáinak nagy része a szemlélet alapján nyilvánvaló, tehát bizonyítás nélkül nyugodtan elfogadható tényeket mond ki. Olyanokat, mint pl. hogy két ponton át mindig lehet egyenes vonalat húzni,

vagy hogy amik (pl. egyenes szakaszok, vagy szögek,) egymásra fektethetők, azok egyenlők. Kivételt képez az az axióma, amely később a *párhuzamosok axiómája* nevét kapta, s amelyet Eukleidész így fogalmaz: ha két egyenes vonalat metsz egy harmadik, s e harmadik egyenes vonal valamelyik oldalán olyan belső szögek keletkeznek a metszéspontokban, amelyeknek összege kisebb, mint két derékszög (ma úgy mondanók: kisebb, mint 180°), akkor az a két egyenes, ha ezen az oldalon elég messzire meghosszabbítjuk őket, metszi egymást (lásd a 10. ábrát. A *párhuzamosok axiómája* elnevezés ennek az axiómának arra az átfogalmazására utal, amely szerint ha két párhuzamos egyenes vonalat metsz egy harmadik egyenes vonal, akkor ennek a harmadik egyenes vonalnak mind-



10. ábra

két oldalán olyan belső szögek keletkeznek a metszéspontokban, amelyeknek összege 180°).

Ez az axióma van olyan bonyolult, vagy még bonyolultabb is, mint számos olyan tétel, amelyet Eukleidész bebizonyított. Ezért már az ókorban felmerült az a probléma, hogyan lehetne a párhuzamosok axiómáját Eukleidész többi axiómája segítségével bebizonyítani. Azonban két évezreden át hiába próbálkoztak a legjobb geométerek a bizonyítással, nem sikerült.

Eközben sokan megpróbálkoztak az indirekt bizonyítás *(reductio ad absurdum)* logikai módszerével is: feltételezték, hogy a párhuzamosok axiómája nem igaz, és igyekeztek bebizonyítani, hogy ez a feltevés képtelenség, mert két egymásnak (abszurd módon) ellentmondó tétel adódik belőle. Mikor azonban néhány lépésen át nem jöttek ki ilyen egymásnak ellentmondó tételek, abbahagyták ezt a próbálkozást is.

A magyar Bolyai János és az orosz Nyikoláj Ivanovics Lobacsevszkij azonban, egymástól függetlenül, messzebb ment a párhuzamosok axiómája tagadása következményeinek levonásában. Ellentmondásra ők sem jutottak, ehelyett azonban az euklideszi geometria tételeitől eltérő geometriai tételek meglepően következetes rendszere, mint Bolyai János apjának, Bolyai Farkasnak írta, „egy új, más világ" bontakozott ki előttük. Ezt a rendszert ma *Bolyai—Lobacsevszkij-féle geometriának*, vagy *hiperbolikus geometriának* hívjuk. Ez volt az első nem-euklideszi geometriai rendszer. (A *hiperbolikus* névnek nem sok köze van a *hiperbola* nevű kúpszelethez; talán arra utal, hogy az euklideszi geometria olyanféle átmenetet képez

a hiperbolikus geometriák és a később Riemann német matematikus által felfedezett másfajta nem-euklideszi geometriai rendszerek, az úgynevezett *elliptikus geometriák* között, mint a parabola a hiperbolák és az ellipszisek között. Hiperbolikus geometria is, elliptikus geometria is egyébként többféle lehet, euklideszi geometria azonban csak egyféle.)

Az a tény, hogy sem Bolyai János, sem Lobacsevszkij nem jutott egymásnak ellentmondó bizonyításához, csak valószínűvé teszi, de biztossá nem, hogy a hiperbolikus geometria axiómarendszere (amely Eukleidész axiómáiból áll, a párhuzamosok axiómája kivételével, és a párhuzamosok axiómájának tagadásából) *ellentmondástalan*, vagyis, hogy ha tovább folytatnók tételek bizonyítását a hiperbolikus geometria axiómáiból kiindulva, akkor sem juthatunk soha, semmiféle logikailag helyes következtetésmód alkalmazásával, egymásnak ellentmondó tételek bizonyításához. Márpedig az új geometria, a Bolyai—Lobacsevszkij-féle geometria számára létkérdés volt, vajon ellentmondástalan-e; mert ha egyszer, további következtetések útján, ellentmondásra vezetne, akkor ez a párhuzamosok axiómájának indirekt bizonyítását szolgáltatná, ami az euklideszi geometria végleges megszilárdítását, egyben azonban a Bolyai—Lobacsevszkij-féle geometria megsemmisülését jelentené.

Cayley angol és Felix Klein német matematikusok bebizonyították, hogy a hiperbolikus geometria axiómarendszere ellentmondástalan, ha az euklideszi geometriáé az; ez az eredmény felvetette azt a kérdést, valóban ellentmondástalan-e az euklideszi geometria axiómarendszere. Hilbert német matematikus bebizonyította, hogy az euklideszi geometria axiómarendszere ellentmondástalan, ha a valós számok aritmetikájáé az; ez az eredmény felvetette azt a kérdést, valóban ellentmondástalan-e a valós számok aritmetikájának axiómarendszere.

Időközben kialakult a halmazelmélet. Sőt, ennek eredeti, G. Cantor-féle formája, amelyet ma *naiv halmazelméletnek* neveznek, s amelyről szintén meg lehet mondani, milyen axiómákat fogad el bizonyítás nélkül többi tételei bizonyításának alapjául, ellentmondásra vezetett. Például be lehetett bizonyítani azt is, hogy van olyan halmaz, amelynek több, vagy legalábbis ugyanannyi eleme van, mint bármely más halmaznak, meg azt is, hogy nincs ilyen halmaz. Újra fel kellett tehát építeni a halmazelméletet, más axiómákra, úgy, hogy az addig felfedezett ellentmondások, a naiv halmazelmélet úgynevezett *antinómiái*, ne adódjanak belőlük, de a halmazelméletnek a matematika többi ágai számára fontos tételei igen. Nyitva maradt azonban az a kérdés, nem adódhatnak-e más úton egymásnak ellentmondó tételek a halmazelmélet új axiómáiból, vagyis, hogy ellentmondástalan-e a halmazelmélet új axiómarendszere. Sőt, minthogy addigra a matematika valamennyi ágát áthatotta a halmazelméleti módszer, felvetődött az a kérdés, nem hatoltak-e be ezzel az ellentmondás csírái a matematika többi ágaiba is, vagyis, hogy ellentmondástalanok-e a matematika különböző fejezeteiben használt axiómarendszerek.

Ahhoz, hogy valamely axiómarendszer ellentmondástalanságát megvizsgálhassuk, figyelembe kell vennünk v a l a m e n n y i logikailag helyes következtetésmódot; mert ha csak egy részüket vesszük figyelembe, nem tudhatjuk, a többiek alkalmazása nem vezethet-e egymásnak ellent-

mondó tételekhez. De melyek az összes logikailag helyes következtetésmódok? Ez nyilván logikai kérdés, de a logika tudománya nem tudott rá választ adni a matematikusnak. A következtetésmódok egy fontos alosztályát, a *kategorikus szillogizmusokat*, pontosabban: azok első három alakzatát (és azok modális változatait is) Arisztotelész megvizsgálta, abból a szempotból, melyek közülük a helyesek. Azt azonban már az ókori görög logikusok is tudták, hogy az arisztotelészi szillogizmusok nem merítik ki valamennyi helyes következtetésmódot. A sztoikus iskola logikusai, később Galenosz, majd a logika tudományának más szakemberei is, egészen a skolasztika rövid felfelé ívelő ágának nagy logikusaiig, további példákat adtak helyes következtetésmódokra, így a feltételes és a diszjunktív következtetéseket, a dilemma különböző eseteit stb. Egyesek megpróbálták ezeket rendszerbe is foglalni, azonban az így létrejött rendszerek még csak meg sem közelítették teljességben az arisztotelészi szillogizmusok rendszerét. Majd jött a skolasztika hosszú hanyatló korszaka, a logika felhasználása a szabad gondolkodás elnyomásának fegyvereként, a gondolkodás gúzsbakötése végett. Ehhez a következtetésnek merev keretek közé való szorítása kellett; a skolasztikusok minden tudományos indokolás nélkül, dogmatikusan kijelentették hát, hogy az arisztotelészi szillogisztika (amelyhez hozzászámították a későbbi logikusok által talált következtetésmódokat is) minden „megengedett" következtetésmódot tartalmaz. Tudományos indokolást nem is fűzhettek ehhez, legfeljebb Arisztotelész tekintélyére hivatkozhattak, hiszen a szaktudományok, különösen a matematika fejlődése akkor már számos olyan következtetésmód alkalmazását kívánta, amelyek nem szerepeltek a skolasztikus logika tankönyveiben felsorolt következtetésmódok között.

A logikának ezen az elmaradásán a szaktudományok fejlődése mögött (ami különben már az ókorban megkezdődött, hiszen pl. Eukleidész definíciói és bizonyításai sem kényszeríthetők az ókori logika szabályai közé,) nem segített az a kétségtelenül jelentős fejlődés sem, ami az újkor logikusai, így pl. Stuart Mill, vagy az ismeretelméleti logika úttörői munkásságához fűződik, Kanttól Hegelig. A marxizmus klasszikusai pedig más, fontosabb feladataik miatt nem jutottak hozzá, hogy új logikai rendszert dolgozzanak ki, akármennyire látták is a hagyományos logika fogyatékosságait, s rá is mutattak azokra.

Ebben a helyzetben a matematikus kénytelen volt maga megvizsgálni a maga módszereivel, a maga szaktudományára való alkalmazásokat tartva szem előtt, hogy melyek az összes logikailag helyes következtetésmódok. Mert hiszen ha olyan következtetésmódok alkalmazásával jutunk is, valamely axiómarendszer axiómáiból kiindulva, egymásnak ellentmondó tételek bizonyításához, amelyeket a hagyományos logika tankönyvei nem regisztráltak a megengedett következtetésmódok között, vagy amelyeket nem tart értékes, ismereteinket gyarapító következtetésmódoknak az ismeretelméleti logika, a kérdéses axiómarendszer mégis csak ellentmondásos; ha ilyen nem regisztrált vagy „értéktelen" következtetésmódok segítségével jutnánk a hiperbolikus geometria axiómáiból kiindulva ellentmondáshoz, az mégis csak a párhuzamosok axiómájának indirekt bizonyítását adná.

Ehhez a vizsgálathoz jött kapóra a matematikusnak, hogy a matematikában alkalmazható következtetésmódok esetén a részben matematikai formulákkal, részben szavakkal kimondott premisszákat és konklúziót az azokat kifejező (ahogy mondani szokás: *formalizáló*), logikai jelöléseket is tartalmazó formulákkal pótoltatta, hiszen e formulákat könnyebb kezelni, mint a részben szavakkal kifejezett ítéleteket. Például azt, ellentmond-e egymásnak két ítélet, nem egészen könnyű eldönteni; ha azonban formulákkal pótoljuk őket, azonnal láthatjuk, úgy keletkezik-e az egyik formula a másikból, hogy eléje tesszük a negáció „¬" jelét. Meg az is kapóra jött a matematikusnak, hogy addigra megvizsgálták a logikai jelölésekkel jelölt műveletek tulajdonságait.

Az axiómarendszerek ellentmondástalanságának kérdése csak egyike volt azoknak a matematika fejlődése során felvetődött kérdéseknek, amelyek a matematikust logikai problémák felvetésére, és minthogy a logika tudománya nem tudott választ adni rájuk, a maga módszereivel való megoldására kényszerítették. E problémák megoldása során meg kellett vizsgálnia a matematika különböző fejezeteiben használt, logikai jelölésekkel is kiegészített jelrendszerek, „formulanyelvek" szerkezetét, ez pedig már olyan kérdésekhez vezetett, amelyek nagyon hasonlítanak a nyelvtudomány egyes kérdéseihez.

Ismeretes, hogy a logika és a nyelvtudomány, ha nem is azonosak, szoros kapcsolatban vannak egymással. Már Arisztotelész is használt olyan szakkifejezéseket a logikában, mint az ítélet *alanya*, *állítmánya*, amelyeket ma, ha nem is teljesen ugyanabban az értelemben, nyelvészeti kifejezéseknek is tekintünk. A matematikai logikus sem kerülhette el, hogy olyan kérdéseket felvessen, mi játssza formuláiban az alany (vagy alanyok), mi az állítmány szerepét. Minthogy a formulanyelvek nyelvtana nem volt még kidolgozva, neki kellett azt is kidolgoznia. Eközben mindjárt figyelembe vehette a logika szempontjait is, hiszen a formulanyelv logikusabb, mint a köznapi nyelv.

A mondatrészek problémája a formulanyelvekben is kapcsolatos a szófajok, vagy inkább — minthogy nincs előírva, hol használjunk a formulákban szóközt (ha az áttekinthetőség fokozására szoktunk is használni), s így a formulanyelvben nehéz megmondani, meddig terjed egy szó — a szintagmafajok problémájával. Ezt is meg kellett tehát a matematikusnak oldania a maga formulanyelveiben. A formulanyelvekben a két legfontosabb szintagmafaj: a *kifejezés* és a *formula*. Kifejezés pl. $a + b$, vagy $(a + b)^2$, vagy $a^2 + 2ab + b^2$. A kifejezések valamilyen matematikai objektumot *jelentenek*, a felsorolt algebrai kifejezések esetén számot, vagy más struktúra-elemet, másféle kifejezések esetén pontot, egyenest, szöget, háromszöget, halmazt stb. A formulák viszont, mint pl. $(a + b)^2 = a^2 + 2ab + b^2$, vagy pl. $a + b < cd$, *állítanak* valamit, ami igaz is lehet, hamis is. (Például az utóbbi formula esetén attól függ, hogy igaz-e vagy nem, hogy milyen értékeket adunk a-nak, b-nek, c-nek és d-nek.) Ebből a szempontból tehát olyanféle különbség van a kifejezések és a formulák között, mint a természetes nyelv *főnévi szintagmái* és *mondat-szintagmái* között. A relációk jelei, mint „=", „<", megint más szintagmafajt alkotnak; a logikai műveletek jelei, mint „∧" „∨", „→", ismét másikat, a

kvantorok (∀, ∃) ismét másikat. Azt már nehezebb megmondani, hogy ezek a természetes nyelv mely szintagmafajainak felelnek meg; pl. az „=" jellel jelölt relációt a magyar és a német melléknévvel fejezi ki *(egyenlő, gleich)*, az angol azonban igével *(equals)*. E szavak mondattani szerepe mindkét esetben állítmány (névszói, illetőleg igei állítmány).

Egy-egy következtetésmód helyessége a premisszák és a konklúzió szerkezetén múlik. Például a kategorikus szillogizmus azon fajtája esetén, amelynek a skolasztikus logika a *Barbara* mnemotechnikai nevet adta (példa rá: minden élőlény halandó, minden ember élőlény, tehát minden ember halandó) az első premissza alanyának meg kell egyeznie a második premissza állítmányával (úgynevezett középfogalom), a konklúzió alanyának a második premissza alanyával, állítmányának pedig az első premissza állítmányával. A matematikai logikus sem kerülhette el, hogy a premisszák és a konklúzió szerkezetét elemezze. Ez tehát ismét nyelvi, morfológiai és szintaktikus problémákat vetett fel a matematikus formulanyelveiben: milyen morfémákból, hogyan épülnek fel a kifejezések, ezekből a formulák, a formulákból összetettebb formulák (sőt, esetleg ismét kifejezések). Például az $a + b$ kifejezés az a, b és a $+$ morfémákból épül fel; az első kettő maga is egyszerű kifejezés, a $+$-jel azonban magában nem. Sőt, helyesírási, jobban mondva: a helyes interpunkcióra vonatkozó kérdések is merülnek fel a matematikus formulanyelveiben; pl. a zárójelek e nyelvben az interpunkciós jelek szerepét töltik be, helyes használatuk, amely a kifejezések és a formulák jelentésének egyértelműségét biztosítja, a matematikus formulanyelveinek grammatikájához tartozik.

Annak a kérdésnek a tisztázásához, melyek az összes helyes következtetésmódok, elegendő volt a premisszák és a konklúzió szerepét játszó formulák, valamint a részeikként fellépő egyszerűbb formulák és kifejezések morfológiai és szintaktikus szerkezetét figyelembe venni. Ez azzal függ össze, hogy erre a kérdésre pusztán formális logikai eszközökkel választ lehet adni. Persze a matematikai logika válaszát nem úgy kell elképzelni, hogy azt mondja: a helyes következtetésmódok ezek és ezek, és felsorolja mindet. Ezt nem is lehet várni, hiszen végtelen sok helyes következtetésmód van. Azonban a matematikus tudja, hogyan lehet egy halmaz elemeit áttekinteni akkor is, ha végtelen sokan vannak. Arra törekszik, hogy olyan algoritmust találjon, amelynek segítségével bármely következtetésmódról el lehet dönteni, helyes-e. Ez csak a legegyszerűbb fajta következtetésmódok esetén sikerül, bár ezek a legegyszerűbb fajták is lényegesen túlmennek az arisztotelészi szillogizmusok, valamint a hipotetikus és diszjunktív következtetések körén. A bonyolultabb fajta következtetésmódok esetén pedig azzal is megelégszik a matematikus, hogy olyan algoritmust találjon, amely sorra szolgáltatja (generálja) az összes érvényes következtetésmódokat, olyanféleképpen, mint ahogy az az algoritmus, amely abban áll, hogy az 1-ből indulunk ki és mindig hozzáadunk 1-et a már megkapott számhoz, sorra szolgáltatja az összes pozitív egész számot. Ugyanis már ilyen algoritmus is elég ahhoz, hogy a matematikus hozzáfogjon axiómarendszerek ellentmondástalanságának vizsgálatához és más olyan vizsgálatokhoz, amelyekhez előzőleg tisztáznia kell, melyek az összes helyes következtetésmódok.

A matematikai logikus egyébként, a nyelvtudóstól eltérően, a morfológiát is hozzászámítja a szintaxishoz, ami meg is felel a *συντάττειν* szó eredeti jelentésének, hiszen ez éppúgy jelentheti szóalakok összeállítását morfémákból, mint mondatok összeállítását szóalakokból. Ennek megfelelően a matematikai logikus az említett tényt így fejezi ki: valamely következtetésmódnak az a tulajdonsága, hogy helyes-e, pusztán *szintaktikus tulajdonság*, tudniillik a következtetésmód premisszái és konklúziója szerepét játszó formulák szintaktikus szerkezetén múlik. Ez azt a tényt is magában foglalja, hogy annak tisztázásához, érvényes-e egy következtetésmód, nem szükséges e formulák *jelentését* figyelembe venni. Más kérdések vizsgálatában viszont nagyon is fontos tudni, mit jelentenek a formulák s azok részei, tehát *szemantikai kérdések* is felmerülnek a matematikai logikában.

18. Persze a matematikus eleinte nem volt tudatában annak, hogy amikor különbséget tesz kifejezés és formula között (ezt a különbséget már régen megtette, a matematikai logika kialakulása előtt is), akkor nyelvi kérdést, a szófajok kérdését tisztázta a maga formulanyelvei számára. Először Carnap osztrák matematikai logikus és neopozitivista filozófus tudatosította a matematikai logika bizonyos problémáinak nyelvi jellegét. Ő terjesztette el a *szintaktikus* és *szemantikai* jelzők használatát a matematikai logika szakkifejezéseiként, bár a szintaktikus és szemantikai tulajdonságok közötti különbségtétel régebbi: Hilbert már megtette ezt a különbséget, amikor *formálisról* és *tartalmiról* beszélt.

Carnap a matematika formulanyelvének, pontosabban, a matematika különböző fejezetei formulanyelvének strukturális, főként szintaktikus vizsgálatából igyekezett általános elméletet is formálni. Evégett definícióit úgy fogalmazta meg, hogy azok mindenféle nyelvre alkalmazhatók legyenek. E definícióiról azonban látszik, hogy elsősorban olyanféle nyelvekre gondol, mint a matematika különböző fejezeteinek formulanyelve, a természetes nyelvekre csak erőltetetten lehet azokat alkalmazni. Carnap gondolatait egyébként a matematikai logikusok nagy része — nem utolsósorban neopozitivista nézetei miatt — kétkedéssel fogadta; jelentőségüket csak akkor ismerték el általánosan, amikor a *programvezérlésű*, elektromechanikus, majd elektronikus *számológépek* megjelenése újabb nyelvi problémák elé állította a matematikust, s Carnap definíciói és megállapításai e problémák megoldására hasznosnak bizonyultak.

E gyorsműködésű számológépeket ugyanis, éppen gyors működésük miatt, nem lehet úgy használni, mint az asztali számológépet: nem lehet megvárni, amíg valamilyen műveletet elvégez a gép s azután eldönteni, esetleg az eredménytől függően, hogy milyen számokon milyen műveletet végezzen el legközelebb. Hiszen a részleteredménynek a gép kezelőjével való közlése (pl. papírra való leírás útján), főleg pedig a gép kezelőjének döntése, lényegesen tovább tartana, mint maga a művelet elvégzése, úgy, hogy a gép gyorsasága kárba veszne. Hanem előre ki kell dolgoznia a matematikusnak a számítás egész menetét, minden eshetőségre gondolva, beleértve az olyanféle döntéseket is, amelyek azt kívánják, hogy a gép aszerint folytassa a számítást egyféleképpen vagy máshogyan, hogy bizo-

nyos előzőleg kiszámított részleteredmények között fennáll-e valamilyen reláció (pl. az egyik kisebb-e a másiknál), vagy nem, meg azt is beleértve, hogy mely műveletek eredményét írja le a gép, mint végeredményt. Az így keletkező programot aztán közölni kell a géppel, a számításhoz szükséges kezdeti számadatokkal együtt. A gép a program utasításait és a kezdeti adatokat „memóriájában" tárolja, míg felhasználásra nem kerülnek; ugyanitt helyezi el a később még felhasználásra kerülő részleteredményeket, addig, míg fel nem kell használnia azokat valamely további műveletnél.

Az azonban, hogy milyen műveletekre lehet utasítani a gépet, a gép műszaki felépítésétől függ. Ez utasítások rendszere a gép „anyanyelve"; ilyen utasításokból kell összeállítani a programot. És viszont, úgy kell a gépet felépíteni, hogy az anyanyelvében szereplő fajta utasításokból mindenféle számítás algoritmusát, sőt, esetleg másféle algoritmust is, pl. egyik nyelvről a másikra való fordításhoz szükséges teendőket, össze lehessen állítani. A gép ezen univerzalitását azzal lehet elérni, hogy a gép anyanyelvét ügyesen állítjuk össze, pl. beleveszünk a *számtani alapműveletekre* vonatkozó utasításokon kívül olyanokat is, amelyek lehetővé teszik a program egyes részeinek *átugrását*, vagy *visszaugrást* a program már végrehajtott részére avégett, hogy újra végrehajtsuk, továbbá olyanokat is, amelyek előzőleg elvégzett műveletek eredményeinek, vagy kezdeti adatoknak, *összehasonlítását* kívánják (pl. abból a szempontból, megegyeznek-e, vagy hogy melyik a nagyobb), és az összeállítás eredményétől teszik függővé, a program melyik utasításához ugorjon a gép, valamint olyanokat is, amelyek arra vonatkoznak, hogy a program valamelyik *utasítását módosítsa* a gép, mielőtt végrehajtaná, vagy mielőtt visszaugrana rá stb. Természetesen mindenfajta, a gép anyanyelvébe felveendő utasítás újabb műszaki problémák megoldását kívánja a gép tervezőjétől.

Az egyik gép anyanyelvén írt programot a másik, más módon felépített gép nem érti meg, hacsak (e másik gép anyanyelvén írt úgynevezett *fordítóprogramnak* a memóriájában való elhelyezésével) „meg nem tanítjuk" arra, hogy hogyan fordítsa le azt a saját anyanyelvére. Számológépközi nyelveket, gépi eszperantót, is kidolgoztak a matematikusok, amelyeket a program összeállítója könnyebben meg tud tanulni, mint az egyes számológépeknek (a műszaki lehetőségek általd eterminált) anyanyelvét, s amelyről, megfelelő fordítóprogramok (úgynevezett *programozó programok*) segítségével, minden egyes számológép a maga anyanyelvére le tudja fordítani a programot. Ilyen gépi eszperantó pl. az ALGOL 60 (< 1960-ban kidolgozott **ALGO**rithmic **L**anguage). Ezt nemcsak arra használják, hogy különböző számológépek számára írjanak programokat, amelyeket azok a maguk anyanyelvére lefordítanak, hanem arra is, hogy az egyik gép kezelője a másik gép kezelőjével tömör alakban közölje, milyen algoritmust alkalmazott valamilyen számítási feladat megoldására.

Az ilyen gépi és gépközi nyelvek már nem olyan egyoldalúak, mint a matematikai formulanyelvek: nemcsak kijelentő mondataik vannak, mint a matematikai formulanyelveknek (amelyek matematikai tételek és bizonyításaik kifejezésére szolgálnak), hanem felszólító mondataik is (a gépnek adandó utasítások). Sőt, főleg ilyenekből állnak; azonban a felszólító mondatok részeiként kijelentő és kérdő mellékmondatok is előfordulnak

bennük. (Ilyenféle módon: „ha $a < b$, akkor vond ki b-ből a-t és az eredményt tedd abba a memóriarekeszbe, ahol eddig c állt; ellenkező esetben a-ból vond ki b-t és az eredményt tedd el abba a memóriarekeszbe, ahol eddig d állt", vagy, kissé módosított ALGOL 60 nyelven: if $a < b$ then $b - a \Rightarrow c$, else $a - b \Rightarrow d$. Itt az $a < b$ feltétel magában kijelentő mondat, amely azonban az egész felszólító mondatban mellékmondat szerepét játssza.)

Az ilyen gépi és gépközi nyelvekkel kapcsolatban újabb nyelvi problémák léptek fel; így pl. nyelvpedagógiai kérdések (egy koppenhágai számítástechnikai központban a nyelvpedagógia tapasztalatainak felhasználásával tanítják meg a programozó matematikusokat az ALGOL 60-ra), továbbá, a természetes nyelvekről természetes nyelvekre való fordítás kérdésétől függetlenül is, fordításelméleti problémák. (Az egyik gépi nyelvről a másikra való gépi fordítás is fordításelméleti problémák megoldását kívánja, csak egyszerűbb problémákét, mint az egyik köznapi nyelvről a másikra való gépi fordítás, mert a gépi nyelvek szabályosabbak, mint a köznapi nyelvek.)

A nyelvtudóst bizonyára megbotránkoztatja, hogy a matematikus formulanyelvről, gépi nyelvről beszél; ez azonban a matematikusnak abból a módszeréből ered, hogy különböző jelenségekben a közöset, az általánosat keresi s evégett a különbségektől (legalább egyelőre) eltekint. Márpedig az eddigiekből világos, hogy azokban a jelrendszerekben, amikre a matematikus gondol, amikor formulanyelvről beszél, van valami közös az igazi nyelvekkel és sem a matematikus, sem a nyelvtudós szempontjából nem kárba veszett fáradság ezt a közöset közelebbről megvizsgálni.

19. A fentiekben, egyes halmazelméleti módszerek illusztrálására, megemlítettem Kulagina néhány gondolatát, amelyeket szófaj- és szintagmafajelméletében alkalmazott. Kulagina elméletének, valamint a hozzá csatlakozó további matematikai nyelvészeti kutatásoknak rendszeres ismertetése azonban nem az én feladatom, ezekkel más előadások foglalkoznak majd ezen a munkaértekezleten.

Mindazonáltal szeretném még felvázolni annak az útnak néhány állomását, amely Kulagina elméletéhez vezetett. Mint már említettem, a matematikus már réges-régen találkozott szófaj-, illetőleg szintagmafajproblémákkal a matematikai formulanyelvekben, és Carnap már tudatosan nyelvi problémaként fogalmazta meg ezeket. Carnap vizsgálataiból világos, hogy a matematika formulanyelveinek szintagmafajai a szintagmák jelentésére való hivatkozás nélkül, tisztán szintaktikusan definiálhatók. Carnap szintagmafaj-definíciójában is már fontos szerepe van annak a kérdésnek, hogy a formulanyelv szintaxisa szerint helyesen felépített formulák mely részei pótolhatók más formula-részekkel úgy, hogy ismét helyesen felépített formulát kapjunk. Ez nem meglepő, hiszen formulák részeinek más formula-részekkel való pótlása, pl. valamely kifejezés behelyettesítése valamely betű (változó) helyébe, vagy valamely egyenlőség egyik oldalának pótlása a másikkal, mindig fontos szerepet játszott a matematikai bizonyításokban. Például ilyen helyettesítéssel (b helyébe $b + c$ behelyettesítésével) kapjuk a már sokszor említett

$$(a + b)^2 = a^2 + 2ab + b^2 \tag{8}$$

azonosságból a bonyolultabb

$$(a + b + c)^2 = a^2 + 2a(b + c) + (b + c)^2 \tag{9}$$

azonosságot, b helyébe c és a helyébe b behelyettesítésével pedig a

$$(b + c)^2 = b^2 + 2bc + c^2 \tag{10}$$

azonosságot (amely ugyanazt fejezi ki, mint a (8) azonosság, csak más jelölésekkel). Ha a (9) azonosság jobboldalán a $(b + c)^2$ kifejezést pótoljuk a ((10) azonosság szerint vele mindig egyenlő értékű) $b^2 + 2bc + c^2$ kifejezéssel, a $2a(b + c)$ kifejezést pedig (a szorzás disztributív tulajdonsága folytán vele mindig egyenlő értékű) $2ab + 2ac$ kifejezéssel, az

$$(a + b + c)^2 = a^2 + 2ab + 2ac + b^2 + 2bc + c^2 \tag{11}$$

azonossághoz jutunk. A (11) azonosság nemcsak bonyolultabb, de általánosabb is, mint a kiindulásul választott (8) azonosság, hiszen ez a (11) azonosság azon speciális esetének felel meg, amikor $c = 0$. Ez a példa azt is mutatja, hogy változó helyébe kifejezés helyettesítésével, és valamely azonosság egyik oldalának másik oldalával való pótlásával, valamely matematikai tételből *általánosabb* tételhez is juthatunk, nemcsak a tétel speciális esetéhez, mint pl. ha a (8) azonosságból, b helyébe 1 helyettesítésével, arra következtetünk, hogy

$$(a + 1)^2 = a^2 + 2a + 1.$$

Azt is mondottam, Carnap megpróbálta definícióit olyan általános alakban fogalmazni, hogy azokat tetszőleges nyelvekre lehessen alkalmazni. Az igazi természetes nyelvekre való alkalmazásuk azonban nem vezet nyelvtudományi szempontból kielégítő szófaj- vagy szintagmafaj-elmélethez, hanem sokkal inkább a matematikai formulanyelvek kategóriáinak a természetes nyelvekre való ráhúzását jelenti.

Márpedig a szófajok és a szintagmafajok problémája a természetes nyelvek szempontjából is fontos, korántsem kielégítően megoldott probléma. A szófajok iskolás definíciója a szavak jelentésére hivatkozik, pl. „a főnév személyt vagy tárgyat jelentő szó", „a melléknév tulajdonságot jelentő szó", „az ige cselekvést, történést vagy állapotot jelentő szó" stb. Világos, hogy ezek a definíciók nyelvtudományi szempontból sem tekinthetők kielégítőknek. Hiszen pl. nehéz megmondani, hol végződik a tulajdonság, és hol kezdődik az állapot, semmi esetre sincs közöttük olyan éles határ, mint a melléknév és ige között. A jelentés nem magyarázza meg, hogy pl. miért fejezi ki a latin ugyanazt, amit a magyar az *egészséges* melléknévvel fejez ki, a *valeo* igével. Vagy mit szóljunk a jelentéstan szempontjából ahhoz, hogy a matematikus ahelyett, hogy egy függvény értéke minden helyen *nagyobb* (középfokú melléknév!), mint egy másiké, hol azt mondja, hogy az előbbi függvény *majorálja* (ige!) a másikat, hol azt, hogy *majoránsa* (főnévként használt melléknévi igenév!) annak, aszerint, ahogy a mondatba jobban beleillik?! (Hogy a nyelvész is megértse e mondatok értelmét, megmondom, hogy *hely* e vonatkozás-

ban pl. valós számot jelent, valamely függvény *értékén* egy helyen pedig olyan függő változónak azt az értékét értjük, amelyet akkor vesz fel, amikor a független változó értéke a kérdéses valós szám, amely függő változónak a független változótól való függését a szóban forgó függvény adja meg. Például ha a függvény az a mód, ahogyan egy szám négyzete függ e számtól, akkor e függvény értéke a 3 helyen 3^2, vagyis 9.)

Ma már a nyelvtudósok is hajlanak arra, hogy a szófajok közötti különbséget ne a jelentéstani különbségben keressék, hanem abban, ami a különböző fajta szavaknak, illetőleg alakjaiknak a mondatban játszott különböző szerepében mutatkozik meg (pl. a *valeo* igét másképpen kell használni ahhoz, hogy a latin mondatba beleilljék, mint a magyarban az *egészséges* főnevet). Más szóval: a szófajok hagyományos szemantikai definíciója helyett szintaktikus definíciót keresnek.

Ebben az irányban haladt a matematikai nyelvészek törekvése is. Nemcsak azért, mert a matematikai formulanyelvekhez hasonló jelrendszerek esetén Carnapnak már sikerült szintaktikus szófajdefiníciót adnia, hanem azért is, mert a gépi fordításra való alkalmazásokat tartották szem előtt. Márpedig ezek, mint már mondottam, azt kívánják, hogy igyekezzünk a szemantikai fogalmakat kiküszöbölni, szintaktikus fogalmakra visszavezetni, mert a gépet sok mindenre „meg lehet tanítani", csak arra nem, hogy megértse azt a szöveget, amelyet feldolgoz. A gépi fordításra való alkalmazások egyúttal, hacsak lehet, még nagyobb szabatosságot igényelnek, mint a nyelvtudomány szempontjai, hiszen a gép csak azt „érti meg", amit teljesen szabatosan közlünk vele. Ez magyarázza meg — a matematikus igen magas fokú igénye mellett a szabatosság tekintetében —, hogy a matematikai nyelvészek fő törekvése e tekintetben az, hogy teljesen szabatos szófaj-, illetőleg szintagmafaj-fogalomhoz jussanak. Ha a szabatosság érdekében fel kellett áldozni valamit, akkor a matematikai nyelvész könnyebben áldozza fel a hagyományos szófajokkal, illetőleg szintagmafajokkal való pontos megegyezést, mint a nyelvtudós. Elvégre a gépi fordítás szempontjából nem nagy baj, ha a szavakat és a szintagmákat a hagyománytól eltérő módon osztjuk fel fajokra, ha az új felosztást is lehet alkalmazni pl. a mondatok elemzésére. Legfeljebb a gépi fordítás célját szolgáló szótárban kell másféle szófajutalást adni, mint a forgalomban levő szótárakban, ez azonban nem baj, hiszen a gépi fordítás céljára úgyis speciális szótárt (jobban mondva morféma- és idiómatárt) kell összeállítani.

A hagyományos szófajoktól (illetőleg szintagmafajoktól) való eltérések miatt a matematikai nyelvészek *szófaj* (illetőleg *szintagmafaj*) helyett rendszerint más szakkifejezéseket használnak. Így pl. Bar-Hillel-szintagmafajok helyett *szintaktikus kategóriákról* beszél, Lambek pedig (minthogy az általa definiált fogalom nemcsak a hagyományos szintagmafaj-fogalmat nem fedi, hanem Bar-Hillel szintaktikus kategória-fogalmát sem teljesen) *szintaktikus típusokról*. Egyébként Kulagina is *típusokról* beszél szófajok helyett.

20. Minthogy a szintaktikus típusok Lambek-féle elmélete a szintaktikus kategóriák Bar-Hillel-féle elméletének továbbfejlesztése, csak Lambek

elméletét fogom röviden ismertetni.[5] Lambek két alapvető szintaktikus típusból indul ki: a (teljes, vagyis nem hiányos) kijelentő mondatok szintaktikus típusát *s*-sel jelöli, a tulajdonnevek egyes számú alanyesetének szintaktikus típusát pedig *n*-nel. Ha két szintagma olyan, hogy bármely kijelentő mondatban bármelyiküket a másikkal pótolva ismét (grammatikailag helyes) kijelentő mondatot kapunk, akkor ugyanazon szintaktikus típusba sorolja őket. Így pl. *n* a szintaktikus típusa az olyan szintagmáknak is, amelyek kijelentő mondatokban pótolhatják tulajdonnév egyes számú alanyesetét és viszont. Ilyen pl. a jelzős tulajdonnevek egyes számú alanyesete. (Például a *János dolgozik* mondatban a *János* szóalakot a *szegény János* szintagmával pótolva a *Szegény János dolgozik* grammatikailag helyesen képezett kijelentő mondatot kapjuk.)

A *dolgozik* igealakhoz és általában minden olyan szintagmához, amelyet bármely egyes számú alanyesetben álló tulajdonnév vagy más, az *n* szintaktikus típushoz tartozó szintagma után téve kijelentő mondatot kapunk, az $n\backslash s$ (olvasd: *n* felett *s*) szintaktikus típust rendeli hozzá Lambek. Általában, ha *x* és *y* olyan kifejezések, amelyek szintaktikus típusokat jelölnek, akkor $x\backslash y$ (olvasd: *x* felett *y*) az olyan szintagmák szintaktikus típusának jele, amelyeket *x* szintaktikus típusú szintagma után téve *y* szintaktikus típusú szintagmát kapunk. Az absztrakt algebra egy szakkifejezését alkalmazva azt mondhatjuk, hogy az $x\backslash y$ szintaktikus típusú szintagmák olyan *jobboperátorok*, amelyek az *x* szintaktikus típusú szintagmákat *y* szintaktikus típusú szintagmákba viszik át. Így minden alanyi ragozású, jelentő mód jelen vagy múlt idő egyes szám harmadik személyű igealak pl. *alszik*, *eszik*, *aludt*, *evett*, vagy bővítménnyel ellátott ilyen igealak, pl. a *jól alszik*, *almát eszik* szintagmák, szintén $n\backslash s$ típusúak, mert *n* típusú szintagmákat *s* típusú szintagmákba átvivő jobboperátorok. (Például *János jól dolgozik*, *János almát eszik* kijelentő mondatok.)

De ez azt is mutatja, hogy az olyan bővítményként alkalmazható szóalakok, mint *jól*, *almát*, az $n\backslash s$ típusú szintagmákat $n\backslash s$ típusú szintagmákba viszik át, ha eléjük tesszük őket, az absztrakt algebra szakkifejezésével: $n\backslash s$ típusú szintagmákat $n\backslash s$ típusú szintagmákba átvivő *baloperátorok*. Ezek szintaktikus típusát Lambek így jelöli: $(n\backslash s)/(n\backslash s)$. Általában x/y (olvasd: *x* alatt *y*) az *y* típusú szintagmákat az *x* típusú szintagmákba átvivő baloperátorok szintaktikus típusának jele, vagyis azoké a szintagmáké, amelyeket *y* típusú szintagmák elé téve *x* típusú szintagmát kapunk. (Ha x/y-ban, vagy $x\backslash y$-ban, *x* vagy *y* nem alapvető szintaktikus típus, akkor egyértelműség kedvéért zárójelbe tesszük.) Így a *szegény* melléknévalak, általában minden melléknév egyes számú alanyesete, n/n típusú, mert *n* típusú szintagmákat *n* típusú szintagmákba átvivő baloperátor.

Persze a tulajdonneveket és általában az *n* típusú szintagmákat tekinthetjük olyan baloperátoroknak is, amelyek $n\backslash s$ típusú szintagmát (pl. *dolgozik*) *s* típusúba (kijelentő mondatba) visznek át, s így $s/(n\backslash s)$ is kifejezi ezek szintaktikus típusát. Így egy és ugyanazon szintagmához különböző szintaktikus típusok is tartozhatnak.

[5] J. Lambek, The mathematics of sentence structure: American Mathematical Monthly 65. 154—70 (1958).

Az is megeshetik, hogy valamely szóalak (vagy szintagma) különböző módon való alkalmazásában kap más és más szintaktikus típust. Például ha az *és* kötőszó két tulajdonnév között áll, újabb *n* típusú szintagmát kapunk (tudniillik a magyarban pl. *János és Péter dolgozik* grammatikailag helyes kijelentő mondat). Így az *és Péter* szókapcsolat *n* típusú szintagmát (*János*) *n* típusú szintagmába (*János és Péter*) átvivő, tehát $n\backslash n$ típusú jobboperátornak tekinthető. Minthogy eszerint az *és* kötőszó az *n* típusú szintagmát (*Péter*) ilyen $n\backslash n$ típusú szókapcsolatba (*és Péter*) átvivő baloperátor, ezért szintaktikus típusa ilyen használata esetén $(n\backslash n)/n$. Ha viszont két kijelentő mondat között áll az *és* kötőszó (pl. *János dolgozik és Péter eszik*), akkor újabb kijelentő mondatot kapunk. Így *és Péter eszik* *s* típusú szintagmát (a *János dolgozik* kijelentő mondatot) *s* típusú szintagmába (a *János dolgozik és Péter eszik* kijelentő mondatba) átvivő jobboperátor, tehát szintaktikus típusa $s\backslash s$. Ennélfogva az *és* kötőszó mondatok között *s* típusú szintagmát (a *Péter eszik* kijelentő mondatot) $s\backslash s$ típusú szintagmába (az *és Péter eszik* szókapcsolatba) átvivő baloperátor, vagyis ilyen szerepében viszont $(s\backslash s)/s$ a szintaktikus típusa. (Lambek az olyanféle szókapcsolatokat is szintagmáknak tekinti, mint *és Péter*, vagy *és Péter eszik*.)

Lambeknek, mint már Bar-Hillelnek is, az a célja ezzel az elmélettel, hogy olyan algoritmust adjon, amellyel szóalakok bármely véges sorozatáról el lehet dönteni, grammatikailag helyesen képezett (teljes kijelentő) mondat-e. (Lényegében már Ajdukiewicz lengyel filozófus is erre törekedett.[6]) Ilyen algoritmust meg is ad, feltéve, hogy rendelkezésünkre áll olyan szótár, amely bármely szóalakhoz megadja annak valamennyi lehetséges szintaktikus típusát. Ilyen szótár összeállításához a fenti példákhoz hasonló további, kiterjedt elemzés szükséges. (Ilyen elemzés persze módosíthatja a fent megállapított szintaktikus típusokat is.) Szükséges továbbá az is, hogy, főleg a tulajdonneveknek az egyes számú alanyesettől különböző alakjai számára, további alapvető szintaktikus típusokat vezessünk be; erre Lambeknek a morfológiailag viszonylag szegény angol nyelvből vett példákban is szüksége volt (a többes számú alak miatt). További problémákat okoz a nem tulajdon-főnevek esetében a névelő szükségessége.

Akármennyire érdekes is Lambek algoritmusa matematikai szempontból, a gépi fordítás szempontjából nincs ilyen algoritmusra szükség, hiszen nyugodtan feltételezhetjük, hogy a forrásnyelvi szöveg mondatai mind grammatikailag helyesen képezett mondatok. (Nem arra való a gépi fordítás, hogy a gépet megpróbáljuk „becsapni".) Igaz, hogy Lambek algoritmusa, mint azt mondja is, felhasználható a mondat elemzése során is, arra pedig a gépi fordításhoz is szükség van. Nyelvelméleti szempontból is helyesebb azonban úgy eljárni, mint Kulagina teszi, hogy (a nyelv szóalak-készletén és minden egyes szó összes alakjainak halmazán kívül) adottnak tekinti a nyelv összes lehetséges (grammatikailag helyes) mondatainak halmazát, hiszen az (ha csak részben is) adott a kérdéses nyelv irodalmával, élő nyelv esetén pedig potenciálisan adott a kérdéses nyelvet anyanyelvük gyanánt beszélő emberek nyelvérzékével. (Ilyen nyelvérzék

[6] K. Ajdukiewicz, Die syntaktische Konnexität: Studia Philosophica 1. 1—27 (1935).

kifejlődik egyébként holt nyelv irodalmának tanulmányozása során is.) Emellett Kulagina annyiban is jobban megközelítette azt, amire a matematikai nyelvész e vonatkozásban törekszik, hogy a szófajokat a szintagmafajoktól függetlenül definiálta, hiszen a szintagmáknak a szóalakok speciális esetei ugyan, de a szavak nem.

21. Örülnék, ha a nyelvész kartársak a következő tanulságokat szűrnék le előadásomból.

A matematika évezredes fejlődése során, főként az algebra és a topológia újabb, a halmazelméleti módszerek által megtermékenyített fejlődésével, de magának a halmazelméletnek és a matematikai logikának kialakulásával is, sok olyan tapasztalatot halmozott fel bonyolult rendszerek szerkezetének vizsgálata terén, amelyek tanulságosak lehetnek a nyelvtudós számára a nyelv strukturális vizsgálata szempontjából is. A matematika ugyan eleinte egészen más természetű és sok szempontból egyszerűbb rendszerek szerkezetét vizsgálta, mint amilyen a nyelv, azonban fejlődése során, különösen a nem mennyiség jellegű elemekből álló rendszereknek a matematikai kutatás körébe való bevonásával, mind jobban közeledett a nyelvhez hasonló és azt bonyolultságban is megközelítő rendszerek vizsgálatához. Sőt, a matematikai logika és a számítástechnika olyan kérdések feltevéséhez is vezetett, amelyek nagyon hasonlítanak a nyelvtudomány által vizsgált kérdésekhez, ha nem is igazi nyelvekre vonatkoznak, hanem csak többé-kevésbé nyelvszerű jelrendszerekre. Ezeknek a kérdéseknek matematikai eszközökkel történt megoldásának módszere is tanulságos lehet a nyelvtudós számára.

Bár a matematika a maga módszereivel leginkább a szinkronikus, strukturális nyelvtudományi kérdésekre való alkalmazásra van jelenleg felkészülve, és a gépi fordítás is elsősorban ilyen kérdések vizsgálatát kívánja, hiszen rendszerint a mi korunkkal szinkrón szövegeket akarunk gépekkel lefordíttatni más nyelvre, a matematikai módszerek természete elvileg nem zárja ki a diakronikus vizsgálatokra való alkalmazásuk lehetőségét sem. Még kevésbé zárja ki a matematikai módszerek természete a szemantikai kérdésekre való alkalmazás lehetőségét, habár a matematikus gyakran törekszik, már csak a gépi fordításra való alkalmazás végett is, szemantikai fogalmaknak szintaktikus és morfológiai fogalmakra való visszavezetésére.

A matematikus ennek megfelelően semlegességet jelent be az olyan vitákban, amelyek a diakronikus és a szinkronikus, vagy a szemantikai és szintaktikus nyelvvizsgálati módszerek fontossági sorrendje körül folynak. Annak eldöntésére, hogy a nyelvtudomány mai állása és fejlődésének iránya elsősorban milyen jellegű nyelvtudományi kutatásokat kíván nyelvtudósainktól, továbbá, hogy melyek azok a kutatások, ahol — ha szabad Antal László kartárs szavait[7] használnom — egy jottányi engedményt sem lehet tenni sem a történetiségnek, sem a jelentésnek, a nyelvtudomány szakemberei hivatottak.

[7] Antal László hozzászólása, Vita a nyelvtudomány elvi kérdéseiről: A Magyar Tudományos Akadémia Nyelv- és Irodalomtudományi Osztályának Közleményei 18. 53—4 (1961).

Papp Ferenc

Nyelvi rendszer, közlési folyamat és ezek néhány matematikai modellje

1. Testszerű és mezőszerű anyag a nyelvben

Nemcsak a matematikának kellett hosszú utat bejárnia addig, míg olyan absztrakt formát ölthetett, mely a legkülönbözőbb tudományok bizonyos tartalmi kérdéseinek is megfelel. Az egyes konkrét tudományoknak is hosszú fejlődést kellett bejárniuk, amíg bennük a matematikai módszerek alkalmazásának igénye egyáltalán felmerülhetett. Így jól tudjuk, hogy a nyelvtudomány akkor vált a szó mai értelmében vett tudománnyá, amikor az egyes nyelvek testszerű anyagi elemeit kezdte vizsgálni: megállapította ezen elemek — hangok, szavak, szószerkezet-típusok stb. — történetét, ezzel párhuzamosan csodálatra méltó pontossággal rekonstruálta az egyes nyelvek — és természetesen az egyes nyelveket beszélő népek — legrégibb történetét, rokonságát. S csak amikor már igen gazdag ilyen jellegű anyag került felhalmozásra, nyílt mód arra, hogy ettől az anyagtól bizonyos értelemben elszakadjon s a testszerű elemek között fennálló, szintén anyagi természetű, objektív v i s z o n y o k a t vizsgálja. E szemlélet számára már nem az egyes testszerű részecskék belső felépítése az érdekes, sőt nem is a testszerű részecskék a maguk anyagi mivoltukban, hanem az a bonyolult viszonyhálózat, mely egy anyagi mezőt alkotva e részecskék között fennáll.

A múlt század nyelvtudományának hatalmas felfedezése az volt, hogy valamely nyelv egyes elemei nem „össze-vissza" változnak —, ami elvileg szintén elképzelhető lenne —, hanem szigorúan érvényesülő, kivételt alig ismerő törvények szerint. Századunk nyelvtudománya hasonlóan nagy horderejű és talán épp annyira meglepő eredményre jutott, amikor megállapította, hogy a sok-sok változás eredményeképpen, még hozzá a hangrendszerben, az alakrendszerben stb. végbement változások összességének eredményeképpen az egyes nyelvek egyes korokban nem valamilyen rendszertelen, össze-vissza káoszt, hanem szigorú szisztémát alkotnak. Ez a következtetés nem mond ellent a józan ész mindennapos tapasztalatának: ha ugyanis az egyes nyelveken nap mint nap rengeteg sok közlés megy végbe minden fennakadás nélkül, ha a beszélő minden pillanatban teljes biztonsággal meg tudja formálni mondanivalóját, s ezt a mondanivalót a címzett mindig meg is érti, akkor okvetlenül kell lennie valamilyen objektív, mind a hallgató, mind a beszélő tudatától független (és ugyanakkor mindkettőjük tudata által elsajátított) r e n d s z e r n e k, mely ezt a közlést lehetővé teszi; — éppenséggel rendszernek és nem össze-vissza-

ságnak, mert az utóbbi esetben a közlés lehetetlen volna. Ugyanakkor meglepő mégis, mert a tömegek által tömegesen, állandóan használt nyelvben állandóan igen sok véletlen apró változás megy végbe, s az ilyen tömeges apró változásokat a naiv tudat szívesen képzeli el rend nélkül valónak, azok eredményét tehát még inkább ilyennek. Mint a természet és társadalom minden más területén, itt is az a helyzet, hogy a rendszert, a törvényszerűséget nem mi visszük bele vizsgálati anyagunkban, hanem csak egyre jobban megismerjük, egyre pontosabban megfogalmazzuk az ott talált objektív törvényszerűségeket.

2. Modell és modellállás

A nyelvtudományban kialakult igénynek s a matematikában megszületett lehetőségnek szükségszerűen találkoznia kellett és találkozott is a matematikai nyelvészetben. A matematikai nyelvészet — legáltalánosabban szólva — modellálással foglalkozik. A strukturális nyelvészettől és a nyelvstatisztikától mint kísérletes tudományaitól bizonyos mérési eredményeket, tényanyagot kap; másrészt a modern matematikából ismer bizonyos sémákat, melyekről a matematikusok sok mindent tudnak — legalábbis többet, mint mi a frissen nyert eredményekről. Ezeket a sémákat nevezzük *modelleknek* s a velük kapcsolatos munkát *modellálásnak*. A modell ezek szerint a valóság tudatosan leegyszerűsített mintája, mely viszonylagos egyszerűségénél fogva lehetővé teszi, hogy a vizsgált jelenséget rajta tanulmányozzuk. A modellálás pedig ebből következően annyit jelent, hogy a matematikai nyelvészet, egyéb egzakt tudományokhoz hasonlóan, a feltárt tényanyag elemei közötti összefüggéseket igyekszik a matematika nyelvén több-kevesebb pontossággal kifejezni (vagyis modellt építeni); a munka következő fázisában e modellen bizonyos átalakításokat hajthat végre; végül az eredményképpen kapott modellt visszaalkalmazza a valóságra, megvizsgálja, mely tények felelnek meg, és melyek nem, a kapott modell előírta követelményeknek. A modellek építése, a modellen végrehajtott műveletek és a modellek visszaalkalmazása a valóságra — a modellálási munka három fő fázisa. (Megjegyzendő, nemcsak absztrakt-matematikai, hanem konkrét-fizikai modellek is elképzelhetők bizonyos, kutatási vagy szemléltető, céllal — ilyen értelemben egy kis repülő-modell is valóban modell a szó általunk használt értelmében.)

Kevés megállapítás váltott ki annyi félelmet, előítéletet, bizalmatlanságot, mint éppen az itt elmondott. A sematizmustól kezdve a durvaságig, a leegyszerűsítéstől kezdve addig, hogy a strukturális és a matematikai nyelvészet csak a „rendes" jelenségekkel foglalkozik és a minden nyelvre annyira jellemző kivételeket elhanyagolja — a legkülönfélébb vádak hangzottak el az ilyen módszerek alkalmazásával szemben. Pedig látnunk kell: a matematika annyira általános „sémákat" ad, hogy abba igazán minden belefér, másrészt oly végtelenül kiművelt formanyelvvel rendelkezik, hogy azzal a legárnyalatibb különbségeket is jól ki lehet fejezni a legapróbb részletekig. Továbbá: számos más tudomány döbbenetesen nagy eredményei éppen úgy születtek, hogy eleinte egyszerűbb, majd egyre pontosabb és „élethűbb" modellek készültek a vizsgált valóságról. A valóság

nem tudatos, hanem történelmileg szükségszerű, az ismeretek hiányából fakadó leegyszerűsítése is rendkívül hasznos lehet: korunk egyik legjelentősebb elmélete, az atom-elmélet máig is nevében hordoz egy korábbi, leegyszerűsített elméletet az anyagok oszthatatlan kis építőköveiről. Az alábbiak talán némileg meg fogják világítani azt, hogy a nyelv és a nyelvi jelenségek modellálása és ezen az úton való megismerése — úgy is mondhatnám: tervszerűen fokozatos megismerése — sem kivétel e folyamat törvényszerűsége alól.[1]

Az általunk alkalmazott modelleknek két fő formája képzelhető el jelenleg: a statisztikai modellek és a strukturális modellek. Az előbbiek esetén

1. táblázat

Statisztikai és strukturális modellek a nyelvrendszer és a közlési folyamat különböző szintjein

		Strukturális modell	Statisztikai modell
1. fonéma	rendszer	*1.1.* véges ponthalmaz n-dimenziós térben	*1.2.* fonémastatisztika
	folyamat	*1.3.* szótag-típusok	*1.4.* szótagtípusok statisztikája
2. morféma	rendszer	*2.1.* morfémaosztály-elmélet	*2.2.* morféma-megterhelés
	folyamat	*2.3.* morfológia	*2.4.* morféma típus-példány viszony
3. szó	rendszer	*3.1.* gráfelméleti modell	*3.2.* gyakorisági szótár
	folyamat	*3.3.* szó típus-példány viszony	*3.4.* szó típus-példány viszony
4. szintagma	rendszer	*4.1.* szintagma-elmélet	*4.2.* szintagma-statisztika
	folyamat	*4.3.* két dimenziós szerkezet egy dimenziós leképezése	*4.4.* szintagma típus-példány viszony

Megjegyzések a táblázathoz. A táblázatban szereplő kurzív számok a szöveg **3.** pontjának alpontjaira utalnak. A táblázatban általában azt tüntettük fel, hogy milyen jellegű kutatásokra épülhetnek a megfelelő modellek, egy esetben azonban **(3.1.)** magát az elképzelhető modellt jeleztük.

a matematikai statisztika fegyvertárába nyúlunk, az utóbbiak esetén a halmazelméletbe. Másrészt akár az első, akár a második fajta modellről van szó, ezeket egyaránt alkalmazhatjuk a nyelvi rendszerre vagy a közlés

[1] Modelleket egyébként eddig is használt a nyelvtudomány: az a l a p n y e l v a maga csillagos formarendszerével a valóságnak többé-kevésbé hű megközelítése, m o d e l l j e .

folyamatára. Tárgyunknak ilyen felosztására a továbbiakban fog talán némi fény derülni. Számunkra elsődlegesen a közlési folyamat van adva. Ezt a folyamatot elemezve felfedezhetjük azt, hogy a közlési folyamat több szintre bontható, s minden egyes szinten megállapíthatók az azt alkotó elemi egységek — a fonémák, a morfémák, a szintagmák. Az ilyen módon feltárt elemeket szintenként rendbe rakhatjuk, s ehhez a strukturális modelleket használhatjuk fel; másrészt bizonyos mennyiségi vizsgálatokat végezhetünk rajtuk, — s itt jönnek segítségünkre a statisztikai modellek. Majd az így nyert s rendszerükben ismert elemek birtokában visszafordulhatunk magához a közlési folyamathoz és megvizsgálhatjuk, milyen szerkezete van a közlési folyamatnak, illetőleg milyen statisztikai-valószínűségi törvényszerűségek lelhetők fel benne. Mivel a nyelv objektíve így oszlik fel, természetes, hogy a vele foglalkozó tudománynak is ilyen legyen a felosztása: az alábbiakban, amikor az egyes szinteken megállapítható nyelvrendszeri, illetőleg közlés-folyamati strukturális és statisztikai modelleket bemutatjuk, ezzel egyben bemutattuk a matematikai nyelvészet felosztását is.

3. Kísérlet a nyelvi modellek osztályozására és csoportosítására*

3.1. A nyelv első szintje, mely elég hosszú ideig a legteljesebben és a legrészletesebben került strukturális és statisztikai kidolgozásra egyaránt, a nyelv hangtestének a szintje. E szint alapegysége a fonéma.

3.1.1. A fonémarendszer strukturális modelljét a következőkben jellemezhetjük.[2] Van egy véges — mindössze néhány tucat — elemet tartalmazó ponthalmazunk: az egyes pontok képviselik az egyes fonémákat. Ezt a ponthalmazt egy több dimenziós térré tehetjük úgy, hogy az egyes dimenziókat a fonémák között meglevő oppozíciók nyújtják; végeredményképpen tehát annyi dimenziós a terünk, ahány oppozíció az adott nyelvben van. Válasszuk ki például az egyik legáltalánosabb oppozíciót: magánhangzó-mássalhangzó. E szempont szerint ponthalmazunk két, vagy inkább három, közös elemet nem tartalmazó részhalmazra bontható: valami vagy magánhangzó (ezt a minőséget jelöljük, mondjuk, +1-gyel), vagy mássalhangzó (ez legyen, mondjuk, —1), vagy pedig irreveláns számára a mássalhangzó-magánhangzó oppozíció — jelöljük ezt 0-val. Hasonló módon felvehetjük a többi fonéma-oppozíciót is és ezek szerint tovább osztályozhatjuk ponthalmazunkat. Végeredményképpen a pontokat egy-egy szám n-essel írhatjuk le: n darab plusz eggyel, mínusz eggyel, illetve nullával. Az egyes nyelvek ezen a szinten különbözhetnek egymástól mindjárt abban, hogy nem azonos számú ponttal — fonémával — rendelkeznek, ugyanakkor abban is, hogy ezeket a pontokat más számú vagy- és más jelentőségű dimenziók (oppozíciók) szerint kell osztályozni.

Vegyünk egy egészen triviális példát. A magyar *b* fonéma és az orosz *b* fonéma, bár realizációik fizikai-eszközfonetikai szemszögből gyakorlatilag azonosnak tekinthetők, a rendszer szempontjából különbözik egymástól.

[2] Vö. például: C. Cherry: On human communication (Cambridge, Mass., 1957) 90 kk.

* Lásd az 1. táblázatot a 77. lapon.

Mert bár mértékszámaik sorra-rendre közösnek bizonyulnak — mind a kettő mássalhangzó, mind a kettő bilabiális, mind a kettő zöngés, mind a kettő explozíva, tehát ezen dimenziókban mind a ketten ugyanazt a plusz egyet vagy mínusz egyet kapják, attól függően csupán, hogy mit nevezünk plusz és mit mínusz egynek, — végül kiugrik egy eltérő komponensük. Az orosz *b* fonéma számára releváns a kemény-lágy oppozíció (ő kemény — legyen ez +1) — a magyar számára viszont ez az oppozíció irreleváns (a magyar *b* fonéma ebben a dimenzióban 0 értéket vesz fel). Az előbb felsorolt sajátságokat jelölve tehát, az orosz *b*-t, mondjuk, a következő számsor határozza meg: —1, +1, +1, +1, +1, a magyar *b*-t meg: —1, +1, +1, +1, 0.

Az így nyert n-dimenziós modellen számos vizsgálatot végezhetünk el. Megvizsgálhatjuk mindjárt az egész rendszert: mennyire gazdaságos, mennyire csekély számú oppozícióban tud bizonyos mennyiségű pontot elhelyezni, és fordítva. Megvizsgálhatjuk n-dimenziós terünk egyes részeit: kiragadhatunk egyes nagyon „népszerű" oppozíciókat (ahol a legkevesebb pont koordinátája 0) és fordítva: egyes pontokról megállapíthatjuk, hogy terünk egészen különálló részében foglalnak helyet (elszigeteltek), másoknak viszont sok komponensük egyezik. E vizsgálatokból következtetéseket vonhatunk le az egyes fonémák realizációinak érthetőségétől kezdve az egész fonémarendszer „gyenge pontjai"-nak megállapításáig (vagyis azon pontok meghatározásáig, ahol e statikus rendszer legkönnyebben megtámadható, megváltoztatható, illetőleg, ahol ilyen változás spontán bekövetkezése a legvárhatóbb), a nyelvoktatás igényeitől kezdve a fonémák és fonémarendszerek absztrakt sajátságainak megállapításáig. Az egyes nyelvek fonémarendszereinek n-dimenziós modelljeit is összehasonlíthatjuk, s így az e szinten mért tipológiai hasonlóság pontos fokozatait állapíthatjuk meg; — ezen absztrakt megállapítás azután az élet számos területén újból alkalmazható.

3.1.2. Ám a fonémarendszer strukturális modellje a valóság számos vonását elhanyagolja. Egy ilyen, nála elhanyagolt vonás-komplexumot tudunk tükröztetni, ha felállítjuk a fonémarendszer statisztikai modelljét is, ahol immár főleg az egyes fonémák megterheléséről (gyakoriságáról) lesz szó. Nálunk ilyen modellekhez igen értékes anyagot adnak Tarnóczy Tamás, Vértes Edit, Fónagy Iván kutatásai.[3] Az előbb nyert strukturális modell és e statisztikai modell összekapcsolása azután már elég pontos képet ad valamely nyelv e téren észlelt jelenségeiről.

3.1.3.—1.4. Amikor a közlési folyamat fonéma-szintű strukturális modelljét felállítjuk, lényegében az egyes nyelvekre jellemző szótag-típusokat rögzítjük rendszerbe; ennek megfelelő statisztikai modellt is építhetünk.

[3] Tarnóczy Tamás, A magyar magánhangzók eloszlása: Magyar Nyelv 38. 352 kk. (1942); Megjegyzések és kiegészítések a magyar magánhangzók eloszlásának kérdéséhez: Magyar Nyelv 39. 369 kk. (1943). Vértes Edit, Adalékok a magyar nyelv hangtani szerkezetéhez: Nyelvtudományi Közlemények 53. 107—52 (1952), 54. 96—140 (1953), 55. 138—80 (1954).

3.2. De lépjünk egy szinttel följebb.

3.2.1. A morféma-rendszer strukturális modelljének magva a morféma-osztály-elmélet. Ezen a szinten egészen más kép fogad bennünket, mint a fonémarendszer struktúrája esetén. Míg a fonémáknál egy-egy pont egy-egy fonémát képviselt, addig itt egy-egy pont egy-egy egész morféma-osztályt képvisel: a legritkább eset, hogy ez az osztály egyetlen tagból álljon. Hogy hány tag van egy-egy osztályban, az igen lényeges és jellemző arra az osztályra. A morfémaosztályok két nagy csoportra oszthatók abból a szempontból, hogy elemeik száma véges-e (ez legyen, mondjuk, +1), avagy végtelen (—1). Az előbbiek között foglalnak helyet az ún. kötött morfémák osztályai (ragok, jelek, képzők), az utóbbiak között a szabad morfémákéi. Ha most e részhalmazokat tovább bontjuk, akkor újabb szempontokat (oppozíciókat, hogy a fonológiában használt terminussal éljünk) kaphatunk: minden egyes szempont egy-egy újabb dimenzió, melyben +1, illetőleg —1 komponenssel foglalnak helyet az egyes pontok.

Egy további szembeszökő vonása ennek a rendszernek olyan, hogy azt még egyelőre nem tudjuk kellő pontossággal ábrázolni a modellen. Ez a sajátság abban rejlik, hogy a különböző taglétszámot felmutató osztályokat ábrázoló (mondjuk — különböző s ú l y ú) pontok között, mind a véges súlyú, mind a végtelen súlyú pontok körül, helyet foglalnak egyes, mindig véges súlyú pontok, melyek ugyanabban a térrészben vannak ugyan, mint az egyik vagy a másik pont, de egyetlen — ritkán egynél több — dimenziót tekintve mégsem ott vannak, ahol a ,,rendes'', a többi véges és végtelen súlyú pontok. Az is kiderül, hogy ebben a rendszerben szükség van — éppen emiatt — egyes kitüntetett dimenziók felvételére, melyen más egységekkel mérünk, mint a többi dimenzióban: ezek a ,,kísérő'' pontok mindig éppen egy-egy kitüntetett dimenzióban térnek el a ,,rendes'' pontoktól, és ezzel messze elkerülnek a súlyos pontoktól. Ezek a ,,kísérő'' pontok nem mások, mint az úgynevezett kivételek; — az eddig ,,kivételnek'' ismert jelenségek nagy része és néhány olyan jelenség vagy anyagdarab, amelyeket eddig még nem tekintettek kivételnek. Példaképpen felhozhatom az irodalmunkban nemrég felvetődött *-bá* elemet (az *olybá*-ból): ez az elem néhány társával együtt egyetlen, de kiemelt dimenzió tekintetében másutt van, mint az összes többi morféma: csak egy-két másik morfémával léphet kapcsolatra, míg az összes többi morfémára a szabad kapcsolódási lehetőség a jellemző. Nemrég kaptuk meg a magyar főnév esetrendszerének következetesen strukturális leírását[4]: ebből a leírásból kiderül, hogy vannak ,,tökéletlen ragok'' és ,,rakoncátlankodó képzők'': így az *-ul* morfémát nem lehetett ragnak tekinteni, mert jel után alig-alig állhat; hasonló a helyzet a *-kor* elemmel; az *-n* ugyanakkor olyan képzőnek bizonyul, mely egy jel után is megjelenhet (például: *erősebben*), ez a helyzet meg a képzőkre nem jellemző.

A ,,súlyos'' pontok és a ,,kísérő'' pontok tehát közel vannak egymáshoz, de nem metrikus értelemben, hanem egy más természetű környezet-fogalom alapulvétele mellett, éspedig a megegyező koordináták nagy száma követ-

[4] Antal László: A magyar esetrendszer (Bp., 1961).

keztében. Vagy más szóval: a kísérő-pontok sajátossága az, hogy valamely igen nagy dimenziójú koordináta-altérre alkotott vetületük megegyezik valamely nagy súlyú pontnak erre az altérre vonatkozó vetületével.

A nyelvtani „kivételek" tehát, mint „kísérő pontok", nagyon szépen ábrázolhatók a modellen. De még ennél többet is mondhatunk. Éppen abból a körülményből kiindulva, hogy erre a szintre jellemzők a kivételek, hogy egyes, „kivételnek" minősített jelenségek a leggyakoribbak közé tartoznak, mások meg a ritkák közé, rámutattunk egy más helyen arra, hogy egyes jelenségcsoportokat vizsgálva tisztán szinkrón anyagból kiindulva meg tudjuk mondani, mi a produktív és mi az improduktív, sőt — ami az eddigihez képest talán előrelépést jelent — ki tudjuk mutatni a „produktivitás" — „improduktivitás" mennyiségileg mérhető fokát is.[5]

3.2.2. A morféma-rendszer statisztikai modellje értékes kiegészítést ad a strukturális modellhez: mint említettük, merev rendszerünkben a mozgást igen sokszor a statisztikai modell képviseli. Így például az orosz igerendszert vizsgálva azt találtuk, hogy e rendszer strukturális modelljében fellelt egyes tagjai kevés különböző tővel, de sok igekötős összetétellel nagy számban szerepelnek a csupán a leggyakoribb szavakat felölelő gyakorisági szótárban, ugyanakkor kis számmal a nyelv egész szókincsében (melyet jól-rosszul például egy közönséges, nem gyakorisági szótár reprezentálhat): ezek egy öregedő osztály tagjai (ilyenek a писать, сказать-féle igék). Másrészt van olyan ige-osztály, melyet rengeteg egyed képvisel a közönséges szótárban, s ezen egyedek legtöbbje külön-külön tőből képzett, de melyből igen kevés lelhető fel a leggyakoribb szavak között (ilyenek az -овать képzővel ellátott igék): ez egy új, fiatal osztály, mely erőteljesen — nagy tömegével — ostromolja a gyakoriság erődjét, egyelőre viszonylag (tudniillik kinti számához képest) nem sok sikerrel.[5]

3.2.3. A morféma-szintű közlési folyamat strukturális modellje nem más, mint a morfológia egzakt rendszere: itt kapjuk ugyanis azt, hogy a közlés folyamán milyen morfémák követik egymást, milyen az az egydimenziós mozaik, amelyet a morfémarendszer tagjaiból szabályosan ki lehet rakni.

3.2.4. Az itt készített statisztikai modell ugyanakkor a morfématikus „type-token" kutatások[6] alapján állítható fel; ilyen kutatás még nem sok van. Statisztikai vizsgálatoknak tudniillik csak akkor lehet nekilátni, ha a strukturális kép világos, ha az alapegység — amit „számlálunk" — vitathatatlanul és egységesen meg van határozva. A fonémával kapcsolatosan ez a feltétel megvan; a morfémáról azonban — a nyelvben elfoglalt centrális jelentősége ellenére, vagy éppen ezért — még nincs olyan biztos, egyértelmű, könnyen formalizálható definíciónk, amelyet a statisztikai számítások alapjául vehetnénk.

[5] Vö. Papp Ferenc, Морфологическая система глагольных основ в современном русском литературном языке: Slavica 2. 127—50 (Debrecen, 1962).

[6] A „type-token" (jeltípus—jelpéldány) kérdésére vonatkozó kutatások klasszikus összefoglaló munkája: G. Herdan: Type-token mathematics (Hága, 1960).

3.3. Épp ellenkező a helyzet a szavak szintjén. A szó korábbi centrális helyét átengedte a morfémának a nyelvelméletben. Ám ha a kutatás írott szövegekből indul ki, és nagyon sokszor ez a helyzet, akkor oly szépen, „maguktól" kiválnak a szavak (mint két spácium közé elhelyezett jelelemsorok), hogy szinte kínálják magukat mind a strukturális, mind — különösen — a statisztikai feldolgozásra.

3.3.1. A szókincs strukturális modelljének megépítésére számos kísérlet történt, eddig kevés eredménnyel. Kalmár László egy előadásában[7] felvetette az itt alkalmazható gráfelméleti modell kérdését; minél többet gondolkozott azóta ezen az ember, annál vonzóbbnak tűnik ez az ötlet, méghozzá nemcsak egy nyelv, hanem több nyelv viszonylatában is.

3.3.2. A szókincs statisztikai modelljét a gyakorisági szótárakra lehet építeni. Az úgynevezett Zipf-törvények[8] e modell tartó pillérei.

3.3.3.—3.4. A közlési folyamat szókincs-szintű statisztikai (és részben strukturális) modelljét a szűkebb értelemben vett „type-token" (jeltípus-jelpéldány) viszony vizsgálatok nyújtják. Kedves és lelkes diák segítőtársaimmal együtt évek óta folytatunk ilyen vizsgálatokat Debrecenben. Éppen ezért talán itt tudnám legjobban megmutatni, mit értünk a statisztikai modellen. A „típus-példány" viszony különféle aspektusai szerint felállított görbék elénk állítják például, egyszerűsített, modell-formában, azt a folyamatot, amelyen egy idegen ajkú olvasó végigmegy, amikor egy művet olvas az általa tanulmányozott nyelven. Nyelvtani nehézségektől, betűolvasási nehézségektől, kiejtési problémáktól stb. eltekint ez a modell, és csak azt mutatja, hogy milyen feladat előtt állna az olvasó, ha még egy szót sem tudna az adott nyelvből: hány különböző szóval találkoznék az első száz szó között; hány új, az első százasban még elő nem fordult, különböző szóval a második száz között; hány új, különböző szóval a harmadik száz között; stb. Hogy ezt a modellt egy vonatkozásban közelebb vigyük a valósághoz, az új típusok puszta számán kívül azt is megállapítjuk, hogy milyenek ezek a típusok (különböző szavak) gyakoriság szempontjából: gyakoriak-e, vagy ritkák. A nem anyanyelvű olvasónak ugyanis nem annyi szót kell megnéznie a szótárban, ahány különböző típussal találkozik, hanem annyit, ahány különböző új r i t k a típussal találkozik, hiszen a gyakoriakat már ismeri. Az, hogy milyen ritkaságú szavakat ismer általában, tehát melyik az a ritkasági szint, amelyen felül el kell kezdenie szótározni, szintén megállapítható, illetőleg a tanuló által ismert szómennyiség és az egyik Zipf-törvény alapján kiszámítható. Más oldalról a „típus-példány"-viszonyt ábrázoló görbéket úgy is fel lehet fogni, mint a művészi alkotás — vagy a tudományos szöveg-írás, a száraz dokumentum-összeállítás, a spontán beszélgetés stb. — „szókeresési" folyamatának

[7] Kalmár László: Lehet-e matematikai módszereket alkalmazni a nyelvtudományban? Előadás a Magyar Nyelvtudományi Társaság általános nyelvészeti és fonetikai szakosztályában 1961. november 24-én. Vö. Magyar Nyelv 58. 528 (1962).

[8] Lásd: G. K. Zipf: Human behavior and the principle of least effort (Cambridge, Mass. 1949).

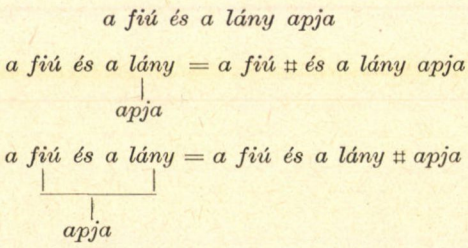
modelljét, annak a folyamatnak a modelljét, amelynek során az író-beszélő közlésében előrehaladva egyre újabb és újabb szavakhoz kénytelen folyamodni (illetőleg tudatosan egyre újabb és újabb szavakhoz fordul, hogy bizonyos stiláris hatást érjen el.) Mennyi itt a megkötöttség (a *chance*) és mennyi az egyéntől függő, szabad választás (a *choice*) eredménye?[9] Nos, ilyen s ehhez hasonló kérdésekkel foglalkozik a szó-szintű statisztikai közlés-folyamat vizsgálat.

3.4. Néhány szót végül a szintagmákról.

3.4.1. A szintagmák rendszerének vagy szintagma-kincsnek a strukturális modellje megszerkeszthető a szintagmatan alapján. Ide vonhatók különösen azok a kutatások, amelyek egyes szintagmatípusok transzformációs lehetőségeit vizsgálják, de sok hasznos meríthető e modell felállítására a „hagyományos" leíró szószerkezet-tanokból is.

3.4.2. Az e szinten megszerkeszthető statisztikai modell szükségképpen az egyes szintagma-típusok megterhelését ábrázolja.

3.4.3. Különösen érdekes a szintagma-szinten a közlési folyamat strukturális modellje. Közismert, hogy amikor — egészen egyszerű példát választva — *a fiú és a lány apja*-féle morfémasorral találkozunk, ez kétféleképpen szegmentálható: *a fiú és a lány # apja* (a fiú és a lány közös apjáról van szó), vagy *a fiú # és a lány apja* (egy fiúról és a lány apjáról van szó).



1. ábra

Természetesen nemcsak birtokviszony esetén, hanem számos más esetben is adódnak hasonló kétértelműségek; különösen fontos ezek között a ragos főnév kétértelmű (igéhez vagy főnévhez fűződő) kapcsolata; ezzel a kérdéssel bővebben Sipőczy Győző foglalkozik.[10] A problémát ezekben az esetekben az okozza, hogy míg a közlés mindig lineáris, tehát egy dimen-

[9] A *choice* és a *chance* szavakat a következő ismert munkára célozva használom: G. Herdan: Language as choice and chance (Groningen, 1956). — Vö. még Papp Ferenc, A stíluselemzés egy mennyiségi mutatókra épített módszere: Filológiai Közlöny 69—85 (1961); Количественный анализ словарной структуры некоторых русских текстов: Вопросы Языкознания 10/6. 93—100 (1961).

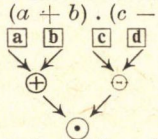
[10] Sipőczy Győző, Az elöljárós kifejezések programozási kérdései: Kötetünkben 231—242.

ziós, addig a közlés szintagma-struktúrája két dimenziós, amennyiben nemcsak egymásra következést, hanem különféle függőségeket is tartalmaz: az *apja* morfémasor egyrészt a *fiú* és a *lány* elemek után áll, másrészt függhet vagy csak a *lány*-tól, vagy mindkettőtől. Más szavakkal azt is lehet mondani, hogy az egymást követő formákon a ragok csak bizonyos műveletek elvégzésére utasítanak, de nem kapunk belőlük tájékoztatást a műveletek elvégzésének sorrendjére, „nincsenek kitéve a zárójelek" (matematikai értelemben — Lees[11] hasonlata).[12] A „zárójelek" hiánya sok esetben nem hat zavarólag: ha egyetlen ragos főnév áll például egy ige mögött, akkor az gyakorlatilag mindig tőle is függ (legalábbis a magyarban); más helyzetekben, így az idézettben, a „zárójelek" hiánya zavart okoz. A beszélt nyelv, igen természetesen, a maga egy dimenziós eszközeivel oldja meg a két dimenziós rendszer rávetítésének problémáját: kisebb-nagyobb szünetekkel utal a második dimenzió viszonyaira, vagyis különböző távolságokkal érzékelteti a „mélységet" (egy kissé hasonló módon érzékelteti például a festő két dimenzióban a három dimenziót).[13] A közlési folyamat szintagma-struktúrájának megoldására korábban a közvetlen összetevők módszere alakult ki; néhány év óta a transzformációs analízis kecsegtet a végleges megoldás reményével.

Fokozódnak a nehézségek, ha két — vagy több — nyelv viszonylatában vizsgáljuk a kérdést: kiderül ugyanis, hogy az egyik nyelvnek e tekintetben is más lehet a „látásmódja", mint a másiknak. Míg például a magyarban a ragos főnév, ha csak van igei természetű szó a közelben, attól függ, addig az oroszban a ragos főnév (viszonyszóval vagy anélkül) igen gyakran függhet valamely főnévtől is. Így aztán a magyar olvasó, amikor olyasféle orosz mondatba ütközik, mint он читал роман в стихах 'verses regényt olvasott', olyan értelmet vél felfedezni ebben az orosz szerkezetben, mint 'versekben (skandálva?) olvasott egy regényt (olvasta a regényt)'. Más esetben egyenesen értelmetlenné válik előtte a mondat (s ez még a jobbik eset, mert ez az értelmetlenség felhívhatja az olvasó figyelmét a bajra, míg az előbbi esetben belenyugodhat a látszat-értelmezésbe s annak korrigálására csak később, a szélesebb kontextus ismeretében, e mondatra újra visszatérve kerülhet sor). Így például az egyik közismert szovjet regényben ezt a mondatot olvashatjuk (a lezuhant pilótáról, miután magához tért): он . . . ощутил . . . острую, жгучую боль в ступнях . . ., majd néhány mondattal később: Боль в ступнях прожгла все его тело.:

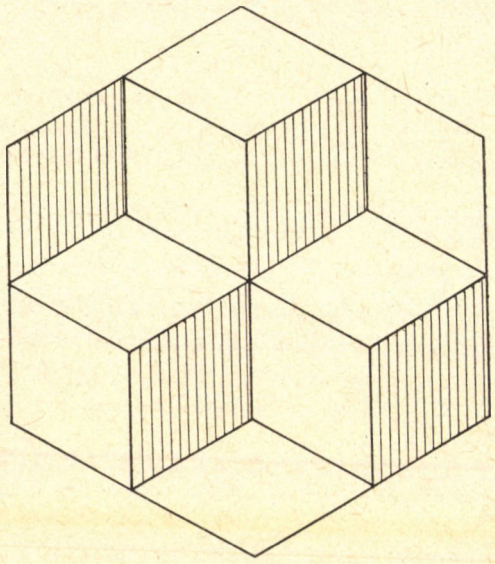
[11] R. B. Lees, О переформулировании трансформационных грамматик: Вопросы Языкознания 10/6. 42—3 (1961).

[12] Kalmár László volt szíves felhívni a figyelmemet arra, hogy amit a matematikus így ír zárójelekkel: $(a + b) . (c - d)$, azt a programozó így látja a Kontorovics-féle gráf alakjában:



[13] Így jár el a matematikus is, amikor olvassa a képleteket és kifejezéseket: *„ápluszbészer — —céminuszdé"*, vagyis szünettel a két tényező között. (Kalmár László szíves közlése.)

'. . . fájdalmat érzett a lábafejében . . . . . A lábafejében érzett fájdalom átjárta egész testét.' Míg az első mondat a magyar „látásmódnak" megfelelő, a в ступнях 'lábafejében' ragos alak az igétől függ, addig a második mondat hasonló alakja már az őt közvetlenül megelőző főnév bővítménye, bár formailag függhetne a mellette álló igétől is. Ám ha az igéhez kapcsoljuk — s erre hajlamos a magyar olvasó —, annak két össze nem egyeztethető bővítménye lesz: a pilóta éles fájdalmat érzett a lábafejében és az egész testében. A figyelmes és az orosz nyelvtan e rejtelmében némileg



2. ábra

járatos olvasó ekkor azonnal korrigálni tudja „optikai csalódását". (Ez az „optikai csalódás" ahhoz a trükkhöz hasonlítható, ami néhány kocka egymásra rajzolásában áll — vagyis egy háromdimenziós tárgy kétdimenziós leképezésében —: egyes nézők egymáson álló kockákat látnak ott, ahol mások egy befelé domborodó szobasarok-kiképzést: vö. a 2. ábrával. Csakhogy az egyes nyelvekben hozzászoktunk az ilyen dolgok bizonyos látásmódjához, és ott nem mindegy, hogy mit „veszünk ki" a leképezésből.)

3.4.4. A szintagma-szintű közlés-folyamati statisztikai modell szükségszerű kiegészítője a szavak vagy a morfémák szintjén végzett „type-token" kutatásnak és az ennek alapján szerkesztett modellnek. Világos ugyanis, hogy az előbb tárgyalt modellek, akár az olvasó-hallgató, akár az író-beszélő szemszögéből nézzük is a dolgot, egy lényeges jelenség-komplexumot szükségszerűen kihagytak: azt ugyanis, hogy a közlés folyamán nem egyszerűen szavak — legyenek azok ritkább vagy gyakrabban használatos szavak — lépnek föl, hanem többé vagy kevésbé bonyolult, többé vagy

kevésbé gyakori szószerkezet-típusok. Mármost mind a megértés, mind az alkotás szempontjából igen lényeges, hogy ezt a körülményt is figyelembe vegyük. Tartalmazhat egy szöveg igen nagy számú, ritka különböző szót, melyek ugyanakkor igen egyszerű és kevés számú típusba foglalható szószerkezetbe állnak össze — ilyenek lehetnek bizonyos szakszövegek (például bizonyos kémiai, matematikai tankönyvek). És megfordítva: állhat valamely szöveg igen gyakori és kevés számú különböző szóból, melyek azonban igen változatos szerkezet-típusokat alkotnak — ilyen felépítésűek a naturális párbeszédek a maguk „hiányos" grammatikai szerkezeteikkel; ez jellemzi például Tolsztoj „Háború és békéjé"-nek nyelvezetét, melyben bámulatra méltóan kevés különböző szót találunk, de annál bonyolultabb nyelvtani formákat, legalábbis a szerzői szövegekben. Ha megint az idegen nyelvet tanuló szempontját vesszük figyelembe: annak az előbbi esetben sokat kell szótároznia, de viszonylag keveset gondolkoznia, az utóbbi esetben viszont a jórészt ismerős szavak olyan meg-megújuló típusú, meglepő kapcsolatokra lépnek egymással, hogy igen nagy problémát jelent a szöveg megértése. Jelenleg többek között éppen azon dolgozunk, hogy az eddig szokásos szó-szintű „típus-példány"-viszonyok mellett ezt a szintagma-szintű statisztikus folyamatot is tükröztessük. Itt kedves munkatársaimmal a kutatás legelején állunk, ezért eredményekről még nem tudok beszámolni.

4. Modellek egymáshoz való viszonya

Egyes modellekkel részletesebben, másokkal éppen csak érintőlegesen foglalkoztunk. Ezt nemcsak technikai okokból tettük: előfordulhat, hogy az általunk felállított osztályozásban egyes osztályok gazdagabbak, érdekesebbek, mások esetleg teljesen üresek; — ez utóbbiakkal nem érdemes foglalkozni csak a rend, a táblázat teljes betöltése érdekében.

Az egyes szinteken felállított modellek kapcsolatban vannak egymással. Közismert például, hogy a fonéma-folyamat statisztikai modellje nagyrészt önmagában is képes „kiugratni" a morfémákat vagy a szavakat.[14] Ha ugyanis azt vizsgáljuk, hogy egy fonéma realizálása után mi a valószínűsége, hogy egy másik fonéma következzék, majd e kettő után egy harmadik és így tovább, tehát a feltételes valószínűségek egy láncolatát tekintjük, akkor kiderül, hogy egyes pontokon egy-egy fonéma valószínűsége — ha az előbb elhangzott fonémák valószínűségét is tekintjük — megközelíti az egyet. Ilyenkor egy morféma végén, illetőleg, ha ilyen magas valószínűségű a következő fonéma (következő fonémák) is — a vége f e l é járunk. Majd ha ez a valószínűség hirtelen lecsökken és minden fonémának a maga rendes, „fajlagos" valószínűsége ugrik elő, akkor morféma-határnál vagyunk s a „rendes" valószínűségével fellépő fonéma az új morféma kezdete. Nyelvészetileg szólva: mielőtt nem ejtettem ki egyetlen szót sem, bármely fonémával kezdhetem a közlést. Mihelyt kiejtettem az első fonémát, mondjuk az *e*-t, már lényegesen szűkítettem a következő valószínűségét. Ha csak nem egyetlen fonémából álló szót mondtam, akkor az első fonéma

[14] Vö.: Z. S. Harris, From phoneme to morpheme: Language 31. 190—222 (1955).

után már csak az *e* fonémával kezdődő szavak második fonémái maradtak, mint lehetséges fonémák. Az *e* után a magyarban aligha következhet magánhangzó (bár van erre is példa: *Európa, EAK* stb.), inkább csak egyes mássalhangzók. Ez utóbbiak tehát „rendes" valószínűségüknél nagyobbal szerepelnek e helyen. Lehet ez a mássalhangzó-fonéma *m* (ha például az *em-ber, em-lék* stb. szavak valamelyike kerül majd kimondásra), a *b* (ha a szó *eb, Ebes* stb. lesz) és így tovább. Ha már kimondtam annyit, hogy *em* vagy *eb*, még inkább nő ezt követően egyes fonémák valószínűsége és csökken másoké — egészen a szó, illetőleg a morféma végéig, ahol aztán megint „mindegy", hogy milyen hang következik, hiszen ott egy új, önálló elem kezdődik.

5. A matematikai nyelvészet viszonya a nyelvtudomány egyéb ágaihoz

A matematikai nyelvészet felvázolt rendszerę szoros kapcsolatban áll a strukturális nyelvészettel. A strukturális nyelvészet a mi kísérletes tudományunk, mely bonyolult módszerekkel, néha gépek igénybevételével is, felfedi a nyelv egyes alapegységeit, „atomjait", bizonyos mérési eredményeket ad nekünk. Ezeket a mérési eredményeket mi, mint a strukturális nyelvészet elméleti tudománya, a vázolt módon rendezzük. (Vö. a kísérleti fizika és az elméleti fizika viszonyával.) — Hasonlóképpen támaszkodik a matematikai nyelvészet a nyelvstatisztikára is. — A matematikai nyelvészetnek a történeti-összehasonlító nyelvészethez való viszonyára itt nem kívánok kitérni, egyébként is csak arra lennék illetékes, hogy elmondjam: mi a mi kutatásaink során mennyi hasznát vesszük nyelvtörténeti ismereteinknek; a kérdés fordítottjára nyilván a történeti-összehasonlító nyelvésznek kell válaszolnia. Nem tartanám azonban különösebb tragédiának, ha az derülne ki, hogy kölcsönösen keveset adunk és adhatunk egymásnak: mind a történeti-összehasonlító, mind a strukturális-matematikai nyelvészet már eddig is, külön-külön, igen sok értékeset nyújtott az emberiségnek, és nyilván fog is még adni; amíg pedig a társadalommal fennálló egyenleg ilyen aktív, nem lehet baj tudományunkkal. A strukturális-matematikai nyelvészet megjelenése, függetlenül a történeti-összehasonlító nyelvészethez fűződő kapcsolatától, ezek szerint annyit jelent, hogy a nyelvészet egészének jelentősége megnőtt a társadalom számára, a nyelvészet egésze váratlan, új területeken és közvetlenül is hasznot tud hajtani a társadalom számára.

6. Matematikai nyelvészet és a társadalom igénye

Mert ha az elmondottak után felmerülne a nyelvész-hallgatóban a kérdés, mire jó mindez, minek kell közismert dolgokat ilyen „nyakatekert" formába önteni, akkor erre az eddigi tapasztalatok alapján is nyugodtan azt válaszolhatjuk, hogy bizonyos közismert, vagy annak tűnő dolgok egzakt kezelése tudományunk ezen ágának létszükséglete. Gépi fordítóink például nem azt kérdezik tőlünk, hogy mire jó mindez, hanem türelmetlenked-

nek: miért nem oldottuk még meg ezeket a problémákat (így például a két dimenziós struktúra egy dimenziós leképezésének problémáját a szintagmaszinten); őket is ezek a kérdések izgatják és egyelőre meg kell elégedniük bizonyos empirikus rész-megoldásokkal.

A nyelv hosszú-hosszú ideig csak hangtestében élt, és ezzel a formával is meg tudta valósítani mind az emberek közötti közvetlen érintkezést, mind — ezen keresztül — a nemzedékek közötti érintkezést, az ember termelési és társadalmi gyakorlatának átadását és megőrzését. De eljött az idő, amikor a nyelvnek ez a formája egyedül már nem volt kielégítő: és ekkor a nyelv ősi formája mellett tartósítható, majd sokszorosítható és tartósítható fényrezgés-formát is öltött, az írás formáját. És napjainkban érjük meg azt, amikor már ez a forma sem elegendő. Ismeretes J. D. Bernal professzor nyilatkozata, aki azt mondotta: a szakpublikációk száma oly mérhetetlenül megnövekedett, hogy ha egy kutatás nem rendkívül költséges, akkor feltehetően érdemesebb újból elvégezni, mint megkeresni, hol van e kutatás eredménye publikálva, s mi az eredmény.[15] Az írott nyelv tehát a maga ellentétébe csapott át: annyit írunk, hogy azt nem lehet mind elolvasni.

E bajon a hatalmas léptekkel fejlődő technika s a mindig fiatal matematika mellett az emberhez oly közel álló humán tudomány, a nyelvészet tud ma már segíteni. Ez a tudomány is, mint a többiek, mindig az emberrel együtt nőtt és változott: a múltba nézett, amikor teremtőjének és szolgájának szeme a múltat kutatta; ma meg körültekint s a célszerű jövőbe néz, eddigi munkáját sem hagyva abba. Ahogy parttalan az idő a maga múlásában, ugyanúgy végtelen az idő minden egyes pillanatában a csodák szinkrón rendszere, melyet a matematikai nyelvészet kutat.

[15] „Számos területen lényegében elértük azt az állapotot, hogy könnyebb egy új tényt kideríteni, vagy egy új elméletet felépíteni, mint megállapítani, hogy nem fedezték-e fel, vagy nem mutatták-e ki ugyanezt már korábban is." J. D. Bernal: Science in history (London, 1954). — Magyarul idézve: Tájékoztató a tudományos kutatás tervezésének, igazgatásának és szervezésének nemzetközi irodalmából, 1961. évf. 3. szám 54.

Kiefer Ferenc

Halmazelméleti és matematikai-logikai modellek a nyelvben[1]

A matematikai nyelvészet nem pusztán, és elsősorban nem azt jelenti, hogy nyelvészeti problémák megoldásához egy-egy matematikai diszciplína módszereit alkalmazzuk. Amikor például valamely nagyobb nyelvcsaládon belüli rokonsági-fok megállapításánál statisztikai módszereket alkalmazunk (glottokronológia), vagy a beszédfolyamatot információelméleti fogalmakkal írjuk le, még nem változott meg a nyelvről mint egészről mint rendszerről alkotott véleményünk, elképzelésünk. A matematikai nyelvészet lényege a nyelv egy új koncepciójában áll, amely a matematikai és nyelvi struktúrák közti hasonlóság felismerésén alapul. A matematikai nyelvészet tehát döntő változást hozott a nyelvről alkotott elképzelésekbe, amikor a különböző matematikai diszciplínák segítségével megalkotta a nyelv különböző matematikai modelljeit. Aszerint, hogy milyen matematikai diszciplínát veszünk alapul modellünk megalkotásánál, beszélhetünk a nyelv halmazelméleti, matematikai-logikai, statisztikai, információelméleti és algebrai modelljéről.

1. A Kulagina-féle halmazelméleti modell

1.1. A modell körvonalazása

A nyelv egy halmazelméleti modellje O. Sz. Kulagina,[2] szovjet matematikus-nyelvésztől származik. A Kulagina-féle modell tulajdonképpen csak a

[1] Előadásom írásbeli változatának címét és szövegét Kalmár László előadásának írásbeli változatához igazítottam, elsősorban az átfedések elkerülése végett.

[2] Vö. O. Sz. Kulagina, Об одном способе определения грамматических понятий на базе теории множеств: Проблемы кибернетики 1. 203—214 (1958). — Lásd bővebben: „A halmazelmélet egy nyelvészeti alkalmazásáról" című tanulmányomat (Általános nyelvészeti tanulmányok. Szerk.: Telegdi Zsigmond. Bp. 1963. 187—200. [A továbbiakban ennek a dolgozatomnak a lapszámára így hivatkozom: Kiefer 187—200]). Ebben a tanulmányban részletesen ismertetem azokat a halmazelméleti fogalmakat, amelyek nyelvészetileg könnyen interpretálhatóak, és amelyeket Kulagina is felhasznál elméletének felépítésében. A tanulmány második részében pedig részletesen tárgyalom a Kulagina-féle elmélet első részét.

A nyelvészeti interpretációnál természetesen mindig óvatosaknak kell lennünk. Amikor például matematikailag, tehát általánosan bebizonyított összefüggések meg-

halmazelmélet alapfogalmait használja fel, bonyolultabb halmazelméleti módszereket nem alkalmaz. Modellje mindenesetre példa a „tiszta alkalmazásra", vagyis arra, amikor egy nyelvi modell megalkotására csak egyetlen egy matematikai diszciplína módszereit alkalmazzuk. A halmazelméleti módszerek általában más (elsősorban matematikai-logikai és algebrai) módszerekkel egészülnek ki. Sokszor célszerűnek látszik komplex nyelvi modellek megalkotása, melyekhez általában több matematikai diszciplína egyidejű alkalmazása vezethet. Többen foglalkoznak a Kulagina-féle halmazelméleti modell ilyen irányú továbbfejlesztésével is.

Kulagina elmélete két nagyobb részre osztható. Az első részben eljut a típus fogalmához, mely megközelítőleg megfelel a szófaj fogalmának és az úgynevezett tipológiai kritériumokig (egyszerűség, egyneműség, szabályosság), a második részben a konfigurációk elméletét tárgyalja.

Kulagina valamely nyelv összes szóalakjainak halmazából indul ki:

$$\Xi = \{\ldots, x, \ldots\}.$$

E halmaz elemeinek bármely sorozatát (kortezs, string, lánc) m o n d a t-nak nevezi (frázis):

$$A = x_1 x_2 \ldots x_n.$$

Az A mondatban egy-egy elem többször is előfordulhat. Ez a mondatfogalom persze nem fedi a hagyományos mondatfogalmat, hiszen az A mondat lehet grammatikailag helytelen mondat is. Hogy a grammatikailag helytelen mondatokat a helyesen képzett mondatoktól megkülönböztesse, Kulagina az utóbbiakat k i t ü n t e t e t t m o n d a t n a k nevezi.[3] A kitüntetett mondatok halmazát pedig a következőképpen jelöli

$$\Theta = \{\ldots, A, \ldots\}.$$

Minden szóalaknak van környezete. Ebbe a környezetbe általában minden olyan szóalak beletartozik, amely ugyanannak a szónak alakja. Valamely x szóalak környezetét Kulagina $\Gamma(x)$-szel jelöli. Γ tehát a szóban forgó nyelvre jellemző környezetfogalomnak, vagy ha tetszik, annak a

---

felelőit keressük a nyelvben, bizonyos mértékig mindig vétünk az általánosság ellen Lehetetlen ugyanis minden esetben minden lehetőséget megvizsgálni, a matematika egzaktság pedig ezt követelné. Egy-egy tulajdonság nem-teljesülésének bebizonyításához természetesen elegendő egyetlen példa is, de egy tulajdonság teljesülésének bebizonyításához minden lehető esetet végig kellene vizsgálni, és csak akkor mondhatnánk határozott igent. Persze, bizonyos mértékig támaszkodhatunk a nyelv strukturális tulajdonságaira, amelyek sokszor csak intuitív szemléletünkben léteznek. Számolnunk kell tehát azzal a ténnyel, hogy a nyelv nem matematikailag definiált, hanem történetileg kialakult, empirikusan adott struktúra.

A Kulagina-féle elmélet előzményeiről és az elmélet néhány gondolatáról Kalmár László is ír kötetünkben (Matematikai és nyelvi struktúrák, 11—74).

[3] A grammatikalitás követelményét — amely nem azonos az „értelmesség" követelményével — lásd még „A matematikai nyelvészet kérdései" című cikkemben: Valóság 6/1. 84—90 (1963).

függvénynek a jele, amely megadja, hogyan függ egy szóalak környezete e szóalaktól.

Kulagina elméletében alapfogalomnak tekinti a szóalak, a kitüntetett mondat és a környezet fogalmát, ezeket tehát nem definiálja. Ebben ismét a matematikai elméletek felépítését követi.[4] Ezek után jut el Kulagina a nyelv következő definíciójához:

A szóalakok adott halmazát környezeteik adott rendszerével és a kitüntetett mondatok adott halmazával együtt nyelvnek nevezzük, és így jelöljük:

$$\Xi\,(\Gamma, \Theta).^{4a}$$

Ha a homonímiától eltekintünk, akkor a környezetek a nyelvet egymástól idegen részhalmazokra osztják. Ez azt jelenti, hogy bármely szóalak csak egy környezetbe tartozik vagy másképpen: nem lehet két környezetnek egyetlen közös eleme sem. Képletben

$$\Gamma_i \cap \Gamma_j = 0 \text{ bármely } i \neq j\text{-re.}$$

O itt az üres halmaz jele, $\cap$ a közös elemek halmaza, az úgynevezett metszet képzéséé.

Természetesen a homonímia problémáját is meg lehetne oldani például indexezéssel, úgy hogy például az $ár_1$, $ár_2$, $ár_3$ szóalakokat különbözőknek tekintenénk.[5]

A nyelvet azonban nemcsak a környezetek alapján lehet egymástól páronként idegen részhalmazokra felbontani. Definiálhatjuk például a szóalakok ekvivalenciáját, és az ekvivalens szóalakokat egy osztályba sorolhatjuk, a nem-ekvivalenseket különbözőkbe. Az így kapott osztályokat Kulagina családoknak nevezi. Nyilvánvaló, hogy a családok is felosztását adják a Ξ halmaznak. De felosztása a Ξ halmaznak az úgynevezett egységfelosztás is, vagyis az a felosztás, amelyben minden osztálynak egyetlen eleme egy szóalak. Ezt az egység-felosztást jelöljük E-vel.

A különböző felosztásoknak megfelelően beszélhetünk az A mondat különböző struktúráiról. Az A mondat Γ-struktúrája például a következő

$$\Gamma(A) = \Gamma(x_1)\,\Gamma(x_2) \ldots \Gamma(x_n).$$

Vagy vegyünk egy konkrét példát: legyen

[4] Lásd részletesebben: Kiefer 187—200.

[4a] A matematikában használatosabb jelölésmód mintájára inkább így lehetne jelölni:
(Ξ, Γ, Θ) vagy $< \Xi, \Gamma, \Theta >$.
A Ξ „kiemelése" a zárójelből a Ξ szóalak-készlet alapvető szerepét hangsúlyozza.

[5] Bar-Hillel például a *poor* típusának megállapításánál a $poor_1$, $poor_2$ és $poor_3$ típusokat is figyelembe veszi. [A quasi-arithmetical notation for syntactic description: Language 29. 50 (1953).]

A = A szőke fiú elmerülten játszik a kertben. Akkor az A mondat Γ-struktúrája a következő

$$\Gamma(A) = \begin{Bmatrix} a \\ az \\ \vdots \end{Bmatrix} \begin{Bmatrix} szőke \\ szőkébb \\ legszőkébb \\ \vdots \end{Bmatrix} \begin{Bmatrix} fiú \\ fiút \\ fiúk \\ \vdots \end{Bmatrix} \begin{Bmatrix} elmerülten \\ elmerültebben \\ \vdots \end{Bmatrix} \begin{Bmatrix} játszik \\ játszunk \\ játszott \\ \vdots \end{Bmatrix}$$

De ugyanannak az A mondatnak vehetjük az E-struktúráját is:

$$\mathrm{E(A)} = \{A\}\ \{szőke\}\ \{fiú\}\ \{elmerülten\}\ \{játszik\}\ \{a\}\ \{kertben\},$$

ahol tehát minden halmaznak egyetlen eleme van.

Amikor tehát valamely mondat struktúráját akarjuk megállapítani, mégpedig valamely felosztás szerinti struktúráját, akkor csak azt kell megadnunk, hogy az adott mondat elemei mely osztályba tartoznak. Az osztályokat egyetlen reprezentáns elem egyértelműen meghatározza. Így az A mondat Γ-struktúrájának megállapításánál elég a következőt írnunk

$$\Gamma(\mathrm{A}) = \Gamma(a)\ \Gamma(szőke)\ \Gamma(fiú)\ \Gamma(elmerülten)\ \Gamma(játszik)\ \Gamma(a)\ \Gamma(kertben).$$

A környezetek egy-egy szóalak esetén különböző nyelvekben nyilván nem lesznek ugyanazok. A magyarban a határozott névelőnek csak két alakját ismerjük. Egyszerűség kedvéért egy környezetbe soroltuk őket.[6] A németben például a határozott névelő környezete nem két elemű, hanem hat elemű halmaz. Hasonló érvényes nyilvánvalóan a jelzők (melléknevek), határozószók stb. környezetére is.

Az A mondat struktúráját Kulagina akkor nevezi kitüntetettnek, ha van olyan kitüntetett mondat, amelynek ez a struktúrája.

Ez azt jelenti, hogy például az

$$\mathrm{A} = Emberek\ élnek\ a\ magas\ hegyekbe\ is$$

E-struktúrája, vagyis

$$\mathrm{E(A)} = \mathrm{E}(emberek)\ \mathrm{E}(élnek)\ \mathrm{E}(a)\ \mathrm{E}(magas)\ \mathrm{E}(hegyekbe)\ \mathrm{E}(is)$$

nem kitüntetett, mert nincs olyan kitüntetett A mondat, amelynek ez lenne az E-struktúrája, hiszen az egyetlen mondat, amelynek ez az E-struktúrája, éppen az A mondat, amely pedig nem kitüntetett.

[6] Lásd bővebben Kiefer 187—200, főképp 192. A besorolás helyessége vitatható: *az* nem ugyanolyan értelemben alakja *a*-nak, mint például a *fiút* a *fiú*-nak.

De kitüntetett például a következő Γ-struktúra, amely ugyanannak az A mondatnak a Γ-struktúrája, vagyis a

$$\Gamma(\mathrm{A}) = \Gamma(emberek)\ \Gamma(élnek)\ \Gamma(a)\ \Gamma(magas)\ \Gamma(hegyekbe)\ \Gamma(is)$$

struktúra, annak ellenére, hogy maga az A mondat nem kitüntetett, de ha a *hegyekbe* szóalak környezetéből a *hegyekben* szóalakot vesszük, akkor az

$$\mathrm{A}' = Emberek\ élnek\ a\ magas\ hegyekben\ is$$

mondat már kitüntetett. Ennek a mondatnak ugyanaz a Γ-struktúrája, mint az előbbinek, tehát van olyan kitüntetett mondat, amelynek a fenti Γ-struktúra a Γ-struktúrája.

Tekintve, hogy itt a nyelv szóalakkészletét jelentő Ξ halmazt már többféleképpen felosztottuk páronként egymástól idegen osztályokra, a felosztás problémáját érdemes általában is megvizsgálni. Kulagina egy tetszés szerinti felosztást vesz, és ezt nevezi B felosztásnak. A B felosztás osztályait pedig $B_1, B_2, \ldots, B_n$-nel jelöli. Azt az osztályt közülük, amelybe valamely adott x szóalak beletartozik, ezenkívül így is jelöli: $B(x)$. Itt is, mint fentebb, azt, hogy valamely x szóalak egy tetszés szerinti B felosztásnál mely osztályba tartozik, egyszerűen magával a kérdéses $B(x)$ osztállyal adhatjuk meg. Így az

$$\mathrm{A} = x_1 x_2 \ldots x_n$$

mondat B-struktúrája a következő lesz:

$$B(A) = B(x_1)\ B(x_2) \ldots B(x_n).$$

A B felosztás osztályai között Kulagina a következő ekvivalenciarelációt definiálja: A B_i osztály B-ekvivalens a B_j osztállyal, ha (1) abból, hogy a $B(A_1)\ B_i\ B\ (A_2)$ kitüntetett B-struktúra, következik, hogy $B(A_1)\ B_j\ B(A_2)$ is az; és viszont, (2) abból, hogy $B(D_1)\ B_j\ B(D_2)$ kitüntetett B-struktúra, következik, hogy $B(D_1)\ B_i\ B(D_2)$ is az. Természetesen itt A_1, A_2, D_1, D_2 tetszés szerinti mondatokat jelöl (bármelyikük lehet üres is); $B\ (A_1)$, $B\ (A_2)$, $B\ (D_1)$, $B\ (D_2)$ pedig a megfelelő B-struktúrákat.

Tekintve, hogy B tetszés szerinti felosztását jelenti a Ξ halmaznak, a B-ekvivalencia speciális eseteként beszélhetünk Γ-, illetőleg E-ekvivalenciáról.[7] Az E-ekvivalencia egyszerűen a szóalakok ekvivalenciáját jelenti: két szóalak ekvivalens, ha bármely kitüntetett mondatból ismét mindig kitüntetett mondat lesz, ha e két szóalak bármelyikét a másikkal pótoljuk. A szóalakok ekvivalenciája módot nyújt — mint láttuk — a Ξ halmaz egy újabb felosztására. Egy osztályba sorolhatók ugyanis az ekvivalens szóalakok, különbözőbe a nem ekvivalensek. Ezt a felosztást Kulagina C felosztásnak (családokra való felosztásnak) nevezi.[8] Nyilvánvaló, hogy nincs két különböző olyan család, amely egymással ekvivalens lenne, vagyis C-ekvivalencia nem létezik.[9]

[7] A Γ-ekvivalenciára példát mutatok: Kiefer 194—195.

[8] Lásd bővebben: Kiefer 192—193.

[9] A matematikus ezt inkább úgy mondja, hogy a C-ekvivalencia a triviális ekvivalencia, amelyben két család akkor és csak akkor ekvivalens, ha azonosak. Hogy ez így van, egész általánosan belátható. Nincs ugyanis olyan kitüntetett mondat,

A Γ-ekvivalencia is módot ad a Ξ halmaz újabb felosztására: egy osztályba akkor és csak akkor kerül két szóalak, ha egymással Γ-ekvivalens környezetekhez tartoznak. Tekintve, hogy ez a felosztás a Γ felosztásból ered, a Γ felosztás deriváltjának nevezzük és a következőképpen jelöljük: Γ' A Γ' felosztást Kulagina típusokra való felosztásnak nevezi és T-vel jelöli. Nyilvánvaló, hogy a típusokra való felosztással bizonyos értelemben helyettesíthető a Ξ halmaznak szófajokra való hagyományos felosztása.[10]

---

amelyben valamely elem C-osztályát egy másik tőle különböző C-osztállyal helyettesítve, a kapott C-struktúra kitüntetett lenne. Legyen például

$$C_1 = \{\ldots, x_i, \ldots\},$$
$$C_2 = \{\ldots, x_j, \ldots\}.$$

Tegyük fel, hogy abból, hogy

$$C(A_1)\, C_1\, C(A_2)$$

kitüntetett struktúra, következik, hogy

$$C(A_1)\, C_2\, C(A_2)$$

is az, és viszont.

E struktúrák kitüntetettsége azt jelenti, hogy van olyan D_1, illetőleg D_2 kitüntetett mondat, amelynek fentiek a C-struktúrái. De akkor nyilván van olyan

$$x_i \in C_1 \text{ és } x'j \in C_2,$$

amelyek ekvivalens szóalakok. Ez viszont azt jelentené, hogy x_i és $x'j$ egy családba, vagyis egy C-osztályba tartoznak. Tehát valóban $C_1 = C_2$. A C felosztás ugyanis felosztása a Ξ halmaznak, ennélfogva a különböző C-osztályok páronként egymástól idegenek.

[10] Ez nem minden nyelvre érvényes egyenlő mértékben. Magyarban nyilván minden főnév környezete Γ-ekvivalens, ezért a főnevek típusa minden főnevet tartalmaz, tehát megegyezik a főnevek halmazával.

A németben viszont például a főnevek halmaza nem egyezik meg a főnév-típussal. Vegyük például a következő mondatot:

A = *Ein braves Kind schläft im Bett*

és a következő környezeteket

$$\Gamma_1 = \left\{\begin{matrix} Kind \\ Kindes \\ Kinder \\ Kinde \\ Kindern \end{matrix}\right\}; \qquad \Gamma_2 = \{Eltern\}.$$

Γ_1 nem ekvivalens Γ_2-vel, mert ha A_1 = *ein braves* és A_2 = *schläft im Bett*, akkor bár a

$$\Gamma(A) = \Gamma(A_1)\, \Gamma_1\, \Gamma(A_2)$$

struktúra kitüntetett (hiszen az A mondat kitüntetett), de az

$$\Gamma(A_1)\, \Gamma_2\, \Gamma(A_2)$$

nem kitüntetett. Tehát a német nyelvben nem egyezik meg a szófajfogalom a típus fogalmával. Hasonló a helyzet az orosz, francia, olasz stb. nyelvben. A magyarhoz hasonló „ideális" eset az ismertebb nyelvek közül csak az angolban fordul elő.

1.2. A tipológiai kritériumok

A fentiek segítségével a nyelvek különböző formális tulajdonságai állapíthatók meg. Ezekhez a következő definíciók révén juthatunk el:

S z a b á l y o s n a k nevezünk egy nyelvet, ha abból, hogy két szóalak egy családba tartozik, mindig következik, hogy egy típusba is tartozik.

E g y n e m ű n e k nevezzük a nyelvet, ha bármely két x, y szóalakra áll a következő: abból, hogy x környezetének és y családjának van közös eleme, következik, hogy y környezetének és x családjának is van közös eleme.

E g y s z e r ű n e k nevezzük a nyelvet, ha egynemű, és ha az x szóalak környezetének és családjának egyetlen közös eleme maga az x szóalak.[11]

Általában minden ismert természetes nyelv szabályos, de könnyű konstruálni olyan mesterséges nyelvet, amely nem szabályos.[12]

Egynemű például a francia és a magyar nyelv, nem egynemű az orosz és a cseh.

[11] E definíciókhoz lásd bővebben: Kiefer 196—200.

[12] Induljunk ki például a következő nyelvből

$$\Xi = \{a_1, a_2, b_1, c_1, c_2, c_3, d_1, d_2\}.$$

Legyen ebben a kitüntetett mondatok halmaza a következő:

$$\Theta = \{a_1b_1, a_1c_2, a_2c_3, a_2d_1, a_1\,b_1d_2, a_1c_2d_2\}\,.$$

A környezetek legyenek per definitionem

$$\Gamma_1 = \{a_1, a_2\}\,, \qquad \Gamma_3 = \{c_1, c_2, c_3\},$$
$$\Gamma_2 = \{b_1\}\,, \qquad \Gamma_4 = \{d_1, d_2\}\,.$$

Ebben a nyelvben csak két többelemű családot találunk

$$C_1 = \{c_3, d_1\}\,, \qquad C_2 = \{b_1, c_2\}\,;$$

a többi család mind egy-egy szóalakból áll.

Nézzük meg most a Γ-ekvivalenciát. Mivel négy környezetünk van, összesen hat lehetőséget kell megvizsgálnunk. Könnyen látható, hogy a hat lehetséges eset közül csak a Γ_2 és Γ_3 környezetek Γ-ekvivalensek, Γ_1 és Γ_4 csak önmagával Γ-ekvivalens.

Ha a Γ-ekvivalens osztályokat egyesítjük, kapjuk a megfelelő típusokat

$$T_1 = \Gamma_1 = \{a_1, a_2\}\,,$$
$$T_2 = \Gamma_2\, U\, \Gamma_3 = \{b_1, c_1, c_2, c_3\}\,,$$
$$T_3 = \Gamma_4 = \{d_1, d_2\}\,.$$

Most tehát meg kell vizsgálnunk, hogy ha két elem egy családba tartozik, akkor szükségképpen egy típusba tartozik-e? A c_3 és d_1 egy családba tartozik, de nem tartozik egy típusba. Ezzel olyan nyelvet konstruáltunk, amelyben abból, hogy két szóalak egy családba tartozik, nem következik szükségképpen, hogy azok egy típusba is tartoznak.

Egyszerű nyelv a magyar, az angol, de nem egyszerű az orosz és a német.[13]

Ezek a tipológiai kritériumok, melyek természetesen másképpen is definiálhatók, illetőleg más kritériumokkal kiegészíthetők,[14] lehetőséget nyújtanak a természetes nyelvek bizonyos osztályozására is. Azonkívül gépi fordításnál is fontos szerepet játszhat egy-egy kritérium teljesülése, illetőleg nemteljesülése.

1.3. Konfiguráció-elmélet

Tekintsük a Ξ halmaz egy tetszés szerinti B felosztását. Az olyan B-struktúrát (amelyet $\widetilde{B}^{(1)}$-gyel jelölünk), nevezzük első fokú B-konfigurációnak, amely legalább két elemet tartalmaz, és amelyhez létezik a B felosztás olyan B_{a1} eleme, hogy a $B(A_1)\ \widetilde{B}^{(1)}\ B(A_2)$ és $B(A_1)\ B_{a1}\ B(A_2)$ B-struktúrák (tetszőleges A_1 és A_2 mondatok esetén) vagy mindketten kitüntetettek, vagy egyik sem. A B_{a1} elemet, amellyel a $\widetilde{B}^{(1)}$-konfigurációt ebben az értelemben helyettesíteni lehet, rezultáns elemnek nevezzük.

A rezultáns elem nincs egyértelműen meghatározva. Nyilvánvalóan, ha B_{a1} rezultáns eleme a $\widetilde{B}^{(1)}$-konfigurációnak, akkor a B felosztás egy vele B-ekvivalens tetszés szerinti eleme szintén rezultáns eleme lesz a $B^{(1)}$-konfigurációnak.

Első fokú B-struktúrának nevezzük, és $B^{(1)}$-gyel jelöljük az olyan B-struktúrát, amely nem tartalmaz első fokú konfigurációt. Első fokú B-struktúrát a következőképpen kaphatunk: Veszünk egy tetszőleges B-struktúrát, vagyis egy tetszőleges A mondat $B(A)$ B-struktúráját; benne minden első fokú konfigurációt rezultáns elemeikkel helyettesítjük, a kapott struktúrában ismét elvégezzük ezt a helyettesítést mindaddig, amíg olyan struktúrát nem kapunk, amelyben már nincs első fokú konfiguráció. Az így kapott első fokú B-struktúrát $B^{(1)}(A)$-val jelöljük. Az első fokú B-struktúrák nyilván az összes B-struktúrák részhalmazát alkotják.

Hasonlóképpen definiálható a másodfokú konfiguráció. Másodfokú konfigurációnak nevezünk olyan B-struktúrát (és $\widetilde{B}^{(2)}$-vel jelöljük), amely legalább két elemet tartalmaz, és amelyhez található a B felosztás olyan B_{a2} eleme ($\widetilde{B}^{(2)}$ rezultáns eleme), hogy tetszés szerinti A_1, A_2 mondatok esetén az

$$B^{(1)}(A_1)\ \widetilde{B}^{(2)}\ B^{(1)}(A_2) \text{ és } B^{(1)}(A_1)\ B_{a2}\ B^{(1)}(A_2)$$

első fokú B-struktúrák vagy mind a ketten kitüntetettek, vagy egyikük sem az.

Másodfokú B-struktúrának nevezzük az olyan B-struktúrát, amely nem tartalmaz másodfokú konfigurációt. Másodfokú B-struktúrát hasonló eljárással kaphatunk, mint első fokú B-struktúrát, csak most minden másodfokú konfigurációt kell rezultáns elemeikkel helyettesíteni. Hason-

[13] Példákat lásd: Kiefer 196—200

[14] A különböző lehetőségeket S. Marcus vizsgálta meg a Tihanyi Kollokviumon (A matematika alapjai, matematikai gépek, és alkalmazásaik, Tihany, 1962. szeptember 11—15) bemutatott előadásában: Typologie des langues et modèles logiques: Acta Mathematica Academiae Scientiarum Hungaricae 14. 269—281 (1963).

lóan definiálható a harmadfokú, negyedfokú stb. konfiguráció és B-struktúra fogalma.[15]

A konfigurációk az egyes szintagma-fajtáknak felelnek meg. Bonyolultabb mondatokból az egyes szintagmák „lebontogatásával" egyszerű mondatokat kapunk. Az első fokú B-struktúra olyan egyszerű mondat struktúrája, amelyben a szintagmákat nem lehet tovább „bontogatni".

Lássunk egy példát. Tekintsük a következő mondatot

$$\mathrm{A} = \textit{Hosszú út vezetett a szép kilátóhoz}$$

Legyen most $B = C$-vel, vagyis a családokra való felosztással. A A mondat C-struktúrája tehát a következő

$\mathrm{C}(\mathrm{A}) = \mathrm{C}$ *(hosszú)* C *(út)* C *(vezetett)* C *(a)* C *(szép)* C *(kilátóhoz)*.

Azt lehetne gondolni, hogy itt

$$\mathrm{C}\ \textit{(hosszú)}\ \mathrm{C}\ \textit{(út)}$$

első fokú konfiguráció és rezultáns eleme C *(út)*. Ez azonban nem áll, hiszen ha a következő C-struktúrát vesszük

$$\mathrm{C}(\mathrm{A}') = \mathrm{C}\ \textit{(nagyon)}\ \mathrm{C}\ \textit{(hosszú)}\ \mathrm{C}\ \textit{(út)}\ \mathrm{C}\ \textit{(vezetett)}\ \mathrm{C}\ \textit{(a)}\ \mathrm{C}\ \textit{(szép)}\ \mathrm{C}\ \textit{(kilátóhoz)}$$

akkor, ha a

$$\mathrm{C}\ \textit{(hosszú)}\ \mathrm{C}\ \textit{(út)}$$

[15] Kulagina tulajdonképpen a konfigurációkat rekurzióval definiálja. Ez azt jelenti, hogy először definiálja az első fokú konfigurációt (és természetesen az első fokú B-struktúrát), majd mindjárt a k-adfokú konfigurációt definiálja (és a k-adfokú B-struktúrát), de ebben a definícióban felhasználja a $(k-1)$-edfokú konfiguráció, illetőleg $(k-1)$-edfokú B-struktúra fogalmát. A k-adfokú konfiguráció definíciója a következőképpen szól:

k-adfokú konfigurációnak olyan B-struktúrát nevezünk (és $\tilde{B}^{(k)}$-val jelöljük), amely legalább két elemet tartalmaz, és amelyhez található a B felosztás olyan Ba_k eleme, hogy tetszés szerinti A_1, A_2 mondatok esetén a

$$B^{(k-1)}(A_1)\ \tilde{B}^{(k)}\ B^{(k-1)}(A_2) \text{ és}$$
$$B^{(k-1)}(A_1)\ B_{ak}\ B^{(k-1)}(A_2)$$

$(k-1)$-edfokú B-struktúrák vagy mind a ketten kitüntetettek, vagy egyikük sem az.

k-adfokú B-struktúrának nevezünk olyan B-struktúrát, amely nem tartalmaz k-adfokú konfigurációt.

$B^k(A)$ az A mondat k-adfokú B-struktúrája, olyan B-struktúra tehát, amely nem tartalmaz k, $k-1$, $k-2$, ... fokú konfigurációt.

Ebből a definícióból visszafelé „lépegetve" kapjuk a definíciót k, $k-1$, $k-2$, ... esetére, míg utolsó lépésként a rekurzió alapjához jutunk, vagyis a $k = 1$ esetre szóló definícióhoz. Vagyis ahelyett, hogy a konfigurációkat $k = 1,2,3,\ldots$ esetre definiálnánk, a k-adfokú konfigurációt a $(k-1)$-edfokú konfiguráció segítségével definiáljuk, ezt ismét a $(k-2)$-edfokú konfiguráció segítségével és így tovább; a másodfokú konfigurációt az első fokú segítségével definiáljuk, az első fokúra pedig külön definíciót adunk.

helyett C *(út)* állna, a fenti *C*-struktúra (*A'* mondat *C*-struktúrája) nem lenne kitüntetett, hiszen a

C(A) = C *(nagyon)* C *(út)* C *(vezetett)* C *(a)* C *(szép)* C *(kilátóhoz)*

C-struktúra nem kitüntetett. Nincs ugyanis olyan kitüntetett mondat, amelynek C(A) lenne a *C*-struktúrája.

Ugyanez vonatkozik a

C *(szép)* C *(kilátóhoz)*

konfigurációra is. Viszont első fokú konfiguráció a

C *(nagyon)* C *(hosszú)*,

mert mindig helyettesíthető

C *(hosszú)*-val,

a kapott *C*-struktúra kitüntetett lesz.

Ha most C (A)-ban minden első fokú konfigurációt rezultáns elemével helyettesítünk, megkapjuk a $C^{(1)}(A)$ első fokú *C*-struktúrát.

Ha csak az első fokú *C*-struktúrákat tekintjük, akkor már

C *(hosszú)* C *(út)*

mindig helyettesíthető

C *(út)*-tal,

tehát

C *(hosszú)* C *(út)*

másodfokú konfiguráció. Ugyancsak másodfokú konfiguráció

C *(szép)* C *(kilátóhoz)*,

mert mindig helyettesíthető minden első fokú *C*-struktúrában

C *(kilátóhoz)*-zal.

Ha a fenti másodfokú konfigurációkat rezultáns elemeikkel helyettesítjük, az *A* mondat $C^{(2)}$ (A) másodfokú *C*-struktúráját kapjuk:

C *(út)* C *(vezetett)* C *(a)* C *(kilátóhoz)*;

ebben a másodfokú *C*-struktúrában a

C *(vezetett)* C *(a)* C *(kilátóhoz)*

helyettesíthető

C *(vezetett)*-tel,

mert C *(út)* egyik eleme például *Éva* is, C *(vezetett)* eleme például *énekelt* és

Éva énekelt

kitüntetett mondat. Ez tehát azt jelenti, hogy

C *(vezetett)* C *(a)* C *(kilátóhoz)*

harmadfokú konfiguráció. Ezt a harmadfokú konfigurációt is helyettesítve rezultáns elemével a következő harmadfokú C-struktúrát kapjuk

C *(Éva)* C *(énekelt)*,

ami ugyanaz, mintha azt írnánk, hogy

C *(út)* C *(vezetett)*.

A rezultáns elem nincs mindig egyértelműen meghatározva, hiszen könnyen látható, hogy például

C *(szép)* C *(kilátóhoz)*

másodfokú konfiguráció rezultáns eleme nemcsak

C *(kilátóhoz)*,

hanem

C *(oda)*

is.

A

C *(lány)* C *(énekelt)*

C-struktúra további konfigurációkat nem tartalmaz. A k-ad fokú kitüntetett B-struktúra, maximális k esetén, általában magmondat (kernel sentence) B-struktúrája.[16]

[16] Ha a

$$B(A) = B(x_1)\, B(x_2) \ldots B(x_n)$$

struktúra $j = 1,2, \ldots, k$ esetén nem tartalmaz j-ed fokú konfigurációt, akkor $B(A)$ k-ad fokú B-struktúra. Ha ez a $B(A)$ struktúra kitüntetett, akkor maximális k esetén magmondat B-struktúrája.

A konfiguráció pontos fokszámának megállapítása adott esetben nem mindig könnyű feladat, sokszor mélyreható analízist követel.[17]

[17] Az orosz nyelvre vonatkozóan Revzin mutat példát a konfigurációkra. Legyen a *B* felosztás a családokra való felosztás, vagyis a *C* felosztás és vegyük a következő *C*-struktúrát: C *(маленькая)* C *(девочка)* C *(долго)* C *(ласкала)* C *(кошку)*, A C *(маленькая)* C *(девочка)* nem első fokú konfiguráció, mert vegyük például a következő struktúrát:

C *(весьма)* C *(маленькая)* C *(девочка)* C *(стояла)*;

akkor a

C *(весьма)* C *(девочка)* C *(стояла)*

C-struktúra nem kitüntetett.

Viszont a

C *(весьма)* C *(маленькая)*

első fokú konfiguráció, mert mindig helyettesíthető

C *(маленькая)*-val (rezultáns elem).

Ha viszont csak az első fokú *C*-struktúrákat tekintjük, akkor a

C *(маленькая)* C *(девочка)*

mindenütt helyettesíthető

C *(девочка)*-val és viszont.

Így

C *(маленькая)* C *(девочка)*

másodfokú konfiguráció, melynek rezultáns eleme

C *(девочка).*

De a

C *(долго)* C *(ласкала)*

is másodfokú konfiguráció, melynek rezultáns eleme

C *(ласкала).*

A másodfokú konfigurációkat rezultáns elemeikkel helyettesítve a következő másodfokú *C*-struktúrát kapjuk

C *(девочка)* C *(ласкала)* C *(кошку).*

Ha csak a másodfokú *C*-struktúrákat tekintjük, akkor

C *(ласкала)* C *(кошку)*

harmadfokú konfiguráció, melynek rezultáns eleme például

C *(стояла).*

Így végül is a következő harmadfokú *C*-struktúrát kapjuk

C *(девочка)* C *(стояла).*

Megjegyezhetjük, hogy például itt sem egyértelmű általában a rezultáns elem, mert a

C *(весьма)* C *(маленькая)*

1.4. A grammatikai függés és függetlenség mértéke

A konfigurációk tanával függ össze az első-, másod- és harmadfajú viszony kérdése. Ehhez viszont ismét néhány precízen definiált fogalomra van szükség.

Legyen $B^{(1)}$ és $B^{(2)}$ a Ξ halmaznak két felosztása. A felső index most nem azt jelenti, mint az előbbi pontban (konfiguráció, illetőleg struktúra fokszámát), hanem csak megkülönböztető index. Képletben

$$\Xi = \bigcup_i B_i^{(1)} \quad \text{és} \quad \Xi = \bigcup_j B_j^{(2)}.$$

Akkor mondjuk, hogy a $B^{(1)}$ felosztás d u r v í t á s a a $B^{(2)}$ felosztásnak (vagy röviden: hogy $B^{(1)}$ d u r v á b b, mint $B^{(2)}$), ha

$$B^{(1)}(x) \supseteq B^{(2)}(x) \text{ minden } x \in \Xi$$

esetén, és létezik legalább egy olyan $x \in \Xi$, hogy

$$B^{(1)}(x) \supset B^{(2)}(x).^{18}$$

Könnyen belátható, hogy ez más szavakkal azt jelenti, hogy a $B^{(1)}$ felosztás bármely osztálya a $B^{(2)}$ felosztás bizonyos számú (egy vagy több) osztályának az egyesítése, de legalább egy közülük több, mint egy ilyen osztály egyesítése. Példaképpen legyen a Ξ halmaz a következő:

$$\Xi = \{a, b, c, d, e, f, g, h, i, j,\}$$

és legyen a $B^{(1)}$ felosztás a következő:

$$B_1^{(1)} = \{a, b, c, d\} \quad \text{és} \quad B_2^{(1)} = \{e, f, g, h, i, j\},$$

a $B^{(2)}$ felosztás pedig

$$B_1^{(2)} = \{a, b\}, \quad B_2^{(2)} = \{c, d\}, \quad B_3^{(2)} = \{e, f\}, \quad B_4^{(2)} = \{g, h\},$$
$$B_5^{(2)} = \{i, j\}.$$

Ez esetben a $B^{(1)}$ felosztás a $B^{(2)}$ felosztás durvítása.

---

első fokú konfiguráció rezultáns eleme nemcsak a

C *(маленькая)*

hanem a

C *(эта)*

is. — Vö. I. I. Revzin: Модели языка (Moszkva, 1962) 122—5.

[18] Ez azt jelenti, hogy bármely x szóalak esetén azok a $B^{(1)}(x)$ és $B^{(2)}(x)$ osztályok, amelyekbe az x szóalak, a $B^{(1)}$, illetőleg a $B^{(2)}$ felosztás esetén esik, a következő viszonyban állnak egymással: $B^{(2)}(x)$ osztály vagy valódi részhalmaza a megfelelő $B^{(1)}(x)$ osztálynak, vagy egybeesik vele, de van legalább egy olyan x szóalak, amelyre ez az egybeesés nem áll fenn.

Tekintsünk két felosztást, $B^{(1)}$-et és $B^{(2)}$-t, és legyen $B^{(1)}$ durvább, mint $B^{(2)}$. Ha a

$$B^{(2)}(y) \subset B^{(1)}(x)$$

feltételből következik, hogy

$$B^{(2)}(y) \underset{B^{(2)}}{\sim} B^{(2)}(x),$$

azaz $B^{(2)}(y)$ $B^{(2)}$-ekvivalens $B^{(2)}(x)$-szel, akkor azt mondjuk, hogy $B^{(1)}$ r e g u l á r i s d u r v í t á s a a $B^{(2)}$-nek. Másképpen ez azt jelenti, hogy $B^{(1)}$ felosztás bármely $B_i^{(1)}$ osztálya esetén a $B^{(2)}$ felosztás azon osztályai, amelyeknek $B_i^{(1)}$ az egyenértéke, egymással $B^{(2)}$-ekvivalensek.

A fenti definícióból következik, hogy a típusokra való felosztás reguláris durvítása a környezetekre való felosztásnak, hiszen a típusok éppen az egymással ekvivalens környezeteket egyesítik magukban.

Azt a megfeleltetést (más néven leképezést), amely minden egyes x szóalaknak a B felosztás azon $B(x)$ osztályát felelteti meg, amelybe az x szóalak beletartozik, így jelöljük:

$$x \to B(x).$$

Ez a megfeleltetés egy olyan leképezést is „indukál" (hoz létre), amely minden A mondatnak a megfelelő B-struktúrát, $B(A)$-t felelteti meg. Ezt az indukált leképezést így jelöljük:

$$A \to B(A).$$

A $B(x)$ osztályt az $x \to B(x)$ leképezésben az x szóalak k é p é n e k, hasonlóan a $B(A)$ B-struktúrát az A mondat képének nevezzük az $A \to B(A)$ leképezésben. Azt is mondjuk, hogy az x szóalak az e r e d e t i j e a $B(x)$-nek, A pedig a $B(A)$-nak a fenti leképezésekben.

Legyen adva a Ξ halmaz két olyan felosztása, $B^{(1)}$ és $B^{(2)}$, ahol $B^{(1)}$ durvítása $B^{(2)}$-nek. Jelöljük $B^{(2)} \to B^{(1)}$-gyel azt a leképezést, amely minden $B^{(2)}$ osztálynak a $B^{(1)}$ felosztásnak azon osztályát felelteti meg, amelynek a $B^{(2)}$ részhalmaza. Ha a $B^{(2)} \to B^{(1)}$ leképezésben egy bizonyos $B^{(2)}$-konfiguráció az eredetije egy $B^{(1)}$-konfigurációnak, továbbá e $B^{(2)}$-konfiguráció rezultáns eleme az eredetije a $B^{(1)}$-konfiguráció rezultáns elemének, akkor azt mondjuk, hogy a kérdéses $B^{(2)}$-konfiguráció a $B^{(1)}$-konfiguráció alárendeltje (szubordináltja), a $B^{(1)}$-konfigurációt pedig alárendelőnek hívjuk.

Tegyük fel, hogy az előbbi példában szereplő $B_1^{(1)}$ és $B_2^{(1)}$ egy $B^{(1)}$-konfiguráció, és hogy az összes $B^{(2)}$-konfiguráció, amely ennek a $B^{(1)}$-konfigurációnak alárendeltje, a következő:

$$B_1^{(2)} B_3^{(2)},\ B_1^{(2)} B_4^{(2)},\ B_1^{(2)} B_5^{(2)},\ B_2^{(2)} B_3^{(2)},\ B_2^{(2)} B_4^{(2)},\ B_2^{(2)} B_5^{(2)}.$$

Jegyezzük meg, hogy ha egy $B^{(2)}$-konfiguráció alárendeltje egy $B^{(1)}$-konfigurációnak, akkor a $B^{(2)}$-konfiguráció első elemének eredetije

a $B^{(1)}$-konfiguráció első elemének része, a $B^{(2)}$-konfiguráció második elemének eredetije a $B^{(1)}$-konfiguráció második elemének része és így tovább. Tehát ha például

$$\{a, b\}\ \{g, h\}$$

egy $\bar{B}^{(2)}$-konfiguráció, és

$$\{a, b, c, d\}\ \{e, f, g, h, i, j\}$$

egy $\bar{B}^{(1)}$-konfiguráció, akkor az, hogy a $\bar{B}^{(2)}$-konfiguráció alárendeltje az adott $\bar{B}^{(1)}$-konfigurációnak, nyilvánvalóan azt jelenti, hogy a $\bar{B}^{(2)}$-konfiguráció első elemének (jelöljük $\bar{B}^{(2)}_1$-gyel) eredetije a $\bar{B}^{(1)}$-konfiguráció első elemében (jelöljük $\bar{B}^{(1)}_1$-gyel) van, és a $\bar{B}^{(2)}$-konfiguráció második elemének (jelöljük $\bar{B}^{(2)}_2$-vel) eredetije a $\bar{B}^{(1)}$-konfiguráció második elemében (jelöljük $\bar{B}^{(1)}_2$-vel) van. Vagy másképpen

$$\bar{B}^{(2)}_1 \to \bar{B}^{(1)}_1 \quad \text{és} \quad \bar{B}^{(2)}_2 \to \bar{B}^{(1)}_2.^{18a}$$

Most minden $B^{(2)}$-konfigurációt egy-egy csoportba osztunk, úgy hogy az egyik csoportban mindegyik $B^{(2)}$-konfiguráció kezdőeleme a $B^{(2)}_1$, a másikban a $B^{(2)}_2$ elem legyen. Tehát

1. csoport: $B^{(2)}_1 B^{(2)}_3$, $B^{(2)}_1 B^{(2)}_4$, $B^{(2)}_1 B^{(2)}_5$,

2. csoport: $B^{(2)}_2 B^{(2)}_3$, $B^{(2)}_2 B^{(2)}_4$, $B^{(2)}_2 B^{(2)}_5$.

Minden csoportra érvényes esetünkben, hogy a második helyen álló elemek képe a $B^{(2)} \to B^{(1)}$ leképezésben (minden egyes elemé!) a $B^{(1)}_2$-ben van és utóbbi minden részhalmaza szerepel a $B^{(2)}$-konfigurációk második helyén. Ebben az esetben azt mondjuk, hogy $B^{(1)}_2$-nek a viszonya a $B^{(1)}_1$-hez harmadfajú a $B^{(2)}$ felosztásra vonatkozóan. Ilyen viszony t e l j e s g r a m m a t i k a i f ü g g e t l e n s é g r e vall $B^{(1)}_1$ és $B^{(1)}_2$ között a $B^{(2)}$ felosztásra vonatkozóan.

Vegyük például a szóalakok halmazának következő részhalmazát:

$$\{\textit{szépen, világosan, látok, hallok, látunk, nézünk, ettünk, ittunk}\}$$

Jelöljük ezeket a szóalakokat az egyszerűség kedvéért rendre x_1, x_2, x_3, x_4, x_5, x_6, x_7 és x_8-cal.

Legyen most a $B^{(1)}$ felosztás a típusokra való felosztás, azaz legyen $B^{(1)} = T$ és legyen a $B^{(2)}$ felosztás a családokra való felosztás, azaz legyen $B^{(2)} = C$. A fentiek szerint a következő osztályokat kapjuk:

[18a] Vö. N. Jaurisová—G. Jauris, Užiti teorie množin v jazykovědě: Slovo a Slovesnost 21. 31—41 (1960).

$$B^{(1)} = \{x_1, x_2\} \cup \{x_3, x_4, x_5, x_6, x_7, x_8\}$$
$$B_1^{(1)} \qquad B_2^{(1)}$$

és

$$B^{(2)} = \{x_1, x_2\} \cup \{x_3, x_4\} \cup \{x_5, x_6\} \cup \{x_7, x_8\}.$$
$$B_1^{(2)} \qquad B_2^{(2)} \qquad B_3^{(2)} \qquad B_4^2$$

Legyen most ismét

$$B_1^{(1)}\ B_2^{(1)}$$

$B^{(1)}$-konfiguráció, akkor az összes olyan $B^{(2)}$-konfiguráció, amely ennek alárendeltje, a következő:

$$\{x_1, x_2\}\ \{x_3, x_4\},$$
$$\{x_1, x_2\}\ \{x_5, x_6\},$$
$$\{x_1, x_2\}\ \{x_7, x_8\}.$$

A $B^{(2)}$-konfigurációk rezultáns elemei a második elemek. Tekintve, hogy a $B^{(2)}$-konfigurációk második elemének képe mind megtalálható $B_2^{(1)}$-ben (ige-halmaz), a $B_2^{(1)}$ viszonya a $B_1^{(1)}$-hez harmadfajú a $B^{(1)}$ felosztásra vonatkozóan. Ez más szóval azt jelenti, hogy az igék típusa és a határozószók típusa a magyarban grammatikailag független.

Most térjünk vissza ismét általános példánkhoz és tegyük fel, hogy az összes olyan $B^{(2)}$-konfiguráció, amely az adott $B^{(1)}$-konfiguráció alárendeltje, a következő:

1. csoport: $B_1^{(2)}\ B_3^{(2)}$, $B_1^{(2)}\ B_4^{(2)}$,

2. csoport: $B_2^{(2)}\ B_3^{(2)}$, $B_2^{(2)}\ B_5^{(2)}$,

Ismét két csoportot kaptunk. Mindkét csoportra érvényes, hogy a második helyen álló elemek — annak ellenére, hogy egynél több, vagyis két különböző elem áll a második helyen — mind eredetijei a $B_2^{(1)}$-nek a $B^{(2)} \rightarrow$ $\rightarrow B^{(1)}$ leképezésben, de utóbbinak nem minden részhalmaza szerepel e $B^{(2)}$-konfigurációk második helyén. Vagyis van olyan elem, amely a fenti konfigurációban nem szerepel, és amelynek képe a $B_2^{(1)}$ részhalmaza. (Tudniillik, az első csoportban a $B_5^{(2)}$, a második csoportban viszont a $B_4^{(2)}$.)

Ilyen esetben azt mondjuk, hogy a $B_2^{(1)}$ és $B_1^{(1)}$ között másodfajú viszony áll fenn a $B^{(2)}$ felosztásra vonatkozóan. Ez azt jelenti, hogy a $B_2^{(1)}$ és $B_1^{(1)}$ között r é s z l e g e s g r a m m a t i k a i f ü g g ő s é g r ő l van szó.

Vegyük a szóalakok halmazának következő részhalmazát:

{nézek, tanítok, lányt, fiút, lányokat, fiúkat, lányok, fiúk}.

Jelöljük e halmaz elemeit egyszerűség kedvéért ismét x_1, x_2, x_3, x_4, x_5, x_6, x_7 és x_8-cal.

Legyen ismét $B^{(1)} = T$ és $B^{(2)} = C$, vagyis a $B^{(1)}$ felosztás egyezzék meg a típusokra való felosztással s a $B^{(2)}$ felosztás pedig a családokra való felosztással. Tehát

$$B^{(1)} = \underset{B_1^{(1)}}{\{x_1, x_2\}} \cup \underset{B_2^{(1)}}{\{x_3, x_4, x_5, x_6, x_7, x_8\}}$$

és

$$B^{(2)} = \underset{B_1^{(2)}}{\{x_1, x_2\}} \cup \underset{B_2^{(2)}}{\{x_3, x_4\}} \cup \underset{B_3^{(2)}}{\{x_5, x_6\}} \cup \underset{B_4^{(2)}}{\{x_7, x_8\}} .$$

$B_1^{(1)}$ $B_2^{(1)}$ ismét $B^{(1)}$-konfiguráció és az összes olyan $B^{(2)}$-konfiguráció, amely a $B^{(2)} \to B^{(1)}$ leképezésben a $B^{(1)}$-konfiguráció alárendeltje, a következő

$$\{x_1, x_2\}\ \{x_3, x_4\}$$

és

$$\{x_1, x_2\}\ \{x_5, x_6\} .$$

Nyilvánvaló ugyanis, hogy a

$$\{x_1, x_2\}\ \{x_7, x_8\}$$

nem lehet $B^{(2)}$-konfiguráció.

A $B^{(2)}$-konfigurációk második helyen álló elemei megtalálhatók a $B_2^{(1)}$-ben, de $B_2^{(1)}$ nem mindegyik eleme szerepel a $B^{(2)}$-konfigurációk második elemeként. Tehát azt mondhatjuk, hogy az igék típusa ($B_1^{(1)}$) és a főnevek típusa ($B_2^{(1)}$) a $B^{(2)}$-felosztásra vonatkozóan másodfajú viszonyban áll, ami nyelvészeti szempontból egy a teljesnél gyengébb grammatikai függést jelent.

Térjünk ismét vissza eredeti általános példánkhoz és tegyük fel, hogy az összes olyan $B^{(2)}$-konfiguráció, amely az adott $B^{(1)}$-konfiguráció alárendeltje, a következő:

1. csoport: $B_1^{(2)}$ $B_3^{(2)}$,

2. csoport: $B_2^{(2)}$ $B_5^{(2)}$.

Minden csoportban csak egy elem található, amelynek képe a $B^{(2)} \to B^{(1}$ leképezésben a $B_2^{(1)}$-ben van, illetőleg minden csoportban csak egy eredetije található $B_2^{(1)}$-nek. Ilyen esetben a $B_2^{(1)}$ viszonya a $B_1^{(1)}$-hez első fajú a $B^{(2)}$-felosztásra vonatkozóan. Ez a viszony egyértelmű nyelvtani függőséget jelent $B_2^{(1)}$ és $B_1^{(1)}$ között a $B^{(2)}$ felosztásra vonatkozóan.

Vegyünk ismét konkrét példát! Legyen most az összes szóalakok halmazának tekintett részhalmaza a következő:

$$\{rossz, jó, enni, állni, eszik, áll, evett, állt\}$$

Egyszerűség kedvéért jelöljük az egyes szóalakokat ismét x_1, x_2, x_3, x_4, x_5, x_7 és x_8-cal.

A $B^{(1)}$-felosztás ismét legyen a típusokra való felosztás, a $B^{(2)}$-felosztás pedig a családokra való felosztás. Akkor

$$B^{(1)} = \underset{B_1^{(1)}}{\{x_1, x_2\}} \cup \underset{B_2^{(1)}}{\{x_3, x_4, x_5, x_6, x_7, x_8\}}$$

és

$$B^{(2)} = \underset{B_1^{(2)}}{\{x_1, x_2\}} \cup \underset{B_2^{(2)}}{\{x_3, x_4\}} \cup \underset{B_3^{(2)}}{\{x_5, x_6\}} \cup \underset{B_4^{(2)}}{\{x_7, x_8\}}.$$

Az egyetlen olyan $B^{(2)}$-konfiguráció, amely a $B^{(2)} \to B^{(1)}$ leképezésben a $B^{(1)}$-konfiguráció alárendeltje, a következő

$$\{x_1, x_2\}\ \{x_3, x_4\}.$$[18b]

Ez azt jelenti — mivel e $B^{(2)}$-konfiguráció második eleme megtalálható a $B_2^{(1)}$-ben, de $B_2^{(1)}$ nem minden részhalmaza található meg a $B^{(2)}$-konfiguráció második elemeként, továbbá csak egyetlen ilyen $B^{(2)}$-konfiguráció létezik — hogy a $B_2^{(1)}$ és $B_1^{(1)}$ között a $B^{(2)}$ felosztásra vonatkozóan első fokú viszony áll fenn. A melléknévtípus tehát jelen esetben egyértelműen függ az igék típusától.[18c]

A fenti példák tulajdonképpen csak a magyar nyelv szóalakkészletének egy nyolc elemű részhalmazára vonatkoznak. Az eljárás azonban kiterjeszthető több elemű részhalmazokra is. Bonyolultabb esetekben természetesen a konfigurációk nem két, hanem több elemből állnak. Ilyenkor a konfigurációk egy határozott helyén álló elemnek általában nemcsak egy, hanem több más elemhez való viszonyát kell vizsgálnunk. Sőt a konfi-

[18b] Ennek a példának egy kis szépséghibája van. Igaz ugyan, hogy

$$\{x_1, x_2\}\ \{x_3, x_4\}$$

konfigurációnak tekinthető a magyarban, de nem a családokra való felosztás, hanem a típusokra való felosztás esetén. Tudniillik, ha $B = T$, vagyis a típusokra való felsorolás esetén

Jó enni

T-konfiguráció, melynek rezultáns eleme például a T (esik). Jobb példa hiányában követtem el ezt a kis pontatlanságot Egyébként a magyar nyelvben igen nehéz lenne a harmadfajú viszonyra jó példát találni, mert a magyar nyelv struktúrája egészen más, mint azoké a nyelveké, amelyek Kulagina szeme előtt lebeghettek, amikor elméletét megalkotta.

[18c] Ha valamely $B^{(1)}$-konfiguráció két eleme között egyik ismertetett reláció sem áll fenn, vagyis a konfiguráció elemei sem első, sem másod-, sem harmadfajú viszonyban nem állnak, akkor azt mondjuk, hogy ezek az elemek az adott konfigurációban nem állnak egymással grammatikai kapcsolatban.

Oroszban első fajú viszonyban áll, ha a típusokra és családokra való felosztást vesszük, a melléknév- és a főnévtípus (egyeztetés). Másodfajú viszonyban áll például a következő kapcsolat „ige + főnév", ahol a családokra való felosztás az ige vonzatai szerint felosztást jelenti. Harmadfajú viszonynak az úgynevezett примыкание felel meg (például: можно играть).

guráción belül nemcsak egy kitüntetett elemnek a konfiguráció többi elemeihez való viszonyát vizsgálhatjuk, hanem több kitüntetett elemnek a nem kitüntetett elemekhez való viszonyát. Ahhoz, hogy a nyelvi struktúra és a struktúra elemei közti összefüggések helyes leírásához jussunk, szükséges is a fenti eljárásnak a teljes nyelvre való kiterjesztése. Természetesen nem tekinthetjük az összes szóalakok halmazát közvetlenül, mindenesetre tekinthetjük annak nagyobb részhalmazát; ugyanakkor a nyelv már ismert strukturális tulajdonságait is figyelembe kell vennünk. Kívánatos lenne, ha a magyar nyelv rendszerét ebből a szempontból is megvizsgálnók.

Természetesen, ha a magyar nyelv egészét vizsgáljuk, általában a fentiekkel ellentétes megállapításokhoz is juthatunk. Sőt, ez akkor is igaz, ha a magyar nyelvnek egy ugyanolyan számosságú, de más részhalmazát vesszük.[18d]

Kulagina erről lényegében csak a következőket mondja: Tekintsük az összes D-konfigurációt, amely alárendeltje egy adott B-konfigurációnak (amennyiben ilyenek találhatók). Az összes D-konfigurációt úgy foglaljuk csoportokba, hogy minden csoportban az első helyen ugyanaz a D_{1i} elem álljon.

Ha minden konfigurációban, amely ugyanabba a csoportba tartozik, a második helyen egy és ugyanaz a D_{2j} elem áll, akkor azt mondjuk, hogy az adott B-konfigurációban B_2 első fajú viszonyban áll B_1-gyel a D felosztásra vonatkozóan (ahol B_1, B_2 a B felosztás két osztálya, B-konfiguráció elemei, és azt is tudjuk, hogy a B felosztás durvítása a D felosztásnak).

Ha azokban a D-konfigurációkban, amelyek ugyanabba a csoportba tartoznak, a második helyen bizonyos B_{2jk}-k állhatnak, amelyeknek száma 1-nél nagyobb, de n-nél kisebb akkor azt mondjuk, hogy az adott B-konfigurációban B_2 másodfajú viszonyban áll B_1-gyel a D felosztásra vonatkozóan.

Ha azokban a D-konfigurációkban, amelyek ugyanabba a csoportba tartoznak, a második helyen tetszés szerinti D_{2j} található, $(1 \leq j \leq n)$ akkor azt mondjuk, hogy az adott B-konfigurációban B_2 harmadfajú viszonyban áll B_1-gyel a D felosztásra vonatkozóan.

[18d] Hogy ez mennyire igaz, csak a fenti példában szereplő halmazt, mint az összes szóalakok részhalmazát kell a következőképpen megválasztanunk:

$\{$*nézek, tanítok, lányt, fiút, lányok, fiúk, lányokat, fiúkat*$\}$.

Ha most ismét $B^{(1)} = T$ és $B^{(2)} = C$, vagyis a $B^{(1)}$ felosztás a típusokra való felosztás és $B^{(2)}$ a családokra való felosztás, akkor ugyanazt kapjuk, mint fent. A $B^{(2)}$-konfigurációk viszont a következők (ugyanazt a jelölést használva, mint fent)

$$\{x_1, x_2\}\ \{x_3, x_4\},$$
$$\{x_1, x_2\}\ \{x_5, x_6\},$$
$$\{x_1, x_2\}\ \{x_7, x_8\}.$$

Vagyis azt kaptuk, hogy a $B^{(2)}$-konfigurációk második helyén álló elemek képei mind benne vannak a $B^{(1)}_2$-ben és $B^{(1)}_2$-nek nincs olyan részhalmaza, amelynek eredetije ne lenne második eleme egy $B^{(2)}$-konfigurációnak, amiből azt a következtetést vonhatnánk le, hogy a magyar nyelvben az igék típusa és a főnevek típusa a $B^{(2)}$ felosztásra vonatkozóan harmadfajú viszonyban áll, ami pedig teljes grammatikai függetlenséget jelent. Ez viszont ellenkezik azzal, amit fent mondottunk. Ha tehát a magyar nyelvből különböző részhalmazokat veszünk, más-más eredményhez juthatunk. A fenti példák csak illusztratív célokat szolgálnak, helyes eredményre csak akkor jutnánk, ha a magyar nyelv egész rendszerét tekintenénk és a szóalakok halmazán kívül még a magyar nyelv bizonyos egyéb strukturális tulajdonságait is figyelembe vennénk.

A magyar nyelv teljes egészének vizsgálatánál persze már előre tudjuk, hogy a grammatikai függés a magyarban korántsem olyan sokrétű, mint azokban a nyelvekben, ahol a nemek közt különbséget teszünk, ahol például a jelző és a jelzett szó közt is első fajú viszony áll fenn. Ha valamely nyelv egész rendszerére kiterjesztenénk e vizsgálatokat, akkor nyilvánvalóvá válna, hogy Kulaginának az első-, másod- és harmadfajú viszonyokra való elmélete nem minden nyelv esetén mond valamit a grammatikai függés milyenségéről. Tulajdonképpen ebben az esetben is arról van szó, hogy Kulagina a grammatikai függés modelljét állítja fel. A modell viszont — tekintve, hogy minden esetben absztrakció eredménye — nem veheti figyelembe az összes lehetőségeket, jelen esetben a grammatikai függés sokrétűségét. Ez azonban nem a modell hibája. Gondoljunk arra, hogy a hagyományos grammatikák sem hibátlanok, és amellett, hogy elsősorban nem a nyelv rendszerével foglalkoznak, még pontatlanok és következetlenek is. Elképzelhető persze, hogy a grammatikai függésről jobb modelleket is tudunk készíteni.[19]

Kulagina elméletének tehát leglényegesebb megállapításait részletesen megvizsgáltuk. A tisztán halmazelméleti módszerek mellett bizonyos esetekben, például annak eldöntésénél, hogy egy adott mondat kitüntetett-e, vagy nem, más módszerekre, elsősorban matematikai-logikai módszerekre van szükség.

2. A matematikai-logikai modell

2.1. A Lambek-féle szintaktikus típusok

A matematikai nyelvészet egyik általános jellegzetessége, hogy formalizál. A nyelvészetben a formalizálás mindig azt jelenti, hogy egy-egy nyelvi elem jelölésére bevezetünk egy értelmes jelet, melyről tehát tudjuk, hogy mit jelöl, hogy mi a jelentése. Majd eltekintünk a bevezetett jel jelentésétől, bizonyos műveleteket definiálunk az egyes jelek közt, tehát számolunk velük. Ennek folytán olyan rendszert kapunk, amely már formális, mesterséges nyelvnek tekinthető. A formalizálásnak azonban mindig csak akkor van meg a jogosultsága, ha eredményeink végül is nyelvészetileg interpretálhatók. Példaképpen nézzük meg a Lambek-féle szintaktikus típusokat, illetőleg az azokon alapuló Lambek-féle rendszert.[19a]

[19] A Kulagina-féle elmélet kritikai elemzése és a konfiguráció-elmélet kapcsolata az amerikai nyelvészek struktúravizsgálataival a tárgya a tihanyi „Matematika alapjai, matematikai gépek, és alkalmazásaik" című kollokviumon tartott előadásomnak (Tihany, 1962. szeptember 11—15). Az előadás címe: Bemerkungen zur Anwendung mengentheoretischer Methoden in der Sprachwissenschaft. — Most itt csak arra utalok, hogy míg a konfigurációelmélet alapján történő analízis a bonyolultabb mondatokat „bontogatja le" egyszerűbbekké, Bar-Hillel, Chomsky és mások egyszerűbb, úgynevezett magmondatokból építik fel bizonyos szabályok segítségével a bonyolultabb mondatokat. A két eljárás közti különbség tulajdonképpen csak az „irányban" van, mindenesetre az utóbbi látszik jobban kezelhetőnek, és eddig is már több eredményt hozónak.

[19a] Vö. J. Lambek következő három művével: The mathematics of sentence structure: American Mathematical Monthly 65.154—70 (1958); On the calculus of syn-

Lambek a szintaktikus típusok fogalmát a szóalakok, illetőleg kifejezések mondatbeli funkciójának leírására vezeti be (kifejezésen szóalakok sorozatát érti). Két primitív típust vezet be (későbbi dolgozataiban többet): minden grammatikailag helyes kijelentő mondat típusa *s*; minden egyes szám alanyesetű főnév (tulajdonképpen csak tulajdonnév, illetőleg mindaz, ami azzal helyettesíthető anélkül, hogy a mondat grammatikailag helyes volta megszűnne) típusa *n*; ehhez később hozzáveszi a főnévi igenév típusát, *i*-t. Természetesen nemcsak három primitív típust lehet felvenni, Lambek is kísérletezik többel (utolsó dolgozatában hármat tart meg). Ezekből a primitív típusokból összetett típusok konstruálhatók három formális művelettel: szorzással, baloldali és jobboldali osztással.
Így például

Péter játszik

típusa *s*, *Péter* típusa *n*, *játszik* típusát n\s [olvasd: n felett s]-nek veszszük, mert *játszik* jobbról „csinál" mondatot *Péter*-ből, ezt az osztást pedig baloldali osztásnak nevezzük.

Kis Péter játszik

típusa is *s*, *játszik* típusa, mint láttuk, n\s, de *Kis Péter* típusa is *n*, tehát *Kis* típusa n/n [olvasd: n alatt n]: *Kis* ugyanis balról „csinál" újabb n-típusú kifejezést *Péter*-ből; e műveletet jobboldali osztásnak nevezzük.

A kis Péter játszik ott

mondathoz a következő típus-sorozat rendelhető hozzá: *A* típusa ugyanaz, mint *kis* típusa, mert hiszen n-típusú *kis Péter*-ből újabb n-típusú kifejezést csinál, vagyis *A kis Péter*-t. Típusa tehát n/n, ugyanaz *kis* típusa, *Péter* típusa *n*, *játszik* típusa n\s, *ott* s-ből (mondatból) jobbról „csinál" újabb mondatot, tehát típusa s\s. Ezek szerint tehát

A kis Péter játszik ott

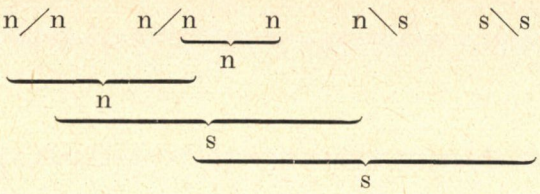
mondatnak megfelelő típus-sor a következő

n/n, n/n, n, n\s, s\s.

Itt két szomszédos típus mindig összeszorozható és az aritmetikából ismert módon egyszerűsíthető, vagyis, ha egy „jobb törtet" (például n/n) jobbról, vagy ha egy „bal törtet" (például n\s) balról szorzunk a nevezőjével (n-nel), a számlálóját kapjuk (az első esetben n-et, a másodikban s-et):

---

tactic types: Structure of language and its mathematical aspects (Szerk. R. Jakobson, Providence, R. I., 1961) 166—178; Contributions to a mathematical analysis of the English verb-phrase: Journal of the Canadian Linguistic Association 5. 83—9 (1959).



Egy jelsorozat (típus-sorozat) akkor és csak akkor szintaktikailag konnex, ha van olyan egyszerűsítési lehetőség, hogy annak végeredménye s legyen. A szintaktikai konnexitásból még nem következik, hogy a típusoknak valamely nyelvben van megfelelőjük, de ha valamely nyelvben egy grammatikailag helyes mondat típus-sorozatát felírjuk, akkor arról általában könnyű belátni, hogy az szintaktikailag konnex.

Bebizonyítható a következő szabályok általános érvényessége:

(1) $x \to x$,

(2) $(xy)z \to x(yz)$, (2′) $x(yz) \to (xy)z$,

(3) $\dfrac{xy \to z}{x \to z/y}$, (3′) $\dfrac{xy \to z}{y \to x\backslash z}$,

(4) $\dfrac{x \to z/y}{xy \to z}$, (4′) $\dfrac{y \to x\backslash z}{xy \to z}$,

(5) $\dfrac{x \to y,\ y \to z}{x \to z}$,

ahol $x \to y$ azt jelenti, hogy minden x-típusú kifejezés egyidejűleg y-típusú is. A vízszintes vonal pedig azt jelenti, hogy a vonal alatt levő következmény következik a vonal felett levő egy vagy több premisszából. Az összetett típusokhoz a következő rekurzív definíció vezet: ha x és y típusok, akkor xy, x/y és $y\backslash x$ is azok. Minden kifejezéshez bizonyos típusok tartoznak (egy, vagy több). Ha az A kifejezés típusa x, és a B kifejezés típusa y, akkor az AB kifejezés típusa $xy \cdot z/y$-nal jelöljük minden olyan A kifejezésnek a típusát, amelyre bármely y típusú B kifejezés esetén AB típusa z. És $x\backslash z$-vel jelöljük minden olyan B kifejezés típusát, amelyre minden x-típusú A kifejezés esetén AB típusa z.

A fenti szabályok közül az (1), (2) és (2′) triviálisan teljesül. A többi viszont a fenti megjegyzések felhasználásával könnyen bebizonyítható. Bizonyítsuk be például a (3) szabályt.

Legyen $xy \to z$, és A típusa x. Akkor minden y-típusú B-re az AB kifejezés típusa xy, tehát z is. Így A típusa z/y, ezért $x \to z/y$. Hasonlóképpen bizonyítható a (3′), (4) és (4′) szabály is. Az (5) szabály ismét triviálisan teljesül.

A fenti rendszer maga is formális nyelvnek tekinthető, mely primitív típusok halmazából három formális műveleti jel és a „$\to$” reláció segítségével épül fel. E nyelvben a (primitív és az összetett) típusok játsszák a

tulajdonnevek, és az $x \to y$ alakú formulák a kijelentő mondatok szerepét. Így e nyelv jeleinek is adható típus: a primitív típusoknak n, a szorzás, bal- és jobboldali osztás jeleinek $(n\backslash n)/n$, a „$\to$" jelnek pedig $(n\backslash s)/n$. — A zárójelet ebben az esetben félreértés veszélye nélkül el lehet hagyni, például akár $n\backslash(n/n)$, akár $(n\backslash n)/n$ helyett lehet $n\backslash n/n$-et írni, mert $(n\backslash n)/n \to n\backslash(n/n)$ és $n\backslash(n/n) \to (n\backslash n)/n$.

Ha az (1), (2), (2′) szabályokat axiómáknak tekintjük, és a (3), (3′), (4), (4′) és (5) összefüggéseket következtetési szabályoknak, akkor deduktív rendszert kapunk, amelyet Lambek újabban asszociatív szintaktikus kalkulusnak nevez.

A fentiekben példát láttunk a formalizálásra. Formalizálás helyett Bar-Hillel, illetőleg Lambek aritmetizálásról beszél. Mi ennek a formalizálásnak vagy aritmetizálásnak az előnye? Előnye Bar-Hillel szerint elsősorban abban áll, hogy teljesen mechanikus módszert nyújt nyelvi kifejezések struktúrájának megállapítására. Erre Bar-Hillel szerint többek között a gépi fordításnál lehet szükség. Bebizonyították ugyanis, hogy konstruálható olyan gép, amely a forrásnyelv bármely mondatának struktúráját meg tudná állapítani, feltéve, hogy ennek a nyelvnek a szintaxisát megfelelő formában (amelyet Bar-Hillel operatívnek nevez) megtanították a gépnek. Ilyen operatív szintaxis konstruálásának első lépcsőfoka az elemi kategóriák (a Lambek-féle primitív típusok) megállapítása.

Lambek azon a véleményen van, hogy az aritmetizálás hozzásegít egy segédnyelv szintaxisának megkonstruálásához, tudniillik ennek a segédnyelvnek tökéletes szintaxissal kell rendelkeznie. A segédnyelvnek aztán ismét a gépi fordításnál van szerepe, ahol közvetítő-nyelvként lehetne használni.

Kétségtelen, hogy ma már a nyelvtudományban is nyilvánvalóvá vált a formalizálás szükségessége. Bar-Hillel[20] kimutatta ugyanis, hogy a legtöbb nyelvtani fogalom (kategória, mint például mondat, szó, alany, tárgy stb.) nem definiálható közvetlenül és explicit módon. Ezeket a fogalmakat kizárólag rekurzív definíciókkal lehetne definiálni.[21] Természetesen ennek előfeltétele — legalábbis egyik előfeltétele — egy bizonyos mértékű formalizálás. A formalizálás általában azért is előnyös, mert bonyolultabb összefüggéseket is képes egyszerű, áttekinthető alakba öltöztetni. Igen fontos továbbá, hogy a formalizált rendszerben, bizonytalan, nem egészen pontos szabályok, meg nem fogalmazott sejtések nem játszhatnak szerepet. Mivel a formalizált rendszer eszerint csak egzakt fogalmakkal dolgozhat, a formalizálás hozzásegít például a grammatikai fogalmak tisztázásához. Sok esetben formalizált rendszer segítségével — tekintve, hogy az ily módon nyert eredmények esetleg többféleképpen is értelmezhetőek — több problémát is oldhatunk meg egyszerre.

[20] Vö. Bar-Hillel, Recursive definitions in empirical sciences: Proceedings of the Eleventh International Congress of Philosophy 5. 160—5 (1953).

[21] A rekurzív definíciókra, illetőleg rekurzív függvényekre vonatkozóan lásd Péter Rózsa: Rekursive Funktionen (Bp., 1957). — Grammatikai fogalmak rekurzív definíciójával Revzin is próbálkozik, a rekurzív definíciók algoritmikus jellegére támaszkodva. Lásd I. I. Revzin, О логической форме лингвистических определений: Применение логики в науке и технике (Moszkva, 1961).

2.2. Az axiomatizálható grammatikák fogalma

A grammatikák halmazelméleti vagy matematikai-logikai modelljei általában axiomatizált rendszerek. Valamely matematikai elmélet axiomatikus felépítése azzal kezdődik általában, hogy felsoroljuk az elmélet alapfogalmait, vagyis azokat a fogalmakat, amelyeket a rendszeren belül nem definiálunk (például Kulagina — mint említettük — matematikai nyelvelméletében alapfogalomnak tekinti a szóalak, a környezet és a kitüntetett mondat fogalmát); továbbá felsoroljuk az elmélet axiómáit, vagyis azon tételeit, amelyeket nem bizonyítunk be. A rendszer további fogalmait már az alapfogalmak segítségével definiáljuk, további tételeit az axiómák segítségével bizonyítjuk. Sok esetben azonban csak azt állapítjuk meg egy rendszerről, hogy axiomatizálható. A tényleges axiomatizálás ugyanis nem mindig könnyű probléma.

Egy axiomatizálható grammatikai modell felépítését a következőképpen képzelhetjük el. Jelöljük G-vel axiomatizálható grammatikai modellünket. G a következő tulajdonságokkal rendelkezhet.

a) Bizonyos elemi jeleket, vagy elemi kifejezéseket tartalmaz.

b) A többi kifejezést ezekből bizonyos képzési szabályok segítségével kaphatjuk meg. (Például ha Lambek rendszerére gondolunk, akkor a primitív típusokból az összetett típusokat — mint fentebb láttuk — bizonyos műveletek segítségével kaphatjuk meg.)

De elképzelhető a következő fajta konstrukció is — amely végső soron Chomskyra[23] vezethető vissza.

Vezessük be a következő jelöléseket:

A = határozott névelő;
Ad = melléknév;
S = többes szám alany esetű főnév (főnévi szintagma);
Adv = határozószó;
O = egyes számú tárgy esetű főnév;
V = többes számú, 3. személyű ige tárgyatlan alakja;
Np = főnévi csoport;
Vp = igei csoport;
( ) = a lehetőség kifejezése, fakultatív szabály;
$\rightarrow$ = „helyettesítsd";
$+$ = két kifejezés együtt.

Ha most ezeknek a jeleknek a halmazából indulunk ki, akkor például a következő szabályt állapíthatjuk meg:

Np + Vp kifejezése lesz G-nek, ha
$Np \rightarrow (A +) S$,
$S \rightarrow Ad + S$,
$Vp \rightarrow (Adv +) V$,
$V \rightarrow V + O$.

[23] Vö. N. Chomsky: Syntactic Structures (Hága, 1957).

Például:

A fiatal gyerekek szívesen játszanak bujócskát.

c) A fenti szabályokhoz hozzá szokták venni még az úgynevezett morfofonématikus szabályokat, amelyek az adott kifejezést fonéma-, illetőleg graféma-láncokká alakítják.

d) A legfontosabbak a transzformációs szabályok, amelyek új kifejezéseket hoznak létre már meglevő kifejezések átalakításával és ezért bizonyos összefüggéseket fejeznek ki kifejezések, illetőleg mondatok között.

Vezessük be a következő jelöléseket:

A = határozott névelő;
D = egyes vagy többes számú részes esetű főnév;
S_g = egyes számú főnév 3. személyű birtokos személyraggal;
S = egyes vagy többes számú alanyesetű főnév;
( ) = szintagma.

Akkor birtokos szerkezetek a következő transzformációval kaphatók kijelentő mondatokból:

$$(A + D + van + S_g) \rightarrow A + S + S_g).$$

Például:

Az ajtónak van zárja → az ajtó zárja

Hasonlóképpen érvényes:

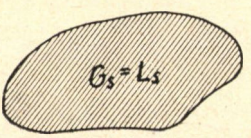
$$(A + S + V) \rightarrow (A + S + S_v + R)$$

ahol A, S jelentése, mint fent, S_v igéből képzett főnév, R egyes szám 3. személyű birtokosrag, V egyes szám harmadik személyű ige. Például:

A kutya ugat → a kutya ugatása.

A G-ben képzett helyes kifejezések halmazát jelöljük Gs-sel. Tekintsünk ugyanakkor egy természetes nyelvet és jelöljük L-lel. Ennek a természetes nyelvnek a mondatainak halmazát jelöljük Ls-sel. Általában a fenti módon konstruált grammatika lehetővé teszi L mondatainak képzését. Ideális eset áll elő, ha G lehetővé teszi L minden mondatának képzését, és Gs minden kifejezése mondata L-nek, vagyis amikor Gs = Ls. Szemléletesen:

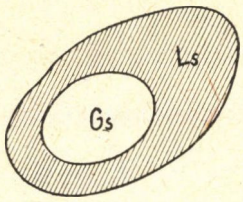
Gs = Ls.



A természetes nyelvek komplikáltsága azonban ezt az ideális esetet gyakorlatilag kizárja. Ezért meg kell elégednünk különböző approximációs lehetőségekkel:

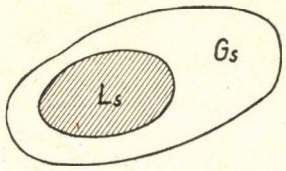
a) Gs minden kifejezése mondata Ls-nek, de Ls-nek van olyan mondata, amely nem kifejezése Gs-nek. Képletben és szemléletesen:

$$Gs \subset Ls.$$



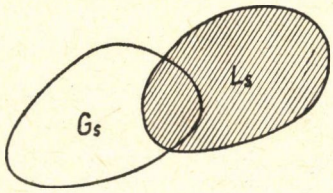
b) Ls minden mondata kifejezése a Gs-nek, de Gs nem minden kifejezése mondata Ls-nek. Képletben és szemléletesen:

$$Gs \supset Ls.$$



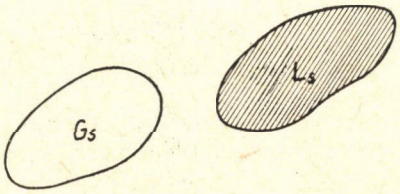
c) Gs-nek van olyan kifejezése, amely mondata Ls-nek, de Gs nem minden kifejezése mondata Ls-nek, továbbá Ls nem minden mondata kifejezése Gs-nek. Képletben és szemléletesen:

$$Gs \cap Ls \neq O,$$
$$Gs \not\subset Ls,$$
$$Gs \not\supset Ls,$$
$$Gs \neq Ls.$$



d) Ezzel szemben nem elégedhetünk meg a következő helyzettel: Gs-nek nincs olyan kifejezése, amely mondata lenne Ls-nek és viszont, Ls-nek nincs olyan mondata, amely kifejezése lenne Gs-nek. Képletben és szemléletesen:

$$Gs \cap Ls = O.$$



A különböző approximációs lehetőségekre Harwood[22] dolgozott ki egzakt módszereket. Az approximációknál nem a teljes Ls-t vette figyelembe, hanem csak annak egy-egy részhalmazát, amelyet Ls_1-gyel jelölhetünk. Ls részhalmazait *korpusznak* szokták nevezni. A probléma tehát úgy is felvethető, hogy hogyan kell kiválasztani Ls-ből olyan korpuszt, amely optimális approximáció alapjául szolgálhat.

A vázolt módszerek ismét példák a halmazelméleti, illetőleg matematikai-logikai módszerek alkalmazására. A különböző modellek tehát különböző approximációi a nyelvi valóságnak. A nyelv matematikai modelljei azonban már eddig is nagymértékben gazdagították a nyelvtudományt. Talán ez a vázlatos ismertetés, valamint kritikai elemzés is hozzájárul ahhoz, hogy az eddig csak a hagyományos módszerekkel kutató nyelvész felismerje a matematikai módszerek által adott lehetőségeket. Az eredményeket nem szabad túlbecsülnünk, hisz sok esetben még csak kísérletekről van szó. De az új koncepció már szinte az egész világon megtermékenyítette a nyelvtudományt.*

[22] Vö. különösen F. W. Harwood, Axiomatic syntax: Language 31. 409—14 (1955).

* Időközben magyar fordításban is megjelent Kulaginának a 2. számú jegyzetben hivatkozott műve. O. Sz. Kulagina, A nyelv egy halmazelméleti modelljéről: Matematikai nyelvészet és gépi fordítás a Szovjetunióban (Bp., 1964., Időszerű Műszaki Dokumentációs Kérdések, 6. szám) 123—146. — *A szerkesztők.*

Dömölki Bálint—Fónagy Iván—Szende Tamás

Köznyelvi hangstatisztikai vizsgálatok*

1. A vizsgálatok jellege és módja

Felmerült a kérdés: vajon mi szükség van arra, hogy Tarnóczy Tamás úttörő fonéma-statisztikái, Vértes Edit alaposan átgondolt, sokoldalú, igényes tanulmányai után a magyar köznyelv hangstatisztikai vizsgálatát ismét műsorra tűzzük?[1] Hogy mi elektronikus számológépekkel dolgozunk és elődeink manuálisan végezték a munkát, nyilvánvalóan nem teszi ezt még indokolttá.

1.1. Voltaképpen mégis az eszközök tökéletesedése volt ezúttal is az első indíték. Csábítottak bennünket az elektronikus számológépek által kínált hatalmas lehetőségek. Ha gépi kapacitást tudunk biztosítani a nyelvstatisztikai munkákra, nincsen akadálya, hogy viszonylag rövid időn belül feldolgozzuk több költő művének jelentékeny részét, prózaírók, tudósok írásait, spontán előadások, beszélgetések szövegét. Megfelelő memóriával rendelkező számológép segítségével úgyszólván egy füst alatt hangtani, verstani, szótani, nyelvtani szempontból elemezhetnénk a szövegeket.

Aczél István elvtárs, a Magyar Tudományos Akadémia Számítástechnikai Központjának akkori vezetője rendelkezésünkre bocsátotta meghatározott időre Intézete M—3-as számológépét. Így jutottunk arra az elhatározásra, hogy megkíséreljük fokozatosan realizálni ezeket az adott lehetőségeket.

1.2. Nem tekinthettük eleve bizonyítottnak, hogy a fonémák eloszlása független a műfajtól, az egyéni stílustól. Ezért úgy döntöttünk, hogy különböző műfajú szövegeket veszünk sorra, és minden műfajon belül két ellentétes temperamentumú szerzőt szemelünk ki.

Eddigi eredményeink Szabó Lőrinc „Tárlat: a Sátán műremekei" című ciklusán és Juhász Gyula 1924 és 1934 között írt válogatott versein (59

*Dömölki Bálint munkája a 3. fejezet; Fónagy Iváné az 1., a 4. és az 5. fejezet és a 2. fejezet 2.1., 2.2. pontja Szende Tamásé a 2. fejezet többi pontja;

[1] Tarnóczy Tamás, A magyar magánhangzók eloszlása: Magyar Nyelv 38. 352 (1942); Megjegyzések és kiegészítések a magyar magánhangzók eloszlásának kérdéséhez: Magyar Nyelv 39. 369 (1943). — Vértes Edit: Nyelvtudományi Közlemények 54. 124—40 (1953), 55. 142—80 (1954).

versen) alapulnak, 6 846, illetőleg 5 295 szón; 34 565, illetőleg 25 429 összesen 59 994) fonémán.

2. A szövegek kiválasztása és előkészítése

Fonéma-statisztikát akarunk készíteni, és nem beszédhang statisztikát.

2.1. A jelentősebb szabad hangváltozatokat — így a *t* retroflex variánsát, az illabiális rövid /a/-t, a rövid /e/-t — csak a hangfelvételek alapján tarthatnánk számon. Szöveg alapján is tekintetbe vehetnénk azonban a mássalhangzó-hasonulást, ahogyan ezt Tarnóczy Tamás (1952) tette, és olyan szöveget helyezhetnénk a gépbe, melyen a *vasgolyó*-t *zs*-vel írtuk át.

Ez az eljárás több szempontból is célszerűtlennek látszott. Az átírás igen nagy vesződséggel járna. De ez lenne még a kisebbik baj. Akadnak komolyabb nehézségek is.

Az asszimiláció vagy kötelező érvényű, vagy fakultatív. Ha nem kötelező érvényű, akkor fakultatív kombinatorikus variánsok közül kell választanunk, és önkényesen, vagy bizonyos valószínűségek alapján kell meghatároznunk, hogy az író „lelki fülében" a *fenség* szó *n*-nel csengett-e, vagy nazalizált magánhangzóval, hogy az *egyszer*-t gondolatban /ts/-vel vagy /c/-vel ejtette-e stb. Ha pedig kötelező az asszimiláció, mint például az *alkudjunk* esetében, akkor meg felesleges *ggy*-nek átírni, hiszen a gép a feltett kérdésre úgyis közli velünk, hogy hány *dj* fordul elő intervokális helyzetben. Sőt: ezt csak akkor tudja közölni, hogyha nem írjuk át „fonetikusan" a szöveget. Ha az *alkudjunk* és a *meggyet* szó /j:/-jét egyaránt *ggy*-nal jelöljük, utólag bajosan lehetne megállapítani, milyen gyakori volt a szövegben a *d* + *j* összeolvadás.

Gondoljunk arra is, hogy bizonyos esetekben vitatott a hasonult hangok fonológiai hovatartozása. A *vasgolyó* szó [ʒ] hangja egyesek szerint a /ʒ/ fonémához, mások szerint a /ʃ/ fonémához tartozik, a prágai iskola nézete szerint a kettő közös archifonémájával azonos. Bármelyik megoldást választjuk, a két másik felfogás hívei számára lehetetlenné tettük, hogy saját elképzeléseik szerint határozhassák meg a kombinatorikus variánsok eloszlását.

A hasonulást tehát éppen azért nem vettük számba az átírásnál, hogy a kombinatórikus variánsok viszonylagos gyakorisága szükség esetén könnyen megállapítható legyen.

Felmerülhet egy elvi ellenvetés. Akik szerint a *vasgolyó* szóban /ʒ/ fonémával van dolgunk, szemünkre vethetnék, hogy ez a statisztika még csak nem is fonemikus, hiszen mi ezt a /ʒ/ fonémát /ʃ/-ként tartjuk nyilván. Ebből a szemszögből nézve statisztikánk valóban nem fonéma-statisztika, hanem morfonéma-statisztika. Olyan morfonéma-statisztika, mely bizonyos hasonulások számbavételével bármikor átalakítható ebben a szűkebb értelemben vett fonéma-statisztikává. Ha ezt az átalakítást végrehajtanánk, módunkban lenne a magyar morfonéma-eloszlást a (szűkebb értelemben vett) fonéma-eloszlással egybevetni. (Ez is egyike marad

azonban annak az igen nagy számú kérdésnek, melyre a mai beszámolónkban nem térhetünk ki.)

Úgy hisszük, azáltal, hogy a számunkra kényelmesebb megoldást választjuk, ezúttal nem okozunk másoknak munkatöbbletet, sőt talán meg is takarítunk némi kényelmetlenséget és bosszúságot a statisztika későbbi használóinak.

Más a helyzet a rövid, zárt /ë/ esetében. Ezt a hangot egyszerűen elsikkasztottuk, követve ebben az irodalmi nyelvet, melyben ugyancsak elsikkadt.

2.2. Ami pedig a hosszú mássalhangzókat illeti: úgy gondoljuk, hogy az *attól* /tt/-je, a *dinnye* /ɲɲ/ hangja bátran felfogható kettős *t*-nek, kettős ɲ-nek, ellentétben a hosszú magánhangzókkal, melyeknek ilyen értelmezése sokkal vitathatóbb. A hosszú mássalhangzókat — az APhI-t követve — kettős mássalhangzóknak írtuk át és nem jelöltük külön betűjellel őket. Ennek ellenére semmi akadálya annak, hogy szükség esetén a kettős (hosszú) mássalhangzók gyakoriságát külön is megállapíthassuk. Erre, a kettős mássalhangzók esztétikai szerepére való tekintettel is szükség lehet olykor, így például költői szövegek elemzésénél.[2]

2.3. Ezeknek megfelelően megtartottuk a fonémák magyar betűjeleit. Ez azzal járt, hogy az *s* + *z* hangkapcsolat egybeesett az */sz/* fonémával, a *z* + *s* a *zs*-vel stb.; kettőzésnél pedig a hosszú *sz* az *s* + *sz*-szel. A géppel megkülönböztettük ezeket, mégpedig úgy, hogy a hangkapcsolatok tagjai közé dőlt vonást (/) ütöttünk: *vas/zár, réz/sín* stb. Elválasztó jellel vontunk határt összetett szavakban, ahol a szótagolásnál hiba történhetett volna *(még/is)*, olyan hangkapcsolatokban, ahol egy magánhangzó és hosszú párja fordult elő egymás mellett (például: *leél* > *leeel* > *le/eel)*, valamint ahol még az értelmezés ezt megkívánta (ha például az *ő*-t jelölő *oe* után még *e* vagy *é* következett).

2.4. Vizsgálataink kiterjedtek a hangcsoportok (digrammok és trigrammok) eloszlására is. Egyelőre nem volt mód valamennyi lehetséges két- vagy háromtagú fonémacsoport számontartására, s ezért — a hangok képzésének helye és módja alapján megkülönböztetett — fonématípusokból indultunk ki, közös nevezőre hozva például a nazális mássalhangzókat, az affrikátákat, a zárhangokat, a réshangokat, a labiális és illabiális magánhangzókat stb.

2.5. A gép megfelelő programozással szótagolni is tud, s megszámlálja, melyik szótagtípusra hány eset jut, meghatározza, hogyan viszonylik az egyes szótagtípusok gyakorisága a szavak hosszához (csökken-e például az egy szótagra eső hangok száma az egy szóra eső szótagok számának növekedésével), más szótagtípusok lépnek-e előtérbe a szavak első, második, *n*-edik szótagjában.

[2] Vértes Edit vizsgálatai szerint a hosszú mássalhangzók különösen gyakoriak a hangutánzó szavakban.

2.6. A szavak hosszát a szótagok és hangok számával fejeztük ki. Meghatároztuk azt is, hogy milyen gyakran fordulnak elő szövegeinkben az egyes szófajok. Minden szó fölé a szófaját jelző kódot írtuk, illetőleg a szó mellé lyukasztottuk a szalagon. Tizenhat lehetőségünk volt, s ezeket így osztottuk el: ige, főnév, főnévi igenév, melléknév, melléknévi igenév, számnév, névmás, határozószó, határozói igenév, módosító szó, névelő, névutó, mellérendelő kötőszó, alárendelő kötőszó, indulatszó és az igétől elváló igekötő (ez utóbbit az jelölte, hogy nem volt kódja).

A szófaj meghatározásában a szóról leválasztott rag- és jelelem előtt megmaradt morfémából indultunk ki. A névmás-határozószókat *(tőlem, hozzád)* egyértelműen névmásnak vettük, a *benne-* és a *rajtam*-félék kivételével, melyeket a határozószók közé soroltunk. A két igéből álló összetett igealakokban a segédigét is külön igének vettük. A kötőjeles összetett szavakban, ha az összetétel egyszeri, távoleső fogalmakat lazán összekötő volt, mindkét tag külön szófaj-jelet kapott.

Nem kellett azonban minden szót szófaj-jellel ellátni. A leggyakoribb szavakat kiválaszthattuk, ezeket a gép memória-rekeszében tárolta, s aztán automatikusan osztotta a megfelelő helyre. Munkatársainknak megadtuk a ,,szófajozási'' szempontokat, felsoroltuk a vitás helyeket, és utána a leírás előtt az egész szöveganyagot ellenőriztük.[3]

A gép feladata volt ezután az előfordulások megszámlálása.

2.7. A gép a munka első fázisában az említett kérdésekre ad választ. A későbbiekben — másfajta programozással — meghatározzuk a mondatfajták gyakoriságát. Megkülönböztettük e célból a mondatrészeket és a tagmondatokat elválasztó vesszőt, mégpedig úgy, hogy az utóbbit kettőztük, a három pontot (...) és pontosvesszőt mellőztük, s az értelemnek megfelelően két vesszőt tettünk a helyébe, vagy zártuk a mondatot. Ezenkívül más és más jellel láttuk el a felszólító, felkiáltó és óhajtó mondatokat, megtartva persze a kérdőjelet és a pontot is.[4] Így meghatározható a mondatfajták viszonylagos gyakoriságán kívül az is, hogy hány kérdőmondat végződik *-e* kérdőszócskával, hány felszólító, felkiáltó, óhajtó mondat kezdődik indulatszóval.

Elhatároltuk egymástól a tagmondatokat. Lehetőségünk van tehát arra, hogy az egyszerű és összetett mondatok arányán kívül megállapítsuk, milyen gyakoriak a két, három és négy tagból álló összetett mondatok, hogyan viszonylik a szavak hossza a mondatok hosszához stb.

2.8. Az is megállapítható, hogy milyen gyakori az e n j a m b e m e n t a két költő versei'ben.

2.9. Gépi úton egyelőre nem volt megoldható a szógyakoriság vizsgálata. Ez jelenleg manuálisan történik, ciklusonként is elkülönített anyagon.

[3] A szempontok meghatározásában, a vitás kérdések eldöntésében Károly Sándor, H. Molnár Ilona, Szende Aladár és Tompa József voltak segítségünkre. Kétes esetekben ,,A mai magyar nyelv rendszere — Leíró nyelvtan'' (Bp., 1961 és 1962) fellapozása után döntöttünk.

[4] A munkának ebben a részében, valamint a szövegnek szalagra vitelében Ferenczy Mária egyetemi hallgató segített sokat.

Az utolsó munkaszakasz a számok formájában kiírt eredmények értelmezése, vagyis a kódjelek visszaalakítása.

Minthogy a gép csak az egyes elemek és jelenségek abszolút gyakoriságát adta meg, valamint azokat, amelyek kontrollul szolgáltak, számológépen százalékoltuk az eredményeket és kiszámítottuk a hiányzó értékeket (például a fonéma előfordulásánál meg van adva az önálló, szó eleji, szó végi és az összes előfordulás, amelyekből a szóközépi már egyszerűen meghatározható). A jövőben a kiíró program módosításával ezeket az utólagos számításokat is ki fogjuk küszöbölni.[5]

3. A technikai megvalósítás

Az elektronikus számológépeket kezdetben kizárólag számítási jellegű feladatok elvégzésére tervezték és használták; ezt a körülményt általában a gépek felépítése is tükrözi. Elég hamar rájöttek azonban arra, hogy ugyanez a technikai apparátus alkalmas egyéb információ-feldolgozási feladatok megoldására is; és napjainkban már elég sok ilyen típusú probléma kerül számológépen megoldásra. A számológépek „számoló" és a feladatok „nem-numerikus" jellegének kétségtelenül fennálló ellentmondása az ilyen feladatoknak a számológépre való alkalmazását a programozók számára feltétlenül érdekes, a szokásos eszközökkel általában nem megoldható feladattá teszi.

Az alábbiakban elsősorban ebből a szempontból fogjuk ismertetni a hangstatisztikai vizsgálatokra vonatkozó feladatnak az M—3 gépen való megoldását, kiemelve azokat a részeket, ahol az alkalmazandó *algoritmusok* felállításának problémája esetleg nemcsak a programozással foglalkozók számára jelenthet érdekességet.[6]

3.1. A feladat megoldása az elektronikus számológépen két lépésben történik:

A) a szövegek feldolgozása, adatgyűjtés;
B) az eredmények táblázatos formában való kiírása.

Ezt az elválasztást az a gyakorlati szempont teszi indokolttá, hogy egy-egy összetartozó szöveg feldolgozása — időhiány és a gép egyéb feladatai miatt — nem egyfolytában történik, és a végleges táblázatos formában való kiírásra csak a feldolgozás végén van szükség.

A továbbiakban lényegében csak az *A)* programmal fogunk foglalkozni, mert a *B)* program felépítését az eredményül kapott táblázatok szerkezete lényegében meghatározza.

[5] Azóta a program jelentékenyen kibővült és módosult. A változások egyúttal azt is jelentik, hogy a számolások minden mozzanatát gép végzi.

[6] Azok számára, akik a számológépek, illetőleg programozásuk iránt érdeklődnek, a következő műveket ajánlhatjuk első tájékozódásra: D. V. McCracken: Digital computer programming (New York, 1957); W. Kämmerer: Ziffernrechenautomaten (Berlin, 1960); A. I. Kitov—N. A. Krinyickij: Электронные вычислительные машины (Moszkva, 1958).

Az *A)* program szintén két részre bomlik;

A1. egy-egy szó bevitele a gép memóriájába, e közben megtörténik:

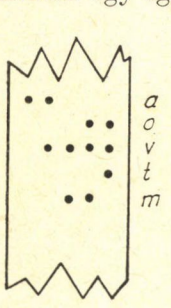
a) a több betűvel jelölt fonémák értelmezése;
b) a szótaghatárok kijelölése;
c) a szóhosszúság számlálása;
d) a szófajokra vonatkozó jelek értelmezése.

A2. a bevitt szó feldolgozása, a gyűjtendő adatok számlálása:

a) minden fonéma után;
b) minden szótag végén;
c) az egész szó végén.

Az *A1.* és *A2.* programrészek szavanként váltakozva kerülnek végrehajtásra: miután az *A1.* program bevitt egy szót és megfelelően átalakítva elhelyezte a gép memóriájába, a gép rátér az *A2.* program végrehajtására, amely a tulajdonképpeni adatgyűjtési munkát végzi el az adott szó alapján. Ezután újra az *A1.* program következik egy új szó bevitelére, stb.

3.2. A vizsgálandó szövegnek a gépbe való bevitele *lyukszalagról* történik. Ennek egy-egy sorában öt lyuk számára van hely (1. ábra), amelyeknek minden kombinációja valamilyen a távírógép billentyűzetén szereplő betűnek, számnak vagy írásjelnek (közös néven *karakternek*) felel meg. Ily módon $2^5 = 32$ karakter megkülönböztetésére volna lehetőség, ez a szám azonban — a közönséges írógépeken használatos „nagybetű-kisbetű"-váltóhoz hasonló technikával — közel kétszeresére növelhető. Így minden karakternek a váltó valamilyen állása, valamint egy 0 és 31 közötti kódszám felel meg. A szöveg bevitelének programja oly módon működik, hogy a szalag minden egyes sorának, azaz minden egyes karakternek a bevitele után a gép rátér egy — az illető karakterhez tartozó — részprogram végrehajtására. (Természetesen több karakterhez is tartozhatik ugyanazon részprogram.)



1. ábra

Így például az *e* betűnek megfelelő részprogramnak meg kell vizsgálnia, hogy

a) az előző karakter az *e* betű volt-e?
b) „ „ „ „ *o* „ „
c) „ „ „ „ *u* „ „

mert ezekben az esetekben tulajdonképpen az *é*, *ö*, illetőleg *ü* fonémák ábrázolásáról volt szó.

A bevitt szó egyes betűit, illetőleg az átalakítás során nyert fonémákat a program a gép memóriájának egymás utáni rekeszeibe egy úgynevezett *szó-listára* helyezi el, és egy külön számláló-rekeszben jelzi, hogy hányadik

rekeszbe kell a következő betűt elhelyezni. Ha olyan helyzet áll elő, hogy a bevitt betű az utolsó elhelyezettel együtt valamilyen több betűvel jelölt fonémának megfelelő kombinációt alkot, akkor a szó-lista utolsó elemét ennek megfelelően módosítjuk, és a számláló-rekesz tartalmát nem növeljük, hiszen nem történt új fonéma elhelyezése.

A számláló-rekesz tartalma egyébként a szó bevitele után mindig eggyel nagyobb a szóban szereplő hangok számánál, így ez látja el az *A1.* program feladatai között *c)* alatt jelzett, a szóhosszúság számlálására vonatkozó funkciót is.

Így például a *tél* szó esetében, amely *teel* formában van a szalagra lyukasztva, az algoritmus működése a következő:

Betű a szalagról		t	e	e	l
Szó-lista		t	t e	t é	t é l
Számláló	1	2	3	3	4

2. ábra

Tekintettel arra, hogy a több betűvel jelzett fonémák egy része ugyanazon betű ismétlődésével van jelölve (például *ee, ii, oeoe)*, annak megvizsgálása, hogy ilyen eset áll-e fenn, egyszerűen a szó-lista két utolsó elemének összehasonlításával történik. (A kódszámok oly módon vannak megválasztva, hogy ilyen esetben a módosítás mindig ugyanazon szám hozzáadását jelentse, így például ha az *e* hang kódszáma 20, akkor az *é* hangé 20 + 40 = 60; ha az *i* kódszáma 14, akkor az *í* hangé 14 + 40 = 54; stb.)

Tekintsünk ennek alapján egy bonyolultabb példát, a *lőcs* szót, amely a szalagon *loeoecs* formában szerepel:

Betű a szalagról		l	o	e	o	e		c	s
Szó-lista		l	l o	l ö	l ö o	l ö ő	l ő	l ő c	l ő cs
Számláló	1	2	3	3	4	4	3	4	4

3. ábra

Látjuk azt, hogy a négy betűvel jelölt *oeoe* fonéma „megfejtéséhez" már szükség volt egy helyen a számláló tartalmának csökkentésére is, de végeredményben a helyes eredményt kaptuk.

Megjegyezzük, hogy minden olyan esetben, amikor a jelölés félreértésre adhatna okot, a megfelelő helyre lyukasztott dőlt vonás (/) jel választja

el egymástól az össze nem tartozó betűket. Így például a *leél* szót *leeel*-nek lyukasztva a gép a szó-listára az *léel* hangokat helyezné, azért ezt *le/eel* formában kell lyukasztani.

3.3. Az *A1.* programhoz tartozik még a szótaghatárok kijelölése is. Ez a következő szabályok alapján történik:

A szó első magánhangzóján kívül minden magánhangzó egy szótaghatárt jelöl ki, éspedig:

ha az előző fonéma is magánhangzó, akkor a határ e mögé a fonéma mögé kerül;

ha az előző fonéma mássalhangzó, akkor a határ ez elé a fonéma elé kerül;

a szó vége mindig egyben szótaghatár is;

a dőlt vonás (/) elválasztójel minden szempontból a „szóvég" szerepét játssza, tehát mindig szótaghatárt jelent, és az utána következő első magánhangzó a szabály szempontjából a szó „első" magánhangzójának tekintendő.

Példák: *előállít*↑

rész/program↑

A dőlt vonás (/), elválasztójel fentiek szerint kettős szerepet játszik: elválasztja egyrészt a félreérthető betű-összetételeket, másrészt pedig a szótagolás szempontjából „rendhagyó" összetett szavakat. Ez a kettősség elvileg néha a szótagolási algoritmus hibás működését eredményezheti; ha ugyanis valahol nem szóösszetétel következtében vagyunk kénytelenek a kettős betűk elválasztására elválasztójelet elhelyezni, akkor a fenti 3) szabály alapján a gép azt is szótaghatárnak, és az 4) szabály szempontjából szóvégnek tekinti. Ha m a g á n h a n g z ó k közé tesszük az elválasztójelet, nincs baj, mert akkor itt amúgyis szótaghatár van. A hibás működés akkor következik be, ha m á s s a l h a n g z ó k sorozata érthető félre a kettős betűk értelmezése szempontjából, és ezért kell közéjük elválasztójelet helyezni (természetesen csak ha n e m szó-összetételről van szó). Ha az elválasztott mássalhangzók után még egy mássalhangzó következik, például *c s z r* akkor, mivel meg kell jelölnünk, hogy *cs-z-r* vagy *c-sz-r* olvasandó-e, a szótaghatár mindenképpen — hibásan — az első mássalhangzó mögé fog kerülni.

Az eddigiekben azonban ilyen betű-összetételt értelmes szóban még nem sikerült találni, és amennyiben léteznének is ilyen tulajdonságú szavak, nem valószínű, hogy gyakoriságuk a s t a t i s z t i k a i eredményeket lényegében befolyásolná.

A szófajokra vonatkozó szám-kódokat az *A1.* program egyszerű átalakítással 1 és 15 közötti szám formájában tárolja. Amennyiben valamely szó után szófaj-jelzés nincs lyukasztva (ilyenkor ez a szám 0 lesz), már az *A2.* program vizsgálja meg, hogy a szó szerepel-e a leggyakoribb szavak előre megadott listáján.

3.4. Az *A2.* program lényegében három részből tevődik össze.

a) Minden fonéma után elvégzendő feladatok. Ide tartozik elsősorban az illető fonéma előfordulásait számláló rekesz tartalmának eggyel való növelése, figyelembe véve, hogy a szó elején, a szó végén vagy a szó belsejében való előfordulásról van-e szó. Gondoskodni kell a trigrammok

vizsgálatáról is, ez oly módon történik, hogy minden egyes fonémának megfelelő memória-rekeszben tárolva van az illető fonéma vizsgálandó fonetikai tulajdonságának kódszáma (0 és 7 közötti szám). Ezekből egy háromjegyű szám áll össze, amely a megfelelő trigramm kódszáma, és az ennek megfelelő memóriarekesz tartalmát növeljük eggyel. Természetesen a szavak első vagy második betűje után erre a műveletre nem kerül sor.

Minden fonémát figyelembe kell venni a szótagtípus kialakításának szempontjából is. Ez — tekintettel arra, hogy minden szótagban egy és csakis egy magánhangzó van — oly módon történik, hogy külön számoljuk a magánhangzó előtti és külön a magánhangzó utáni mássalhangzókat, és az így kapott számokból egy kétjegyű számot állítunk össze. Így például a 12 kódszám egy C-V-C-C (vagyis: mássalhangzó-magánhangzó-mássalhangzó-mássalhangzó) típusú szótagnak felel meg.

b) Minden szótag végén az előbbiek szerint kialakított kódszámnak megfelelő rekesz tartalmát növeljük eggyel; ismét külön vizsgálva a szó első, utolsó és belső szótagjait.

c) A szó végén elvégzendő feladatok közé tartozik a már előzőleg kialakított fonémák, illetőleg szótagok szerinti szóhosszúsághoz, illetőleg a szófaj-jelhez tartozó memóriarekeszek tartalmának növelése (meg nem jelölt szófaj esetén a gyakori szavak „szótárából" való kikeresés segítségével), valamint a szó utolsó fonémájára, trigrammjára, illetőleg szótagjára vonatkozó teendők elvégzése.

3.5. A gépi megvalósításról elmondottakat összefoglalva láthatjuk, hogy a kívánt vizsgálatok elvégzése teljes mértékben gépesíthető feladat, és az elektronikus számológép ilyen feladatoknál jól alkalmazható.

A tapasztalatok szerint a feldolgozás sebessége kb. 0,6 fonéma/sec, ami első látásra nem látszik nagy sebességnek; figyelembe kell azonban venni a vizsgálat szempontjának sokféleségét, valamint azt a körülményt, hogy az M—3 gép jelenleg az egészen kis teljesítményű elektronikus számológépek közé tartozik: műveleti sebessége 25—30 alapművelet/sec.

A gép 1962 első felére tervbe vett bővítése után a fenti sebesség kb. 30—40-szeresére fog növekedni és lehetővé fog válni több, az eddigi feldolgozásból a memória-kapacitás hiánya miatt kimaradt vizsgálati szempont figyelembevétele.

A sebességi adatokkal kapcsolatban érdemes megemlíteni, hogy a gépi program kipróbálásához szükségessé vált 1000 hangból álló szöveg kézi feldolgozása néhány napi munkát vett igénybe.

4. Az eredmények értékelése

4.1. Az egyes hangok viszonylagos gyakoriságának eltérése az esetek többségében 1—2 tizednyi vagy ennél is kevesebb, tehát nem szignifikáns (a 0,63—0,53%-os szóráson belül esik, mely a megvizsgált hangok számából adódik).[7]

[7] A szórást a hagyományos $100 \frac{1}{\sqrt{N}}$ százalékból kiindulva határoztuk meg.

Nem adódnak nagyobb különbségek általában akkor sem, ha a hangokat csoportosítjuk és a magas és mély nyelvállású magánhangzókat hasonlítjuk össze, vagy a mássalhangzókat vetjük egybe képzésük helye és módja szerint.

Csak néhány esetben szignifikáns az eltérés:

4.12. Szabó Lőrinc vizsgált ciklusában viszonylag több a zöngétlen mássalhangzó[8]

	Sz. L.	J. Gy.
zöngétlen	36,58	34,21

4.13. Szabó Lőrinc versciklusában viszonylag nagyobb helyet foglalnak el a par excellence „kemény" másalhangzók /*t, k, r*/, mint Juhász Gyula verseiben, ahol a „lágy" mássalhangzók /*m, n, l*/ lépnek némileg előtérbe:

	Sz. L.	J. Gy.
kemény	29,16	27,63
lágy	31,57	32,84

4.14. Juhász Gyula verseiben viszonylag gyakoribbak a hosszú magánhangzók, Szabó Lőrinc versciklusához képest, melyben a rövidek erősebben dominálnak:

	Sz. L.	J. Gy.
rövid	76,98	74,54
hosszú	23,02	25,46

4.15. Szabó Lőrincnél több a labiális magánhangzó:

Sz. L.	J. Gy.
47,29	45,78

4.16. Több az úgynevezett „mély" vagy sötét színű magánhangzó is:

Sz. L.	J. Gy.
46,30	45,00

4.2 Hogyan értékeljük ezeket az eltéréseket? Ha nem kellene most beszámolót tartanunk, azt mondhatnánk: ne értékeljük egyelőre, várjuk meg a többi ciklus feldolgozásának eredményét.

De tegyük fel, hogy a további vizsgálatok nem módosítják minőségileg, csak megszilárdítják a kisebb anyag alapján elért eredményeinket. Az értel-

[8] Itt és a továbbiakban a megadott számok külön a mássalhangzókon belül, külön a magánhangzókon belül számított százalékok. Ugyanis bennünket az egyes mássalhangzó-típusok, illetőleg az egyes magánhangzó-típusok viszonylagos gyakorisága érdekel egy másik mássalhangzó-, illetőleg magánhangzó-típushoz képest.

mezés akkor sem könnyebb feladat. A különbség legfeljebb abban áll, hogy bizonytalan konstrukcióinkat biztosabb alapokra építjük.

Amikor áttérünk a fénémaeloszlás leírásáról a szignifikáns eltérések értelmezésére, tökéletesen megváltoznak munkánk feltételei. A határon innen a kijelentések egyértelműen igazak vagy tévesek. Bárki utánaszámolhat annak, vajon valóban 29,158 százalékot tesz-e ki együttesen Szabó Lőrinc verseiben a /k/, /t/ és /r/. A határon túl, az értelmezések birodalmában azonban a kijelentéseink legfeljebb többé vagy kevésbé valószínűek lehetnek.

A strukturalista nyelvészek többsége nem is vállalkozik éppen ezért ilyen „külföldi" utakra. És ez kétségtelenül indokolt. De indokolható talán a határátlépés is. Elsősorban arra a történelmi tényre hivatkozhatnánk, hogy a nyelvtudomány mai területe sorozatos határátlépések nyomán alakult ki, és nincsen olyan tartománya, melyben nem jártunk volna eleinte ingatag talajon. A mai hangtani vagy akusztikai terminológiánkban burjánzó metaforák (világos, sötét, kemény, lágy hangok stb.) arra emlékeztetnek, hogy egy időben a hangtan is definiálatlan, bizonytalan körvonalú fogalmakon alapult. Az egyes területek meghódítását „tudományelőttes" kísérletezések előzték meg. Azt hiszem, előbb lesz a stilisztikából tudomány, ha a tévedések bőséges kockázatát vállalva, csak valószínűsíthető kijelentések segítségével kíséreljük megvetni a stílus jelentéstanának alapjait, így a hangstilisztikáét is.

Természetesen tudnunk és tudatnunk kell, hogy amit egyes szerzők hangos stílusának jellemző jegyeiről mondhatunk, ideiglenes jelleggel bír. (Egy kitűnő osztrák regényíró, Heimito von Doderer módszerét követve, kívánatos lenne más színű tintával írni azt, amit magunk sem tekintünk feltevésnél többnek. Szóban történő beszámolónkban azonban legfeljebb nyomatékosan felhívhatjuk a figyelmet arra, hogy a számszerű eredményeken alapuló következtetéseink semmiképpen sem matematikai pontosságúak.)

4.21. A költők jellegzetes hangspektrumát kétféleképpen is értelmezhetjük: közvetve és közvetlenül. Az előbbi esetben az egyes hangok viszonylagos gyakoriságát bizonyos morfémák gyakoribb használatára vezetjük vissza.[9]

Ezek után már csak azt kell megértenünk, értelmeznünk, miért gyakoribb a költőnél, az írónál a hangspektrumot módosító egyik vagy másik morféma.

4.22. Ezt a viszonylag könnyen és biztonságosan járható utat nem követhetjük ezúttal, mivel a morfémák eloszlásának feldolgozása még folyamatban van.

[9] Így járt el például Vértes Edit (az 1. jegyzetben hivatkozott dolgozatában), amikor Ady Endre nyelvében az /m/ gyakoriságát az egyes szám első személyét jelző *m* morfémára hivatkozva — az egyes szám első személynek, az „én"-nek előtérbe állításával magyarázza; azután az /i/ gyakoriságát az *ifjú* és a *fiatal* szó gyakoriságával; az *-uz*-digramm gyakoriságát pedig — Veres Péter nyelvében — a *búza* szó gyakoriságával magyarázza.

„Marad a másik út: közvetlenül a hangokból, a hangok képzési és akusztikai sajátságaiból kiindulva kell magyaráznunk az eltéréseket, mintha nem is volnának morfémák, mintha a hangok közvetlenül szerepelnének a beszédben.

Ezt a paradoxális eljárást egy régebbi tapasztalatunk indokolhatja némiképpen. A /k/, /t/, /r/ fonéma például előtérbe kerül Szabó Lőrinc „Mérget, revolvert!" című költeményében; az /l/, az /m/, az /n/ dominál az elégikus alaphangulatú „Valami örök"-ben. Ugyancsak a „kemény" mássalhangzók uralkodnak Petőfi agresszív jellegű verseiben, szemben a szelíd hangúakkal, ahol az /m/ és /l/ szerepelt viszonylag gyakrabban. Ez az összefüggés nincsen a magyar szókincshez kötve. Hasonlóképpen módosul a megfelelő érzelmek hatására a hangspektrum Rückert, V. Hugo és Verlaine verseiben is.[10]

Ha a hangspektrum — a morfémák áthidalásával —, összefüggésbe hozható bizonyos érzelmekkel, bizonyos hangulatokkal, feltehető, hogy egyes költők és írók sajátos fonéma-eloszlásának is van önálló „mondanivalója", hogy összefüggésben áll az illető költő sajátos lelki alkatával, hogy a lelki alkatában domináló érzelmek, attitűdök kifejezője. Ennek a hipotézisnek az alapján a „kemény" mássalhangzók gyakorisága bizonyos lelki tartásbeli keménység tükröződéseként fogható fel. Ösztönösen is így értelmezzük egyesek keményebb, feszesebb ejtésmódját. A hangspektrum eltolódása a feszesebb képzésű zöngétlen hangok irányában (ez Szabó Lőrinc hangspektrumának egy másik szignifikáns sajátossága) a feszesebb ejtés írott vetületének tekinthető.

4.23. Ha a beszédtempót akarnánk a versek írott nyelvére átkódolni, a hosszú magánhangzók szaporításával kellene fékeznünk, a rövidek gyakorításával gyorsítanunk az írott nyelv virtuális tempóját. Ennek az elvnek megfelelően Szabó Lőrinc költői nyelvének — a rövid magánhangzók gyakorisága következtében — releváns ritmikai sajátságát, a gyors tempót, hasonlóképpen értékelhetnénk, mint a gyorsan pergő beszédet. A hosszú magánhangzónak dominanciáját Juhász Gyula verseiben pedig a lassú beszédtempó írott vetületének tekinthetnénk. (Más — bár nem érdektelen — kérdés, hogy a két költő mindennapi beszédében is érvényesült-e ez a tendencia.)

4.24. A „mély" vokálisok viszonylagos gyakorisága Szabó Lőrinc verseiben a vizsgált ciklus (A Sátán műremekei) komor alaphangjával függhet össze.

4.25. A labiális magánhangzók előtérbe kerülését ezúttal magyarázatlanul hagynánk. Azok számára, akik járatosak a mélylélektanban, utalnék arra, hogy e verscikluson vörös fonálként húzódik végig az éhség (pénzéhség, az életéhség, szereleméhség) motívuma.

[10] Mind a magyar, mind a német és francia statisztika húsz—harminc ezer hangot tartalmazó verses anyagon alapult; vö. Fónagy Iván: A költői nyelv hangtanából (Bp., 1959) 37—50.

4.3. A szótagtípusok eloszlása részben megfelel a várakozásoknak.

	J. Gy.	Sz. L.	Ady[11]	V. P.[11]
V	9,11	*10,60*	8,3	8,6
VC	11,18	10,41	9,5	11,1
VCC	0,64	0,85	0,3	0,7
VCCC	0,01	0,20	—	—
CV	33,51	*35,77*	32,5	34,2
CVC	38,57	35,20	44,4	41,2
CVCC	6,68	5,19	4,75	4,0
CVCCC	0,08	0,36	—	—
CCV	0,09	*0,59*	0,2	0,1
CCVC	0,05	0,28	0,1	0,1
CCVCC	0,01	*0,48*	—	—

A gyakorisági sorrend Juhász Gyula verseiben ugyanaz mint Ady verseiben és Veres Péter prózájában. Szabó Lőrinc azonban eltér némiképpen ettől a „normától". Verseiben a másik három szerzőhöz képest gyakoriak a nyílt szótagok (V és CV), s ugyanakkor viszonylag gyakoriak bizonyos ritkább, mássalhangzó-torlódásos szótagtípusok is (VCC, CCV). A hangkarakterológia mai fejletlenségi fokán — a szóstatisztika eredményeinek ismerete nélkül — legfeljebb a mássalhangzó-torlódás és a kemény konszonantizmus kapcsolatára utalhatnánk. A legfeltűnőbb formai sajátság: a CV (és V) típus dominanciája, egyelőre megmagyarázatlan marad.

4.4. A szavak szótageloszlását tekintve a következő eredményt kaptuk:

	1.	2.	3.	4.	5.	6.	7.
Sz. L.	38,77	27,33	20,72	*9,70*	*2,69*	0,60	0,18
J. Gy.	39,57	31,70	21,12	6,70	0,80	0,08	0,04
Ady[11]	35	40	18	6	0,8	0,13	
Arany[11]	38	41	15	6	0,1	0,05	
Petőfi[11]	40	31	21	7	1,0	0,3	

Juhász Gyula verseiben tehát a kétszótagú szavak viszonylag gyakoribbak, mint Szabó Lőrinc verseiben, melyekben meglehetősen gyakoriak a négyszótagú szavak és nagyobb szerepet játszanak az öt- és többszótagúak is, mint Juhásznál, Petőfinél, Aranynál, Adynál. Tarnóczy Tamás vizsgálatai nyomán ismeretes, hogy a három- és többszótagú szavak gyakorisága a tudományos próza jellegzetessége. Alighanem Szabó Lőrinc erősen fogalmi jellegű költészetével függ össze ez a „prózai" szótageloszlás. Szerepe lehet ebben Szabó Lőrinc tömörítő tendenciájának, gondoljunk olyan összetett szavakra, mint *ágyúkoronázta* (242), *gyémántvértezetben* (242), *istengyilkoló* (220).[12]

[11] Vértes Editnek az 1. jegyzetben hivatkozott tanulmánya alapján idézzük.
[12] A számok Szabó Lőrinc „Összegyűjtött Versei"-nek (Bp., 1960) lapszámai.

4.5. Megvizsgáltuk egyúttal a szófajok megterhelését is:

	Sz. L.	J. Gy.
főnév	26,621	*28,261*
ige	13,274	*15,480*
névmás	7,783	6,511
módosító szó	3,373	2,161
főnévi igenév	1,300	0,628
számnév	1,168	0,798
igekötő	1,110	0,741
(indulatszó	0,701	0,477)

A többi szófajnál nem tapasztaltunk szignifikáns eltérést. A mondatok alapelemét képező főnév és ige Juhász Gyula többnyire elemi eszközökkel operáló és ható verseiben gyakoribb. A főnévi igenév, a névmás, számnév, módosító szó, igekötő, indulatszó Szabó Lőrincnél. A morféma-statisztika lezárása előtt bajos lenne az eloszlást értelmezni. (Nem tudjuk egyebek között például, hogy melyik névmás gyakori Szabó Lőrinc verseiben: a birtokos, a személyes, a mutató névmás-e, és ha a két előbbi, akkor melyik személyben. Hasonlóképpen tudnunk kell, hogy a perfektiváló *meg-* igekötő lép-e előtérbe, vagy mások is, és melyek azok.) A számnév és a főnévi igenév gyakorisága a poeta doctus stílussajátsága is lehetne, az indulatszóé talán a tudós költővel harcban álló szenvedélyes emberé.

5. A téma további problémái, lehetőségei

Előrebocsátottuk, hogy indulóban levő munkáról számolunk be. Helytelen lenne éppen ezért a statisztikai vizsgálatok általános jellegét, célszerűségét vagy célszerűtlenségét a mi igen kezdeti eredményeink alapján megítélni. A fenti adatokat inkább azért említettük meg, hogy a vitához fogódzót adjunk.

Ugyancsak ezért ismertetjük röviden terveinket, célkitűzéseinket.

Esztétikai és lélektani jellegű stíluselemzéshez s mindenekelőtt a mai magyar köznyelv leírásához szeretnénk anyagot szolgáltatni.

5.1. A költő a vers formájával valami olyasmit is elmond, ami a vers tartalmából hiányzik. A forma — mely mennyiségi mutatókkal jellemezhető — a tartalomnak ezúttal nem pusztán hordozója, hanem maga is tartalom, közlemény, a főszólamot kísérő második szólam. A kemény konszonanciának, a pergő ritmusnak (azaz: a rövid magánhangzók gyakoriságának), az áthajlásoknak, a hosszú szavak, számnevek, egyes mondatszerkezetek vagy meghatározott morfémák előtérbe kerülésének sajátos mondanivalója van.

Ehhez, a formában, tehát látszólag a felszínen levő mondanivalóhoz valójában igen nehezen jutunk csak hozzá. A viszonylag legbiztosabb út a forma statisztikus elemzésén át vezet. Amikor egyes versciklusokról eloszlási táblázatokat készítünk, ezt azzal a titkos reménnyel tesszük, hogy a „rideg számok" mögött izgalmas közlemény rejlik: feltéve, ha megtaláljuk hozzá a kulcsot.

5.2. És ha nem találjuk meg a kulcsot? Akkor talán megtalálják később mások. Addig is: meghatároztuk egyes költők nyelvi fiziognómiáját, megrajzoltuk nyelvük jellegzetes vonásait, még ha nem is tudjuk, mi rejlik ezek mögött a vonások mögött. (Valószínű, hogy a mennyiségi mutatók önmagukban is használhatók, így például vitatott eredetű művek szerzőjének azonosításához.)

5.3. Szabó Lőrinc és Juhász Gyula verseinek részleges feldolgozása után a költők esszéinek (és, lehetőség szerint, leveleinek) egy részét szeretnénk azonos módszerekkel megvizsgálni. Ennek során el kellene dőlnie annak, hogy milyen eloszlási sajátságok tekinthetők egyénieknek, és melyek a téma, a műfaj függvényei.

Hogy az egyéniség determináló szerepét a műfajétól elhatárolhassuk, minden műfajon belül két erősen eltérő jellemalkatú szerző művét elemezzük (Szabó Dezső, illetőleg Móra Ferenc; Lázíczius Gyula, illetőleg Vértes O. András — szelídebb fonetikust nem találtunk.)

Ha a jellemalkat valóban módosítja egyes nyelvi jelenségek eloszlását, feltehető, hogy bizonyos jellemvonások bizonyos nyelvi sajátságokkal hozhatók összefüggésbe. Ha valóban vannak ilyen jellemtől függő mennyiségi mutatók, akkor a nyelvi statisztika a pszicho-diagnosztikában is felhasználható. (Annyi bizonyos, hogy beszédünk fonémaeloszlását vagy szófajeloszlását bajosabb akaratlagosan módosítani, mint a kézírást.)

5.4. De mindennek a nyelvész számára alig van mondanivalója. Ugyanez az út vezet azonban egy-egy szövegtípus, egy-egy „műfaj" és végső soron, nagy vargabetűvel, a mai köznyelv számszerű jellemzéséhez.

A szövegek tartalmával, esztétikumával, egyéni vonásaival nem foglalkozó nyelvész számára sem közömbös, miben különbözik a lírai költészet nyelve az értekező próza, a sajtó vagy a köznapi beszélgetés nyelvétől. Módosul-e „műfajonként" a fonémák, a szótagtípusok, a szófajok eloszlása, a szavak, a mondatok hossza, szerkezete?

5.5. Nemcsak az egyes „műfajok" nyelve mutathat számszerű eltérést. Eltérő lehet egyes elemek eloszlása t á r s a d a l m i r é t e g e n k é n t, vagy n e m z e d é k e k szerint. (Tudjuk például, hogy az *ágyút* /u:/-ja, a *fésűt*/y:/-je, a *házból* /o:/-ja sokkal gyakrabban rövidül meg a fiatalok beszédében, mint az idősebbekében.[13]

5.6. De vajon mi közük a nyelvhez az ilyen mennyiségi mutatóknak, amikor a nyelv csak minőségi különbségeket ismer?

A nyelvtudomány kialakulása óta tudjuk, hogy a nyelv állandóan változik. Ehhez ma már azt is hozzátehetjük, hogy ez a változás a nyelv bármely síkján egyidőben érvényes változatok harcán keresztül zajlik le. A [fe:ʃy:t] és [fe:ʃyt], a *lány* és *leány*, a *gyerek* és *gyermek*, a *mert* és *mivel*, az *iszok* és *iszom* váltakozása a nyelvi változás szinkrón vetülete. Az eltérő

[13] Vö.: Magdics Klára, A szóvégi magánhangzók rövidülése a köznyelvben: Nyelvtudományi Közlemények 62. 301—24 (1960).

értékű, de azonos jelentésű változatok egyidejű jelenléte leplezi el a változást, s nem a folyamat fokozatos volta, lassúsága. A permanens váltakozás, a régi és az új egyidejű jelenléte a nyelvben azt a téves látszatot kelti, hogy a magyar köznyelv nem változott az 1900-as évek óta. S valóban, a *gyerek* szót elvétve megtaláljuk már az 1900-as évek folyamán írott levelekben, novellákban, másfelől a *gyermek* vagy az *atya* szóról sem állíthatjuk, hogy hiányzik a mai köznyelvből, sőt: nem állíthatjuk ezt még a *halld* alakról sem. A változatok — a *gyerek* és *gyermek*, az *apa* és *atya*, a *halld* és *halljad* — gyakorisága változott csak meg, és ezen keresztül a változat stiláris értéke. Ennek következtében munkatársam, Magdics Klára csodálkozva nézne rám, ha a fiamról szólva a *gyermek* szót használnám; csak valami diákos tréfából mondhatná ma az apjának a fiú, hogy „Felöltözöm, *atya*".[14] Még kevésbé mondaná a mai utcaseprő a társának, hogy „*Halld*, Sándor", még akkor sem, ha regényben szerepel. Holott Bródy Sándor utcaseprője még bátran szólhatott így a század elején: „Táti, *halld* csak, ez a vén kötélrevaló . . ."[15] S nem véletlen, hogy egyetlen mai slágerben sem mondja magáról a sláger hőse, hogy „*mily bohó*" volt.

A nyelvi statisztika — amint ezt Papp László[16] kitűnő tanulmánya is megmutatta — nagymértékben elősegítheti a nyelvtörténeti vizsgálatokat is. A minőségi, nyelvi változáshoz vezető mennyiségi változásokat csak statisztikusan tudjuk leírni. Csak az ilyen „holt számok" segítségével ragadható meg ezért eleven valóságában a köznyelv.[17]

[14] Ezt a mondatot Abonyi Lajosnak „A plivai vár" című novellájából idézzük: Hét, 1898, 196.

[15] Egy tragédia 17.

[16] Vö. Papp László: Nyelvjárástörténet és nyelvi statisztika (Bp., 1963).

[17] Az előadás elhangzása óta mind a program (vizsgálati szempontok, az eredmények kiírásának módja), mind pedig a vizsgálandó anyag jelentősen kibővült. Időközben ugyanis lehetővé vált, hogy az M—3 gép az eredményeket a közvetlen felhasználás céljaira sokkal alkalmasabb módon írja ki táblázatokban. Ennek megfelelően új program készült, melyet Kalmár Ágota és Bánkfalvi Zsolt egyetemi hallgatók készítettek.

A program módosítása révén pontosabb ismereteket szerezhetünk a szótagtípusok eloszlására vonatkozólag. Meg tudjuk ugyanis azt is határozni, hogy melyik szótagtípus milyen hosszú szó hányadik szótagját alkotja, vagyis meghatározhatjuk a szótag-típusok eloszlását a szó hosszának függvényében aszerint is, hogy hányadik helyen áll.

A *B.* program megvalósítása főleg a mondatszerkezetekre vonatkozóan hoz majd jellemző eredményeket. Meghatározhatjuk az egyszerű és az összetett mondatok arányát; az összetett mondatokban meghatározható a tagmondatok száma (a a tagmondatokon belül a szavak száma) és ugyanakkor az összetett mondaton belüli helyük is. Költői szövegben meghatározható az „enjambement"-ok száma is. Ez az új program most (1963 elején) van kipróbálás alatt. Eddig elkészült Szabó Lőrinc „Tárlat: a Sátán műremekei" és „Mérget, revolvert" című ciklusának szógyakorisági vizsgálata; most folyik Juhász Gyula verseinek a feldolgozása. A vizsgálandó szövegeket Szabó Dezső, Móra Ferenc és Tóth Árpád műveivel egészítettük ki; ezek szalagra lyukasztva várják a vizsgálatokat.

Kozma László

Nyelvstatisztikai automata

1. Bevezetés

Egy ilyen berendezésnek célja a magyar nyelv statisztikai felvételeinek elkészítésében a nyelvészek segítségére lenni. Egy gépi berendezés — mint általában az automaták — az embernél lényegesen gyorsabban, megbízhatóbban és olcsóbban készíti el a felméréseket és szabadítja meg egy egyébként fárasztó és unalmas elfoglaltságtól az embert.

Gépi feldolgozás céljaira a vizsgálandó szöveget olyan formában kell leírni, hogy az elektromosan leolvasható legyen. E célra legalkalmasabb a géptávíró-rendszerben használt lyukasztott szalag. Ennél jobbat is el lehetne képzelni, azonban ez létezik és célszerű felhasználni. A Hollerith és hasonló lyukkártya-rendszerek ugyancsak alkalmasak, de lényegesen drágább berendezések, mint a géptávírók. A kiértékeléshez pedig mindenképpen kapcsoló áramkörök kellenek.

Egy ilyen géptávíró a gépelés folyamán két feladatot lát el egyidőben: egyrészt leírja papírlapon olvashatóan a gépelt szöveget, másrészt pedig szalagon, lyukasztás révén — elektromos leolvasás céljából — rögzíti a feldolgozandó anyagot.

E referátumban röviden ismertetni kívánom a szalagon, lyukasztás révén rögzített információ lényeges vonásait, majd vázlatosan leírom a berendezés általános működését, az ilyen berendezés feldolgozó kapacitását, továbbá azt, hogy milyen statisztikák készíthetők segítségével, és hogy e statisztikák elkészítése mennyi időt vesz igénybe. Végül pedig meg fogom adni, hogy egy ilyen berendezés realizálása milyen feltételektől függ.

2. Lyukasztott szalag

A géptávíróban használatos szalag 17 mm széles, vékony (kb. 1/15 mm vastagságú) papírból készül. E szélesség mellett a papíron 5 lyuknak megfelelő hely van, ami azt jelenti, hogy egy sorban $2^5 = 32$-féle kombináció rögzíthető. (A 6. kisebb méretű lyuk információt nem tartalmaz, hanem a szalag vezetésére szolgál.) Egy lyuk 1 bit információt tárol.

A 32-féle jel a géptávíróban sem elégséges, és ezért azt a rendszert dolgozták ki, hogy — a közönséges írógépnél bevált módszer szerint — minden kalapácson 2 jel van, és a kalapácsokat egy billentyű lenyomásával lehet felemelni vagy visszasüllyeszteni. Így a géptávíróban is van egy

lyukkombináció, amely a kalapácsokat felemeli, és van egy másik, amely visszaejti. A géptávíróban az egyik csoportot *betűknek*, a másik csoportot *jeleknek* nevezik, azonban a mi céljainkra helyesebb *A* és *B* csoport elnevezést használni. A rendelkezésünkre álló 32-féle kombinációból k e t t ő t felhasználunk tehát arra, hogy a soron következő jelek egyik vagy másik csoportba tartozását rögzítsük. Elvben tehát a kombinációk számát ily módon meg lehet duplázni, gyakorlatilag azonban — a géptávíró jellegéből következően — bizonyos kombinációkra van szükség, amelyeknek mindkét csoportban közvetlenül elérhetőknek kell lenniük. Ilyenek: a sorváltás jelzése; a soremelés jelzése; a szóköz jelzése; az üres papír jelzése (ha egy lyuk sincs a papíron). Továbbá a már említett két jelzés: az *A* csoport jelei következnek; a *B* csoport jelei következnek. Így tehát a szalagkombinációk száma:

$$N = 2/2^5 - 6/ = 52$$

A magyar betűknek, illetőleg hangoknak elosztása a két csoportba a nyelvészek kívánságainak megfelelően történhetik. A meglevő géptávíró kalapácsain levő karakterek egy részét mindenképpen ki kellene cserélni és a felírásokat hasonlóképpen megváltoztatni. Ez nem nagy munka, és ha a nyelvészeknek különleges kívánságaik vannak, akkor akár valamennyi karaktert módosítani lehet. Én példaképpen összeállítottam a két csoport jeleit oly módon, hogy az *A* csoportba tettem a leggyakrabban előforduló 26 hangot, míg a második csoportba a dupla betűknek megfelelő hangokat, bizonyos hosszú magánhangzókat, valamint a szükséges írásjeleket. Összesen tehát 52 ilyen használható kombináció lehetséges.

A szalagon két lyuksor közti osztás távolság 2,5 mm, vagyis cm-enként 4 jelet lehet rögzíteni. Egy tekercs 250 m hosszú szalagot tartalmaz, amire tehát 100 000 lyukkombináció fér rá. Tekintettel arra, hogy a már említett csoport átváltási és egyéb jelek kiesése miatt nem lehet a szalagot 100 százalékosan kihasználni, ezért feltételezhetően egy ilyen üres tekercsre kb. 80 000 jelet, azaz 2 ívnyi terjedelmű szöveget lehet rályukasztani. Ha a vizsgálatnak 2 ívnél nagyobb terjedelemre kell kiterjednie, akkor semmi akadálya nincs egy második vagy további szalag-tekercs elkészítésének, amelyeket azután egymás után kell a gépben feldolgozni.

3. Blokkdiagram

3.1. A berendezést képező szerelvényeket egy blokkdiagram mutatja. E szerint a teljes berendezés négy fő részből áll: géptávíró, beállító mező, letapogató készülék, kiértékelő áramkörök.

A beállító rész kapcsoló hüvelyekből álló mező, amelyen dugaszolás révén be lehet állítani előre, hogy a gép — kapacitásán belül — milyen fajta kombinációkat vizsgáljon, illetőleg másfajta vizsgálatokat végezzen el egyidejűleg. Ezen a kapcsoló mezőn felül az 52-féle lehetséges jelnek megfelelő lyuksorozatot találunk, míg a kapcsoló mező alsó része olyan kapcsoló hüvelyeket tartalmaz, amelyek a bemenetet képviselik a kiértékelő áramkörökhöz. Ez az elrendezés teljes rugalmasságot biztosít, a vizs-

gálandó betűk (hangok) vagy betűkombinációk (hangkombinációk) beállítása rendkívül gyorsan és egyértelműen történhet meg.

A berendezés statisztikai felvevő kapacitását le kell rögzíteni előre azért, mert a kiviteli terveket ennek megfelelően kell elkészíteni. Elvben nincs határa az egyidejű vizsgálatoknak, de természetesen minél többet vizsgálunk egyidőben, annál több szerelvény szükséges. Természetesen az egyes kiértékelő áramkörök kevesebb jel vizsgálatára is használhatók, mint amire előkészítettük őket. Tehát például egy négyes kombináció vizsgálatára készült áramkör a dugaszolás révén hármas vagy kettes kombináció vizsgálatára is alkalmas.

A letapogató olyan készülék, amely a szalag lyukasztásait érzékeli és a lyukaknak megfelelően 5 párhuzamos vezetéken elektromos kódot ad ki. A letapogatóhoz tartozik egy dekódoló áramkör, amelynek feladata rögzíteni a kódoknak *A* vagy *B* csoportba való tartozását és lebontani a kódot 52 darab kimenetre, amely kimenetek az 52 előfordulható jelnek felelnek meg. A dekódoló áramkör tehát a kapott kód alapján aktiválja az 52 vezeték egyikét.

A kiértékelő rész tartalmazza a következő áramköröket:

a) Mindenekelőtt a tárolókat. A tárolók olyan áramkörök, amelyek a dugaszolás révén meghatározott kombinációk előfordulási gyakoriságát vannak hivatva megállapítani. Ha tehát a leolvasott kód azt a betűt (hangot) tartalmazza, amelyikre a tároló be van állítva, akkor ezt a tényt rögzíti és várja a következő jelet. Ha ez és a többi előre dugaszolással meghatározott jel érkezik megfelelő sorrendben, akkor az áramkör megállapítja a dugaszolt kombináció azonosságát és egy számláló jelfogó működtetésével regisztrálja e tényt. Ha más jel (betű) érkezne közben, akkor a tároló kiürül és az ellenőrzést újból kezdi. A számláló jelfogók a telefontechnikából ismert szerelvények; számláló kapacitásuk 10 000, ami után a számlálást újból nullától kezdik. Visszaállítani őket nullára nem lehet, ezért a számláló jelfogók állását a vizsgálat indulásakor, majd a vizsgálat befejezése után le kell olvasni és a kettőnek a különbsége adja meg a kívánt statisztikai adatot.

b) A kiértékelő részhez tartozik az úgynevezett átlagoló áramkör, amely a rögzített szöveg egyes szavainak átlagos betűhosszát adja meg, valamint azt is, hogy a szóhosszak az átlagos szóhossztól felfelé és lefelé mennyit térnek el, vagyis egy szórási görbe készítésére alkalmas. E célra a berendezés 15 számláló jelfogót tartalmaz, amelyen leolvasható, hogy a szövegben hány szó fordult elő 1, 2, 3, . . . 14, 15 (és több) betűből. Ugyanez a berendezés arra is felhasználható, hogy a szöveg mondatainak számát rögzítse és megállapítsa, hogy a mondatokban előforduló szavak számának mennyi az átlaga és mennyi a szórása felfelé és lefelé.

A berendezéshez tartozik még olyan áramkör, amely gondoskodik a letapogató készülék automatikus léptetéséről és egyáltalában az egész berendezés működésének vezérléséről.

3.2. A berendezésben felhasznált alkatrészek a híradástechnika ismert építőelemei: jelfogók, elektroncsövek, kondenzátorok, diódák, ellenállások stb.

A jelfogók működési sebessége, valamint a letapogatónak a léptetési ideje határozzák meg az egész berendezés működési sebességét. A berendezést el lehetne készíteni tisztán elektronikus elemekből is, azonban a letapogatónak szükségszerűen egy mechanikai szerkezetnek kell lennie, amely az elektronikusan megvalósítható áramkörök nagy sebességének kihasználását nem tenné lehetővé, s ezért célszerűbb az áramköröket csak részben elekronikusan, nagyobbrészt pedig elektromágneses módon megvalósítani.

A felhasznált elemekből következik, hogy az elérhető sebesség kb. 4 sor leolvasása pro másodperc, illetőleg másodpercenkénti 1 cm-es szalag leolvasása, amiből következik, hogy a 250 m hosszú tekercs leolvasási ideje 25 000 mp, azaz kereken 7 óra. Minthogy a kiértékelő áramkörök párhuzamosan egymástól függetlenül tudnak működni, így célszerű egy leolvasás alkalmából valamennyi ilyen tároló áramkört, sőt az átlagoló áramkört is üzembehelyezni (dugaszolás révén). Ily módon a berendezés a lyukasztott szöveget kb. 60 szempontból tudja egyidejűleg vizsgálni, vagyis a másodpercenként végzett vizsgálatok száma 240 körül van, ami nyilvánvalóan legalább százszorosa az egy ember által végezhető felmérésnek. Amellett az ilyen munka végzése fárasztó, és ha egyidejűleg több szempontból kell a szöveget vizsgálni, akkor vagy nagyon lelassul a felmérés, vagy pedig hibákkal kell számolni. A gépi berendezés tehát ilyen esetben nagy segítséget jelent.

4. A teljes berendezés

A teljes berendezés az eddig elmondottakból következtethetően a következő részeket tartalmazza.

a) A távgépírót, amely a lyukasztott szalag előállítására szolgál. Ez a rész elválasztható a feldolgozó berendezéstől és földrajzilag is máshova vihető. Egy ilyen géptávírót kaptam a Posta Távirda Hivatalától. A gépet át kell alakítani, mint ahogy korábban jeleztem, a karakterek kicserélése céljából a nyelvészek igényeinek megfelelően.

b) A következő rész a letapogató, amelyet szintén a Távirdától kaptam és amelyet már átalakítottunk a nyelvvizsgáló berendezés céljaira, azaz arra, hogy a berendezés által igényelt elektromos kódokat adja vissza.

A tároló és átlagoló áramkörök a következő alkatrészmennyiségeket tartalmazzák:

400 db kapcsoló jelfogót (amilyeneket a BHG tömegszerűen gyárt telefonközpontok céljaira);

80 db számláló jelfogót (ugyancsak a telefonközpontokban használatos tömegcikk);

30 db dupla triódát (szintén tömegcikk).

Mindezeket a szerelvényeket, valamint a hozzájuk tartozó egyéb áramköri elemeket (diódákat, kondenzátorokat, ellenállásokat stb.) egy lehetőleg pormentesen elzárt szekrénybe kell beszerelni, amelynek a méretei: kb. 180 cm magasság és 60 cm szélesség.

5. Helyzetkép

A géptávírót és a letapogatót 1960 márciusában, tehát két évvel ezelőtt kaptam meg a Távirda Hivataltól. Utána felvettem a kapcsolatot a Nyelvtudományi Intézettel, személy szerint Fónagy Iván elvtárssal, és megbeszéltük egy ilyen berendezés alapelveit, főleg a nyelvészek részéről támasztható igényeket. 1960 szeptemberében írtam a Magyar Tudományos Akadémia Műszaki Tudományok Osztályának és kértem kb. 60 000 forintnyi hitelkeretet a berendezés elkészítésére. A Műszaki Osztály válasza 1960. november 11-i keltezésű és lakonikusan csak annyit tartalmazott, hogy a pótcéltámogatási forint kérés kerethiány miatt nem teljesíthető. Ilyen előzmények után 1961. évre a Vezetékes Híradástechnikai Tanszék akadémiai tervében mégis beállítottam a nyelvvizsgáló berendezést is. Az Osztályvezetőség 1961. február 17-én a témát elfogadta. Ekkor az áramköri tervek már nagyjából elkészültek, csak éppen valakinek szorgalmaznia kellett volna a berendezés kivitelét. Sajnálatos módon az én hivatali elfoglaltságom akkoriban sokszorozódott meg, így elsősorban az oktatási reformmal kapcsolatos sokoldalú tevékenységem megakadályozott abban, hogy a témával továbbra is foglalkozni tudjak, a tanszéken pedig nem volt senki, akit ez különösebben érdekelt volna. Minthogy a nyelvészek részéről sem érkezett érdeklődés, illetőleg különösebb sürgetés, a dolog elaludt.

Az év elején, amikor szóba került ennek a kongresszusnak megtartása (eredetileg Debrecenben), újból elővettem az anyagot, az áramköröket átnéztem, és a még hiányzó elvi munkát befejeztem. A hátralevő mérnöki munka most már csupán az, hogy el kell készíteni az elrendezési rajzokat, valamint a készen levő elvi kapcsolások alapján a kábelezési listákat. Érdeklődésemre közölték, hogy a BHG kb. három hónap alatt le tudná szállítani a szükséges jelfogókat és ugyancsak kb. három hónap lenne szükséges a berendezés összeállítására, bekábelezésére és kipróbálására. Vagyis az egészet egy félév alatt üzembe lehetne helyezni.

Ami a költségeket illet, egy előzetes becslés a jelenlegi elvi rajzok alapján azt mutatja, hogy a teljes összeg kb. 80 000 forintot tenne ki. Tanszékem a folyó évre kb. 50 000 forinttal nagyobb hitelt kapott a Magyar Tudományos Akadémiától, mint amennyit kértünk, azonban egy másik munkára, és nem erre a nyelvvizsgáló berendezésre, amit az idei költségigénylésünkbe nem vettünk be.

A berendezés sorsát illetően a Nyelvtudományi Intézetnek a véleménye lenne döntő. Amennyiben egy ilyen berendezést szükségesnek tartanak, akkor meg lehetne kérni a Műszaki Tudományok Osztályát, hogy az említett 50 000 forintos hitelt írják át erre a nyelvvizsgáló berendezésre és azután még valahonnan kellene kb. 30 000 forint hitelt szerezni.[1]

[1] A berendezés elkészítéséhez szükséges anyagot az előadás óta eltelt idő alatt a tanszék megrendelte a Beloiannisz Híradástechnikai Gyártól; az anyag leszállítására 1963 februárjában került sor. A berendezés kábelezése most folyik és üzembe helyezése 1964 nyarára várható.

Kisebb közlemények*

Antal László

1. Bevezetésképpen szeretnék emlékeztetni a tavalyi konferencián elhangzott felszólalásom utolsó mondataira. Ezek ugyanis igen széles körű elterjedtségnek örvendenek, ha nem is mindig egyetértő elterjedtségnek. Azt mondtam akkor — azt és nem mást —, hogy eljött az ideje annak, hogy írjunk egy olyan magyar nyelvtant, amely egy jottányi engedményt sem tesz a maga helyén oly értékes történetiségnek, s amely következetesen a formális elemzés szempontjaira épül, tehát egy jottányi engedményt sem tesz a jelentésnek.[1] Ezt mondtam és ebben az összefüggésben; ezért a jövőben azt szeretném, ha mindazok, akik idézik, így idéznék.

2. A mai előadásokban ismét szó esett t a r t a l o m és f o r m a viszonyáról a nyelvben. Ezzel kapcsolatban a következő meggondolásoknak szeretnék hangot adni.

2.1. Minden nyelvközösségben minden időben rendkívül nagyszámú nyelvi megnyilatkozás hangzik el. Ezeket a megnyilatkozásokat — mondatokat — igen sokféle szempontból elemezhetjük. Én pillanatnyilag a következő osztályozás lehetőségét vetném fel: A naponta elhangzó mondatokat feloszthatjuk egyfelől olyan mondatokra, amelyeket mindenki mondhat mindenkinek. Ilyenek: *Esik az eső; Éhes vagyok; Indul a vonatom ;* stb. Ezek mellett van a mondatoknak egy másik nagy csoportja, azoknak a mondatoknak a csoportja, amelyeket nem mindenki mondhat mindenkinek. Ilyenek a csillagászat mondatai, amelyeket csak csillagász mondhat csillagásznak, mert más nem értené meg, vagy a matematika mondatai, de ilyenek az asztalosmesterség speciális mondatai is. Az előbbi mondatokat u n i v e r z á l i s, ez

* Az itt következő „kisebb közlemények" (eredetileg hozzászólások) a kötet első felének tanulmányaihoz (előadásaihoz), illetőleg az azokban fölvetett gondolatokhoz kapcsolódnak. — *A szerkesztők.*

[1] „A nyelvészetnek az a feladata, hogy a strukturalista vizsgálatokat maximálisan felvirágoztassa, hazánkban első feladatként a magyar nyelv strukturális elemzését, tehát egy valóban modern magyar nyelvtan létrehozását jelenti. 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 jottányi engedményt sem a maga helyén oly értékes történetiségnek. A másik, hogy a formális elemzésen alapuljon, és ne tegyen egy jottányi engedményt sem a jelentésnek." A Magyar Tudományos Akadémia Nyelv- és Irodalomtudományi Osztályának Közleményei 18. 54 (1961).

utóbbiakat partikuláris mondatoknak nevezhetjük. Mármost felvetődik a kérdés, hogy a nyelvtudomány a mondatok eme két nagy csoportja esetében mit csinálhat a mondatok tartalmával.

Az nyilvánvaló, hogy a partikuláris mondatok tartalmát a nyelvtudomány nem vizsgálhatja. A csillagászat vagy a matematika mondataihoz csak csillagász vagy matematikus szólhat hozzá tartalmi szempontból. Az univerzális mondatok tartalmát már a nyelvész is vizsgálhatná, de nemcsak ő, hanem még jó néhány más tudományág képviselője. Mármost nyilvánvaló, hogy a nyelvtudomány csak azt vizsgálhatja a mondatok esetében, amit minden mondat esetében meg tud vizsgálni és minden mondat esetében csak ő tud megvizsgálni. Mint láttuk, a partikuláris mondatok tartalmát a nyelvtudomány nem vizsgálhatja, az univerzális mondatok tartalmát pedig nemcsak ő vizsgálhatja. Ennek megfelelően a tartalom vizsgálata eleve és egyértelműen kívül esik a nyelvtudomány feladatkörén.

2.2. Gondolatban most ismét osszuk fel a naponta elhangzó mondatokat két nagy csoportra. Az egyik csoportba tartoznak azok a mondatok, amelyek a nyelvről szólnak, a másikba pedig azok, amelyek nem a nyelvről szólnak. Nyilvánvaló, hogy az első csoport mondatai, tehát azok, amelyek a nyelvről szólnak, teszik ki a nyelvtudományt. Mármost a nyelvtudomány mondatainak is van tartalma, De miből állnak ezek a tartalmak? Láttuk, hogy a nyelvtudomány nem vizsgálhatj, más mondatok tartalmát. Ily módon a nyelvtudomány mondatainak a tartalma a többi mondat formájára vonatkozik, arról szól. Ami tehát egyébként formaa az a nyelvtudomány számára tartalom.[2]

3. Szeretnék egy-két szót mondani a formális logikával kapcsolatban is A modern logika jeles művelői, elsősorban pedig Carnap, gyakran nevezik a maguk tudományágát szemantikának. Carnapnak egy könyve is van ilyen címmel: „Bevezetés a szemantikába."[3] Mivel a nyelvtudományban is *szemantika* a 'jelentéstan' neve, egyeseket a terminológia megtévesztett. S azt mondták: Íme, már a formális logika is foglalkozik jelentéstannal, hát a nyelvészek is figyelembe vehetnék a jelentést a nyelvtani elemzésben. (A jeruzsálemi egyetem világhírű tudósa, Bar-Hillel is beugrott a Carnap-féle „szemantiká"-nak.[4]) Nos, ha Carnap vagy más logikusok úgynevezett szemantikáit felnyitjuk, világosan látnunk kell, hogy annak semmi köze a nyelvészek által művelt jelentéstanhoz. Carnapék — mint minden logikai kutató — a gondolkozás formáival foglalkoznak, azokkal a formális feltételekkel, amelyek között egy ítélet lehet igaz vagy hamis. Például a következő két ítélet — *Boston az Egyesült Államok legnagyobb városa ; Nem Boston az Egyesült Államok legnagyobb városa* — szerintük oly módon zárja ki egymást, hogy vagy az egyik igaz, vagy a másik. Ezzel szemben e két mondat a nyelvtudomány szempontjából nem kizáró viszonyban áll egymással szemben. Az igaz meg a hamis a nyelvtudománynak nem kategóriája. A két mondat csak különböző a nyelvtudomány szempontjából.

[2] Részletesebben fejtem ki e gondolatot „Content, Meaning, and Understanding" című kis könyvemben (Hága, 1964)

[3] R. Carnap: Introduction to Semantics (Cambridge, Mass., 1942).

[4] Vö. pl. Y. Bar-Hillel, Logical syntax and semantics: Language 30. 230—7 (1954).

Az egyikben van egy morféma-többlet *(nem)*, a másikban nincs meg. A nyelvtudomány szempontjából tehát egész más típusú a különbség.

4. Nagy figyelemmel és haszonnal hallgattam végig Kalmár László professzor úr előadását. Mindenki, akinek Magyarországon a nyelvtudomány haladása szívügye, Kalmár professzor urat tiszteli, mert sokat segített nekünk, és hisszük, hogy a jövőben is fog segíteni. Éppen ezért rendkívül fájlalom, hogy előadásában — nyilván időhiány miatt — nem kerülhetett sor a matematikai logika formális eljárásainak az ismertetésére. Ugyanis a modern nyelvtan tudatosan formális utakon halad, s e tekintetben különösen sokat tanulhatnánk a matematikai logikától, annak elegáns formális apparátusától.

Petőfi S. János

A homonimiának és a nyelv halmazelméleti modelljének kapcsolatával kívánok foglalkozni, megmutatva a nyelv egy olyan meghatározását, amely a homonimiákat különösen figyelembe veszi. Két dolgot kell előrebocsátanom.

Közismert tény, hogy a homonimia meghatározása az általános nyelvészetben nem egyértelmű. Bárczi Géza állapítja meg, hogy a „homonimia is azok közé a műszavak közé tartozik, amelyekről mindnyájan úgy érezzük, hogy teljesen világosan látjuk fogalmi tartalmukat, mégis ha utánafürkészünk, kiderül, hogy egyenként közel sem ugyanazt értjük rajta."[1] E kérdés természetesen — éppen bonyolultsága következtében — a nyelv halmazelméleti modelljében is előkerül. Lehet, hogy ennek megmutatása nem visz közelebb a megoldáshoz, hogy ez csupán más szavakkal, más jelekkel mondja ugyanazt, amit már eddig is mondtak. Meg kell azonban jegyeznünk, hogy nem két tudomány között, hanem egy tudományon belül is egy kérdés más megfogalmazása annak megoldásához is hozzásegíthet. A matematikában ez az „egy nyelven belüli fordítás" több új eredményhez vezetett.

Ami pedig a halmazelméleti modellt illeti, itt is marad több nyitott kérdés. Én csak egy halmazelméleti modellnek a „morfológiai tér" meghatározását mutatom meg, nem vizsgálva azt, hogy ez a modell bármely nyelvre, vagy csak bizonyos nyelvekre (esetleg egyetlen nyelvre) alkalmas-e. A bemutatott példákat, ellenpéldákat tehát ilyen jelleggel kell néznünk.

Ezek után nézzük meg a homonimák és egy halmazelméleti modell kapcsolatát. Kiefer Ferenc, Kulagina alapján, a magyar nyelvre a következő szócsaládokat írta fel:

C_1 *asztal, szék, . . .*
C_2 *asztalt, széket, . . .*
C_3 *jön, megy . . .*
. . . . . .

és megjegyezte, hogy ezek — Kulagina feltevése szerint — egymástól idegen halmazokat kell, hogy alkossanak, azaz olyanokat, amelyeknek közös elemük nincs.[2]

[1] Bárczi Géza, A szótári homonimia kérdéséhez: Magyar Nyelv 54.42 (1958).
[2] Kiefer Ferenc, Halmazelméleti és matematikai-logikai modellek a nyelvben: Kötetünkben 89—115.

Szinte tetszés szerinti számban sorolhatók fel azonban olyan szavak, amelyek két, esetleg több szócsaládba is tartozhatnak. A C_1 és C_3 család például az alábbi módon folytatható:

C_1 *asztal, szék, . . . ég, tűz, fog, dob, szív, ér, vár*
C_3 *jön, megy, . . . ég, tűz, fog, dob, szív, ér, vár*

Egyidejűleg több családhoz tartozó szavakat is felsorolhatunk, amelyek például az Értelmező Szótárban mint *bár*[1], *bár*[2], *fél*[1], *fél*[2], *ír*[1], *ír*[2], *ír*[3], . . . találhatók (a szótárban feltüntetett homonima-index természetesen nem egyezik azoknak a családoknak a sorszámával, amelyekhez ezek a szavak besorolhatók). Olyan különlegességszámba menő homonimákat is említhetünk, mint például

ár[1] 'valamely áru értéke'
ár[2] 'áramló víztömeg'
ár[3] 'szerszám a bőrfeldolgozó kisiparban'
ár[4] 'területmérték'

Ezeket az előbb felírt szócsaládok közül egyesekben egy, másokban több szóalak képviseli, például:

C_1 .....*ár*$^{(1, 2, 3, 4)}$
C_3 .....*árat*^{1} *árt*$^{(2, 3, 4)}$, . . .

Itt az árt szóalak még grammatikai homonimával is rendelkezik.

A homonimák zavarják a szófaj Kulagina-féle halmazelméleti meghatározását is és nem marad helytálló a „szabályos nyelv" meghatározása sem.[3] A

szék, asztal, ház, házat

sor alapján a szék és a házat egy típusba tartozik, a

szék, asztal, vár, vártam

sor alapján ugyanígy egy típusba tartozna a *szék* és a *vártam ;* mert a követelménynek, hogy tudniillik bármely két egymás utáni szó ugyanabba a családba, vagy ugyanabba a környezetbe tartozzék, ez a sor is eleget tesz (ha csak a szóalakot tekintjük).

Ezeknek a problémáknak jó része megszüntethető, ha a valamilyen nyelvi szempont alapján megkülönböztetett homonimákat megkülönböztetjük például indexszel is. Nem elképzelhetetlen az sem, hogy a családok és környezetek segítségével is meghatározható, hogy hány azonos alakú, de eltérő jelentésű szót képvisel ugyanaz a szóalak. Egy-egy homonima környezetét, mint megkülönböztetőt, az Értelmező Szótár is közli, például *ár*[1] *(-at, -a), ár*[2] *(-t, -ak, -ja).*

Egy ilyen értelmű indexelés bevezetése megszünteti az említett ellenpéldák okozta ellentmondásokat is.

A nyelv homonimiát is figyelembe vevő felépítésére Solomon Marcus román kutató tesz kísérletet.[4] Modellje topológiai modell. A topológia ismertetésével itt nem foglalkozom. (Meg kell jegyeznem, hogy egyszer már Vértes Edit is kapcsolatba hozta a topológiát a nyelvvel)[5]

[3] Kiefer Ferenc: i.m. 95.
[4] S. Marcus, Structures linguistiques et structures topologiques: Revue de Mathématiques Pures et Appliquées 6. 501—6 (1961).
[5] Vértes Edit, Szótározott szavaink száma és a szókészlet nagysága: Magyar Nyelvőr 79. 329—38 (1955), a topológiáról: 337.

Az ismertetésre kerülő részhez a topológiai fogalmak közül elegendő csupán a nyílt halmaz fogalmát ismerni, ez pedig a következőkből adódik: Annak, hogy az M halmaz nyílt legyen, szükséges és elegendő feltétele, hogy M bármely pontjának legyen olyan környezete, amely teljes egészében M-hez tartozik. — A nyelv szóalakjainak halmaza — a Kulagina-féle környezetfogalom esetén — felfogható ilyen halmazként, ugyanis ha valamely szóalakot tekintünk „pont"-nak, nyilvánvaló, hogy környezete is a nyelvhez tartozik.

Egy topológiai teret nemcsak azzal lehet megadni, hogy megmondjuk, mely elemek a pontjai és minden egyes pontjának melyek (vagyis a tér mely részhalmazai) a környezetei, hanem azzal is — és a modern topológiában ez a szokásos —, hogy megadjuk, mely részhalmazait tekintjük nyíltaknak.

Marcus szerint: „Egy nyelv szóalakoknak olyan E halmaza [un lexique], amely három struktúrával rendelkezik: morfológia, szintaxis és fonológia."[6] Nézzük a legegyszerűbb struktúra, a morfológiai tér meghatározását. Előre kell bocsátanom, hogy Marcus nyelvi példák nélkül, tisztán matematikai felépítést ad. A meghatározásban szereplő „egy nyelv" kitétel arra enged ugyan következtetni, hogy ezt bármely nyelvre érvényesnek tekinti, ami azonban csak alapos ellenőrzéssel lenne eldönthető. Ennek figyelembevételével interpretálom elméletét magyar példák alapján.

Jelöljük I_x-szel egy adott x szóalak homonimáit jelző indexek halmazát és i_x-szel az I_x halmaz elemeinek számát. Például:

ha $x = dob$ $\quad I_x = 1, 2 \quad i_x = 2$

$x = ár$ $\quad I_x = 1, 2, 3, 4 \quad i_x = 4$

Ha valamely x esetén $i_x = 1$, akkor e szóalaknak nincs homonimája, ha pedig egy nyelv bármely x szóalakja esetén $i_x = 1$, akkor e nyelvben nincsenek homonimák (Marcus mindenütt a *l'homonymie lexique partielle* kifejezést használja). Ez azt mutatja, hogy először is nem foglalkozik homonim mondatokkal (például: *Megharapta a menyét);* másodszor pedig nem tekinti homonimiának, ha két különböző jelentésű szó valamennyi alakja megegyezik, mint például (valószínűleg) $ár^2$, $ár^3$, $ár^4$ esetén.

Alkossuk meg most minden $x \in E$ paradigmáinak — vagyis az x szóalak által képviselt szó valamennyi alakjának — halmazát és jelöljük ezt $A_\alpha(x)$-szel, akkor az $x \in A_\alpha(x)$ reláció minden $\alpha \in I_x$-re teljesül. Így minden x-hez egy bizonyos $\{A_\alpha(x)\}_{\alpha \in Ix}$ halmazt (illetőleg halmazokból álló halmazt) rendelünk hozzá, amelynek elemei az E részhalmazai. Az $x = dob$ esetén ez a következőt jelenti:

$A_1(dob) = \{dob, dobot, dobnak \ldots\} \quad \alpha = 1 \in I_{dob} = \{1, 2\}$[7]

$A_2(dob) = \{dobok, dobsz, dob, \ldots\} \quad \alpha = 2 \in I_{dob} = \{1, 2\}.$

Több $A(dob)$ halmaz nincs, mert $i_{dob} = 2$, tehát

$$\{A_\alpha(dob)\}\alpha \in I_{dob} = \Big\{\{dob, dobot, dobnak, \ldots\}, \{dobok, dobsz, dob, \ldots\}\Big\}.$$

[6] S. Marcus: i. m. 501.

[7] Antal Lászlónak a magyar főnév paradigmáinak számára vonatkozó megállapításait elfogadva, ez egy 714 elemű halmaz; vö. Antal László: A magyar esetrendszer (Bp., 1961) 50.

Képezzük most ezeknek az $A_\alpha(x)$ halmazoknak az egyesítési halmazát:

$$V(x) = \bigcup_{\alpha \in I_x} A_\alpha(x).$$

Amíg az $\{A_\alpha(x)\}\alpha \in I_x$ halmaz elemei együttesen egy x szóalak összes homonímáját, illetőleg ezek összes paradigmáit tartalmazzák (tehát több azonos szóalakot is), addig a $V(x) = \bigcup A_\alpha(x)$ halmaz csak különböző szóalakokat tartalmaz, például az $x = ár$ esetén

$$\{A_\alpha(ár)\}\alpha \in I_{ár} = \Big\{\{ár, árat, árnak, árral, \ldots ára, \ldots\},$$

$$\{ár, árt, árnak, árral, \ldots árja, \ldots\}\Big\},$$

$$V(ár) = \{ár, árat, árt, árnak, árral, \ldots ára, árja, \ldots\}.$$

Ezek után a nyílt halmazokat Marcus a következő módon értelmezi: az E halmaz egy G részhalmaza akkor és csak akkor nyílt, ha bármely $x \in G$ esetén $V(x) \subset G$. Az így meghatározott topológiával rendelkező E halmazt E_M-mel jelöli, és morfológiai térnek nevezi.

Ahogyan a bevezetőben említettem, csak ismertetni kívántam ezt a fajta halmazelméleti felépítést — illetőleg ennek egy kis részét —, különösebb vizsgálat tárgyává tenni azonban nem. Erről különben is csak akkor lehetne szó, ha a nyelv Marcus-féle topológiai modelljét egészében megismernők. Akkor természetesen azt is meg kellene vizsgálni, hogy mennyiben lehet egyáltalában egy nyelvet mint rendszert topologikus térnek tekinteni, más szóval, hogy milyen axiómáknak tesz eleget a nyílt halmaz fenti Marcus-féle fogalma.

Egy-két rövid megjegyzés még a halmazelmélet nyelvészeti alkalmazásaihoz. Nyelvi elemek halmazait megalkotni nem azt jelenti csupán, hogy bizonyos sokaságot bizonyos elv szerint csoportosítok és esetleg (algebrai, topológiai, halmazelméleti) szerkezetét vizsgálom, hanem az is érdekelhet bennünket, hogy egy-egy ilyen halmaznak hány eleme van. (Gondoljunk például egy szó paradigmáinak halmazára.) De nemcsak olyan jellegű halmazokat lehet alkotni, mint amilyenekkel a nyelv halmazelméleti modelljei foglalkoznak.

Lehet például x = valamely metrum;

$$A(x) = \text{e metrum összes lehetséges ritmikai realizációinak halmaza.}$$

Megalkothatók esetleg az $A_i(x)$ halmazok is, ahol az i futóindex például különböző nyelveket jelenthet. Bármely $A_i(x)$ természetesen részhalmaza lesz az $A(x)$-nek.

Talán az sem lenne érdektelen, ha a stilisztikai kutatások céljára felhasználható gépekkel végrehajtanának olyan mondattípus vizsgálatot, amelynek eredményeképpen egy statisztikailag súlyozott mondattípus-halmazt — úgy is mondhatnám, egy mondattípus-gyakoriság szótárt — lehetne összeállítani. Költői művek ilyen fajta vizsgálata egy valóban objektív alapokon nyugvó versmondattan megalkotását is elősegítené, amit jelenleg még az úgynevezett nagy európai nyelvek is nélkülöznek.

Gáldi László

1. Véleményem szerint sokkal nagyobb óvatosság lenne ajánlatos Guiraud* számításainak a francia nyelvre vonatkozó kiindulópontjaival kapcsolatban. Ami elhangzott, abból sajnos nem derült ki, hogy Guiraud kiindulópontja a nem is túlságosan tudományos igényű, de közismert Petit Larousse volt, amelyről azt mondja, hogy körülbelül 50 000 szó van benne. Ezenkívül Guiraud azt is állítja, hogy a kis Larousse-nak vannak rövidített változatai, amelyekben 30 000 szó van, és ezek közül az átlagos műveltségű francia feltételezhetően tud mintegy 24 000 szót. Így kap Guiraud egy olyan kiindulópontot, amelyhez mint átlagos szókészlethez különböző írók szókincsét próbálja hasonlítani. Sajnálatos tény, hogy a 24 000 „ismert szó" hipotézise semmi pozitívumon — legalábbis az én számomra — nem alapul. Még kevésbé alapul valami kézzelfoghatón az, amit például Molière-ről olvasunk. Hogy csak ezt az egy kérdést ragadjam ki, Guiraud nem végzett semmiféle pozitív számítást — még a meglevő Molière-szótárak alapján sem! —, hanem a „Femmes savantes"-ból, tehát egy verses darabból ragadott ki három részletet, és ennek a három részletnek alapján néhány sorban eldönti azt, hogy Molière-nak kb. 20 000 szava van, tehát ennyi és ennyi százalékkal kevesebb, mint az átlagos francia szókészlet. Hadd mondjak még egy idevágó példát. Rendkívül aggodalmat keltő Guiraudnak az a megállapítása is, hogy az ófranciában nyilván kevesebb szó volt, mint a modern franciában, mivel a „Chanson de Roland"-ban szerinte — Guiraud megint semmiféle forrásra sem hivatkozik! — körülbelül 4000 szó volt. Ehhez csak annyit szeretnék hozzátenni, hogy nyilvánvalóan az ófrancia korban, tehát a francia középkorban, a halászat, a tengerészet, a földművelés rendkívül virágzó foglalkozási ágak voltak; mármost tessék megnézni, mindebből mit tükröz a Chanson de Roland! Nem evidens-e, hogy az ófrancia nyelvnek a reális, élőnyelvi szókészletére nem egyetlen, de különböző vidékekről való munkákból is csak nagy óvatossággal lehet következtetni?

Csak röviden jelzem azt, hogy Guiraud statisztikáival kapcsolatban fogalmunk sincs arról, mit tekint szónak, pontosabban 24 000 szóegyednek. Tehát ha például a *penser* infinitivus mellett van *pensé* participium és van *le penser* mint főnevesített főnévi igenév, akkor most ő ezt egy szónak veszi vagy három szónak? Nyilvánvalóan a statisztika szempontjából döntő fontosságú momentum volna az is, vajon például a *faux* melléknév mellett *faussement* és *le faux* külön szónak veendő-e vagy sem. Mindezt különösen azért kell felvetnünk, mert Guiraud kiindulópontja, sajnos, a legtudománytalanabb francia szótárak egyike. S azért is érdemes mindezt elmondani, mert

* Vö. „A szerkesztők bevezetésé"-nek **5.** pontjával.

ha majd egyszer megbízható magyar szóstatisztikát akarunk készíteni, akkor mindenekelőtt az idevágó külföldi irodalom részletes és tudományos bírálatát kell elvégeznünk.

2. Említést nyert futólagosan az a tény, hogy eddig is készültek már, legalábbis a nyugati nyelvekre, nagyobb igényű szóstatisztikák; nem történt azonban említés egy olyan könyvről, amelyet talán érdemes volna szintén figyelembe venni. Ez tudniillik Helen Eatonnek négy nyelvre (angol, francia, német és spanyol) vonatkozó összeállítása; ebben mindegyik nyelvnek 6000 leggyakoribb szavát találjuk meg, párhuzamos szójegyzékekben, még pedig úgy, hogy 1000, illetőleg 500 szavanként a szógyakoriság 1a-val és 1b-vel van jelezve. (Erre még később visszatérek.) Ez a könyv juttatta eszembe a következő eljárást: mi lenne, ha nem csupán globálisan számolnók meg, hogy egy írónak hány szava van, és ennek alapján foglalkoznánk az író szókészletével, hanem azt is megnéznők, milyen módon is rendeződnek el egy írónál a különböző gyakorisági csoportba tartozó szavak. Hogy ez mit jelentene, legyen szabad egyetlen példával illusztrálnom; mindössze két sorról van szó. Valéry egyik legismertebb verse a „Le cimetière marin"; ennek első szakasza így kezdődik:

Ce toit tranquille où marchent les colombes
Entre les pins palpite, entre les tombes.

Tartalmilag csak annyit, hogy itt Cète (másképp Sette) városról, ahol a költő született, illetve annak egy dombon fekvő, a tengerre néző temetőjéről van szó. Kosztolányi ezt a szöveget a harmincas években így fordította:

Nyitott tető, hol járnak a galambok,
Fenyő s sírok közt, s ő remeg alattok;
Ráhág a dél, tűzzel belengeni . . ."

A versforma részletes tagolása ezúttal felesleges; az egyik előttem szóló szavaihoz kapcsolódva elég annyit jeleznem, hogy a versforma a középkor óta ismert décasyllabenak egyik szép modern példája; már ez a két sor is jól szemlélteti a décasyllabe két legfontosabb variánsát: a 4, 6, illetve a 6, 4 tagolást; az 1. sor az olasz terminológia szerint „endecasillabo a minore", a 2. sor pedig „endecasillabo a maggiore". De maradjunk pillanatnyilag a szóstatisztikánál. Ha Eaton könyve szerint megnézzük, hogy az ebben a szövegben előforduló szóegyedek hogyan állnak az első 6000 szó szempontjából, a következő számokat írhatjuk fel. Rangsorolás szempontjából a *ce* névmás nyilvánvalóan a legközönségesebb elem, tehát elől van, az 1/a csoportban. A *toit* 'tető' szó már kissé messzebb esik (2a); érdekes módon a 'nyugodt' jelentésű *tranquille* melléknév megint az első ezren innenre kerül; 500-on innen van az *où* szócska, a *marcher* ige, valamint a *les* névelő. De most hirtelen ugrás következik: a *colombe* 'galamb' szó 6000 és 6500 között van! Megnéztem kuriózumképpen a *colombe* szó szinonimáját: ha a verssor végén a közönségesebb szinomima, *pigeon* volna, ez a szó 5a jelzést kapott volna, tehát 4000 és 4500 között lenne. *Colombe* tehát valamivel ritkább szó, s egyben itt átvitt jelentésű (a.m. 'vitorlás bárka'). Most lássuk a második sort. *Entre* és *les* persze nagyon elöl vannak, sőt érdekes módon a *pin* 'fenyő' szó is, hiszen Franciaországban sok fenyő van. A *palpite* 'dobog, remeg' (itt inkább: 'reszket, csillámlik') ige 6a jelzésű. Utána megint közönséges szavak vannak, végül a 2. sor utolsó szava, ha nem is nagyon ritka, de mindenesetre nem a legközönségesebb s nem is az első 1000 szó közt foglal helyet. Azt hihetnők, hogy a mai beszélt nyelvben a *tombeau* szó közönségesebb, mint *tombe;* a statisztika szerint viszont *tombeau* valamivel hátrább került.

Mindenesetre látjuk ebből az egyetlen példából is, milyen különös hullámzás van szinte minden sorban a közönségesebb és ritkább szavak közt, és látjuk azt is, hogy a francia vers intonációja — ellentétben a magyarral — rendszerint felfelé szalad; a végső, legerőteljesebb, leghangsúlyosabb pontra gyakorta olyan szó kerül, amely az általános szófrekvencia szerint viszonylag ritkább.

3. Szeretném a figyelmet arra is felhívni, hogy sajnos nyelvstatisztikailag rendkívül nehéz megragadni a szavaknak a „szemantikai mélységét". „Szemantikai mélységnek" nevezem azt, hogy például *palpiter* a statisztika szerint nyilvánvalóan csak mint a szív dobogására vonatkozó, majdnem biológiai műszó volt regisztrálva; szó sem volt tehát arról, hogy valaki a tengernek a déli napfényben való csillogását szintén ezzel az igével fejezze ki. Talán nem túlzás, ha azt mondom, hogy ezúttal valóságos szemantikai „hapax legomenonnal" van dolgunk. Hadd említsem meg, hogy végeztem egy másik kísérletet is; ezúttal nem mutatom be részletesen, hiszen bárki más szintén elvégezheti a megfelelő adatok segítségével. Elolvastam Mallarmének közismert „Les fenêtres" (Az ablakok) című, valóban szimbolista, sőt szinte programversét, és megnéztem — mivel nemrégiben más vonatkozásban úgyis a jelzőkkel foglalkoztam — vajon az itt előforduló jelzők hogyan oszlanak meg gyakorisági fokok szerint. Mallarmé költeménye jelzőkkel meglehetősen túlzsúfolt szöveg: alig 40 sorban 25 jelző van. Sajnos, pillanatnyilag nincsenek számításaink arra nézve, mennyi szokott a jelző átlagos frekvenciája lenni, de ha jelzőn csupán a melléknévi jelzőt értjük, akkor nyilvánvaló, hogy stílusnemek, valamint közlési formák szerint igen nagyok lehetnek az eltérések. Mindenesetre Mallarmé 25 jelzőjéből a Helen Eaton-féle statisztika szerint az első kategóriákba esik 19; csak 6 került távolabbi kategóriákba, de még ezek is olyan közönséges szavak, mint például a 3b-be tartozó *tiède*, a 4a-ba *ivre*, az 5a-ban levő *beau* és mások. Mit lehet ebből következtetni? Azt, hogy nyilvánvalóan még Mallarménál is, aki pedig a legkülönösebb nyelvhasználatú francia költők egyike, a szókészlet e l e m e i legnagyobbrészt — persze százalékosan nem merek semmit sem mondani — a nyelv a l a p s z ó k i n c s é b ő l származnak, viszont a költői hatás elsősorban a s z a v a k k o m b i n á c i ó j á n, tehát sajátos költői felhasználásán alapul.

4. Végül szeretnék vitába szállni az egyik elhangzott előadásnak azzal a megállapításával, hogy a közönséges szavak kevésbé kifejező erejűek. Egy ilyenféle mondat hangzott el: „Kinek tűnt fel az, hogy Adynál milyen sűrűn fordul elő a *nap* vagy a *nagy* szó?" Stilisztikai szempontból e szavakkal kapcsolatban is minden a kontextuson, a szövegösszefüggésen múlik. Ezt talán nem is kellene magyarázni, de azért hadd idézzek legalább két konkrét példát, éppen Adyból. A *nap* szónak expresszív használatát jól szemlélteti a következő szöveg: „Szerelem, óh, uzsorás kölcsön, / Fölfölesdeklő *napjaimnak* / Rosszkor érkezett drága rossza, / Mégis — mégis csak visszahozza / Nagy fizetségeit a kínnak / Egy-egy boldog, hitető *napod*." Ezt írta Ady 1918-ban; sokkal hamarább, 1909-ben a *nagy* szónak igen szép használatára nyújtott példát: „Áldassanak bennünk A kifeslett vérrózsák, Bánat, kín, szenvedés, Mert Jézus volt a jóság S a *nagy*, szent türelem." Min múlik ennek az utolsó szónak költőisége? Egyszerűen azon, — de ez, azt hiszem, messze túlesik a nyelvstatisztikának a jelenlegi megbeszélésünk körébe tartozó problémáin —, hogy a *türelem* szónak sem együtt, sem külön nemigen szokott a *nagy* és a *szent* jelzője versben efféle, ünnepélyes intonációt kapni. A szöveg legegyszerűbb elemei is egymást kölcsönösen erősítik, s így emelkedik végül egy különös szó, a költő szava, a szépség tisztultabb ormaira.

Vértes Edit

1. Némi irigységgel hallgattam Fónagy Iványék és Kozma professzor előadását. Magam is végeztem ugyanis — mint többször szóba is került — statisztikai számításokat; és nekünk a Központi Fizikai Kutató Intézet Akusztikai és Ultrahang Csoportjában ezeket a számolásokat betűről betűre, illetőleg hangról hangra fonetikus átértékelésben kellett elvégeznünk. Hallottuk, hogy a munkaidő hányad része alatt tudta volna mindezt egy gép elvégezni, még hozzá nemcsak hogy kevesebb munkabefektetéssel, de nyilván pontosabban is. Ezek a számlálások ugyanis rendkívül fárasztóak, könnyen belefáradtunk; és éppen ezért előfordult, hogy olyan eredmények is jöttek ki, amelyek számítás nélkül pontosabbak, eleve nyilvánvalóak voltak. Eredményként kaptuk például, hogy a megvizsgált szövegben k ö r ü l b e l ü l ugyanannyi szó kezdődött, mint fejeződött be. Előfordult ugyanis, hogy hol egy szó kezdetét, hol pedig egy szó befejezését felejtettük el följegyezni, és így szerencsére legalább még nagyságrendileg kaptunk azonos eredményt, de nem szám szerint ugyanannyit. Ilyen eltérés gépnél, ha csak valami hiba nem fordul elő, nem következhetik be.

Ugyanakkor azonban az irigylésre méltó technikai előnyök mellett bizonyos aggályaim is vannak a munka teljes gépesítésével szemben. Kétségtelen, hogy a nyelvész az ismertetett gépesítés esetében akkor fog a munkához, amikor a munka élvezetesebb része következik, a gépies rész már befejeződött. Nem kerül azonban a feldolgozó olyan közvetlen és intenzív kapcsolatba a szöveggel, mint az, aki maga végezte a számolást, és így talán nagyon sok, nem is jelentéktelen részlet elkerülheti a figyelmét.

2. Hogy valamely költő vagy író nyelvében mi az egyéni, illetőleg hogy mi a jellemző az egyéni stílusára, arra frekvenciaszámokkal csak akkor hivatkozhatunk, hogyha ezeket megbízható nyelvstatisztikai normákkal tudjuk összehasonlítani. Méghozzá nemcsak olyan normákkal, hogy egy hangnak ennyi vagy ennyi a nyelvi előfordulása, hanem az előfordulást hibahatárral, szórással együtt kell venni, és csak abban az esetben tarthatjuk az eltérést jellegzetes túltengésnek vagy hiánynak, ha hiba- és szóráshatár figyelembevételével ezen fölül vagy ezen alul esik egy frekvenciaszám. Ha viszont egy bizonyos szó gyakori előfordulása okozza ezt az eltérést, akkor ezt nem lehet fonéma-statisztikánál figyelembe venni, hanem már eleve ki kell ezt a szót a statisztikázásnál zárni. Itt két példát kreáltam, amelyek így önmagukban elég erőltetettek, de rávilágítanak arra, amit mondani szeretnék. Ha mondjuk a „Jó estét gyerekek" című műsorszámot dolgozná fel a gép statisztikusan, elhangozhatnék

egy mese, amely mesének a hősét „*Vörös Körömtő Töhötöm, földből jövő hős ördögkölyök*"-nek neveznék. Elképzelhető, hogy ennek az ördögkölyöknek a neve minden mondatban alanyként, tárgyként vagy határozóként előfordul. Köztudomású, hogy a magyar nyelvben az *ö* nem gyakori hang. Ebben az ötperces mesében rendkívül felemelné, teljesen meghamisítaná a statisztikát mind a *ö*-re, mind pedig az *ö*-kapcsolataira, ha ezt a nevet bennehagynák a vizsgálandó szövegben. Tehát nyelvileg használható statisztikát, ami a nyelvi valószínűséget tükrözi, csak úgy lehetne készíteni, hogyha ezt a nevet kihagyjuk belőle, legföljebb, mondjuk, egyszer lehetne figyelembe venni; azonban a nevet, mint az egész mese hangulatára jellemző elemet külön ki kellene emelni. Ugyanakkor, képzeljük, hogy egy szatírikus regénynek a hősét *Fekete-selyem-bekecses Elek*-nek hívják, ahol csupa *e* van egy hosszú névben. Nem valószínű, hogy egy regényben minden mondatban előfordulna a név. Azonkívül az *e* olyan gyakori hang, hogy ha az illető regényhős nevében még átlagon felüli gyakoriságú *e* szerepel is, például tíz darab *e*, az *e*-k arányszámát semmiképp sem ugratná egy néhány száz lapos regényben ez a név annyira ki, mint az előbb említett öt perces mesében az *ö* hangokét a csupa *ö* betűs név. Ha tehát itt nem hagyják ki ezt a nevet, ez nem hamisítja meg a nyelvi statisztikai képet.

3. A nyelvi statisztika kiugró értékeit[1] előidéző nyelvi okok felkutatására Tarnóczynak Adyra vonatkozó és saját magamnak Veres Péter nyelvéről készített statisztikáim egybevetésekor (az összehasonlító táblázatokhoz írt jegyzeteimben) tettem kísérletet.[2] Ezek közül az okok közül egyet, mely stiláris-expresszív kifejező eszközök nyelvstatisztikai vizsgálatában módszertani szempontból nem látszik érdektelennek, most részletesebben kifejtek, mint annak idején tettem.

Az *m* hang statisztikus gyakorisága Ady nyelvében teljesen megfelel a nyelvi normáknak, mert Tarnóczy szerint 33 218 hangnak a 4,1 százaléka *m;* a Tolnai, Mikes, Nemes, Tarnóczy és saját vizsgálataimon alapuló, eddigi ismereteink szerint legmegbízhatóbbnak tekinthető nyelvi norma pedig 4,12 százalék. Ha viszont a szókezdő vagy a szóvégi hangok eloszlásában vetjük egybe az *m* százalékarányát az Ady és Veres Péter nyelvére vonatkozó vizsgálatokban, ezek a százalékarányok már nem felelnek meg egymásnak a valószínűségszámítási hibahatáron belül. A szókezdetre vonatkozó százalékarányok még elég közel állnak egymáshoz, Adynál ugyanis a szókezdetek 11,65 százaléka, Veresnél pedig 12,55 százaléka *m*. Nagyobb az eltérés a szóvégen, Adynál 7,33 százalék, Veresnél pedig csak 3,79 százalék az *m*, ez utóbbi eltérést az *-m* egyes szám első személyű igei és birtokos személyrag gyakorisága Ady verseiben indokolja. Ha viszont az *m* hang szókezdő szóközépi és szóvégi előfordulási arányát, továbbá a szóközépi *m* kapcsolódását az előtte, illetőleg utána levő magán- és mássalhangzókhoz vizsgáljuk, a százalékos eltérések olyan nagyok, hogy az eltérések indokolása nélkül a nyert eredmények szinte a statisztikus vizsgálat csődjét látszanak alátámasztani, Tarnóczy 4290[3] *m* hangot, azaz az *m* 5443 előfordulási helyzetét vizsgálta meg ilyen szempontból. Vizsgálódásainak eredményeit, százalékra átszámítva, egybevetettem Veres Péter nyelvére vonatkozó 1208 *m* hang

[1] Az itt következő pont a felszólalásban — előzetes felkészülés hiányában — nem hangzott el ilyen részletesen; a következőkre vö. Adalékok a magyar nyelv hangtani szerkezetéhez: Nyelvtudományi Közlemények 53. 107—52 (1952), 54. 96—140 (1953), 55. 138—180 (1954).

[2] Nyelvtudományi Közlemények 54. 124—140 (1953), 55. 142—180 (1954).

[3] A 4754-es szám (a Nyelvtudományi Közlemények 54. 125. lapján) sajtóhiba.

alapján kapott eredményeimmel. A százalékarányok olyan nagyfokú eltéréseket mutattak, hogy feltétlenül szükségesnek látszott az Adyra vonatkozó számításokat úgy revideálnom, hogy az eltérést okozó, gyanúba vett egyes szám első személyű birtokos és igei személyragokat külön statisztikáztam. Az Ady kötetnek a Tarnóczy munkacsoportja által feldolgozott lapjain 5487 *m*-et találtam, ebből 2179 (39,7 százalék) szókezdő, 1445 (26,3 százalék) szóközépi és 1863 (34 százalék) szóvégi helyzetben. Ezek a számarányok a manuális számításkor feltétlenül adódó, kifáradásból és figyelmetlenségből származó hibahatáron belül teljesen egyeznek a Tarnóczy adataiból kiszámított arányokkal, mely szerint Adynál

m- : *-m-* : *-m* = 39,1 : 26,9 : 34, tehát ≈ 39,7 : 26,3 : 34.

A Veresnél kapott arányok azonban teljesen mások:

m- : *-m-* : *-m* = 53,2 : 31 : 15,8,

illetőleg a 658 adat alapján készített kontrollnál 56,7 : 26,3 : 17. Az általam készített statisztika szerint Adynál 1568 *m*, azaz az adatok 28,6 százaléka első személyű személyrag, s ebből is 1350, azaz 24,6 százalék szóvégi helyzetben fordul elő, s csak 218 *m*, azaz az összes adatok 4 százaléka a szóközépi helyzetben megszámolt első személyű birtokos személyrag. Első személyű *-m* személyrag — szóvégi és szóközépi helyzetben egyaránt — előfordul természetesen Veres Péternél is; hiba lenne tehát az Adynál megszámlált valamennyi első személyű személyragot egyéni jellegzetességnek tekinteni, s ezeket elhagyva a többi *m* arányát a Veres Péternél kapott arányokkal egybevetni. Megpróbáltam azért olyan „korrekciót" alkalmazni, hogy az Adynál és Veres Péter művéből nyert számarányoknál a szóvégi *m*-ek százalékszáma körülbelül azonos legyen. Ezért arra gondoltam, hogy a szóvégi *m* hangokból 1207-et, azaz az összes *m* hangok 22 százalékát Ady egocentrikus lírai költészete számszerű jellemzőjének tekintem, s ha az ezek elhagyása után maradó *m* hangok aránya nagyjából megfelel a Veres prózájában talált aránynak, akkor a

m

	Szókezdő			Közbülső			Végső			Összes		
	Ady		Veres	Ady		Veres	Ady		Veres	Ady		Veres
Adatszám	2179	2179	685	1445	1445	318	1863	656	205	5487	4280	1208
Magánhangzó + + *m*	X	X	X	20,4	25,6	23,8	34,0	15,3	15,8	54,4	40,9	39,6
Mássalhangzó + + *m*	X	X	X	6,5	8,1	7,2	—	—	—	6,5	8,1	7,2
m + magán-hangzó	39,1	50,9	53,2	19,8	24,8	22,0	X	X	X	58,9	75,7	75,2
m + mással-hangzó	—	—	—	7,1	8,9	9,0	X	X	X	7,1	8,9	9,0
Összes helyzet	39,1	50,9	53,2	53,8	67,4	62,0	34,0	15,3	15,8	126,7	133,7	131,0
Összes hang	39,1	50,9	53,2	26,9	33,7	31,0	34,0	15,3	15,8	100,0	100,0	100,0

nagy eltérés okozóját sikerült kielemeznem. Ha a 2179 szókezdő és 1445 szóközépi *m* mellett csak 1863—1207 = 656 szóvégi *m*-et veszünk tekintetbe, akkor az 5487—1207 = 4280 *m* megoszlási aránya

$$m\text{-}:\text{-}m\text{-}:\text{-}m = 50{,}9 : 33{,}7 : 15{,}3$$

s ez már igen jól megfelel a Veres Péternél talált arányoknak. Ha a szóközépi *m*-eknek az előtte álló magán-, illetőleg mássalhangzóhoz való kapcsolódási arányát (20,4 : 6,5) és az utána álló magán- és mássalhangzóhoz való kapcsolódását (19,8 : 7,1) arányosan átszámítjuk úgy, hogy az arányszámok összege 26,9 helyett 33,7 legyen, akkor az Ady nyelvére vonatkozó eredmények — a stílusjellemzőnek tekintett szóvégi *m*-ek elhagyása után — szinte azonosak a Veres Péter prózájára vonatkozó eredményekkel.

Hasonlóképpen az *m* hang kapcsolatai — a szóeleji, szóközépi és szóvégi helyzet figyelembevétele nélkül — az előtte, illetőleg az utána álló hanggal jól egyezik Adynál és Veres Péternél, ha Ady adatainak 22 százalékát, — mindet szóvégi helyzetben, mint stílusjellemző elemet, nem vesszük a nyelvre jellemző arányok megállapításakor tekintetbe.

	Ady		Veres
Adatok száma	5487	4280	1208
Magánhangzó + *m*	54,4	41,5	39,6
Mássalhangzó + *m*	6,5	8,3	7,2
ø + *m* (szókezdet)	39,1	50,2	53,2

	Ady		Veres
Adatok száma	5487	4280	1208
m + magánhangzó	58,9	75,5	75,2
m + mássalhangzó	7,1	9,1	9,0
m + ø (szóvég)	34,0	15,4	15,8

Azt hiszem, hasonló statisztikai vizsgálódási eljárásokkal ki lehetne egy-egy költő vagy író nyelvének egyik vagy másik sajátosságát mutatni, ha van elég megbízható „nyelvi norma", mellyel egybe lehet vetni. E nyelvi norma megállapítása gépek segítségével viszont már nem látszik elérhetetlen vágyálomnak.

4. Szokás a statisztikáknál arra hivatkozni, hogy ha egy költőnél sok a magas hang, akkor világosabb, derűs hangulatú, ha sok a mély hang, akkor sötétebb tónusú a vers. Hirtelen ellenpéldaként csak két közismert verssorra gondoltam.

„Sűrű sötét az éj,
dühöng a déli szél".

A két névelőn kívül csupa magas hang van benne, tehát 83,3 százalékban magas hang, és csak 16,6 százalékban mély hang. Azt hiszem, nincs olyan magyarul tudó ember,

aki „üde, derűs, bájos" vagy más hasonló jelzővel illetné ezt a két verssort. Vannak tehát speciális esetek, amelyeket nem lehet statisztikusan elbírálni, hanem feltétlenül ki kell emelni mint statisztikus vizsgálatra alkalmatlanokat. Természetesen, ha Aranynak az egész életművét vizsgáljuk statisztikusan, biztos vagyok benne, hogy ez a két sor nem befolyásolja a magas és mély hangok arányát. Ebben az adott esetben a sötét hangulatot nem a mély hangok hordozzák, ehhez kétség nem fér.

5. Az utóbbi időben sokszor kérdezték meg a matematikai vonatkozású előadások hozzászólói, hogy hol kezdődnek a nagy számok, honnan érvényes a nagy számok törvénye, és meddig nem. Erre megpróbáltam a választ megadni, Veres Péter nyelvezetéről írt statisztikai összeállításomban; ez a matematikai képletek között szerepelt (azt hiszem, a nyelvészek ezt át szokták lapozni). Így elkerülte a figyelmüket, hogy minden egyes problémánál külön-külön kell megállapítani, hogy adott eset a nagy számok törvényével mérhető-e, vagy sem. Minthogy például majdnem 10 százalék az *a* hangok aránya a magyar fonéma- vagy hangkészletben, így ha megszámlálunk egy körülbelül 10 000 hangból álló szöveget, 1000 körül kapunk *a* hangot. Ha egy bizonyos szöveget megnéztünk, és abban 990 vagy 1000 *a* hangot találtunk, nagy bizonyossággal állíthatjuk azt, hogy az a szöveg körülbelül 10 000 hangot tartalmaz. Ugyanakkor, ha egy 10 000 hangból álló szöveget megszámlálunk, a *zs* hangoknak az eddigi statisztikai eredmények szerint 6, *c*-nek 22, *ty*-nek 10-nek kell benne lennie. Ha viszont úgy nézzük, hogy elkezdünk olvasni, ha 6 *zs*-t találtunk, semmi esetre sem mondhatjuk, hogy körülbelül 10 000 hangnyi szöveget olvastunk el. Mert, ha például Adynak a Móricz Zsigmondhoz és Zsókához írt versét olvassuk el, már többet találunk benne mint 6 *zs*-t, és a két vers távolról sem tartalmaz 10 000 hangot.[4] Tehát, ha a nagyból indulunk ki, akkor a ritka hangokra megkapjuk a statisztikus átlagot, de a kevés hangból nem következtethetünk arra, hogy mennyit néztünk át.

6. Papp Ferenc előadásában szerepelt az, hogy „*a fiú meg a lány apja*" mondat két síkon értelmezhető, vagy hogy a fiúnak meg a lánynak, azaz mindkettőjüknek az apja (ezt ő lefelé húzott vonallal két dimenzióban ábrázolta), vagy pedig hogy csak a lánynak az apja. Úgy vélem, az értelmezés nyugodtan elvégezhető, ha matematikailag akarunk gondolkodni, egy dimenzióban: ha mind a kettőnek az apja, akkor a fiú meg a lány zárójelben lesz: *(a fiú meg a lány) apja ;* azaz a zárójel azt jelenti, hogy a kettőt előbb adjuk össze, és aztán kettőjükre vonatkozik, hogy az *apja*. Ha viszont azt írom, hogy „*a fiú meg a lány apja*", azaz, hogy nincs zárójellel jelezve, hogy az apa kettőjükre vonatkozik, akkor csak a lány apjáról van szó, meg egy fiúról. A zárójel ugyanis a matematikában, mint tudjuk, a műveletek sorrendjére vonatkozik.

7. Egy kérdésem volna Soltész Katalinhoz.* Guiraud képletei között van egy a francia nyelvre vonatkoztatott képlet, melyben a szókészlet nagysága $2n$-nel van jelölve, minthogy a fele formaszó, a fele pedig tartalmas szó. Tekintettel arra, hogy ez a megoszlás statisztikailag értendő, tehát a fele nem azt jelenti, hogy 50—50 százalék,

[4] A „Levél-féle Móricz Zsigmondhoz" 11, a „Zsóka búcsúzója" 9 *zs* hangot tartalmaz; ezek közül 9 a *Zsigmond, Zsiga*, illetőleg 8 a *Zsóka* névben fordul elő. A további 2, illetőleg 1 *zs* a *rózsafa, zsivajja*, illetőleg *varázzsal* szavakban fordul elő, azaz nyilvánvalóan ritkán használt szavakban.

* Vö. „A szerkesztők bevezetésé"-nek **5.** pontjával.

hanem épp úgy lehet 49—51 százalék, 48—52 százalék, ezek még mindig eleget tesznek a feltételnek. Semmi esetre sem érzem helyénvalónak, hogy olyan képlettel operáljunk, ahol a képlet azt mutatja, hogy a szókészlet nagysága p á r o s s z á m. A képlet ugyanis matematikailag azt jelenti, hogy a formaszavak és tartalmas szavak száma összegének feltétlen páros számnak kell lennie, ami pedig nyelvileg abszurdum. Ugyanakkor egy másik képletben, mely az előbbi következménye, a $2n$ a nevezőben négyzetgyök alatt szerepel, ahol meglehetősen kényelmetlen a négyzetgyök alatt egy kettes. Fölöslegesen bonyolítja tehát a kettes a képleteket. Ha már minden áron hangsúlyozni akarjuk a két részletet, akkor inkább összeggel $(n_1 + n_2)$ kellene ezt kifejeznünk.

Ehhez kapcsolódik egy az előadás során felírt francia példára vonatkozó megjegyzésem. Tekintettel arra, hogy egy mondatban — különösen a franciában — rendkívül gyakoriak a formaszók, 50 százalékon felüliek, így természetes és a szórendi követelményből is adódik, hogy a szavak egy része az ezres frekvencián felül van, minthogy ezek formaszavak. Érdekes volna tudni, hogy vajon a formaszavak hol helyezkednek el a statisztikában. Valószínűnek tartom, hogy mind ezer felett, és alig lehet ezer felett az egyéb szavak számára hely. Ez tehát éppen ezért jelenti azt, hogy a tartalmas szavaknak feltétlenül magasabb a rendszámuk, mint a minden mondatban szükségképp szereplő formaszavaknak.

8. Végül még egy megjegyzésem volna Petőfi S. János hozzászólásához, kitől meglepődve hallottam, hogy nekem egy cikkemben topológiai gondolat szerepelt nyelvészeti síkra vonatkoztatva. Tekintettel arra, hogy Kerékjártó Béla tanítványa vagyok, ki úgy gondolom, nemcsak magyar, hanem nemzetközi viszonylatban is topológiából feltétlenül a vezetők közé tartozott, és minthogy én nála topológiából és csoportelméletből dolgoztam, úgy képzelem, hogyha nem is alkottam a topológiában, de annyit talán értek topológiához, hogy tudjam, mi a topológia. Nagyon sokat gondolkodtam rajta, sőt Tarnóczyval sokat vitatkoztunk is rajta, hogy lehetne a topológiát nyelvészeti gondolkodásban használni, tudomásom szerint azonban nekem erre soha sehol sem sikerült semmi utat és módot találnom. Egy hasonlat alkalmazása még nem nevezhető a topológia nyelvészeti felhasználásának.

Mikó Pálné

1. A szakirodalmi szövegek (ideológiai, műszaki, természettudományi, filológiai stb. szövegek) aránya a megjelenő könyv- és folyóirat-kiadványok egészében. Az úgynevezett „közepes nehézségű" szakszöveg; a meghatározás lazasága, az ebből eredő hátrányok az idegen nyelvek tanításában.

2. A nyelvi tények jellegzetességei komplex rendszerükben és kölcsönös viszonyukban. A konkrét, egzakt adatok kvalitatív és kvantitatív vizsgálata; csak ez lehet a pontosabb szabályok és törvényszerűségek megformálásának forrása.

3. Véges számú nyelvi elemnek végtelen számú lehetséges kombinációja. Valamennyi jelenség megvizsgálásának korlátai. A „szegmentumok" (vagyis szöveg-típusok) elemzésének lehetőségei és előnyei. A valószínűségszámítás felhasználásáról; az általános törvényszerűségek és az esetleges (egyedi) jelenségek viszonyáról, megoszlásáról.

4. Meghatározott mennyiségű nyelvi „nyersanyag" viszonylagos értéke. A szakirodalmi szövegek közvetlen megfigyelése: a szövegek homogeneitásának és semleges voltának problémája. A lényeges és lényegtelen tényezők szétválasztása a szövegstatisztikában.

5. Nyelvstatisztikai törvényszerűségek és a nyelvelemek gyakorisága. A gyakoriság mérése; összehasonlító módszerek és eljárások. A hibák százaléka; a nyelvstatisztikai vizsgálatok abszolút és viszonylagos pontosságának kérdése.

6. Bemutatásra kerülnek a vizsgált adatok nyelvi szintek szerint: a mondatok hossza; a szó-osztályok szerinti megoszlás, illetőleg a szóállomány adatai; a szavak hossza (a toldalékolás jellegzetességei és a morfémák megoszlása); a szintaktikai tagolás (a közvetlen összetevők és a transzformáció módszere szerint); egyéb tényezők.

* Ez a kivonat a szerző — előadás méretű — hozzászólása helyett áll itt. Az előadás (Nem szépprózai szövegek vizsgálatának néhány mennyiségi mutatójáról) a program lezárása után került bejelentésre, csak részben kerülhetett bemutatásra. Teljes egészében máshol jelenik meg. — *A szerkesztők.*

Papp László

Azt a statisztikai vizsgálatot, amelyről a „Vita a nyelvtudomány elvi kérdéseiről" című konferencián röviden hírt adtam, 400 különböző, összesen 230 000 szótag terjedelmű szövegen elvégeztem.[1] A munka során bennem felmerült és csak matematikusok által megoldható problémákat érintettem egy könyvemben[2] ugyan a feldolgozásban, de megoldásra nem jutottam, bár matematikus-statisztikussal[3] is konzultáltam ez ügyben. — Engedjék meg tehát, hogy itt és most előadjam kérdéseimet.

Módszerem a „type—token" viszony[4] vizsgálatának némileg módosított változata. Nem szókészleti tényeket, hanem bizonyos hangtani jelenségeket vizsgáltam; említett munkámban a zárt *í*-zést.

Dialektológiai munkálatainkban rendkívül gyümölcsözőnek bizonyult az az eljárás, amely a hangtani jelenségeket a normatív nyelvváltozathoz viszonyítva, szembenállásonként, példák és ellenpéldák egybevetésével méri fel és ábrázolja. Tehát például valamely „corpus" — én ugyan szívesebben használom a „vizsgálati egység" terminust —, valamely vizsgálati egység zárt *í*-zését, labiális *ö*-zését, nyílt *ë*-zését stb. úgy írjuk le, hogy a köznyelvi *é*-hez, *ë (e)*-hez, *i*-hez viszonyítva példának tekintjük az *í*-ző, *ö*-ző, *ë*-ző morfémát *(szíp, embör, vëlla)*, ellenpéldának a köznyelvivel azonos hangú morfémát *(szép, embër* vagy *ember, villa)*. Ha felmérjük számszerűen egy vizsgálati egység (ez lehet egy tetszőleges terjedelmű szöveg, de lehet egy író életműve vagy egy nyelvjárás is) bármely hangtani jelenségének adatait, a példák és ellenpéldák alapján pozitív és negatív értékeket kapunk. E két szám abszolút értékének összege adja a kérdéses jelenség előfordulási számát az illető „corpus"-ban. Például: egy szövegben a zárt *í*-zés lehető eseteinek száma 300; ebből ténylegesen *í*-ző 100, *é*-ző 200. Az arányt százalékokban kifejezve: + (példa) 33,33 százalék; — (ellenpélda) 66,66 százalék.

Ez azonban csak általános első tájékoztatásra elegendő. Vizsgálnunk kell azt is, hogy az abszolút számú előfordulásra hány különböző jel esik. (Jelen — ebben a

[1] A Magyar Tudományos Akadémia Nyelv- és Irodalomtudományi Osztályának Közleményei 18. 85—7 (1961).

[2] Papp László: Nyelvjárástörténet és nyelvi statisztika (Bp., 1963). — Vö. még: Papp László, Application de la statistique linguistique aux recherches de dialectologie historique: Acta Linguistica 12. 67—94 (1962).

[3] Dr. Szalay Sándorral (az MTA Közgazdaságtudományi Intézete).

[4] A „type—token" viszonyra vonatkozóan vö. Antal László: Magyar Nyelvőr 83. 307—11 (1959).

vonatkozásban — m o r f é m á t értek.) Ha ezt a vizsgálatot is elvégezzük, a következő összefüggéseket tapasztaljuk.

A jelenség körébe vonható morfémák száma csak kivételes esetben, igen rövid szövegekben lehet azonos az előfordulási számmal, általában kisebb, és mennél terjedelmesebb a vizsgálati egység, annál nagyobb a különbség a két érték között. Ez természetesen magától értetődik. — Ha megállapítottuk, hogy hány különböző jel (morféma) jelentkezése adja meg az abszolút előfordulást, a következő lépésként meg kell állapítanunk azt is, hogy ezek a jelek hogyan oszlanak meg a példák és ellenpéldák között. Itt már nemcsak pozitív és negatív értékeket kapunk, hanem $\{\pm\}$ értékeket is, minthogy ugyanaz a jel előfordulhat a példák és az ellenpéldák között is. Ha tehát az előbbi elméleti eset 300 adatát — mondjuk — 50 különböző jel adja, ez 50 jel közül *i*-ző lehet 20 (= 40 százalék), *é*-ző 20 (= 40 százalék) és $é \sim i$ váltakozású lehet 10 (= 20 százalék).

Minden vizsgálati egység bármely jelenségét, amelyet példák és ellenpéldák alapján vizsgálhatunk, öt abszolút számmal és a belőlük adódó két összértékkel vagy — talán még pontosabban — öt százalékszámmal jellemezhetjük:

$$+A \text{ és } -B = C;$$

$$+D \text{ és } -E \text{ és } \pm F = G;$$

illetőleg:

$$+m \text{ és } -m;$$

$$+c \text{ és } -c \text{ és } \pm c.$$

(A verzális betűk számlálással nyert értékek szimbólumai, a kurrens betűk százalékszámokat jelölnek.)

Nyilvánvaló, hogy ennyi számértéket jellemzésre felhasználni túlságosan nehézkes. Vagy ezek között kell keresnünk egy olyat, amely a leghívebben vagy a viszonylag leghívebben jellemezheti a vizsgálati egység szóban forgó hangtani jelenségét, vagy pedig ezek között az elemek között kell — ha lehetséges — megállapítani és megformulázni olyan összefüggést, hogy a kapott számérték — egy konstans — legyen jellemző mutatószáma a kérdéses jelenségnek. — Én az említett vizsgálat során úgy találtam, hogy a számlálással és számítással kapott értékek közül leghívebb mutató a $-c$ értéke (a jelekre vonatkoztatott statisztika ellenpéldáinak, csak *é*-ző alakban előforduló morfémáinak százalékszáma). Ezzel az értékkel fordított arányban van az illető szöveg *i*-zésének (vagy bármely más hangtani jelenségnek) mértéke.

Így hozzájutottam tehát egy objektív mutatószámhoz, de rendkívül hosszadalmas számlálás és fáradságos számítás után. Azt is el kell ismernem, hogy ez a $-c$ érték csak viszonylag a leghívebb kifejezője a vizsgálati egység kérdéses jelenségének, a többi százalékszám is, sőt az abszolút számok is fontosak.

Az tehát a kérésem a matematikusokhoz, hogy segítsenek megformulázni egyértelműbben és jellemzőbben a számlálással nyert értékek közötti összefüggést. Ha lehet ezek között az értékek között képlettel kifejezhető reális összefüggést találni, akkor a „corpus"-ok jellemző hangtani jelenségeit számokkal fejezhetjük ki; ezek a számok objektívek lennének, egymással összevethetők lennének; a hosszúpályi nyelvjárás hangtani állapotát számszerűen össze lehetne vetni a békési nyelvjárásával, Pázmány nyelvhasználatának hangtani állapotát Telegdiével, Ady fiatalkori munkáinak hang-

tani állapotát későbbi munkáiával; stb. — Képzeljünk el egy nyelvjárási hangtani leírást, amelynek összefoglalása egy ilyenféle táblázat:

$$K_{i} = a$$

$$K_{é} = b$$

$$K_{ë} = c$$

$$K_{ö} = d$$

$$K_{ü} = e \quad \text{stb.}$$

Most csak a $-c$-nek megfelelő százalékszám helyettesíti ezt a K állandót, ez azonban — újra hangsúlyozom — csak viszonylag a legjellemzőbb érték.[5]

„Nyelvi statisztikai vizsgálatok" címen 1962-ben összeállítottam egy körülbelül öt ívnyi tanulmányt. Ebben az *i*-zés vizsgálatának alapjául vett négyszáz különböző XVI. századi irat anyagán tárgyalok bizonyos hangtani-alaktani szembenállásokat, valamint a feltételes módú igealakok formai problémáit. Ez a dolgozatom — valószínűleg két részletben — a „Magyar Nyelvjárások" évkönyvben fog megjelenni.

Dezső László

A statisztikai módszerek egyaránt alkalmazhatók szinkrón és diakrón szempontú vizsgálatoknál, akár hagyományos, akár strukturális módszerekkel végezzük a vizsgálatokat. Az alábbiakban igen röviden a statisztikai módszerek olyan nyelvtörténeti alkalmazásáról szólok, amellyel az ukrán nyelvtörténeti anyag vizsgálatánál magunk is élünk.

A modern nyelvtudomány átfogó és ugyanakkor részletekben pontos ismereteket kíván nyerni nagyobb nyelvemlékekről és nyelvemlék-csoportokról. E cél érdekében fel kell használnia a statisztikai módszereket. Nyelvészeti és statisztikai szempontból egyaránt helyesen választott minták megbízhatóan tájékoztatnak bennünket nagyobb terjedelmű szövegek nyelvi sajátosságairól.

Az egyes nyelvfejlődési tendenciák vizsgálatában már régebb óta alkalmazták az egyszerűbb statisztikai módszereket (viszonyszámítás, egyszerű átlagok). Az egyes jelenségek fejlődéséről ugyanis csak akkor kaphatunk megbízható képet, ha a szóbajöhető adatokat pontosan felmérjük. A mennyiségi viszonyok pontos számbavétele nem jelentheti a vizsgálódási szempontok szegényítését. Például a keleti szláv *-a*-s végződések *(-am, -ami, -ach)* elterjedéséről a hím-, semleges- és nőnemű lágytövű ragozásban a keleti szláv nyelvterületen csak akkor kaphatunk majd pontos képet, ha az egyes emlékekben és emlékcsoportokban számba veszik majd az *-a*-s végződések és a velük konkurrens végződések elterjedtségét, szem előtt tartva az egyes nyelvjárások és az irodalmi nyelv sajátosságait.

A statisztikai módszerek mindenütt alkalmazandók, ahol a jelenségek mennyiségi vonatkozásai szerepet játszanak. Különös fontosságot kapnak azonban ott, ahol valamely jelenség magyarázatára számos és előre pontosabban meg nem határozható erejű összetevő tételezhető fel. Ilyenkor a vizsgált jelenség statisztikai jellegű. Többékevésbé ilyenek például a morfológiai jelenségek: új végződések megjelenése és elterjedése.

A modern nyelvszemlélet lényeges jellemző sajátossága az, hogy nem elégszik meg az egyes jelenségek meglétének konstatálásával, hanem annak pontos mennyiségi vonatkozásait is fel akarja mérni. Az egyes fejlődési tendenciák mennyiségi vonatkozásairól már beszéltünk. Ezen kívül szükségünk van bizonyos alapvető gyakorisági adatokra: ilyenek az egyes szófajok, és azokon belül a történeti és műfaji szempontból fontos nyelvtani kategóriák elterjedtségének gyakorisága, az egyes mondatfajták (mellé- és alárendelő) és az alárendelő mondatfajták gyakorisági viszonyai, mert ezek jól jellemzik az emléket műfaja és kora szempontjából, sőt a szerzőre is jellemzők

lehetnek. A szókincs jellemzésére szolgálnak az egyes tárgykörökbe és történeti rétegekbe tartozó szavak, jövevényszavak számszerű adatai, valamint az úgynevezett szókincsgazdálkodás.

Az eddigiekből világos, hogy véleményünk szerint statisztikai módszereket alkalmazni kell a nyelvtörténeti vizsgálatok minden területén. A legegyszerűbb statisztikai jellemzőkön túlmenve ki kell kísérletezni bonyolultabb statisztikai vizsgálatok (például index- és trend-számítás) nyelvtörténeti alkalmazását.[1] Érdekes lenne ugyanis a különböző nyelvjárásterületeken írott emlékek alaktani és mondattani rendszerére jellemző komplex mutatószámok keresése, amelyek egymással összevethetők lennének. Ezenkívül az ugyanazon nyelvjárásban különböző időben keletkezett emlékek nyelvtörténeti fejlődését illusztráló sorok megállapítása és azok összevetése más terület nyelvfejlődését mutató adatokkal szintén hozzásegíthet a nyelvtörténet jobb ismeretéhez. Természetesen a statisztikai vizsgálat értékét nem annyira azok matematikai bonyolultsága, mint inkább nyelvészetileg helyes alkalmazása adja meg.

[1] Vö.: Köves Pál—Párniczky Gábor: Általános statisztika (Bp., 1960) 228—313.

Szépe György

1. Sajnálatos, hogy a matematikai-nyelvészeti vitában aránylag kevés matematikus vett részt. Ez arra mutat, hogy Magyarországon egyelőre még nagyon szűk a nyelvészek és matematikusok együttműködésének az alapja. Munkaértekezletünk megrendezésének éppen az lett volna egyik célja, hogy interdiszciplináris közeledést eredményezzen: a matematikusok kedvet kapjanak nyelvi kérdéseken való elmélkedésre, a nyelvészek pedig a matematikai módszerek további, eredményesebb tanulmányozására.

A matematikai nyelvészet nyilvánvalóan nem érdektelen sem a matematika, sem a lingvisztika számára. Remélhetőleg gazdagítja a matematika tudományát: elsősorban egy új alkalmazott matematikai területtel vagy ágazattal, de nem lehetetlen, hogy magát az alaptudományt is némely tekintetben. Ennek a gazdagításnak az adja meg a lehetőségét, hogy a matematikai nyelvészet fölvethet olyan kérdéseket, amelyek eddig nem fordultak elő a matematika történetében. De természetesen bizonyos eredmények — főként az algebrai nyelvészet területén — egyszere számíthatók eredménynek a matematika és a nyelvtudomány számára. Ilyen „kettős" eredmények már eddig is voltak.

A matematikai nyelvészetet azonban a nyelvészet részének, egyik korszerű és magas igényű részének tartom. (Mint ahogyan a „matematikai biológia", vagy a „matematikai közgazdaságtan" is a biológia és a közgazdaságtan része.) Természetesen a matematikai nyelvészethez alapos matematikai és logikai ismeretek is szükségesek a nyelvészeti ismeretek mellett.[1]

* Az elhangzott hozzászólás kibővített változata.

[1] A márciusi konferencia óta tisztult a matematikai nyelvészet fogalma. Főleg az 1962. szeptemberében Tihanyban a Bolyai János Matematikai Társulat által megrendezett „A matematika alapjai, matematikai gépek, és alkalmazásaik" című kollokvium matematikai nyelvészeti és gépifordítási szekciójának (vö. Magyar Nyelvőr 87. 157—8 (1963)) tapasztalatai járultak ehhez hozzá. — Most (1962 legvégén) úgy látom, hogy a nyelvi szerkezet leírásában és magyarázatában két szinten lehetséges a matematika felhasználása. Az első, az „alacsonyabb" szint: egy konkrét természetes nyelv grammatikájának elkészítésében felhasználhatók a matematika (és a modern logika) számos ágazatának eredményei mint modellek. A másik, a „magasabb" szint: a különböző grammatikai modellek közül (általában, és a természetes nyelvekre vonatkozóan különösen) csupán a matematika segítségével lehet teoretikus értékű döntés útján a jobb (az aránylag „erősebb") modellt, koncepciót kiválasztani. Természetesen a döntés előkészítése és bizonyítása nagymértékű hozzájárulást jelent a grammatikai metateória kialakításához.

2. A *modell* az ismeretelméletnek és a tudományos kutatás módszertanának egyik újonnan központi helyre került kategóriája.[2] A nyelvészeti (pontosabban: a matematikai nyelvészeti) modelleknek a Papp Ferenc által adott kettéválasztása strukturális (kvalitatív) és statisztikai (kvantitatív) modellekre — sokaktól elfogadott és első megközelítésként megfelelő is. Továbbra is fennáll a nyelvi szerkezet egységes modelljének megkonstruálására vonatkozó igény; Papp Ferenc felfogása szerint ez nyilván elsősorban a strukturális és statisztikai modellek összehangolásának formájában végezhető el. — Természetesen a strukturális és statisztikai modellekre való szétbontás elvi alapjainak vizsgálata szintén feladat marad.

2.1. A Papp Ferenc által táblázatba[3] rendezett „kis modellek" a fonéma, a morféma, a szó és a szintagma szintjén jellemzik a nyelv szerkezetét. Az egyes szintek közötti kapcsolat nem került kifejtésre. A szintek bemutatásának sorrendje azonban *analitikus* és *induktív*[4] eljárásmódra utal, amely a kisebb nyelvi egységekből (illetőleg variánsokból) elemzi ki a közvetlenül „fölötte" álló szint nagyobb egységét (illetőleg konstánsát). A másik eljárás viszont *szintétikus* és *deduktív*; ez a modell a mondatot veszi az első alapfogalomnak, s abból vezeti le az alacsonyabb szintek egységeit, kategóriáit.[4]

Ez az „induktív" ~ „deduktív", illetőleg „szintétikus" ~ „analitikus" dilemma a statikus nyelvészet körén belül marad. A dinamikusabb, generatív grammatikai koncepció elsősorban nem az egyes nyelvi szintek leírására, hanem a mondatok szerkezetének minden szinten történő magyarázatára helyezi a hangsúlyt.

A generatív grammatika Chomsky-féle második modellje[5] — amely lényegében a közvetlen összetevők módszerének, vagyis a mondatelemzési eljárásnak generatívvá fejlesztett változata — a Papp Ferenc által fölvázolt valamennyi szintet átfogó dinamikus modell. Bebizonyosodott azonban, hogy még ez a modell is elégtelen: nem képes valamennyi mondat megalkotására szabályokat adni.

A generatív grammatika transzformációs modellje[6] azonban pótolja ezt a hiányosságot. A transzformációs szabályok (a szintaktikai, illetőleg a morfofonematikus transzformációs szabályok) segítségével — legalább is elvileg — valamennyi mondat szerkezete megalkotható. A transzformációs szabályok azonban nem sorolhatók be a hivatkozott összefoglaló táblázat egyik szintjébe sem, hiszen ide csupán nyelvi egységek (illetőleg a rájuk vonatkozó statikus rendszerezések, konfigurációk megalkotása, esetleg kombinatorikus csoportosítása) férhetnek bele. Nincs más megoldás, mint a statikus modellnek olyan továbbfejlesztése, ahol mind a szintaktikai transz-

[2] A modellek kérdésében az aktuális nyelvészeti nézeteket (és főleg szakirodalmat) lásd: A generatív grammatika transzformációs modelljéről: Általános nyelvészeti tanulmányok (Bp, 1963) 280—1.

[3] Lásd: Kötetünkben a 77. lapon.

[4] Vö. Alphonse Juilland: Outline of a general theory of structural relations (Hága, 1961) 10; vö. még: W. Haas: Word 16. 259 (1960). — A terminusok szokatlan használata Juilland művéből származik.

[5] N. Chomsky: Syntatic structures (Hága, 1957); lásd a „Phrase Structure" (26—33) és a „Limitations of Phrase Structure Description" (34—48) című fejezeteket. — Magyarul vö. a 2. jegyzetben hivatkozott tanulmányom 3. fejezetével; „Három generatív grammatikai modell" (280—8).

[6] Vö. N. Chomsky, Three Models for the Description of Language: I. R. E. Transactions on Information Theory IT—2. 113—24 (1956); valamint az 5. jegyzet hivatkozásaival.

formációk, mind pedig a morfofonematikus transzformációk saját nyelvi szintet kapnak.[7] Ez azonban már nem statikus, hanem dinamikus modell.

2.2. Papp Ferenc nemcsak a strukturális és statisztikai módszerű modellek kettéválasztását végezte el. A modelleket aszerint is osztályozta, hogy a nyelvi rendszerre (a saussure-i értelemben vett paradigmatikus tengelyre, vagy az információelmélet „code" kategóriájára), illetőleg a közlési folyamatra (a saussure-i értelemben vett szintagmatikus tengelyre, vagy az információelmélet „message" kategóriájára) vonatkoznak. Így a következő négyféle lehetőséget kapjuk valamennyi szinten:

a nyelvi rendszer statisztikai modellje [R1]	a nyelvi rendszer strukturális modellje [R2]
a közlési folyamat statisztikai modellje [K1]	a közlési folyamat strukturális modellje [K2]

Ebben a zárt diagrammban a lehetőségek teljesen szimmetrikusak; ez föltétlenül előnyös, mivel így szisztematikusabb módon vizsgálhatjuk végig őket. A probléma természetesen az, hogy a nyelv szempontjából mit ér ez a megközelítés.

Véleményem szerint a nyelvben korántsincs olyan szimmetria, mint a táblázaton. Az utolsó évek tapasztalatai alapján úgy látom, hogy az [*R* 2] szempont sokkal termékenyebb, mint az [*R* 1]; viszont a [*K* 1] vizsgálata hasznosabb, mint a [*K* 2] vizsgálata. (Azt hiszem, ebben egyetértünk Papp Ferenccel.) Ez annyit jelent, hogy két modellt (természetesen dinamikus és valamennyi szintre kiterjedő modellt) tartok elméletileg érdekesnek: elsősorban a nyelvi rendszernek strukturális modelljét, másodsorban a közlési folyamatnak (és megvalósulásának: a szövegnek) statisztikai modelljét. Az elsőt szerintem megnyugtató módon megvalósítja a generatív grammatika transzformációs modellje. A másodikra a következő pontban térek ki.

3. A statisztikai, illetőleg bizonyos egyszerűbb „információelméleti" modellek[8] felhasználása terén némileg csökkenőben van az elmúlt évek optimizmusa. Ma már nagyon sok módszer, megközelítés ismeretes. Egyre jobban kialakul az is, hogy a nyelvészet milyen ágazatai számára, továbbá milyen nyelvészeten kívüli alkalmazások számára használhatók fel a szövegstatisztika, illetőleg a szövegen kívüli nyelvi elemek statisztikája (s a kétféle szempont kombinálása).[9] Ezek az eredmények azon-

[7] Vö. N. Chomsky: Syntactic Structures 46; R. B. Lees, Language 33. 388—91 (1957), „The anatomy of grammar" című rész.

[8] Itt nem kívánok különbséget tenni a „statisztikai' és az „információelméleti" módszerek között — jól tudva természetesen, hogy ezek korántsem azonosak. (Vö. az O. A. Ahmanova szerkesztésében megjelent: О точных методах исследования языка című kötettel (Moszkva, 1961).) — Papp Ferenc eljárásához kapcsolódva és csak a lokális egyszerűség miatt állítom ezeket a modelleket együttesen szembe a strukturális modellekkel. Ennek a fajta kettéválasztásnak a megalkotója alighanem Warren Plath, Mathematical linguistics: Trends in European and American linguistics 1930—1960 (Utrecht/Antwerp, 1961) 21—57.

[9] Vö. G. Herdan: Language as choice and chance (Groningen, 1956). Uő: Type-token mathematics. — Kitűnő magyar bevezető az információelmélet nyelvi felhasználására: Fónagy Iván, Beszéd és valószínűség: Magyar Nyelvőr 86. 309—20 (1962).

ban lingvisztikai szempontból egyelőre — legföljebb — periferiális jelentőségűek.[10] A nyelvi szerkezet centrális tudománya, a grammatika egyelőre még kevés hasznát látta a nyelvstatisztikának.

Ismeretesek a fonéma- (illetőleg betű-)statisztikák. Sőt ismeretes Harrisnek egy olyan elgondolása is, amely a morfémán belüli (fonematikus) szegmentumok következési valószínűsége alapján következtet a morfémahatárokra.[11] Azonban csak mostanában kezdődött meg a morfematikus alapú statisztikázás, amely a nyelvi rendszer, a grammatika szempontjából érdekes. A szintaktikai statisztika is a próbálkozások legelején tart;[12] még semmilyen jelentősebb eredményt nem hozott. A nyomtatott szöveg írott szavainak statisztikus feldolgozása aránylag előrehaladott; de eredményei nagyon nehezen hozhatók kapcsolatba a nyelvi rendszer mélyebb szabályaival.

A szövegnek, tehát bármilyen zárt írott vagy beszélt nyelvi produktumnak a statisztikai vizsgálata természetesen szükséges és hasznos. Ez a vizsgálat általában a jelenségek felületi vizsgálata; azonban megfelelő hipotézisek segítségével lehatolhat a lényeg szintjére is. — A vizsgálatok eredménye leginkább a szöveg valamilyen szempontú jellemzése lehet. A szövegben található bármilyen egység (betű, szó, mondat; fonéma, morféma, mondat; stb.) bármilyen helyzetben alkalmas arra, hogy megszámlálhassuk. Bizonyos nagyságrend után előre megjósolható az ugyanabban a szövegfajtában a már megszámláltakkal azonos típusú példányok előfordulásának valószínűsége. A különböző szövegeken belül más és más statisztikai jellemzőket találhatunk. Abban az esetben pedig, ha a szövegeknek (bármilyen szempontból megállapított) típusai és azoknak statisztikus jellemzői között rendszeres kapcsolat található, akkor a (meghatározott szempontból) vizsgált szövegtípust jellemeztük. Mivel az illető szövegtípus jellemzése az egyes konkrét szövegdarabok jellemzésének összegezése, általánosítása útján történik (és általában más konkurrens szövegtípusokkal szemben), ezért ezt a jellemzést m a k r o s z k o p i k u s nak nevezhetjük el.

Makroszkopikusan jellemezhető például a „költői nyelv" más közlési formákkal szemben. Ezen belül makroszkopikusan jellemezhető két költő nyelve egymáshoz viszonyítva; továbbá egy költő oeuvre-je fejlődésének különböző korszakában vagy műfajok szerinti csoportosításban.

De ugyanígy kereshetünk összehasonlítási szempontokat bármilyen (tehát nem költői) szövegek között a szövegek szerzőjének vagy mondanivalójának tárgykörök szerinti (szociológiai, pszichológiai, földrajzi és még sok másféle) eltérése alapján. Sőt, így vizsgálható kvantitatív módszerekkel az egyes nyelvek, nyelvtípusok, sőt esetleg nyelvcsaládok különbsége is — megfelelő, reprezentatív szövegek alapján. — A kvantitatív megközelítést azonban mindig megelőzi egy kvalitatív kérdésföltevés. S még így is csupán tájékoztató értékű vagy előmunkálat értékű a statisztikai vizsgálat a kvantitatív eredmények kvalitatív interpretálása nélkül. Igen örvendetes, hogy a magyarországi nyelvstatisztikai kísérletek ezt a szempontot erőteljesen érvényesítik.

[10] Vö. R. B. Lees (bíráló tanulmány), Apostel — B. Mandelbrot—A. Morf, „Logique, Langage et théorie de l'information": Language 35. 271—303 (1959).

[11] Vö. Z. S. Harris, From phoneme to morpheme: Language 31. 190—222 (1955). — Magyarul: Antal László: Magyar Nyelvőr 86. 200—2 (1962).

[12] Vö. például E. R. Gammon, A statistical study of English syntax: Preprints of Papers for the Ninth International Congress of Linguists (Cambridge, Mass., 1962) 121.

Dömölki Bálint, Fónagy Iván és Szende Tamás vizsgálataiban a két költő szembeállítása során nyert eredmények egyaránt érdeklik a nyelvészetet, a lélektant, az esztétikát és az irodalomtörténetet. Ugyanígy lehet ez minden jól elgondolt nyelvstatisztikai vizsgálat esetében is. Ezért gondolom azt, hogy a nyelvstatisztika, főleg a szövegstatisztika minden olyan diszciplínának elsőrendű és korszerű segédtudományává válhatna, amelynek vizsgálati tárgya valamilyen formában a szöveg. (Nem beszélve a híradástechnikai alkalmazásról.)

A filológusok nagyon hamar felismerték a nyelvstatisztika fontosságát, talán hamarabb, mint a matematikai statisztika megszületett volna. A Biblia-fordításokkal kapcsolatos konkordanciák és indexek, valamint a verstan formális vizsgálatának kvantitatív segédmunkálatai a kor színvonalán álló filológiai segédeszközök voltak.[13] A nagyteljesítményű matematikai gépek azonban elvi változást jelentenek ezen a téren is; ma már nem egy élet magányos cédulázásába kerül egy téma statisztikai előkészítése, hanem minden témának megvolna rá a lehetősége, hogy gyorsan, biztosan és megfelelő mennyiségben végeztesse el ezeket a szövegvizsgálatokat. Remélhetőleg — Fónagy és mások úttörő vizsgálatai nyomán — a magyarországi filológiában is rendszeresen felhasználásra kerül a statisztikai segéd-módszer.

4. A nyelvészetnek csupán még két ágazatát említem meg, ahol esetleg haszonnal használható fel valamilyen matematikai segítség.

Elsőként a d i a l e k t o l ó g i á t említem. A Magyar Nyelvjárások Atlasza szerkesztésének tudományos előkészítése során[14] nagy mennyiségű magyar nyelvtörténeti és nyelvjárási anyag került feldolgozásra. Deme László az anyagot némileg statisztikailag is feldolgozta, de könyvében csupán az ellenpéldáknak adja meg a számszerű mutatóit. Nyilván különösebb fáradság nélkül megadhatta volna az összes példák számszerű adatait is; akkoriban azonban ez nálunk nem volt divat. Egyébként sem biztos, hogy minden monográfiában minden tételnek számszerű adatait teljesen be kell mutatni, hiszen a számszerűség csak útmutató jellegű. Ha van kvalitatív szabályunk, akkor sokszor fölösleges részletként elhagyhatók a mennyiségi adatok. Viszont azt, hogy m i k o r m i l y e n számítások elvégzése, közlése szükséges, újabb kísérleteknek kell majd eldönteniük.

A történeti és összehasonlító módszer területén a g l o t t o k r o n o l ó g i a (lexikostatisztika) vizsgálatában alig tettük meg a konkrét munka kezdő lépéseit. Az elmélet bemutatásáról, sőt kritikájáról, már több írás is megjelent magyar szerző tollából.[15] Az uráli, illetőleg az altaji nyelvek rokonsági viszonyainak és történetének vizsgálatában azonban — magyar kutatótól — még kísérleti fokon sem kerültek ezek a módszerek alkalmazásra. A módszer elvi elutasítása vagy néma mellőzése

[13] Vö. P. Guiraud: Bibliographie critique de la linguistique statistique (Utrecht/Anvers, 1954) 18—40.

[14] Vö. Deme László: Nyelvatlaszunk funkciója és további problémái (Bp., 1956), a Tüzetes rész III. fejezete: „Az összesítő megterheltséglapok anyaga (A magyar nyelvjárások hangtörvényeinek csoportosítása)" (99—280). — Más irányú fontos kísérlet: Imre Samu, Nyelvjárási jelenségek statisztikai feldolgozása (Módszertani kísérlet): Magyar nyelvjárások 8. 11—26 (1962). — Legutóbb: Papp László: Nyelvjárástörténet és nyelvi statisztika (Bp., 1963) 186 lap.

[15] Fodor István, A statisztikai módszer alkalmazásának néhány kérdése: Magyar Nyelvőr 84. 196—203 (1960). — Uő: Current Anthropology 3. 131—4 (1962). A kérdés legutolsó összefoglalása: D. H. Hymes, Lexicostatistics so far: Current Anthropology 1. 3—44 (1960).

helyett helyesebbnek látszik annak kipróbálása és megvitatása. Erre egyébként megvan a szándék az említett szakterületek kutatóiban is; a matematikus-nyelvészek segítsége nélkül azonban nehezen indulhatnak el.

5. A vita során (de ezen a vitán kívül is) fölmerült a j e l e n t é s problémája a modern nyelvészeten belül. A kérdés tárgyalását mindenesetre megnehezíti az, hogy a hazai nyelvész-közvéleménynek ki nem elégítő tájékoztatása következtében néha nagyon vegyes értékű nézetek, sőt félreértések kavarognak e téma körül.

Az alapkérdésben, tudniillik abban, hogy v a n nyelvi jelentés, azt hiszem mindenki egyetért. A nézeteltérések ott kezdődnek, hogy milyen természetű a jelentés, és hogy mennyiben kell figyelembe venni a jelentést a grammatikai kategóriák és szabályok kialakításában. Itt csupán a második alkérdésre kívánok kitérni.

Véleményem szerint a jelentés részben benne van a grammatikai kategóriákban, épp a grammatikai ketagóriák (szabályok) adják meg a jelentés jelentékeny részének a formáját. Ez azonban nem azt jelenti, hogy az előzetesen megállapított jelentéskategóriák alapján lehet a grammatikai kategóriákat kialakítani.

Ezt már csak azért sem lehet, mivel egyelőre nincsen megnyugtató módon kiépített (nyelvészeti) szemantikai elmélet. Bizonyos kiindulópontok eléggé világosan láthatók,[16] sőt az autonóm nyelvészeti-szemantikai teória előkészítése is megkezdődött.[17] Természetesen egy szemantikai teóriának — elvileg — nem lehet a feladata, hogy nem szemantikai kategóriák magyarázatát adja.

A Chomsky-féle generatív grammatika transzformációs szabályai viszont rávilágítanak a mondat-szemantika bizonyos aspektusaira is.[18] A transzformációs szabály úgy is interpretálható, mint az átalakítandó és az átalakított szerkezetek jelentésének meghatározott viszonya. Ennyiben a grammatikaelmélet hozzájárul a szemantikaelmélet egyes részeinek kidolgozásához, de korántsem veheti át annak feladatait.

Külön probléma a szemantika „matematizálása". Ez eddig főleg az úgynevezett referent-szemantikára vonatkozott. Az erre irányuló erőfeszítések idáig nem hoztak számottevő nyelvészeti eredményt, inkább a pszcihológia (vagy az úgynevezett pszicholingvisztika) területén mozogtak.[19] — Természetesen a szemantika „matematizálása" egyáltalában nem reménytelen feladat. Viszont úgy gondolom, hogy egy megfelelő szemantikai teória kidolgozása, vagy legalább körvonalazása e l ő t t vagy n é l k ü l ez egyelőre még nem időszerű.

[16] Magyarul vö. Telegdi Zsigmond: Bevezetés a nyelvtudományba (Bp., 1961, Kézirat gyanánt) 15—19; Antal László, Jel, jelentés, szövegösszefüggés: Magyar Nyelvőr 85. 203—8 (1961); legújabban ugyanő: Questions of meaning (Hága, 1963).

[17] Lásd Paul Ziff: Semantic Analysis (Ithaca, N. Y., 1960); Magyarul vö. Terestyéni Ferenc, A leíró jelentéstan kérdéséhez és módszeréhez: Magyar Nyelv 58. 172—83, 326—35 (1962).

[18] Vö. N. Chomsky, Logical Structures in Language: American Documentation 8. 284—91 (1957).

[19] Vö. például: Charles E. Osgood—George J. Suci—Percy H. Tannenbaum: The measurement of meaning (Urbana, Ill., 1957); magyar ismertetése: Antal László: Magyar Nyelvőr 82. 504—5 (1958).

*Kalmár László**

Szeretnék magam is hozzászólni az elhangzott előadásokhoz, ugyanakkor válaszolnom kell a vitában az én előadásomhoz elhangzott hozzászólásokra. Úgy gondolom, nem okoz zavart, ha e kétféle teendőmet nem választom ketté forma szerint is.

1. Papp Ferenc előadásával kapcsolatban az a véleményem, hogy a nyelvi jelenségeknek és folyamatoknak azok a matematikai modelljei, amelyeket igyekezett olyan rendszerbe foglalni, amelynek a hiányai — a kémiai elemek Mengyelejev-féle periodikus rendszeréhez hasonlóan — jelzik, hol van még felfedezni való, valójában csak oktatási célra való, szemléltető jellegű mikromodellek. Avégett, hogy a nyelvi jelenségeknek és folyamatoknak tudományos kutatáshoz is felhasználható modelljéhez jussunk, ezeket a mikromodelleket bonyolultabb, ezért esetleg szemléltetésre már nem is alkalmas makromodellé kell szintetizálnunk. Amennyiben valamelyik mikromodell nem illik bele ebbe a makromodellbe, másikkal kell pótolni. Ebben a szintetizáló munkában is útmutatással szolgálhat a matematika, amely olyan halmazok vizsgálatával is foglalkozik, amelyek egyidejűleg többféle struktúrával is „fel vannak szerelve". Így például a folytonos csoportok olyan halmazok, amelyekben egy művelet is van definiálva, amelyre nézve a halmaz csoportot alkot, meg egy környezetfogalom is, amelynek szempontjából topológiai tér; még pedig nem egymástól függetlenül, hanem — és ebben áll az algebrai és a topológiai struktúra szintézise — oly módon, hogy a kérdéses művelet a környezetfogalomra nézve folytonos legyen.

Annak a kérdésnek szempontjából, lehet-e távolságfogalmat bevezetni a megegyező koordináták száma alapján, utalhatok az úgynevezett Baire-féle 0-tér fogalmára. Ez egy speciális topológiai, sőt metrikus tér; pontjai az egész számokból álló végtelen sorozatok, s egy ilyen sorozat egy másik ilyen sorozatnak annál szűkebb környezetébe esik, minél több tagig megegyezik vele (az első tagtól kezdve). Az ilyen környezetfogalom alapja ugyanaz a gondolat, mint ami Papp Ferenc előadásában is felmerült.

2. Kiefer Ferenc előadásában felvetődött az a kérdés, mennyiben alkalmazhatók az axiomatizált grammatikák az élő nyelvre. E kérédéskörbe tartozik az a kérdés is, mennyiben jogos fiktív szavakat adjungálni az élő nyelvhez avégett, hogy csökkentsük a kivételek számát. Például az orosz *ребенок* 'gyermek' szó többes alany-

* Az elhangzott hozzászólásnak kikerekített és (főleg a 6. és 7. pontban) bővített írásbeli változata.

esete, *ребята* rendhagyó, azonban ez a szóalak a *ребято* főnév szabályos többes alanyesete lehetne. Ilyen főnév azonban az élő orosz nyelvben nincs. Mennyiben jogos mégis beírni a szótárba a fiktív *ребято* 'gyermek' szót? Ha a szótár csak oroszról magyarra való fordításra szolgál, akkor jogos. Mert akkor a forrásnyelvi szöveg úgy sem tartalmazza a *ребято* szóalakot; viszont a *ребята* szóalakot a szótár alapján a *ребято* szó többes alanyesetének tekinti, tehát helyesen 'gyermekek'-nek fordítja a fordító (ember, vagy gép). Ha azonban a szótár magyarról oroszra való fordításra szolgál, akkor a *ребято* fiktív szó adjungálása nem jogos, mert azt a veszélyt rejtheti magában, hogy a fordító a *gyermek* szóalakot *ребенок* helyett a nem létező *ребято* szóalakkal fordítja. Ez a példa mutatja, hogy az élő nyelvet bizonyos szempontból meg lehet közelíteni olyan fiktív nyelvekkel, amelyekben kevesebb kivétel van, mint az élő nyelvben, amelyek grammatikája tehát közelebb áll valamely axiomatizált grammatikához, mint az élő nyelv tényleges grammatikája.

3. Petőfi S. János hozzászólása S. Marcus halmazelmélet-topológiai nyelvmodelljének ismertetésével hasznosan egészítette ki előadásomat és Kiefer Ferenc előadását. A homonimia-indexek alkalmazása a gépi fordításnál nagyon érdekes gondolat, de hátramarad az a kérdés, hogyan határozza meg azokat a gép. Persze lehetséges az a megoldás, hogy ezeket egy előredaktor írja be a forrásnyelvi szövegbe; evégett nem kell ismernie a célnyelvet, de jól kell ismernie a forrásnyelvet, hogy tudja, mikor van szó homonimiáról, ezenkívül ismernie kell a homonimia-indexeket megadó értelmező szótárt. De ha nem maradunk ennél a primitív megoldásnál, akkor a gép csak a kétes szó környezete, pontosabban az őt tartalmazó szintagmák alapján tudhatja megállapítani a homonimia-indexet. Ha viszont Marcus modellje helyett Kulaginát tesszük meg a fordítóprogram alapjának, akkor ugyanez a kérdés másképp jelentkezik. Az olyan kétes fajú szó, mint *ég*, nem tartozik az olyan kétségtelen főnevek családjába, mint *asztal, szék*, hiszen például *jól ég a fa* grammatikailag helyes mondat, de *jól asztal a fa* nem az. De az olyan kétségtelen igék családjába sem tartozik, mint *ugat, nyávog*, mert például *az ég zeng* grammatikailag helyes mondat, de *az ugat zeng* nem az. Hanem külön „főnév is, ige is lehet" családot alkot a *tűz, fal* stb. szavakkal együtt. Ha a gép a szótárból azt olvassa ki, hogy egy szó, például *ég*, ilyen hibridcsaládba tartozik, akkor ismét a kérdéses szó szintagma-környezetei alapján kell eldöntenie, hogy az adott mondatban például *burns* vagy *sky*-e a helyes fordítása.

4. Dömölki Bálint, Fónagy Iván és Szende Tamás előadásához megjegyzem, érdemes volna a hangstatisztikai és egyéb nyelvészeti vizsgálatokhoz legalkalmasabb kódolási eljárást megkeresni. Amint Kozma László professzor előadásából láttuk, a szabványos telex-perforátoron könnyen meg lehet akár 52 fonémát is különböztetni. Ahhoz, hogy ezzel a lehetőséggel élhessünk, mindössze a billentyűzetet kell kicserélni, ami könnyű dolog. Érdemes volna ezt megtenni és a hazai számológépeket nemcsak alfanumerikus, hanem diakritikus alfa-numerikus bemenettel és kimenettel is ellátni, hiszen tudjuk, hogy a nyelvész lépten-nyomon használ diakritikus jeleket. De ha már a megszokott kódolásnál maradunk, akkor is célszerűbb volna például a hosszú /ő/-t. [*oeoe*] helyett, a kétbetűs mássalhangzók kettőzésének mintájára, [*ooe*]- vel átírni. Ezáltal nemcsak egy-egy szalag-sort takarítanánk meg minden egyes hosszú *ő*-nél, hanem az is elesnék, hogy az [*oeo*] betűcsoport után csak a következő [*e*]-nél derül ki, hogy nem /öo/ fonémacsoportról, hanem hosszú /ő/-ről van szó. Ehelyett, ha két

[o] betű jön egymás után, csak azt kell megnézni, [e] jön-e utánuk, vagy nem; ha igen, hosszú /ő/-ről, ha nem, hosszú /ó/-ról van szó. (Az esetleges /óe/ fonémacsoport esetén kötőjelet kell az [oo] és az [e] közé írni.)

Arra a kérdésre, hogy a fonéma-statisztika a morféma-statisztikából következik-e, vagy önálló (például pszichológiai) oka van-e, valószínűleg az a válasz, hogy ez is, az is. Például annak, hogy valaki sok bilabiális hangot használ, az is lehet az oka, hogy gyakran használja a *mi* morfémát, meg az is, hogy örömet okoz neki két ajka összeérintésének elképzelése. A kettőt egymástól a maradékok módszerével lehet elkülöníteni: feltesszük, hogy csak az egyik ok áll fenn, megnézzük, ebből milyen statisztika következnék, s akkor az e statisztikától való eltérés, amennyiben szignifikáns, utal arra, hogy a másik ok is fennáll. Így fedezték fel a Neptun bolygót is. Ezzel kapcsolatban szeretném megkérdezni Fónagy Iántól, megvizsgálta-e, hogy a különböző költők fonéma-statisztikájának eltérése szignifikáns-e, mert a normális statisztikai ingadozás is okozhat bizonyos eltérést.

Arról, hogy a szó végének közelségét jelzi az, hogy hirtelen megnő a következő fonéma kitalálásának valószínűsége, eszembe jut, hogy amikor görög kiegészítő érettségire jelentkeztünk, tanárunk arra intett, ne legyünk olyan mulyák, mint az a volt tanítványa, aki Aisopos tyúkmeséjét kapta tőle fordítani az érettségin. Az első mondatot elkezdte fordítani: volt egyszer egy asszony, volt neki egy tyúkja, amely neki minden nap egy tojást . . . ; majd így folytatta: tanár úr kérem, ezt az igét nem ismerem. Ez példa arra, hogy hasonló jelenség mondat-szinten is mutatkozik.

Érdekes volt az is, amit Fónagy Iván a diakrónia szinkronikus vetületének megmutatkozásáról mondott a morféma-statisztika tükrében. Az a jelenség, hogy szinkronikus vizsgálat eredményéből lehet következtetni diakronikus jelenségekre is, a természettudósok körében közismert. Így például az úgynevezett statisztikus mechanikában, amely a gázak viselkedését írja le oly pontosan, abból kiindulva, hogy azok nagyon sok száguldozó, hol egymásnak, hol a gázat tartalmazó edény falának ütköző molekulából állnak, fontos szerepe van az úgynevezett ergodikus elvnek. Ez a ma már matematikai módszerekkel is megalapozott, de már megalapozása előtt is léptennyomon használt elv azt mondja ki, hogy azt, egy-egy molekula száguldozása során (diakronikusan) idejének hányad részét tölti az edény valamely adott részében (amelybe hol beszáguld, hol ismét kiszáguld belőle), úgy lehet kiszámítani, hogy megállapítjuk, hogy valamely időpontban (szinkronikusan) a gáz összes molekuláinak hányad része tartózkodik éppen az edénynek ebben a részében. Hasonló elven alapul a mikrobiológusnak az az eljárása, amellyel valamely csak mikroszkópon át észlelhető szervezet, például a spirochaeta, fejlődése menetét megállapítja. Az ilyen szervezet csak különleges festési eljárás alkalmazása után válik mikroszkopikusan megfigyelhetővé. Ehelyett sok spirochaetát figyel meg egyetlen időpontban (szinkronikusan); ezek között van spirochaeta-csecsemő is, ifjú is, élte javakorát élő és öreg spirochaeta is. A mikrobiológus aztán a spirochaeták megfigyelt, különböző fejlődési fokát időrendbe szedi, így állítja össze egy-egy spirochaeta (diakronikus) fejlődése menetét.

A diakrónia tükröződése szinkronikus vetületében arra vall, nem helyes a kettőt egymással olyan élesen szembeállítani, mint ahogy nyelvész körökben szokás.

5. Kozma László professzor előadásával kapcsolatban megkérdezem, nem volna-e jobb a lyukasztott szalagot mechanikus érintkezők segítségével való letapogatás helyett fotodiódás berendezés segítségével leolvasni. Ehhez elég volna öt fotodióda,

ami még a leolvasó berendezéssé való összeépítéssel együtt sem kerül olyan sokba, hogy nagyon megdrágítaná a tervezett, véleményem szerint kivitelezni feltétlenül érdemes nyelvstatisztikai berendezést. Persze a fotodiódás leolvasás sem küszöböli ki a mechanikai mozgást, hiszen a szalagot el kell mozgatni a fényforrás és a fényjelet elektromos jellé alakító fotodiódák között, hogy a szalagnak sorra mindegyik sorát leolvassa (azaz, ahol lyuk van a szalagon, amelyen keresztül fény éri a fotodiódát, adjon elektromos jelet, ahol pedig nincs lyuk, tehát a szalag eltakarja a fényforrást a fotodiódától, ne adjon elektromos jelet a megfelelő vezetékre). Azonban nem mindegy, hogy a mechanikai mozgás során a szalagot az egyik orsóról a másikra áttekercselő motor teljes forgási sebességével mozgatjuk-e, vagy pedig a letapogatást végző érintkezőknek a súrlódása (amelyek ott, ahol lyuk van, ezen át érintik egymást, ahol pedig nincs lyuk, a szalag elszigeteli őket egymástól) fékezi-e a mozgást.

Amennyiben azért nem érdemes a feldolgozandó szövegnek a statisztikai berendezésbe való bevitelét meggyorsítani, még ha ez olcsón lehetséges volna is, mert a feldolgozás úgyis lassú, akkor érdemes volna a feldolgozást is meggyorsítani úgy, hogy számláló jelfogók helyett elektronikus alkatrészeket használunk. Nem hiszem, hogy ilyen elektronikus nyelvstatisztikai berendezés tervezése túlságosan nagy gondot okozna Kozma professzornak, aki nemcsak a hagyományos, elektromechanikus telefónia kitűnő szakembere, hanem akinek a tanszékén folynak a legeredményesebb kutatások elektronikus telefonközpont tervezésére nézve. Úgy gondolom, nem jelentene lényeges költségtöbbletet sem ilyen elektronikus feldolgozás, hiszen egy kettős trióda ma már olcsóbb, mint egy szabványos telefon-jelfogó, s így egy elektronikus számláló áramkör sem kerülhet sokkal többe, mint egy számláló jelfogó. Márpedig, ha nem túlságosan nagy költséggel meg lehetne gyorsítani a szövegek bevitelét és feldolgozását, akkor érdemes lenne, mert minél gyorsabb a feldolgozás, annál hosszabb szövegeken lehet a statisztikát végezni s így annál kevésbé kell félni a „kis számok gonosz játékától", a statisztikailag várt eredménytől való lényeges eltérésektől, amelyek abból adódnak, hogy nem eléggé nagy számú adaton végeztük a statisztikai feldolgozást.

6. Antal Lászlónak igaza van: pontatlanul idéztem, amit a tavalyi nyelvtudományi vitán mondott arról, milyen legyen a magyar nyelv összeállítandó strukturális nyelvtana. (Voltaképpen nem is idéztem, hanem név nélkül, parafrázisszerűen emlegettem, ami a tavalyi vitaülés kissé későn kezembe kapott anyagának felületes átnézése során erről a fejembe maradt.)[1] De kérem, ő is pontosan idézze, amit én mondtam előadásomban. Én azt mondtam, a nyelvtudomány szakemberei hivatottak a szinkrónia és a diakrónia, a szintaxis és a szemantika fontossági sorrendjére vonatkozó vita eldöntésére, a matematikus erre nem illetékes. Márpedig ilyen vita van, hiszen vannak nyelvészek, akik szerint a diakrónia fontos — Antal László is a maga helyén értékes történetiségről beszélt — és vannak, akik nagyon is fontos nyelvtudományi kérdésnek tartják a szavak és mondatok jelentésének kérdését. E vitában, mint mondtam, a matematikus semlegességet jelent be. Ha szintaktikus célokra kellenek a nyelvésznek a matematikai módszerek, rendelkezésükre áll, de akkor is rendelkezésükre fog állni, ha az derül ki a vitából, hogy a szemantika is van olyan fontos a nyelvtudomány számára, hogy érdemes megvizsgálni, nem lehet-e ott is siettetni a fejlődést matematikai módszerek alkalmazásával. Ilyen módszerek alkalmazása valóban lehetséges

[1] Az előadásom átdolgozott kéziratában már kijavítottam a félreérthető parafrázist.

szemantikai kérdésekre is. Ugyanis a matematikai logika nemcsak szintaktikus, hanem szemantikai kérdéseket is vizsgál. (Ennek részletezésére, valamint a matematikai logika formális eljárásainak ismertetésére és a matematikai logikának a nyelvtudományhoz való viszonyának összefoglalására, amit Antal László jogosan hiányolt, nem jutott idő előadásomban.) Egy-egy *formalizált nyelvet* úgy definiálunk a matematikai logikában, hogy megadjuk az ábécéjét: alapjeleinek listáját, esetleg különböző (voltaképpen szófaji hovatartozásukra utaló) névvel ellátott alosztályokra osztva az alapjeleket (például számjegyek, változók, aritmetikai műveletek jelei, relációk jelei, logikai műveletek jelei, kvantorok). Azután megadjuk, hogy az ezen alapjelekből alkotott jelláncok (vagyis véges jelsorozatok) közül melyeket tekintünk kifejezéseknek, melyeket formuláknak (vagy esetleg bonyolultabb alakzatoknak, mint például az utasítások és a deklarációk a számológépközi nyelvekben). Ezen a megadáson persze nem a kifejezések és formulák felsorolását kell érteni, hiszen végtelen sokan vannak, hanem a kifejezések és a formulák halmazának olyan jellemzését, amelynek segítségével ezek a halmazok a matematikai logika további céljai szempontjából kezelhetőkké, elemeik áttekinthetőkké válnak. Ilyen jellemzésük állhat például olyan algoritmusból, amelynek segítségével bármely jelláncról el lehet dönteni, kiefejezés-e, formula-e, vagy egyik sem; vagy gyakrabban olyan algoritmusból, amelynek segítségével sorra minden kifejezést és minden formulát meg lehet kapni (generálni lehet), olyanféle, csak bonyolultabb módon, mint ahogy az összes (végtelen sok) természetes számot úgy generálhatjuk, hogy kiindulunk az 1-ből és mindig 1-et adunk hozzá az előzőleg generált számhoz.

Az így definiált formalizált nyelvet a matematikai logikus azonban csak kiindulásul, ahogy mondani szokás, *primitív keretül* (angolul: primitive frame) használja egy, már a logikát is tartalmazó *formális rendszer* definíciójához. Ez a definíció abban áll, hogy megadjuk, hogy a formulák közül melyek számítanak axiómáknak, s a formulákat formulákba átvivő transzformációk közül melyek számítanak következtetési szabályoknak. (Ha véges számú axióma vagy következtetési szabály van, akkor úgy adjuk meg ezt, hogy felsoroljuk őket. De a matematikai logika foglalkozik olyan bonyolultabb formális rendszerekkel is, amelyekben végtelen sok axióma van. Ilyenkor azt, hogy mely formulák számítanak axiómának, ismét olyan algoritmus segítségével adhatjuk meg, amelynek alkalmazásával bármely formuláról el lehet dönteni, axióma-e, vagy olyan algoritmus segítségével amely generálja az összes axiómákat, például úgy, hogy olyan (nem feltétlenül a primitív kerethez tartozó) formulákat, úgynevezett axióma-szkémákat adunk meg, amelyekből megkapjuk az axiómákat, ha bizonyos, az axióma-szkémában szereplő jelek, úgynevezett metaváltozók helyébe sorra behelyettesítjük a primitív keret összes formuláit.) A formális rendszer tételei, más szóval bebizonyítható formulái azok a formulák, amelyek úgy generálhatók, hogy axiómákból indulunk ki és mindig újra és újra alkalmazzuk a következtetési szabályokat az előzőleg már generált formulákra.

A formális rendszer e logikai része még mindig tisztán formális, ahogy a matematikai logikus mondja, szintaktikus jellegű. Azonban ezután rendszerint a formális rendszer által leírt (formalizált) matematikai elmélet *interpretációjának* kérdése következik (ha azt nem előbb, már a primitív kerettel kapcsolatban vetettük fel). Ez pedig már a szemantikához tartozik. Az interpretáció megadása végett definiálnunk kell, hogy mit jelent (pl. valamely halmaz mely elemét jelenti) egy-egy kifejezés, és hogy mit jelent egy-egy formula (legalábbis azt, mikor számít igaznak, mikor hamisnak, amit jelent). Csak e szemantikus jellegű definíciók alapján

vethetünk fel olyan kérdéseket, melyek a formális rendszer összes adekvát, vagyis olyan interpretációi, amelyekben a bebizonyítható formulák mindegyike igaz ítéletet jelent, vagy hogy teljes-e a formális rendszer valamely interpretációra nézve, vagyis minden olyan formula bebizonyítható-e, amely igaz ítéletet jelent. (Egy-egy formális rendszernek általában nagyon sokféle adekvát interpretációja lehetséges: ebben tükröződik a matematika széles körű alkalmazhatósága, aminek a nyelvészeti kérdésekre való alkalmazhatóság csak speciális esete.)

Persze nemcsak a matematikai logika foglalkozik szemantikai kérdésekkel, hanem a gépi és gépközi nyelvekkel kapcsolatban is lépnek fel ilyen kérdések és egyébként is érdekli a matematikusokat, hogy mit jelentenek a formuláik. És ha az derül ki az erre illetékes nyelvtudósok vitája során, hogy a nyelvész feladata továbbra is a szavak és a mondatok jelentésével is foglalkozni, és ha a nyelvész ehhez igényli, hogy a matematikus rendelkezésére bocsássa azokat a tapasztalatait, amelyeket formalizált nyelvek szemantikus vizsgálata során szerzett, akkor szívesen rendelkezésére bocsátjuk, mert meg vagyunk arról győződve, hogy a nyelvész is tanulhat belőlük.

Antal László érve, hogy tudniillik a szakmai mondat jelentése a szaktudósra tartozik, és az olyan közönséges mondatok jelentésével, mint hogy *éhes vagyok*, sem épp a nyelvésznek kell foglalkoznia, élesen elhatárolja a nyelvtudományt attól, ami nem a nyelvész dolga. Lehet, hogy ilyen elhatárolásra szükség volt a maga idejében, de ma nem időszerű. Mi volna, ha a matematikus is ilyen mereven elhatárolná a matematikát a többi tudománytól és például azt mondaná, hogy a halmazelmélet pontos axiomatikus felépítése az én feladatom, de a nyelvészetre való alkalmazásával törődjék más?! Akkor én sem tartottam volna itt előadást. Ugye, nem ezt várják tőlünk, matematikusoktól, a nyelvész kollégák? A mi korunk a nagy szintézisek, a határtudományok, a kibernetikához hasonló komplex tudományok kifejlődésének kora. Márpedig a határterületeket mindkét oldalról kell gondozni, csak így jöhet létre eredményes fejlődésüket biztosító együttműködés a két határos terület tudósai között. Nos, ha a jelentéstan nem a „tiszta nyelvészethez" tartozik, akkor hát a nyelvészetnek más tudományokkal való határterületéhez tartozik, s így a nyelvésznek is gondoznia kell, meg annak a másik fajta embernek is, akire Antal László szerint tartozik a vizsgált mondat jelentése, éppúgy, mint ahogy a matematikai nyelvészetet is gondoznia kell a matematikusnak is, meg a nyelvésznek is.

7. Két hozzászólás is: Gáldi László professzoré és Mikó Pálnéé hangsúlyozta azt, ami különben Papp Ferenc előadásában is szerepelt, hogy csak a többiektől pontosan elhatárolt egyedekről lehet statisztikát csinálni. Így addig nem lehet a magyar szavakról statisztikát csinálni, amíg pontosan meg nem mondjuk, mit értünk magyar szón, és hogy két szópéldány mikor ugyanannak a szónak két példánya, mikor két különböző szónak egy-egy példánya. Ezzel kapcsolatban megemlítem, hogy ilyen elhatárolás nemcsak a statisztika, hanem minden tudományos vizsgálat előfeltétele. Így a jelentéstant is azért félnek sokan a nyelvtudományhoz számítani, mert úgy vélik, nem is lehetne pontosan definiálni, mikor egyezik meg két szónak, vagy két mondatnak a jelentése. Kétségtelen, hogy a tudományos szinonima-elmélet előfeltétele a korszerű értelemben tudományos szemantikának. Az, hogy vannak szigorú értelemben egyjelentésű szavak mellett rokonértelmű szavak is, nem lehet akadálya a tudományos szinonimia-elméletnek, hiszen például két valós szám is lehet pontosan egyenlő is és csak többé-kevésbé megközelítőleg egyenlő is, mégis van a valós számoknak is szabatos tudományos elmélete.

8. Vértes Editnek igaza van a statisztikailag várttól eltérő jelenségekkel. Ezek nagyon tanulságosak lehetnek, ha megtaláljuk az eltérés okát. Ilyen jelenség az, hogy a *Sűrű sötét az éj, dühöng a déli szél* verssorban a két névelő kivételével minden mássalhangzó magas, ez a sor mégsem szelíd, hanem borzongató hangulatot kelt. Nekem az a véleményem, ebben a verssorban már előre kísért az érckakas csikorgásának magas, de kellemetlenül magas hangja, Arany János már előre hallotta ezt a csikorgó hangot, azért használt az első verssorban is sok magas, de nem szelíden, hanem így, egy tömegben kellemetlenül magas /ü/ és /ö/ hangot. Persze, hogy a szavak jelentése is hozzájárul a verssor borzongató hangulatához a fonémák mellett. Érdekes lenne analizálni, melyiknek mekkora a szerepe.

9. Ami a topológia alkalmazását illeti a nyelvészetben, igaza van Vértes Editnek abban, hogy nála csak hasonlatról, nem a mai értelemben vett matematikai nyelvészeti modellről volt szó. Azt azonban, hogy kétségbe vonja a topológia alkalmazásának lehetőségét a nyelvészetben, annak tudom be, hogy Kerékjártó Béla professzor tanítványa volt, aki folytonossági topológiával foglalkozott, míg itt nyilván diszkrét, vagy „összetapadt" topológiákról van szó. (Például egy szónak egyetlen Kulaginaféle környezete van és az egyben összes alakjának is egyetlen környezete, úgy, hogy egy szó különböző alakjait nem lehet környezetekkel szétválasztani még a T_0 terekben kikötött szétválasztási axióma értelmében sem. De hiszen Kulagina nem arra a célra vezette be a környezetfogalmat, hogy annak segítségével szétválassza egy szó különböző alakjait, hanem ellenkezőleg, arra a célra, hogy azonosítsa őket, eltekintsen attól, hogy a szó melyik alakjáról van szó.)

10. Papp László rossz helyre adresszálta a kérését, hogy a matematikus segítsen neki öt, statisztikai adatokból adódó, világos jelentésű paraméterből egyetlen számadatot csinálni. Előadásomból kitűnik, hogy a matematikus már régóta elfogad adatnak olyat is, amit nem egy, hanem több szám jellemez, mint például a vektorokat, akár a több dimenziós térbeli vektorokat is, és örül, ha csak ötdimenziós és nem $5 \cdot 10^{23}$, vagy még több dimenziós térbeli vektorról van szó, mint a statisztikus mechanikában. Persze ez nem zárja ki, hogy lehetséges, hogy az adott esetben van jó jellemző egydimenziós számadat is, csak nem biztos, hogy van.

Kozma László

Kalmár László hozzászólására válaszolok. — Kétségtelen, hogy fotodiódás letapogatással lényegesen meg lehet gyorsítani a leolvasást, azonban ehhez elektronikus számlálószerkezet szükséges, ami elég drága áramkör. Tekintettel arra, hogy a Tanszék rendelkezik az elektromos távírógépeknél használatos mechanikus letapogató berendezéssel, és számláló jelfogókat is tud készen kapni az ipartól, ennélfogva a Tanszék számára sokkal egyszerűbb és kevesebb mérnöki munkát igénylő feladat, hogyha a berendezést elektromechanikusan oldja meg. Az így elérhető sebesség több nagyságrenddel meghaladja az ember által végezhető munkasebességet és az előkészítés feladata (a lyukasztott szalag elkészítése) adja a Nyelvtudományi Intézet számára a legnagyobb munkát, amihez képest az automatikusan folyó leolvasás lassú vagy gyors volta kisebb jelentőségű.

Fónagy Iván

1. Egy gondolatot szeretnénk kiemelni az igen sok közül Rényi Alfréd előadásából.*

Szó esett a beszédet kísérő jelenségek hírértékéről. Hírértéke van, amint Rényi professzor mondta, a telefonbeszélgetést kísérő (és zavaró) hálózati zörejeknek is. Nyilván különbséget kell tennünk a szó tágabb és szűkebb értelmében vett hírérték között. A szó tágabb értelmében — az emberi fül számára érzékelhető frekvencia- és intenzitáskülönbségeket véve jelkészletként — nincsen különbség a telefonba mondott szó és a szót érthetetlenné tevő zaj között. A szó szűkebb, nyelvészeti és lényegében köznyelvi értelmében — a magyarban megkülönböztetett beszédhangokat, pontosabban az ezeket megkülönböztető akusztikai jegyeket véve alapul — a beszéd szembeállítható a hírértékkel nem bíró csatornazajjal. Ha a zaj forrása nem a csatorna, hanem az emberi beszédszerv, akkor ez a zaj a szónak szűkebb értelmében is lehet informatív. Amikor ismerős helyen kopogtat valaki és a „*Ki az?*" kérdésre azt feleli, hogy „*Én*", akkor mindaz, ami beszédében zajnak minősíthető (a normától eltérő artikuláció, pöszeség, rekedtség, dadogás, monoton hanglejtés) nagyon is informatív: ez kölcsönöz éppen értelmet a különben semmitmondó személyes névmásnak. Vannak hírértékkel bíró alkalmi torzítások is: a düh hatására bekövetkező rekedtség vagy szaggatott beszéd, a vágyat tükröző vagy máskor megvetést mutató nazalizálása orális hangoknak stb. Ezek a zajelemek az idők folyamán a jelkészlet elemeivé válhatnak. A mássalhangzók és magánhangzók affektív megnyújtása *(a ppiszok, ce bbandit; uugyan, aber natüürlich)* például a nyelvi kód szerves része.

Ezzel párhuzamosan felvethető a kérdés, vajon különbséget lehetne-e tenni a kommunikáció szempontjából nélkülözhetetlen, vagy legalábbis hasznos és a felesleges, zavaró redundancia között is.

2. Minden hangstatisztika alapvető kérdése: a konkrét beszédhangokból induljunk-e ki, vagy a fonémákból? Mi a fonémákat vettük alapul. Ennek egyik oka, hogy a munkának ebben a fázisában nyomtatott szövegeket dolgozunk fel. Bonyolult módon és kétes eredménnyel lehetne a szöveg alapján a kiejtést rekonstruálni (hiszen még a kontextuális, helyzethez kötött variánsok is sokszor fakultatívok, *tea* ~ *teja*, *etyszer* ~ *eccer (egyszer)* stb.

De van ennél súlyosabb ellenérv is: ha fonemikus szöveget fonemikus átírás szerint viszünk a szalagra, ha *háztól*-t, *adtá*-t írunk, gépünk megfelelő program segít-

* Lásd Kötetünkben 245—251. — *A szerkesztők.*

ségével bármikor közölheti, hogy *-zt-*, *-dt-*, azaz hány zöngétlen alakban szereplő zöngés fonéma van a szövegben. Ha azonban *hásztól*-t, *attam*-ot írunk, ezt már sokkalta nehezebb megállapítani.

A fonemikus átírás mellett szól az a körülmény is, hogy a fonemikus átírás (a „broad transcription") elvei világosabban állnak előttünk, mint a fonetikai átírás („narrow transcription") elvei. Végül: a fonemikus átírás sokkal kevesebb munkával jár. Az egyszerűbb, könnyebb megoldás, véleményünk szerint, ezúttal helyesebb és elvszerűbb is.

3. Lényeges kérdést vetett fel Gáldi László és Vértes Edit. Bajos a jellegzetes egyéni hangzóeloszlást esztétikai szempontból értékelni, amíg nem ismerjük a morfémák vagy szavak gyakorisági eloszlását. Hiszen a hangzóeloszlás nyilván a morféma-eloszlás függvénye. Szerencsénkre verses műfajt elemeztünk. A versekben pedig ez a függés kétirányú. A morféma-eloszlás nem kisebb mértékben függ a hangzóeloszlástól, más szóval a tartalom is függ a formától. A forma kétszeresen determinált. Ezért nem alkalmazható Vértes Editnek egyébként talán helyes elve verses szövegek esetében. Kétségtelenül szerepe van az ördögfiak nevének abban, hogy sok az *r* azokban a jelenetekben, melyekben *Kurrah*, *Berreh*, *Duzzog* szerepel. Mégsem lenne helyes a neveket első előfordulásuk után kizárni a statisztikából. Nem véletlen, hogy Vörösmarty éppen ilyen hangzású neveket adott az ördögfiókáknak. (Vagy legalábbis szerencsés költői véletlen.) Vagy: helyes lenne-e Kosztolányi „Iloná"-jának hangzóeloszlását vizsgálva 18 *Iloná*-ból 17-et elhagyni?

4. Végül a J. Soltész Katalin előadásának vitájához* kapcsolódva Guiraud „mentségére" megemlíteném, hogy nem feledkezett meg a rímszavak fontosságáról. Kimutatta például, hogy a könnyen és természetesen rímelő költő verseiben rímhelyzetben sem gyakoribbak a ritka szavak, mint más helyzetben.

* Vö. „A szerkesztők bevezetésé"-nek **5.** pontjával.

*Kiss Károly**

1. Dömölki Bálint, Fónagy Iván és Szende Tamás köznyelvi hangstatisztikai vizsgálatairól tartott előadásához csak annyit kívánok megjegyezni, hogy rendkívül tiszteletre méltó egy ilyen vállalkozás, és bár mind a költészetnek, mind az említett költőknek magam is nagy tisztelője vagyok, a köznyelvi hangstatisztika szempontjából azonban semmiképpen sem mondható jellemzőnek sem az illető költők, sem semmilyen más költői mű nyelve. Helyesebb lenne köznyelvi hangstatisztikai vizsgálatokat *ténylegesen elhangzott beszédek* vagy *beszélgetések* köréből végezni.

2. Nagy örömmel hallottam J. Soltész Katalin „A szókincs statisztikai vizsgálata" című előadásának bevezetőjében a szóstatisztikai munkák vázlatos áttekintése során Nemes Zoltán egymillió szótagon alapuló munkájának megemlítését,[1] annál is inkább, mert eddig — különösen a megjelent nyelvészeti közleményekben — azt kellett tapasztalni, hogy a cikkírók egyáltalán nem vettek tudomást erről a munkáról és eredményeiről, vagy ha tudomásul vették is eredményeit, nem tekintették egyenrangúnak hasonlíthatatlanul kisebb szövegből nyelvészek által készített statisztikákkal.

A gyakoriság-szótár elkészítése a legnagyobb mértékben érdekli a gyorsírókat, hiszen nekik napi munkájuk során kell megküzdeniük a nyelv állandó változása által felvetett nehézségekkel. A felszabadulás utáni nyelv gyakorisági vizsgálataihoz jó adalékul szolgálhatnak az azóta megjelent gyorsírási rövidítésgyűjtemények, amelyek összeállításában természetszerűen elsősorban a gyakoriság, az előfordulás volt a fő tényező.

3. Végül a határozati javaslathoz szólva, úgyis mint gyorsíró és úgyis mint magyar szakos tanár, örömmel üdvözlöm a javaslatot, mind a tematikai, mind a szervezési kérdések, mind pedig az anyagi kérdések vonatkozásában. Annak a reményemnek szeretnék kifejezést adni, hogy a statisztikai feldolgozásban, továbbá az információelmélet adta új lehetőségeknek a nyelvészeti vizsgálatokban való fokozottabb felhasználásában, a gyakorisági szótár vagy fordított betűrendi szótár elkészítésében, a statisztikai vizsgálatok szövegelőkészítésében — hogy csak néhány főbb pontot említsünk — a gyorsírók részéről is akadnak, akik részt tudnak venni.

* Írásban benyújtott hozzászólás.

[1] Nemes Zoltán: Szóstatisztika egymillió szótagot felölelő újságszövegek alapján (Bp., 1939) = Az Egységes Magyar Gyorsírás Könyvtára 190. szám.

Hell György

Az orosz—magyar gépi fordítás elméleti és gyakorlati kérdései

Sajnos, egyelőre nemvagyunk abban a helyzetben, hogy az orosz—magyar gépi fordítás eredményeit gépen fordított szövegeken szemléltethessük. Még ott sem tartunk, hogy a kérdést elméletileg minden részletében tisztázottnak tekinthessük. A kezdetén vagyunk csak az orosz—magyar gépi fordításnak, és csak kezdeti eredményekről számolhatok be.

A következőkben szeretném ismertetni az orosz—magyar gépi fordítás eredményeit egészen a kezdettől fogva. Nem azért, mintha ezzel szemléltetni akarnám, hogy a gépi fordításhoz túl nehéz út vezet el; sőt ellenkezőleg, szükséges, hogy a gépi fordítással kapcsolatban még esetleg meglevő fölösleges rejtélyesség alaptalansága nyilvánvalóvá legyen.

1. A fordításra használható gépek

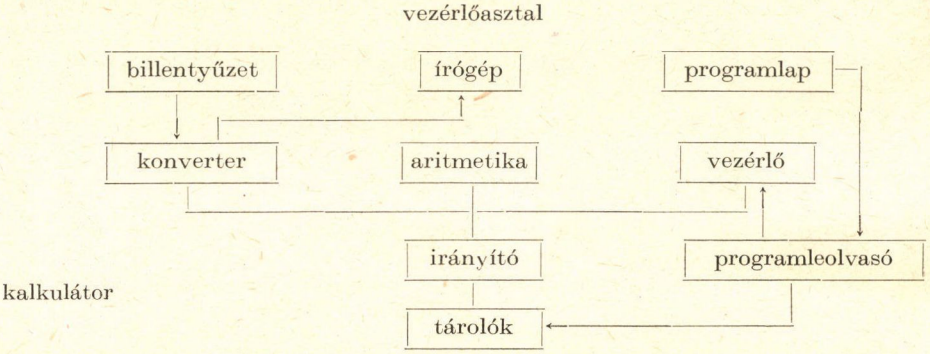
A gépi fordítás mint erősen gyakorlati jellegű kérdés, már kezdetben is egy használható algoritmus összeállítására ösztönzött. Ehhez, a fordításra használható gépek műveleti lehetőségeinek megismerésén kívül, egy olyan gépre is szükség volt, amelyen az algoritmust ki lehet próbálni. A nálunk meglevő „Ural" típusú elektronikus számológépek helyett célszerűbbnek látszott ezt a kezdeti kísérletezést egy egyszerűbben programozható, könnyebben hozzáférhető, de ugyanakkor a nagy gépek lehetőségeit is lényegében tartalmazó, bár lassabban működő gépen elvégezni. Erre a célra igen alkalmasnak bizonyult a Kozma László tanszékvezető egyetemi tanár által tervezett és a Budapesti Műszaki Egyetem Vezetékes Híradástechnikai Tanszékén található MESZ—1 típusú jelfogós számológép.[1]

A fordítás felépítésének áttekintése céljából szükségesnek látszik, hogy nagy vonásokban vázoljuk a fordításhoz használatos gépek, jelen esetben a MESZ—1 gép elvi felépítését és a fordítás menetét.[2]

[1] Itt is meg szeretném köszönni azt az előzékenységet és megértést, amellyel a Tanszék a gépet az orosz—magyar fordítás kezdeti algoritmusainak kipróbálása céljából rendelkezésemre bocsátotta. Külön köszönöm Frajka Béla tanársegéd és Bognár Zoltán demonstrátor közvetlen segítségét.

[2] Kozma László, A Műszaki Egyetem első digitális számológépe: Budapesti Műszaki Egyetem tudományos évkönyve (Bp., 1960) 82—96.

A számológép két nagy egységre osztható: a vezérlő asztal és a kalkulátor egységeire. A vezérlő asztal könnyen áttekinthető; itt történik az adatok bevitele (a tárolók feltöltése) a billentyűzet gombjainak egymásutáni lenyomásával és a programot tartalmazó lap elhelyezése. Az írógép a gépből kijövő eredményt rögzíti. A billentyűzeten keresztül beadagolt értékek tízes számrendszerben vannak megadva, ezeket a konverter kettes számrendszerű számokká alakítja át és a gép ebben a rendszerben számol. A programlapon rögzített utasításokat (a műveletet előkészítő utasításokat, műveleti utasításokat, kivitelt stb.) a programleolvasó érzékeli és küldi tovább a tárolókba, illetőleg a vezérlőbe. A tulajdonképpeni műveleteket az aritmetikai egység végzi, a kész eredmények a konverteren keresztül lépnek ki a kettes számrendszerből a tízes számrendszerbe visszaalakítva.



1. ábra

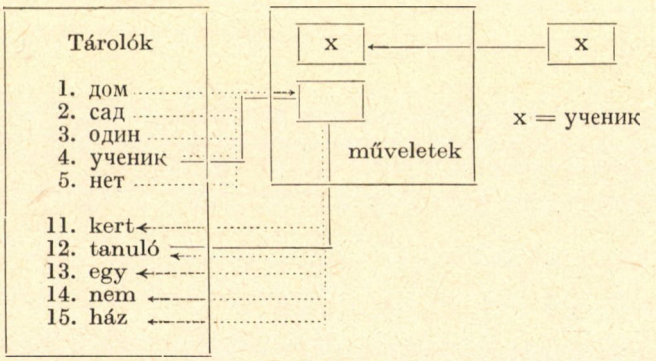
2. A fordítás általános menete

A gépet a fordításhoz elő kell készíteni. Ez abban áll, hogy elhelyezzük benne mindazokat az adatokat, amelyek a fordítás megvalósításához nélkülözhetetlenek. Ilyen adatok azok a magyar szavak, amelyeket a gépnek fordítás alakjában ki kell adnia. Ezek a szavak nem teljes alakban szerepelnek a gép tárolóiban, nemcsak azért nem, mert ez lehetetlen lenne, de célszerűtlen is egy-egy szónak minden ragozott alakját tárolni. A tőmorfémák formájában tárolt szókincs viszont szükségessé teszi azt is, hogy a kész szóalak összeállításának utasításait is elhelyezzük a gépben. Ezek az utasítások olyan jellegűek, hogy a fordítandó szöveg elemzésére támaszkodva a két nyelv között fennálló fordítási szabályok szigorú figyelembevételével alakítják ki a célnyelv szövegét. Ezzel tulajdonképpen mindjárt el is érkeztünk a gépi fordítás lényegéhez: olyan elemzési módokat kell alkalmaznunk, mellyekkel megkapjuk a forrásnyelvnek a mondanivalót hűen tartalmazó szerkezeti egységeit, olyan fordítási szabályokat kell előírnunk, amelyek formális-mechanikus alkalmazási lehetőségüknél fogva biztosítják a szerkezeti egységekben megragadott forrásnyelvi tartalom hiánytalan célnyelvi visszaadását. Az adatok tárolása mindennek a megvaló-

sítását célozza. Szükségünk van egy forrásnyelvi szótárra és szabályokra, mellyekkel a fordítandó szöveg jelentést és viszonyítást hordozó elemeit föl tudjuk deríteni, és adott esetben konkrét szerepüket meg tudjuk állapítani.[3]

A gép előkészítésében a szóanyag elhelyezése (tárolása) adja látszólag a legkisebb problémát, viszont első pillantásra nehezen megvalósíthatónak tűnik az a követelmény, hogy az elemzés és fordítás szabályait is tárolnunk kell. Mindez azonban a gép által elvégezhető műveletek megfelelő kombinálásával elérhető.

Minden olyan gép, melyet eddig gépi fordítás céljaira használtak, elsősorban matematikai feladatok megoldására szolgál, és ezért elsősorban matematikai alapműveleteket, ezek összetett változatait, valamint néhány logikai műveletet képes végezni. A gépi fordításban a legnagyobb szerepe az összehasonlítás műveletének van. Egy megfelelően összeválogatott összehasonlító művelet-sorozat segítségével kiválaszthatjuk a megfelelő célnyelvi szavakat. Az eljárás lényegének bemutatására tegyük fel például, hogy a géppel az ismeretlen „ученик" szót akarjuk lefordítani.[4] A fordítást most korlátozott terjedelmű gépi szótárral hajtjuk végre, melynek orosz részében a *дом, сад, один, ученик, нет* szavak, a magyar részben: *kert, tanuló, egy, nem, ház* szavak vannak elhelyezve. Mint látjuk, a szavak „párhuzamos" elrendezése nem fontos, viszont lényeges, hogy a megfelelő orosz és magyar szavak között olyan „kapcsolatot" teremtsek, mellyel az ismeretlen fordítandó szónak bármely orosz szóval történő eredményes összehasonlítása után a gép „adja ki" (gépelje ki) a megfelelő jelentésű magyar szót. Egy ilyen „kapcsolat" elektronikus gépeken megoldható és a fordítás most már abból áll, hogy a valóban ismeretlen szót egymás után minden forrásnyelvi szótári szóval összehasonlítom olyan feltétel mellett, hogy pozitív válasz esetén érvénybe lép az említett utasításszerű kapcsolat.



2. ábra

[3] P. Sz. Kuznyecov—A. A. Ljapunov—A. A. Reformackij, Основные проблемы машинного перевода: Вопросы Языкознания 5/5. 107—110 (1956).

[4] A munkaértekezlet után jelent meg: Hell György, A gépi fordítás néhány kérdése: Műszaki Könyvtárosok Tájékoztatója 9/5. 1—14 (1962).

A gép által elvégzendő műveletek sorát a fordítási algoritmus tartalmazza.

3. A MESz—1-re írt program

Lehetetlen vállalkozás lett volna mindjárt kezdetben egy olyan orosz—magyar algoritmus összeállítását célul kitűzni, mellyel tetszés szerinti szövegeket le lehet majd fordítani. Célszerűbbnek látszott először csak a legfontosabb és a fordítás szempontjából legkönnyebben megoldható nyelvtant tartalmazó mondatok fordítási algoritmusát leírni és később a mondatok grammatikáját célszerűen bővítve eljutni a bonyolultabb mondatszerkezetek megoldásáig. Így a megoldandó feladatok nehézségi fokát is magunk tudtuk megválasztani.

Így készültek algoritmusok toldalékos és elöljárós orosz főnevek fordítására, majd egyszerűbb mondatok fordítására is. A első géppel fordított mondat *(Решает ли машина задачу?)*[5] egy ragozott ige és két toldalékos nőnemű főnév elemzését tartalmazta. A gépi szótár orosz része a *ли* szócska mellett a *реша*, *машин* és *задач* töveket, valamint az *ет*, *а* és *у* toldalékokat tartalmazta.

A szavak tárolásához minden betű (cirill és latin) négy kettes számrendszerben írt számjegyből (bit)[6] álló kódot kapott (például: Ш = 1100, з = 1010; D = 1001, K = 0011).

A MESZ—1 adottságai ezen a téren a felvett módszertani egyszerűsítéseken kívül még további megszorításokat tettek szükségessé. A gépnek ugyanis összesen 28 emlékező tárolója van, melyekbe egyenként 27 bitnek megfelelő adatot lehet tárolni. Ha a cirill ábécé 33 betűjét akarom kettes számrendszerben kifejezni, úgy ehhez egy betű kódolása céljából 6 bitre lenne szükség, ami csak legfeljebb négy betűs szavak tárolását engedte volna meg. A cirill ábécének csak 16 betűjét felhasználva négy bitből kapunk egy betűt, és ez megfelelően összeválogatott hat betűs orosz és magyar szavak alkalmazását tette lehetővé. Eltérően a nagyobb elektronikus számológépektől, a MESZ—1 tárolóiba a szavakat nem betűnként kellett beadagolni, hanem egyszerre az egész szót tízes számrendszerben. Ehhez a szónak megfelelő kettes számrendszerben írt számsort tízes számrendszerbe kellett átírni, beadagolni, és a gép automatikusan kettes számrendszerben (azaz a betűknek megfelelően) helyezte el a tárolókban a szavakat.

A szótári egyeztetéshez a fordítandó mondat toldalékos szavairól le kellett vágni a toldalékokat. A program a levágás műveletét betűnként írta elő a szó végétől kezdve, és egy-egy betű levágása után a gép összehasonlítást végzett. A betű levágását a MESZ—1 gépen tizedes vesszőig való osztással $\left(\left[\frac{x}{2^4}\right]; \left[\frac{x}{2^8}\right]; \left[\frac{x}{2^{12}}\right]\right)$; a rag visszaállítását osztással, szorzással

[5] Azért volt célszerű kijelentő mondat helyett kérdő mondatot venni, mert ennek fordítása kevesebb változatot ad.

[6] Egy *bit* a kettes számrendszerben egy számjegyet jelent, tehát 1 vagy 0 számjegyet.

és kivonással: $x - \left[\frac{x}{2^4}\right] \cdot 2^4$ (a szögletes zárójel azt jelenti, hogy a nevezővel csak a tizedes vesszőig szabad osztani) lehetett elérni. A rag elemzése a szótő alapján történt, a program megfelelő rendszerű felépítésével. Ez lényegében azt jelentette, hogy külön ragelemzést kapott az ige ragja, és külön elemzést a nőnemű főnév végződései, ami az adott esetben mindig egyetlen értelmezéshez vezetett. Ez a módszer az indulásnál fölöslegessé tette azt, hogy a szótárban a szótövekre vonatkozóan nyelvtani kiegészítő jelzéseket helyezzünk el, ami a gép kapacitása miatt amúgy is lehetetlen lett volna.

A szótári rész beadagolása a tárolók egymásutáni feltöltésével történt, a lefordítandó mondat szavait egyenként kellett tárolni. A fordításhoz 97 műveletre volt szükség, amelyeket megfelelő lapon lyukasztással lehetett előírni. A fordítást a gép tízes számrendszerű számokban (külön a magyar szótövet és külön a ragot) gépelte ki, melyeket a kettes számrendszerbe alakítva minden négy számjegy egy-egy latin betűt adott. (A fordítás: *Megoldja-e a gép a leckét?* A *feladat* szó nem fért el egy tárolóban.)

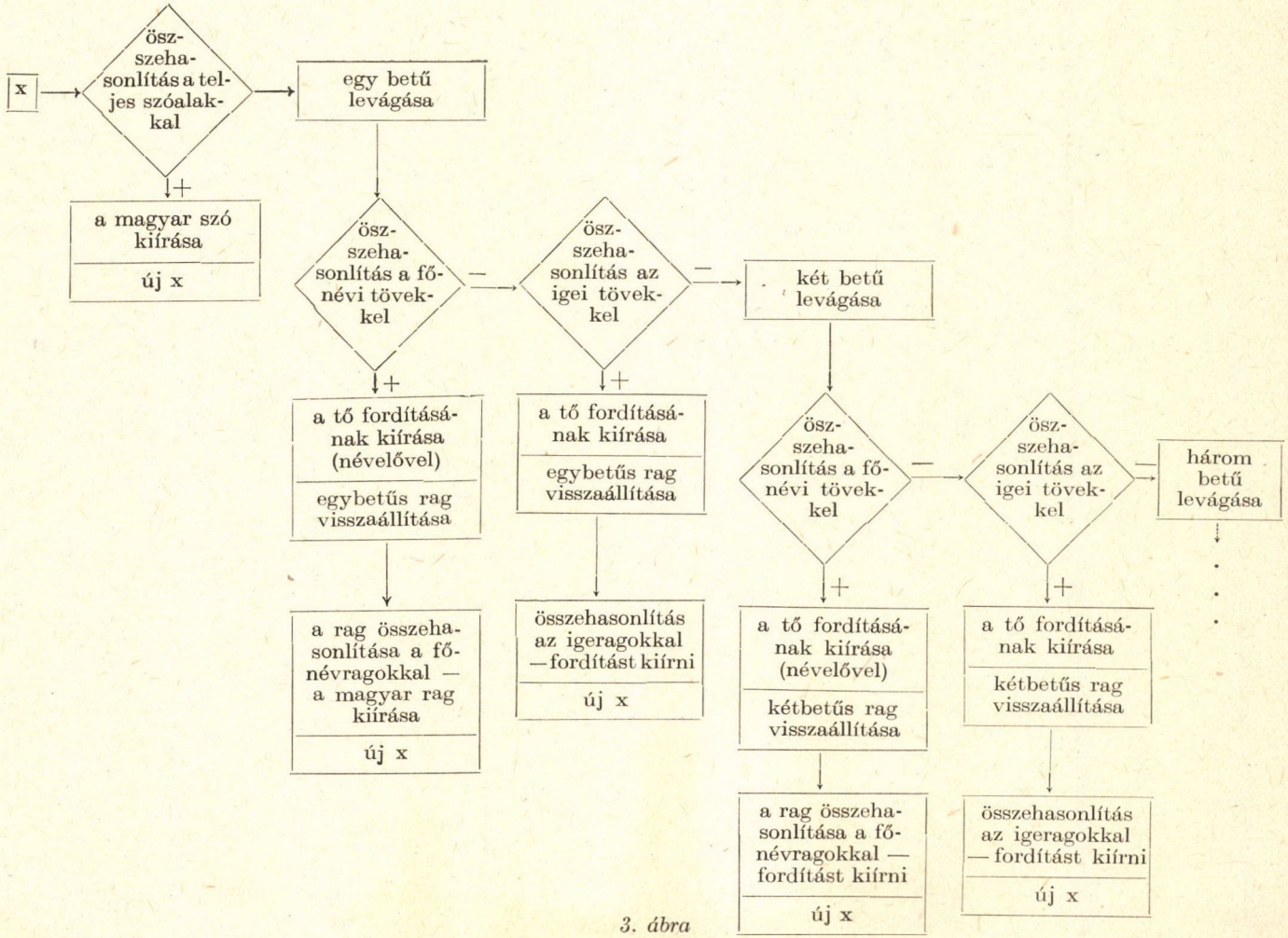
A műveletek egymásutánját a 182. lapon látható utasítási vázlat mutatja.

Látjuk, hogy a szavak morfológiai elemzéséhez a program a szótőből indul ki, és a szótövek különböző csoportjaira vonatkozó végződéseket és ragokat egymástól eltérően az adott szóra jellemző végződéssorral azonosítja. Ez a fordítási menet igen egyszerű és fölépítésében magán viseli az első kísérlet minden kezdetlegességét. Az adottságok következtében mindig a legegyszerűbb megoldást választja (például a rag levágásában és azonosításában) és azonos jellegű műveleteket ismétlésben tartalmaz, de ez a szótárban a szavak mellé már el nem helyezhető nyelvtani jelzések hiányában szükségszerű volt. A szótár grammatikai információit ebben az esetben a logikai műveletek (az „igen" és „nem" között történő választás) rendszere helyettesíti.

4. Grammatikai információk

A gyakorlat, további programok készítése azt mutatta, hogy a nyelvtani jelzések még ilyen műveleti elrendezéssel sem mellőzhetők. Már aránylag egyszerű mondatszerkezetekben szükség volt arra, hogy egy toldalékmorféma magyar fordítását ne csak a szótő és elöljáró, hanem a már lefordított szavak mondattani szerepe alapján döntsük el. Például az egyező alakú alany- és tárgyesetű orosz főnevek magyar megfelelőjének megállapításához tudnunk kell azt, hogy az adott szó előtt volt-e már állítmány, illetőleg alany a mondatban, és milyen jellegű az állítmány. A birtokos esetű szó mondattani szerepének eldöntésekor is fontos tudnunk, hogy előtte főnév áll-e, vagy ige stb. Ebből a célból a lefordított szavak mondattani szerepéről, azok sorrendjével egyezően nyelvtani feljegyzéseket is kellett vezetni, hogy szükség esetén az elemzés megvalósítható legyen.

Ezek a nyelvtani feljegyzések jellegükben erősen eltérnek azoktól, melyekről korábban szó volt, amelyeket a szótárban kell elhelyezni, hogy segítségükkel például a ragelemzést el lehessen végezni. Most külön tárolunk


x
össz-szeha-sonlítás a tel-jes szóalak-kal
+
a magyar szó kiírása
új x
egy betű levágása
össz-szeha-sonlítás a fő-névi tövek-kel
−
össz-szeha-sonlítás az igei tövek-kel
−
két betű levágása
+
a tő fordításá-nak kiírása (névelővel)
egybetűs rag visszaállítása
a rag összeha-sonlítása a fő-névragokkal — a magyar rag kiírása
új x
+
a tő fordításá-nak kiírása
egybetűs rag visszaállítása
összehasonlítás az igeragokkal —fordítást kiírni
új x
össz-szeha-sonlítás a fő-névi tövek-kel
−
össz-szeha-sonlítás az igei tövek-kel
−
három betű levágása
+
a tő fordításá-nak kiírása (névelővel)
kétbetűs rag visszaállítása
a rag összeha-sonlítása a fő-névragokkal — fordítást kiírni
új x
+
a tő fordításá-nak kiírása
kétbetűs rag visszaállítása
összehasonlítás az igeragokkal — fordítást kiírni
új x

3. ábra

grammatikai jegyeket (morfológiai és szintaktikai jelzéseket), amelyekből az egyes szavaknak már említett elemzése alapján a fordított szöveggel párhuzamosan feljegyzéssort készítünk. Ez a feljegyzéssor vagy információsor nemcsak egy-egy adott szó fordításakor volt felhasználható, de ennek alapján (mivel a fordított szöveggel párhuzamosan haladt) esetleg szórendi vagy morfológiai változtatásokat is lehetett a már lefordított, de még ki nem gépelt szövegen eszközölni.

A mondatok nyelvtani felépítettségének bonyolultsága és még inkább az orosz—magyar gépi fordításban rendkívül fontos szerepet játszó szórendcsere vezetett arra a gondolatra, hogy célszerűbb előbb a fordítandó szövegről egy kielégítő információt nyújtó nyelvtani feljegyzéssort összeállítani, amelynek alapján a magyar szöveg egyetlen fordítási művelettel összeállítható. Az orosz—magyar gépi fordítás ilyen alapon tulajdonképpen ugyanúgy két műveletcsoportra oszlik: az analízis és szintézis műveleteire, mint több, már meglevő fordító algoritmus. Az analízis során összeállítjuk az információsort, a szintézis során a végleges célnyelvi szöveget.[7]

5. A szavankénti fordítás

Az analízis és szintézis alapján történő gépi fordítás nem egyedülálló módszer, azonban sok fordító csoport ilyen alapon dolgozik. A gépi fordítás kezdeti időszakából való az úgynevezett szavankénti fordítás módszere, mely ugyan meghökkentően kezdetleges elgondolásra épül, de számot kell vetnünk vele, mert továbbfejlesztésével egy jellegzetes fordítási módszer alakult ki, amelyet előnyeinek megismerése nélkül nem mellőzhetünk. A módszer elsősorban a gépi fordítás első gondolatát felvető szakemberek beállítottságából, és annak az időszaknak technikai körülményeiből alakult ki. W. Weaver felfogása, mely szerint minden idegen nyelvű szöveg az angolnak különlegesen rejtjelezett formája, elegendőnek tart egy gépi szótárat, amellyel az idegen szöveg megfejthető.[8] Az elgondolás lényegén az sem változtatott sokat, hogy a szótárat morféma-tárrá alakították át. A fordítás során a forrásnyelv egyes morfémáit egymástól függetlenül, elszigetelve fordítja le az automatikus szótár, és a többértelműségből adódó fordítási változatok egymás mellé kerülnek. Az a tény, hogy az így nyert szósor orosz—angol viszonylatban a szakemberek számára használható fordítást adott, tekintélyt biztosított ennek a fordítási módnak, és kiindulási alapja lehetett egy fordítási gyakorlatnak, melyet végső alakjában ma már nem annyira szóról szóra fordításnak kell neveznünk, hanem inkább közvetlen fordításnak. A továbbfejlesztés tulajdonképpen a szöveg többszöri átjavítását jelenti, melyet az első fordításkor a célnyelvi szavakkal együtt a morféma-tárból vett nyelvtani kiegészítő jelzések elemzése alapján végeznek el.[9] Az első kiegészítő elemzésben a főnév a végső határ,

[7] D. Ju. Panov: Автоматический перевод (Moszkva, 1958); E. Delavenay: An introduction to machine translation (London, 1960).

[8] Machine translation of languages (Szerk. W. N. Locke—A. D. Booth, New York, 1955).

[9] Linguistic and engineering studies in the automatic translation of scientific Russian into English. Technical report, Phase I (Seattle, 1960).

ennek esetét határozza meg a program az előtte álló elöljáró, illetőleg melléknévi, névmási jelzők alapján. A kiegészítő elemzés eredményeképpen a célnyelvi szövegből kihagyják a helytelennek bizonyuló szavakat, elöljárókat, toldalékokat. A 2. és 3. elemzéssel a szöveg-környezet alapján határozzák meg a még pontosabb morfológiai adatokat, illetőleg a szintaktikai szerepet. Végeredményben egy nem teljesen egyértelmű, de kevés zavaró kétértelműséget tartalmazó célnyelvi szöveget kapunk. Az ismételten megvalósított javítások és kiigazítások tulajdonképpen szintagmatikai elemzések, melyekkel elsősorban a viszonyító morfémák többértelműségét lehet kiküszöbölni. (A lexikai többértelműség ennél nehezebb feladat elé állítja a gépi fordítást.)

6. Analízis a gépi fordításban

A továbbiakban külön kívánok foglalkozni az analízis és szintézis műveleteivel, azok jellegével és a velük kapcsolatos elméleti kérdésekkel. Az analízis a szókereséssel kezdődik és a teljes információsor összeállításával végződik. Az információsor akkor tekinthető teljesnek, ha tartalmazza a célnyelvi szó szótári helyét és azoknak az utasításoknak a programszámát, amelyekkel a végleges célnyelvi szöveget megkapjuk. Ez tulajdonképpen azt jelenti, hogy az információsor tartalmazni fogja a végleges fordítási utasítások összeállításához nélkülözhetetlen nyelvtani információkat, amelyeket a forrásnyelvi szöveg elemzése során kaptunk.[10] Az adatokat kódszámokkal rögzítjük, hogy ne kelljen állandóan a forrásnyelvi szöveghez fordulnunk. Ha az információsort a fordítás emberi munkájával akarnám összehasonlítani, azt kellene mondanom, hogy ez az információsor olyasmi, mint ami a fordító fejében van meg egy mondat elolvasása után. Érti a mondatot, de még nem fogalmazta meg a fordítást. Ebből a hasonlatból az is következik, hogy az információsornak mindazt tartalmaznia kell, ami a fordításhoz nélkülözhetetlen (célnyelvi viszonylatban is), de csak annyit kell tartalmaznia, többet nem. Az információsor összeállítását megelőzően éppen azért végig kell gondolni, milyen forrásnyelvi adatokra van szükségünk a célnyelvi szöveg összeállításához. A forrásnyelvi szöveg elemzésének szempontjait a célnyelvi szöveget figyelembe véve választjuk meg.[11]

Mint az emberi fordításban, itt is szükségem van a mondatot alkotó szóértékű morfémákra és ezek viszonyát kifejező viszonyító morfémákra. A szóról szóra fordításban az említett módon mindkét morfémacsoport előforduló egyedeit automatikus szótározással lefordítom. Az információsor esetén csak a szóértékű morfémákat fordítjuk le, a viszonyító morfémákat általában nem, csak akkor, ha a jelentésük megállapításához szükséges pontos szintaktikai viszonyt a mondat szavai között tisztáztuk. Ezt a viszonyt a szavakhoz mellékelt külön jelzések, nyelvtani kódok[12] fejezik

[10] Vö. D. Ju. Panovnak a 7. jegyzetben hivatkozott műve 32. lapjával.

[11] I. A. Melcsuk, Некоторые вопросы машинного перевода за рубежом: Доклады по машинному переводу (Moszkva, 1961)

ki. Ezeknek a nyelvtani kódoknak kell olyan bőségeseknek lenniük, hogy a fordításhoz biztos támpontot nyújtsanak.

7. A gépi szótár

Az „analízis" legelső művelete, a szókeresés, mint láttuk, abban áll, hogy a forrásnyelv minden szavát egyenként egyeztetjük a gépi szótár forrásnyelvi szókészletével, és egyező szó esetében „kiadjuk" a megfelelő célnyelvi szót. A fordítógépek szótárában általában nem teljes szavakat, hanem csak szótöveket helyeznek el toldalékok nélkül; ami azt jelenti, hogy a gépi szókeresés során az eredeti szöveg toldalékos szóalakjáról a toldalékot el kell távolítani ahhoz, hogy a szótőt megkaphassuk. Újabban képzők nélküli szóalakokkal is működnek gépek, de ebben az esetben a képzők értelmezésére külön szabályokat kell összeállítani. A gépi szótár szótövei azonban nem a hagyományos nyelvészetben megismert szótövek, hanem ún. grafikus tövek, melyek a szó minden írott alakjában változatlanul megmaradnak.[13] Míg a nyelvtani alakokat ismerő ember számára elegendő egy szó egyetlen szótövének megadása és a tanult hangtani változások alapján például a *сиж* szótőben felismeri a *сидеть* ige *сид* szótövének hangváltozásos alakját, addig a gép erre külön program nélkül nem képes és minden tővariánst külön kell elhelyezni.

A toldalékok leválasztása a gépi fordítás gyakorlatában többféle módon történhet. A legegyszerűbb eljárás az, amikor a gép a szótárban meg nem talált szó végéről egyenként betűket vág le, és minden betűlevágás után összehasonlítást végez: nem talál-e a megmaradt szóalakkal egyező tövet a szótárban. Egy másik módszer szerint a gép a hosszabb tövektől a rövidebbek felé elrendezett szótáron végigviszi a toldalékos szóalakot, és a legnagyobb egyezést mutató szótári szó lesz a keresett „tő" a megmaradó rész pedig a „rag".[14] Az orosz—magyar gépi fordításban célszerűbbnek és gyorsabbnak látszik az a megoldás, hogy a szóról előbb eltávolítjuk a teljes ragot (ha van), és a megmaradt szótövet azzal a szótárrésszel hasonlítjuk össze, amelyből a tő (megfelelő elrendezés mellett) a rag alapján várható.[15] Ezzel a szókeresésre fordított idő lecsökken.[16]

Az orosz szó utolsó betűje alapján kielégítő biztonsággal megállapítható a toldaléknak nemcsak a jelenléte, de maga a teljes toldalék is. Mégis az ilyen elemzés fenntartás nélküli érvényesítése újabb tövek bevezetését és kiegészítő elemzés leírását tenné szükségessé például ilyen szavakban, mint *предмет, калий, для, часть* stb., ezért egy kísérleti program részére célszerűnek látszik az orosz—magyar gépi szótárt két részre osztani. Az

[12] A *kódok* tulajdonképpen rendszerbe foglalt (vagy minden rendszer nélkül adott) számok, amelyek mintegy rejtjelezett formában fejeznek ki valamit (a mi esetünkben szintaktikai viszonyokat).

[13] I. K. Belszkaja, О принципах построения словаря для машинного перевода: Вопросы Языкознания 8/3. 98—106 (1959).

[14] O. Sz. Kulagina—I. A. Melcsuk, Машинный перевод с французского на русский: Вопросы Языкознания 5/5. 111—120 (1956).

[15] A. G. Oettinger: Automatic language translation (Cambridge, Mass., 1960).

[16] A szótározásnak itt ismertetett menete a gyakorlati megvalósítás során a technikai szempontok érvényesítése következtében erősen módosulhat.

első rész a „teljes alakú szavak" szótárrésze, amely az ilyen toldaléklevágás szempontjából nehezebben kezelhető szavakat, valamint a paradigmákat nem képező szavakat és az idiomatikus kifejezéseket (a közbevetés jellegűeket is) tartalmazza. A szótár második része tartalmazza a már említett grafikus töveket. A szavak elrendezésére legjobbnak az az eljárás tűnik, hogy a töveket szófajonként és ezen belül betűrendben helyezzük el. Így nemcsak a szókeresés idejét tudjuk esetleg lecsökkenteni, hanem a szótár bővítése is egyszerűbben megoldható.

A gépi szótárral összefüggő másik lényeges kérdés: milyen jellegű és terjedelmű legyen a szótár? Mivel a fordítás szakmai szövegekkel történik, szinte minden kutatócsoportban szakterületi, úgynevezett mikroszótárakat használnak.[17] Ebben a szótárban nagy számmal vannak általános jelentésű szavak, de vannak a szakterülethez tartozó különleges szakjelentésű szavak is (a gépek korlátozott befogadó képessége mellett a szavak többértelműsége is megkívánja a mikroszótárak alkalmazását). A szókészletet a szavak előfordulási gyakorisága alapján válogatják össze, de vigyázni kell arra, hogy a szó gyakorisága nem jelenti a szavaknak a szöveg értelmezése szempontjából számított fontosságát is. Ezért csak az általános jelentésű szavakat veszik gyakoriság alapján az adott szótárba.

8. A szavak többértelműsége

Mielőtt ezt a gondolatot folytatnám, ki szeretnék térni röviden a szavak többértelműségéből származó kérdésekre, mégpedig a lexikai többértelműség kérdésére. Ennek megoldása még ma sem nevezhető véglegesnek.[18] Többféle elgondolás létezik; ezekből elsőnek a legegyszerűbb megoldást említem, amikor a gép a többjelentésű szó minden jelentését egymásután kiadja, és az utószerkesztő feladata, hogy a megfelelőt kiválassza. A gépi fordítás próbálkozásai ezzel a módszerrel indultak, s bár vannak ma már célravezető eljárások, egyes esetekben az orosz—magyar gépi fordításban is ezt az eljárást fogjuk alkalmazni. Egyelőre például nincs gyakorlatilag kielégítő megoldás annak kiderítésére, hogy a mutató névmásként használt egyes szám alany- és tárgyesetű, többes szám alanyesetű személyes névmás mikor jelez személyt és mikor tárgyat. (A többi alaknál ez nem lényeges.) A kérdéssel I. N. Selimova is foglalkozott, de eredményei főleg a kérdés bonyolultságát tárták föl.[19]

A többértelműség kérdéséhez tartozik az idiomatikus kifejezések fordítása. Érdekes módon ezeknek a megcsontosodott szóalakoknak a fordításáról Bar-Hillel azt mondja, hogy kielégítő eredményt kapunk akkor is,

[17] Vö. I. K. Belszkajának a 13. jegyzetben hivatkozott művével.

[18] A. Koutsoudas—R. Korfhage: Mechanical translation and the problem of multiple meaning: Mechanical Translation 3. 46—51 (1956); T. N. Molosnaja—V. A. Purto—I. I. Revzin—V. Ju. Rozencvejg, Некоторые лингвистические вопросы машинного перевода: Вопросы Языкознания 6/1. 101—106 (1957).

[19] N. Selimova, Установление синтаксических связей предложных групп в русском языке: Лингвистические исследования по машинному переводу (Moszkva, 1961).

ha szóról szóra fordítjuk őket, mert több nyelven végzett vizsgálatai azt mutatták, hogy a tükörszavakhoz hasonlóan igen elterjedtek a tükör-idiómák is, és szóról szóra történő fordításuk a legrosszabb esetben is csak nehezen érthető lesz. Az orosz—magyar gépi fordításban nem ezt az elgondolást használjuk. Az orosz—magyar gépi szótár orosz részében külön jelzéssel látjuk el azokat a szavakat, melyek idiomatikus kifejezések kezdő szavaként szerepelnek. Minden ilyen „nyitószó" felbukkanásakor a gép a további szótározáshoz azt az utasítást követi, hogy a fordítandó szöveg következő szavát vagy szavait raglevágás nélkül a szótárban külön elhelyezett kifejezésekkel hasonlítja össze. Ha a „nyitószó" a következő szóval vagy szavakkal együtt megegyezik az idióma-gyűjteményben tárolt valamelyik kifejezéssel, akkor az ennek megfelelő fordítást kapjuk. Ha a jelzett „nyitószó" utáni szavak ilyen egyezést nem mutatnak, a szótározás az általános szabályok szerint fog tovább folyni. A többértelműségnek szókapcsolati összefüggések alapján meg nem oldható esetei a gépi fordításnak továbbra is a legnehezebb kérdéseihez tartoznak.[20]

9. Morfológiai és szintagmatikus elemzés

A szótározás során nyert magyar célnyelvi szóalak nem kerül be az információsorba a maga teljességében. Elegendő, ha ott a megfelelő magyar szó tárolójának számát helyezzük el. Esetleges későbbi elemzés céljából a forrásnyelvi szó tárolójának számát is elhelyezzük.

Az információsor legfontosabb adatai az adott szó morfológiai, szintagmatikai és szintaktikai jegyeire vonatkoznak. Az elsőknek egy részét, a nem változó morfológiai tulajdonságokat a szótárban készen találjuk. Ilyen a főnév neme, ragozási csoportja (ezen belül az is, hogy élő, vagy élettelen dolgot jelent-e), az igék folyamatos vagy befejezett volta. A változó morfológiai tulajdonságokat (például a főnév száma, esete; a melléknév neme, száma, esete; az ige módja, száma, személye stb.), amennyiben azok az adott szó elemzéséből nem határozhatók meg egyértelműen, a szövegösszefüggés elemzéséből kell vennünk. Ezt az elemzést, melyet szintagmatikai elemzésnek nevezünk, az orosz—magyar gépi fordítás programja két részre bontja. Az első részt morfológiai-szintagmatikai elemzésnek nevezném, a másodikat szintagmatikai-szintaktikai elemzésnek. A két elemzés a fordítási programnak két különböző helyére van beépítve és nem közvetlenül egymás után következnek, mint például a georgetowni (és egyéb) programban. A két elemzés szétválasztása és az elsőnek kibővítése előnyösnek látszik, mert így egyszerűbben és biztosabban juthatunk el a szó teljes grammatikai jelentésének megállapításához.

A szó teljes grammatikai jelentése, ami a helyes fordításhoz okvetlenül szükséges, legtöbbször nem állapítható meg csak az elszigetelt szó elemzése alapján, hanem a szókapcsolatokat is tekintetbe kell venni. A szókapcso-

[20] T. N. Molosnaja, Вопросы различения омонимов при машинном переводе с английского языка на русский: Проблемы кибернетики 1. 215—221 (1958); Semantic resolution: Proceedings of the National Symposium on Machine Translation held at the University of California, 1960.

latok feltárását viszont nem végezhetjük el olyan szövegeken, melyeknek szavairól még semmit sem tudunk. Éppen ezért a fordítás menetében a szintagmatikai elemzés előtt morfológiai elemzés is szerepel, melyet azonban csak az elszigetelt szavakra írnak elő a programok. A georgetowni program például a mondat szavait ábécé-sorrendbe szedi, és így végzi el a szótározást egy szintén ábécé-sorrendbe szedett szótárban. — Csak ezután következik az egész mondat terjedelmében a szintagmatikai elemzés. A washingtoni program szintén elszigetelten, csak előfordulási sorrendjükben elemzi a mondat szavait, és bizonyos grammatikai információk megadásán kívül kiírja a megtalált szótári alakhoz tartozó célnyelvi fordítást vagy fordítási lehetőségeket is.

Az orosz—magyar fordítás programja más módon igyekszik célhoz jutni. Az első morfológiai elemzés itt is a mondat szavainak sorrendjében halad, de már ez az első elemzés sem elszigetelt jellegű, hanem minden szükséges esetben számításba vesszük a már korábban elemzett szók grammatikai információit is. Joggal nevezhető ez a művelet morfológiai-szintagmatikus jellegűnek, mert a minél teljesebb grammatikai jelentés megállapítása érdekében beszámítja azokat a szintagmatikus kapcsolatokat is, melyeket a szó (elsősorban a főnév) az őt megelőző szavakkal alkot. Nyilvánvaló, hogy ezzel az első elemzéssel nem mindig fogunk kielégítő eredményre jutni, és a szó teljes grammatikai jelentését csak a második fogja feltárni. Ebben nincsenek olyan megkötöttségek, mint az elsőben: nemcsak a mondat megelőző szavaira terjed ki, hanem felöleli az egész (egyszerű vagy összetett) mondatot elejétől végig. Sorra veszi az összes lehetséges szintagmatikus kapcsolatokat, a mondat egészében érvényesülő vonzatokat. Mindez akkor lesz igazán eredményes és megbízható, ha pontos és bő morfológiai adatokra támaszkodik. Az orosz—magyar gépi fordítás programja ezt egyrészt az első morfológiai elemzésnek már említett szintagmatikus kibővítésével igyekszik biztosítani, másrészt még a szintagmatikus elemzés elé mondatelemzést is beiktat. Az összetett mondat elemzése így nemcsak könnyebb, hanem gyorsabb is lesz.

A mondatelemzés szükségességét már az első jelentősebb gépi fordítási elméleti munkák is hangoztatták; gyakorlati alkalmazására tulajdonképpen csak az utóbbi években került sor (az 1958-as washingtoni beszámoló[21] például még nem tartalmaz ilyet).

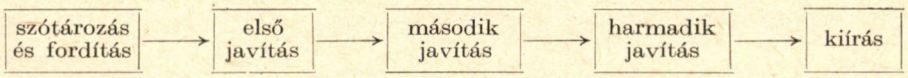
Az orosz—magyar gépi fordításban a mondatelemzés kiindulópontja, mint másutt is, az írásjel. Ha az írásjel után nem jön közbevetés, a már kiszótározott szakaszt mondatelemzésnek vetjük alá. Számon vesszük, van-e állítmány, illetőleg alany a vizsgált szakaszban, és az elemzésbe belevonjuk a szakaszt kezdő írásjel jellemző adatait is. (Ez tulajdonképpen a vessző előtti szakasz jellegéről ad felvilágosítást.) Az elemzésből két értékű eredményt kapunk a vizsgált írásjel jellegére vonatkozóan (például fölsorolás vagy megszakítás, fölsorolás vagy mondathatár stb.), amely a további szótározás alapján a vessző után álló kötőszóból, vagy igenévből, vagy akár a következő szakasz elemzéséből tisztázódik.

[21] Vö. a 9. jegyzetben hivatkozott művel.

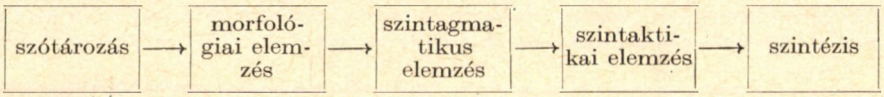
10. Szintaktikai elemzés; az információsor tartalma

Az állítmány elemzésére a következő módszert használjuk: külön kódot írunk azokhoz a szavakhoz, melyek előre megállapíthatóan a mondat állítmányaként (vagy annak valamilyen részeként) fognak szerepelni. Statisztikai valószínűséget is alapul véve az ilyen szavak és szócsoportok száma mintegy 24. A program az ilyen kódolású szavakat „kiemeli" a mondatból és külön összegyűjti. A mondatszakasz elemzésekor az a feladatunk, hogy a kiemelt szavak kódszáma és ezek kombinációja alapján megállapíthassuk az állítmányt, annak típusát és magyar fordítását. A kapott adatok a mondatszakasz elemzésekor döntő szerepet játszanak. Az utolsó mondatszakasz elemzése után visszatérünk a mondat elejére, és most már a teljes mondaton belül végezzük el a második elemzést, melynek szintagmatikai és egyúttal szintaktikai jellege is lesz már.

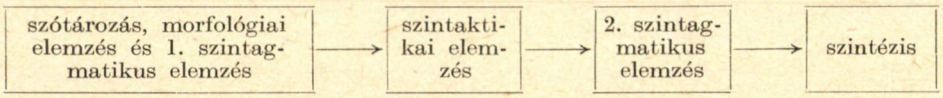
A gépi fordítás általános menete szavankénti fordítás esetén:



A fordítás menete a georgetowni csoport algoritmusa alapján:



Az orosz—magyar fordítás felépítése:



Már korábban szó volt arról, hogy az információsor nemcsak morfológiai adatokat, hanem szintaktikai-szintagmatikai kapcsolatokat is tartalmaz. Ezeket az adatokat a szavak mellé tett és a végleges fordításra vonatkozó kódok tartalmazzák. Így például két egyformán birtokviszonyban álló szó mellett különböző fordítási kódok állnak attól függően, hogy a szó milyen szintagmatikai kapcsolatban szerepel.

Az információsor tehát tulajdonképpen kettős: az egyik a szavakra vonatkozik, a másik mondatszerkezetet ad. Az egyes szavakra vonatkozólag a következőket tartalmazza:

(1) a forrásnyelv szavának tárolószáma;
(2) a forrásnyelv szavának alapvető morfológiai adatai;
(3) a forrásnyelv szavának további morfológiai adatai;
(4) a forrásnyelv szavának szintagmatikai adatai;
(5) a magyar megfelelő tároló-száma.

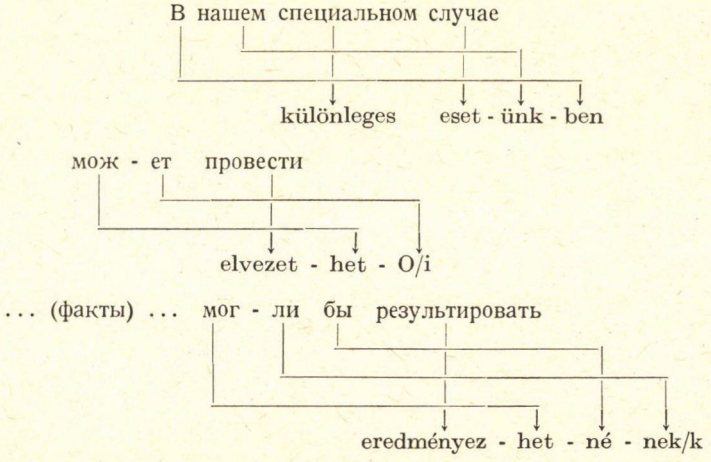
A forrásnyelvi szöveg egyes mondattani szakaszaira vonatkozólag a következő adatokat tartalmazza az információsor:

(1) a szakasz „kezdőszavának" tároló-száma;
(2) a szakasz mondattani jellege;
(3) a szakaszban található állítmány(ok) típusa;
(4) az állítmány(ok) tárolószáma;
(5) a szakaszban található alany(ok) jellege;
(6) a szakasz „végső szavának" tároló-száma.

11. A magyar szöveg szintézise

Az információsor összeállításával az analitikai műveletek sora tulajdonképpen lezárult. A szintétikus műveletek legfontosabb kérdése az orosz—magyar gépi fordításban a szórend kialakítása. Nem minden gépi fordításban ez a legfontosabb kérdés. Az orosz—angol programban a szórend kérdését például harmadrendű kérdésnek tartják, de csak saját viszonylatban, mert például az amerikai Reifler már igen korán megjegyzi,[22] hogy az automatikus szótárral nyert szóról szóra fordítás csak akkor használható, ha a morfémák fordítási sorrendje a célnyelv természetes szórendjéhez igazodik. Az orosz—magyar gépi fordítás kérdésében éppen a helyes magyar szórend biztosítása jelentette a legnagyobb, mondhatnám döntő jelentőségű feladatot.

Az orosz és a magyar nyelv közötti szórendi különbségből kifolyólag a szöveg megértése szempontjából zavarólag hat, hogy a ragozó nyelvtípushoz tartozó magyar nyelv több orosz szókategóriát toldalékokkal fejez ki. A program ezeket a szavakat „toldalék értékű grammatikai viszony-



4. ábra

[22] Lásd a 9. jegyzetben hivatkozott gyűjteményben.

szók'' címen külön utasításrendszer alapján dolgozza fel. Az orosz—magyar gépi fordításban ilyennek számítanak az elöljárók, a birtokos névmások, a lehetőséget és a műveltetést kifejező igei segédszók, valamint a feltételezést kifejező бы partikula. Az ezeknek megfelelő ragok, illetőleg képzők az ige, illetőleg a főnév toldalékai közé kerülnek a magyar nyelvhelyesség sorrendjének megfelelően. A toldalék értékű grammatikai viszonyszók közül a birtokos névmás a többes szám jelét is megváltoztatja. (Lásd a 4. ábrát.)

A grammatikai viszonyszóknál nehezebben megoldható feladatot adnak bizonyos jelzős szókapcsolatok. Ennek szemléltetéséhez gondolnunk kell arra, hogy a magyar mondatban a jelzők közül csak a ragozatlan jelzőnek van kötött szórendi helyzete (melléknévi, számnévi, névmási jelző, ragozatlan birtokos jelző). Ezek ugyanis közvetlenül a jelzett főnév előtt helyezkednek el. A ragos birtokos jelző már például a jelzett szó mögé kerülhet közvetlenül, vagy akár távolabb is. Ha a birtokszónak melléknévi igeneves szerkezete van, akkor a birtokos szó szintén távolabb áll a birtokszó előtt, és mindig ragos.

A magyar mondatban a szerkezetileg a ragozott igéhez tartozó szavak a mondatban bárhol elhelyezkedhetnek, és mindig az igéhez fognak kapcsolódni. Ez a szórendi kötetlenség vonatkozik a különböző határozóragos főnevekre is, amelyekkel az orosz—magyar gépi fordítás szempontjából részletesebben kell foglalkoznunk. A magyar mondatnak ez a tulajdonsága sajátságos, hiszen más indoeurópai nyelvekhez hasonlóan az oroszban is például egy elöljárós vagy esetragos főnév nem mindig a mondat igei állítmányával alkot szószerkezetet, hanem az őt megelőző főnévvel vagy a megelőző főnevek valamelyikével.[23] Ugyanakkor az ilyen hátravetett, nem egyeztetett jelzőt adó főnév alaktanilag semmiben sem különbözik az igéhez tartozó határozótól, illetőleg a közvetett tárgyi kiegészítőtől.

Az orosz—magyar gépi fordítás során az ilyen mondatok miatt két nehézséget kell legyőznünk. A birtokos jelzőt szórendileg a jelzett szó elé kell helyezni. A nem egyeztetett, hátravetett határozói jelzőket azonosítani kell és szükség esetén szintén a jelzett szó elé kell vinni jelzői determinánsként. (Nem okvetlenül szükséges a határozói jelzőt szórendileg áthelyezni az igei állítmányt nem tartalmazó mondatokban, sőt bizonyos igés mondatokban sem.[24]

Az oroszban a birtokos jelzőt adó főnév néha egyedül, néha bővítményekkel áll a jelzett szó után, amelynek szintén lehetnek saját bővítményei. Mindez az 5. ábra szórendi változtatásait teszi szükségessé.[25]

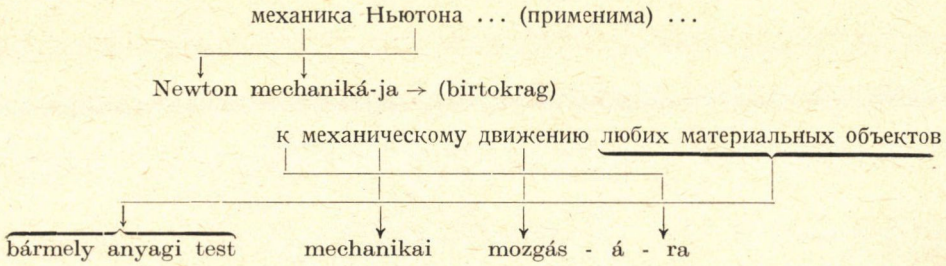
Az ilyen jellegű szórendi változtatásokhoz az orosz—magyar gépi fordításban külön meg kell alkotni a birtokszó, majd a birtokos szó melléknéves, névmásos, illetőleg számneves jelzős szókapcsolatát, hogy a szórendcserét a teljes szókapcsolattal lehessen elvégezni. A forrásnyelvi szó-

[23] Papp Ferenc, Трансформационный анализ русских присубстантивных конструкций с зависимой частью-существительным: Slavica: 1. 55—83 (Debrecen, 1961).

[24] Hell György—Sipőczy Győző: Orosz műszaki szövegek fordítása (Műszaki egyetemi jegyzet, Bp., 1961).

[25] Hell György, О нескольких характерных чертах программы МП с русского языка на венгерский: Slavica 2. 151—64 (1962).

szerkezetek földerítéséhez és a célnyelvi megfelelő kialakításához az úgynevezett „nyitó- és zárószavak rendszerét" használjuk.



5. ábra

12. A nyitó- és zárószavak rendszere

A nyitó- és zárószavak rendszere azoknak a szavaknak, szófajoknak és szabályoknak az összességét jelenti, amelyekkel a gépi fordításban meg tudjuk határozni az összetartozó szavak kapcsolatának kezdetét és végét. Minden nyitószó a szószerkezetek megalkotására szolgáló program kezdetét jelenti, melyet a zárószó fejez be. A program a szókapcsolatok megalkotása mellett megadja a szószerkezetek egymáshoz való viszonyát is, amelynek alapján a szórendi változtatás történik. Mivel a szószerkezetek megalkotásával egy időben a szók a mondattanilag helyes ragokat is megkapják, a nyitószók megválasztása a magyar nyelvi sajátságoknak megfelelően történt, és az elöljáró, valamint a többi grammatikai viszonyszó nem számít nyitószónak.

A szószerkezetek nyitó- és zárószavait szófaji kategória, vagy egyedi tulajdonság alapján kapjuk:

Nyitószók:

(1) melléknév: *a)* melléknévi jelzős főnevekhez,
b) mellékneves szerkezetekhez,
c) középfokú hasonlító szerkezetekhez,
d) főnévi igeneves szerkezetekhez;
(2) melléknévi és határozói igenevek igeneves szerkezetekhez, illetőleg a melléknévi igenév jelzős főnévhez;
(3) névmások: egyeztetett jelzős főnevekhez;
(4) számnevek (tő- és sorszámnevek) egyeztetett jelzős főnevekhez;
(5) főnév: *a)* birtokviszonyhoz,
b) hátravezetett nem egyeztetett, jelzős szerkezetekhez,
c) főnévi igeneves szerkezetekhez;

(6) bizonyos igék főnévi igeneves szerkezetekhez;
(7) határozószók melléknevekhez és melléknévi igeneves szerkezetekhez;
(8) módosító szók melléknévi igeneves szerkezetekhez;
(9) felsoroló kötőszó és vessző a hasonlító és magyarázó kötőszókhoz.

Látható, hogy csaknem minden szófajhoz tartozó szó szerepelhet nyitószóként. Az elöljáró nem nyitószó, mivel egy orosz elöljárós főnév a magyar célnyelv szempontjából nem ad szószerkezetet.

A nyitószavak rendszere a fordításban előreható tényezőként működik: egy nyitószó megjelenése olyan programozás kezdetét jelenti, mellyel fölkészülünk a nyitószóra jellemző szószerkezet fogadására. Ezzel az eljárással kettős eredményt érünk el: egyrészt a szószerkezetek kialakítását a sorrendileg elöl álló tagtól kezdjük meg, és nincs szükség a később jövő alaptagból kiinduló bonyolult visszakeresésre; másrészt a szószerkezet be nem következése esetén értesülünk arról, hogy az adott szónak esetleg különleges szerepe van a mondatban. Az utóbbi információ különösen a melléknevek és névmások esetében hasznos, mert a várt szókapcsolat elmaradása azt jelzi, hogy a teljes alakú melléknévnek esetleg állítmányi szerepe van, vagy pedig főnévként használatos, illetőleg főnév után álló, szórendileg szokatlan helyzetű melléknévi jelző.

Előfordulhat az, hogy a nyitószóval jelzett szószerkezetnek még nincs vége, de egy másik szó újabb szószerkezetre ad lehetőséget (például: birtokviszony, hátravetett határozói alakú jelzős szerkezet, igeneves szerkezetek stb.). Ilyen esetben a szószerkezet akkor tekinthető lezártnak, ha a legutolsó zárószó is sorra került már.

Zárószók: Minden olyan szó, melyhez nincs nyitószó, azaz, amely a megelőző szóval nem alkot szórendileg kötött szókapcsolatot, zárószónak számít. A zárószó olyan értelmű információt ad, hogy az őt megelőző szó egy (esetleges) szerkezet utolsó tagja. A leggyakoribb zárószók:

(1) személyes névmások;
(2) ragozott igék;
(3) határozószók, ha nem melléknévvel alkotnak szókapcsolatot;
(4) nem nyitószót jelentő vessző.

A nyitószók és zárószók, mint már említettem, nemcsak az összetartozó szószerkezeteket adják meg, hanem azoknak egymáshoz való viszonyát is. A gépi fordítás során a lefordított szavak nem közvetlenül a forrásnyelvi szöveg szavainak sorrendjében épülnek bele a célnyelvi szövegbe, hanem először különleges elrendezésű tárolókba kerülnek, ahol a nyitószók és zárószók szabályainak segítségével földerített mellé- és alárendelő kapcsolatoknak megfelelően helyezkednek el. A zárószó felléptével a szavak (most már mondattanilag helyes toldalékokkal) minden alkalommal ugyanabban az előre meghatározott sorrendben kerülnek ki ezekből a rendező tárolókból; a behelyezés és a kivétel együttes eredményeként a helyes magyar szórendet kapjuk.

A nyitó- és zárószavakkal a birtokos jelzős szerkezetekhez hasonlóan oldjuk meg a hátravetett határozói alakú jelzők fordítását is. Ezekben a szerkezetekben különös nehézséget csak a hátravetett jelző azonosítása jelent; ennek elmulasztása nem egyszerűen magyartalanságot eredményez, de gyakran félreértésre is vezet.[26] Az ige nélküli mondatok közül elsősorban a címekben nincs szükség ilyen célú elemzésre, mert ige hiányában a hátravetett határozói alakú jelző a magyar mondatokban is olyan szókapcsolatot alkot, mint az oroszban. Az igei állítmány mellett azonban elvileg minden főnév után álló és vesszővel nem elválasztott elöljárós, illetőleg esetragos főnevet kiterjedt elemzésnek kell alávetni, hogy pontos szintagmatikus kapcsolatát megállapíthassuk. Ez gyakran igen körülményes feladat, amit, ha a főneveknek csak alaktani tulajdonságait vesszük tekintetbe, sokszor nem is tudunk megoldani. Így például tisztán formai eszközökkel nehéz eldönteni, hogy a „... были исследованы ... в двигателях с воспламенением ..." szósorban a „... с воспламенением ..." elöljárós főnév az állítmány határozója-e, vagy pedig (a magyar nyelvtani terminológia jelölésével) az előtte álló főnév határozói alakú jelzője.

13. A szószerkezetek transzformációja

Az orosz—magyar gépi fordításban eddig általános szabálynak tekintettük azt a tételt, hogy szóról szóra fordítunk, vagyis az orosz szöveg szófajait a toldalék értékű grammatikai viszonyszóktól eltekintve ugyanolyan szófajjal adjuk vissza a fordításban. A szószerkezetekre vonatkozóan ez az elv azt jelenti, hogy az orosz nyelv szókapcsolatai a szórendtől eltekintve ugyanolyan szókapcsolatokat adnak a magyarban is. A szószerkezetek egyszerűbb eseteitől eltekintve ez az elv az orosz—magyar gépi fordításban éppúgy nem valósítható meg, mint az egyéb, már meglevő fordítási rendszerekben. Több fordítási központ külföldön a szószerkezetek fordítására szókapcsolat-mintákat állított össze, melyek a forrásnyelv szószerkezeteihez hozzákapcsolják a célnyelv megfelelő szókapcsolatait, és a fordítás ezeknek a mintáknak az alapján történik. T. N. Molosnaja az angol—orosz fordításhoz kb. 7 000 szószerkezet-mintát használ fel, melyek nemcsak a fordítási lehetőségeket adják meg, hanem az angol szöveg szófaji és szószerkezeti elemzésére is ezek a minták szolgálnak.[27]

Az orosz forrásnyelv gazdag alaktana feleslegessé teszi számunkra a szókapcsolatoknak ilyen módszerrel való tisztázását, mert a szavak alaktani jegyeiből nemcsak azok szófaji hovatartozására, de egymással alkotott szintagmatikus kapcsolataira vonatkozóan is elegendő információt kapunk. Mégis bizonyos szószerkezeti típusokra, rendszerünkben a nyitó-

[26] Például: *Были исследованы процессы сжигания топлива в двигателях с воспламенением от сжатия.* a határozóragos főnevek lehetséges jelzői szerepének figyelembevétele nélkül: „*Vizsgáltuk az üzemanyag égésfolyamatait a motorokban sűrítéses gyújtással.*" — A nyelvtanilag helyes fordítás: „*Vizsgáltuk az üzemanyagnak sűrítéses gyújtású motorokban végbemenő égésfolyamatait.*"

[27] T. N. Molosnaja, Некоторые вопросы синтаксиса в связи с машинным переводом с английского на русский: Вопросы Языкознания 6/4. 102—110 (1957).

szók bizonyos csoportjaira és a szószerkezetek bonyolult eseteire az orosz—magyar gépi fordításban is célszerű alkalmazni transzformációs szószerkezet-mintákat. A cél mindenesetre az, hogy a nyitószók alapján közvetlen programozással (minták nélkül) oldjuk meg a transzformációt, hogy ne legyen szükség az eléggé időrabló szerkezet-azonosításra és utólagos átalakításra. A közvetlen programozás azt jelenti, hogy bizonyos szintagmaszerkezeteket előre megadott transzformációval fordítunk, de úgy, hogy a transzformációs fordítást még a szerkezet teljes egészének azonosítása előtt megkezdjük. A megfelelő nyitószó alapján ugyanis már kezdetben is elegendő információt kapunk a szerkezet jellegéről és a program a fordítás további menetére kielégítő utasítást fog nyújtani.

Közvetlenül a nyitószó alapján alakítjuk át például a következő szerkezeteket: *a)* hátravetett igeneves szerkezeteket; *b)* főnévi igeneves szerkezeteket. Például: *Силы, преобразующие природу . . .* : *Az erők, amelyek átalakítják a természetet . . . . . позволяет переводить текст . . .* ; *. . . lehetővé teszi, hogy a szöveget lefordítsuk . . .*

Szószerkezet-mintákkal oldjuk meg

a) a négy-ötszörös birtokviszonyokat,

b) a bonyolult, birtokviszonnyal bővített, hátravetett jelzős szerkezeteket.

Például: *В случае подробного рассмотрения движений точек твердого тела . . .* Az igei alapjelentést kifejező főnév igévé való átképzésével: *Abban az esetben, ha részletesen megvizsgáljuk a merev test pontjainak mozgását . . . Метод определения понятия количества информации . . .*; Szóösszetétellel és melléknév-képzéssel: *Az információmennyiség fogalmának meghatározási módja . . .*

Terjedelmes hátravetett határozói jelzős szerkezet fordítása (a határozói jelző a többszörös birtokviszony alaptagjára vonatkozik): *При изображении строения атомов с указанием только энергетических уровней . . .* : A mondat fordítása transzformáció nélkül: *Az atomok szerkezetének csak az energetikai szintek feltüntetésével való ábrázolásakor . . .* — A mondat fordítása transzformációval: *Ha az atomok szerkezetét csak az energetikai szintek feltüntetésével ábrázoljuk . . .*

14. Néhány sajátságosan magyar kérdés

A gépi fordítás már eddig megvalósított algoritmusaihoz hasonlóan a fordítás szintétikus műveleteihez, amikor a gép a célnyelvi szöveget állítja össze, tárolnunk kell a magyar szavaknak mindazon alaktani tulajdonságait is, melyek a végleges fordított szöveg generálásához szükségesek. A magyar nyelv toldalék-gazdagsága következtében a ragozható szavakhoz, különösen a névszókhoz (a főnévi igenevekhez is) sokféle alaktani ismertető jegyet kell adni. A névszótövek sokféleségének jelzése a magyar célnyelvű fordításnak különleges vonása. Alaktani ismertető jegyeket nemcsak a szavakról, hanem a ragok bizonyos csoportjáról is fel kell jegyezni aszerint, hogy melyik tőhöz milyen módon (kötőhangzóval vagy anélkül) járulnak.

A magyar nyelv gazdag alaktani sajátságai már a fordítás kezdetén felvetik azt a kérdést, hogy a forrásnyelvből elegendő információt kapunk-e mondattanilag helyes fordítás megalkotásához. Bár az orosz nyelv szintén sok alaktani vonással rendelkezik, vannak olyan fordítási kérdések, amelyekre csak igen kevés információt, illetőleg semmilyen információt sem kapunk. A legnehezebben megoldható információs kérdéseket a már említett hátravetetett határozói alakú jelzős szókapcsolatok egyértelmű megállapítása mellett a névelő használatának, a tárgyas ragozásnak, a visszaható és szenvedő igék megkülönböztetésének, az igekötő elválasztásának kérdései jelentik. Mindezek az orosz — magyar gépi fordítás megoldásra váró feladatai közé tartoznak.

Kónyi Sándor

A magyar—orosz gépi fordítás néhány problémája

Tanulmányom címe: A magyar—orosz gépi fordítás néhány problémája. Ez — legalább is szerintem — azt jelenti, hogy csak n é h á n y problémát érintsek. Azonban menet közben (tegnap) kiderült, hogy „kívántatik" néhány szó a magyar—orosz gépi fordításról á l t a l á b a n is. Egészen röviden tehát erről is!

1. Melcsuk úttörő munkája a magyar—orosz gépi fordításban

Ismeretes, hogy milyen döbbenetes hatást tett nálunk, amikor híre jött, hogy Moszkvában magyar nyelvről is fordítottak.[1] Ha ugyan volt, aki megdöbbenjen . . . Mindenesetre döbbenetes volt a c s e n d, amely ezután következett.

Mi késztette Melcsukot arra (mert ő volt a fordító; előtte csak Richens és Booth[2] fordított le összesen húsz nyelv közt egy magyar mondatot is, — de ez egészen primitív próbálkozás volt), hogy a francia után a magyart vegye sorra, holott ő maga gyenge magyar nyelvtudással rendelkezik?

Az, hogy a magyar ragozó nyelv, mint a Szovjetunió sok nemzetiségének a nyelve, az, hogy a magyarban gyakoriak az összetett szavak és az elváló igekötők, mint a német nyelvben. — A magyarban speciális probléma az, hogy míg más „boldogabb" nyelvben a szótő megtalálása utáni maradék mindig r a g (elnagyolva mondom persze), addig a magyarban a maradék személyrag is lehet, viszonyrag is, esetleg a kettő együtt: *ház-unk-ban*, vagy az összetett szó utótagja: *ház-tető*. — Míg oroszról angolra (lásd az Egyesült Államokban Harper[3]) vagy franciáról oroszra (lásd Kulagina és

[1] I. A. Melcsuk, О машинном переводе с венгерского языка на русский: Проблемы кибернетики 222—64 (1958).

[2] R. H. Richens—A. D. Booth, Some methods of mechanized translation: Machine translation of languages (New York—London, 1955) 24—46. — A magyar mondat a következő: „Az apróbogyójú fajták, úgy termésmennyiségben, mint száraz anyaghozamban, felülmúlják a nagybogyójú fajtákat." (37.)

[3] Kenneth E. Harper, A preliminary study of Russian: Machine translation of languages 66—85.

Melcsuk[4]) — legalább is kezdetben — lehetett fordítani a szórend problémájának elhanyagolásával is (Harper semmit, Kulagina és Melcsuk nagyon keveset változtatott a szórenden), addig oroszról magyarra, vagy magyarról oroszra fordítva, a szavakat teljesen össze-vissza kell csereberélni, mert különben alig akad érthető mondat. — A magyar főnevek tőtára majdnem kizárólag a szótári alakokat tartalmazza (aránylag ritka a *szó : szavak, bokor : bokrok*, még ritkább a *teher : terhek* típus), szemben az orosz főnevek tőtárával, ahol csak a zéró végződéses *стол* típus szerepel szótári alakban, míg a többi csak a gépi értelemben vett (grafikus) tőalakban fordul elő: *знани-, черт-* stb.

Melcsuk olyan tökéletesen kidolgozta a magyar—orosz gépi fordítás algoritmusát, hogy azon jóformán még annyi rés sem akad, amelyen át a műhelytitkokba bepillanthatnánk. Csak helyeselni lehet mind az alapelveket, melyekre épít, mind a kidolgozást. (A magas- és mélyhangrendű szavak megkülönböztetését; a váltakozva való fordulást hol a tőtárhoz, hol a toldaléktárhoz; a főnévi és melléknévi toldalékok két oszlopos elrendezését; az igei toldalékok közt a magas hangrendű magánhangzók, illetőleg az *s, sz, z* átváltoztatását, a kiegészítő utalások, „sajátosságok", posztpozíció-kulcsok alkalmazását; valamint azt a lényeges megállapítást, hogy a gépi fordítás első és legfontosabb lépése a szintagmák tagjainak összekeresése kell, hogy legyen, stb.)

Melcsuk ugyan nem tart igényt sem teljességre, sem véglegességre, mégis meg kell állapítani, hogy ha valaha szélesebb alapokon akarunk magyar—orosz gépi fordítást végezni, ezt az általa tört utakon indulva kell tennünk, mivel módszere továbbfejlesztésre is kiválóan alkalmas.

Mivel Melcsuk a kísérleti, első próbálkozásokhoz képest eléggé részletesen kidolgozta a magyar—orosz gépi fordítás algoritmusát, nem kívánom az ő cikkét részletesen bemutatni, hanem igyekszem — tőlem telhetőleg — valami újat mondani. Teszem ezt két irányban: *a)* a Melcsuk-féle cikkhez fűzött kritikai észrevételek formájában; *b)* továbbfejlesztve Melcsuk munkáját, egy kiragadott problémát részletesen tárgyalva.

2. Kritikai észrevételek Melcsuk magyar—orosz algoritmusához

Megvallom őszintén, engem személy szerint érdekelt a dolog, hogy lehet-e hibát találni egy gépi fordító tételeiben, hiszen ő kísérletileg, a géppel ellenőrizheti magát. Úgy vettem észre, lehet!

Hangsúlyozni kívánom azonban, hogy kritikai megjegyzéseimmel a Melcsuk által vizsgált körön belül maradok, vagyis sohasem vetem föl a kérdést például olyasféleképpen, hogy „mi lenne, ha ezt vagy azt a táblázatot t o v á b b i toldalékokkal bővítenők?"

[4] O. Sz. Kulagina—I. A. Melcsuk, Машинный перевод с французского языка на русский: Вопросы Языкознания 5/5. 111—21 (1956).

2.1. A főnévi toldalékok táblázatában (233[5]) az *-ai, -ei* toldaléktömb egy darabban nem szerepel, hanem csak különválasztva *-a, -e* egyes szám 3. személyű birtokos személyragra és *-i* többesjelre. Nyilván úgy gondolja Melcsuk, hogy a kettő összeadásából jön ki a helyes eredmény: egyes szám 3. személyű birtokos több birtoka. (Melcsuk példái: *módosulásaival, nehézségeire*, de hiányzanak az olyan példák, mint *fái, kapui.*) Ennek az eljárásnak azonban (megfelelő példák beiktatása esetén) az lenne a következménye, hogy például — mivel a táblázat szerint *-i* és *-k* egyaránt csak t ö b b e s s z á mot jelent — a *hajói* szó teljesen egyenértékűvé válna a *hajók* szóval, vagyis a *hajói* szó nem utalna birtokosra. Ha azután olyan főnevekkel is szemközt találná magát a gép, amelyek végén például az *-aink* toldaléktömb van, a toldaléktömbnek ez a felbontása egyáltalában nem lenne tanácsos, mivel a gép egyszerre elemezne ki belőle egyes szám 3. személyű birtokost *(-a)*, többes számot *(-i)*, meg többes szám 1. személyű birtokost *(-nk)* ; s így ellentmondásba keveredne.

2.2. Az igei toldalékok táblázatában (235) az azonos alakú toldalékok megkülönböztetése nem eléggé teljes, például nincs tekintetbe véve, hogy az ikes igék *-ik* ragja nemcsak egyes szám 3. személyben fordul elő, hanem — tárgyas ragozás esetén — többes szám 3. személyben is: *ők eszik.*

Ha a *-t* időjelet összeadjuk az *-ák, -ak, -a, -uk, -unk* személyragokkal, akkor helyes eredményre jutunk, azonban ha a *-j-*t adjuk össze az itt számbajöhető *-ák, -a, -on, -uk, -unk* ragokkal, akkor megállapíthatjuk, hogy egyedül a *-jon* mutat (helyesen) felszólító módot; a *-junk* nem lehet kijelentő módú, csak felszólító; a *-ják, -ja, -juk* kijelentő módú toldalékokul szerepelnek a táblázatban, de mivel e toldalékokban egybeesik a kijelentő és felszólító mód, ezt a VIII. A. 1. б. pontnak (247) mind a háromnál tekintetbe kellene vennie. (Az említett pont ugyanis csak a *-j-uk* toldalékkal számol az esetleges grammatikai homonimia szempontjából.)

2.3. Ugyancsak az igei toldalékok táblázatában (235) nyilván tévedésből szerepel a *-janak, -junk* és *-on* toldalékok után a megjegyzés-rovatban az, hogy utánuk *бы*-t kell tenni, ugyanis ez csak a feltételes módú *-nák, -nánk* stb.-re áll *(они читали бы* stb.); a felszólító módú *-janak, -junk* és *-jon* orosz megfelelője (mellékmondatban a múlt idő) után azonban nem kell *бы*-t tenni, hanem az előtte álló *hogy* kötőszót kell *чтобы*-val fordítani. — Ugyanez vonatkozik az igék fordításának szabályairól szóló fejtegetésre is (247).

2.4. Az igéknél általában beválik, ha a toldalékban az *e*-t *a*-ra, *ö*-t *o*-ra, *ő*-t *ó*-ra, *ü*-t *u*-ra, sőt az *s, sz, z* betűket *j*-re változtatjuk (mivel hiányoznak az 1. és 2. személyű *-em, -ek, -ed, -itek* stb. ragok, amelyekben ez nem hajtható sikeresen végre). Azonban meg kell jegyeznem, hogy a 3. személyű ragok között is akad néhány, amely a fenti változtatás(ok)on átesve, a gép számára a felismerhetetlenségig eltorzul. A *-jen, -sen, -szen, -zen* ragokból például *-jan* lesz *-jon* helyett, a *-sék*-ből *-ják* lesz *-jon* vagy (a felszólító

[5] A kerek zárójelben levő puszta szám az 1. jegyzetben említett mű lapszámát jelöli.

értelmű) *-ják* helyett (például *essék* és *keressék*, átforgatva a *vár* ige megfelelő alakjaivá, egyaránt *várják* lesz, *várjon* és *várják* helyett.)

2.5. A számnevek kiegészítő utalásában (245) nincs tekintbe véve az, hogy az 1, 2, 3, 4-re végződő tőszámnevek közül a 11, 12, 13, 14, illetőleg az erre végződő számnevek az orosz főnév többes számú birtokos esetét kívánják.

2.6. Az *-ú, -ű* képző után (a 226. lapon levő megjegyzéstől eltérően) állhat további toldalék (például *ellenkező irányúnak*).

2.7. A *-ható, -hető* végű igenév nemcsak állítmányi értelemben használatos (vö. 236), hanem jelzőül is.

3. A magyar szavak helyes tagolása

Főként az analízisnek két, speciálisan magyar problémájával szeretnék foglalkozni: az összetett szavak kérdésével és a ragok (s általában a toldalékok) kielemzésével. Ez a két probléma szorosan össze is függ egymással, hiszen végeredményben mindkettő a szó helyes tagolásának problémájával azonos. Mivel magyar szavak elemzéséről lesz szó, azt hiszem, ez a kérdés nem csupán a magyar—orosz fordítás viszonylatában figyelemre méltó.

3.1. Helyesen állapítja meg Melcsuk, hogy ha magyar—orosz gépi fordítást végzünk, a keresést egyszerre kell folytatni a tőtárban és a toldaléktárban, váltakozva fordulva hol az egyikhez, hol a másikhoz.

A keresés Melcsuknál lényegében a következőképpen folyik. Előbb a tőtárban keresünk, majd a szótő megtalálása után, az esetleges „maradékot" keressük a toldaléktárban. Ha a maradék nem található, megpróbáljuk első betűjét *v*-re változtatni, s úgy keresni (a *-val, -vel* vagy a *-vá, -vé* ragnak valamely hasonult alakját sejtjük a maradékban). Ha a keresés így sem eredményes — összetett szóra gyanakodva — újra a tőtárhoz fordulunk. Az összetett szó utótagját is megtalálva, ha még mindig van maradék, azt újra a toldaléktárban keressük.

A keresésnek ez a rendje csak a főnevekre vonatkozik, mivel Melcsuk egyéb összetett szókra nem terjeszti ki figyelmét. Való igaz, hogy az összetett szó a főnevek közt a leggyakoribb, s az igék közt elő sem fordul (az egyetlen *rúgkapál* kivételével); viszont véleményem szerint a későbbiek folyamán számításba kell venni az olyan összetett szavakat is, amelyek előtagja ugyan főnév, de igéből képzett főnév (vagy esetleg más szófajhoz tartozó szó): *tapasztalatcsere, szerkesztésmód, írógép* stb.

Mivel Melcsuk számol az összetett főnevekkel, a főnévi és az igei toldaléktár nála lényegesen különbözik. Az igei toldalékok közt egyszerűsítő összevonást hajt végre, amikor azt a szabályt iktatja be, hogy a toldalékban az *e*-t *a*-ra, *ö*-t *o*-ra, *ő*-t *ó*-ra, *ü*-t *u*-ra, sőt az *s, sz, z* betűket *j*-re kell változtatni. Ez nemcsak azért lehetséges, mert a **beszélnak* szóról ugyanúgy jelen idő többes szám 3. személyt stb. állapít meg a gép — a toldaléktárban való keresés után —, mint ahogy a *beszélnek* igealakról kellene meg-

állapítania, hanem azért is, mivel — ahogy már említettem — összetett igék nincsenek, és igei előtag után más szófajiságú utótag sem következhet, legfeljebb olyan szervetlen összetett szókban, mint az *adjonisten* ; — de hát az ilyenek a gépi fordítás szókincsétől csillagászati távolságra vannak.

A főnévi toldalékok között azonban ez az egyszerűsítés nem hajtható végre, mert ha itt is átforgatnók mély hangrendbe a ragokat és más toldalékokat, nemcsak olyan (aránylag ritka) meglepetések érnének, hogy a *fésütök* helyett **fésütok*-ot, *lopótök* helyett **lopótok*-ot kapnánk, hanem — elváltozott magánhangzói miatt — meg se találná a gép az összetett szó utótagját. Ezért Melcsuk a főnevek (és melléknevek) toldalékait két párhuzamos oszlopba elrendezve közli (külön a mély- s külön a magas hangrendűeket). Ez a módszer a kísérleti fordítások folyamán nagyszerűen be is vált, s mindenképpen csak helyeselhető. Így fordította le Melcsuk géppel az olyan „sima" összetett szavakat, mint például a *hangrendszereket*, de a (gépi szempontból) komplikáltabbakat is, például *nyelvelemek*. Tudni való, hogy a gép számára az olyan összetett szó komplikált, amelynek utótagja teljes egészében, vagy első betűiben (esetleg csupán első betűjében) azonos alakú valamely toldalékkal. Kisebb egyszerűsítéssel bemutatva az előbbi két összetett szón Melcsuk eljárását: A *hang* szót megtalálva a tőtárban, a mély hangrendű toldalékok között keressük a *rendszereket* szót, s mivel annak eleje nem vág egybe egyetlen toldalékkal sem, a **vendszereket* szóval való sikertelen próbálkozás után visszatérünk a tőtárhoz, ott megtaláljuk a *rendszer* szót. A hátralevő maradék fölbontása többes jelre és tárgyragra már simán megy. — A *nyelvelemek* szó utótagját a gép — utasítás szerint — szintén a toldalékok közt kezdi keresni, s ott részben meg is találja, az *-e* (birtokos személyrag) formájában. Utána tovább keresi a toldaléktárban a **lemek*-et, majd a **vemek*-et, s csak mikor ezt sem találja, állítja vissza az egész előbbi maradékot, s most már a tőtárban keresi — sikeresen — az *elemek*-et. — Eddig Melcsuk eljárása. (A *nyelvel* ige, amely a bonyodalmat okozza, nem szerepel a szótárában.)

3.2. Most már Melcsukon túltekintve — véleményem szerint — meggondolandó a következő. Ha bővül a tőtár, illetőleg kiterjedtebb fordításokat végzünk, rohamosan nő a valószínűsége annak, hogy valamely, gépi szempontból összetettnek számító szó utótagja, vagy az utótagnak legalább az első néhány betűje egybeesik valamely toldalékkal (esetleg személyragból és viszonyragból összetett toldalékkal). Ha ilyenkor a toldaléktárban kezdi keresni a gép a maradékot, két eset lehetséges: *a)* vagy megtalálja a maradékot teljes egészében (például *vaskor*), s akkor esetleg kész a téves analízis; *b)* vagy csak részben találja meg (például *szöveg-e-lemzés*). — Ha csak részben találja meg, három eset lehetséges: ezt a mostani „második maradékot" *a)* vagy egyáltalában nem találja (például *-lemzés*, a *szövegelemzés*-ből), vagy részben találja meg (például a *vég-e-re-d-m-é-n-y* szó *-redmény* „második maradékából" megtalálja a *re*-t); *c)* vagy egészben megtalálja (például *láb-i-nak*). Az *a)* pont esetében a gép még helyreállíthatja az első maradékot, s újra keresheti a tőtárban ; a *b)* pont esetében

a gép tovább robog a hamis nyomon; a *c)* pont esetében tévesen analizált (s következésképpen csak tévesen fordíthat).

Hosszabb szóval kapcsolatban talán még tovább lehetne folytatni, hogy „ismét három eset lehetséges" (hiszen például a *végeredmény* szó utótagjából a gép nem kevesebb, mint h a t toldalékot tud kielemezni).

3.3. Az ilyen példákat látva, az a gondolatom támadt, hogy talán jobb lenne változtatni a keresés sorrendjén: keresse a gép a szó megtalálása után (legalább is akkor, ha az főnévnek bizonyult), a maradékot újra a t ő t á r b a n, s csak ezután forduljon szükség esetén a toldaléktárhoz! Ennek az eljárásnak persze vannak előnyei, s vannak hátrányai. Előnyei főként az összetett szavak elemzésében mutatkoznak meg: az olyan gépi szempontból komplikált összetett szavak is zökkenő nélkül analizálhatók, mint az *áramátalakító*, *kendődarab*, *alapokmány*, *hajójárat* stb.

Azonban kézenfekvő a legfőbb ellenvetés. A gépi szempontból összetetteknek számító szók a szókincsnek aránylag kis százalékát képezik, s nagy lenne az időveszteség és az a kerülő, amelyet a tőtáron keresztül tennénk, hiszen kétségtelen, hogy a szótőről leváló „maradék" legtöbbször toldalék szokott lenni, nem pedig újabb tő.

3.4. Lenne ugyan egy univerzális megoldás, amely jóformán minden bajnak orvossága lehetne: n e m a n a l i z á l n i az összetett szavakat, hanem egységként kezelni őket. Így cselekszik Melcsuk is bizonyos esetekben: a *nyelvtudomány* szót például nem bontja elő- és utótagra (mint a *nyelvelem* szót), hanem egy darabban szerepelteti a tőtárban. Azonban ha általános szabályként kimondanók, hogy „nem ismerünk összetett szót!" — esztelen pazarlást vinnénk végbe a gép befogadóképességével. Ez magyar—orosz viszonylatban is áll, de még inkább vonatkozik például a magyar—német gépi fordításra. Ez a megoldás tehát nem jöhet számításba. Azonkívül arra is kell gondolnunk, hogy valamikor talán túljutunk a kísérleti fordításokon, s akkor a gép már nemcsak kész, megszokott, minden szótárban megtalálható összetett szavakkal találja magát szemben, hanem alkalmilag összeállt szavakkal is. Végül: ez az univerzális megoldás nem is orvossága m i n d e n bajnak, ugyanis a homonimákkal így sem tudna megbirkózni a gép. A *karó-ra* ~ *kar-óra*, *fal-ó* ~ *fa-ló* így is megmarad problémának. Azonban a homonimák megfejtése mindenképpen egy külön szabályrendszer feladata.

3.5. Mégis csak helyesebb hát Melcsuk eljárása; azonban — úgy gondolom — tökéletesítésre szorul. A maradékot a toldaléktárban kell keresni, azonban — véleményem szerint — az analízis alkalmával figyelembe kell venni a s z ó e l e m e k k a p c s o l ó d á s á n a k s o r r e n d j é t é s m ó d j á t. A szóelemek kapcsolódásának sorrendjét nem nehéz néhány szabályba foglalni, s ha ezekre „megtanítottuk" a gépet, játszva leleplezi az esetleges „értelmetlenségeket". Ezután már csak egy utasításra van szükség, hogy értelmetlenség esetén pedig állítsa vissza az egész maradékot, s keresse a tőtárban. Így például a *vasakaratnak* szót elemezve, az *-ak* többes szám-jel után a gép *-a* birtokos személyragra bukkan. Ez a

szabályok értelmében nem lehetséges, ezért visszaállítja az egész *akaratnak* maradékot, és a tőtárban keresi. A *hajójárat* szó maradékából előbb egy *-ja* birtokos személyragot, majd egy ragot *(-ra)* elemez ki a gép; ez eddig még szabályos, de azután egy újabb rag *(-t)* következne, — ez nem lehetséges, tehát visszaáll az egész *járat* maradék. — A *végeredmény* szóval kapcsolatban a *d* betűig jutva állapítja meg a gép, hogy az *eredmény* maradékot a tőtárban kell keresnie, ugyanis a *-d* már második (!) birtokos személyrag lenne. — A *kendődarab* szó elemzésekor az értelmetlenséget abban találja a gép, hogy az **-arab* „második maradék" egyáltalában nem található (se egészben, se részben) a magas hangrendű toldalékok között. — A *viharhozó* szóban a *-hoz* rag után szokatlanul *-ó* képző következne.

3.6. Ezzel az itt ismertetett módszerrel — megítélésem szerint — olyan szorosra lehet húzni az ellenőrzés hálóját, hogy azon már csak a homonimák jutnak át, továbbá a *kor* főnévvel azonos alakú *-kor* rag (az *órakor* és *vaskor* közös *kor* elemének megkülönböztetését csak jelentéstani alapon lehet megoldani). Rendkívül ritkán számításba jöhet itt még a *-tok* és *-tök* birtokos személyrag: az *ollótok* például még elképzelhető összetett szónak, a *hajótok* már alig. Általában azok a toldalékok, amelyek valamely értelmes magyar szóval azonos alakúak, többnyire jelentéstani, néha ragozástani okok miatt nem jöhetnek számításba összetett szó utótagjaként. Ide értem a személyragból vagy többesjelből és viszonyragból összetett toldalékokat is. Nem tudnék például olyan összetett főnevet mondani, amelyben a gyakori *-ok* többes jellel azonos alakú *ok* főnév (ragozástani alapokon le nem leplezhetően) utótagul szerepelne. Az *ében* szó csak előtag lehet *(ébenfa)*, utótag nem; „bátran" analizálhatja ezért a gép például a *szellemében* szót. Talán még két szó marad kétértelmű: a *népének* (habár elképzelhető *legényének, gyermekének* is), meg a *lábinak* (ez utóbbi csak addig, míg a szóelemek kapcsolódásának módját nem veszi figyelembe a gép.) Végül még van egy rag, illetőleg igekötő, a *be*, amely ellenáll az elemzésnek, mert az egyik összetett szóban az előtaghoz, a másikban az utótaghoz tartozik: i g é n y b e*vétel* ~ *termény*b e v é t e l. Ezek egyedi elbírálást kívánnak. (Majdnem hasonlóan viselkedik az *át* is, de ha tüzetesebben megvizsgáljuk, ez mindig az utótaghoz tartozik: *áramátalakító ;* ha az előtaghoz tartozna, külön íródna a szó: *áramát alakító*.)

3.7. Ez a módszer analóg módon eggyel magasabb szinten is alkalmazható. Ritkán van ugyan rá szükség, többnyire olyankor, mikor valamilyen képző szól közbe. Vegyük példának az egyszer már szerepelt *nyelvelemek* szót. Melcsuk „különösebb baj nélkül" analizálta. Azonban ha a szótár bővül, és történetesen a *nyelvel* igét is tartalmazza, a gép „menthetetlenül" a hosszabb tőből indul ki és az i g e i toldaléktárban fogja keresni a maradékot, sőt tetézve a bajt, ott az *-em* ragot meg is találja (feltéve persze, hogy az igei toldaléktár is teljesebb a Melcsukénál).

Itt ismét a szóelemek kapcsolódásának szabályait kell elővenni, s akkor a megfelelően elrendezett igei toldaléktárból megállapítható, hogy az iktelen, tárgyatlan *nyelvel* igének nem lehet *-em* ragja, „vissza" kell tehát térni egy r ö v i d e b b t ő h ö z, azt keresni a t ő t á r b a n, s abból

kiindulva kell újra elemezni. — Hasonló az eljárás például a *szóról* ragos főnévvel, ha a szótárban ott van a *szór* ige is.

3.8. Már többször említettem, hogy figyelembe kell venni a szóelemek kapcsolódásának a módját is. Erre számtalanszor szükség van. Így különböztethetők meg egymástól disztributív analízis nélkül az *áramat* és *áramot*, az *írt* és *írott*, *fala* és *falja*, sőt *indítók* és *indítóok*, *kezét* és *kézét* — pusztán ragozástani alapon. A megfelelő jelzéseket a szócikkekben kell elhelyezni, s ezeket egyeztetni a toldaléktárakkal. — Megjegyzem, hogy esetleg magasabb fokon ezek egy részére nem lesz szükség, mert a mondattani elemzés tesz különbséget például az *írt* és *írott* között.*

* Időközben magyar fordításban is megjelent Melcsuknak az 1. számú jegyzetben hivatkozott műve. I. A. Melcsuk, A magyarról oroszra történő gépi fordítás szabályainak ismertetése: Matematikai nyelvészet és gépi fordítás a Szovjetunióban (Bp., 1964., Időszerű Műszaki Dokumentációs Kérdések, 6. szám) 147—214. — *A szerkesztők.*

Hadas Ferenc

Az orosz—magyar nyelvtani szótárról

1. A szótár felépítése és anyaga

Az orosz—magyar nyelvtani szótár[1] a következő részekből áll: előszó, tájékoztató a szótár használatához, rövidítések jegyzéke, a szótár teste, az orosz ábécé és — tartalomjegyzék gyanánt — a címszavak jegyzéke.

A Szótár a fordított szótáraknak egy különleges típusát képviseli. Azon fajta kétnyelvű szótár ez, mely alfabetikus elrendezésben közli az orosz szavaknak összes olyan végződéseit, illetőleg végződéscsoportjait, melyeknek valamilyen nyelvtani funkciójuk van, s egyben megadja ezeknek magyar ekvivalensét is. A végződések utolsó betűi a szótárban ábécésorrendben helyezkednek el, azok minden betűje pedig jobbról balra haladva szintén ábécé-sorrendben foglal helyet. A szótár az írott orosz nyelvet veszi alapul, mindig leírt betűsorokkal operál, s ennyiben több helyen eltér a hangtestre épülő nyelvtanoktól.[2]

A szótár tipografiailag azonosan tagolt könyvoldalakon helyezkedik el. Minden oldal fejléccel rendelkezik, mely mindenütt azonos rovatokat tartalmaz.

A szótár testén végigvonuló fejléc adatai a következők:

1.1. Az első hasáb a példaszó szótári alakját tartalmazza. Ennek megadását azért tartottuk fontosnak, mert így a bizonyos nyelvtani alakban álló szó szótári alakja támpontul szolgál a példaszók között nem szereplő, de a szövegben szereplő szó szótári alakjának megtalálásánál. Így például a *стола* szóalak nem szerepel a szótárban. Annak nyelvtani helyét és szótári alakját azonban könnyen megállapíthatjuk az *-a* címszó alatt szereplő *хлеба* szó alapján. A szótári alakok megtalálásában minden esetben járható az analógiás út.

A példaszó címszóban megadott végződésének általában a szótári alakban zárójelbe tett szórész felel meg. Zárójel hiánya esetén **∅**-végződésről van szó. A nyelvtani szótárban meg nem adott szó szótári alakjának kere-

[1] Hadas Ferenc: Orosz—magyar nyelvtani szótár (Bp., 1961) 164 lap. — A munka orosz nyelvű autoreferátumszerű ismertetése: О русско-венгерском грамматическом словаре: Slavica 2. 165—76 (1962).

[2] Az írott nyelvre épülő grammatikák létjogosultságára és jelentőségére nézve vö. Т. М. Nyikolajeva, Письменная речь и специфика ее изучения: Вопросы Языкознания 10/3. 78—86 (1961).

sésénél mindig a címszó nélküli szórészhez kell tenni a zárójelbe tett szótári alak végződését, hogy megkapjam a keresett szó szótári alakját (a kétnyelvű szótárban).

Zárójelet azonban a szótári alak belsejében is alkalmazunk betűkiesésnél: *од(и)н — одним*, vagy éppen akkor, ha a szó végén *-ся* áll. Így a *бо(ять)ся* esetén a *ять* betűkapcsolat zárójelbe került, mivel a ragozás során a *ять* helyébe más elem(ek) kerülnek (vö. például: *бо-ящ-их-ся)*. A példaszavak szótári alakjában — a szükséges esetekben — a zárójelen kívül különböző más jeleket is alkalmazunk. Így a kieső betű jele a ': *г'н(ать)* — *г-о-н-им*, a betűváltozást meg az adott betű aláhúzása érzékelteti: рас(ти) — рос. A szótári alak után álló szám- és betűjelzések a Hadrovics—Gáldi-féle szótár[3] ragozási táblázataira utalnak. Gyakorlatilag azonban a nyelvtani szótár ezek nélkül is elképzelhető.

1.2. A II. jelű hasábok (a példaszótól balra és jobbra) az adott nyelvtani alakban levő példaszó kifejezésbe, illetőleg mondatba ágyazására adnak lehetőséget. Mivel szótárunk gyakorlati célt szolgál, helytelen lenne, ha a példaszavakat elszigetelve mutatnánk be. Jóllehet szótárunk az elszigetelten álló szó nyelvtani jelentését is világosan meghatározza, ez mégsem elegendő a magában álló szó teljes(ebb) nyelvi funkciójának jelölésére. A nyelvtannak mint szerkezettannak számos olyan ténye van — mindenekelőtt a mondattan területén —, amelyeket semmiféle hagyományos szótári formába nem lehet önteni. Nyelvi elemek szintaktikai relációi azonban szimbólumokkal kifejezhetők és rendszerezhetők is; erre tettünk kísérletet a nyelvtani szótárban a „II." jelű hasábok igénybevételével, természetesen a nyelvtani szótár adta korlátolt lehetőségeken belül. A kifejezésbe vagy mondatba ágyazást nem alkalmaztuk olyan esetekben, amikor a nyelvi funkció magától értetődik, például a felszólító mód alakjainál.

1.3. A példaszó az 1, 2, 3, 4, 5, 6 számokkal jelzett hasábban helyezkedik el, itt adjuk meg a szó strukturális felépítését.

Az 1. alatt a szó állandó része szerepel. Az általunk így nevezett szórész nem egyezik meg a hagyományos értelemben vett tővel (основа), amely magában foglalja a tővégződést is. A tővégződést a nyelvtani szótár külön kezeli — annál is inkább, mert a szó állandó része és a tővégződés között is foglalhatnak helyet bizonyos elemek. Másrészt a tővégződést azért is külön kell kezelni, mert abban a különböző hangváltozások vetületeként — adott esetekben — betűváltozások mehetnek végbe.

Mivel a szótár célja az, hogy lehetőséget adjunk a szavak ragozott formáinak megtalálására, a szó állandó részét célszerűbb a szó ragozott alakjából levezetni, nem pedig a szótári alakból. Egyes esetekben a szó állandó része megegyezik a hagyományos tő tővégződés nélküli részével *(сто-л)*, más esetekben viszont azzal nem egyezik meg: *с ж -е-чь—с о ж-ж-ёт*

[3] Orosz—magyar szótár. Szerkesztette Hadrovics László és Gáldi László. Második, átdolgozott és bővített kiadás (Bp., 1959).

лс-ать—ш-л-ём, л о-жь—л ж-ж-и. Vannak olyan esetek is, amikor a szó állandó része attól függ, hogy annak milyen nyelvtani alakjából hámozzuk azt ki, például *м -р-у-м ё -р.*

Az orosz nyelv esetében — éppen a hagyományosan értelmezett tő elején, belsejében vagy végén végbemenő hangváltozások lehetőségére való tekintettel — egy ilyen nyelvtani szótár megszerkesztésénél elengedhetetlen volt a hagyományosan értelmezett tő elvi és gyakorlati félreállítása, jobban mondva felbontása.

A 2. alatti hasábot azoknál a szavaknál töltjük ki, melyek ragozott alakjában b i z o n y o s b e t ű k vagy b e t ű k a p c s o l a t o k u g r a n a k be, például *гнать — гоним.* Ide kerülnek a *-е-, -о-, -ер-, -ен-* stb. elemek, de a *ь* is, mely hangot ugyan nem fed, de az írott nyelvben önálló jelként kerül alkalmazásra.

A 3. számmal jelzett rovatba a hagyományos t ő u t o l s ó b e t ű j e (vagy betűi) kerül. Jóllehet ez a hagyományosan értelmezett tő része — a fentebbi indokoknál fogva — a szótárban leválasztásra került, önálló nyelvi elemként lép fel. A szótári alak tővégződése a ragozás során gyakran — leggyakrabban az igéknél — változik, éppen ezért volt szükség e rovat beállítására. A tővégződés változására a szótári alakban speciális jellel utalunk.

A 4. számú rovatba a h a g y o m á n y o s k é p z ő kerül. Ezen képzők minden szükséges esetben jelentkeznek a címszóban is.

Az 5. számú rovat a szavak n y e l v t a n i v é g z ő d é s é t tartalmazza, a 6. számú pedig az igéknél kerül feltöltésre a *-ся*, illetőleg a *-сь* elemek esetében.

1.4. A III. számú hasáb a m a g y a r f o r d í t á s t tartalmazza. Hasonlóan a hagyományos kétnyelvű szótárhoz, a fordítás azt a célt szolgálja, hogy az adott orosz példaszó nyelvtani alakjának (végződés vagy végződés és képző) magyar megfelelőjét fordításban lássuk. Ez a hasáb a példaszó fordítása mellett világos választ ad a szerkezetbe (kifejezésbe, típusmondatba) ágyazott példaszó fordítására nézve is. A fordítás szövegében kiemeljük az orosz nyelvi elemeknek megfelelő magyar nyelvi elemeket s ezáltal a nyelvtani szótár — bizonyos vonatkozásokban — az orosz—magyar összevető nyelvtan funkcióját is betölti.

2. A szótár címszavai

A szótár 492 címszót, azaz nyelvtani végződést, illetőleg végződéscsoportot tartalmaz. Ezek típusokra oszthatók.

Vannak 1, 2, 3 és 4 elemből álló címszavak, figyelembe véve azok szerkezetét: *-а, -ё-г, -ущ-ему-ся, -и-те-сь-ка.*

A címszavak alkotó részei a következők lehetnek:

a) a tő utolsó betűje: *-в (гне-в);*
b) a beugró betűk és a tő utolsó betűje: *-о-к(дос-о-к);*
c) képző: *-ӣ (чита-ӣ);*

d) képző és képző: *-ŭ-ся (бо-ŭ-ся);*
e) a tő utolsó betűje és képző: *-з-ся (гры-з-ся);*
f) a végződés: *-ют (да-ют);*
g) képző és végződés: *-ш-ие (грёб-ш-ие);*
h) a végződés és képző: *-ёте-сь (сме-ёте-сь);*
i) a képző és a végződés és képző: *-и-те-сь (ошиб-и-те-сь).*

Az utolsó típusú címszónál jegyezzük meg, hogy a felszólító mód alakja struktúrájának meghatározásánál a *-ŭ*, *-ь*, *-и* elemeket képzőnek, míg a *-те* elemet végződésnek tekintjük. A felszólító alakhoz kapcsolódó *-ка* módosító szócskát a példaszótól leválasztva, a II. számú rovat jobboldali hasábján helyeztük el, mivel számára értelmetlen dolog lett volna új rovatot nyitni. A III. számú rovat baloldali hasábján — ugyancsak a példaszó rovatán kívül — helyeztük el az *он*, *она*, *оно*, *они* névmások ragozott alakjaihoz előljárószó mellett kapcsolódó, szóeleji *н*-elemét.

Érdemes itt kitérni arra, hogy szótárunk címszavai jóval nagyobb mennyiséget képviselnek, mint amennyi nyelvtani végződést a hagyományos nyelvtani rendszerezés alapján számontartanak. Ez — többek között — abból fakad, hogy egyes képző-morfémákat (melléknévi igenév képzőjét és a hozzá hasonló képzőket) mi a végződések szerves részének tekintjük. Ezen túlmenően a beugró betűket szintén a szavak önálló részeként ítéljük meg. A szótár céljának megfelelően a tőhangra végződő, nem szótári alakban álló szavak utolsó betűi (tehát a tővégződés) — függetlenül attól, hogy a szótári alakon kívül milyen nyelvtani funkciót töltenek be (főnévi tárgy-: *хлеб*, és birtokos eset: *стран*, ige múlt ideje: *умер*, melléknév rövid alakja: *угрюм)* — önálló címszóként szerepelnek.

Ha csak a főnevet ragadom ki, megállapíthatjuk, hogy a főnévi végződés hagyományos rendszerezése mindössze 28 végződést ölel fel. Ezzel szemben az orosz—magyar nyelvtani szótár a főnevek tekintetében 110 főnévi végződést (címszót) tartalmaz.

A különféle beszédrészek azonos végződéseit a szótár elhatárolja egymástól, sőt gyakran egy szófajon belül is disztingvál. Gyakorlatilag ez azt jelenti, hogy például az *-а* címszó hétszer szerepel egymás után, mivel ezzel a végződéssel rendelkezhet főnév az egyes számú birtokos és tárgyesetben, a többes számú alanyesetben, de ezen kívül lehet ilyen végződése birtokos melléknévnek, a melléknév rövid alakjának, a névmásnak, a számnévnek és a határozói igenévnek. Az adott funkciót megfelelő rövidítésekkel nyelvtani módon is körülírjuk, adatoljuk, azon túlmenőleg, hogy a példaszók nyelvtani helyét a magyar fordítással is szemléltetjük.

3. A szótár célja

Az „Orosz—magyar nyelvtani szótár" elsősorban a fordítás munkáját van hivatva elősegíteni orosz nyelvről magyar nyelvre. Annak rövidített, az iskola követelményeit figyelembe vevő, a tanuló színvonalára méretezett változata hasznos segédeszközt jelenthet középiskolai és egyetemi tanulóifjúságunk kezében, valamint a magántanulás vonalán.

A nyelvtani szótár a nyelvoktatásban csak segédeszköz szerepét töltheti be, semmiképpen sem tekinthető pedagógiai ,,csodaszernek'' a nyelvoktatás helyettesítésére. A nyelvtanításban a szótár megfelelő változatának segítségével végzett ,,mechanikus'' fordítási gyakorlatok fejlesztenék a tanulók logikai-analitikus készségeit s ez hozzájárulna az orosz nyelv formális elemeinek gyorsabb elsajátításához. Nem véletlen, hogy erre vonatkozólag több utalást találunk a szovjet matematikus nyelvészek megnyilatkozásaiban.[4]

Az ,,Orosz—magyar nyelvtani szótár''-t felhasználhatják a tudomány és technika legkülönfélébb területein dolgozó szakemberek is, akik orosz nyelven óhajtják figyelemmel kísérni a szakirodalmat, de az orosz nyelvet nem ismerik kielégítő módon. Ezeknél a szakembereknél a nyelvtani szótár az ,,emberi fejre alkalmazott gépi fordítási adattár'' funkcióját töltheti be.

Elképzelhető — és gyakorlatilag kivitelezhető —, hogy megfelelő szövegstatisztikai vizsgálatok alapján kisebb terjedelmű nyelvtani ,,szakszótárak'' készüljenek, mivel nem mindenfajta szakszöveg igényli a nyelvben fellelhető összes végződés felhasználását.

Szeretnénk még megemlíteni, hogy az ,,Orosz—magyar nyelvtani szótár'' sikeres felhasználása a fordítás munkájában — véleményem szerint — feltételez bizonyos, minimálisnak nevezhető orosz nyelvi tudást, és bizonyos nyelvészeti műveltséget az anyanyelv tekintetében.

4. A szótár elméleti vonatkozásai

Az ,,Orosz—magyar nyelvtani szótár'' megerősíti azt az álláspontot, hogy a nyelv grammatikája formalizálható. Meg kell azonban mondani, hogy a nyelv formalizálása eddig a gépi fordítást szolgáló algoritmusok formájában testesült meg. Ezzel ellentétben az ,,Orosz—magyar nyelvtani szótár'' nem a gépi fordítási algoritmusok formájában teljesíti a nyelv formalizálását. Szótárunk nem elektronikus gépek, hanem az emberi fej számára készült.

Ezek hangsúlyozása mellett azonban feltétlenül ki kell jelentenünk, hogy a szótár anyaga munkabázist teremthet egy oroszról magyarra fordító gép számára. Annak megállapítása, hogy milyen formában, hogyan történhet ez, még előttünk álló feladat.

Felmerült az a probléma, hogy helyes-e ezt a szótárt ,,nyelvtani szótár''-nak nevezni. Ez a cím a hagyományos kétnyelvű (fordítói) szótárak elnevezésének analógiájára született meg, melyek elnevezésében mindkét nyelvet megjelöljük. Az egyik vélemény szerint a jelenleginél pontosabb cím lenne a következő: ,,Nyelvtani funkciójú orosz szóvégződések ábécérendes orosz—magyar gyűjteménye''.

[4] Sz. K. Saumjan, Лингвистические проблемы кибернетики: Вопросы Философии 14/9. 120—31 (1960). — Megjegyzendő, hogy az elmúlt évek folyamán a Szovjetunióban is kísérleteztek egy hasonló, tisztán mechanikus fordítást lehetővé tevő módszerrel — angolról oroszra; vö. A. A. Zaliznyak, Опыт обучения англо-русскому переводу с помощью алгоритма. Питання прикладної лінгвістик (Csernovic, 1960) 63—5.

Érdeklődésre tarthat számot a szótár abból a szempontból is, hogy eléggé megbízható alapot nyújt egy orosz—magyar összevető nyelvtan kidolgozásához, mivel végigköveti a két nyelv azonos, hasonló vagy eltérő nyelvtani kifejező eszközeit az azonos gondolati tartalom kifejezésében. S ennek didaktikai haszna sem vitatható.

*

Végezetül hadd említsük meg, hogy napról-napra növekszik a nyelvtudomány különféle gyakorlati alkalmazásának szerepe. De nemcsak a nyelvtudomány ad segítséget a gyakorlatnak (gépi fordítás, nyelvoktatás stb.), hanem a gyakorlat követelményei is visszahatnak a nyelvtudomány fejlődésére. Az „Orosz—magyar nyelvtani szótár"-t a gyakorlat igényelte, a szótár gyakorlati céllal készült. Elméleti vonatkozásban a szótár felvet egy sor problémát, melyeknek egy részét beszámolónkban érintettünk. Ezeknek a kérdéseknek a vizsgálata elengedhetetlen, amihez viszont szilárd bázist teremtettek a gépi fordítás elméleti és gyakorlati eredményei.

Szépe György

A magyarországi gépi fordítás néhány kérdése[1]

1. Bevezető megjegyzések

A gépi fordítást a gyakorlatban úgy tekintik, mint az információfeldolgozás egyik legfontosabb részét. (Az információ-feldolgozás többek közt a dokumentáció korszerű, nagyüzemi módszerű formájának az alapja.) Az információ-feldolgozás a tudomány és a termelés számára ma már elengedhetetlen. A világgazdaság kialakulása nemzetközivé tette a tudományt és a technikát, s a rohamosan növekvő számú idegen nyelvű publikációk feldolgozása, lefordítása ma már minden nemzet gazdaságának, kultúrájának létkérdése. A régi módon szervezett fordítói üzemek, bármilye jó munkát végezzenek is, lassanként nem tudnak lépést tartani a szükségletekkel. Fölmerült az igény a gyorsabb és olcsóbb fordításra. Ma már nem a tudomány adja föl önmagának a kérdést erre vonatkozóan, hanem és társadalmi gyakorlat követeli meg ennek a feladatnak a megoldását. Világszerte eldöntöttnek tekinthető az a kérdés, hogy szükséges a gépi fordítás.

Ma már senki előtt sem kétséges az sem, hogy a gépi fordítás lehetséges, elvégezhető feladat.[2] Viszont az is nyilvánvaló, hogy rendkívül bonyolult előkészítést igényel; több szakma koncentrált erőkifejtése szükséges hozzá.

[1] A referátum feolvasása és nyomdába adása között eltelt időben több kérdésben módosult a véleményem. Ebben leginkább a Y. Bar-Hillel professzorral való többszöri és kiadós eszmecserék játszottak szerepet (1962. szeptember 11 és 20 között). A tanulmány azonban csaknem teljesen az eredeti szövegében került nyomdába; helyenként a jegyzetekben tértem ki felfogásom alakulására.

[2] A gépi fordítás lehetőségét legutóbb Mortimer Taube vonta kétségbe „Computer and common sense" (New York, 1962) című könyvében. Taube érvelése azonban nem meggyőző. — Sokkal megalapozottabbak Y. Bar-Hillelnek több helyt kifejtett bizonyos kételyei, vö. például: Some linguistic obstacles to machine translation: Advances in Computers (New York—London, 1960) 1. 146—57; Uő, A demonstration of the nonfeasibility of fully automatic high quality translation: uo. 158—63. — Vö. még: „A fordítógép nem pótolhatja az emberi intelligencia működését" című riportot Y. Bar-Hillel professzorral: Csongrád Megyei Hírlap (szentesi kiadása) 1962. október 4, 4. lap. — Természetesen Y. Bar-Hillel nem vonja kétségbe a gépi fordítás lehetőségét. Mindössze azt hangoztatja több helyütt, hogy a gép utólagos fésülés nélkül aligha tud hibátlan, teljesen automatikus fordítást létrehozni. Itt jegyzem meg azt is, hogy a szovjet kutatók Bar-Hillelnél jóval optimistábbak ebben a tekintetben is.

A nyelvészek tudják, hogy nekik is — közvetlenül vagy közvetve — elő kell segíteniük ezt az induló nagy kollektív vállalkozást. Azonban azt is jól tudják, hogy a feladatnak csupán egyik része esik rájuk.

A következőkben néhány olyan problémát szeretnék érinteni, amellyel foglalkoznunk kell, ha Magyarországon is hozzálátunk a gépi fordítás szervezett előkészítéséhez. Beszámolóm inkább áttekintő, tájékoztató jellegű lesz, mintsem elemző; és inkább problémafölvető, mint saját eredményekről való beszámoló. Problémafölvetésem pedig jóformán csak nyelvészeti szempontú lehet.

2. A külföldi gépi fordítási munkálatok

A külföldi gépi fordítási munkálatok áttekintését megnehezíti a hatalmas mennyiségű és egyre szaporodó szakirodalom.[3] Sajnos, ennek csupán egy töredéke található meg hazai könyvtárainkban. A gépi fordításnak szentelt öt folyóirat, időszaki kiadvány közül — tudomásom szerint — egyetlen egy sem található meg teljes sorozatban a budapesti közkönyvtárakban.[4] Mindez nemcsak az én beszámolómnak, hanem az egész hazai munkánknak ideiglenes és föltételes jelleget kölcsönöz.

A rendelkezésemre álló adatok szerint gépi fordítással foglalkoztak, illetőleg foglalkoznak a következő országokban (betűrendben): Brazília, Csehszlovákia, Egyesült Államok, Franciaország, Japán, Jugoszlávia, Kína, Lengyelország, Nagy-Britannia, Német Demokratikus Köztársaság, Német Szövetségi Köztársaság, Norvégia, Olaszország, Románia, Svédország és Szovjetunió. (Magyarország — ezek szerint — valószínűleg a 17. ország, amely hozzálátott a gépi fordítás kérdéséhez.[5]) Ismertetésemben csupán

[3] A gépi fordítás irodalmát 1959-ig összefoglalja: Allen Kent, Machine literature searching and translation. An analytical review: Information Retrieval and Machine Translation (New York—London, 1960) 136—236; E. és K. Delavenay: Bibliography of Machine Translation (Hága, 1960). Az előző mű folytatása (a nem szovjet kutatásokra vonatkozóan): R. D. Ravics: Библиография зарубежных работ по машинному переводу (1960—1961 гг.) (Moszkva, 1962). — Állandó jellegű tájékoztatást ad az amerikai National Science Foundation félévenként kiadásra kerülő tájékoztatója: Current Research and Development in Scientific Documentation (Washington, D. C., 1962. május: 10. kötet); továbbá a gépi fordítás több folyóirata. — Vö. még: G. Mounin—A. Dewèze, Ressources bibliographiques concernant la traduction automatique: La Traduction Automatique 1961. 2. szám 10—12.

[4] A gépi fordításnak egészben (vagy jó részben) szentelt folyóiratok a következők: Mechanical Translation (Cambridge, Massachusetts, 1954-től; szerkeszti: Victor Yngve); Бюллэтэнь объединения по машинному переводу (Moszkva, 1957-től); Új címe: Машинный перевод и прикладная лингвистика; La Traduction Automatique (Párizs, 1960-tól; szerkeszti: E. Delavenay); Mathematical Linguistics and Automatic Translation (Cambridge, Mass., 1959-től, a Harvard University, Computation Laboratory kiadásában — Soviet developments in Mechanical Translation (Washington, D.C., 1957-től). — Sajnos, csupán az első három jutott el — csonkán! — Budapestre. — A felsorolt forrásokban valamennyi általam felvetett kérdésre bőven található utalás.

[5] A lista persze hozzávetőleges, a sorszámnak nincsen semmilyen különösebb értéke. Azóta például olyan országokból kaptunk hírt, mint Ausztria, Belgium, Hollandia, India, Mexikó. Ilyen sokaság esetén a sorrendnek már nincs jelentősége. Ez a névsor arra szolgál, hogy megmutassa a gépi fordítással való foglalkozás általános voltát.

a legnagyobb országok, valamint a szomszédos szocialista országok tevékenységére térek ki röviden.

2.1. A gépi fordítási munkálatok kétségtelenül a Szovjetunióban legnagyobb méretűek.[6] Az első kísérletek viszonylag későn indultak, azonban hamarosan sikeres próbafordításokhoz vezettek. 1955-ben a Szovjetunió Tudományos Akadémiája Finommechanikai és Számítástechnikai Intézetben végrehajtották az első eredményes vállalkozást angol tudományos és technikai szövegek oroszra fordítására.[7]

Ma Moszkvában, Leningrádban és más egyetemi városokban valószínűleg mintegy harminc munkacsoportban több száz nyelvész és matematikus kutató foglalkozik az automatikus fordítás tudományos előkészítésével. Külön egyesületük, saját folyóiratuk van,[8] számos egyetemen és főiskolán előadások és szemináriumok foglalkoznak ezzel a témakörrel. Munkájuk rendkívül sokoldalú; a különféle alapról elinduló és más-más módszerrel dolgozó kutatók gyakran cserélik ki vitákon tapasztalataikat.

A szovjet gépi fordítási munkálatok egyik fő jellemzője az, hogy egyensúlyban kívánják tartani az elvi kutatásokat és a gyakorlati kísérleteket. A számos csoport közül két iskola profilja rajzolódik ki határozottabban; az egyik Moszkvában, a másik Leningrádban működik. A moszkvai iskola több akadémiai intézet (például a finommechanikai és számítástechnikai, a matematikai, a nyelvtudományi intézet)[9] csoportjait, valamint felsőoktatási intézetek dolgozóit (elsősorban az I. számú idegen nyelvi főiskola oktatóit) foglalja magában. Az iskola igen erős teoretikus alapokkal (általános nyelvészet és matematika) rendelkezik, és erősen grammatikai irányzatú. Ez azt jelenti, hogy ma már egyáltalán nem elégednek meg az auto-

[6] A szovjet gépi fordítási kutatásokról igen jó összefoglalást ad Papp Ferencnek „A matematikai nyelvészet és gépi fordítás a Szovjetunióban" című műve (Bp., 1864,. Időszerű Műszaki Dokumentációs Kérdések 6. szám); uő, Új irányzatok a szovjet nyelvtudományban: Nyelvtudományi Közlemények 61. 392—410 (1959). — Egyéb ajánlható összefoglalások: V. Ivanov, Report on the MT Work in the USSR: Proceedings of the VIII International Congress of Linguists (Oslo, 1958) 533—5; O. S. Kulagina—A. I. Martynova—T. M. Nikolaeva, Mechanical Translation at the Academy of Sciences of the URSS: Information Retrieval and Machine Translation (New York—London, 1960) 867—79; Kenneth Harper, Soviet Research in Machine Translation: Proceedings of the National Symposium on Machine Translation (Szerk.: H. P. Edmundson; Englewood Cliffs, N. J., 1961) 2—12. — Itt jegyzem meg, hogy az amúgy is kényszerűen hézagos hivatkozásokat erősen megrostáltam: a frissebb, fontosabb és hozzáférhetőbb, valamint a nem gépi fordítók számára érthetőbb műveket részesítettem előnyben.

[7] Vö. D. Ju. Panov: Автоматический перевод (Moszkva, 1956) 44 lap; angol fordítása: D. Yu. Panov: Automatic Translation (London—Oxford—New York—Paris, 1960. Pergamon Press) 73 lap. — Vö. még D. Ju. Panov, A gépi fordítás és az ember: Tájékoztató a tudományos kutatás tervezésének, igazgatásának és szervezésének nemzetközi irodalmáról 1961. 1—2. szám, 52—8. lap; függelék: Gépi fordítási kísérletek angolból oroszra a SzTA BESzM elektronikus számológépével; uo., 58—61. lap.

[8] Az egyesület neve: Объединение по машинному переводу. A folyóiratot lásd a 4. jegyzetben.

[9] Az intézmények pontos nevét és címét lásd Papp Ferencnek a 6. jegyzetben hivatkozott összefoglalásában, valamint a 3. jegyzetben felsorolt összefoglaló művekben.

mata szótár (vagy más szóval: „szóról szóra") jellegű fordítással. Számottevő az iskolának a fordításelmélettel és egyéb gépi fordításhoz szükséges ágazatokkal, például a szókincs rendszerezésével, grafematikával való foglalkozása is. Az iskola egyik vezető egyénisége, I. A. Melcsuk, az elméleti gépi fordítási kutatásban is alapvető jelentőségű munkát végzett.

A moszkvai iskolai gyakorlati kísérletei általában a közvetlen fordításra (tehát az egyik nyelvről a másikra való átfordításra) irányultak. Úgynevezett szimulatív kísérleteket (vagyis a gépi feladatnak ember által való végrehajtását), illetőleg gépi kísérleteket hajtottak végre több nyelvről, így a magyarról is.

A leningrádi iskola fő jellemzője az interlingua segítségével végrehajtandó fordításra való törekvés. Andrejevnek a leningrádi egyetemen működő Gépi Fordítási Technikai Laboratóriuma[10] talán világszerte a legnagyobb ilyen jellegű intézmény, több mint száz belső és külső munkatárssal dolgozik. Munkájuk két irányú: egyrészt az orosznak és egész sor más nyelvnek „független analízisét", illetőleg „független szintézisét" készítik el; másrészt az interlinguára, illetőleg a gépi instrukciók „nyelvére" vonatkozóan végeznek beható kutatásokat. Elképzelésük az, hogy minden jelentősebb nyelvet egzakt elemzésnek vetnek alá. Az elemzések eredményéül egy közvetítő nyelvet kapnak, amely a megvizsgált egyes nyelvek tulajdonságait tartalmazza. A fordítás iránya pedig ez volna:

forrásnyelv → közvetítő nyelv → célnyelv;
vagyis: analízis → átkódolás → szintézis.

Érdekes módon a leningrádiaknak — úgy tűnik — kisebb az érdeklődésük a gyakorlati kísérletek iránt.

2.2. Az Egyesült Államokban is szinte minden jelentős egyetemi központban foglalkoznak valamilyen formában a gépi fordítással. A kutatások jó részét az állam (elsősorban a honvédelmi szervek) és az ipar finanszírozza. Az egyes csoportok között csak a folyóiratok és a gyakori konferenciák tartják fenn a kapcsolatot.

Az első sikeres fordítási kísérletet 1954-ben hajtották végre a Georgetown Egyetem és az International Business Machine Corporation közös vállalkozásaként, Washingtonban. Az aránylag csekély orosz szövegnek angolra fordítása elsősorban a gépi fordítás lehetségességét kívánta demonstrálni. A kísérlet nyelvészeti részének kidolgozója a magyar származású Paul L. Garvin professzor volt.[11]

[10] Papp Ferenc összefoglalásán kívül lásd: N. D. Andrejev, Basic Directions of Work at the Experimental Laboratory of Machine Translation: Information Retrieval and Machine Translation (New York—London, 1961) 687—91. — A csoport cikkgyűjteménye: Материалы по машинному переводу (Szerk.: N. D. Andrejev; Leningrád, 1958) I. kötet, 228 lap; II. kötet (1963) 196 lap.

[11] Vö. Paul L. Garvin, Machine Translation: Proceedings of the VIII International Congress of Linguists (Oslo, 1958) 502—10. — A Georgetown Egyetem jelenlegi munkáját lásd: Michael Zarechnak—A. F. R. Brown, Current Research at Georgetown University: Proceedings of the National Symposium on Machine Translation 63—87. — Ennek a washingtoni csoportnak számos sokszorosított kiadványa megtalálható az MTA Számítástechnikai Központjának könyvtárában.

Az — itt csak egy-két jellemző példában hivatkozott — csoportok kétféle elgondolás alapján dolgoznak. A Georgetown Egyetem és a Massachusetts Institute of Technology[12] az alapos grammatikai előkészítés hívei. (Az utóbbi intézményben dolgozik — de nem közvetlenül a gépi fordítás kérdésein — Noam Chomsky, a modern nyelvtudomány egyik kiemelkedő teoretikusa.) Ők alapos strukturális elemzés útján kívánják megközelíteni a feladatot. Más csoportok, közülük például a Harvard Egyetem[13] és a seattle-i Washington Egyetem[14] csoportja, valamint a Ramo-Wooldridge Társaság[15] csoportja alapos és részletes technikai bázison inkább az automata szótárt kívánják fordítássá bővíteni. A seattle-i csoport különlegessége, hogy az elektronikus számológépek óriási memóriaképességére[16] épít: fotoszkopikus módszerrel minden orosz ragozott névszói és igealakot szótárszerűen rögzítenek, és az így fennmaradó, aránylag „csekély" számú grammatikai kérdést is különféle módszerek eklektikus alkalmazásával kívánják megoldani. Persze az automata szótár nagyvonalúsága a fordítás problémáinak körülbelül 80—90 százalékát oldhatja csak meg. Érdekes módon ezek a grammatikától távolabb álló csoportok behatóan foglalkoznak szemantikával. — A többi közül kiemelkedik még a Pennsylvania Egyetemen Z. S. Harris[17] irányításával működő csoport,

[12] W. N. Locke—Victor H. Yngve, Research in Translation by Machine at M. I. T.: Proceedings of the National Symposium on Machine Translation 510—14. — Újabban: Victor H. Yngve, MT at the Massachusetts Institute of Technology: Proceedings of the National Symposium on Machine Translation 126—32.

[13] A Harvard Egyetemen folyó munkára vö.: A. G. Oettinger: Automatic language translation (Cambridge, Mass., 1960) XIX + 380 lap. — Vö. még: Bibliography of Work of the Harvard Computation Laboratory on Mathematical Linguistics and Automatic Translation (Cambridge, Mass., 1959). — Legújabban: Anthony G. Oettinger—Murray E. Sherry, Current research on automatic translation at Harvard University and predictive syntactic analysis: Proceedings of the National Symposium on Machine Translation 173—82. Itt látható, hogy a Harvard-csoport az utolsó két évben megváltoztatta munkamódszerét: átvette Ida Rhodes (National Bureau of Standards, Washington, D. C.) „prediktív analízis" elnevezésű eljárását.

[14] Erwin Reifler, The machine translation project at the University of Washington, Seattle, Washington, USA: Proceedings of the VIII International Congress of Linguists (Oslo, 1958) 514—18; Linguistic and Engineering Studies in Automatic Language Translation of Scientific Russian into English (Seattle, Washington, 1958) 658 lap; Erwin Reifler, Current research at the University of Washington: Proceedings of the National Symposium on Machine Translation 155—59.

[15] Az intézmény teljes neve: Thompson Ramo—Wooldridge Inc. — Újabb munkabeszámoló: Jules Mersel, Research in machine translation at Ramo—Wooldridge: Proceedings of the National Symposium on Machine Translation 26—38. — A konferencia óta P. L. Garvin szívességéből eljutottak hozzám a Ramo—Wooldridge laboratórium legújabb kiadványai. Ezek azt mutatják, hogy a csoport a legutolsó években új irányú, sajátos, magas elméleti igényű munkamódszert alakított ki.

[16] A permanens memóriakapacitás $30 \cdot 10^6$ bit. — A memóriakapacitás nagyságára vonatkozóan vö.: Gilbert King:, The problem of lexical storage: Report of the Eighth Annual Round Table Conference on Linguistics and Language Teaching. No. 10.

[17] Tevékenységük rövid összefoglalását lásd: Current Research and Development in Scientific Documentation, No. 10. 285—87. — A csoport kiadványsorozatai: „Transformations and Discourse Analysis Papers" és „Papers on Formal Linguistics"; ezek Magyarországon csupán a beszámoló megírása után váltak részben hozzáférhetővé.

valamint az Illinois Egyetemen[18] R. B. Lees vezette csoport; mindkét kollektíva elektronikus számológéppel végzi el az angol nyelv szintaktikai elemzését és szintézisét. Nekik nem (közvetlen) céljuk a gépi fordítás, azonban talán éppen ezért az ő analízisük lesz a legmegbízhatóbb azok számára, akik angolból akarnak fordítani.

Itt jegyezzük meg, hogy az amerikaiak az esetek túlnyomó többségében csak az oroszról angolra történő fordítás kérdéseivel foglalkoznak. Nagy-Britanniában viszont — talán éppen az amerikaiak választásának következtében — jelentős szerepet kapnak egyéb kontinentális nyelvek és a közvetítő nyelv (az interlingua) problémái is.

2.3. Nagy-Britanniában két csoport tevékenysége jelent újdonságot valamennyi eddig ismertetettéhez képest. Az első a londoni egyetem Birbeck Kollégiumának Numerikus Automatizálási Intézetében[19] működik A. D. Booth vezetésével. Valamennyi csoport közül ez leginkább technikai orientációjú. Vezetője számológép-szakember; s tevékenységük a gépi fordításon kívül kiterjed a filológiának szinte valamennyi gépesítési lehetőségére, például a stilisztikai elemzésre is. Különösebb nyelvészeti elemzést azonban csak a csoport egyetlen volt nyelvész tagja végzett. Boothnak kétségtelenül úttörő érdemei vannak a gépi fordításban; egyrészt a híres Weaver-féle memorandum[20] létrejöttében van szerepe, másrészt pedig ő alapította az első szervezetten működő gépi fordítási csoportot még a negyvenes évek legvégén.[21] A kezdeményezés tehát műszaki oldalról történt.

A Cambridge-i Nyelvi Kutató Egység[22] egyébként komplex tevékenységének különlegessége a thesaurus-módszer. A thesaurus Roget tárgykörök szerint elrendezett szótárja alapján fejlődött ki — általános szemantikai rendszerré, vagy talán pontosabban: interlingua-lexikonná. A thesaurus egységek statikus leírását a kombinatorikus logikával egészítik ki. Az interlingua-szótárhoz még egy interlingua-szintaxis is hozzátartozik. A csoport elvi irányítója Margaret Masterman filozófus. — Moszkvában Rozencvejg

[18] A csoport idevágó tevékenységéről 1962. szeptember 13-án Tihanyban („A matematika alapjai, matematikai gépek, és alkalmazásaik" című kollokviumon) beszámolt R. B. Lees „Automatic Generation of Natural-Language Sentences" címmel. — Az előadás a kollokvium beszámoló kötetében jelenik meg.

[19] Department of Numerical Automation, Birbeck College, University of London. — A csoport tevékenységére vö. A. D. Booth—L. Brandwood—J. P. Cleave: Mechanical Resolution of Linguistic Problems (New York—London, 1958).

[20] Warren Weaver, Translation: Machine Translation of Languages (Szerk.: W. N. Locke—A. D. Booth; Cambridge, Mass.—New York, 1955) 14—23. — Ez az 1948-ban kéziratos memorandum a mostani gépi fordítási munkálatoknak első fölvetése. — A régebbi javaslatoknak nincs a mai munkával összefüggése.

[21] Vö. A. D. Booth, The History and Recent Progress of Machine Translation: Aspects of Translation = Studies in Communication, No. 2 (London, 1958) 88—104.

[22] Cambridge Language Research Unit. — A munkálatok összefoglalására vö.: M. A. K. Halliday, Report on MT at the CLRU: Proceedings of the VIII International Congress of Linguists 527—3; A. F. Parker-Rhodes, Some recent work on thesauric and interlingual methods in machine translation: Information Retrieval and Machine Translation 923—34. — Vö. még: Report of Colloquium of Cambridge Language Research Group, August 1955: Mechanical Translation 3. 2—28 (1955).

csoportja,[23] Milánóban pedig az operacionalista filozófiai iskola egyik csoportja[24] dolgozik hasonló irányban.

2.4. A szomszédos szocialista országok közül háromban volt alkalmam — más irányú elfoglaltságom mellett — röviden betekinteni az ott folyó gépi fordítási előkészületekbe. E téren — mint a modern nyelvtudomány legtöbb területén — élen járnak Csehszlovákiában.[25] A prágai Károly Egyetem bölcsészkarán a „Cseh nyelvi, általános nyelvészeti és fonetikai" tanszéken belül működik egy gépifordítási alosztály. 5 nyelvész, 1 matematikus és 7—8 cédulázó dolgozik rendkívül intenzíven az angolról csehre való fordítás kérdésein. A csoport tagjait fölmentették más irányú egyetemi kötelezettségük alól, csupán a gépi fordítás köréből tartanak szemináriumokat és speciálkollégiumokat. Valamennyit elküldték hosszabb időre külföldre, főleg a Szovjetunióba. Nagy arányú könyvtári és technikai beruházásokat eszközöltek: nemrég kaptak például egy alfa-numerikus kártya-lyukasztógépet. Körülbelül három évi előkészület után a csoport sikeres próbafordítást hajtott végre angolról csehre a Számítógépek Kutató Intézetével karöltve.

A fordítást csekély szövegen végezték el; jelentőségét elsősorban demonstratívnak tekintik. Azóta teljesen az interlingua alapjára helyezkedtek. Elkészítették a cseh nyelv független analízisét és szintézisét. Jelenleg egy nagyobb szövegen közvetítő nyelv (interlingua) segítségével végrehajtandó, szintén angolról csehre történő fordítás előkészítésén dolgoznak. Ennek elvégzése 1962 első felében várható.

Kiterjedt elméleti munkát is végeznek a fordításelmélet és a komparatív szinkrón grammatika terén. Az interlingua-kutatásban felhasználják a prágai iskola tipológiai hagyományait. Több nyelvnek — köztük a magyarnak is — szemantikai, morfológiai és szintaktikai tipológiai katalógusát állították össze az interlingua számára. A végleges algoritmusokat a transzformációs analízis módszereinek felhasználásával kívánják elkészíteni. Tevékenységük kiterjed az információ-feldolgozás területére és az információs nyelv problémájára is. — Elkezdték az angol és az orosz elektrotechnikai szókincsnek a fordítás szempontjából való vizsgálatát.

Brnóban is megkezdik a közeljövőben a gépi fordítási munkálatokat.[26] Számunkra ennek a csoportnak tevékenysége különösen fontos lehet,

[23] Sz. K. Saumjan szóbeli közlése 1962 januárjában; 1963 végén személyesen is alkalmam volt megmerkedni a csoport munkájával.

[24] Vö.: Enrico Albani—Silvio Ceccato—Enrico Marinetti: Classification, rules, and code of an operational grammar for mechanical translation: Information Retrieval and Machine Translation 693—753. — Sajnos a „Centro di Cibernetica e di Attività Linguistiche (Università degli Studi di Milano)" munkájának könyv-méretű összefoglalása: „Linguistic analysis and programming for mechanical translation (Mechanical translation and thought)' (Milano, 1960, Feltrinelli, 238 lap) csupán nemrég jutott el Budapestre.

[25] A csehszlovákiai gépi fordítási munkálatokat a helyszínen tanulmányoztam 1961 május folyamán, majd rövid ideig 1962 januárjában. — Frissebb adatokat (bibliográfiával) lásd: Klauszer Judit—Jan Průcha, Matematikai és alkalmazott nyelvészet Csehszlovákiában: Nyelvtudományi Közlemények 64. 371—78 (1962). Vö. még Прикладная лингвистика в Чехословакии: Машинный перевод и прикладная лингвистика, № 7.

[26] J. Jelínek brnói docens személyes közlése 1961 májusában.

mivel — értesüléseim szerint — a csehről idegen nyelvre történő közvetlen fordítást kívánják előkészíteni.

2.5. A Német Demokratikus Köztársaságban a Német Tudományos Akadémia külön elnökségi bizottságot hozott létre a matematikai nyelvészet és a gépi fordítás irányítására.[27] A bizottság alá két önálló kutató csoport tartozik. Az egyik csoport a német strukturális grammatikával foglalkozik.[28] A másik csoport[29] feladata elsősorban a gépi fordítás általános tanulmányozása és az oroszról németre, valamint angolról németre való gépi fordítás előkészítése. A csoport néhány éve alakult, s most alig hogy túljutottak a tájékozódás, a bibliográfia elkészítése és a kezdeti dokumentáció szakaszán.

Konkrét munkájuk azonban már így is igen biztató. Két alcsoportra oszlanak. Az egyik az oroszról való fordítást készíti el a leningrádi Andrejevféle interlingua segítségével. A másik alcsoport (vezetője egy elektrotechnikában járatos nyelvész szakember, G. Agricola) elkészített egy 40 000 szóból álló angol—német automata szótárt, amely a műszaki területek közös szóanyagát tartalmazza. A feladat egészen eredeti; megoldásához minden iskola nyújtott nekik valamit, a legtöbbet E. Reifler seattle-i csoportja. — A csoportban 9 főállású és 10 mellékállású munkatárs dolgozik: nyelvészek és matematikusok.

Ezenkívül a Humboldt Egyetem Fonetikai és Kommunikációkutatási Intézetében[30] foglalkoznak a közvetítő nyelvnek (az interlinguának) a problémájával. G. F. Meier professzor — az egyetemi intézetnek a vezetője — tizennégy nyelvre (köztük a magyarra) kívánja kiterjeszteni a nemzetközi interlingua-együttműködést. Tervei szerint nekünk, magyaroknak az orvosi nyelv szemantikai problémáit kellene kidolgoznunk az interlingua számára.

2.6. A Román Népköztársaságban a temesvári műegyetemen[31] és a kolozsvári Babeș-Bolyai Tudományegyetem bölcsészkarán működik gépi for-

[27] Kommission für Mathematische Linguistik und Automatische Übersetzung — A német gépi fordítás kérdéseit — igen röviden — a helyszínen tanulmányoztam 1962 januárjában.

[28] Arbeitsstelle für strukturelle Grammatik. — A csoport kiadványsorozata: Studia Grammatica.

[29] Arbeitsstelle für Mathematische Linguistik und Maschinenübersetzung. — Problémabeszámolójuk: Erhard Agricola, Methoden und Probleme der automatischen Übersetzung: Die Technik hilft der wissenschaftlichen Arbeit = Forschung —Lehre—Praxis (Berlin, 1964.) 4. szám, 153—71. — Azóta — mint Telegdi Zsigmond tájékoztatásából 1962 végén hallom — ez a csoport már jóval messzebb haladt.

[30] Institut für Phonetik und Kommunikationswissenschaft. — A terveket G. F. Meier személyesen közölte 1962 januárjában.

[31] Institut Politechnic din Timișoară. — Azóta különféle forrásokból többet tudunk erről a csoportról. 1962 júliusában Domonkos Erika sikeres próbafordítást hajtott végre angolról románra: Raport asupra algoritmului de traducere automată din limbă engleză în limbă română (Bukarest, 1962, Academia Republicii Populare Romîne, Comisia de Lingvistică Matematică) 22 lap; megjelent angol és orosz fordításban is. — Vö. még: Erica Domonkoș, Algoritmul de traducere englezo—română pentru calculatorul M. E. C. I. P. T.: Probleme de automatizare 3. 105—108 (1960).

dítási csoport.[32] Én csak az utóbbi munkáját tudtam tanulmányozni: Kolozsváron Schveiger Pál vezetésével román és magyar kutatók együttes munkával készítenek elő egy oroszból románra és magyarra egyszerre történő „iker"-fordítást. Az elágazó algoritmus világszerte egyedülálló kísérlet. Az ilyen módon történő praktikus fordításon kívül egy Melcsukéhoz hasonló közvetítő nyelv (interlingua) kialakítása is vezette őket ebben a kísérletben.

3. Hazai törekvések a gépi fordítás terén

A következőkben a gépi fordítás magyarországi történetét, helyesebben előtörténetét próbálom meg vázlatosan összefoglalni — elsősorban a nyelvészet szempontjából. Csak a közvetlen előtörténetre gondolok; hiszen távolabbról, közvetve ide tartozna a magyar tudomány és technika több eredményének felsorolása, bemutatása. A hazai matematika, kibernetika és természetesen a nyelvtudomány színvonala — valamint a magyar tudomány szervezete — teszi számunkra lehetővé, hogy a reális megvalósítás igényével hozzá tudjunk fogni a gépi fordítás problémájának megoldásához.

A gépi fordításra irányuló közvetlen törekvések először népszerűsítő ismertetések formájában nyilvánultak meg. Azonban szinte az első perctől megvolt az igény az elmélet mélyebb bemutatására, illetőleg gyakorlati alkalmazására.

A Magyar Nyelvtudományi Társaság általános nyelvészeti szakosztálya 1958 őszétől vitaüléseket szentel a szinkrón nyelvészet időszerű kérdéseinek, így a gépi fordításnak is. Fodor István,[33] Papp Ferenc,[33] majd Szalai Sándor több előadásában is ismertette a gépi fordítást. Kalmár László matematikai-nyelvészeti és Rényi Alfréd információelméleti előadása is bővítette a gépi fordítással foglalkozni kívánók látókörét.

A Magyar Tudományos Akadémia Nyelvtudományi Intézetében Tarján Rezső és Szalai Sándor ismertette külön előadásban a gépi fordítás kérdéseit. Szalai Sándor 1960 őszén és 1961 tavaszán hat hónapos szemináriumot tartott az intézetben: a matematikai logikáról és az információelméletről különös tekintettel a gépi fordításra. Szalainak különleges érdemei vannak a gépi fordításra vonatkozó ismeretek elterjesztésében és a szervezett fordítói munka sürgetésében. Szalai gyakorlati munkát is végzett e téren: alighanem ő az első Magyarországon, aki — a Nyelvtudományi Társaságban tartott előadásához való illusztrációként — a gépi fordítás céljaira egy algoritmus-részletet kidolgozott

[32] Problémabeszámoló: Paul Schveiger, Unele probleme ale traducerii cu maşina electronică de calcul I.: Studia Universitatis Babeş-Bolyai, Series IV, Fasciculus 2 92—8 (1960).

[33] Vö. Fodor István, A modern nyelvtudomány főbb problémái: Magyar Nyelvőr 82. 460—75 (1958); a gépi fordításról: 471—2; a cikknek ezt a részét a Valóság is közölte 1958. év. 1. számában 45—49. Papp Ferenc, Új irányzatok a szovjet nyelvtudományban: Nyelvtudományi Közlemények 61. 392—410 (1959); a gépi fordításról: 403—8.

1960 nyarán a magyar tudomány távlati tervébe is belekerült a gépi fordítás. Azóta a távlati terveket jóváhagyták. A gépi fordítás tehát már nemcsak kívánságunk, hanem kötelességünk is.[34]

1960 decemberében a Nyelvtudományi Intézet Telegdi Zsigmond vezetésével összehívott „a távlati tervnek egyes, a nyelvtudomány, a logika és a kibernetika együttműködését feltételező pontjainak megbeszélésére" egy interdiszciplináris értekezletet. Az értekezlet beszámolóit Kalmár László és Papp Ferenc tartották. Jelen munkaértekezletünk voltaképpen ennek a megbeszélésnek a folytatása.

A Nyelvtudományi Intézetben bizonyos szervező munka is megindult, hogy megkeresse és összegyűjtse az ezen a területen szétszórtan működő szakembereket. Kónyi Sándor orosz szakos tanár elsősorban pedagógiai célokból fogott hozzá egy magyar—orosz nyelvtan szótárszerű feldolgozásához.[35] (A gépi fordítás szükségleteit is figyelembe vevő könyvméretű magyar—orosz fordítási algoritmusa is elkészült s jelenleg kiadásra vár.)

Hadas Ferenc ugyancsak pedagógiai célzattal, de az automatikus fordításra is tekintettel készítette el „Orosz—magyar nyelvtani szótár"-át, amely sokszorosításban megjelent.[36] — Itt jegyzem meg, hogy Hadas műve előtt jelent meg Gábor Mihálynak „Nemzetközi fordítókulcs" című német—magyar, magyar—német s angol—magyar nyelvtani szótára, amelyről Melcsuk elismerő bírálatot írt a moszkvai gépi fordítási közlönyben.[37]

Hell György, valamint Sipőczy Győző, a Műszaki Egyetem Gépészkari Idegennyelvi Lektorátusának tanárai 1961-ben gyakorlatilag is hozzáláttak bizonyos gépi fordítási kísérletek elvégzéséhez. Az év végén a Vezetékes Híradástechnikai Tanszék jelfogós gépén egy orosz mondatot magyarra fordítottak. Fordításuk ezen az oktatási célokat szolgáló gépen természetesen kísérleti jellegű volt. Azonban módszerük (szóról szóra való fordítás grammatikai korrekcióval) elvileg azonos szinten áll több külföldi kísérlettel. 1962 elején Hell György tört idős alkalmazásban a Magyar Tudományos Akadémia Számítástechnikai Központjába került. Itt Dömölki Bálint segítségével a Központ M—3 típusú gépén végez kísérleti előkészítő munkát[38] az orosz—magyar gépi fordításhoz.

[34] A 19.01.05. témacsoport (Kibernetikai módszerek alkalmazása a nyelvészetben) 02. pontja: A gépi fordítás elméleti előkészítése.

[35] Kónyi Sándor, A kibernetika és a fordítás gépesítése: Természettudományi Közlöny 4. 26—30 (1960), pályadíjnyertes ismeretterjesztő közlemény. — Vö. még uő.: Magyar—orosz nyelvtani szótár: Az idegen nyelvek tanítása. 4. 177—82 (1961).

[36] Vö.: Hadas Ferenc: Orosz—magyar nyelvtani szótár (Bp., 1961, kézirat). Felsőoktatási Jegyzetellátó Vállalat. 164 lap; Hadas Ferenc, Az orosz—magyar nyelvtani szótárról: Kötetünkben 205—210; vö. még: Hadas Ferenc: О русско-венгерском грамматическом словаре: Slavica 2. 165—76 (1962).

[37] A bibliográfiai adatokat, valamint Melcsuk véleményét lásd kötetünkben 275, 276.

[38] Vö. Hell György, Az orosz—magyar gépi fordítás elméleti és gyakorlati kérdései: Kötetünkben 177—196; Sipőczy Győző, Az elöljárós kifejezések programozási kérdései: Kötetünkben 231—242. — Vö. még: Hell György, О нескольких характерных чертах программа МП с русского языка на венгерский: Slavica 2. 151—64 (1962); Hell György: A gépi fordítás néhány kérdése: Műszaki Könyvtárosok Tájékoztatója 9/5. 1—13 (1962).

A gépi fordítás nyelvészeti kérdéseinek területén önálló magyar tudományos publikáció — 1962 márciusáig — még nem jelent meg. Figyelemre méltóak Papp Ferenc beszámolói a szovjet munkáról.[39] A Magyar Nyelvőr ismertető írásain kívül az Akadémiai Könyvtár tudományszervezési tájékoztatója tartja kisebb-nagyobb közlemények formájában napirenden a gépi fordítást, az információ-feldolgozás témakörén belül. „Fordítógépek a tudomány szolgálatában" című jól sikerült összeállításában részleteket közöl Panov ismert orosz fordítógép-szakember tanulmányából, és ismerteti az első sikeres orosz gépi fordítási kísérletet.[40]

A magyar nyelvvel kapcsolatos gépi fordítás legjelentősebb publikációja Moszkvában jelent meg. I. A. Melcsuk „A magyar nyelvről az oroszra történő gépi fordításról" című tanulmánya[41] a gépi fordítás általános elmélete számára is igen jelentős. A magyarról oroszra és más idegen nyelvre való gépi fordításhoz jó alapot nyújt Melcsuk fenti írása.

4. A gépi fordítás néhány elvi és módszertani kérdése

A gépi fordítás elvi és módszertani kérdéseinek bemutatását megkönnyíti és megnehezíti az a tény, hogy erről még több előadás is szól. A fölösleges ismétlések elkerülése végett számos pont bemutatásától eltekintek. Mindössze három problémát vázolok föl egészen röviden.

4.1. Az első dilemma az elmélet és a gyakorlat viszonyára vonatkozik. A gépi fordításban a sokat publikáló nyugati csoportok két részre oszlanak. Az egyik fajta irányított alapkutatásnak tekinti a gépi fordítást, a másik pedig a számológépek egy sajátos területen történő alkalmazásának. (Ezzel a kétféleséggel részben — de korántsem föltétlenül — egyezik a nyelvészeti és a „műszaki" orientációjú csoportok különbsége.) Az elvi előkészítés hívei olyan nyelvészek, akik a feladatot annyira részletes és megbízható algoritmus kidolgozásában látják, amelynek segítségével a gép teljes (ahogy ők mondják: legalább 95 százalékos megbízhatóságú) fordítást tudjon elvégezni. Mivel pedig a meglevő grammatikák nem adnak választ az algoritmushoz szükséges minden kérdésre,

[39] Vö. a 6. és 33. jegyzetben hivatkozott munkáját. — Vö. még: Papp Ferenc, A matematikai módszerek alkalmazásáról nyelvtudományunkban: A Magyar Tudományos Akadémia Nyelv- és Irodalomtudományi Osztályának Közleményei 17. 289—306 (1961), a gépi fordításról: 299—301; vö. Pappnak még ugyanebben a folyóiratban tett megjegyzését a gépi fordításról: 18. 102 (1962).

[40] Tájékoztató a tudományos kutatás tervezésének, igazgatásának és szervezésének nemzetközi irodalmáról 1961. évi 1—2. számában, 52—61. lapon; vö. a 17. jegyzet végével. — Szépe György ismertetése, Émile Delavenay „An Introduction to Machine Translation" című művéről: Magyar Nyelvőr 85. 487—89 (1961).

[41] I. A. Melcsuk, О машинном переводе с венгерского языка на русский: Проблемы кибернетики: 1. 222—64 (1958). — Vö.: Kónyi Sándor, A magyar—orosz gépi fordítás néhány problémája: Kötetünkben 197—204. — Lásd Melcsuk másik magyar vonatkozású írását is: Некоторые выводы общего характера в связи с машинным переводом с венгерского языка: Бюллэтэнь объедения по проблемам машинного перевода 6. 34—43 (1958).

éppen ezért elsősorban a grammatikai elemzés alkotja ezen csoportok jelenlegi kutatási területét. Ők nem kívánnak egyelőre demonstratív jelentőségű próbafordításokat bemutatni. — A másik tábor viszont először fordítani akar: akárhogy, de ténylegesen fordítani. Természetesen már ez is jelentékeny előmunkálatokat igényel. A fő súlyt azonban ők a már meglevő kidolgozott fordítási folyamat bővítésére, javítására helyezik. Erre a legjobb példa az amerikai Ramo—Wooldridge csoport néhány év előtti félautomata fordítóprogram-javító módszere.[42] Ez az elképzelés rekurzív műveletek sorozatával fokról fokra kívánja emelni a fordítás megbízhatóságát. A seattle-i egyetem módszere is némileg ehhez hasonlít. — Ennek a dilemmának a feloldását a moszkvaiak jól szervezett komplex kutatásaiban láthatjuk. Viszont minden iskolától, még a szélsőségesektől is van mit tanulnunk, hogy kialakítsuk a magunk által követendő módszer részleteit is.[43]

4.2. A második probléma a gépi szótár és a gépi nyelvtan viszonya. Természetesen minden kutató egyetért abban, hogy a lexikon és a grammatika együttesen alkotja a nyelvet, mégis egyesek az előbbire, mások az utóbbira helyezik a hangsúlyt a gépi fordítást előkészítő munkájukban. A grammatikus nyelvészek, a matematikusok és a logikusok általában a nyelvtant, a szemantikus nyelvészek, a filozófusok és a mérnökök viszont a szótárt tekintik fontosabbnak a legtöbb esetben. Akik a szótárból indulnak ki, azok óriási memóriájú számológépekre építenek, amelyekbe egy nyelv (sőt két nyelv) valamennyi szavát és szóalakját beadagolják; s a fordítás első szakasza lényegében a szóalakok megfelelésének a kikeresése. A további szakaszokat alkotják a grammatikai finomítás koncentrikus körei. Természetesen, ha a szóra építenek, akkor számos szemantikai problémát kell megoldaniuk, mégpedig elsősorban a kontextus segítségével. — Akik a nyelvtanból indulnak ki, azok megelégednének aránylag kisebb memóriájú gépekkel is; ők viszont a nagyszámú és gyors kombinatorikus műveleteket igénylik. Ők is tisztában vannak azzal, hogy szótár is kell a fordításhoz; azonban egyféle szöveg lefordításához nem szükséges az egész szókészlet közreműködése. Ehhez elegendőnek tartják — az állandóan készenlétben tartott „alapszókincs"-en kívül — az egyes szakmák szókincsét tartalmazó mikroszótárak elkészítését. Elektrotechnikai szöveg esetén tehát elegendő, ha az elektrotechnikai

[42] Vö. 14. és 15. jegyzettel. A Ramo—Wooldridge-csoport néhány év előtti koncepcióját lásd például: Experimental machine translation of Russian to English (Canoga Park, Calif., 1958) 101 lap.

[43] A konferencia óta fellendülőben van a gépi fordítás az MTA Számítástechnikai Központjában. 1962 áprilisában megalakult itt a Gépi Fordítási Szeminárium; vö.: Sipőczy Győző, A gépi fordító szemináriumunkájáról: Magyar Nyelvőr 87. 272—5 (1963). — A legújabb munkáról beszámolók (1962. szeptember 13-án Tihanyban „A matematika alapjai, matematikai gépek, és alkalmazásaik" című kollokviumon): Hell György: Prinzipieller Aufbau der russisch—ungarischen Maschinenübersetzung; Sipőczy Győző: Problems of programming some grammatically homonymic forms. — Számos részletes munkabeszámoló sajtó alatt van különböző orgánumokban.

mikroszótárt tartalmazó memóriatároló egységet használjuk, mivel itt nem valószínű, hogy előforduljanak a mélyépítés körébe tartozó műszavak és jelentésárnyalatok. Természetesen a grammatikus módszer hívei a szóalakokat tőre és toldalékra bontják, s a szuffixumokat a grammatika körébe vonják. — A probléma elvi részében alighanem a grammatika elsőbbségét hangoztatóinak van igazuk.[44] Semmilyen kiváló grammatikai módszer mellett nem haszontalan azonban, ha a gép memóriája is tágas. Igen gyakran előfordulhat ugyanis, hogy egy tudományos közlemény túllépi a mikroszótár hatáskörét. Az egész szókincsnek, a thesaurusnak a fogalomköri felosztása, a seattle-i csoport tudományszakok szerinti szóanyagcsoportosítása és a mikroszótár elgondolás valószínűleg egy nevezőre hozhatók. A kérdést természetesen a számológépek fejlődése is befolyásolja.[45]

4.3. A harmadik probléma a közvetlen (például oroszról magyarra) történő fordítás és a közvetítő nyelv (az interlingua) segítségével történő fordítás viszonya. — Az interlingua kutatás legintenzívebben a szocialista országokban folyik. Ez nem véletlen. A szocialista országok között megvalósítható az a nagyarányú nemzetközi együttműködés, amelyet az interlingua módszerű fordítás megkövetel. Továbbá a szocialista országokban nemcsak az idegen nyelvről a saját nyelvre, hanem a saját nyelvről az idegen nyelvre való fordítást is fontosnak tartják.[46] — Az interlingua melletti főérv igen egyszerű. Egy nyelvről egy másikra való fordítás esetén egy fordítási folyamatra van szükség. (Ha mindkét nyelvről kölcsönösen egymásra fordítunk, akkor kettő lesz a fordítási műveletek száma.) 10 (illetőleg általában n) nyelv egymás közti közvetlen és kölcsönös fordítása esetén azonban már 90 (illetőleg n (n—1)) fordítási folyamatra van szükség. Ha viszont egy közvetítő nyelvet alkotunk meg, akkor minden egyes nyelvről erre az interlinguára kell fordítani, a másik nyelv pedig az interlinguáról fordít az anyanyelvre. Ha interlingua segítségével fordítunk, akkor egy nyelvnek a másikra való fordítása esetén egy fordítási folyamat helyett kettőre van szükség, tehát ez kétszeres munkát jelent. 10 (illetőleg n) nyelv kölcsönös fordítása esetén azonban 90 (illetőleg $n(n$—1)) folyamat helyett csak 20-ra (illetőleg 2 n-re), ez pedig óriási megtakarítást jelent. Világos, hogy az interlingua csak ott gazdaságos, ahol a nyelveknek szélesebb körére terveznek fordítást.

De mi is ez az interlingua? Van olyan nézet, hogy ha egy meglevő természetes nyelvet használunk fel közvetítő nyelvnek, akkor még kevesebb

[44] A grammatika gépi fordítási alkalmazásának kérdéseire vö.: Y. Bar-Hillelnek a 2. jegyzetben hivatkozott műveit. — Vö. még: Andreas Koutsoudas eltérő véleményével: Information Retrieval and Machine Translation 767.

[45] Kalmár Lászlótól tudtam meg, hogy már a mai fejlődés is lehetővé tesz 100 000 000 alfa-numerikus karakter (körülbelül hatszor annyi bit) nagyságrendű permanens memóriát egy millió dollár alatti áron. Az úgynevezett molekuláris elektronika fejlődése ezt a kapacitást előreláthatólag nagyságrendekkel emelni fogja.

[46] Vö. P. Sgall: Über internationale Zusammenarbeit in der Maschinenübersetzung (előadás 1962. szeptember 13-án „A matematika alapjai, matematikai gépek, és alkalmazásaik" című kollokviumon); Dezső László: A csehszlovák közvetítő nyelv nyelvtana (előadás 1962. december 11-én az MTA Számítástechnikai Központ Gépi Fordítási Szemináriumán).

folyamatra van szükségünk, 10 (illetőleg n) nyelv esetében csak 18 (illetőleg $2(n-1)$) folyamatra. Ez a könnyítési javaslat azonban „megkönnyíti" az interlingua gondolatát minden előnyétől, amivel a minden esetlegességnek kitett természetes nyelvekkel szemben rendelkezik. Igaz, hogy nem kis feladat egy mesterséges közvetítő nyelvnek a kidolgozása, azonban elméletileg ez helyesebb útnak látszik, mint a könnyítési javaslat.

Az interlingua természetére vonatkozóan két vélemény áll egymással szemben. Melcsuk a fordításban szóba jövő nyelvek közös tulajdonságaiból készítené el az interlingua grammatikáját (természetesen ez egy kategóriahálózat volna). Andrejev (és „mutatis mutandis" Margaret Masterman) pedig a szóba jövő nyelvek tulajdonságainak összességét teszi az interlingua alapjává. (Andrejev kutatásai a szemantikára is kiterjednek.) Halmazelméleti terminológiával: Melcsuk interlinguája az egyes nyelvek tulajdonságai halmazának metszetén, Andrejevé pedig a halmazok összegén alapul.[47] Andrejev megoldása biztosabbnak, de fáradságosabbnak látszik; Melcsuké viszont egyszerűbb, elegánsabb és bizonyos nem elérhetetlen föltételek mellett ugyanolyan megbízható lehet. — Természetesen a közvetítő nyelv módszere az interlingua kialakítása mellett igényt formál alapos független nyelvi analízisek és szintézisek elkészítésére. Az interlingua módszer sokat kér a grammatikától, és sokat ad a nyelvkutatásnak. — A közvetlen fordítás és az interlingua viszonyában a jövő — azt hiszem — az interlingua segítségével történő fordítást helyezi majd előtérbe. Természetesen ez nem azt jelenti, hogy az interlingua segítségével történő fordítás előkészítése már most közvetlenül feladatunk volna. Egyelőre erre már csak azért sem kerülhet sor, mivel a tanulás számára jobbnak látszik a közvetlen fordítás módszere, s mivel az interlingua-kutatások ma még nem érték el azt a szintet, hogy nyugodtan rájuk lehetne építeni.

5. A gépi fordítás és a nyelvészet viszonya

A gépi fordítás és a nyelvészet viszonyát nem lehet azzal a sommás megállapítással elintézni, hogy a gépi fordítás csupán „alkalmazott tudomány", vagy „alkalmazás"; tehát csupán felhasználja a nyelvészet eredményeit, illetőleg a nyelvészet és más alaptudományok eredményeit. Még kevésbé igaz, hogy a gépi fordítás a nyelvtudományon kívül eső tevékenység, mérnöki probléma; se nem kérhet a nyelvészettől, se nem adhat a nyelvészetnek semmit. — Azt hiszem, közelebb járunk az igazsághoz, ha úgy tekintjük, hogy a gépi fordítás olyan komplex diszciplína, amely több irányú és szoros kapcsolatban van a nyelvészettel. Attól kapja a nyelvvel

[47] N. D. Andrejev, Машинный перевод и проблема языка-посредника: Вопросы Языкознания 6/5. 117—21. (1957); N. D. Andrejev—Sz. Ja. Fityialov, Язык-посредник машинного перевода и принципы его построения: Тезисы конференции по математической лингвистике (Leningrád, 1959) 53—60. — I. A. Melcsuk, Модель языка-посредника для машинного перевода: Тезисы конференции по машинному переводу (Moszkva, 1958) 20—23. — Ennek a kérdésnek a bemutatásában Papp Ferencnek (a 6. jegyzetben hivatkozott) összefoglalása volt a forrásom.

kapcsolatos műveleteinek alapját; s ezért nagy és súlyos igényeket támaszt a szinkrón nyelvészetnek szinte minden ágazata iránt. Másrészt azonban a gép sajátos tevékenysége bizonyos mértékig — és sajátos szempontból — ellenőrzi a szinkrón nyelvészet megállapításainak egzaktságát.[48] A nyelvi analízishez tartozó egyik szintézisnek is tekinthető a gépi fordítás. Az pedig ismeretes, hogy milyen bizonyító ereje van a szintézisnek az analízist, illetőleg az analízis alapjául szolgáló elméletet illetően. A gépi fordítás már előkészítésének stádiumában is sok eredménnyel gazdagította az orosz[49] és az angol szinkrón nyelvészetet — a gépi fordítással közvetlenül össze nem függő ágazatokban is.

A gépi fordítás grammatikája a nyelvészet grammatikájától két vonatkozásban tér el: a gépi fordítás szövegeinek területe több szempontból szűkebb, mint a köznyelv egészéé, tehát a gép grammatikája is szűkebb. (Melcsuk algoritmusában például nem szerepel az igék egyes és többes számú tegező alakja, hiszen a műszaki nyelvben nem valószínű ezeknek az előfordulása.) A gép grammatikájának másik sajátossága az, hogy fogalmai és szabályai általában sokkal mechanikusabbak (például itt a szótő definíciója egyelőre körülbelül a következő: a toldalékok leválasztása után megmaradt leghosszabb közös betűsor). Azt hiszem, hogy sem ez a két különbség, sem a többi nem elvi jelentőségű. A gép grammatikája — kellő kidolgozás után — valószínűleg majdnem azonos lesz egy írott nyelvre épített dinamikus, generatív grammatikával. — Természetesen idáig a gépi fordítási kísérletek a legváltozatosabb grammatikák, illetőleg nyelvészeti elképzelések alapján mentek végbe. Valószínűleg bármilyen grammatika alapján lehet bizonyos eredményeket elérni ezen a téren. Sőt nem minősíthető teljesen eredménytelennek a jóformán grammatika nélküli automatikus-szótárszerű megközelítés sem. Csakhogy az inadekvát grammatikák által elért eredmények költségesebbek és tökéletlenebbek lesznek. A szóról szóra összeillesztéssel szemben a morfémáról morfémára történő szembesítés például ugrásszerűen megjavíthatja az eredményeket már a munkálat kezdeti szakaszában is. (Lényegében ezt jelenti raglebontás bevezetése az automatikus szótárhoz képest.) Természetesen a szóval mint az írott nyelv egyik alapegységével továbbra is számolni kell. A kérdés lényege az, hogy mennél egyszerűbben formalizálható és mennél dinamikusabb egy grammatika, annál inkább felhasználható a gépi fordítás céljaira is.[49a]

[48] A gépi fordítás és a nyelvészet viszonyára vö. még „A generatív grammatika transzformációs modelljéről" című tanulmányomat: Általános nyelvészeti tanulmányok (Szerk.: Telegdi Zsigmond; Bp., 1963) 278. — Itt jegyzem meg, hogy a nemgépfordító nyelvész T. A. Sebeok a gépi fordításnak a nyelvészetre való „feedback"-jéről (kettőjükből álló rendszeren belüli visszacsatolásáról) beszél: Proceedings of the National Symposium on Machine Translation 194. — A szűkebb értelemben vett nyelvtudományon kívül a gépi fordítás nyilvánvalóan hatással lesz a filológiai, sőt tágabbról a humán diszciplínák egész körére is, vö. W. P. Lehmann, Machine translation: agent of the humanities: Studies in Literature and Language 3. 163—70 (1961).

[49] Vö.: T. M. Nyikolajeva: О русском языке в зарубежных работах по машинному переводу: Вопросы Языкознания 10/5. 122—30 (1961).

[49a] Kalmár Lászlónak köszönöm a problémának következő tömör megfogalmazását: „Minden generatív grammatika több lehetőség közötti szabad választások

Mivel a gép egyelőre csak írott szövegeket fordít — bár ismeretesek beszédfordítási kísérletek is —, a gépi fordítás megköveteli a grafematikával való foglalkozást.[50] Itt részben olyan részletkérdések megoldásáról van szó, mint például a transzkripció, a központozás, az összetett szavak írása, a homográfia kérdése stb. A grafematika kutatásában azonban alapvető a beszéd és az írás viszonyának, izomorfizmusának és különbségének világos szabályokba foglalása. Az írott nyelvben nem számolhatunk sem a hangsúllyal, sem a hanglejtéssel, sem a beszédszünettel: itt a szórend kérdéseit az írott szóláncolat tagjainak formális elemzése révén kell megállapítani.

A gépi fordítás egyik további igénye a szakmai szókincs feldolgozására és a terminológia szabványosítására irányul.[51] Itt tekintettel kell lenni az egyes szakmák stilisztikai konvencióira is. Természetes a mikroszótárak tárgykörének elkülönítése és belső elrendezése némileg a külső tárgyi világ nyelvi vetülete; elkészítésében a szótárszerkesztő mellett szükség van a „lexikoníró" és a „tájékoztató könyvtáros" szempontjaira és esetleg munkaerejére is. Tisztábban nyelvészeti feladat az általános szókincs és a szakmai területek közös szókészletének minőségi és mennyiségi elrendezése.

Itt említem a szemantikával kapcsolatos általános problémát. Ez abban áll, hogy egyelőre nincsen korszerű leíró és nyelvészeti szemantika. Nagyon problematikus, hogy a nyelven kívüli kiindulású szemantikából végül is mi használható fel a gépi fordításban. Természetesen amit ebből formalizálni lehet, mégpedig a grammatikai algoritmussal összhangban, az programozható is, tehát felhasználható. Nagyon valószínű azonban, hogy a modern grammatikából, valamint a transzfer-grammatikából kiinduló szemantikai vizsgálatok hozzák majd meg e kérdések kielégítő megoldását.[52]

A gépi fordítók nem feledkezhetnek meg arról, hogy a gépi fordítás a fordítás egyik fajtája. Éppen ezért a fordításelmélet és fordítói gyakorlat eredményeit állandóan felhasználják.[53] A fordítógép munkafolya-

---

sorozatán alapul (ettől tudja a nyelv különböző mondatait generálni). Itt arról van szó, hogy a célnyelv generatív grammatikájának szabad választásait a transzfer grammatika vezérelje a forrásnyelvi mondat alapján, hogy ne akármilyen célnyelvi mondat, hanem épp a forrásnyelvi mondat fordítása generálódjék."

[50] Vö. például: R. A. Crossland, Graphic linguistics and its terminology: Mechanical Translation 3. 8—11 (1955).

[51] Vö.: Allen Kent, Practical goals for common language standards: Information Retrieval and Machine Translation 1275—80. — Ezen a téren felhasználandók a magyar műszaki nyelvvel kapcsolatos kutatások eredményei is: Anyanyelvi műveltségünk (Szerk.: Benkő Loránd, Bp., 1960) 77—134; Grétsy László—Wacha Imre, A műszaki nyelv művelése: Nyelvőr Füzetek 8. szám (Bp., 1961).

[52] A szemantikával kapcsolatban lásd még Kötetünkben (291) tett megjegyzéseimet is.

[53] A fordítás nyelvészeti vonatkozására vö.: International Journal of American Linguistics 20. kötetének (1954) 4. számát („Translation issue") Z. S. Harris, C. F. Voegelin, C. F. Hockett és mások közreműködésével; C. Rabin, The linguistics of translation: Aspects of Translation=Studies in Communication, No. 2. (London, 1958)

mata a kézifordító munkáját utánozza. Csakhogy az ember helyét itt egy logikai automata foglalja el. Ebben áll a gépi fordítás előnye, de hátránya is. Mert bármennyire is egyenletesebb, olcsóbb és gyorsabb az ideális gépi fordítás, a fordítást elvégző gép azonban merevebb az emberi fordítónál: csak azt tudja, amire közvetlenül vagy közvetve megtanítják. A heurisztikus gépek, vagyis a programjukat heurisztikus módszerekkel fejlesztő univerzális számológépek remélhetőleg sokkal intelligensebbek lesznek. Esetleg „megtanulhatnak" olyasmit is, amit egy rutinos fordító évtizedek munkája alatt halmozott fel „ösztönösen".[54]

A grammatika és a fordításelmélet találkozási pontján helyezkedik el a nyelvészetnek az az ágazata, amely a gépi fordítás számára legközvetlenebb fontosságú. Ez az összevető nyelvtan, vagy újabb divatos nevén a transzfer grammatika.[55] A transzfer grammatika Harris klasszikus fogalmazásában két vagy több grammatika strukturális különbségének szabályokba foglalása. Ha ez a transzfer-grammatika kiegészül egy generatív résszel is, akkor nagyjából megkaptuk a gépi fordítás grammatikai bázisát.[56]

A transzfer grammatikában résztvevő nyelvek szaporításával eljutunk a Harris által „Z" (vagyis: végső) grammatikának nevezett formációhoz. Ez voltaképpen azonos Melcsuk interlingua modelljével. — Ugyanerre az eredményre vezethet — más kiindulópontról — a nyelvi tipológia, amely a genetikus kapcsolattól függetlenül foglalkozik szinkronikus (illetőleg pankronikus) viszonylatban nyelvi struktúrák összevetésével.[57]

123—45; Roman Jakobson, On linguistic aspects of translation: On translation (Szerk.: Reuben A. Brouwer; Cambridge, Mass., 1959) 232—39. — Legutóbb az Anthropological Linguistics foglalkozott a fordítással; a vitát összefoglalta: D. H. Hymes, Discussion of the symposium on translation between language and culture: Anthropological Linguistics 2/2. 81—4 (1960). — Az 1962. évi nemzetközi nyelvész kongresszuson hangzott el N. P. Andrejev „Linguistic aspects of translation" című referátuma (8 sokszorosított lap).

[54] A nyelvészeti elemzéssel (de még nem a gépi fordítással) kapcsolatos első heurisztikus programozási elgondolás: Paul L. Garvin: Automatic linguistic analysis — a heuristic problem: On Linguistic Method (Hága, 1964) 78—97; elhangzott az 1961-i teddingtoni konferencián.

[55] Az összevető transzfer grammatikára nézve magyarul vö. Papp Ferenc hozzászólását: Általános nyelvészet, stilisztika, nyelvjárástörténet (Szerk.: Kniezsa István, Bp., 1956) 81—87, valamint a 6. jegyzetben hivatkozott összefoglalását. — A transzfer grammatikára idegen nyelven vö.: Z. S. Harris, Transfer grammar: International Journal of American Linguistics 20. 259—70 (1954).

[56] Ilyen jellegű munkálatok folynak a Massachusetts Institute of Technology Yngvevezette csoportjában; vö. Edward S. Klima három tanulmányát: Structure at the lexical level and its implications for transfer grammar (Teddington, Middlesex, 1961) 13 lap; Correspondance at the grammatical level: Quarterly Process Report No. 64 (January 15, 1962, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Mass.) 212—29; Relatedness between grammatical systems (Research Laboratory of Electronics, April 16, 1962) 14 lap.

[57] A tipológiára modern összefoglalás: Roman Jakobson, Typological Studies and Their Contribution to Historical Comparative Linguistics: Proceedings of the VIII International Congress of Linguistics (Oslo, 1958) 16—25. — Legújabban lásd a T. N. Molosnaja által szerkesztett tanulmánygyűjteményt: Структурно-типологические исслевдования (Moszkva, 1962) 297 lap. — A matematikai nyelvészet tipológiai vonatkozásainak szintézis-kísérlete: Solomon Marcus, Typologie des langues et modèles logiques: Acta Mathematica Academiae Scientiarium Hungarcae 14. 269—281 (1963).

A transzfer-grammatikát kiegészíti az úgynevezett kontakt-nyelvészet,[58] vagyis a gyakorlatban konkrétan érintkező, interferáló nyelvek sokoldalú vizsgálata. A kontakt nyelvészet — úgy gondolom — főleg a szókincs, valamint a szemantika területén segítheti a gépi fordítást.

6. A hazai gépi fordítási munka néhány kérdése

A következőkben röviden a hazai gépi fordítás időszerű gyakorlati problémáira térnék ki. Úgy vélem, hogy a társadalmi gyakorlat és a nemzetközi tudományos fejlődés fölvetette a problémát. Ezek után szervezett formában hozzá kell látni a megoldás alapos előkészítéséhez.

A gépi fordítási munkálatok nálunk a következő intézményeken belül összpontosulnak: a Magyar Tudományos Akadémia Számítástechnikai Központjában (és a Műszaki Egyetem Gépészkari Idegennyelvű Lektorátusán), valamint a debreceni Kossuth Lajos Tudományegyetemen. A további munka szervezeti feltétele, hogy a csekély létszámú és mellékhivatású csoportok szilárd és megfelelő szélességű kereteket kapjanak — megfelelő személyi, költségvetési és gépi lehetőségekkel együtt. Ugyancsak fontos, hogy olyan grammatikát tudjunk a fordítógépesek kezébe adni, amelyre biztosan építhetnek; ennek elkészítése természetesen csak egy nyelvészeti intézményen belül kialakuló kollektíva feladata lehet.

Mivel a gépi fordítás előkészítése nem egy helyen folyik (esetleg más munkacsoportok megalakulásának lehetőségével is kell számolnunk), szükség volna valamilyen szervre, amely a hasonló területen működő csoportok munkáját összehangolná.

A gépi fordítás műszaki kérdéseiről nyilván nálam hivatottabbak, a műszakiak fognak szólni. Az mindenesetre a nyelvészek számára is világos, hogy a jelenleg hazánkban rendelkezésre álló elektronikus számológépek csupán kísérleti célokra megfelelők. Üzemszerű fordításra a jelenlegieknél jóval nagyobb kapacitású univerzális számológépek, vagy pedig speciális fordítógépek szükségesek. A kísérletek jó részét: az algoritmusok kipróbálását és korrigálását természetesen a jelenlegi gépeken is el lehet végezni. Azonban ha a gépi fordítást komolyan vesszük, akkor figyelembe kell venni végleges megoldásának gépi igényét is.

Arra a kérdésre szeretnék rátérni: milyen nyelvből és milyen tárgykörből lehetséges és szükséges a fordítás. A lehetségességen azt értem, melyik nyelvből állnak rendelkezésre megfelelő nyelvész kutatók, akik föltehetően nem zárkóznak el ilyen jellegű feladatoktól. A szükségességhez az Országos Fordító és Fordításhitelesítő Iroda 1960. évi fordítási

---

hazánkban is bemutatásra került az 1962. szeptemberi tihanyi kollokviumon. — A gépi fordítás és a tipológia kapcsolatára vö.: P. Sgall—P. Novák, Пражская типология и модель языка: Тезисы конференции по математической лингвистике (Leningrád, 1959) 11—2; Dezső László: Заметки о связе МП с типологией (előadás 1962. szeptember 13-án Tihanyban „A matematika alapjai, matematikai gépek, és alkalmazásaik" című kollokviumon).

[58] A kontakt nyelvészet klasszikus összefoglalása: Uriel Weinreich: Languages in Contact. Finding and Problems (New York, 1953), bő bibliográfiával; magyar ismertetése: Filológiai Közlöny 7. 439—40 (1961).

kimutatása[59] — és bizonyos egyéb meggondolások — alapján tudok hozzászólni. Az említett iroda 1960. évi magyarra való fordítási teljesítménye szabvány ezer oldalban a következő volt:

oroszról	94	csehről	24	a további nyelvek sorrendje: szlovák,
németről	82	franciáról	19	bolgár, svéd, olasz, szerb, spanyol, hol-
angolról	54	lengyelről	17;	land, ukrán, kínai.

Megjegyzendő, hogy az irodán kívül még több fordítóiroda és vállalalat fordítórészleg működik az országban. A műszaki szövegek magyarra történő fordításáról a Műszaki Könyvtár fordításnyilvántartó részlegének van áttekintése.[60]

Ezek az adatok és egyéb megfontolások bizonyítják annak helyességét, hogy fordítóink elsősorban az oroszról magyarra való fordítás előkészítését kezdték meg. Ezenkívül idővel foglalkozni kell a további öt nyelvből való fordítás kérdésével is. A német, a cseh és a lengyel a szocialista országok közti együttműködés keretében interlinguával vagy közvetlen fordítással is elvégezhető (bár lengyelből egyelőre nincsen erre érdeklődő szakemberünk). A két nagy nyugati nyelv fordítása (különösen az angolé) nyilván nagyon fontos; elvégzésüket megkönnyíti az, hogy szovjet kollégáink igen sokat elvégeztek ezeknek fordítás céljára történő elemzéséből. Természetesen sokat tanulhatunk az amerikaiak által elvégzett oroszanalízisekből is. Aránylag jó független analízisek és szintézisek állnak rendelkezésre az oroszból, az angolból és a csehből. Sajnos a német terén még nem megfelelő a helyzet; mire azonban az időszerű lesz, számolni lehet a strukturális német gramatika nagyarányú föllendülésével.

A fentiek alapján azt hiszem, hogy első feladatunk továbbra is az oroszról magyarra való fordítás marad, idővel azonban meg kell kezdeni a németről, az angolról és talán a csehről való fordítás előkészítését is. Természetesen igen ajánlatos volna továbbfejleszteni a Melcsuk által elkezdett magyarról oroszra, s általában idegen nyelvre történő fordítás munkáját. Ez összekapcsolható az interlingua-fordítás számára szükséges független magyar analízis és szintézis elvégzésével, amely a gépi kivonatolás és az információ-feldolgozás más fajtái számára is haszonnal jár. Azt talán mondani sem kell, hogy a független analízis és a szintézis visszavitt bennünket az alapkutatáshoz: a korszerű strukturális magyar grammatika elkészítéséhez.[61]

A fordítás t á r g y k ö r e inek a problémája már sokkal nehezebb. Már maga a tárgykörök szerinti csoportosítás is igen vitatható és ingadozik világszerte. Az említett kimutatás[59] mindenesetre tizennégy tárgykörre osztja a mintegy háromszázezer szabványoldalnyi fordított anyagot. A hét legnagyobb tárgykör a következő (ezer szabványoldalban):

[59] Egy 1961. május 20-i keltű hivatalos levélben. — Ez úton is köszönjük az Országos Fordító és Fordításhitelesítő Irodának — közelebbről Havas Bencének — a kimutatás összeállítását.

[60] Vö. Körner Györgyné hozzászólásával: Kötetünkben 262—264.

[61] 1962 utolsó negyedében az MTA Nyelvtudományi Intézetén belül megalakult az Általános és strukturális nyelvészeti Csoport, melynek egyik feladata a magyar nyelv formális (generatív) grammatikájának előkészítése.

81 gépgyártási technológia (gép- és gépjárműgyártás, anyagvizsgálat);
42 elektrotechnika, finommechanika, híradástechnika, film, műszergyártás és villamos berendezések;
40 bányászat, kohászat, energiatermelés;
27 építőipar (magas- és mélyépítés, építőipari gépek, fűtés, hűtés);
22 politikai gazdaságtan, statisztika, kereskedelem, pénzügy;
18 vegyiipar, gyógyszeripar;
14 mezőgazdaság és élelmiszeripar.

A fenti tárgyköri megosztás a szakosított fordítóirodák adatai révén nyilván radikálisan átalakul. Azt a következtetést azonban már így is levonhatjuk, hogy ezek a tárgykörök igen változatosak. S általában a termelés körébe vágnak: az előkészítés grammatikai oldalának egyaránt változatos szövegek elemzésére van szüksége, ha a gyakorlat igényeit ki akarja elégíteni. A mikroszótárak összeállítása, beosztása külön — több szakterület összefogásán alapuló — kutatást igényel.

Végül arról ejtenék néhány szót, hogy a gépi fordítás előkészítésében körülbelüli milyen munkaszakaszokkal számolhatunk. (Természetesen mindez nagyon föltételes jellegű; sok mindentől, főleg a rendelkezésre álló lehetőségektől függ.) Az első szakasz a tanulás szakasza. Ehhez elsősorban a teljes bibliográfia elkészítése és lehetőleg a teljes külföldi szakirodalom beszerzése szükséges. Ebben a szakaszban — de később is — rendkívüli jelentősége van egy ilyen induló tudományágazat számára (az érdemleges ideig tartó) külföldi tanulmányutaknak.[62] Természetesen a már megindult fordítási kísérleteket ebben a periódusban is folytatni kell. — A további szakaszok elsősorban az anyagi, szervezeti és személyi lehetőségektől függenek. Annyi már most világos, hogy az előkészítés és az üzemszerű fordítás szakaszai elkülönülnek; az előkészítésen belül először nagyobb súly esik az elméleti előkészítésre, később pedig a kísérletekre. Az elmélet és a gyakorlat egysége ezen a téren is az egyenletes fejlődés biztosítéka.

*

Befejezésül csak annyit, hogy a gépi fordítás nem kis munka. Több szakma számos szakemberének koncentrált munkája szükséges elvégzéséhez. Ezt a munkát azonban senki más nem végzi el helyettünk sem a magyar nyelvre, a magyar nyelvről idegen nyelvre. Ezt nekünk magunknak kell itt, Magyarországon elvégeznünk.

[62] Azóta a magyarországi gépi fordítási munkálatokra nagy hatást gyakoroltak a Tihanyban 1962. szeptember 11—15-én tartott „A matematika alapjai, matematikai gépek, és alkalmazásaik" című konferencia „Matematikai nyelvészet és gépi fordítás" szekciójának ülései, valamint az üléseken kívül szakadatlanul folyt konzultáció. Különösen nagy jelentősége volt Y. Bar-Hillel előadásának „Recent developments in algebraic and computational linguistics"; valamint szintén Y. Bar-Hillel által a Budapesti Műszaki Egyetemen 1962. szeptember 17-én tartott előadásoknak, illetőleg konzultációknak. Vö. Magyar Nyelvőr 87. 157—58 (1963).

Sipőczy Győző

Az elöljárós kifejezések programozási kérdései

1. Az elöljárós kifejezések kettős funkciója

Minden fordításnak, így a gépi fordításnak is a pontos és megbízható morfológiai és szintaktikai elemzésen kell alapulnia. Ezt a kettős elemző munkát a gép csak azoknak a megkülönböztető és megkülönböztethető nyelvi jeleknek a segítségével tudja végezni, amelyek vagy mint külön morfémák, vagy mint a tőmorfémák után felsorolt információk a tárolókban kapnak helyet. Az információk egy része (szófaji különbségekre, vonzatokra stb. utaló jelek) lehetővé teszi, hogy a gép a nyitó- és zárószók rendszere alapján megállapítsa egy-egy adott szóalak mondatbeli funkcióját.

Sok esetben azonban ezek az információk vagy nem adják meg a helyes értelmezéshez szükséges összes tájékoztatást, vagy még a szintaktikai kapcsolatok megállapítására sem szolgáltatnak elegendő felvilágosítást. Erre a legjellemzőbb példákat a gépi fordítás szempontjából nálunk elsősorban fontos nyelvek — orosz, német, angol — elöljárós kifejezései adják. Mivel az elöljárós kifejezések gyakoriságuknál fogva a nyelv rendszerében centrális helyet foglalnak el, a helyes értelmezésükhöz szükséges információrendszer létrehozását a gépi fordítás egyik kulcskérdésének kell tekintenünk.

1.1. Köztudomású, hogy az indoeurópai nyelvekben az elöljárós kifejezés jelzőként és határozóként is szerepelhet attól függően, hogy főnévvel vagy melléknévvel, illetőleg igével alkot-e szintagmát. A kétféle szintaktikai kapcsolat megkülönböztetése pusztán formai információk alapján lehetetlen. Sőt az elöljárós kifejezés szintaktikai szerepének ismerete sem mindig elegendő a helyes értelmezéshez. Ezt igazolja egy közismert és sokszor idézett angol példa:

A „*He was kissed by the girl*" és „*He was kissed by the bridge*" mondatpár formailag semmiben sem különbözik: mindkettőben szenvedő ige van, utána pedig *by* elöljárós főnév. Az ambiguitást az okozza, hogy az egyik főnév eredet-, a másik helyhatározó. Ezt a fordítás szempontjából fontos különbséget a gép pusztán formai információk alapján nem tudja érzékelni. A fordítót a formai azonosság nem téveszti meg: tudja, hogy az ige „cselekvője" csak élőlény lehet, és ismerve a főnév „élőlény", illetőleg „hely" jelentését, a szöveget helyesen fogja értelmezni.[1] Mivel a gép fordításkor végső fokon ugyanazokat a műveleteket végzi, mint az ember, csak akkor

várhatunk tőle helyes fordítást, ha „képessé tesszük" a formában meg nem mutatkozó, de az értelmezés szempontjából döntő különbségek „észlelésére", azaz szemantikai tájékozódásra.

1.2. Az említett példában az elöljárós főnevek határozói szerepe világos volt, a kétértelműség abból adódott, hogy az információk nem tájékoztattak a határozók jellegéről. Nézzünk most — orosz példán — olyan esetet, amikor a formai jegyek a kifejezés mondatbeli funkciójának megállapítására sem elegendők: Az

«Они установили печь на новом заводе»
és *«Они установили печь на твёрдом топливе»*

mondatok közül az elsőben az elöljárós kifejezés az igével alkot határozós szintagmát, a másodikban pedig a печь főnév jelzője. Világos, hogy a döntés ismét csak szemantikai alapon lehetséges: a „helyet" jelentő főnév az ige határozója, az „anyagnév" a főnév jelzője.

1.3. A gépi fordítással foglalkozók lépten-nyomon emlegetik a szemantikai információk szükségességét. Luitgard és Alex Wundheiler cikke[2] angol—német példák alapján az igék szemantikai információinak fontosságáról beszél. Szemantikai megoldásokra utalt Szalai Sándor a Nyelvtudományi Társaságban tartott előadásában az „It is too hot to eat" angol mondattal kapcsolatban. (A mondat többértelműségét csak a környező mondatok szemantikai vizsgálata útján lehet feloldani.)

Az orosz—magyar gépi fordítás szempontjából figyelemre méltó L. V. Selimova egyik cikke.[3] Az elöljárós kifejezések szintaktikai kapcsolatainak megállapítására szolgáló programját nem közvetlenül fordítási céllal szerkesztette, így célnyelvi szempontokat nem kellett figyelembe vennie. A főnév után álló elöljárós kifejezés szintaktikai kapcsolatait az elöljárók szemantikai csoportosításával oldja meg, de figyelembe veszi a főnevek igei vagy nem igei jellegét és a „tárgy" jelentésű, valamint a „mozgást és térbeli kiterjedést" jelentő főnevek esetében a szemantikai ismertető jegyek szintaktikai kapcsolatot meghatározó szerepét is.

2. A mondatbeli funkció megállapításához szükséges információk

Az orosz—magyar gépi fordítás a nyitó- és zárószók rendszere alapjára épül. Ha tehát a szótári rész tárolóiban mindazokat a szavakat és tőmor-

[1] Vö. N. Chomsky: Syntactic Structures (Hága, 1957).
[2] Luitgard és Alex Wundheiler, Some logical concepts for syntax: Machine translation of languages (New York, 1955).
[3] I. N. Selimova, Установление синтаксических связей предложных групп в русском языке: Лингвистические исследования по машиному переводу, II (Moszkva, 1961, VINITI).

fémákat, amelyek elöljárós kifejezések nyitószavai[4] lehetnek, úgy kódoljuk, hogy nyitószó szerepükre vonatkozóan elegendő információt kapjunk, segítségükkel az elöljárós kifejezések szintaktikai kapcsolata pontosan meghatározható még akkor is, ha az elöljárós kifejezés nem közvetlenül a nyitószó után van. A tulajdonképpeni problémát az elöljárós kifejezést megelőző főnév jelenti: erre később külön visszatérek.

2.1. Az elöljárós kifejezéseket határozó kapcsolatban a következő nyitószók után találjuk:

2.11. Ragozott ige: . . . зависит от массы . . .

2.12. Főnévi igenév: . . . говорить о движении . . .

2.13. Teljes alakú melléknévi igenév: . . . получающейся в коксовальных печах . . .

2.14. Rövid alakú melléknévi igenév: . . . помещены в таблице . . .

2.15. Teljes alakú melléknév: . . . характерные для этого общества . . .

2.16. Rövid alakú melléknév: . . . целесообразно в механике . . .

2.17. Határozói igenév: . . . преобразовываясь в кинетическую энергию . . .

2.18. »нет«: . . . трения нет в узлах . . .

2.2. Határozós kapcsolatban van az elöljárós kifejezés akkor is, ha előtte nem nyitószó áll (kivéve természetesen a főnevet és melléknevet követő elöljárós kifejezést). Nem nyitószók az elöljárós kifejezés előtt a következők:

2.21. «можно, надо, нужно»: . . . можно без ощутимой погрешности принимать — Itt a nyitószó a fenti szavakhoz tartozó főnévi vagy melléknévi igenév.

2.22. Módosítószó vagy határozószó: К тому же в подогретом состоянии . . . она . . . близка . . . . . найдены почти во всех странах.

2.23. Névmás: . . . эксплуатация его для промышленности . . . — Az utóbbi két esetben a nyitószót az elöljárós kifejezéstől balra vagy jobbra kell keresni.

2.24. Pont, pontosvessző vagy kötőszó: Для более рационального проведения . . . — . . . при этом . . . истощается . . . — . . . что в современном машиностроении применяются . . . — A nyitószó csak az elöljárós kifejezéstől jobbra lehet.

[4] *Nyitószó*-nak nevezem itt mindazokat a szavakat, amelyek mellett — akár jelzős, akár határozós kapcsolatban — bővítményként elöljárós kifejezés szerepelhet.

2.25. Vessző: ... на расстояниях, во много раз больших ... — Следовательно, на камень, ..., наложена ... — Ha az elöljárós kifejezést vessző előzi meg, az információk alapján először a vessző szerepét kell eldönteni, a nyitószót csak ezután lehet keresni.

2.3. Térjünk át ezek után arra az esetre, amikor az elöljárós kifejezés főnév után áll. Ilyenkor vagy az előtte levő főnévvel alkot jelzős szintagmát, vagy a mondatban levő igével határozós szintagmát. Ha előtte birtokviszony van, tartozhat a birtokviszony valamelyik tagjához (mint jelző), vagy az igéhez (mint határozó). Hogy az említettek közül melyikhez tartozik, azt az előtte álló főnév, az elöljáró után az elöljárós kifejezésben levő főnév, illetőleg az ige jelentése szabja meg. Világos ugyanis, hogy az egyes főnevek és igék csak egyes elöljárókkal és egyes — meghatározott — fogalomkörökhöz tartozó főnevekkel alkothatnak szintagmát. Például a *вопрос* főnév után természetes egy *o* elöljárós főnév, mint annak jelzője, de ugyanezzel a főnévvel egy *под* elöljárós kapcsolat szokatlan lenne. Ha tehát a *вопрос* szó után olyan elöljárós kifejezés van, amellyel egyáltalán nem, vagy csak nagyon ritkán alkot szintagmát, ugyanakkor viszont természetesebb szintaktikai kapcsolatban áll az igével, az adott elöljárós kifejezés a mondatban levő ige határozója lesz.

2.31. Egészen természetes, hogy egyetlen — nyitószóként szereplő — főnév az elöljáró után álló szó jelentésétől függően több elöljárós kifejezés jelzett szava is lehet. Például a *письмо* főnév mellett jelzői szerepben gyakoriak a következők:

из	és valamilyen helyet jelölő szó;
от	és személyt vagy intézményt jelölő szó;
к	és személyt vagy intézményt jelölő szó;
о	és a tartalmat jelölő szó;
в	és helyet jelölő szó, stb.[5]

Ez annyit jelent, hogy az egyes szavakat vagy tőmorfémákat olyan információkkal kell ellátnunk, amelyekből — szemantikai jellemzőkön kívül — az is megállapítható, hogy az adott szó vagy tőmorféma melyik elöljárókkal alkothat szintaktikai kapcsolatot.

Külön kell kezelnünk az igéből képzett vagy cselekvéssel kapcsolatos szókat (ide sorolva a *вопрос*, *письмо*, *свободный*, stb. típusú szavakat is, mert ezek ugyanazokkal az elöljárókkal kerülhetnek szintaktikai kapcsolatba, mint az eredeti ige, vagy az az ige, amelynek jelentésével a szó kapcsolatban áll.

2.32. A *вопрос* szóval összefüggésben már utaltam arra, hogy a szintaktikai kapcsolatok megállapításához nem elég az adott nyitószóval szintaktikai viszonyban elképzelhető elöljárók ismerete, az is fontos,

[5] Papp Ferenc—Páll Erna—Péter Mihály—Szamuelly Tiborné: Mai orosz nyelv II. (Bp., 1958).

hogy a nyitószó az elöljárók közül — jelentésénél fogva — melyikkel kerülhet „természetesebb" kapcsolatba. Ennek megállapítására az előfordulási gyakoriság ad jó támpontot. A kódolás szempontjából ez azzal a követelménnyel jár, hogy az információknak a nyitószóval szintaktikai kapcsolatot alkotó elöljárókkal adott tájékoztatáson túl, azoknak gyakorisági sorrendjéről is kellő felvilágosítást kell adniok.

Ennek a sorrendi tájékoztatásnak a fontossága azokból az esetekből világlik ki legtisztábban, amikor az elöljárós kifejezést birtokviszony előzi meg. Ilyenkor a következő lehetőségek vannak:

2.321. Nyitószó a birtokosszó. Ez lehet igei: . . . обобщения действий над . . . величинами . . .; és nem igei: . . . число печей на твёрдом топливе . . .

2.322. Nyitószó a birtokszó. Ez ismét lehet igei: . . . Развитие машины в СССР; és nem igei: . . . патенты разных стран на паровые машины . . .

2.323. A nyitószó nem a birtokviszonyban van: . . . производить предварительную газификацию твёрдого топлива в газогенераторных установках.

A birtokviszony után álló elöljárós kifejezés szintaktikai hovatartozását és jellegét csak akkor lehet megállapítani, ha tudjuk, hogy az elöljárós kifejezés a mondat állítmányával, a birtokos szóval vagy a birtokszóval van-e szorosabb szemantikai kapcsolatban. Ehhez pedig megfelelő információra van szükség.

2.33. És ez még mind kevés. A bevezető példákban is láttuk már, hogy magának az elöljárónak a jelentése is sokszor az utána álló főnév jelentésétől függ, és így szemantikai információkra akkor is szükségünk van, amikor az elöljárós kifejezés szintaktikai szerepe egyébként világos. A tárgyesettel járó в elöljáró például — határozói kapcsolatban — helyhatározót jelöl, ha utána helyet jelentő főnév van, ha az utána álló főnév időt jelöl, a kapcsolat időhatározói lesz. Az információknak ezekre a lehetőségekre is kellő felvilágosítást kell adniok.

2.4. Az elmondottak figyelembe vételével összegezhetjük azokat a követelményeket, amelyeknek az elöljárós kifejezések szintaktikai kapcsolatainak megállapításához és értelmezéséhez szükséges információknak eleget kell tenniök:

2.41. Rá kell mutatniuk a szó igei vagy nem igei jellegére.

2.42. Tájékoztatniok kell a szóval szintaktikai kapcsolatban előforduló összes elöljáróról.

2.43. Az előfordulási szám csökkenő sorrendjében felvilágosítást kell adniok a szóval szintaktikai kapcsolatban szereplő elöljárók gyakoriságáról.

2.44. Magukban kell foglalniok az értelmezéshez szükséges szemantikai adatokat.

A gépek tárolókapacitásának korlátozott volta és egyéb információk helyigénye a megoldási lehetőségeknek szűk határokat szab. A feladat olyan információrendszer létrehozása, amelynek egy-egy tagja kis számú bitet köt le, de mégis valamennyi felsorolt követelményről kellő tájékoztatást ad.

2.5. A következőkben szeretnék felsorolni néhány elképzelhető megoldási módot.

2.51. Több témakörből vett műszaki szövegekben 23 olyan elöljárót találtam, amelyek mind jelzői, mind határozói szintagmában előfordulnak. Az információrendszer létrehozásának legegyszerűbb módja az lenne, ha a szótár tárolóiban a szavak és tőmorfémák után sorban közölnénk mind a 23 elöljárót és 1-gyel vagy 0-val jelölnénk, hogy az elöljáró alkothat-e szintagmát velük, vagy nem. Ez a 23 bitet lekötő információ azonban nem adna tájékoztatást sem az elöljárók sorrendi változatairól, sem a szó vagy tőmorféma igei vagy nem igei jellegéről, sem jelentéséről.

2.52. Az elöljárók sorrendjét is figyelembe vevő megoldást a 23 elöljáró permutációi adnák. 23 faktoriálisa 10^{22} nagyságrendű szám. Ennyi változat kifejezésre a kettes számrendszerben 75 jegyű számra, azaz 75 bitre lenne szükség. És akkor még külön jel kellene a szemantikai adatokra.

2.53. Ha az említett 23 elöljáró közül kivesszük azt a hetet, amelynek a magyarban jelzői szerepben *-i* melléknévképzővel ellátható névutók (névutó-melléknevek) felelnek meg *(között-közötti; alatt-alatti;* stb.) és megelégszünk ezek határozói fordításával jelzői szerepben is olyan formán, hogy például a *„két pont közötti távolság"* helyett *„távolság két pont között"* fordítást kapjuk, akkor „csak" 16 faktoriálisával kell számolnunk.. Ez a 14 jegyű szám az információrendszerben 45 bitet igényelne ismét anélkül, hogy a jelentésről tájékoztatna.

2.54. Ha igényeinket még tovább csökkentjük, és kiemeljük a szó vagy tőmorféma — mondjuk — 5 legjellemzőbb elöljáróját és csak ezekkel számolunk, hogy a sorrendi változatokra is tekintettel vagyunk, akkor a 23-nak ötödosztályú variációira van szükségünk. Ennek kifejezésére a kettes számrendszerben még mindig 22 bitre lenne szükség, a jelentés figyelembevétele nélkül.

2.55. Ugyanezzel az öt elöljáróval a sorrendváltozás igénye nélkül — 23 ötödosztályú kombinációiként — az információkhoz szükséges bitek számát 16-ra „csökkenthetjük". Ez — természetesen — még mindig nagyon sok, különösen, ha tekintetbe vesszük, hogy milyen kevés tájékoztatást ad.

A felsorolt lehetőségek a gyakorlatban nem valósíthatók meg, mert igen nagy bit-számot igényelnek és egy sincs közöttük, amely a szintaktikai

kapcsolatok megállapításához és az értelmezéshez szükséges valamennyi információt megadná.

2.56. A gyakorlatban is alkalmazható megoldás a következő elgondolás alapján valósítható meg.

A szavakat a szemantikai szempontok figyelembevételével oly módon kell csoportosítanunk, hogy az azonos elöljárókkal szintagmát alkotók kerüljenek egy csoportba. Ezek a csoportok természetesen nem felelnek meg a hagyományos „szótartományok"-nak, nincsenek tekintettel pl. a szófaji eltérésekre, a szorosabb értelemben vett szójelentésre, stb. Éppen ezért talán a „szóhalmaz" elnevezés illetné meg őket leginkább. A szóhalmazok összeállításakor az elöljárókon kívül figyelembe kell venni az értelmezéshez szükséges jelentéskategóriákat, valamint a szó igei vagy nem igei jellegét.[6] Ugyancsak több témakörből vett műszaki szövegek elemzése alapján úgy találtam, hogy az említett követelmények kielégítésére 52 szóhalmaz szükséges. Ezek a következők:

1. Cselekvést vagy történést jelentő szók helyi vagy időbeli kapcsolatban.
2. Cselekvést jelentő szók képes helyhatározós kapcsolatban.
3. Cselekvést jelentő szók célhatározós kapcsolatban.
4. A tárgyra irányuló befolyást, hatást jelentő szók.
5. Mozgatást vagy mozgást jelentő szók.
6. Változást jelentő szók.
7. Létezést jelentő szók.
8. Akaratot, kívánságot, követelményt kifejező szók.
9. Felosztást, felbontást jelentő szók.
10. Lehetőséget vagy szükségességet jelentő szók.
11. Állapotot jelentő szók.
12. Elhelyezést, elhelyezkedést jelentő szók.
13. Tulajdonságot jelentő szók.
14. Tartalmi vagy szerkezeti sajátosságokat jelentő szók.
15. Cselekvés eredményeképpen létrejövő tulajdonságot jelentő szók.
16. Társas kapcsolatot (ismeretség stb.) jelentő szók.
17. Távolságot, távolodást vagy távolítást jelentő szók.
18. Közelséget, közeledést vagy közelítést jelentő szók.
19. Irányt jelentő szók.
20. Tudást, ismeretanyagot jelentő szók.
21. A cselekvésre ható befolyást, okot, módot, stb. jelentő szók.
22. Függőségi viszonyt jelentő szók.
23. Hasonlóságot jelentő szók.
24. Különbséget jelentő szók.
25. Jelzést jelentő szók.
26. Tárgyakat jelentő szók, amelyeknél a térbeli elhelyezkedés dominál.
27. Anyagnevek.
28. Valamilyen célt szolgáló eszközök nevei.
29. Alakot jelentő szók.
30. Mennyiséget jelentő szók.
31. Tulajdonságot jelentő szók.
32. Mennyiségileg mérhető tulajdonságot jelentő szók.
33. Helyet jelentő szók.
34. Időegységet jelentő szók.
35. Személyt vagy közösséget jelentő szók.

[6] A felsorolásban a szavak i g e i vagy n e m - i g e i jellegét a szóhalmazok sorrendje mutatja: az első 25 halmaz igei jellegű, azaz igéket és igei alapjelentésű más szókat tartalmaz, a többi halmaz tagjai nem-igei jellegűek. Ez a sorrendi megoldás indokolja, hogy azonos szemantikai kategóriába tartozó halmazok két helyen is szerepelnek. Például az „Irányt jelentő szók' halmaza megtalálható a 19. és a 44. sorszám alatt.

36. Felületet jelentő szók.
37. Felosztást, felbontást, részt jelentő szók.
38. Lehetőséget vagy szükségességet jelentő szók.
39. Tartalmi vagy szerkezeti sajátosságokat jelentő szók.
40. Cselekvés eredményeképpen létrejövő tulajdonságot jelentő szók.
41. Társas kapcsolatot (ismeretség stb.) jelentő szók.
42. Távolságot, távolodást vagy távolítást jelentő szók.
43. Közelséget, közeledést vagy közelítést jelentő szók.
44. Irányt jelentő szók.
45. Tudományágat jelentő szók.
46. Tudást, ismeretanyagot jelentő szók.
47. A cselekvésre ható befolyást, okot, módot, stb. jelentő szók.
48. Függőségi viszonyt jelentő szók.
49. Korlátozó jellegű intézkedést jelentő szók.
50. Hasonlóságot jelentő szók.
51. Különbséget jelentő szók.
52. Jelzést jelentő szók.

A szóhalmazok után a gyakoriság csökkenő sorrendjében szerepelnek azok az elöljárók, amelyek a szóhalmaz tagjaival szintaktikai kapcsolatba kerülhetnek. Így például a „cselekvést vagy történést jelentő szók helyi vagy időbeli kapcsolatban" szóhalmaz után az alábbi elöljárókat találjuk: *в, на, с, при, без, у, перед, за, после;* a „Valamely célt szolgáló eszközök nevei" után a *для, с, в, на, без, при,* stb. elöljárók lesznek.

Mivel a kettes számrendszerben 5 számjegy csupán 32 lehetőséget, 6 számjegy pedig 64 lehetőséget képes kifejezni, ezért az 52 szóhalmaz jelöléséhez 6 bitre van szükség. Ez a 6 bites információ — megfelelő programba dolgozás után — kellő tájékoztatást ad mind a szó igei vagy nem igei voltáról, mind a szintaktikai kapcsolatok megállapításához és az értelmezéshez szükséges szemantikai különbségekről.

A 6 bites információ szerencsésen alkalmazkodik a gép egységrendszeréhez, hiszen egy-egy betű jelöléséhez is 6 bit szükséges, így a műveletek (ragleválasztás, stb.) mindig 6 bites egységekkel történnek; s a tárolókapacitást sem terheli meg túlságosan, hiszen pusztán egy betű jelölésére szükséges bit-számot foglal le.

A főnevek leírt csoportosítása sem a szóhalmazok, sem az elöljárók szempontjából nem tekinthető véglegesnek. A végleges szóhalmazokat csak akkor lehet összeállítani, amikor majd tudjuk, hogy mely szavak kerülnek a szótár tárolóiba. Mivel a 64, vagy — ha a 0-t mint információ-viselő jelet mellőzzük — 63 lehetőség közül az említett szóhalmazok csak 52-t foglalnak le, még 11 szabad hely rendelkezésre áll azoknak a szavaknak a számára, amelyek esetleg ebbe az 52-es keretbe nem sorolhatók be.

Láttuk, hogy a szintaktikai kapcsolatok sokszor csak az elöljárók sorrendje alapján állapíthatók meg, ezért itt nagyon fontos, hogy a sorrend összeállítása igen sok szövegen végzett, minden szótárba kerülő szóra kiterjedő elemzés alapján történjék.

Az egyes szavakra vonatkozó adatok gyűjtését már elkezdtem. Ez a meglehetősen hosszú időt igénylő munka egyúttal a műszaki szövegekben előforduló szavak gyakoriságáról is részletes adatokat fog eredményezni, és ezzel nagymértékben megkönnyíti a szótár és a mikroszótárak összeállítását.

2.6. A kérdés most már az, hogyan lehet az említett 6 bites információkat felhasználni a gyakorlatban. Vegyük a legnehezebb esetet: amikor az elöljárós kifejezés előtt birtokviszony van. Láttuk, hogy ilyenkor az elöljárós kifejezés szintaktikai kapcsolatának eldöntéséhez három tényezőt: az állítmányt, a birtokos szót és a birtokszót kell figyelembe venni. Ha az egyes elöljárók feldolgozásának programját úgy állítjuk össze, hogy a lépések a szóhalmazok mellett szereplő elöljárócsoportok előfordulási gyakoriságának csökkenő sorrendjében követik egymást, a gépnek csak azt kell megállapítania, hogy az adott elöljáró programjában a szintagma tagjainak megfelelő szóhalmazok közül melyik szerepel első helyen, azaz csak a szintagmában szereplő szavak szóhalmazáig vezető lépések számát kell összehasonlítania. Ez egyszerű kivonással elvégezhető. Az elöljárós kifejezés azzal a szóval alkot szintagmát, amelyik mellett a gép a kisebb számot találja. Ha ez főnév, a kapcsolat jelzői, ellenkező esetben határozói.

A szintaktikai kapcsolat meghatározása után következik az elöljáró jelentésének megállapítása az elöljárós kifejezésben levő szó szemantikai információja alapján. Láttuk már, hogy az elöljáró jelentése sok esetben az utána álló főnév jelentésétől függ. Hadd említsek itt még egy példát: ha a *письмо* szó után *из* elöljáróval helyet jelentő főnév van, az elöljáró a levél keletkezési helyére utal, de ha az *из* előtt valamilyen anyagból készíthető tárgy, utána pedig anyagot jelentő főnév van, azt fogja jelölni, milyen anyagból készült az adott tárgy vagy alkatrész; például: *деталь из стали* — 'acélalkatrész'. Az egyes elöljárók feldolgozása programjainak pontosan tartalmazniuk kell azokat a jelentésváltozatokat, amelyeket az elöljárós kifejezésben levő szó okoz.

Ennek a rendszernek az a tulajdonsága, hogy az elöljárók sorrendjének összeállításában statisztikai adatokra támaszkodik, és így valószínűségi alapon dönti el a szintaktikai kapcsolatokat. Az eddig megvizsgált példák azt mutatják, hogy a döntés helyességének valószínűsége igen nagy, a tévedés alig lehet több néhány százaléknál. A szóhalmazok további felbontásával ezt az arányt is csökkenteni lehetne, de ez már 7 bites információrendszert igényelne (127, illetőleg 128 szóhalmaz jelölési lehetőségével). A 7 bites információrendszer az M—3-as és a Magyarországon használt egyéb gépekben a „betű"-határ túllépését jelentené, megbontaná a tárolók egységrendszerét, és megnehezítené a programozást.

3. Az orosz hátravetett jelzők fordításához szükséges információk

A szintaktikai kapcsolat megállapítása csak a határozós szintagma esetén elegendő a fordításhoz, jelzős szintagma esetén további meggondolások szükségesek. Tudjuk ugyanis, hogy az elöljárós főnévvel kifejezett hátravetett jelzőnek a magyarban több fordítási lehetősége van. Ezek közül legfontosabbak az *-i*, *-s*, vagy *-ú*, *-ű* képzős melléknév, birtokviszony, összetett szó, kiegészítés jelzősítő igenévvel (néhány esetben határozószóval), néha pedig határozós szerkezet.[7]

[7] Vö. az 5. jegyzetben hivatkozott művel.

Hogy ezek közül a lehetőségek közül a gépnek melyiket kell választania, az ismét szemantikai kérdés. Ez azt jelenti, hogy a magyar szókészletet tartalmazó tárolókban szintén külön jelekkel kell jelölnünk, hogy az adott főnév jelzői szerepben a sok lehetőség közül melyikkel fordítható. Annak az anyagnak a neve például, amiből valami készül, összetett szó első tagja lesz: *vastengely*. Ha azonban az anyagot jelentő főnév edényt jelölő főnév jelzője, *-s* képzős melléknévvel kell fordítanunk: *olajoskanna*, stb. Külön kell jelölnünk azokat a jelzősítő igeneveket is, amelyekkel a jelzős kapcsolat a magyarban bővülni fog. Például ha a szintagma nyitószava *munka* és az elöljáró után helyjelölő főnév van, a jelzősítő igenév *folyó* lesz. Ugyanilyen helyzetben az *út* jelentésű szó jelzősítő igeneve *vezető*.

Vannak esetek, amikor ugyanannak a szónak jelzői szerepben többféle fordítási lehetősége van. A valamivel ellátottságot kifejező *c* elöljáró magyar megfelelője például *-s* melléknévképző, *furatos golyó*, *menetes szár*, stb. Ha azonban a szintagma utótagjának egyeztetett jelzője van, már *-ú*, *-ű* melléknévképzővel kell fordítani: *nagyfuratú golyó*, *körkeresztmetszetű rúd*, stb. Kellő információk és utasításrendszer alapján a gép ezeket a különbségeket vissza tudja adni. A programozás szempontjából legegyszerűbb, ha — jóllehet a magyarosság rovására — ilyenkor is kiegészítő szavakkal élünk: *furattal rendelkező golyó*, *nagy furattal rendelkező golyó*, stb.[8]

4. Az elöljárós kifejezések fordításának programja

Itt mindjárt felvetődik az a kérdés, hogy vajon minden elöljáróhoz külön program tartozik-e, vagy — mivel a szintaktikai kapcsolatok meghatározásához szükséges lépések az összes elöljáróra vonatkozóan azonosak — nem lenne-e előnyösebb közös program alapján elindulni, és csak a szintaktikai kapcsolat megállapítása után áttérni külön programra.

A közös program alkalmazása azzal a hátránnyal jár, hogy minden megállapított határozó-kapcsolat esetén a programban fel kell sorolni mind a 42 elöljárót. Külön programok esetén viszont az első 13 lépés (ennyiféle helyzetben lehet ugyanis az elöljárós kifejezés) minden programban külön szerepel; ez hosszabbá teszi a programot, a döntéshez vezető lépések száma azonban kevesebb lesz, mint hogyha közös programból indulnánk ki. A leggazdaságosabb megoldást olyan közös induló program adja, amely kapcsoló utasítással (switch) való szétágaztatás segítségével tér át az egyes előljárók programjára.

4.1. Vannak olyan elöljárós kifejezések, amelyek — szólásszerűen — mindig azonos formában jelentkeznek, vagy amelyeknek állandó vonzatuk van;

[8] Mind az elöljáróknak az elöljárós kifejezésben levő főnévtől függő jelentéseinek összeállításához, mind pedig a magyar szótárrész információinak megalkotásához nagy segítséget nyújtanak azok a transzformációk, amelyeket Papp Ferenc végzett: Трансформационный анализ русских присубстантивных конструкций с зависимой частью-существительным: Slavica 1. 55—83 (Debrecen, 1961). A gépi fordítás műszaki igényeinek megfelelően természetesen a pusztán szépirodalomban előforduló lehetőségeket a program összeállításakor figyelmen kívül lehet hagyni.

magyar megfelelőjük is állandó. Ezeket a szótár-résznek tartalmaznia kell. Műszaki szövegben gyakoriak a következők: *иметь дело с, наряду с, применительно с, сравнительно с, в зависимости от, в качестве, в комплексе с, в особенности, в отличие от, в отрыве от, в свою очередь, в пределе, в результате, в связи с, в большинстве случаев, в частности, в существенной части, в ущерб, за счёт, к тому же, на основе, по отношению к, по сравнению с, при помощи, с помощью, с точки зрения.* A gép ezeket azonnal megtalálja, lefordítja, ilyenkor további elemzésre nincs szükség. Az elöljárók feldolgozása programjában ezek szerepelnek első helyen, hogy az elkerülhető elemzések mellőzésével az algoritmus rövidebb lehessen.

4.2. Ezután következik az elöljárós kifejezés mondatbeli szerepének megállapítása a nyitószók alapján. Az algoritmus itt is rövidebb lesz, ha a program a nyitószókat az előfordulási gyakoriság fordított sorrendjében tartalmazza.

4.3. Ha az elöljárós kifejezés előtt nem nyitószó van, a nyitószót a szótárban a szó vagy tőmorféma mellett levő információk alapján kell megkeresni. A fenti esetekben — a főnév utáni helyzet kivételével — a kapcsolat határozói.

4.4. Ha az elöljárós kifejezés előtt főnév van, a szintaktikai kapcsolat megállapításához a következő lépések szükségesek:

4.41. Ha a főnév igei jellegű, a gép a szemantikai információk alapján megállapítja, hogy az elöljárós kifejezés a mondat állítmányához, vagy az igei jellegű főnévhez tartozik-e. Ha mindkettőhöz tartozhat, a döntés a már ismertetett módon az elöljárók sorrendje alapján történik. Ennek eredményeképpen fordítja a kifejezést határozónak vagy jelzőnek.

4.42. Ha a főnév nem igei jellegű, meg kell állapítani, hogy az elöljáró és az elöljárós kifejezésben levő főnév jelentése lehetővé teszi-e, hogy a kifejezés az adott főnévvel szintagmát alkosson. Ha igen, akkor azt is meg kell állapítani, hogy az elöljárós kifejezés nem alkothat-e szintagmát ugyanakkor az igével is. Ha a kifejezés mind az igével, mind az előtte álló főnévvel alkothat szintagmát, ismét az elöljárók sorrendje, azaz a szóhalmazhoz vezető lépések száma dönt.

4.43. Ha az elöljárós kifejezés előtt birtokviszony van, a döntést a gép a már korábban ismertetett lehetőségek figyelembevételével és a vázolt sorrendi összehasonlítások alapján hozza meg.

4.5. A szintaktikai kapcsolatok megállapításához vezető lépések után a program felsorolja az adott elöljáró mellett számbajöhető összes főnév információs számát, és a gép ennek, valamint a megfelelő magyar adatot tartalmazó tárolónak az információja alapján megállapítja az elöljáró jelentését és a kifejezés fordítási módját.

4.6. Még oly gondos programkészítés mellett is előfordulhat, hogy a gép a mondat egyik szavában sem találja meg az elöljárós kifejezés nyitószavát. Ezért a program utolsó utasítása mindig az, hogy a kifejezést határozónak kell fordítani. A fel nem ismert szintaktikai kapcsolat egyébként mindig vagy a program hibáját jelzi, vagy az információk elégtelenségére mutat. Ilyenkor a feltárt hiányosságot utólag korrigálni kell.

A program ismertetésében nem tértem ki az elképzelhető egyéb programozási lehetőségekre, nem vettem sorba a közölt programvázlat valamennyi lépését sem, hiszen ez meglehetősen sok időt vett volna igénybe. Talán az elmondottakból is kiviláglik, milyen utasításokra van szüksége a gépnek ahhoz, hogy az elöljárós kifejezéseket az ismertetett információrendszer felhasználásával megbízhatóan fordítsa.

*

Az elmondottakban felvetett megoldási lehetőség használhatóságáról — kísérletek alapján — a gyakorlat fog ítéletet mondani. Mielőtt azonban ezekre a kísérletekre sor kerülhetne, be kell fejezni az egyes szóhalmazokhoz tartozó elöljárók sorrendjére vonatkozó vizsgálatokat. (Ezt a statisztikai alapon összeállított sorrendet a gépi kísérletek alapján esetleg utólag módosítani kell.) Ezek a vizsgálatok mind a szóhalmazokkal szintagmát alkotó elöljárókra, mind az elöljárókkal kapcsolatba kerülő szóhalmazokra vonatkozóan megadják azokat a jellemző adatokat, amelyek alapján az egyes elöljárók részletes, minden műszaki szöveg fordítására alkalmas programját el lehet készíteni.*

* Az előadás átdolgozott orosz változata időközben megjelent: Вопросы програм мирования рулсских предложных оборотов: Slavica 3. 59—72 (Debrecen, 1963).

Szalai Sándor

Elméleti és kísérleti adalékok a magyar nyelvű tudományos közlemények gépi kivonatolásának kérdéséhez* *(Kivonat)*

A gépi kivonatolás kérdését — a gépi fordítás kérdéséhez hasonlóan — a tudományos dokumentáció méretarányainak hatalmas megnövekedése tette aktuálissá. A kérdés megoldását — vagy legalábbis megközelítését — a modern matematikai-logikai, valószínűségszámítási és különösen információelméleti megfontolásoknak a nyelvi anyagra való alkalmazása teszi lehetővé. Ennek a nyelvi anyagnak (elsősorban a tudományos közlés írott nyelvi anyagáról beszélünk!) vannak bizonyos általános, a konkrét közlési nyelvtől független sajátosságai, s vannak olyan sajátosságai is, amelyek magából a konkrét közlési nyelvből — mondjuk éppen a magyar nyelvből — erednek. Úgy látszik, hogy a jó eredményű gépi kivonatolást leginkább mindkétféle sajátosság együttes kihasználásával lehet megvalósítani. Nálunk tehát az ilyen irányú kutatásoknak általános nyelvészeti és magyar nyelvészeti tanulságai egyaránt lehetnek, s ez szolgálhat talán indokul arra, hogy néhány elméleti és kísérleti adalékkal igyekezzünk itt hozzájárulni a magyar nyelvű tudományos közlemények gépi kivonatolása módszertanának tisztázásához.

Az előadás bevezetőleg áttekintést nyújt a gépi kivonatolás elméletének az információ-visszakeresés (information-retrieval) és a mutatókészítés (indexing) elméletével való kapcsolatáról, majd pedig a „nyelven belüli fordítás" sajátos változataként elemzi a kivonatkészítési eljárást. Bemutatja, hogy a szógyakoriság és az együttes szóelőfordulás bizonyos ismérveinek felhasználásával hogyan sikerült különböző kutatóknak többé-kevésbé hatékony — bár egyelőre még csak kísérleti jellegű — automatikus módszereket kialakítani meghatározott diszkurzív szövegtípusok (elsősorban tudományos közlemények) tartalmi jellemzőinek — deszkriptorainak — megválogatására, illetve a releváns szöveghelyek szelektálására. Az elért eredmény jóságát különféle mutatószámokkal lehet jellemezni (sűrítési index, relevancia-index stb.).

Mindezen módszerek gyakorlati alkalmazását — géprevitelét — rendkívül megnehezíti, hogy nagyon „memóriaigényesek", nevezetesen szükségessé teszik a köznyelvi szókincs leggyakoribb (s ugyanakkor legkisebb szövegelemző erővel, „deszkrip-

* Az elhangzott előadás alapjául szolgáló tanulmány időközben teljes terjedelmében könyvalakban megjelent: „Gépi kivonatolás. Elméleti és kísérleti adalékok a magyar nyelvű tudományos közlemények gépi kivonatolásának kérdéséhez" (Bp., 1963., Időszerű Műszaki Dokumentációs Kérdések, 5. szám).

tor-értékkel" rendelkező) szavainak az elektronikus számológép memóriájában való tárolását. Ez teszi ugyanis lehetővé, hogy a gép különbséget tudjon tenni az olyan szók között, amelyek tartalmi okokból gyakoriak a szövegben (relevánsak, magas deszkriptor-értékűek, mint például a szöveg „címszavai", „műszavai", „tárgyszavai"), s olyan szavak között, amelyek trivialitásuknál fogva fordulnak elő nagyszámú ismétlődéssel (*a, hogy, ez, csak* stb.).

Az előadó ismerteti egy saját módszertani elgondolását, amely — úgy látszik — bizonyos határok között lehetővé teszi „tárgyaló" jellegű (például tudományos) szövegek esetében, hogy a köznyelvi szógyakoriságokkal való összehasonlítás nélkül is automatikusan meg lehessen különböztetni a relevánsan és a triviálisan gyakori szókat. Közelebbről: a szöveg relevánsan gyakori szavainak bizonyos automatikus „feldúsítása" és „súlyozása" tisztán szövegen belüli ismérvek alapján a jelek szerint mellőzhetővé teszi a köznyelvi szógyakoriságok memorizálását, s nem nyújt lényegesen kedvezőtlenebb eredményt, mint az említett nagy memória-igényű (és jelentős műveleti többletet is igénylő) eddig ismert gépi kivonatolási eljárások. Az eljárás — legalábbis magyar nyelvű alkalmazásában — nem feltételezi a gépi fordítás viszonylag bonyolult szóvégződés-lebontási és más grammatikai technikáit, viszont „preventíven" alkalmazható a gépi fordításon belül több értelműségi alternatívák számának csökkentésére. Magától értetődően kísérleti eljárásról van szó, amely még jelentős mértékben fejlesztésre szorul, illetve jelenleg még csak az első ellenőrzés stádiumában tart.

Az előadó néhány szimultatív eszközökkel készített szövegkivonat-mintán mutatja be az eljárás eredményeit és fogyatékosságait, s utal a gépi kivonatkészítés problémájának néhány érdekesnek látszó általános, illetőleg magyar nyelvészeti vonatkozására.

Rényi Alfréd

Információelmélet és nyelvészet

1. Először arról szeretnék néhány szót szólni, hogy mi is az információelmélet. Lehet beszélni információelméletről szűkebb és tágabb értelemben. A szűkebb értelemben vett információelmélet a matematikának, közelebbről a valószínűségszámításnak egy új fejezete, amely az utolsó tizenöt évben jött létre és amelynek létrehozása Claude Shannon[1] nevéhez fűződik. A matematikának ez a fejezete a híradástechnikában előforduló információ-átviteli folyamatok matematikai elméleteként alakult ki; az információ-továbbítás, információ-tárolás, információfeldolgozás problémái alkotják tárgyát.

A szűkebb értelemben vett információelmélet tehát az üzenetközvetítéssel foglalkozik (függetlenül attól, hogy az üzenetközvetítés ember és ember között, vagy ember és gép, vagy gép és gép között történik) és azt vizsgálja, hogyan lehet az információt a legcélszerűbb alakban bizonyos zavaró külső tényezők („zajok") jelenlétében közvetíteni, hogy kell ehhez „kódolni", hogy kell azután a kapott üzenetet „dekódolni", stb. Ezt a szűkebb értelemben vett információelméletet a maga területén (a híradástechnikában, az elektronikus számológépek, az automatizálás elméletében stb.) nagy sikerrel alkalmazzák.

Ez a speciálisabb értelemben vett információelmélet is sok tekintetben nyújthat segítséget a nyelvtudománynak, azonban a nyelvésznek tulajdonképpen többre van szüksége: olyan matematikai elméletre, amely a nyelv mélyebb összefüggéseit is képes megragadni. Azt az elméletet, amely az információt nemcsak kvantitatív szempontból, hanem tartalmilag is vizsgálja, nevezhetjük tágabb értelemben vett információelméletnek. Már itt hangsúlyoznom kell azonban, hogy míg a szűkebb értelemben vett információelmélet egy kidolgozott matematikai diszciplína, amelyben rengeteg nyitott probléma van ugyan, de azért már ma is egy kerek és ha nem is lezárt, de fejlett elmélet, ezzel szemben az, amit tágabb értelemben vett információelméletnek nevezek, csak az első lépéseknél tart. Ez a tágabb értelemben vett információelmélet lényegében úgy jött létre, hogy észrevették, hogy a szűkebb értelemben vett információelméleti

[1] C. E. Shannon—W. Weaver: The mathematical theory of communication (Urbana, Ill., 1949).

eredményeknek a hordereje még nagyobb, mint amire a kiindulását képező problémákból következtetni lehetne. Ezzel persze korántsem akarom lebecsülni azoknak a műszaki problémáknak a jelentőségét, amelyek megoldására az információelmélet létrejött, hiszen ezek a maguk területén nagyon fontos problémák. A matematikában azonban nem meglepő, hogy egy területen kidolgozott módszerekről később kiderül, hogy egész távoleső területeken is felhasználhatók. Ugyanakkor persze ki szokott derülni az is, hogy az új területen való felhasználás érdekében e módszerek általánosítására, kiegészítésére, újabb kutatásokkal való továbbfejlesztésére van szükség. Ez a helyzet ma az információelmélet terén is. A tágabb értelemben vett információelmélet tárgyát képezheti a nyelv nemcsak mint szimbólumok sorozata (mert mint ilyen, már a szűkebb értelemben vett információelméletnek is tárgyát képezi), hanem mint tartalommal, jelentéssel bíró jelsorozat. A nyelv statisztikai tulajdonságai, statisztikai törvényei önmagukban is fontosak és érdekesek, és feltétlenül az eddiginél több figyelmet érdemelnek. Azonban a matematikai nyelvészetben ma ennél még többre törekszenek, tudniillik, hogy valamilyen módon a nyelv tartalmát is matematikailag tárgyalják.

Shannon nagyon nyomatékosan hangsúlyozta, hogy amikor egy mértékszámot adott meg, amivel az információ mennyiségét mérjük, akkor az információnak pusztán a mennyiségét vizsgálta, teljesen eltekintve a tartalmától, még inkább az értékétől. Az általánosabb értelemben vett információelméletnek tárgyát képezheti nemcsak a nyelv, hanem egy egész sereg más emberi tevékenység is, például az, ahogy egy orvos diagnózist állít fel egy betegről, a különböző vizsgálatokból nyert információkat összegezi, összehasonlítja, feldolgozza. Tárgyát képezheti ennek a tágabb értelemben vett információelméletnek például az a folyamat, ahogy egy vizsgálóbíró nyomoz. Tárgyát képezi továbbá az átöröklés, mint (genetikai) információátadási folyamat; ez egyik viszonylag eléggé kialakult fejezete ennek az általános információelméletnek. Tárgyát képezhetik a dokumentáció gépi módszerei, továbbá a tanulás folyamata, beleértve nemcsak az ember, hanem a gép „tanulását" is. Ez a tágabb értelemben vett információelmélet kétségtelenül egy rendkívül érdekes és izgalmas téma, de — még egyszer hangsúlyozom — sokkal kevésbé kialakult állapotban van ma, mint a szűkebb értelemben vett információelmélet.

Az előadásban először a szűkebb értelemben vett információelméletről szeretnék egész röviden beszélni, azután néhány szót szólni a matematikai nyelvészet perspektíváiról. Mielőtt erre rátérnék, szeretném azt hangsúlyozni, hogy a tágabb értelemben vett információelmélet csak a szűkebb értelemben vett információelméletre épülhet, hiszen nem indulhatunk ki másból, mint ami van, és csak arra próbálhatunk valami újat építeni. A kétfajta információelmélet közötti különbséget illetőleg döntőnek az információ tartalmának a kérdését érzem: amíg a tartalommal nem foglalkoztunk, vizsgálatunk a szűkebb értelemben vett információelméletbe tartozik, de ha a tartalommal is foglalkozunk, az már a tágabb értelemben vett információelméletbe tartozik. Egy hasonlattal lehet legjobban megvilágítani a helyzetet; ha valaki elmegy a postára táviratot feladni, akkor a postatisztviselő megszámolja a szavak számát, kirója a díjat és utána

leadja a táviratot Morse-jelekkel, igyekszik ezt hibátlanul továbbítani; azzal, hogy a táviratban mi van, ő nem foglalkozik (és nem is volna helyes, ha foglalkozna). Ma már technikailag az is megvalósítható, hogy az egész folyamat gépesítve legyen, és a távirat úgy érkezzék meg a címzetthez, hogy azt senki más ne is lássa; így ma már lehetséges, hogy a levéltitok mintájára a „távirat-titok" is létrejöjjön, és ez kívánatos is volna. A táviratovábbítással kapcsolatos minden probléma a szűkebb értelemben vett információelméletbe tartozik. Általános információelméleti jellegű tevékenység volna azonban a távirat-feladással kapcsolatban az olyan postatisztviselő munkája, akinek az volna a feladata, hogy tanácsot ad az embereknek, hogy hogyan lehet a táviratot más szavakkal, rövidebben megfogalmazni. Elvileg elképzelhető az is, hogy ezt a feladatot egy gép végezze el. Kétségtelen, hogy az információ tartalmának kérdése sokkal bonyolultabb, komplexebb dolog, mint a puszta mennyiségének a kérdése. Így például az információmennyiség teljesen objektív fogalom, míg a tartalom és a jelentés kérdésénél sokkal nehezebb a szubjektív és objektív elemeket elválasztani.

2. Ezen bevezetés után most a szűkebb értelemben vett információelméletről szeretnék néhány szót szólni. Ennek az elméletnek a kiindulópontja az a kérdés volt, hogy hogyan lehet az információ mennyiségét számszerűleg kifejezni, tehát egy mérőszámot adni arra, hogy egy közlés mennyi információt tartalmaz. Ebből a célból először egy egységre van szükség: az információ egységére. Ilyen egység, mint közismert, az úgynevezett „bit", amelyen az az információ értendő, amely egyetlen olyan jellel kifejezhető, amelynek két lehetséges értéke van, a 0 vagy az 1. Más szavakkal: az információ egysége egy olyan válaszban foglalt információ mennyisége, amely válasz *igen* vagy *nem* szóval kifejezhető, teljesen függetlenül a kérdés tartalmától. Egy tetszőleges közlésben foglalt információ mennyiségét számszerűleg úgy lehet mérni, hogy azt a közlést kifejezzük ilyen két értékű (0 vagy 1 értéket felvenni képes) jelek sorozatával és a jelek számával mérjük az információ mennyiségét. Ez minden fajta közlés esetében elvileg is és gyakorlatilag is lehetséges. Amennyiben számokról van szó, akkor át lehet írni a számokat a kettes számrendszerbe. Ha az információ írott szöveg, meg lehet számozni a betűket és írásjeleket (a szóközt is beleértve), és ki lehet fejezni a betűk (írásjelek) sorszámát a kettes számrendszerben. Persze minden közlés sokféleképpen fejezhető ki 0 és 1 jelek sorozatával; egy közlésben foglalt információ mennyiségén azt értjük, hogy a közlést a leggazdaságosabb módon kifejezve 0 és 1 jelekkel, hány jelből fog állni. Könnyen be lehet látni, hogy ha egy „nyelv"-ben n jel van (n „betűből" álló ábécéje van a nyelvnek), akkor e „nyelv" egy betűjének átírásához átlagosan $\log_2 n$ 0- vagy 1-jel szükséges ($\log_2 n$ az n szám 2 alapú logaritmusát jelöli). Az $I = \log_2 n$ formulát nevezik Hartley-féle formulának. A Hartley-formula csak abban az esetben alkalmazható, ha a szóban forgó nyelv „betűi" egyformán valószínűek, azaz általában mind körülbelül ugyanolyan gyakran fordulnak elő. Az általános esetben, amikor a „nyelv" betűinek valószínűségei a $P_1, P_2, \ldots, P_n$ számok, a nyelv egy betűjének átírásához átlagban

$$I = P_1 \log_2 \frac{1}{P_1} + P_2 \log_2 \frac{1}{P_2} + \ldots + P_n \log_2 \frac{1}{P_n}$$

0- vagy 1-jelre van szükség; ez az úgynevezett Shannon-féle formula. Ezzel még korántsem merült ki az információelméletben használt információ-mértékszámoknak az összessége; a Shannon-félétől különböző mértékszámok is vannak,[2] melyek néha hasznosak ugyanezen információnak a mérésére. De vannak más fogalmak is, amelyekre itt nem térhetek ki részletesen, mint például a feltételes információ, az információnyereség és a relatív információ fogalma.

Az információelmélet legalaposabban kidolgozott ága a zajos csatornán keresztül való információ-továbbítás elmélete. A következő sémával szokták jellemezni ezt a feladatot: adva van egy információ-forrás, amely bizonyos jeleket (szimbolumokat) produkál, ezeket kell továbbítani egy „zajos" csatornán keresztül. Azon, hogy a csatorna „zajos", az értendő, hogy a továbbítás folyamán a jelek eltorzulhatnak, és a felvett jel nem lesz azonos a leadott jellel. Továbbá egy ilyen csatornának megvan az a jellegzetessége, hogy csak meghatározott típusú jelek közvetítésére képes, és csak meghatározott típusú jeleket szolgáltat; szóval egy ilyen csatornának van egy bemenő ábécéje és van egy kimenő ábécéje. Így az információ-forrás által produkált jeleket valamilyen módon át kell alakítani olyan jelekké, amelyeket a csatorna képes továbbítani; ez a kódolás. Ezt oly módon kell elvégezni, hogy ez az átalakítás a zajokkal szemben minél „ellenállóbb" legyen. Azután a csatorna másik végénél a felvett jeleket dekódolni kell, tehát át kell alakítani őket olyan jelekké, amilyeneket a felvevő kíván. Például a televíziónál az eredeti jel egy kép, ezt a képet felbontja a felvevőgép először is kis elemekre, ezeket az elemeket azután átalakítja elektromágneses hullámokká, ezek az elektromágneses hullámok terjednek a téren át, ami ez esetben a „csatorna" szerepét tölti be; a légköri zavarok, a légköri elektromosság, egyéb adók működése következtében ezek a jelek kissé eltorzulva érkeznek meg a televíziós készülékhez, amely visszaalakítja őket képpé.

Az információelméletben szerepel egy nagyon fontos fogalom, amiről még szólni szeretnék, ez a r e d u n d a n c i a fogalma. Az előbb azt hangsúlyoztam, hogy egy közleményben foglalt információ mennyiségének a megállapításánál azt nem akárhogy kell átírni 0 és 1 értékű jelekből álló sorozattá, hanem a leggazdaságosabb módon. Mármost arról, hogy egy szöveg az adott jelkészlet segítségével a leggazdaságosabb módon van-e kifejezve, nemcsak 0 és 1 jelekből álló szöveg, hanem tetszőleges

[2] Vö. Rényi Alfréd, Az információelmélet néhány alapvető kérdése: A Magyar Tudományos Akadémia Matematikai és Fizikai Osztályának Közleményei 10. 251—282 (1960); On measures of entropy and information: Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability (University of California Press, 1961), I, 547—61; Az információ-akkumuláció statisztikus törvényszerűségeiről: A Magyar Tudományos Akadémia Matematikai és Fizikai Osztályának Közleményei 12. 15—31 (1962); Egy információelméleti problémáról: A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei 6 B. 505—16 (1961).

szimbólumokból álló szöveg esetében is lehet beszélni. Ha a szimbólumok felhasználása nem a leggazdaságosabban történik, akkor azt mondjuk, hogy a szöveg redundáns; mégpedig a redundancia számszerűleg azt adja meg, hogy a szövegnek tulajdonképpen hányadrésze az, ami felesleges, tehát hányadrészével válna rövidebbé a szöveg, ha a leggazdaságosabban használnánk fel a rendelkezésünkre álló jelkészletet.

Ebben az értelemben a természetes nyelvek persze erősen redundánsak; ezt azonban semmiképpen sem szabad a nyelv rovására írni. Éppen ellenkezőleg, a redundanciának a nyelvben rendkívül pozitív szerepe van, és egy redundancia nélküli nyelv gyakorlatilag használhatatlan volna. A redundanciára tehát szükség van, például azért, mert ha egy mesterséges nyelv abszolút redundanciamentes volna, az azt jelentené, hogy nemcsak az igaz benne, hogy minden szimbólum egyformán valószínű, hanem az is igaz, hogy egyes szimbólumok egymástól teljesen függetlenül követhetik egymást. Egy ilyen „nyelv"-ben tehát minden elképzelhető jelsorozat értelmes szöveg. Viszont nyilvánvaló, hogy a betű- (illetőleg hang- vagy fonéma-)sorozatok túlnyomó része kiejthetetlen. Tehát már ezért sem képzelhető el egy redundancia nélküli nyelv. Másodszor, ha egy „nyelv"-ben a betűk tetszőlegesen választott sorozatának értelme volna, ezáltal nagyon megnőne a félreértés lehetősége. Azáltal, hogy a természetes nyelv redundáns, elég nagy százalék sajtóhibát szinte teljes biztonsággal korrigálni lehet. Szóval, bár az információelméletben általában célkitűzés a felesleges redundancia megszüntetése, ez nem lehet olyan célkitűzés, amit a természetes nyelveknek követnie kellene.

A természetes nyelv redundanciája teszi lehetővé, hogy megértsünk egy olyan szöveget, amelynek — azáltal, hogy esetleg a helyiségben zaj van — csak egy elég kis százalékát halljuk ténylegesen, talán csak 50—60 százalékát. Ha tisztában vagyunk is azzal, hogy a redundancia a természetes nyelvben szükséges és pozitív funkciói vannak, mindenesetre érdemes azt tudni, hogy ez a redundancia egy szövegben mekkora. Arra vonatkozólag, hogy ezt hogyan lehet meghatározni, a Magyar Nyelvtudományi Társaságban tartott előadásomban[3] ismertettem egy eljárást. Hangsúlyozni kell ezzel kapcsolatban, hogy a redundancia függ a szövegtípustól. Lehet beszélni általában a magyar nyelv redundanciájáról, de ezen belül lehet külön a szépirodalomnak vagy a napi sajtónak a redundanciájáról, a szépirodalmon belül lehet az egyes írók stílusának a redundanciájáról beszélni, és ezek a számértékek egymástól elég nagy mértékben eltérhetnek. A redundancia-meghatározás is felhasználható az egyes stílusok közötti különbségek kimutatására. A redundancia kérdését tulajdonképpen mint tipikus példát említettem arra, hogy milyen eredményeket adhat a szűkebb értelemben vett információelmélet fogalmainak alkalmazása a természetes nyelvekre. Kétségtelenül ez is vezet kézzelfogható, nyelvtudományi, sőt esetleg még irodalmi szempontból is érdekes eredményekre.

[3] Információelmélet és nyelvtudomány. Előadás a Magyar Nyelvtudományi Társaság általános nyelvészeti és fonetikai szakosztályában 1962. február 20-án. Vö. Magyar Nyelv 59. 121 (1962).

Szeretném azt is hangsúlyozni, hogy a redundanciába a nyelvtani struktúra nagyon erősen belejátszik, szóval a redundancia nemcsak egyszerűen a betű-, illetőleg szóstatisztikától függ. Hiszen ha egy bizonyos szónak néhány betűje meg van adva, akkor gyakran a nyelvtani szabályok alapján ki lehet találni a hiányzó betűket. Tehát a redundancia fogalmára nem lehet azt mondani, hogy ez csak a nyelv felszínes jelenségeit ragadja meg, mint mondjuk a betű-gyakoriságok. De azért ez a fogalom még mindig nem hatol túlságosan mélyre.

Mint mondottam, az általános információelmélet még csak a kezdeti lépéseket tette meg; gondolok itt például Bar-Hillel munkáira. Még távol vagyunk attól például, hogy olyan kérdéseket, mint az irodalmi nyelvben az utalásokkal való információközlést, matematikailag tárgyalni tudnánk. Egy szó tartalmához ugyanis hozzátartoznak azok az összefüggések, amelyekben az illető szót a mindennapi életben használják, azok az összefüggések, amelyekben az illető szó (különösen egy-egy ritka szó) valamilyen közismert klasszikus irodalmi műben szerepel, és ezek segítségével az író fel tudja idézni az egyes szavakhoz az irodalmi köztudatban fűződő hangulatot, atmoszférát. Különösen gyakran élnek ezzel az eszközzel — amelyet rezonancia-keltésnek is nevezhetünk — a költők. Az információ-közlésnek sok ilyen rejtettebb módja van, amelyet szintén nem szabad figyelmen kívül hagyni a nyelv információelméleti vizsgálatánál.

3. Befejezésül csak egész röviden még arról szeretnék beszélni, hogy mit várhatunk az információelmélet nyelvészeti alkalmazásaitól. Először is szeretném még egyszer hangsúlyozni, hogy annak, amit szűkebb értelemben vett információelméletnek neveztem, az alkalmazása (vagyis lényegében a nyelv statisztikai vizsgálata) önmagában is rendkívül értékes felvilágosításokat adhat, és nemcsak hasznos, hanem egyenesen nélkülözhetetlen például a gépi fordításban, a szótárkészítésben, de a stílusvizsgálatoknál is. Azonban van ennél egy talán még jelentősebb perspektívája az információelmélet nyelvészeti alkalmazásainak, mégpedig az, hogy az információelmélet a természetes nyelveket beállítja egy szélesebb kategóriába, más információközlési módszerek közé, és így lehetővé teszi az összehasonlítást, az egyezések és a kontrasztok kiemelését, és ezáltal elősegíti a nyelvre vonatkozó ismereteink elmélyítését.

Melyek azok az információközlési módszerek, amelyekkel párhuzamba lehet állítani a nyelvet? Ezek közt említem először a matematikának a formanyelvét, és a matematikai logikának a formanyelvét, továbbá a különböző mesterséges nyelveket, mint például az ALGOL, vagy a Lincos.[4]

A matematikai nyelvészet eredményei felhasználhatók a nyelvtanulás racionálisabb alapon való megszervezésénél is. Arra gondolok, hogy — azt hiszem — mindenki, aki valamilyen idegen nyelvet megtanult, azt legtöbbször nem úgy tanulta meg, hogy végig követte a nyelvtanulás előírt tankönyvszerű menetét, hanem egy bizonyos fokon „szabadúszóvá" vált, elkezdett olvasni, eleinte esetleg még szótárral, később azonban már szó-

[4] H. Freudenthal: Lincos. Design of a language for cosmic intercourse (Amsterdam, 1960).

tár nélkül. Aki így tesz, tulajdonképpen azt a statisztikai törvényszerűséget használja, ha nem is mindig tudatosan, hogy ha egy könyvet olvasok egy idegen nyelven, akkor az új szavaknak a száma, amelyekkel találkozom, exponenciálisan csökken. Az új szavakkal pedig az ember többször is találkozik és az összefüggésből összehasonlítással ki lehet találni a szónak a jelentését (amennyiben a szavak nagy részét ismeri az olvasó). Tehát a nyelv redundanciáját fel lehet használni a nyelv tanulásánál is.

Meg vagyok róla győződve, hogy az elkövetkező években és évtizedekben még tucatszám fognak születni a legkülönbözőbb célokra szolgáló mesterséges nyelvek és hogy ezeknek a mesterséges nyelveknek a létrehozása révén megértünk majd bizonyos dolgokat a természetes nyelvekről, olyan dolgokat, amelyekre különben nem jöttünk volna rá; ugyanúgy, ahogy a számológépek megalkotása útján értettük meg az agyműködés bizonyos jellegzetességeit. Úgy hiszem, ilyen tekintetben is nagy perspektívái vannak a matematikai nyelvészetnek.

Ruzsiczky Éva

Az elektromechanikus adatfeldolgozó gép nyelvtudományi alkalmazásának néhány kérdéséről

1. Bevezetés

1. Nyelvtudományunkban egyre nagyobb tért hódít az a meggyőződés, hogy a statisztikai módszert az eddiginél fokozottabb mértékben kell alkalmazni. Nagyobb arányú statisztikázás alig képzelhető el egyes kutatók vagy akár kutatócsoportok eddig szokásos cédulázó anyaggyűjtése alapján; így akarva-akaratlanul fölvetődik a kérdés: a megbízható eredményekhez okvetlenül szükséges adatmennyiség előállításában vajon nem kaphatnánk-e s ha igen, milyen jellegű és mértékű segítséget kaphatnánk a modern technikától, pontosabban az adatfeldolgozó gépektől. Így nem lesz talán haszon nélkül, ha beszámolok az adatfeldolgozó gép nyelvtudományi alkalmazásának arról a néhány kérdéséről, amellyel 1960. szeptemberében Berlinben nyílt alkalmam megismerkedni, részint közvetve,[1] részint közvetlenül azon a munkaértekezleten, amelyet a Wörterbuch der deutschen Gegenwartssprache munkatársai Wissmann professzornak és a Grimm szótár több munkatársának részvételével, Ruth Klappenbach osztályvezetőnek a vezetésével abból az alkalomból tartottak, hogy az épp akkortájt vásárolt adatfeldolgozó gép üzembe helyezése időszerűvé vált.

2. A lyukkártya-módszer

Legelőször is néhány szót az úgynevezett lyukkártyákról, illetőleg a lyukkártya módszerű adatfeldolgozásról, ennek gépeiről, pontosabban gépi egységeiről, s teljesítőképességükről.[2] A lyukkártyák nem egyebek, mint olyan bal felső sarkukon vágott, 187,4 × 82,5 mm nagyságú, 9,18 mm vastagságú kartotéklapok, amelyeken szinte végtelen mennyiségű adatot föl lehet dolgozni. 80 vagy 90 rubrika van rajtuk (attól függően, hogy milyen gyártmányú a gép), mindegyik rubrika fel van osztva 10 részre, a 10 részt egy-egy szám jelzi:

[1] A (Mitteilungsblatt für die Arbeiter der deutschen Akademie der Wissenschaften zu Berlin cikkeiből: 6. 243—50, 255—58 (1960).

[2] Legutolsó magyar összefoglalása: Glacz Béla—Haraszti Ferenc—Zrinszky Oszkár: A lyukkártya-rendszer általános ismertetése (Kézirat, Bp., 1961. Sokszorosítva kiadja a Központi Statisztikai Hivatal Ügyvitelgépesítési Főosztálya).

1	1	1	1	1	1	1	1	1	1	
2	2	2	2	2	2	2	2	2	2	
3	3	3	3	3	3	3	3	3	3	
4	4	4	4	4	4	4	4	4	4	
5	5	5	5	5	5	5	5	5	5	
6	6	6	6	6	6	6	6	6	6	
7	7	7	7	7	7	7	7	7	7	
8	8	8	8	8	8	8	8	8	8	
9	9	9	9	9	9	9	9	9	9	
1	2	3	4	5	6	7	8	9	10	stb.

A 80, illetőleg 90 db 10 számjegyből álló oszlop hatalmas számot ad ki, minimálisan olyat, amelyben az egyes (illetőleg az egyesek) után 79 vagy 89 zérus következik, tudniillik hátulról visszafelé haladva az első oszlop az egyeseket, a második a tizeseket, a harmadik a százasokat, a negyedik az ezreseket, az ötödik a tízezreseket, a hatodik a százezreseket jelzi — s a sort folytathatjuk a 80., illetőleg a 90. oszlopig. Minden egyes számhoz — 1-től egészen a 10^{79}, illetőleg a 10^{89}-ig — egy-egy adatot lehet hozzárendelni olyanformán, hogy egy előre kidolgozott rendszer, p r o g r a m szerint a kartotékra vezetendő adatnak megfelelő számot a géppel kilyukasztjuk, illetőleg kilyukasztatjuk. (Ha egy följegyzendő adatnak mondjuk 35 782 a száma, hátulról visszafelé az első oszlopban kilyukasztjuk a kettest, a második oszlopban a nyolcast, a harmadik oszlopban a hetest, a negyedik oszlopban az ötöst, az ötödik oszlopban a hármast. A gyakorlatban persze nemigen szokás program kialakításában 1-től a 10^{79} vagy 10^{89}-ig elmenni; a 80, illetőleg a 90 oszlopot különböző szempontok szerint 2—3, esetleg több oszlopot magába foglaló osztályokra lehet osztani, így egy-egy kártya annyiféle szempontnak megfelelően alkalmazható, ahányféle osztályra oszlik.

A gépi adatfeldolgozáshoz elsősorban lyukasztó gépre, rendező gépre, táblázó (tabelláló) gépre és válogató gépre (kollátorra) van szükség, de kívánatos még ellenőrző lyukasztó (kontroll) gépnek, másoló gépnek és feliratozó gépnek (interpreter) az alkalmazása is. A l y u k a s z t ó g é p arra szolgál, hogy vele a lyukkártyákra a programnak megfelelően rávezessük, belelyukasszuk a kívánt adatokat. A r e n d e z ő g é p a kész, azaz kilyukasztott kártyákat berendezi a magazinba, a beléjük lyukasztott szempontok szerint; a lyukkártyákat annyiszor kell a gépen átengedni, ahány számjegyből áll a rendezés alapjául szolgáló szempont a programban. A t á b l á z ó g é p e k tulajdonképpen íróműszerkezettel ellátott számológépek, s különféle összesítések készítésére alkalmasak. Kétféle módon összesítenek: List-menetben és Tab-menetben (az első esetben az összegen kívül minden egyes kártyából kiírják az adatokat, a második esetben azonban csak csoportonként, pontosabban csoportváltozáskor a csoport első kártyájából). Némely táblázó gépek csak számadatok kiírására alkalmasak, mások szöveges adatokat is ki tudnak írni; az előbbieket numerikus, az utóbbiakat alfabetikus táblázó gépeknek nevezik. A v á l o g a t ó g é p n e k az a feladata, hogy a magazinból a megadott szempontoknak megfelelően ki-, illetőleg összeválogassa az adatokat, a kártyákat.

Az ellenőrző gépek hasonlítanak a lyukasztó gépekhez, csakhogy nem lehet velük lyukasztani, arra alkalmasak, hogy a már kilyukasztott kártyákat ellenőrizzük velük. A másoló gépek a kártyák teljes vagy részleges átmásolására valók. A feliratozó gép mintegy hatvan írókarral rendelkezik; az írókarok számok és betűk felírására egyaránt alkalmasak, s olyankor használatosak, amikor a kártyákat gép nélkül kívánjuk leolvasni vagy kartotékként is föl akarjuk használni.

A típustól függően a lyukasztó gép óránként mintegy 8—10 000, a rendező gép 25—40 000, sőt 60—120 000, a táblázó gép 6—8 000, a válogató gép 12—78 000, a feliratozó gép 2000—6000 kártya feldolgozására képes. Nyilvánvaló: ilyen óriási teljesítő képességű gépek alkalmazása csak akkor fizetődik ki, ha hatalmas arányú munkára, adatfeldolgozásra vállalkozunk. Innen érthető, hogy Besançon vagy Gallarate — mint ismeretes — valóságos adatfeldolgozó teleppé épült ki: a munkatársak a teljes szókincs, továbbá a teljes alak- és mondattan feldolgozását tűzték ki célul (Besançonban naponta mintegy 30—50 000 kártyát állítanak elő), úgyhogy a jövőben minden romanistának készen, korszakok, szerzők, nyelvjárások stb. szerint rendezve s megszámolva rendelkezésére állnak majd a szükséges adatok, csak ki kell őket értékelnie.[3]

A Német Tudományos Akadémián jelenleg egyelőre három szempontunkból fontos területen folyik az adatfeldolgozó gépek lehetőségeinek kikísérletezése, sőt már kiaknázása is: a bibliográfiában, a nyelvtanírásban és a szótárírásban.[4]

3. A lyukkártya felhasználása a könyvészetben

A bibliográfiában az első feladat a helyszám följegyzése. Ha valamely szakterületen mondjuk 100 000-re tervezzük a „numerus currens"-t, az első öt oszlopból alakítunk számára osztályt. A szerző nevét kétféleképpen is feltüntethetjük: vagy úgy, hogy a szerzők nevéhez egyenként hozzárendelünk egy-egy számot s ezt lyukasztjuk, vagy oly módon, hogy a különféle betűknek juttatunk egy-egy számot s ezekből lyukasztással mintegy kirakjuk a szerző nevét. Ez utóbbi úgynevezett betűszámos (alfabetikus-numerikus) eljárásnak előnye, hogy a név a kulcs ismeretében rögtön leolvasható, nem kell visszakeresni. Hasonló módszerrel rákerülhet a kartotéklapra a mű címe, tárgya, sőt esetleg rövid tartalma is, aztán a megjelenés helye, éve. — Természetesen ezek az adatok a kézi erővel készített kartonokra is rákerülhetnek és rá is kerülnek. A gépi úton előállított kartotéklapoknak az előnye voltaképpen jobb felhasználhatóságuk-

[3] A fenti külföldi munkálatok könyvészete megtalálható például: E. és K. Delavenay: Bibliography of mechanical translation (Hága, 1960).

[4] A berlini Német Tudományos Akadémia keretében működik egy külön ezzel a témakörrel foglalkozó munkabizottság: Komitee für technische Hilfsmittel und maschinelle Informationsverarbeitung. A bizottság munkáinak összefoglalása: Die Technik hilft der wissenschaftlichen Arbeit in: Forschung—Lehre—Praxis. Schriftenreihe der Gewerkschaft Wissenschaft, 1962. 4. füzet; magyar ismertetése: Szépe György: A Magyar Tudományos Akadémia Nyelv- és Irodalomtudományi Osztályának Közleményei 19. 364—66 (1962).

ban rejlik. Mechanikusan lehet őket szelektálni. Ha tehát szükségünk van mondjuk valamely szerzőnek összes művére, cikkére vagy valamely kérdés egész szakirodalmára, ezt a gép szinte pillanatok alatt elő tudja nekünk adni a magazinból. Szükség esetén bonyolultabb feladatokat is rá lehet a gépre bízni. Például ilyesmit: adja elő a finnugor nyelvek képzőinek, azon belül külön a vogul és a magyar nyelv képzőinek eredetére vagy alaki problémáira vonatkozó szakirodalmat, 1920-tól 1960-ig, a szerzők neve szerint ábécérendbe szedve. Az ilyen feladatokat a gép többszörös rendezés (szortírozás) útján oldja meg; az anyag minden egyes rendezési (szortírozási) folyamattal korlátozódik.

Persze a gépi bibliográfiának nem egyetlen előnye az, hogy igen gyorsan besorolja és kiemeli a kártyákat. Mondhatni „minőségi" előnye is van: segítségével egészen a dokumentációs anyag megjelenéséig mind kiküszöbölődnek a közbülső munkafolyamatok hibalehetőségei; miután a kártyákat előállítjuk és ellenőrizzük, később semmiféle hiba nem fordulhat elő (föltéve, hogy a gép műszakilag kifogástalanul működik). A gép ugyanis nemcsak arra képes, hogy a kártyákat először szakterületenként, aztán szerzők szerint elrendezze, kiemelje, hanem arra is, hogy az elrendezett, illetőleg kiemelt kártyákat „visszafordítsa" írott nyelvre, áttegye gép-, illetőleg nyomtatott írásba, ezt aztán fotokópiák útján át lehet vinni matricákra, s már lehet is kezdeni a nyomtatást, mégpedig korrektúra nélkül.

4. A lyukkártya-módszer alkalmazása a nyelvtani munkálatokban

4. Az a munkaközösség, amely Berlinben jelenleg a lyukkártyáknak a nyelvtani munkálatokban való felhasználásán fáradozik, a mai német nyelv strukturális elemzését tőzte ki célul.[5] Az alapvető kérdések ezzel kapcsolatban a következők: milyen egységeket foglal magában a nyelv, a nyelvi rendszer; milyen viszony van az egyes egységek között; miképpen lehet az egységekből megnyilatkozást formálni? Az egyes egységek két hierarchiát képviselnek. Az első a fonémákból, a szótagokból és a hangsúlyozási egységekből adódik. Erre ráépül a második, amelynek egységei — mindig magasabb szinten — a morfémák, a morfémák kombinációi, azaz a szavak, a szavak kombinációi, vagyis a szerkezetek (mint amilyenek a névszós csoportok, az igés csoportok) s végül maguk a megnyilatkozási egységek. (A morfémák egybeeshetnek bizonyos fonémákkal vagy a szótagokkal, de egybeesésük egyáltalán nem szükségszerű, következésképpen egy bizonyos morfémának sem mindig ugyanazok a morfémacsoportok felelnek meg.) Nyilvánvaló: a második hierarchia komplikáltabb, mint az első, tehát főként azt kell vizsgálni, hogy ennek milyen a struktúrája, illetőleg az első hierarchiához való viszonya. A fő problémák; milyen szófajok, morfémakapcsolatok milyen szerkezetté kapcsolódhatnak össze; milyen fonéma-, illetőleg szótagláncok milyen morféma-

[5] M. Bierwisch, Die Verwendung von Lochkarten bei der grammatischen Analyse: Die Technik hilft der wissenschaftlichen Arbeit 94—102; a munkaközösség neve: Arbeitsstelle für Strukturelle Grammatik der Deutschen Akademie der Wissenschaften zu Berlin.

konstrukcióknak felelnek meg? — A nyelvtan csak akkor felelhet meg rendeltetésének, ha a nyelv minden megnyilatkozási egységének a felépítését ellentmondásmentesen meg tudja magyarázni, ha a nyelv minden jelenségének alkatát meg tudja világítani. Egy ilyen nyelvtan csak hatalmas adatmennyiség statisztikai kiértékelésén alapulhat; létrehozásához ugyanis minden síkon valamennyi jelenséget állandóan át kell tudni tekinteni, a nyelvre jellemző megnyilatkozásoknak hatalmas tömegét fel kell dolgozni, ki kell elemezni. Érthető tehát, hogy Berlinben a munkatársak lyukkártyákat kívánnak az adatgyűjtéshez felhasználni: a lyukkártyák lehetővé teszik nemcsak az adatok gyors átrendezését, hanem a különféle szempontok szerinti párhuzamos elemzését is. De ne vágjunk a kérdés elébe, haladjunk sorjában.

Az elmondottakból következik, hogy a munkaközösség tagjai elsősorban a második hierarchia analizálásához szükséges adatokat szándékoznak lyukkártyákra rávezetni. A különböző egységek számára más-másféle lyukkártyamagazinokat kívánnak alkalmazni. A lyukkártyák három osztályra tagolódnak, az igés csoportok, az egyéb (elsősorban a névszós) csoportok és a morfémák számára. Gyakorlatilag mindig az egész megnyilatkozási egység egészében rákerül a kártyára. Minden kártya annyi példányban készül, ahány egységet az elemzett megnyilatkozási egység tartalmaz. Ez lehetővé teszi, hogy minden egység a maga helyére kerüljön a magazinban. E megnyilatkozási egységből például: *Im Steigen kamen mir wieder Zweifel, ob der alte Mann uns wirklich Einsicht in die unschätzbaren Schriftstücke gewähren würde*, egy kártya az *ob der alte Mann uns ... gewähren würde* résszel az igés szerkezetek magazinjába, más három kártya az *Im Steigen, Zweifel ob der alte Mann uns ... gewähren würde, Einsicht in die unschätzbaren Schriftstücke* részeknek megfelelően a névszós szerkezetek magazinjába kerül. A 3., vagyis a morfémáknak való magazin megtervezésekor elengedhetetlen bizonyos ésszerű korlátozás: nem volna kifizetődő például minden egyes többes számnak vagy múlt időnek stb. egy-egy külön kártyát juttatni. Elégséges, ha a szótőknek juttatnak egyet: ezekből ugyanis kiolvasható, hogy az említett s a hozzájuk hasonló toldalékok mi módon kapcsolódnak.

A jelzett egységek természetesen előre kidolgozott k u l c s, helyesebben kulcsrendszer szerint kerülnek a kártyákra. A kulcsnak először is minél differenciáltabbnak kell lennie. Aztán lehetővé kell tennie, hogy az egységeket típusonként a sorrendjük szerint egyaránt ki lehessen keresni. Végül minden egyes szempont számára, amely szerint csak kérdést lehet tenni, biztosítania kell egy vagy több számoszlopot. E föltételeknek megfelelően például a névszós csoportok számára való kulcs először azt a kérdést veti fel, hogy vajon a csoport önálló megnyilatkozás-e vagy valamely magasabb konstrukció része; hogy benne az egy vagy több névszó milyen esetben áll. A válaszok (tudniillik, hogy — példánkat véve alapul — *a)* önálló megnyilatkozás, *b)* igés szerkezet alanya; accusativus, nominativus stb.) egy-egy nekik fenntartott oszlop megfelelő helyére kerülnek. Aztán lehet kérdezni azt is, hogy a névszós kapcsolat milyen egységekből áll. Itt számbajöhetnek: a névszói mag (többnyire alany), egy vagy több jelző, határozó, értelmező, prepozíciós csoport stb. Mindezek a szá-

mukra lefoglalt oszlopon egy-egy helyet kapnak, annak megfelelően, hogy az egység a szerkezetben a hányadik helyet foglalja el. Természetesen: a két oszlopcsoporthoz még tartalékoszlopok is csatlakoznak, így még más, újabb szempontoknak megfelelő adatokat is lehet a kártyára rávezetni, s a szempontok szinte korlátlanul szaporíthatók (már tudniillik a kártya nyújtotta lehetőségek felől nézve a dolgot). — A megnyilatkozási egységen, illetőleg a belőle kielemzett adatokon kívül természetesen még a forrásjelzések (a szerző neve, a mű címe, megjelenési helye, ideje stb.) is rákerülnek a kártyára a nekik kijelölt oszlopba, rendezhető módon, megadott kulcs szerint.

A kártyák fölépítéséből már következtetni lehet felhasználásuk, pontosabban szortírozásuk módjára. Tegyük fel, hogy tudni akarjuk a következőket: ha egy névszós csoportban van három jelző, ezek milyen sorrendben következnek vagy következhetnek egymás után, milyen viszonyban vannak egymással; egy szerkezet egyszerre hány egységet tartalmazhat s milyeneket; mikor következik tárgyesetben álló névszó vagy jelzői értékű genitivus prepozíciós csoport után? A géppel ennek megfelelően nagyon gyorsan ki lehet kerestetni *a)* azokat a névszós csoportokat, amelyek önálló megnyilatkozások; *b)* ezek közül azokat, amelyek nem alanyok; *c)* az accusativusban, genitivusban álló névszókat. — A gép az adatokat rögtön számolva szolgáltatja, így azokat azonnal lehet különféle szempontokból statisztikailag értékelni: a névszós csoportok hány százaléka tartalmaz egy, két, három jelzőt; milyen gyakran használjuk az egyik vagy másik morfémát; hogy viszonylik egymáshoz a különféle morfémák gyakorisága; milyen gyakoriak mondjuk az igétlen mondatok?

Nem kétséges: a lyukkártyás módszerű nyelvtani adatgyűjtés voltaképpen nem egyszerűbb, sőt még bonyolultabb is, mint a szokásos, cédulázó módszerű. Előnye nem annyira az adatok rögzítésében észlelhető, sokkal inkább abban, hogy a feldolgozott anyagról egyszerre minden szempontból világos áttekintést nyújt, s így lehetővé teszi az adatok módszeres, statisztikai kiértékelését. Mivel a táblázó gép egy listára rávezeti — folyó szövegbe áttéve — az adatokat, mégpedig a kívánt példányszámban, lehetővé válik, hogy ugyanazt az anyagmennyiséget egyszerre több kutató is elemezze a saját szempontjai szerint. Közben a kártyák visszavándorolhatnak a magazinba, hogy ott készen álljanak esetleges újabb szortírozási műveletekre. Végeredményben tehát a lyukkártyák segítségével egyetlen alapgyűjtéssel szinte mindenfajta nyelvtani kutatásnak meg lehet vetni az alapját.

5. A lyukkártya-módszer alkalmazása a lexikográfiában

A Wörterbuch der deutschen Gegenwartssprache munkatársai a lexikográfiai egységesítéshez, továbbá a példa-, illetőleg idézetanyag ellenőrzéséhez (esetleg valamilyen okból való kicseréléséhez) veszik igénybe az adatfeldolgozó gép segítségét. Évégre kidolgoztak egy rendszert, egy programot. Pontosabban szólva először is megállapították a szortírozás (vagy ha úgy tetszik az egyeztetés, ellenőrzés)

szempontjait s kijelölték az ezeknek szükséges osztályokat, szám szerint tizennégyet, azután felsorakoztatták — egy-egy számmal ellátva — az egyes szempontok alá foglalt, illetőleg a nekik megfelelő osztályokba besorolható kategóriáit. A kategóriákat mind fölösleges volna nekem is felsorolnom, de néhány csoportjukat, valamint osztályozásuk szempontjait azért talán érdemes bemutatnom. Előrebocsátom: a lyukkártya első hat oszlopa maguknak a kidolgozandó szavaknak van osztályként fenntartva, a következő két oszlop pedig az egy-egy szóhoz tartozó kártyák száma részére képez osztályt: a több jelentésű szavaknak minden egyes jelentéséről külön kártya készül. Ilyenformán a tulajdonképpeni osztályok a 9. oszloppal kezdődnek. A 9. és a 10. oszlop az úgynevezett stílusszférák (Stilsphären) osztályát alkotja. Ilyen minősítések tartoznak ide: költői, emelkedett, vulgáris, köznyelvi, illetőleg beszélt nyelvi stb. A következő négy oszlop a stilisztikai színeződés (Stilfarbungen) alapján összefoglalható minősítéseknek van lefoglalva, mint lekicsinylő, régies, durva, ünnepélyes, papír ízű, tréfás, gúnyos, bizalmas stb. A soron következő három oszlopba az időbeli vonatkozások szerinti minősítések kerülnek, azaz az ilyenek: új, fölelevenített, elavult, régies szó, divatszó, a nácizmus szava stb. Ezután négy oszlop a földrajzi megoszlás szerinti minősítések osztályát alkotja, mint: berlini, északnémet, délnémet, svájci, osztrák stb. Most hat oszlop a 134 szakterületnek, például mezőgazdasági, orvosi stb. van osztályul lefoglalva. A következő három oszlop a foglalkozásneveket, testrészneveket, országneveket, mitológiai neveket, növény- és állatneveket stb. magában foglaló osztályt alkotja. Ezután két oszlop a szavaknak végződések szerinti osztályozása számára van lefoglalva. A 34—35. oszlopba a sajátos használatra vonatkozó tudnivalók vannak beosztva, tehát az ilyenek: milyen vonzattal él; képes, átvitt értelemben, ritkán, csak tagadólag használatos-e a szó vagy jelentés. Ezután két oszlop a kiejtésre, hangsúlyozásra vonatkozó megjegyzések részére szolgál. A 38—39. oszlopba a nyelvtani tudnivalók kerülnek: hogy milyen szófajú a szó, erős vagy gyenge ragozású-e, mi a többes száma, hogy kell fokozni stb. A 40—41. oszlopba a szó eredetére vonatkozó tudnivalók sorolhatók. A 42. oszlop a kidolgozott címszó tagolódását van hivatva feltüntetni, azaz hogy vannak-e római, arab számos, ábécés, görögbetűs jelentései stb. A 44. oszlopba a jelentéstípusok, illetőleg értelmezésfajták kerülnek, tehát például hogy a jelentés ki van-e dolgozva vagy csak szinonimával, esetleg utalószóval van-e jellemezve. Végül három oszlopba az idézett, felhasznált 227 szerző neve van beosztva.

A gyakorlati kivitelt a következő fázisok szerint képzelik el: *a)* A szócikk szerzője rögtön analizálja a kidolgozott szócikket s a kulcsszámokat (piros ceruzával) rávezeti vagy rögtön a lyukkártyára, vagy egy külön listára. *b)* A lyukkártyákba gépi erővel beleütik a megadott kulcsszámnak megfelelő lyukakat. Kétféle színű kartont alkalmaznak: egyet a már nyomdakész, sőt a már nyomdában levő s egyet a még munkában levő, kéziratos szócikkek számára. Amikor a még nem egységesített, tehát nem nyomdakész címszók kártyáit az egységesítés különféle szempontjai szerint ismételten átrendezik, azok a kártyák, amelyek átmennek a különféle egységesítési folyamatokon, azzal egyidejűleg színt is cserélnek. Mire valamennyi számbajöhető szempontból megtörténik az egységesítés, az egyes

címszókhoz tartozó kártyák mind más színűre cserélődnek. — A szükséges szortírozásokat erre kiképzett technikusok végzik a gépen; a szótáríró mindig a már együvé tartozó csomókat kapja kézbe. Megjegyzendő: a „Wörterbuch der deutschen Gegenwartssprache" munkatársai nem külön kartonokra gépeltetik a címszó egyes jelentéseit, jelentésárnyalatait, egységeit, mint nálunk az Értelmező Szótár munkatársai, hanem ívpapírra, folyó szövegként. Innen érhető: az egységesítéshez az adatfeldolgozó gép segítségét nélkülözhetetlennek tartják.

6. Feladatok, lehetőségek

6. A bibliográfia „gépesítésé"-t hazai viszonylatban elsősorban a könyvtárak feladatának látom. (Mindazáltal nem tartom lehetetlennek azt sem, hogy idővel egy-egy kutatóintézet, így a Nyelvtudományi Intézet is bevezetheti a szakirodalom bibliográfiai feldolgozásába, a dokumentációba a lyukkártya módszert.)

Az adatfeldolgozó gépeknek nyelvtani célra való felhasználhatóságát — forrásaim alapján — csak a strukturális elemzések szempontjából próbáltam — ha bemutatni nem is, hát legalább vázolni. Természetesen el tudom azonban képzelni, hogy a gépet megfelelő programozással más szempontú nyelvtani kutatásban is kitűnően lehet hasznosítani. Hogy a magam területén maradjak, számomra nemcsak nagy öröm lenne, hanem nagyon nagy megnyugvás is, ha például egy tüzetes nyelvtan számára a képzők gyakoriságát, termékenységét, jelentésárnyalatait olyan nagy mennyiségű konkrét statisztikai adat alapján állapíthatnám meg, amennyinek a gyűjtésére „A mai magyar nyelv rendszere" szóképzésünket tárgyaló fejezetének a megírásakor a hagyományos adatgyűjtési módszerrel nem vállalkozhattam.[6]

De igen gyümölcsözőnek tartanám nagy íróink, költőink nyelvének lyukkártya módszerű feldolgozását stilisztikai szempontból is. Az a hatalmas anyagmennyiség ugyanis, amely rendelkezésünkre állna, módot nyújtana rá, hogy ne csak érdekes jelenségekre figyeljünk fel, hanem egzakt módon megvizsgáljuk koronként és egyénenként mondjuk a költők jelzőhasználatát, sajátos szórendjét, mondatszerkesztését, hogy konkrétabban elkülönítsük a különféle stílusfajtákat, a korstílusokat, az egyéni stílusokat stb.

J. Soltész Katalin fölvetette a gyakoriság-szótár gondolatát.[7] Ha ez a gondolat megérlelődne, úgy hiszem, azon is érdemes lenne gondolkozni, vajon nem lehetne-e elkészítésébe nekünk is gépi segítséget bevonni. De egyéb soron következő szótári munkákkal kapcsolatban is föltétlenül

[6] A mai magyar nyelv rendszere. Leíró nyelvtan (Bp., 1961) „A szóképzés" című fejezet: I, 333—420.

[7] Lásd J. Soltész Katalin, Guiraud statisztikai módszere a szókincs vizsgálatában: Általános nyelvészeti tanulmányok 1. 263—272 (1963). Vö. még a konferencia határozati javaslataival: A Magyar Tudományos Akadémia Nyelv- és Irodalomtudományi Osztályának Közleményei 19. 335—39 (1962).

számba veendő dolognak vélem a gépi szortírozás gyorsaságát. Ezzel kapcsolatban legyen szabad megemlítenem: Tübingenben azon a konferencián, amelyet az irodalmi elemzésnek és a szótárírásnak gépi módszereiről több mint ötven szakember részvételével tartottak, egyidejűleg három táblázó gépet is működtettek. Ezek a címszó után percenként 150 soros sebességgel jegyezték az előforduló ragozott formákat három különböző műből, pontos forrásjelzéssel. Végül összeszámolták az adatokat: az egyik Goethe kötetből például 54 979 szóalak került ki: 5763 címszó 8318 különféle ragozott formában.[8]

Természetesen az elektromechanikus adatfeldolgozó gépek nyelvészeti felhasználásának lehetőségét nemcsak a módszer, illetőleg a módszernek a nyelvtudományban való alkalmazhatósága felől lehet, kell taglalni; meg kell vizsgálni a kérdést az adott pénzügyi, műszaki feltételek szempontjából is. Erre azonban én nem vállalkozhatom, annál kevésbé, mivel jóformán csak a figyelmet kívántam felhívni a lyukkártyás módszerre.

Befejezésül annyit, hogy a Német Tudományos Akadémián még nem záródtak le azok a viták, amelyek az általam ismertetett kérdések körül kialakultak; az adatfeldolgozó gépnek s a lyukkártyáknak vannak már föltétlen hívei, de még ellenzői is. S azt, hogy a gép gép, tehát hogy a gép nem tud gondolkozni, csak azt tudja, azt tudhatja, amit beleépítettek, még azok is hangoztatják, akik alkalmazása előnyeiről már meg vannak győződve.

[8] Lásd: Spectrum (Mitteilungsblatt für die Mitarbeiter der deutschen Akademie der Wissenschaften zu Berlin) 7. 81—4 (1961).

Kisebb közlemények*

Körner Györgyné

A Magyar Tudományos Akadémia Nyelvtudományi Intézete a gépi fordítás előkészítésével kapcsolatban azt a kívánságát fejezte ki, hogy munkájának előkészítésére az Országos Műszaki Könyvtár keretén belül működő Országos Fordításnyilvántartás egy meghatározott időszakról adjon értékelést a birtokában levő adatokról.

Tájékoztatásul meg kell említenem, hogy az Országos Fordításnyilvántartás országos feladatot lát el. 1949 óta ez a szerv nyilvántart minden fordítást, legyen az műszaki, közgazdasági, mezőgazdasági, orvostudományi vagy egyéb tárgyú szakfordítás. A Fordításnyilvántartás munkáját népgazdasági rendelet hozta létre, és azóta minisztertanácsi rendelet szabályozza működését.

A fordításnyilvántartás közel tizenhárom évi működése alatt katalógusok egész sorát építette ki. Így van katalógusa: szerzők szerint az önálló művekről, folyóirat szerint a folyóirat-cikkekről, gépek használati utasításáról, konferenciákról és kongresszusokról, szabványokról és szabadalmakról.

Csupán a folyóirat-katalógusból merített adatokat dolgoztam föl, mint a legjobban koncentrált anyagot. Ebből is az 1959. és 1960. évi műszaki folyóiratokból készült fordításokat. Az 1961. évet még nem lehet feldolgozni, mivel az 1961-es folyóiratokból a fordítások még folyamatosan várhatók, a katalógus állománya különösen ebből az évfolyamból naponta bővül, ahogyan az újabb bejelentések érkeznek. Az 1959-es 1960-as éveket már nagyjából lezártnak tekinthetjük, és ebből viszonylag a legfrissebb és megközelítően teljes anyagot vázolhatjuk.

A műszaki folyóiratokból fordított cikkek feldolgozásánál 12 fő szakot vettem alapul. Ezek a következők:

(1) alaptudományok (ideértve matematika, fizika, csillagászat);
(2) elektrotechnika és híradástechnika;
(3) közlekedés;
(4) gépipar (műszerek, berendezések, szerszámok stb.);
(5) élelmiszeripar;
(6) kohászat;
(7) bányászat, energia (geológia, geodézia);
(8) építőipar, építészet;

* Az itt következő „kisebb közlemények" (eredetileg hozzászólások) főként a kötet második felének tanulmányaihoz (előadásaihoz), illetőleg az azokban fölvetett gondolatokhoz kapcsolódnak. — *A szerkesztők.*

(9) üzemgazdaság (automatizálás, balesetvédelem);
(10) könnyűipar (bőr, textil, ruházat, cipő, fa, papír);
(11) vegyiipar, kémia;
(12) könyvtárügy (illetőleg ennek műszaki vonatkozásai).

A kutatás során tizenöt nyelvből találtunk a fenti körön belül fordításokat. Itt jegyzem meg, hogy számszerűen összevontam az angol nyelvben az amerikai, a brit és az ausztráliai anyagot, ugyanígy a német nyelvben a Német Demokratikus Köztársaság, a Német Szövetségi Köztársaság, Ausztria és Svájc német anyagát.

A Fordításnyilvántartás 1958-ban összesen 19 500, 1959-ben 24 100, 1960-ban 28 500, 1961-ben pedig — amit csak a fejlődés dokumentálására jegyzek meg — 31 100 db fordítás-bejelentést dolgozott fel. Ez azt jelenti, hogy az 1958-as évet alapul véve 1959-re 122 százalékra, 1960-ban 145 százalékra, 1961-ben pedig 159 százalék fölé emelkedett a fordítás-bejelentések száma. Az 1959. évi 24 100 db fordítás-bejelentésből 14 150 volt a műszaki folyóirat fordításainak darabszáma, vagyis 58 százalék.

Az 1959. évi 14 150 db műszaki fordításból a nyelvi megoszlás a következőképpen alakult:

német nyelvből	4540 db, 32,1%	cseh és szlovák nyelvből	874 db, 6,2%
angol nyelvből	4029 db, 28 %	francia nyelvből	861 db, 6,1%
orosz nyelvből	2885 db, 20,4%	lengyel nyelvből	468 db, 3,3%

Az említett 12 fő szakot vizsgálva, a 14 150 db megoszlása a következő:

(1) kohászat.....	2980 db, 15 %	(4) könnyűipar	1826 db, 12 %
(2) gépipar	1922 db, 13,5%	(5) építőipar	1626 db, 11,5%
(3) vegyiipar	1846 db, 13 %	(6) elektrotechnika ..	1626 db, 11,5%

A többi szak 9 százalék alatt van.

A szakmákon belül a nyelvek sorrendje a következőképpen alakul:

kohászat: angol, orosz, német;
gépipar: angol, német, orosz;
vegyiipar: angol, német, orosz;
könnyűipar: német, angol, orosz;
építőipar: német, orosz, angol;
elektrotechnika: német, angol, orosz.

Az 1960. évi 28 500 számú összbejelentésből 14 614 db volt a műszaki folyóirat-cikkek fordításának darabszáma, ami 51,5 százaléknak felel meg. Ezek nyelvi megoszlása a következő:

német nyelvből	4926 db, 34,5%	francia nyelvből	844 db, 5,7%
angol nyelvből	3948 db, 27,1%	cseh és szlovák nyelvből	821 db, 5,6%
orosz nyelvből	2994 db, 20,6%	lengyel nyelvből	626 db, 4,3%

A szakmai megoszlás 1960-ban így alakul:

(1) gépipar ..... 2529 db, 17,3%
(2) kohászat ..... 2110 db, 14,4%
(3) vegyiipar ..... 1839 db, 12,6%
(4) könnyűipar ..... 1741 db, 11,9%
(5) építőipar ..... 1586 db, 10,8%
A többi szak 10 százalék alatt van.

A szakon belül a nyelvek sorrendje a következő képet mutatja:

gépipar: angol, német, orosz;
kohászat: angol, orosz, német;
vegyiipar: angol, német, orosz;
könnyűipar: német, angol, orosz;
építőipar: német, orosz, angol.

A két év adatait összevetve azonos eredményt látunk a következőkben. A nyelvek sorrendje változatlanul: német, angol, orosz. Ez a három nyelv képviseli a fordításoknak több mint 80 százalékát. A fennmaradó tizenkét nyelv együttesen és mindkét évben 20 százalék alatt van.

A szakok szerinti felosztásban — mint láttuk — már nincsenek olyan szakok, nyelvek, amelyek dominálnak; azonban mindkét évben azt tapasztalhatjuk, hogy azoknál a szakoknál, ahol a fordítások száma a tizenkét szakon belül eléri az összes fordítások 10 százalékát, a nyelvek sorrendjében mindig az angol, a német és az orosz foglalja el az első három helyet.

A megadott pontos számadatok mellett figyelembe kell venni még azt is, hogy a műszaki anyaghoz tartoznak még a gépek használati utasításai, a műszaki könyvek, a műszaki kongresszusok és konferenciák, valamint a szabványok és a szabadalmak fordításai is. Ezek nincsenek benne a fenti részletezésben, és ezért csupán tájékoztatásul közlöm róluk a következő adatokat. Az 1959. évre megadott 58 százalék műszaki anyag mennyisége az össz-bejelentések számához viszonyítva 68—70 százalékra, az 1960. évi 51,5 százalék pedig 65—67 százalékra becsülhető.

*Péter Rózsa**

Én csak Kalmár László szövegének Szalai Sándor bemutatta kivonatáról szeretnék néhány szót mondani, mégpedig arra hívni fel a figyelmet, hogy a matematika alapjaira vonatkozó szövegekben nem azok a triviális szavak, mint más szövegekben. Például az *és* itt nem triviális szó, hiszen a konjunkció műveletének kifejezése, és ez okozza sűrű előfordulását. De nemcsak az *és* ilyen, hanem például Bertrand Russell egy könyvének egy fejezetcíme[1] is: „*Az, amely*". Az *az, amely* ezen a területen lényeges szerepet játszik; az *az* és az *amely* sűrű előfordulását itt ez indokolja. Azonkívül, ha axiómákat kielégítő modellekről beszélünk, azokat sohasem sak egy, hanem több modell elégíti ki, tehát itt nem az *amely*, hanem az *olyan* az, mi lényeges szerepben sokszor előfordul.

* Péter Rózsa észrevételeit Szalai Sándor figyelembe vette tanulmánya végleges szövegének elkészítésében; lásd: Kötetünkben 308—309 és másutt is. — *A szerkesztők.*

[1] Bertrand Russell: Introduction to mathematical philosophy (London és New York, 1919).

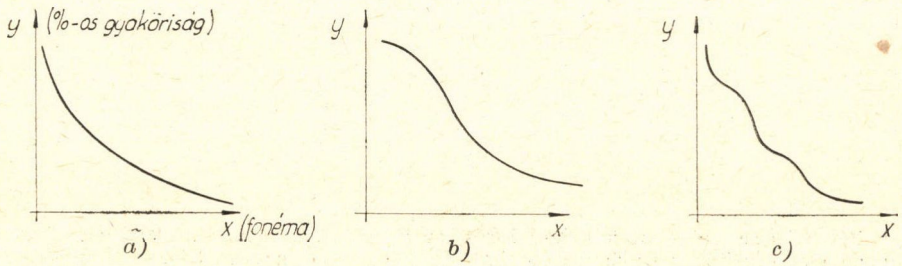
*Hajtman Béla**

1. Egy kialakulatlan, alig megfogalmazott ötletet szeretnék felvetni. Nemcsak azt nem tudom, kidolgoztak-e már hasonló eljárásokat,[1] de azt sem, hogy az egész módszer egyáltalán használható lesz-e a nyelvészek számára. Mindenesetre nekem, mint matematikusnak úgy tűnik, hogy ezen a területen támogatni tudnánk a nyelvészek kutatásait.

Évekkel ezelőtt egy nyelvész barátom, Róna Tas András fordult hozzám néhány kérdéssel bizonyos görbék matematikai jellemzésével kapcsolatosan. Ezeket a görbéket ő nyelvstatisztikai adatokból nyerte (a monguor nyelv tibeti jövevényszavainak vizsgálatánál, többek között, a különböző szókezdő fonémák előfordulásának gyakoriságát ábrázolta) és előadásában az így kapott görbét matematikailag is jellemezni akarta.

Ezzel kapcsolatban vetődött föl bennem a gondolat, hogy hasonló görbét más, a nyelvre jellemző adatok alapján is fel lehet venni, és a görbe matematikai elemzéséből esetleg nyelvészeti köevtkeztetéseket is lehet levonni, vagy más úton nyert ilyen következtetéseket alá lehet támasztani.

2. Nézzünk egy példát. Ábrázolni akarjuk, tegyük fel, egy nyelv összes fonémáinak előfordulási gyakoriságát. Az „x-tengelyen" egyenletes közökben felvesszük a fonémákat, mégpedig olyan sorrendben, hogy a leggyakrabban előforduló legyen az első, a második leggyakoribb a második, és így tovább; az „y-tengelyre" pedig felmérjük az előfordulások gyakoriságát, százalékban. Az így nyert pontokat azután (először csak durván, szemre) összekötjük. Így egy csökkenő görbét kapunk, valami ezekhez hasonlót:



* Az elhangzott hozzászólásnak az írásbeli közlés számára kikerekített változata.
[1] Mint azóta értesültem, B. Mandelbrot foglalkozott hasonló problémával.

Vegyük először az *a)* típust, mint legvalószínűbbet. (Az említett esetben is, a szókezdő fonémák vizsgálata ilyen görbéhez vezetett.) Kétféle feltevés kézenfekvő ilyen görbetípus esetén: az, hogy a görbe hiperbolikus, vagy az, hogy exponenciális jellegű. Az első esetben $y = \frac{a}{x+b}$ alakú, a másodikban $y = a \cdot e^{-bx}$ alakú egyenlettel igyekszünk megközelíteni. Ha semmi különös okunk nincs eleve egyik vagy másik hipotézis mellett dönteni, azt fogadjuk el, amelyik a pontokra jobban illeszkedő görbét szolgáltat. A feladat (ha a görbe alakjára vonatkozóan elfogadtuk valamelyik hipotézist) az a és b paraméterek meghatározására korlátozódik. Ezeket természetesen úgy igyekszünk meghatározni, hogy a görbének a pontoktól való eltérése[2] minimális legyen. Ezt az egész eljárást nevezzük a matematikában *kiegyenlítésnek*.

Valamivel nehezebb feladat a pontok kiegyenlítése, ha azok olyanféleképpen helyezkednek el, ahogyan a *b)* és *c)* ábra görbéi mutatják. Az ilyen görbék parabolikus jellegűek; de hogy hányadfokú legyen a parabola, azt igen nehéz megválasztani. Ha túlságosan magasfokú, akkor pontosabb ugyan a közelítés ($n + 1$ pont esetén van olyan n-ed fokú parabola, amelyik valamennyi ponton átmegy), de nehezen kezelhető — és nagyon sok a paraméter is (eggyel több, mint a parabola fokszáma; az egyenlet alakja ugyanis $y = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n$). Előfordulhat, hogy nem törekszünk a görbe teljes jellemzésére, csak néhány dolgot állapítunk meg (például hogy van-e inflexiós pont[3] és hány; a görbe konvex vagy konkáv szakaszszal kezdődik-e; mekkora a konvex szakaszok összhosszúsága, illetőleg ennek aránya a konkávok összhosszához; stb.) attól függően, hogy nyelvészeti szempontból mire akarjuk az egészet használni, milyen más görbével vagy görbékkel hasonlítjuk össze.

Természetesen nemcsak a fonémák előfordulásának gyakoriságát lehet így ábrázolni, hanem más előfordulási gyakoriságot (esetleg egyéb, nyelvre — vagy egy szerző stílusára — jellemző adatokat) is. Így például az összes morfémák előfordulását, a szóhosszúságokat[4] stb. Matematikai szempontból annál jobban kezelhető a probléma, annál biztosabb eredményre vezet a kiegyenlítés, minél több pontunk van. Ezért jobb nézni például a morfémák gyakoriságát a fonémákénál. De még több pontot nyerünk, ha például az összes lehetséges fonéma-párok vagy három (esetleg több) elemű betűcsoportok előfordulásának gyakoriságáról készítünk statisztikát és ezt ábrázoljuk.

3. Végül szeretnék néhány szót szólni arról, hogy mire lehetne ezt az egész módszert felhasználni. Nyelvtörténeti és összehasonlító nyelvészeti alkalmazás látszik a legvalószínűbbnek. Gondolok itt például arra, hogy nyelvek kölcsönhatásának vizsgálatakor egy — teszem azt — „hiperbolikus" eredetű nyelvre erős hatást gyakorol valamelyik más, mondjuk „exponenciális" vagy „parabolikus" nyelv. A hatásnak megfelelően a görbe is (feltehetően) egyre inkább eltér az eredeti hiperbolikus formától. Esetleg egy ismeretlen hovatartozású nyelvről meg lehet állapítani (legalábbis kizárásos alapon, negatív döntésekkel), hogy melyeknek lehet vagy nem lehet rokona. Ha például az összes indoeurópai nyelvek fonémagyakorisága hiperbolikus jelleget

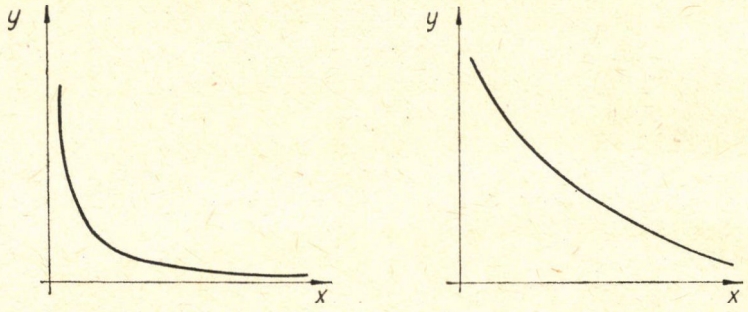
[2] Ezt nézhetjük többféleképpen is (például az abszolút eltérések összegével), de a legáltalánosabban használt mérőszám a négyzetes eltérések átlaga.

[3] Olyan pont, ahol a görbe konkávból konvexbe megy át, vagy megfordítva.

[4] Szó- vagy mondathosszúságok vizsgálatánál a csökkenő görbével való ábrázolás helyett jobbnak látszik a szokásos eloszlási görbe felvétele. Ilyen görbék összehasonlítására a matematikai statisztikának jól kidolgozott módszerei vannak.

mutat, másfajta görbéjű nyelv rokonait más nyelvcsaládban kell keresnünk. Természetesen, az egyes jellemzők (fonéma, morféma, hangcsoport, stb.) megválasztása szerint más és más lehet egy nyelv „görbe-jellege". Éppen a feladat fogja diktálni azt, hogy melyik jellemzőt kell választanunk; előfordulhat, hogy a minél nagyobb differenciáltság a célunk, de az is lehet, hogy minél nagyobb „családokat" akarunk rokonvonásokkal jellemezve összetartozónak tekinteni.

A nagyobb differenciáltságot gyakran úgy tudjuk elérni, hogy figyelembe vesszük a paraméterek értékét is. Egy görbetípuson belül ez (sokszor lényeges!) különbséget jelenthet. Várható, hogy a történeti fejlődést is a paraméterek tükrözik majd. Elképzelhető (és tisztán gondolati úton, a kívülálló — és nem a nyelvész szakember! — logikájával meg is indokolható), hogy (az *a)* típusú görbét véve alapul) a „primitívebb" nyelvek görbéje „meredekebb" lesz, a fejlettebbeké „laposabb". Valahogy így:



Hiperbolikus görbe esetén ez a jelenség az *a* paraméter nagyságában tükröződik. Eszerint tehát a nyelv fejlettségi fokát (ugyanazon a típuson belül, természetesen) egyetlen számmal lehetne jellemezni.

Persze, mint mondtam, ez mind csak ötlet, és nyelvészetileg a legteljesebb mértékben megalapozatlan.

Dömölki Bálint

A Hell György előadásában ismertetett orosz—magyar gépi fordítással kapcsolatban szeretnék néhány szót szólni az első, számológépen valóban végrehajtott kísérletről. Itt a feladat az volt, hogy a beadott orosz szöveg minden egyes szaváról megállapítsuk: a szó utolsó néhány betűje megegyezik-e az orosz nyelvben használatos végződések valamelyikével. Megjegyzendő, hogy a gép jelenlegi állapotában külön szótár használata nélkül nem tudja megállapítani azt, hogy például a *предмет* szóban az *-ет* végződés-e, vagy pedig a szótő része; egyelőre tehát kizárólag formai megállapításokra vagyunk utalva. Gyakorlatban a gépi fordítás szempontjából ez nem okoz majd problémát, mert ott úgyis szükségünk lesz majd szótárra; és az elképzelésünk az, hogy a végződés leválasztásának művelete előtt a szavakat a szótáron végigfuttatva a végződés nélküli szavakat már ki lehet szűrni.

A Hell György által összeállított orosz végződés-típusok száma körülbelül negyven volt; egy típusnak vettük a grammatikailag azonos végződéseket, például az *-ому* és az *-ему* szuffixumokat. Az összes megkülönböztethető végződések száma hetven körül járhat; tehát elég nagy szám ahhoz, hogy a sorozatos összehasonlítások módszerével meglehetősen hosszadalmas algoritmust kapjunk. Ugyancsak elég hosszadalmas — bár az említettnél rövidebb — Melcsuk algoritmusa. Ő a szó végét a végződések szótárával egybefuttatva megnézi, hol kapunk — lehetőleg teljes — egyezést. Itt is körülbelül harmincöt összehasonlításra van szükség (hetven különböző végződéstípusnál az átlagosat véve), ami elég tekintélyes műveletszám.

A Magyar Tudományos Akadémia Számítástechnikai Központjában kidolgoztunk egy olyan módszert, melynek segítségével egy-egy művelettel több szempont vizsgálatát lehet egyszerre elvégezni.[1] Ezzel a módszerrel működött a Hell-féle fordítási algoritmushoz kapcsolódó kísérleti végződés-felismerő program is. Úgy látszik, ez a program gyorsabb, mint az említett módszerek. A módszer részleteiről itt nem kívánok részletesebben beszélni, mert ez túlságosan programozási-szakmai jellegű. Megemlítem azonban, hogy kiderült: ez a módszer egy általánosabban megfogalmazható feladat megoldására is alkalmas. A további gépi fordítási és egyéb algoritmusoknál esetleg a feladatok megfogalmazásánál ezt figyelembe lehet venni, és a problémákat majd ennek megfelelően lehet megfogalmazni.

[1] Vö. Dömölki Bálint, A jelsorozatok tulajdonságainak felismerésére szolgáló algoritmusokról: A Magyar Tudományos Akadémia Számítástechnikai Központjának Tájékoztatója, 8. szám (Bp., 1962).

Nézzük meg ezért, hogy mi volt a feladat a végződések leválasztásában.

Ha például az orosz *-ание, -яние, -ение* végződéseket akarjuk felismerni, amelyek grammatikailag azonosan kezelendők, akkor a feladat a következő „táblázattal" adható meg:

$$\begin{bmatrix} \text{a} \\ \text{яние} \\ \text{e} \end{bmatrix}$$

A feladat azoknak az orosz szavaknak a kiválasztása, melyeknek az utolsó betűje megtalálható az utolsó oszlopban, az utolsó előtti betűje pedig az utolsó előtti oszlopban, stb.

Ha most betűk helyett általánosabb értelemben vett jeleket (például szófajok, mondatrészek kódjait, írásjeleket stb.) veszünk, illetőleg szavak helyett ilyen jeleknek tetszőleges sorozatait, akkor általánosabb struktúrákat (például grammatikai szerkezeteket) lehet felismerni ennek az algoritmusnak a segítségével, mégpedig az eddig használatos módszerekhez képest általában nagyobb sebességgel.

Dezső László

Az orosz tudományos szövegek fordításának egyik lényeges problémájával, a *-ся* képzős igékkel képzett szerkezetek magyarra fordításával foglalkozom röviden, s ennek kapcsán kitérek a gépi fordítási vizsgálatok egy-két, a nyelvészeti kutatás szempontjából lényeges kérdésére.

1. A *-ся* képzős igékkel alkotott szerkezetek igen gyakoriak az orosz tudományos és műszaki szövegekben, s mivel a magyarban adekvát igekategória-megfelelésük nincsen, gépi fordításuk nehézségekbe ütközik. A gépi fordító csoport első céljának a matematikai szövegek fordítását tűzte ki, így e probléma megoldásához is matematikai szakszövegből gyűjtöttem anyagot. Két művet[1] vizsgáltam meg rétegzett mintavétellel. Megvizsgáltam mintegy 130 igével alakított 750 orosz szerkezetet s magyar fordítását, és erre alapozom mondanivalómat.

A *-ся* igés szerkezetek gépi fordításához szükség van a matematikai szövegekben előforduló minden egyes igének — tartozzon az a közös irodalmi nyelvi, az általános tudományos vagy a speciálisan matematikai szókincsbe — adekvát magyar lexikai megfelelőjére és a fordításával kapcsolatos nyelvtani és szórendi ismeretekre. Ezenkívül összefoglaló képpel kell rendelkeznünk a *-ся* képzős igékkel alkotott szerkezetek fordításakor előforduló nyelvtani és szórendi változásokról, mert csak ezek birtokában tudjuk megválasztani a legmegfelelőbb fordítási eljárásokat. Így tehát meg kell szereznünk az egyes orosz és magyar igékre vonatkozó úgynevezett mikroszintaktikai ismereteket, és azokat a mennyiségi vonatkozások figyelembevételével összegeznünk kell. Az igés szerkezetek összevető elemzésére a formális elemzés nyújtja a legjobb lehetőséget. Bár itt mondat értékű szerkezetekről van szó, a mondatrészek szerinti elemzés fölösleges, sőt a két nyelv eltérő elemzési szempontjai miatt erősen zavaró is lenne. A szerkezetek leírásához a transzformációs nyelvtanból ismert módokat használjuk némi változtatással. Itt meg kell jegyeznem, hogy a transzformációt igen régóta alkalmazzák a nyelvészek, szerepe azonban szükségszerűen

[1] P. Sz. Alekszandrov: Введение в общую теорию множеств и функций (Moszkva—Leningrád, 1948); magyar fordítása: Bevezetés a halmazok és függvények általános elméletébe (fordította: Bizám György, Bp., 1952). — V. V. Sztepanov: Курс дифференциальных уравнений (Moszkva—Leningrád, 1950); magyar fordítása: A differenciálegyenletek tankönyve (fordította: Földes István, Körmendi István és Lovass-Nagy Viktor, Bp., 1955).

megnő a strukturális elemzésben. A transzformáció alkalmazása nem jelenti azt, hogy N. Chomsky transzformációs koncepciójának[2] részleteihez kell minden tekintetben tartani magunkat.

Az orosz szerkezetek két fő típusra oszthatók. Az első $C^1_u\ \Gamma_{ся}$, tehát az egyes jelzésű főnév alanyesetben, a *-ся* képzős ige általában harmadik személyben áll, például *внутренность . . . выбрасывается*. A második: $C^1_u\ \Gamma_{ся}\ [C^2_K]$, tehát az egyes jelzésű főnév alanyesetben, a *-ся* képzős ige általában harmadik személyben, a kettes jelzésű főnév, amely vonzata az igének, valamely indirekt esetben van, például: *последовательность . . . называется ε-цепю*. Ennek a két fő típusnak vannak szórendi variánsai: $\Gamma_{ся}\ C^1_u$: *берется треугольник*, tehát elöl áll a *-ся* képzős ige, s utána következik a főnév, és $[C^2_K]\Gamma_{ся}\ C^1_u$; például *примерами исчерпывается . . . разнообразие*, ahol első helyen áll a kettes jelzésű főnév, utána jön a *-ся* képzős ige, míg az egyes jelzésű főnév az utolsó helyre kerül. Ez utóbbi típusnak természetesen más, kevésbé lényeges szórendi variánsa is van. Ezen két fő típuson kívül vannak egyebek is, például, ahol az igéhez egy másik infinitivusban álló ige, *что* kötőszavas mellékmondat járul stb. Az egyenes szórendű $C^1_u\ \Gamma_{ся}\ [C^2_K]$ szerkezet gyakoribb, és általában az egyenes szórend az uralkodó. Megjegyezzük, hogy a szerkezet egyes tagjait gyakran választják el szavak, amelyek valamilyen mondattani kapcsolatban vannak velük, vagy közbevetetett szavak és szerkezetek, például a *parciális differenciálegyenlet* orosz meghatározása így hangzik: *Уравнение, связывающее искомую функцию, независимые переменные, и частные производные от искомой функции, называется уравнением с частными производными*, ahol az egyes jelzésű főnév és az ige között 11 szó van.

A $[C^2_K]$-tagos, tehát a kettes jelzésű főnevet tartalmazó szerkezeteket, tovább lehet csoportosítani transzformációs lehetőségeik szerint, amire itt nem térek ki.

Inkább a magyar fordítással kapcsolatos néhány kérdésről szólok. Az orosz *-ся* képzős igéknek gyakran nincs meg az ekvivalens magyar párjuk, ezért a fordításukra felhasználandó magyar ige más alakban áll, mint orosz megfelelője. Az igekategóriák megváltoztatása magával hozhatja a velük kapcsolatos főnevek alakváltozását is. Így például ez a tipikus szerkezet: *последовательность . . . называется ε-цепю*, nem lehet olyan magyar szerkezettel fordítani, amelynek tagjai az orosszal egyező grammatikai kategóriákban állnának, tehát: $C^1_u\Gamma_{ся}[C^2{}_T] \rightarrow$ $\rightarrow {}^*F^1_N\ I\ [F^2{}_{val}]$, azaz az orosz egyes jelzésű főnév nominativus, a *-ся* képzős ige egyes szám 3. személyű, a kettes jelzésű főnév instrumentalis szerkezetet nem lehet magyar egyes jelzésű főnév alanyeset, ige egyes szám 3. személy, *-val*-ragos kettes jelzésű főnév szerkezettel fordítani, mert a helyes fordítás: . . . *sorozatot ε-láncnak nevezzük*, tehát $F^1_A\ [F^2{}_{nak}]I_{\text{pl. 1.}}$ azaz az egyes jelzésű főnév accusativusban, az ige többes 1. személyben, kettes jelzésű főnév pedig *-nak* raggal áll. A fordítással kapcsolatos ilyen alakváltoztatást interlinguális transzformációnak nevezhetnénk. A nyelvtani transzformációkon kívül igen gyakran a szórend megváltoztatására is szükség van, mint a fenti példában is látható. A gépi fordítás során arra kell törekednünk, hogy az interlingualis transzformációk száma minimális legyen, s ha mégis elkerülhetetlenek, olyanokat alkalmazzunk, amelyek géppel könnyebben megoldhatók. Természetesen ez csak annyira valósítható meg, amennyire ezt a nyelvhelyesség követelménye megengedi.

[2] Vö. N. Chomsky: Syntactic Structures (Hága, 1957).

A vizsgált igék negyed részében nem kellett interlingualis transzformációt végrehajtani, s nem volt szükség szórendi változtatásra sem. Az igék háromnegyed részében azonban vagy az ige, vagy a főnév alakjában, illetőleg a szórendben változtatást kellett eszközölni.

Az ige változásai kétfélék lehetnek. Az egyik esetben — ilyen volt a fenti — az orosz egyes szám 3. személy helyett a magyarban többes szám 1. személy áll: *называется — nevezzük*. Ez magával hozza azt, hogy az orosz alanyesetű főnév helyett a magyarban tárgyesetű áll $(C^1_u \rightarrow F^2_A)$, ami a fordításban újabb nehézséget okoz. A vizsgált igék húsz százalékában került sor ilyen változtatásra. Az ige személyének megváltoztatása elkerülhető akkor, ha ható igét alkalmazunk. A ható ige egyes szám 1. személyű alakja vagy még gyakrabban jelen idejű melléknévi igenevének nominativusi alakja alkalmas arra, hogy vele a medialis és passzív orosz *-ся* képzős igék egyes szám 3. személyű alakjait fordítsuk. Ebben az esetben az egyes jelzésű főnevet nem kell transzformálni. Most nem foglalkozom azzal a kérdéssel, mikor használhatunk ható igét, és mikor nem. Röviden csak arra utalok, hogy a ható igét nem alkalmazhatjuk akkor, ha konkrét cselekvésről van szó, ezzel szemben használható akkor, ha általános megállapítást teszünk, ami tudományos szövegekben elég gyakori, mert az igék harminc százalékában megengedhető a ható igés fordítás.

Vegyünk egy példát a szóban forgó jelenség illusztrálására. Az első oroszról magyarra fordított mondat némileg megváltoztatott alakját vesszük: *машина решает задачу*. Ennek passzív szerkezetű transzformját fogjuk elemezni: *задача решается машиной*, képlete $C^1_u \Gamma_{ся}\ C^2_T$, tehát az egyes jelzésű főnév nominativusi, a *-ся* képzős ige egyes szám 3. személyű, a kettes jelzésű főnév instrumentalisi alakkal bír. Egyszerűség kedvéért vegyünk egy olyan fordítást, ahol a szórend az orosszal megegyezik: *A feladatot megoldja a gép*, képlete $F^1_A\ I\ F^2_N$; tehát az egyes jelzésű főnév accusativusban, az ige ugyanabban a számban és személyben, a második főnév nominativusban áll. Ha ez a szerkezet olyan kontextusban áll, amelyben konkrét cselekményről van szó, akkor nem használhatunk ható igét, de bizonyos szövegösszefüggésben az orosz szerkezet fordítható így is: *a feladat megoldható géppel*, képlete: $F^1_N I_{\text{ható}}\ F_{\text{val}}$, tehát csak az igén kell változtatást eszközölni. Ez utóbbi fordítás azonban csak általános megállapítás esetén lehetséges. Felmerül azonban a kérdés: ha a ható ige alkalmazása ilyen, a gépi megoldás szempontjából bonyolult összefüggéstől függ, nem kell-e eltekinteni tőle. A végleges választ nagyszámú példa statisztikai elemzése adja majd meg. Ha ugyanis valamely ige mindig fordítható ható igével, akkor ez a megoldás jelölendő ki számára, ha kétféle megoldás lehetséges, de az ige személyének transzformációja mindig megfelelő, az utóbbinál kell maradni.

2. A szórendről röviden. Mint fentebb említettem, az orosz szerkezet tagjainak szórendje elég szigorú, és néhány pontosan leírható típusra oszlik. A magyar szórenden az eddig vizsgált igékkel képzett szerkezetek csaknem felében változtatást kellett eszközölni. A leggyakrabban előforduló orosz szerkezetek fordítása esetén a magyarban változtatásra van szükség. Anyagunk korlátozott volta miatt nem vállalkozhatunk arra, hogy a magyar szórend kérdését érdemben vizsgáljuk. A magyar szórend a gépi fordítás legnehezebb problémái közé tartozik, s igen jó lenne, ha e téren segítséget kapnánk magyar nyelvészeinktől.

3. A gépi fordítási kutatások során lépésről lépésre kell kidolgoznunk az orosz matematikai szövegek fordításának egyes kérdéseit. De ugyanakkor, amikor le kell szűkí-

teni a megoldandó feladatot, elengedhetetlen, hogy ne végezzünk az adott feladat keretein túlmutató tájékozódó jellegű kutatásokat is. Így szükségesnek tartom, hogy képet alkossunk arról, milyen eltérést okoz eddigi megállapításainkhoz képest a bizonyos mértékig elütő jellegű szövegek, mint például politikai irodalom fordítása. Feltételezhető, hogy a matematikai szövegek fordítása során nyert nyelvtani ismeretek nagyjából és egészéből elégségesek lesznek a politikai szövegek fordításához is, de nem érdektelen erről pontosabb képet alkotni.

Jelenleg nem tűzhetjük ki célul azt, hogy az oroszon kívül más szláv nyelvről történő gépi fordítással is foglalkozzunk. Mindazonáltal igen tanulságos lesz, ha az orosz és magyar nyelv tipológiai szempontból jellemző kérdéseinek vizsgálatakor tájékozódásképpen kiterjed elemzésünk más szláv nyelvű szövegekre is.

Az egyes kérdések vizsgálata során fel kell használnunk azokat az eredményeket, amelyeket más országokban már elértek. Különösen fontosak számunkra az oroszról történő fordítással kapcsolatos kutatások, elsősorban orosz—angol vonatkozásban.

A nemzetközi gépi fordítási munkálatok, véleményem szerint, jelentősen hozzá fognak járulni ahhoz, hogy a t i p o l ó g i a i kutatások vizsgálati szempontjaik, módszereik, valamint a vizsgált nyelvek terén egyaránt gazdagodjanak és elfoglalják helyüket a három másik kutatási irány: a leíró, történeti és összehasonlító nyelvészet mellett.

Kutatásainkban lépésről lépésre előrehaladó, konkrét feladatokat megoldó tervszerű munkát végezve, állandóan szem előtt kell tartanunk az ebből adódó általános e l m é l e t i k é r d é s e k e t. Fel kell használnunk mindazt, amit az úgynevezett „hagyományos" módszerekkel végzett kutatások elértek, mert ez nekünk a nyelvi ismeretek gazdag tárházát jelenti, amit komoly hiba lenne lebecsülni. Ugyanakkor azonban bátran kell élnünk azokkal a módszerekkel, amelyeket a strukturális nyelvészet tárt fel.

Remélem, hogy a gépi fordítási munkálatok ugyanakkor, amikor egy népgazdasági szempontból fontos feladatot oldanak meg, némileg hozzájárulnak majd ahhoz, hogy az elemzett nyelvek struktúráját jobban megismerjük, tipológiai sajátságaikról pontosabb képet alkossunk.

Gábor Mihály

Több, mint félszáz esztendeje tanítok nyelveket: németet, franciát, angolt. E hosszú idő folyamán látnom kellett, hogy például a németül tanulók micsoda keserves kínlódásokon mennek keresztül, és milyen kevés az eredmény, amelyet el lehet érni. Százalékokban talán ki sem lehet fejezni, hányan vannak azok, akik valóban meg is tanulnak helyesen németül, legalább fordítani. Látva ezeket a nehézségeket, keresni kezdtem a módját annak, hogyan lehetne ezen segíteni.

Felmerült bennem az a gondolat, hogy a nyelvtani végződések betűrendbe szedésével valószínűleg meg lehetne szerkeszteni egy olyan módszert, amelynek segítségével az illető tanuló vagy fordító a szövegben előforduló végződések nyomán megtalálja azt a nyelvtani útmutatást, amire szüksége van. Huszonöt esztendeig tartottak kutatásaim és kísérleteim, amíg ezt az egyszerű, retrospektíve, persze, igen egyszerű megoldást megtaláltam. Amíg azonban ez sikerült, addig mérhetetlenül sokat kellett keresnem, kutatnom.

Munkámnak, amikor készen lett, a „Nemzetközi Fordítókulcs" címet adtam.[1] Alapelve a következő: Ragozó nyelveken írott szövegek szavakból és az azokhoz fűzött ragokból, jelekből, képzőkből, gyűjtőnéven végződésekből állanak. Az ilyen nyelveken írott szövegekben tehát a két elemet külön tudjuk választani.

Ha mármost valamely ilyen nyelv nyelvtanának összes ragjait, jeleit és képzőit minden számban, esetben, fokban, alakban, módban, időben és személyben azok három utolsó betűje szerint szoros betűrendbe tesszük s mellékük felsorakoztatjuk fordításukat a kívánt idegen nyelven; olyan „kulcsot" tudunk alkotni, melynek segítségével a fordítani kívánónak csak az lesz a feladata, hogy a fordítandó szövegben talált olyan szavakhoz, amelyeket a szótárban nem talál meg szövegbeli formájukban, hozzáfűzi a szótárban nem szereplő, de a kulcsban fellelhető végződésük jelentését, vagy ezek eltérő szerkezetű fordítását (például rag helyett viszonyszó).

Ezen a módon képesek vagyunk először is minden nyelvtani tudás nélkül is kifogástalan fordításokat készíteni; és másodszor ugyanúgy tudjuk magunkat, fogyatékos nyelvtani tudás mellett, feltétlen biztonsággal ellenőrizni.

Az idegen nyelv szórendjét e módszerben olyan táblázat segítségével alkotjuk meg, mely a mondatrészek egymásután való helyezését a szórend szabályainak isme-

[1] A „Nemzetközi Fordítókulcs" eddig megjelent részei: német—magyar rész (1957); magyar—német rész (1957); angol—magyar rész (1958).

rete nélkül, és így minden gondolkozás nélkül teszi lehetővé, feltéve, ha a mű utasításait pontosan követjük. Nem lehet kétség az iránt, hogy a nyelvtani tudást teljes mértékben pótolhatja a szótövek, ragok, jelek, képzők és mondatrészek — építőkövek — tisztára mechanikus, gépies, mozaikszerű összeállításának lehetősége, megfelelő módszer és utasítások alapján. Hiszen e nyelvtani elemek mozgását és elhelyezkedését jól ismert előírások szabályozzák: a nyelvtan szabályai; a táblázatok minden egyes rovata és sora egy-egy szabály, s az új a módszerben az, hogy e szabályokat azoknak előzetes ismerete nélkül, e sorokra és rovatokra hivatkozó betűk és számok segítségével gépiesen alkalmazni tudjuk.

Munkámmal megismerkedett I. A. Melcsuk, a modern matematikai és alkalmazott nyelvészet ismert szovjet kutatója.[2] I. A. Melcsuk azzal is megtisztelt, hogy művemről külön recenziót írt.[3]

[2] Mivel több szempontból is tanulságos, szükségesnek tartom az eredeti (1958. november 6-i keltű) francia nyelvű levél érdemi részének idézését magyar fordításban. „Munkáját a gépi fordítás szempontjából elemeztem. Éppen ezért lehet, hogy ítéletem túlságosan speciális és szubjektív.

Ön nagyon érdekes munkát végzett, amely elméleti síkon épp oly tanulságos, mint amilyen hasznos gyakorlatilag. Ön különböző nyelvek grammatikájának formalizálását tűzi ki célul maga elé, olyan grammatikákat, amelyek szótárokkal kombinálva lehetővé tennék egy idegen szövegnek a lefordítását az illető nyelv ismerete nélkül.

Ön sehol sem említi a gépi fordítást; úgy látszik, hogy nem ismeri annak részleteit. Ha ez igaz, akkor csak csodálni lehet gondolatainak eredetiségét és mélységét. A hagyományos nyelvészetben nincs meg a hajlandóság erre a formalizálásra, amelyre Ön bizonyító példát nyújt, és ami — szerintem — a nyelv egész tudományos kutatásának a központi kérdése. Ön vállalkozott a grammatikának erre a formalizálására, s ennek magára kell vonnia a nyelvészek figyelmét is.

Ami munkálatainak gyakorlati oldalát illeti, úgy gondolom, hogy az Ön fordítókulcsai alapján grammatikai függelékeket lehetne készíteni a kétnyelvű és talán a többnyelvű, főleg műszaki szótárak számára, amelyek lehetővé tennék a nem nyelvész olvasók (tehát mérnökök, tudósok, technikusok stb.) számára azt, hogy a nyelv ismerete nélkül fordítsanak le egy szakszöveget."

[3] A felhasználás módjáról a következő olvasható a recenzióban: „A fordítókulcs alapul szolgálhat az elektronikus fordítógépbe betáplálandó nyelvtani anyag megszerkesztéséhez. A kulcs elsősorban tudományos, műszaki szövegek olvasásához kíván segítséget nyújtani ismeretlen vagy kevéssé ismert nyelveken. Ilyen segédeszközre valóban szükségük van a legkülönbözőbb tudományos és műszaki területeken dolgozó szakembereknek, akik több nyelven kívánják figyelemmel kísérni a szakmai irodalmat, de akik nem ismerik kellő fokon az illető nyelveket. Feltételezhető, hogy a fordítókulcs ezek számára komoly segítséget nyújt. Ezenkívül alkalmazható a kulcs mint oktatási segédeszköz. A nyelv rendszeres oktatása mellett a kulcs segítségével végzett mechanikus fordítási gyakorlatok után következő tartalmi magyarázatok fejleszthetik a tanulók logikai elemző képességeit s hozzásegíthetik őket ahhoz, hogy elsajátítsák a nyelv formális szerkezetét . . . Ha valamely szakembernek nincsen alkalma ahhoz, hogy rendszeresen foglalkozzék a nyelvvel, de mégis akar idegen nyelvű szakirodalmat olvasni, a kulcs segítségére lehet a szükséges szövegek lefordításában, sok esetben még tanácsadóra sem lesz szüksége és ugyanakkor lehetővé teszi, hogy fokozatosan önállóan is elsajátítsa a nyelvet, minthogy a kulcsot rendszeresen használó olvasó önkéntelenül is megjegyez magának nyelvtani szabályokat és egyes nyelvtani szerkezeteket."

I. A. Melcsuk: Машинный перевод и прикладная лингвистика (Moszkva, 1959), 2 (9) füzet 59—69. [Az orosz szöveg fordítását Suara Róbertnek köszönöm.] — Itt jegyzem még meg, hogy munkámról a modern nyelvtudomány egy másik ismert szovjet képviselője is nyilatkozott már. Vö. Sz. K. Saumjan, Лингвистические проблемы кибернетики и структурная лингвистика: Вопросы Философии: 14/9. 120—31 (1960).

Kelemen József

1. A nyelvvel kapcsolatos halmazok matematikai értelmezésének kérdése

Hallottunk itt arról, hogy a nyelvvel kapcsolatos halmazok matematikailag vizsgálhatók, de jellegüknek megállapításához, tisztázásához feltétlenül nyelvészeti szempontokra van szükség. Így nyelvészetileg kell meghatározni azt is, melyek a zárt és melyek a nyitott halmazok a nyelvben, s a zártságot és a nyitottságot hogyan kell értenünk.

1.1. A fonémarendszer tagjaiból alkotott halmazt zártnak, a morfémarendszer tagjaiból alkotott halmazt nyitottnak tarthatjuk abban az értelemben, hogy a vizsgált nyelv fonémái — meghatározott korban — számszerűleg meghatározhatók, morfémái számszerűleg nem határozhatók meg.

1.2. A számszerűleg meghatározható, tehát zárt halmazt alkotó fonémákkal szemben, a fonémák realizációjának, a beszédhangok fiziológiai-fizikai változatainak menynyisége számszerűleg nem határozható meg, mert számuk aszerint nagyobbodik vagy kisebbedik, ahogy a finomabb árnyalatú hangszín- és időtartam-különbségeket különféle fokig tekintetbe vesszük vagy többé-kevésbé elhanyagoljuk. Ha ezeket a beszédhang-változatokat fonetikai tényeknek, a fonémákat fonológiai (vagy fonológiai-fonetikai) tényeknek tekintjük, akkor azt mondhatjuk, hogy a beszédhangok a vizsgált nyelvben meghatározott korban zárt fonológiai halmazt (zárt fonémahalmazt) és nyitott fonetikai halmazt (nyitott fonémarealizáció-halmazt) alkotnak.

1.3. Nyelvtörténeti szempontból a fonémák szinkrón szempontú zárt halmaza is nyitott halmaznak minősül abban az értelemben, hogy a szinkrón fonémaállomány egyes elemei korábbi korszakokban ismeretlenek, más elemei későbbi korszakok fonémarendszeréből hiányoznak. Új fonémák keletkezésével, meglevő fonémák eltűnésével a fonémaállomány és a fonémarendszer is változik. A mai magyar nyelv fonémarendszeréről például régóta tudjuk, hogy az *ö* fonéma,[1] a *zs* fonéma[2] (vö. pl. *zsaru : saru*, *zseb : seb*, *zsír : sír*) nyelvünk külön életében fejlődött ki, az *ë* fonéma nyelvünk régibb korszakaiban közmagyar sajátság lehetett, a mai magyar nyelv-

[1] Losonczi Zoltán, Az *ö*-zés története: Nyelvtudományi Közlemények 44. 373—406 (1917), 45. 45—116 (1917), 45. 195—266 (1918).

[2] Kniezsa István, A magyar *zs* hang eredete: Magyar Nyelv 29. 94—104, 142—57 (1933).

járások jelentős részéből viszont hiányzik.[3] A fonémarendszernek történeti szempontból való nyitottsága természetesen nem olyan fokú, mint a morfémák halmazának nyitottsága, hanem sokkal kisebb mérvű. Ez a fonémák számának alacsony voltával függ össze.

1.4. A morfémák halmazának nyitottsága sem egyenlő mértékű minden morfématípusban. A járulékmorfémák (toldalékmorfémák) halmaza szinkrón szempontból viszonylag kevés morfémából áll s aránylag kevéssé nyitott: inkább csak a tőmorfémából járulékmorfémává válás foka és egyes járulékmorfémák elavultságának foka szempontjából nyitott ez a halmaz. A tőmorfémák, a szavak halmaza viszont megszámlálhatatlanul sok egyedből áll[4] s nagymértékben nyitott.

1.5. Történeti szempontból a járulékmorfémák halmaza is, a tőmorfémák, a szavak halmaza is nyitott, de a változás lefolyása szempontjából itt is lényeges különbséget találunk. A járulékmorféma-állomány csak lassan változik: az új képzők, jelek, ragok hosszú idő alatt alakulnak ki, a meglevők hosszú idő alatt halnak ki. A tőmorfémák, a szavak állománya szinte állandó mozgásban van: az új életkörülmények, az új szükségletek állandóan új szavakkal gyarapítják a nyelvet; a tőmorfémák, a szavak közt nem ritka jelenség a rövid idő alatt való elavulás.

1.6. De tovább mehetünk, s a tőmorfémák, a szavak állományáról is megállapíthatjuk, hogy ez sem tekinthető minden részletében egyenlő mértékben nyitott halmaznak. Ha például a tartalmas szavak összességét a formaszók összességével állítjuk szembe, akkor azt látjuk, hogy a formaszók halmaza — a számnevektől eltekintve — viszonylag zárt, alig egy-két száz szóból áll, problémát inkább csak az átmeneti esetek jelentenek. A formaszók kialakulása és elavulása is lassú folyamat eredménye, mint a járulékmorfémáké. A tartalmas szavak halmaza viszont nyitott, s állandó mozgásban van.

1.7. A mondatszók[5] halmaza is nyitott, mind szinkrón, mind diakrón szempontból, a kategórián belül viszont nem minden alcsoport mondható nyitottnak. Az érzelem- és akaratkifejező mondatszók halmaza szinkrón szempontból viszonylag zártnak tekinthető, a mutatók, a kérdők és felelők mindössze néhány egyedből álló zárt halmazt alkotnak; a hangutánzó mondatszók halmaza viszont pontos számmal meg nem határozható nyitott halmaz.

1.8. Ha a szókincs különféle rétegeinek gyakoriságát vizsgáljuk, akkor azt tapasztaljuk, hogy a formaszók csoportjára a nagyobb gyakoriságra mutató szám, a tartalmas szók csoportjára a ritkább előfordulásra mutató szám jellemző. Néhány tucat vagy legfeljebb néhány száz formaszó együttese csaknem ugyanolyan százalékarány-

[3] Laziczius Gyula: Bevezetés a fonológiába (Bp., 1932) különösen 50., 62—74, — A köznyelv *ë* [*e*] fonémájára vö. Kelemen József: Népünk és Nyelvünk 7. 208—9, 210. (1935) Nagy J. Béla: Magyar Nyelv 37. 82—96 (1941). Elekfi László: Magyar Nyelvőr 85. 159—65 (1961).

[4] Kelemen József, Hány szó van nyelvünkben?: Magyar Nyelvőr 78. 68—74 (1954); Uő, A magyar szókincs nagysága: Nyelvművelő (Bp., 1956) 37—44, és Iskolai nyelvművelő (Bp., 1959) 29—42.

[5] Kelemen József, A mondatszókról: Nyelvtani tanulmányok (Bp., 1961) 103—114; Uő, A szótárírás elmélete és gyakorlata a Magyar Nyelv Értelmező Szótárában (Bp., 1962) című kötetben 27 és 86—7.

ban fordul elő a folyamatos szövegben, mint a tartalmas szavak ezrei, tízezrei vagy százezrei. A Szalai Sándor által idézett adatok szerint Nemes Zoltán Szóstatisztikájában[6] a leggyakoribb 25 szóból („önállóan", azaz nem összetétel tagjaként használt szóból) 21 formaszó s még 3 (a *van*, *volt* és a *kell* ige) formaszóként is felfogható, s ez a 24 szó az összes előfordulásoknak 22,39 százalékát, közel egy negyedét jelenti. A Kalmár László szövegéből idézett „triviális" szavak mind formaszavak s az összes előfordulások 20 százalékát jelentik; a Verzár-szövegből idézett „triviális" szavak a *van* kivételével mind formaszavak s együtt az előfordulások 24 százalékát (a *van* nélkül 23,20 százalékát) adják; az Aranyból idézett „triviális" szavak, melyek az összes előfordulások 25,90 százalékát teszik ki, az *év* (1,10 százalék), a *van* (0,90 százalék) és a *volt* (0,70 százalék) kivételével mind formaszavak, az előfordulások 23,20 százaléka — sőt ha a *van* és *volt* igét is közéjük számítjuk, 24,80 százaléka —, tehát itt is a formaszókra esik.[7]

2. A leíró, a „szinkronikus" jelentéstan kérdése[8]

2.1. Nem érthetünk egyet azokkal, akiknek az a véleményük, hogy a „leíró" jelentéstan, azaz a szinkronikus jelentéstan kérdéseiben nem láthatunk mást, mint a for-

[6] Nemes Zoltán: Szóstatisztika egymillió szótagot felölelő újságszövegek alapján (Bp., é. n. [1939/41]) 55., 343. A feldolgozott szövegek 473 813 szóelőfordulást tartalmaznak. — Szalai a Nemes összeállításában a 13. helyen 1746 előfordulással szereplő *magyar* szót elmondott előadásában — nyilván elnézésből — kihagyta.

[7] Arany János leggyakoribb 25 szava összes költeményeiben az Ambrus Lajos-féle kéziratos Arany-szóstatisztika — (Arany János költészetének szóstatisztikája . . . a Ráth-féle 12 kötetes kiadás 1., 2., 3., 4. és 9. köteteiben felvett összes költeményeiből. I—IV. füzet, 1951. 860 lap, a Magyar Tudományos Akadémia Nyelvtudományi Intézetének könyvtárában) — adatai alapján végzett számításaim szerint az előfordulásoknak 26,075 százalékát adja, tehát több mint egy negyedét, s ezek egyetlen kivétellel, a *nagy* szó kivételével mind formaszók. Az első 24 formaszó (a *nagy* kivételével számított első 24 szó) az összes előfordulásoknak 25,56 százalékát, tehát még mindig több mint egy negyedét adja. A *van*, a *lesz* és a *nagy* kivételével számított első 22 szó előfordulási százaléka 23,84. (Az Ambrustól kimutatott 271 172 szóelőfordulásból 70 710, a *nagy* nélkül 69 337, a *nagy*, a *van* és a *lesz* nélkül 64 652 szóelőfordulás.) Az első 25 szó az előfordulások számának kisebbedése szerint: 1. *a, az* [névelő és mutató névmás] 24 983 előfordulással, 2. *nem* [tagadószó] 4500, 3. *s* 3772, 4. *van* 3301, 5. *is* 3060, 6. *hogy* [kötőszó és kérdő határozószó] 2998, 7. *egy* [névelő és számnév] 2622, 8. *és* 2411, 9. *de* 2146, 10. *ez* 2082, 11. *mint* 2023, 12. *ha* [kötőszó és kérdő határozószó] 1682, 13. *ki* [névmás] 1568, 14? *mi* [kérdő névmás] kb. 1400 (?), 15. *lesz* 1384, 16. *nagy* 1373, 17. *meg* [kötőszó és határozószó] 1187, 18. *csak* 1174, 19. *mely* 1109, 20. *már* 1086, 21. *ő* 1040, 22. *maga* 1033, 23. *még* 1005, 24. *én* 917, 25. *most* 864, összesen 70710 (?) előfordulással. Az Ambrus-féle szóstatisztika egyik fogyatékossága, hogy bizonyos esetekben a homonimákat nem választotta szét, ezért szerepel a fenti kimutatásban a feltételesen 14. helyre került *mi* kérdőnévmás körülbelüli számmal; az Ambrusnál szereplő „mi [kérdőnévmás és személynévmás] 997" + „minket 10" + „mibe, miben 10", „miből 11", „mihez 4", „mije 1", „miként, mikint 33 + 1 [=] 34", „mikép, miképen 80", „miket 1", „min 10", „minél 11", „mire 97", „miről 12", „mit 465", „mitől, mitül 12 + 2 [=] 14", „mivé 3", „mivel 89" helyett. A kb. 1400 számtól fölfelé vagy lefelé való néhány száz adatnyi eltérés módosítja a *mi* helyét a gyakorisági rangsorban, de az első 25 szóba mindenképpen beletartozik, s az első huszonöt szó százalékos arányán nem változtat lényegesen. Az Ambrusnál 1084 előfordulással szereplő *e* szót figyelmen kívül hagytam a leggyakoribb 25 szó összeállításában, mert feltehető, hogy az *-e* kérdőszó és az *e* mutató névmás adatainak szétválasztása után mindegyik szó előfordulási száma 864 alatt marad.

[8] Hozzászólásomnak ez a része a szerkesztők kívánságára — mint az értekezlet tárgyával távolabbi kapcsolatban álló gondolatmenet — itt rövidített alakban kerül közlésre. A teljesebb szöveg valamelyik folyóiratban fog megjelenni.

mális logikai elemzés kérdéseit. A szinkronikus jelentéstannak sajátos problematikája van, amely bizonyos pontokon egyezik a formális logikai elemzés problematikájával, de egészében különbözik tőle.

2.2. A szinkronikus jelentéstan egyik alapvető kérdése a jelentéstani egységek megállapítása. A nyelvi eszközökkel jelölt jelentéstani egységet *szemémá*-nak nevezem. A szemémának egyik legáltalánosabb, legismertebb típusa az a jelelentés-egység, amelyet a szótárak mint *egy* jelentést szoktak meghatározni; például *alma* 'az almafának kb. gyermekököl nagyságú, gömbölyded, húsos, leves belű, apró magvú ehető termése' (ÉrtSz. 1. jelentés).

A szónál nagyobb jelentéstani egység: szókapcsolat, kifejezés is jelölhet egy-egy szemémát; például az *Erisz almája* szókapcsolat jelentése 'valamely viszályt vagy nézeteltérést kiváltó ok' (ÉrtSz. 4. jelentés).

A járulék-morfémákhoz is kapcsolódik egy vagy több szeméma. A járulék-morfémához tartozó szemémát néha önálló szóval is jelölhetjük; például a *várok, nézék*-féle igealakok *-k (-o-k, -ë-k)* morfémájához az a szeméma tartozik, amelyet az *én* szóval is jelölhetünk. Vannak olyan járulék-morfémák is, amelyekhez olyan szeméma tartozik, amelyet egyetlen szóval nem szoktunk jelölni, csak körülírással határozhatunk meg; ilyen például az *almás, körtés*-féle szók *-s* morfémája.

Azok a szemémák, amelyeket egy-egy szóval, szókapcsolattal jelölünk, önállóbbak, határozottabbak, könnyebben megfoghatók, mint azok, amelyeket csak járulékmorfémákkal jelölünk.

2.3. A szinkronikus jelentéstan másik fontos kérdése a jelentés-struktúra vizsgálata a jelentéstartalom fontos és kevésbé fontos részeinek elválasztása, a szeméma határainak jellemzése (elmosódó vagy élesebb határok); a jelentés tartalmi és hangulati részének elemzése, a tartalmi részt és a hangulati értéket meghatározó tényezők vizsgálata.

2.4. A többértelműség (többjelentésűség) terén is találunk meghatározható törvényszerűségeket. Ilyen az a jelenség, hogy a szorosabban összefüggő szinonimák gyakran bizonyos fokig párhuzamos többértelműséget (többjelentésűséget) mutatnak; például a *gyár, hivatal, üzem* szók mindegyike jelenti a magyarban magán az intézményen kívül az intézmény helyét is, az intézményben dolgozó személyek összességét is. A *bab, borsó, lencse* szók nemcsak a növényt, hanem annak termését s a belőle készített ételt is jelentik.

A viszonyítást jelölő szók, főleg melléknevek, határozószók, névutók, kötőszók a többértelműségnek (többjelentésűségnek) más típusait mutatják, de ezeken belül is vannak olyan kisebb csoportok, amelyeknek tagjai egymáshoz hasonló jelentéshálózatot mutatnak; például *alsó, fölső; lë, lënn, alant, fël, fënn; alá, fölë; de, hanëm.* Sajátos szemémaviszonyokat találunk a névmások és a névmási határozószók körében.

2.5. A többértelmű (többjelentésű) szavak jelentés-struktúrájának sajátos kérdése a szóhoz tartozó jelentések belső rendje: az „alapjelentés" és a származtatott jelentések viszonya.

2.6. Tárgya a szinkronikus jelentéstannak a többszófajúság szemantikai vizsgálata is.

2.7. Kevésbé kidolgozott ága a jelentéstannak a szeméma-korrelációk kérdése: a szinonimia és az antonimia vizsgálata, a szűkebb és tágabb fogalomköröknek, a szócsaládok tagjainak, a paradigmatikus viszonyban levő szavaknak és morfémáknak jelentéstani összefüggése. Ezek tanulmányozása nagyrészt a jövő jelentéstanának kutatóira vár.

3. A nyelv modelljének kérdése

A Papp Ferenctől bemutatott modellt kelleténél vázlatosabbnak tartom, azért — Kalmár László megjegyzésétől is bátorítva — az alábbiakban egy komplexebb modell vázlatát mutatom be. A nyelvi megnyilatkozás linearitását, a beszélés és az írás lefolyását figyelembe véve a megnyilatkozást, a hangok és betűk egymásutánját többféle egységre bonthatjuk.

3.1. A legkisebb egységet akusztikailag is, nyelvileg is a hangok jelentik. Az egymás után következő hangok összessége a nyilatkozat bizonyos terjedelmétől mint minimumtól függően különféle kombinációs egységeket hoz létre. — A hangsorból bizonyos szempontból kiemelkedik és magasabbrendű egységnek tekinthető a hangkapcsolat, legalábbis a hangkapcsolatoknak bizonyos típusai, mint például *mb*, *mp*, *nd*, *nt*, *ŋg*, *ŋk* vagy például a diftongusok. — A hangok és hangkapcsolatok közül ismét kiemelkedik egy sajátos típus: a szótag. Ez állhat egyetlen hangból is, több hang kapcsolatából is. A szótag nyelvi szerepe még nem tekinthető teljesen tisztázottnak, de nyelveket megkülönböztető és a nyilatkozatot a hangoknál nagyobb egységekre tagoló szerepe kétségtelen. — A nyilatkozat ritkán egyetlen szótagból, rendszerint azonban hosszabb vagy rövidebb szótagsorból áll. — A szótagnál nagyobb egység az ütem vagy szakasz vagy — kevésbé találó névvel — szólam. Az ütem ritkán egyetlen szótagból áll, rendszerint azonban egy hangsúlyos szótaghoz tartozó egy vagy több hangsúlytalan szótag alkot egy ütemet. — Az egyetlen ütemből álló nyilatkozat is ritkább a két vagy kettőnél több ütemből (ütemsorból) álló nyilatkozatnál. Az ütemsor egyes ütemei bizonyos rendező elv szerint követik egymást, gyakran a főhangsúlyos és mellékhangsúlyos ütemek meghatározott összefüggése (például mellékhangsúlyos — főhangsúlyos, vagy főhangsúlyos — mellékhangsúlyos, vagy mellékhangsúlyos — főhangsúlyos — mellékhangsúlyos sorrend) alapján.

3.2. Ha a hosszabb nyelvi megnyilatkozás jelentéstani oldalát nézzük, akkor a szemémák szerinti összefüggést találjuk. A nyilatkozat rendszerint szemémasorként fogható fel, azzal a megjegyzéssel, hogy a szemémának elrendeződése nem annyira mechanikusan lineáris, mint a hangok, szótagok elrendeződése. — A szemémasorból rendszerint szintén kiemelkednek olyan szemémakapcsolatok, amelyek szorosabb nyelvi egységként jelentkeznek. A szemémakapcsolat jellemző típusa az egyetlen morfémával (egyetlen szóval) kifejezett szemémakapcsolat és a szóalakkal mint morfémakapcsolattal kifejezett szemémakapcsolat, például *kanca* 'nőstény ló', *mén* vagy *csődör* vagy *monyas* 'hím ló'; *látjuk* stb. A hosszabb nyilatkozat rendszerint szemémakapcsolatok egész sorából vagy szemémák és szemémakapcsolatok sorából áll.

3.3. A szemémakapcsolat már átvezet bennünket a következő szinthez, a morfémák szintjéhez. A nyilatkozat néha csak egyetlen morfémából áll, rendszerint azonban egész morfémasort elemezhetünk ki a nyilatkozatból. A morfémasorból kiemelkedő

magasabbrendű egység a morfémakapcsolat, illetve ennek egyik jellegzetes faja, a szóalak. Ha azt mondom: *látom*, ebben egyrészt a *l-á-t-o-m* hangokat, illetőleg a *lá* és *tom* szótagokat, másrészt a *lát-* igetővel és az *én* szóval jelölt szemémákat látjuk, de látjuk benne ezen kívül a *lát- (láto-)* tőmorfémát és az *-m (-om)* járulékmorfémát, valamint a tőmorféma és a járulékmorféma sajátos kapcsolatát is. A szóalakban tehát együtt van a fonológiai (fonetikai), a szemantikai tagoltság és a morfológiai tagoltság. — A hosszabb nyilatkozat rendszerint meghatározott számú és minőségű szóalaksorból áll.

3.4. A szóalaksorból kiemelkedő magasabbrendű egység a szintagma mint meghatározott mondattani viszonyban levő két (esetleg több) szónak, illetőleg szóalaknak kapcsolata. A hosszabb nyilatkozatot alkotó szintagmasorból gyakran egy vagy több szintagmakapcsolatot elemezhetünk ki. A hosszabb mondat szintagmák kapcsolatából épül fel.

3.5. A morfémák és a szintagmák szintjénél magasabb rendű egység a nyelvi megnyilatkozásban a mondat, az a legmagasabb szintű egység, amelyben a hangtani-szemantikai-morfológiai-szintagmatikus egységek még szoros kapcsolatban vannak egymással, s amelynek egésze meglehetősen nagy szerepet játszik a hangtani, a szemantikai, a morfológiai és a szintagmatikus egységek és összefüggések rendjének kialakításában. Lehet ugyan mondat egyetlen szó vagy szóalak is — például mondatszók esetében, mint *Jaj! Csitt!* vagy a *Könyvtár* [mint a könyvtárhelyiség ajtaján levő felirat]. *Vizet!* [mint felszólítás] *Lassan!* vagy a *Várlak*-féle szómondatok, — de legtöbbször több szóalak, illetőleg több szintagma vagy több szintagmakapcsolat alkot egy mondatot. — Két vagy több szorosabban összetartozó mondatból mint tagmondatokból álló egység az összetett mondat: az a mondattípus, amely a tagmondatokban a szorosabb értelemben vett mondattani egység alacsonyabb fokú típusát is, a mondat egészében a mondattani egység magasabb típusát is magában foglalja. — A mondatnál magasabbrendű egység a mondatsor és az összefüggő mondatsorokból álló megnyilatkozásbeli egységek. Ezek már inkább stilisztikai, mint nyelvtani egységek.

Visszatekintve az elmondottakra, azt mondhatjuk, a vázoltak adják a nyelvi megnyilatkozás lineáris tagozódásának azokat az elemeit és elemi összefüggéseit, amelyekből felépül az egész nyelvi megnyilatkozás, bármilyen hosszú legyen is az. Ezeket az egységeket és ezeknek egymás fölé rétegeződését a következő ábrával szemléltethetjük. (Lásd a következő lapon).

Az itt érintett egységeknek (hang, szótag, ütem; szeméma, szemémakapcsolat; morféma, szóalak; szintagma, szintagmakapcsolat; mondat, mondatkapcsolat) természetesen nemcsak lineáris összefüggéseit és felépítését, hanem rendszertani összefüggéseit is vizsgálnia kell a nyelvtudománynak. De az egyes egységek rendszertani elemzése sajátos, a linearitástól lényegesen különböző problémákat vet fel, azért azokkal tt nem foglalkozom.

IV.	m o n d a t s o r
	mondatkapcsolat (összetett mondat)
	mondat
III.	s z i n t a g m a k a p c s o l a t - s o r
	szintagmakapcsolat (szószerkezetekből álló kapcsolat)
	s z i n t a g m a - s o r
	szintagma (szószerkezet)
II.	s z ó a l a k - s o r
	szóalak (morfémakapcsolat)
	m o r f é m a - s o r
	morféma
I. B.	s z e m é m a k a p c s o l a t - s o r
	szemémakapcsolat
	s z e m é m a - s o r
	szeméma
I. A.	ü t e m - s o r
	ütem
	s z ó t a g - s o r
	szótag
	h a n g - s o r
	hang

Deme László

1. Amikor matematikai nyelvészetről beszélünk — s főleg, ha e beszélgetésben nemcsak, sőt nem is elsősorban matematikai nyelvészek vesznek részt, hanem matematikusok és nyelvészek is — hasznosnak látszik meghatározni: mit értünk és mit értsünk — vagy legalább így: ki mit ért — a *matematikai nyelvészet* kifejezésen. Félő ugyanis, hogy a szót nem mindenki használja azonos tartalommal, s így — azt hívén, hogy egymáshoz szólunk azonos témáról — könnyen megeshet, hogy tulajdonképpen egymás mellé beszélünk, s ráadásul nem is ugyanarról.

Engedtessék hát meg nekem egy distinkció, s döntessék majd el róla: jogos és helyes-e.

Úgy látom: matematikus és bizonyos tekintetben matematikai nyelvész kollégáink szóhasználatában ez az új irányzat nem pusztán új módszereket jelent, hanem legalább olyan mértékben új szempontokat s ezzel új szemléletet, talán túlzás nélkül mondom: új e l m é l e t e t is. Lehet, hogy ebben nincs igazam. Az viszont kétségtelen, hogy ebben a matematikai nyelvészetben a matematika nemcsak módszereket szolgáltat, hanem s z e m p o n t o k a t is: a halmazelmélet segítségével való vizsgálat, úgy látom, túlmegy az új módszer bevezetésén, s magát a kérdésfeltevést szabja meg az eddigitől eltérően; lényegében tehát problematikában is új útra lép.

Merészség volna tőlem azt megítélni vagy akár csak megjósolni, hogy ez közelebb visz-e bennünket a nyelv lényegének marxista — tehát valóságos saját természetének megfelelő — megragadásához, s részleteinek az eddiginél helyesebb megismeréséhez; vagy pedig tévútra kanyarít. Nem is ennek eldöntése vagy feszegetése a célom: szerintem minden új törekvést támogatni kell, hogy minél hamarabb bebizonyíthassa igazát vagy téves voltát. Csak annyit akartam mondani: hiányos ismereteim s a matematikusok tájékoztatásai alapján a matematikai nyelvészetet elsősorban új szempontokat felvető stúdiumnak látom.

S mi az ezzel szemben, amihez jó néhány nyelvész kollégánk szolgáltatott itt adalékokat abban a hiszemben, hogy ők is matematikai nyelvészetről beszélnek? Talán megbocsátják nekem a szójátékot, ha azt mondom: amiről ők beszélnek, az nem matematikai nyelvészet, hanem inkább *nyelvészeti matematika;* azaz nem matematikai s z e m p o n t o k é s m ó d s z e r e k alkalmazása a nyelv vizsgálatában, hanem a hagyományos nyelvészet által feltett kérdéseknek belsően nyelvészeti szempontú és módszerű, csak éppen matematikai vagy matematikus jellegű eljárásokkal kibővített és pontosabbá tett vizsgálata. Itt a s z e m p o n t föltétlenül továbbra is a nyelvészé. Statisztikát készít — persze matematikai módszerekkel,

s ha majd géppel, akkor még könnyebben és pontosabban — nyelvi tényekről; de hogy *mi*rő*l*, és az eredményt *hogyan értékeli*, azt továbbra is a nyelvtudomány eddigi szempontjai szabják meg számára.

Ami az ilyen statisztikáknak, képletek alkalmazásának stb. értékelését illeti, itt talán már nem kötelező rám egészen az az óvatosság, amelyet a matematikai nyelvészetről nyilatkozván illendőnek éreztem. A nyelvészeti matematika, vagyis a statisztika és a sablonozás egyáltalában nem idegen tőlünk, s nem is új számunkra: a szókincs tárgyi rétegeinek eredetstatisztikájával próbáltam magam is egyszer megoldani — s már nem is úttörőként — a népi csizmadiamesterség eredetének kérdését;[1] s a hangátvetések statisztikai értékelésével például azt a kérdést: hogyan hat az írásbeliség rögzítettsége a hangátvetés újabb eseteinek keletkezési lehetőségeire.[2] És a statisztikai értékelést egyre többen alkalmazzák mások is. Ami meg a sablonozást illeti: a továbbiakban erről a szórenddel kapcsolatosan lesz még szó.

Remélem, sikerült világossá tennem fenti eszmefuttatásom szándékát. Lényege ez: a nyelvészeti szempontok matematikai módszerű vizsgálata máris kipróbált, s szerintem föltétlenül gyümölcsöző dolog; alkalmazzuk az eddiginél bátrabban, gyakrabban és pontosabban. A matematikai szempontoknak a nyelvre való alkalmazása új próbálkozás; segítsük bebizonyítania önmagáról, hogy helyes és hasznos-e. — Csak azt ne higgyük, hogy aki az *i*-zést a hagyományos kérdésfeltevésen belül a szokásos adatfelsoroláson kívül statisztikai módszerekkel is megvizsgálja, az már matematikai nyelvészetet művel, s munkája sikerével mintegy igazolja akár a kibernetika, akár a halmazelmélet nyelvre való alkalmazásának helyességét.

2. S ha már belekezdtünk a distinkciókba, megkísérelnék még egy másikat is. Nálunk a strukturalista vizsgálatot és a matematikai nyelvészetet személy szerint azonos kollégák képviselik, vagy legalábbis propagálják; s ezért bennük is, meg a kívüllállókban is úgy rögződött a dolog, hogy ez a kettő lényegében azonos, legföljebb csak elnevezésében s bizonyos árnyalataiban más. — Nos, jó volna tudnunk: valóban úgy van-e. Vállalja-e mondjuk Kalmár László vagy Frey Tamás Harrist, és viszont. Ehhez még hozzászólni sem merek, de a kérdés feltevése talán jogos: ezt még az alábbiakban is igyekszem megtámogatni.

3. Még egy distinkció; utolsónak. Gyakran van szó mostanában a fordítógép elkészítésének és nyelvi kiszolgálásának kérdéséről. S ez föltétlenül helyes: a gépi fordítás is olyan új elgondolás, amelyet meg kell kísérelni, ki kell próbálni, hogy tisztába jöhessünk reális lehetőségeivel és értékével.

Világos, hogy egy olyan szerkezetnek, amely az *igen* és a *nem* kategóriáján kívül semmi egyebet nem ismer — bár ennek eldöntését igen nagy termelékenységgel végzi végeláthatatlan sorozatokban — másféle nyelvtanra van szüksége, mint az emberi olvasóknak, akiknek agya a barkochba játékon kívül még igen sokféle és sokrétű műveletegyüttesre is képes. A feladat tehát nem lehet más, mint lebontani a hagyományos nyelvtant erre a kéttagú — de szerencsére sokszor ismételhetően kéttagú — rendszerre; s ehhez persze formalizálni kell mindent, ami a hagyományos feldolgozásban funkcionális, mert a gép a funkciót nem érzi, csak a formális elemeket, amelyek azt kifejezik, hordozzák.

[1] Magyar Nyelv 43. 226 (1947).
[2] Magyar Nyelv 41. 55 (1945).

Nos, tehát: a hagyományos és a fordítógép számára kidolgozandó nyelvtan evidensen két külön dolog. Azok, akik most a fordítógép nyelvi kiszolgálásának ügyét vállalják, vagy legalábbis propagálják, a megoldást a nyelvtan strukturális — de talán helyesebb, ha így nevezem: strukturalista — átdolgozásában, illetőleg újrafeldolgozásában keresik. S ez az, aminek helyességét illetően nekem vannak bizonyos szubjektív — be nem bizonyított, de meg sem cáfolt — kétségeim. A képletet nem úgy látom, mint ők, hogy: $H \neq St = G$, azaz a hagyományos nyelvtan nem azonos a strukturalista feldolgozással, ez viszont azonos a gépi fordítás nyelvi kiszolgálásának megoldásával; hanem így: $H \neq St \neq G$. Vagyis: a hagyományos nyelvtan és a fordítógéptől felvetett új igények közötti különbség áthidalásában a strukturalista feldolgozás nem a megoldást jelentő ellenpólus, hanem egy közbülső harmadik.

Persze részemről ez is csak intuicionális alapú feltevés. De mi az alapja? Az, hogy a strukturális vizsgálat — éppen mint „-izmus", s ezt képviselői is hangsúlyozzák — nem módszer elsősorban, hanem szemléletmód; végső eredményeiben nem feldolgozási mód döntően, hanem elvi váz. Legalábbis azt tapasztaljuk: magyar és általunk ismert külföldi strukturalistáink elsősorban elméleti eredményeket produkálnak, nem lebontási szempontokat adnak. Egy példát csak! Az *-n* melléknévi modálisrag a hagyományos felfogás szerint rag, egyik strukturalistánk szerint alighanem képző.[3] Most ne döntsük el, kinek van igaza. De szögezzük le, hogy ez a fordítógép számára teljesen közömbös kérdés, mivel az nem ismer sem ragot, sem képzőt, csak szótövet, és toldalékot. S nyilván a szótövek közé kerülnek a magyarban a ritkább képzőkkel alkotott derivátumok; viszont az egységes toldalék kategóriába az igen nagy produktivitású *-ás, -és* képző is. — De ha ez így van, akkor mennyiben vitt közelebb a fordítógép nyelvi kiszolgálásának megoldásához e ponton a strukturalista nyelvtan? Nem valami közbülső harmadiknak bizonyult-e valóban? S ha igen, akkor miért a fordítógép igényeire való hivatkozással indokoljuk ezt a vizsgálati módot?

4. Egy közbevetett megjegyzés. Az „a tergo" szótár, amiről mostanában többször is szó esett, érdekes és nyilván hasznos dolog. De azt ennek megcsinálásával sem venném eldöntöttnek, hogy a fordítógépnek föltétlenül hátulról kell bontania a magyarban is, s így az „a tergo" szótár eleve a magyar fordítógép szótára. Nyelvünk oly világosan morfémaszintetizáló szerkezetű, hogy az elülről való bontás mellett legalább ugyanannyi érv szól. Ha tehát az „a tergo" szótár elkészítését a fordítógép miatt tervezzük, először nagyon meg kell vitatnunk: valóban fogjuk-e majd használni hozzá.

5. Néhány konkrétumot az elmondottakhoz a szórend sablonozásáról, és főleg ennek nehézségeiről.

Jól tudjuk: a szórend a magyarban meglehetősen szabad. Ez persze nem azt jelenti, hogy a szórend területén kötött értékek és szigorú szabályok nincsenek: a szabadság csak a morfológiai jelölő funkció és a grammatikai kötöttségek hiányát jelzi.

Nézzük példán! Az angolban „The boy gives his friend a book" ezt jelenti: „A fiú ad a barátj*ának* egy könyv*et*". Hogy az ige előtti ragtalan főnév alany, az ige utáni első részeshatározó, második pedig tárgy, azt kizárólag szórendi helyük mutatja. Megcserélni nem lehet őket, mert az „A book gives the boy his friend" (ami a magyarban természetes: „Egy könyvet ad a fiú a barátjának") az angolban szó szerint ezt jelen-

[3] Antal László: Magyar Nyelv 56. 54 (1960).

tené: „Egy könyv adja a fiúnak a barátját". Itt a szórendnek morfológiai jelölő, funkciója van, s ezt a gép könnyen megérti, hiszen azt beléje tudjuk programozni, hogy az angolból az ige előtti főnevet ragtalan főnévvel, az ige utánit tárgyragossal, illetőleg ha — mint itt — kettő van, az elsőt *-nak, -nek* ragossal, a másodikat tárgyragossal kell fordítania; s ugyanezt meg tudja csinálni fordítva is.

Az oroszban: „*Мы идём* в кино": a szórend az alany—állítmány—határozó egymásutánján alapul. Az előre emelt határozó sem változtat ezen: „*Вечером мы идём* в кино". Ez grammatikai kötöttség, nem morfológiai; de a gépet nem zavarja, mert a mondatrészeket szófajukról (toldalékolási osztályukról) fel tudja ismerni.

A németben „*Wir waren* im Zoo" látszólag ugyanúgy alany—állítmány kötöttségű szórend, mint az előbbi orosz példáé. De a „Gestern *waren wir* im Zoo" rámutat arra, hogy ez csak látszat, az előbbinek és ennek nem az alany—állítmány elve a közös alapja, hanem az, hogy az állítmány (vagy ragozott része) mindig a második helyen áll. De a géppel ezt is meg lehet értetni.

A magyarban sincs nagyobb baj a nyomatéktalan mondatoknál. Igaz: „A játékosok *levonulnak* a pályáról"; „A pályáról *levonulnak* a játékosok"; „A játékosok a pályáról *levonulnak*" : mind a három lehetséges megoldás, és mind a három nyomatéktalan. S ráadásul mindegyikhez járulhat még közömbös nyomatékú határozó: „Tegnap a játékosok levonultak a pályáról"; stb. — De a statisztikai vizsgálatok alighanem azt mutatnák ki, hogy a közömbös stílusú értekező prózában — s a fordítógépet úgyis nyilván erre akarjuk majd programozni — az igazán nyomatéktalan mondatok tipikus szórendi megoldása: $(H+)\ A + \acute{A} + B$; esetleg $(H + B +)\ A + \acute{A} + B$; ahol: H = előre emelt határozó; B = bővítmény (határozó vagy tárgy); A = alany; $\acute{A}$ = állítmány; s a sok példából kihámozható az is, hogy az állítmány meglehetős biztonsággal helyezhető az utolsó előtti helyre, az alany pedig eléje. Persze ez stilisztikai árnyalatokat jócskán elhanyagoló szabály; de mégis fogódzó; s az ennek alapján létrehozható megoldás általában nem rossz; s mindezt a fordítógép is meg tudja emészteni.

De ehhez persze a gépnek azt el kell döntenie, hogy a mondat nyomatéktalan valóban; s ez már nem olyan könnyű a számára. Ha a magyar „célnyelv", talán az idegen előzmény segít. Ha „bemenő nyelv" a magyar, az igei egyenes szórend ad eligazítást; — feltéve persze, hogy az állítmány igekötős, nominoverbális, összetett, tehát olyan, amelyiknek lehet egyenes és fordított szórendje.

S nem is ez a legnagyobb nehézség, hanem az, hogy a szórendnek a magyarban két egészen különböző síkja van, s mindegyik síkon egészen mások a törvények. Az egyik sík a szabad mondatrészeké, a másik az alárendelt szintagmatagoké.

Az előbbi példában szabad mondatrészek voltak csak. De ha ezeket bővítjük, bővítményük alárendelt szintagmatag: „Tegnap az órát Jóska felakasztotta a falra" = $= H + T + A + \acute{A} + H$*; a „Vasárnap délután[1] a régóta megjavított órát[2] az éppen otthon tartózkodó Jóska[3] felakasztotta[4] a szoba frissen festett falára[5]" ugyancsak: $H + T + A + \acute{A} + H$; de részletezve már: $hH + hjT + hhjA + \acute{A} + b(hj)H$; s ebben a képletben az önálló (nagybetűs) mondatrészek továbbra is szabad és független változók; de az alárendelt szintagmatagok (a kisbetűsek) csak fölérendeltjükkel

* A nagy- és kisbetűs szimbólumok a szó mondatrészi minőségét jelölik; a + mellérendelt, a közvetlen kapcsolás alárendelt helyzetükre utal. Tehát b(hj)H = olyan független határozó, amelynek határozóval bővített tulajdonságjelzője, s az egész komplexusnak még birtokosjelzője van. — Az indexszámok a komplex szabad mondatrészek végét jelzik.

együtt mozoghatnak. — A gépnek tehát két síkon kell dolgoznia: először össze kell állítania az egyes komplex szabad mondatrészeket (például: *hhjA*), s azután e komplexusokat a szabad mondatrészek szabályai alapján el kell helyeznie a mondatban.

Ez a feladat bonyolult, de nem megoldhatatlan, mert a mondatrészek és szintagmatagok alaktani úton vagy környezetleírással meghatározhatók; s a komplexusok belső rendje eléggé kötött: a jelző értékű hátravetett határozót és a *-nak, -nek* ragos birtokost kivéve a szintagma alárendelt tagja szorosan fölérendeltje előtt áll; a névelő az egész komplexus előtt, ha a legfőbb dominánsé, a szintagmatag előtt, ha csak az övé; stb. — „A bányatérképezés elveinek és módszereinek egységesítése és a bányászati mérések hibahatárainak megállapítása[1] az államosítások után[2] egyik fontos feladatává[3] vált[4] egész bányászatunknak, de különösen szénbányászatunknak[5]": e mondat elnagyolt képlete $(A + A) + H + Á + (H + H)$, tehát alanya és utolsó határozója maga is kételemű; komplex képlete pedig: $[b(b+b)\ A + (jb)bA] + H + jjH + Á + (jH + hH)$.

Még ez is csak programozás kérdése volna, ha a mondat nem volna nyomatékos. De bizony az: az állítmány előtti határozó főhangsúlyos. Márpedig a nyomatékos mondatokhoz érve — és sajnos ezek az értekező prózában a mondatok állományának négyötödénél is nagyobb részét teszik — a magyarban megszűnik a mondatrészi jelleggel való meghatározhatóság, s egy egészen más elv uralkodik el. Íme: „Forgácsmentes alakítással[1] hideg vagy meleg állapotban[2] munkálják meg[3] az anyagot[4]" = $= H + \boldsymbol{H} + Á + T^*$; „Ilyen esetben[1] nem anyagleválasztás[2] történik[3] a megmunkáláskor[4] . . . $= H + \boldsymbol{A} + Á + H$; itt a mondatrészi minőség semmit sem mond az elrendezés szabályairól. Itt a legfőbb törvények a következők: 1. gyakran áll a mondat élén egy olyan bevezető rész (= *B*), amelyik vagy az előző mondatból vezet át, vagy valami kevéssé súlyos határozó; 2. ez után állhat még egy előre emelt, új mozzanatot közlő, de nem főhangsúlyos rész (= *Ú*; példáink között nincs ilyen); 3. utánuk a főhangsúlyos rész a végére tapadt állítmánnyal (= ***F***); ebbe mindig beleértendő az állítmány is, illetőleg ragozott része; 4. s végül ez után a csekély súlyú egyéb mondatrészek (= *E*). A nyomatékos mondat általános képlete tehát: $B + Ú + \boldsymbol{F} + E$ s itt teljesen közömbös, hogy melyiknek milyen a mondatrészi minősége. Íme: a fenti két példa egyaránt $B + \mathbf{F} + E$ képletű, pedig az elsőben határozó, a másodikban alany a mondathangsúlyos rész; az elsőben tárgy áll a mondat végén, a másodikban határozó.

Nos: a nyomatéktalan mondatnak $(H +) A + Á + B$ képletét a gép meg tudja érteni; de a döntő többségben jelentkező nyomatékos mondatnak $B + Ú + \boldsymbol{F} + E$ általános képletével már nem tud mit kezdeni, mert ahhoz értelmi elemzés kell, amire képtelen.

S ez példa arra is: a strukturalista feldolgozás, ami formalizál ugyan, de emberek számára, valóban nem szolgálja közvetlenül a hagyományos nyelvtannak a fordítógép nyelvére való átültetését; hanem a kettő között egy harmadik: itt éppen inkább elvisz tőle, mint közelebb hozzá.

Persze, ami nehéz és megoldatlan, az nem föltétlenül lehetetlen és megoldhatatlan. Csak el kell dönteni, melyiket tartsuk sürgősebbnek: a gép számára való feldolgozást, vagy — ami korántsem azonos vele, sőt nem is lépés feléje: — a strukturalista módszerűt.

* A félkövér szimbólum az adott mondatrész főhangsúlyos voltára utal.

6. S végül még valamit, a „nemzetközi fordítókulcs" kérdéséhez.

Körülbelül tíz éve ismerkedtem meg Gábor Mihály nemzetközi fordítókulcsával, amikor bírálatra küldték meg nekem. Akkor még fordítógépről egyikünk sem hallott; a kísérletet a szerző emberi használatra szánta, s így bíráltam el én is. Mint emberi használatra szánt segédeszköz, ez a kísérlet érdekes; de megdöbbentő és hajmeresztő. Erre a célra annak találom ma is, s nem látom a gyakorlati értelmét annak, hogy valaki nyelvtudás nélkül leüljön fordítani, azaz — fejszével javítson zongorát. — De a fordítógéppel folytatott kísérletek fényében azt meg lehetne és meg is kellene nézni: az ott megkísérelt formalizálások ehhez nem használhatók-e fel valamilyen módon.

Így mondom: „ehhez"; és: „valamilyen módon". Mert azt semmiképpen sem gondolnám, hogy Blériot a Csatorna átrepülésével tulajdonképpen Daedalus és Ikarus legendáját igazolta volna.

*Szépe György**

1. Rényi Alfréd előadásának a témája nagyon érdekes a modern nyelvészet számára. A szűkebb és tágabb értelemben vett információelmélet különbsége párhuzamba hozható a fonéma és a morféma szintjének különbségével.[1]

A fonéma szintjén nem nyelvi jelekkel, hanem úgynevezett „figurákkal"[2] dolgozunk; ennek a szintnek az egységei nem bírnak jelentéssel. (A fonémának a jelentéshez való viszonyát természetesen számos iskola, különféle megközelítéssel és változatos terminológiával igyekszik — de általában ilyen értelemben — tisztázni.) A nálunk legelterjedtebb terminológia szerint: a fonémának nincs saját „jelentése", csupán megkülönböztető funkciója, amely a morfémák és/vagy szavak differenciálására vonatkozik. — Tehát a fonológia általában a „jelentés" figyelembevétele nélkül dolgozik. Vannak persze olyan törekvések, amelyek épp a fonémák elhatárolásában kívánják igénybe venni a „jelentés" segítségét. (A fonetikában természetesen irreleváns a „jelentés".)

A morféma szintjén nyelvi jelekkel dolgozunk; ennek a szintnek az egységei — a legtöbb iskola felfogása szerint — „jelentéssel" bírnak. (Más megfogalmazásban: a jel a jelentés kifejezése; megint más felfogásban: a jelentés a jel használatának a szabálya.[3]) A morfémák elhatárolásában, meghatározásában s a morfémákkal való egyéb műveletek során általában figyelembe veszik a jelentést is. — (Természetesen ha a *morféma* terminus egy formális rendszeren foglal helyet, akkor definíciója is formális.)

A szűkebb értelemben vett információelmélet bizonyos tételeit már eddig is eredményesen használták fel a legmodernebb fonológiában. Jakobsonnak és társainak a

* Az elhangzott hozzászólásnak némileg kibővített, rendszeresebb változata.

[1] A párhuzam nyelvészeti oldalára vö. például: André Martinet, La double articulation linguistique: Recherches structurales = Travaux du Cercle Linguistique de Copenhague 5. 30—7 (1949).

[2] Vö. Louis Hjelmslev: Prolegomena to a theory of language (Baltimore, 1953) = = Supplement to International Journal of American Linguistics, 19. kötet, 1. szám, 29, 34 stb. — Újabban: Antal László, Jel, jelentés, szövegösszefüggés: Magyar Nyelvőr 85. 203—8 (1961).

[3] Vö. Charles W. Morris, Foundations of the theory of signs: International encyclopedia of unified science (Chicago, 1955) 101 kk. — Telegdi Zsigmond: Bevezetés a nyelvtudományba I. (Bp., 1961. Kézirat gyanánt) 15—9. — Lásd még Antal Lászlónak a 2. jegyzetben hivatkozott cikkét is.

fonematikus megkülönböztető jegyekre épített fonológiai elmélete voltaképpen az információelméleten alapul.[4]

A morféma szintjén talán viszont azért nem volt jelentősebb hatása az információelméletnek, mivel a tágabb értelemben vett változata még idáig kevésbé volt kidolgozva. Nyilvánvaló, hogy a nyelv egészének elmélete számára nagy jelentőséggel bírna ennek a kidolgozása is.

2. Kelemen József hozzászólásában érintette azt a kérdést, hogy a strukturalisták — közelebbről Antal László — elhanyagolják a jelentés szempontjait. Azt hiszem, hogy a magyarországi „strukturalista" nyelvészek, közelebbről pedig Antal László, nem vádolhatók ilyesmivel. Antal László írt Magyarországon Gombocz Zoltán műve óta először önálló szemantikai könyvet, amely rövidesen meg is jelenik.[5]

A hazai modern nyelvészek — azt hiszem, mindnyájan — nagyon fontosnak tartják a szemantikát. De nem akármilyen szemantikát! A pszichológiai vagy ontológiai kategoriákon, vagy ezen-azon alapuló jelentéstannal kapcsolatban a bírálat látszik fontosnak. Úgy tudom, Antal könyve is ezt tartja egyik fő feladatának. Ezekkel szemben tartjuk fontosnak a nyelvészeti jelentéstan kiépítését. Antal is ehhez kívánt hozzájárulni kísérletével; természetesen az más kérdés, hogy célját mennyiben érte, érhette el. Véleményem szerint az ugyanis még nem egészen biztos, hogy jelenleg teljesen időszerű, tehát a siker reményében megkísérelhető feladat a szemantikai teória megalkotása. Azt viszont kétségtelennek tartom, hogy a nyelv egészének összefoglaló elmélete csonka volna szemantikai teória nélkül.[6]

3. Deme László hozzászólásával kapcsolatban azt hiszem egyetértünk vele abban, hogy a matematikai nyelvészet is nyelvészet — csak matematika segítségével folytatott nyelvészet —, de még nem találkozott a filológusok statisztikai módszereivel. A fejlődés konvergáló iránya egyre inkább látható. Egyelőre még nem világos, hogyan kerül majd kapcsolatba a statisztikai és a nem statisztikai módszerek alkalmazása. Papp Ferenc és mások megpróbálták ezt a kérdést alaposabban megvilágítani.[7]

Úgy gondolom, hogy a matematikai nyelvészet nem más, mint a nyelvészet bármely fajtájának, optimális esetben a modern „strukturális" nyelvészetnek egy, a matematika segítségét is igénybe vevő gazdagabb fajtája.

[4] R. Jakobson—C. G. M. Fant—M. Halle: Preliminaries to speech analysis (Cambridge, Mass., 1952); R. Jakobson—M. Halle: Fundamentals of language (Hága, 1956). — Kritikus hazai bemutatásuk: Fodor István, A megkülönböztető elemek jelentősége: Nyelvtudományi Közlemények 63. 129—60 (1961).

[5] Időközben megjelent Antal László könyve: Questions of meaning (Hága, 1963). — A másik hivatkozott mű: Gombocz Zoltán: Jelentéstan (Pécs, 1926). — Itt említem, hogy az ötvenes évek végefelé alkalmam volt kéziratban olvasni Martinkó Andrásnak egy könyvméretű jelentéstani dolgozatát, ez azonban — sajnos — nem került kinyomtatásra és így közforgalomba sem.

[6] Természetesen — ettől függetlenül — továbbra is fontosnak tartom annak hangsúlyozását, hogy a nyelv szerkezetének leírásában, a grammatikában téves a szemantikai kiindulás. Úgy látszik, hogy míg az egzakt, formális grammatika nagy segítségül lehet a szemantika számos aspektusának kidolgozásában, addig a szemantika — legalább is egyelőre — a grammatikának nem tud segítséget nyújtani. Sőt: a jelentésre való hivatkozás gyakran megakadályozza a további grammatikai kutatást. Erre vö. N. Chomsky véleményét: Third Texas conference on problems on linguistic analysis in English (Austin, Texas, 1962) 59.

[7] Vö. Papp Ferenc, Nyelvi rendszer, közlési folyamat és ezek néhány matematikai modellje: Kötetünkben 75—88; továbbá Papp előadásához kapcsolódó hozzászólásomat: Kötetünkben 160—164. Mindkét helyen található továbbiirodalmi utalás is.

Papp Ferenc

Egyetlen igen tanulságos kérdéssel kívánok foglalkozni, nevezetesen a strukturalizmusnak a gépekhez való viszonyával.

1. Rényi Alfréd, Hell György, Sipőczy Győző előadásait hallgatva, mindannyian még jobban meggyőződhettünk arról, mennyire alaptalan az a korábban nem egyszer elhangzott vélemény, hogy a strukturalisták „a gépből születtek", a „strukturalisták csupán a gép kiszolgálói", „a strukturalisták nem is igazi nyelvészek, csak egy szűk, praktikus ág igényeinek kielégítésén fáradoznak". Egészen szembetűnő volt az elhangzott előadások után az a közismert tény, hogy az idézetthez hasonló megállapítások nem helytállóak. Hiszen évtizedek óta volt már strukturalizmus, volt következetesen formális elemzés a nyelv különböző szintjein. Sőt már igen fejlett módszerek, e módszerrel elért jelentős eredmények is voltak, amikor először feltűntek az információk feldolgozására alkalmas gépek s amikor megszületett az ezzel kapcsolatos, speciális-technikai és általános-matematikai elmélet. A nyelvi elemzés gyakorlata s az azt támogató új nyelvelmélet tehát évtizedekkel megelőzte a modern technika konkrét igényeit és lehetőségeit.

De ez egyáltalán nem volt baj, sőt. Nem lehetne megérteni a gépi fordítás terén elért kezdeti kiugró sikereket, ha nem tudnánk, hogy e sikerek mögött több évtized ilyen irányú kutatása rejlik. Természetesen azt, hogy ezek a kutatások „ilyen irányúak" voltak, a legutóbbi évekig maguk a kutatást végző szakemberek sem tudhatták; ők pusztán korunk általános szelleméből kiindulva igyekeztek egyes jelenségek számára egzakt és formális definíciót adni.

Hadd világítsam meg ezt egyetlen példán. A korszerű morfológiai elemzés arra tanít bennünket, hogy azt mondjuk: a magyar parancsoló módot kifejező morféma a {J} morféma, s mi ezt a morfémát véljük felfedezni még az olyan alakokban is, mint *egyél*, *légy*, *hozz* stb.; csupán megjegyezzük, hogy ez utóbbi esetekben a {J} morféma így és így realizálódik. S íme, most kiderült, hogy a gép számára készült nyelvtanba is valahogy így kerülnek be a szabályok: az *egyél*-nél utalást talál a gép arra, hogy az *esz* + *j*, a *légy* = *van* + *j* stb. A gép tehát — de, ismételjük, évtizedekkel később — igazolta a pusztán formális, egzakt eljárásmódot.

S nemcsak az egyes konkrét esetekben, hanem egészen általánosan is igazolta az ilyen módszert. Itt van például az az elméleti alapkövetelmény, amely szerint a grammatika összeállításakor egy jottányit sem szabad engednünk a jelentésnek. S valóban, mit érne a gép az olyan definíciókkal, amely szerint a főnév „az a szófaj, amely ilyen

és ilyen formális kritériumok mellett ezt és ezt — például tárgyfogalmat — f e j e z k i", a mondat, ilyen és ilyen formális kritériumok mellett (melyek olykor megemlíttetnek, olykor nem a hagyományos leíró nyelvtanokban) — „viszonylag kerek gondolatot fejez ki" stb.? A gép, sajnos, képtelen megérteni, hogy mi mit fejez ki. S a modern leíró nyelvészek, szerencsére, ezt nem is sejtve — pontosabban: egyáltalán nem gondolva erre — már évtizedek óta pontosan úgy járnak el, hogy a nyelvi jelenségeket nem aszerint definiálják, osztályozzák stb., hogy „mi mit fejez ki".

Hozzá kell tennem, némileg saját előadásomhoz is kapcsolódva, hogy e gépi igénynyel teljesen egybeesik az elméleti nyelvész igénye: ha a nyelvstatisztika számára vagy a strukturális modellek számára nem tisztán formális anyagot kapunk, „értelmi" momentumokkal csak egy kissé is szennyezett anyagot kapunk, úgy ezzel az anyaggal nem tudunk mit kezdeni, ebben az esetben az egész anyaggyűjtő munkát újból kell kezdeni. Ha tréfálni akarnék, azt mondhatnám: ám tessék levonni a számunkra kedvezőtlen konzekvenciát abból, hogy a mi igényeink teljesen megegyeznek a gépek igényeivel. De ha nem kívánunk élcelődni, akkor beláthatjuk, hogy ez nem speciálisan „gépi" vagy „strukturalista nyelvelméleti" igény külön-külön — az egzaktság és a maximális formalizáltság egyszerűen a korszerűség elengedhetetlen követelménye, méghozzá nemcsak a nyelvtudomány számára. Ahol nincs lehetőség ezen igény kielégítésére, ott kérdéses, hogy — egyáltalán vagy egyelőre — beszélhetünk-e olyan értelemben tudományról, amilyen értelemben a természettudományok azok.

2. Amikor ezt egészen határozottan leszögeztük, ehhez még okvetlenül hozzá kell tennünk valamit, ami ugyan implicite benne van a mondottakban, de nem árt azt kissé ki is fejteni belőlük. Ez a szemantika kérdése. Már többször kifejtettük, hogy a jelentésnek „jottányit sem engedés" korántsem jelenti a szemantika lebecsülését, sőt ellenkezőleg: mi a szemantikával mint külön önálló tudománnyal kívánunk foglalkozni. Persze a szemantikával is a fentebb mondottak értelmében kell eljárnunk, ha tudománynak tartjuk: a lehető legkövetkezetesebben meg kell maradnunk a szemantika saját anyagán belül a szemantikai jelenségek vizsgálata során — ez a „formalizmus" követelménye —, és a lehető legegzaktabb leírásra kell törekednünk. Úgy gondolom, Sipőczy Győző előadása jó példát szolgáltathatott arra, hogyan kell erről az új oldalról megközelítenünk e nagyon régi kérdés-komplexumot. Egyébként Sipőczy előadása abból a szempontból is érdekes, hogy ő mint gyakorló gépi fordító botlott a jelentés kérdésébe, és — helyesen — úgy látta, hogy már e kezdeti fokon sem tudunk a gépi fordítás gyakorlatában tovább haladni, ha a szemantikának legalábbis egyes kérdéseit meg nem oldjuk. Hogyan, kérdezhetné gúnyosan valaki, hát mégis csak kell a gép számára is a jelentés? Bizony, kell. Csakhogy nem abban a formában, hogy a jelentés magától értetődő s így derűre-borúra használható ott is, ahol tulajdonképpen alaki jelenségekről van szó; egzakt formában, amit aztán a gép is meg tud érteni.

Mindenki előtt nyilvánvaló, hogy a jel akkor jel, ha jelent is valamit, hogy a jelrendszernek, definíció szerint, rendelkeznie kell egy denotátum-rendszerrel is. Ezt az elméleti meggondolást ismét szépen alátámasztja, mint látjuk, a gépi gyakorlat. A két jelenségsort — a formáét és a jelentését — összekevernünk egy cseppet sem szabad, a lehető legkövetkezetesebben el kell őket választanunk egymástól, erre már fentebb is utaltunk. De mind a két jelenségsort figyelembe kell vennünk, ez is kétségtelen. Meg kell mondanunk, hogy ha a „hagyományos" nyelvészek, vagy közülük egyesek, joggal voltak vádolhatók azzal, hogy részint keverték definícióik-

ban a két jelenség-sort, részint túlbecsülték a jelentés-oldalt, akkor a strukturalisták közül egy időben többen abba a hibába estek, hogy a jelentéssel egyáltalán nem kívántak foglalkozni. Nos, a gépi gyakorlat nemcsak a „hagyományos" nyelvészeti álláspontot bírálja, hanem a jelentést mellőzni kívánó „modern" álláspontot is.

3. Befejezésül Kónyi Sándor érdekes előadására kívánok egészen röviden reflektálni. Ez az előadás újból felhívta a figyelmünket arra, milyen felmérhetetlenül sok kincs rejlik a mai magyar nyelv rendszerében, mint rendszerben. Éppen napjainkban örömmel lehetünk tanúi annak, hogy új, mélyenszántó és részben összefoglaló munkák jelennek meg a mai magyar nyelv rendszeréről.[1] De nyilván nem lehetünk megelégedve ezekkel az eredményekkel: azon kell fáradoznunk, hogy nyelvünk igen érdekes és tanulságos történetének tanulmányozása mellett egyre jobban megismerjük mai állapotát is; ez mind gyakorlati szempontból lesz hasznos, mind pedig az általános nyelvelmélet számára fog igen érdekes és értékes anyagot szolgáltatni.

[1] A mai magyar nyelv rendszere. Leíró nyelvtan. I—II. (Bp., 1961, 1962). — Antal László: A magyar esetrendszer (Bp., 1961).

Kalmár László

1. Először a gépi fordítás kérdéséhez szeretnék hozzászólni. Nagyon érdekesnek tartom Hell György és Sipőczy Győző ügyes ötleteit, amelyek lehetővé tették, hogy a hazánkban rendelkezésre álló kis teljesítményű elektronikus számológépeken is eredmények jöjjenek létre a gépi fordítás terén. Értékes gyöngyszemek ezek, olyanok, mint amiket a kagyló hoz létre, ha valamely kellemetlen inger éri. Itt a kellemetlen ingert az jelentette, hogy még csak nem is mertünk arra gondolni, hogy valaha nagyobb teljesítményű gépen is végezhetünk ilyen kutatásokat. Márpedig igenis gondoljunk erre, és harcoljunk nagyobb teljesítményű gépekért, különben félő, hogy a gépi fordításra vonatkozó kutatásunk deformálódik, mert azt hisszük, hogy annak csupa ilyen „ügyeskedésből" kell állnia. Persze semmiképpen sem kárba veszett munka az ilyen ügyes ötletek kitalálása, hiszen az ilyen ötletek nagyobb gép fordítási teljesítményét is fokozni tudják.

Kis teljesítményű számológép alkalmas lehet gépi fordítási kísérletekre, avégett, hogy megtanuljuk, hogyan lehet egyáltalában géppel fordítani, de nyilvánvaló, hogy üzemszerű gépi fordításhoz nagy teljesítményű gép kell. Sőt, véleményem szerint helytelen az üzemszerű gépi fordításhoz számológépeket igénybe venni csak azért, mert azok rendelkezésre állnak. Helytelen volna például a szótő-leválasztást üzemszerű gépi fordítás esetén is osztás segítségével végezni csak azért, mert a számológépek tudnak osztani. Az üzemszerű gépi fordításhoz nagy teljesítményű c é l - g é p e t, vagyis külön e célra készült nagy teljesítményű fordítógépet a legcélszerűbb használni. Ennek a gépnek nem kell tudnia például osztani, de a szótövet sokkal direktebb módszerrel le kell tudnia választani és gyors kereső berendezéssel is el kell látva lennie, hogy a szótöveket, a toldalékokat, az idiómákat gyorsan ki tudja keresni a szótárból, a toldaléktárból, illetőleg az idiómatárból. Óriási memóriával is kell rendelkeznie, hogy mindezek beleférjenek. Azonban ennek a memóriának az elektronikus számológépek úgynevezett operatív memóriájához csak annyiban kell hasonlítania, hogy a hozzáférési idejének nagyon rövidnek kell lennie, egyébként permanens memória lehet, amelybe csak egyszer teszünk be rengeteg adatot, de aztán nem változtatjuk azokat. Ebbe a permanens memóriába kerülnek a fordítás során esetleg szükségessé válható eljárásokra vonatkozó programrészek, úgynevezett szubrutinok is, amelyeket persze mindig csak a megfelelő esetben alkalmazunk. Sőt, ugyanide kerül maga a fordító program is, amely a fordítás során minden esetben végrehajtandó eljárások programján kívül arra vonatkozó utasításokat is tartalmaz, mikor melyik szubrutint „hívja fel" a gép. Az operatív memóriának csak akkora kapacitásúnak

kell lennie, hogy a forrásnyelvi szövegnek — amelyet a külső memóriában lehet tárolni — még a leghosszabb mondata is beleférjen, amikor annak a lefordítására kerül sor, emellett a mondat egyes szavaira vonatkozó, a fordításhoz szükséges, a gép által a szótőtárból vagy a toldaléktárból kikeresendő információk is beleférjenek még, továbbá a mondatnak ezen információk alapján történő átalakítása során keletkező „félkész-termékek" is, egészen a „késztermékig", a mondat célnyelvi fordításáig. (Eközben természetesen minden olyan adat, amire már nem lesz szükség a mondat fordításának további lépései során, törölhető az operatív memóriából, s helyére újabb adatokat lehet tenni.)

Azt viszont túlzásnak tartom, hogy néhol olyan óriási permanens memóriára számítanak, amelybe akkora hatalmas — persze szintén gépi úton összeállítandó — szótár is belefér, amelyben minden szóalak benne van. Ez megkerülése a morfológiai kérdéseknek, vagy úgy is mondhatnám, egyoldalú kísérlet a szótározás segítségével való megoldásukra. Még ha meg is engedik majd egyszer a műszaki lehetőségek az ilyen megoldást, akkor sem lesz célszerű alkalmazni. Hiszen az ilyen megoldás nem használja ki azt a könnyítést, amit a morfológiai szabályok lehetővé tesznek, hanem úgy jár el, mintha minden szó rendhagyó volna, vagyis mintha a forrásnyelv szóalakjainak teljesen véletlenszerűen felelnének meg a célnyelv szóalakjai.

A szóhalmazok módszerét nagyon érdekes módszernek tartom. Akkor is helyes módszernek tartanám, ha nem kellene annyira takarékoskodni minden egyes bittel.

2. Nagyon örvendetesnek tartom, hogy Kónyi Sándor alaposan átgondolta és meg is bírálta I. A. Melcsuk magyar—orosz fordítási algoritmusát. Néhány éve találkoztam először Melcsukkal, A. A. Ljapunov professzor hozott össze vele. Melcsuk akkor átadta nekem a magyar—orosz fordítás algoritmusáról szóló dolgozatának részletes, sokszorosított változatát és nagyon kért, vigyem el Magyarországra, és kérjek meg egy, a gépi fordítás kérdése iránt érdeklődő nyelvészt, bírálja meg. Ő ugyanis biztos benne, hogy sok hiba van az algoritmusában, hiszen, amint látom — mondta — törve beszéli a magyar nyelvet, ennnyire is úgy tanult meg magyarul, hogy elolvasta Bárczi Géza két könyvét (A magyar szókincs eredete; Bevezetés a nyelvtudományba). Nagy segítségére lenne, mondta, ha valaki rámutatna algoritmusa gyengéire, hogy kijavíthassa; magyar anyanyelvű nyelvészek segítsége nélkül nem tudja munkáját folytatni. Hazahoztam a dolgozatát, házaltam vele, de sajnos, eleinte a visszautasítás legkülönbözőbb módjaival találkoztam, biztosan azért, mert rossz helyen kopogtattam. Az egyik szegedi nyelvész például hazavitte, megígérte, hogy jól átnézi, de másnap reggel már viszahozta azzal, hogy tökéletesnek tartja, pedig nyilvánvaló, hogy az éjjel nem olvasta el a vaskos dolgozatot. Úgy gondolom, kötelességünk Melcsuknak megadni a kért segítséget, épp ezért örülök annak, hogy Kónyi kartárs most már tényleg megadja.

Kónyi előadásában a magyar forrásnyelv okozta különleges nehézségeket emelte ki; Melcsuk viszont a magyar forrásnyelv előnyeit hangsúlyozta. Kétségtelen, hogy a magyar nyelv morfológiai gazdagsága nagy előny a gépi fordítás szempontjából; a magyar szóra szinte „rá van írva" minden, a fordításhoz szükséges információ. Hasonlítsuk össze a magyar nyelvet az angollal ebből a szempontból: ott még annak a megállapításához is nagy „környezettanulmány" kell, hogy egy-egy szó melyik szófajhoz tartozik.

Kónyi előadása azt is mutatja, hogy a magyar nyelv abból a szempontból is egyedülálló, hogy szinte „mindenre" szolgáltat példát, ennélfogva ellenpéldákat is, amelyek feltárják egy-egy fordítási algoritmus gyengéit s ezzel továbbfejlesztésének útját is. Kónyi példái nagyon szellemesek, látszik, hogy kitűnő ismerője a magyar nyelvnek. Érdemes volna az ilyen példákat, amelyek nemcsak a nyelvészek számára tanulságosak, népszerűsítő folyóiratban, szélesebb olvasóközönség számára is közzétenni. Hogy csak egy példát említsek, nekem arról, hogy a *hajójárat* szó morfológiai elemzése során az utolsó *t* betű felborítja az elemzés addigi útját *(hajó-, hajója-, hajójára)* az jutott eszembe, hogy ha egy csupa 9-es számjegyből álló számhoz 1-et hozzáadunk, ez az utolsó helyértéken álló 1-es „felborítja" a sok 9-est, mindegyikből 0 lesz, és eléjük kerül egy 1-es. Például:

$$\begin{array}{r} 999999999 \\ 1 \\ \hline 1000000000 \end{array}$$

Az elektronikus számológépek természetesen be vannak rendezkedve ilyesmire, így nem okoz nekik nehézséget, hogy ilyen esetben vissza kell fordulniok az elemzés addig követett útjáról és új útra kell térniük.

3. A mikroszótár-rendszer kérdéséről, amelyet Szépe György vetett fel, az a véleményem, hogy a gép számára használható komoly makroszótár összeállítása nagyon sok időt venne igénybe, azért, ha minél gyorsabban be akarunk számolni valamilyen eredményről, nincs más lehetőség, mint viszonylag rövid idő alatt, kevés munkával elkészíthető mikroszótárt használni. Azonban a mikroszótár-rendszer előnyeit elméleti úton bebizonyítani próbálni nem más, mint, ahogy a német mondja, „aus der Not eine Tugend machen". Véleményem szerint a mikroszótár-rendszer csak kísérleti célokra jó, végleges megoldásnak még akkor sem lehet tekinteni, ha kizárólag tudományos szövegek gépi fordítására gondolunk. Korunk tudományos fejlődésére ugyanis az jellemző, hogy a két-két tudományszak határán újabb tudományok alakulnak ki, mint például a biokémia, vagy a geobotanika, vagy — a matematikai nyelvészet. Sőt, ha szabad az *interlingua* szó mintájára így mondanom, ma már „interscientiák" is kezdenek kialakulni több tudományszak határán, olyanok, mint a kibernetika. Ha megnézzük a szakirodalmat, azt látjuk, mind több és több publikáció esik két vagy több tudományszak határterületére. Már pedig ilyen dolgozatok fordításához nem elég egyetlen mikroszótárra támaszkodni, lehetséges, hogy a mikroszótárak összességét kell felhasználnunk, az pedig nem más, mint tárgykörök szerint is felbontott makroszótár. Az ilyen berendezésű makroszótárnak persze lehetnek előnyei, ilyen szempontból a mikroszótárak összeállítása hasznos előkészítő munkának tekinthető.

4. Az interlingua kérdésével kapcsolatban, amely ugyancsak Szépe György előadásában szerepelt, azt hiszem, csak terminológiai eltérés okozhatta azt, hogy a Melcsuk-féle metszet-típusú és az Andrejev-féle egyesítés-típusú interlinguák két ellentétes lehetőségnek tűntek fel. Hiszen az interlingua használatának csak akkor van értelme, ha a forrásnyelvi mondatot, mihelyt lefordítottuk interlinguára, törölhetjük a gép memóriájából. De akkor az interlingua-mondatnak minden olyan információt tartalmaznia kell, amely a szóba jövő célnyelvek valamelyike számára releváns, hiszen

nem nyúlhatnak vissza ilyen információ szerzése végett a forrásnyelvi mondathoz. Ez viszont azt jelenti, hogy az interlinguának az egyes célnyelvek releváns grammatikai viszonya halmaza egyesítését kell tükröznie.

Lehet persze, hogy arról van szó, hogy olyan tökéletes interlinguát, amely minden, az orosz, az angol, a francia, a német, a kínai, a magyar, az indonéz és a többi nyelvek számára releváns viszonyt tükröz, csak nagyon nehezen, nagyon hosszú idő alatt lehetne összeállítani. Melcsuk talán ezért vetette fel azt a gondolatot, hogy először kezdjük a munkát metszet-típusú interlinguával. Ez esetben egy-egy célnyelvi fordításhoz az interlingua-mondaton kívül valamennyi, a forrásnyelvi mondat alapján összeállított indexet is figyelembe kell venni, amely a célnyelv szempontjából releváns grammatikai információt tartalmaz. Már pedig a metszet-típusú interlingua-mondat, hozzávéve minden olyan indexet, amely valamelyik célnyelv számára releváns információt ad meg, a forrásnyelvi mondatnak egy egyesítés-típusú interlinguára fordításának tekinthető.[1]

Ebből már világos, hogy én mindenképpen az Andrejev-féle egyesítés-típusú interlingua híve vagyok, a Melcsuk-féle metszet-típusú interlinguát csak átmenetileg hasznos első közelítésnek tekintem.

Egyébként az ember is valamiféle interlinguán keresztül fordít, még akkor is, ha csak egy nyelvről fordít egy másikra. Ugyanis először megértjük a forrásnyelvi szöveget, tehát lefordítjuk egy „fogalmi nyelvre" és erről, mint valami interlinguáról, fordítjuk le a célnyelvre. Ennek a fogalmi interlinguának a szerepe csak azért nem feltűnő, mert a fogalmi interlingua-szöveg formailag hasonlít a forrásnyelvi szövegre, de lényegében mégis más, mert már megértett szöveg.[2]

5. Papp Ferenc érintette azt a kérdést, lehet-e egyáltalában szerepe a diakróniának a gépi nyelvészetben. Ezzel kapcsolatban felvetek egy ötletet. Érdekes kísérlet volna olyan fordítási programot (vagy legalább részprogramot) összeállítani, amely felhasználja a történeti-összehasonlító nyelvtudomány eredményeit. Ez a program a mai forrásnyelvi szövegnek helyreállítaná a régi alakját, lefordítaná a célnyelv ősére, majd a célnyelvi régi alakokból levezetné a maiakat. A gép memóriájában benne kellene lennie, hogy egy-egy forrásnyelvi szónak mi volt a régi alakja, s hogy a célnyelvben egy-egy változás hogyan ment végbe.[3] Ilyen kísérlet bebizonyítaná, hogy nem képtelenség, hogy a gépi fordítás algoritmusát az összehasonlító-történeti nyelv-

[1] Az interlingua ugyanis nem szükségképpen a hagyományos nyelvekhez hasonló nyelv, mint amilyen az eszperantó, hanem olyan nyelv is lehet, amelynek a mondatai részben grammatikai információt tükröző indexekből állnak.

[2] Itt szeretném megjegyezni, hogy nem értek egyet a Sztálin-féle nyelvtudományi cikknek a gondolkodás és a nyelv viszonyáról szóló nézeteivel, amelyeket egyébként, kissé más formában, már Leibniz is vallott. Minden matematikus tapasztalta, hogy milyen nehéz a már eredményre vezetett matematikai gondolkodás menetét utólag szavakba, képletekbe öntenie, főleg olyankor, amikor közben új fogalmakat is kellett alkotnia, amelyeknek addig nem volt nevük. Ilyenkor csak a matematikai szak- és formulanyelv bővítése, mintegy szakmai nyelvújítás után sikerül ez a szavakba és képletekbe öntés. Nos, hát addig milyen nyelven gondolkodott? Véleményem szerint valamiféle fogalmi nyelven. Ez a fogalmi nyelv szintén anyagi valóság, ha még egyelőre nagyon kevéssé ismerjük is.

[3] Melcsuk lényegében így jár el, amikor a magyar elemző algoritmusában egységesíti a magas és mély hangú toldalékokat. Csakhogy ő a mély hangú alakot használja egységes alakként, holott a nyelvtörténet tanulsága szerint a legtöbb esetben a magas hangú alak volt a régebbi.

tudomány eredményeire építsük. Legfeljebb az derülne ki, hogy így kevésbé célszerű algoritmust kapnánk, mint a szinkronikus nyelvtudomány alapján, mint ahogy az élő fordító sincs, amikor például oroszból magyarra fordít, feltétlenül tudatában annak, hogy egy-egy orosz szó hogyan hangzott az ószlávban, vagy hogy egy-egy magyar képző hogyan érte el mai alakját. Azonban az olyan egyoldalú beállítás ellen, hogy elvileg lehetetlen volna felhasználni a történeti nyelvtudomány eredményeit például a gépi fordításban, határozottan tiltakozom.

Szépe György előadásához még annyit, hogy amikor számba vette, milyen nyelvekről kell majd géppel fordítani magyarra, akkor nem vette figyelembe, hogy egyes nyelveknek, igaz, jelenleg még csekély a jelentősége a mi szempontunkból, de növekvőben van. Ilyen például az arab nyelv, ilyen a kínai is, de gondolnunk kell a gyarmati sorból újonnan felszabadult afrikai országokban beszélt nagyobb nyelvekre is.

6. Ami a nyelvtani szótár kérdését illeti, én nagyon fontosnak tartom Hadas Ferenc kutatásait is, de hasonlóan Gábor Mihály kutatásait is. Hasznos volna, ha az ilyen nyelvtani szótárak nyomtatásban is megjelennének. Amennyiben más nyelvre vonatkozóan is vannak ilyenek, azokat is fel kellene kutatni és kiadni. Természetesen a nyomdára való előkészítés során ki kell javítani e szótárak fogyatékosságait, evégett előzőleg alaposan meg kell azokat bírálni. Például az olyanféle példakifejezéseket, mint a Hadas-féle szótár 45. oldalán (—айш — ими címszó alatt) „/a/ lecke, amelyet a *legszigorú-bb* tanítók készítettek," jobb példákkal kell pótolni. (Vajon hol készítették, talán pedagógus továbbképző tanfolyamon? Lehet, hogy az orosz példakifejezés, *урок, сделанный строжайшими учителями*, jó; csak úgy kellene magyarra fordítani, hogy „/az az/ óra, amelyet a legszigorúbb tanítók tartottak".) Az sem világos nekem, mire jó a tővégződést leválasztani olyankor, amikor nincs semmilyen betoldás. Ha pusztán azért, hogy legyen milyen címszó alá tenni a példát, akkor jobb volna ilyen esetben „üres címszót" használni.

7. Meg voltam döbbenve, milyen jó kivonatot szolgáltatott Szalai Sándor algoritmusa előadásom kivonatáról, holott alig igényel memóriakapacitást (mindössze azt a néhány szót kell a gépnek megjegyeznie, amelyeket bizonyos esetekben figyelmen kívül kell hagynia, például azokat, amelyeket nem szabad „szomszédságot" elválasztó szavaknak tekintenie). Ha azt a feladatot kaptam volna, húzzam alá előadáskivonatom mondatai közül azt a négyet, amelyet a legfontosabbnak tartok, meg vagyok róla győződve, hogy ugyanezt a négy mondatot húztam volna alá. Azonban, mint erre már Szalai is rámutatott, ez részben annak tudható be, hogy stílusom nem eléggé változatos. Így például a Szalai-féle algoritmus által kiválasztott négy mondat közül az utolsónak nagymértékben növelte az indexét az a körülmény, hogy háromszor is használom a *vizsgál* szó különböző bokrozatait („Az előadó Carnap, Bar-Hillel, Lambek és Kulagina *vizsgálatainak* rövid ismertetésével kívánja illusztrálni azt a tézist, hogy a matematikus által a maga *vizsgálatai* körébe tartozó struktúrák tanulmányozására kifejlesztett halmazelméleti módszerek a nyelvi struktúrák *vizsgálatára* is felhasználhatók"). Ha a második helyen *vizsgálatai* helyett azt írtam volna, hogy *kutatásai*, a harmadik helyen pedig ahelyett, hogy *vizsgálatára*, azt írtam volna, hogy *tanulmányozására*, akkor a megfelelő szavak súlya 3,1 százalékról leszállt volna 1 százalékra, és így e mondat indexe lényegesen csökkent volna, annak ellenére, hogy valóban nagyon lényeges mondanivalókat tartalmaz. (Persze azért a *struktúra*, *halmaz*, *nyelv* szavak bokrozatai is lényegesen emelték e mondat indexét.) Az első mondat

viszont, ahol — úgy látszik — még vigyáztam a stílusra (a *vizsgálni* szóalak mellett a *tanulmányozásához* és a *kutatni* szóalakokat használtam), ennek ellenére belekerült a Szalai-féle kivonatba, mert ott valóban tömören szó van mindarról, amiről az egész (kivonatolt) kivonatban.

8. Érdekesnek tartom, hogy Rényi Alfréd aszerint határolja el a speciális (matematikai) információelméletet az általános információelmélettől, hogy az előbbit csak szintaktikus kérdések érdeklik, míg az utóbbi az információ tartalmával is foglalkozik. Ebből azonban az is kiderül, hogy az információelméletnek a matematika számára releváns része is szélesebb annál, hogy pusztán a szintaktikus kérdések területére volna korlátozható, hiszen például a matematikai formulák szolgáltatta információ tartalma is érdekli a matematikust. Ezért helyesebb volna a speciális információelméletet szintaktikus, kvantitatív, vagy még inkább statisztikai információelméletnek nevezni és a matematikai információelmélethez hozzászámítani azt a még ki nem dolgozott kvalitatív információelméletet is, amely matematikai formulanyelvek esetén, ahol az információ tartalma is matematikai, igyekszik az információmennyiség vizsgálatán kívül megragadni azt a pluszt is, amelyet a tartalom figyelembevétele szolgáltat. Erre nézve vannak már heurisztikus gondolataim, meg fogom próbálni azokat kidolgozni.[4] Lehetségesnek tartom, hogy ezen az úton olyan matematikai modellt találunk, amely pl. természetes nyelven kifejezett, nem matematikai tartalmú információ esetén is alkalmazható, s így előre jutunk majd az általános információelmélet kidolgozása felé is.

Rényi hangsúlyozta, hogy a redundancia nem mindig káros „terjengősség", hanem hasznos is lehet. Ehhez néhány illusztráció. A tízes (és hasonlóan a kettes és a többi) számrendszerben gyakorlatilag semmi redundancia nincs. (Az egyetlen lehetséges redundancia abból ered, hogy 0 számjegyeket lehet a szám elé írni, de a számológépeknél még ez sem szerepel, hiszen ha rövidebb számról van szó, mint amekkora a rekeszbe fér, akkor automatikusan a megfelelő számú 0-t képzeli elé a gép.) Mármost épp emiatt sokkal nehezebb telefonkönyvet, vagy menetrendet korrigálni, mint egy betűkből, szavakból álló szöveget: a sajtóhiba nem szökik az ember szemébe hiszen a számoknak hibásan is van értelmük. Hasonlóan, ha csak „approximatíve', tudok egy telefonszámot, mint az ismert viccben, nem érek vele semmit, nem tudom' mit tárcsázzak.

Jól tudják az elektronikus számológépek konstruktőrei, hogy a redundancia hiánya a kettes számrendszerben, bár gazdaságos, de hiba forrása is lehet, különösen akkor, ha sok számadatot viszünk be a gépbe, vagy pl. a külső memóriából viszünk át sok számot a belsőbe. Sok szám esetén ugyanis megnő annak a valószínűsége, hogy valamelyikük hibásan kerül be a gép memóriájába. Épp ezért ilyenkor a közönséges kettes számrendszer helyett némely gépben olyan jelkulcsrendszert használnak, amelybe már be van építve hasznos redundancia, oly módon, hogy a gép automatikusan észreveszi, ha valamely szám hibásan megy be a memóriájába (úgynevezett hibajelző kód). Sőt, minthogy a hiba puszta észrevétele magában véve keveset ér, olyan jelkulcsrendszert is használunk, amely még nagyobb beépített redundanciá-

[4] Lásd Kalmár László, A kvalitatív információelmélet problémái: A Magyar Tudományos Akadémia Matematikai és Fizikai Tudományok Osztálya Közleményei 12. 293—301 (1962).

jánál fogva lehetővé teszi, hogy a gép automatikusan ki is javítsa az észrevett hibát (úgynevezett hibajavító kód).

Azt, hogy a nyelvi információ mennyiségét a nyelvtan is növeli, Wu Jün-c'eng kínai matematikus egy érdekes megjegyzésével szeretném illusztrálni. Azt mondta egyszer, hogy amikor ő franciául beszél, úgy érzi, hogy a felhasznált információmennyiségnek körülbelül csak a fele áll rendelkezésére arra, hogy mondanivalóját kifejezze, a másik felét arra kell használnia, hogy állandóan azt mondja, hogy „kérem, én jól tudok franciául; kérem, én jól tudok franciául; kérem, én jól tudok franciául", mert a francia nyelvtan szabályainak betartása csak erről ad információt. Nem véletlen, hogy kínai mondta ezt, aki szinte nyelvtan nélkül fejezi ki anyanyelvén a maga mondanivalóját, és az sem véletlen, hogy a francia nyelvről mondta, amelynek „feleslegesen bonyolultnak" tűnik a nyelvtana egy kínai számára.

A kínai nyelv, főleg a klasszikus kínai irodalom nyelve azt is jól illusztrálja, amit Rényi a célzások információaktivizáló szerepéről mondott. Egyetlen hieroglifa gyakran egy egész történelmi eseményt idéz fel a művelt kínai tudatában, és ezzel a lehetőséggel bőségesen él is a kínai irodalmi nyelv. Épp ezért a kínai irodalmat csak alapos történelmi tudással lehet megérteni. Hasonló módon alkalmazzák a célzás, az utalás eszközét a diákok között kifejlődött „kollégiumi tolvajnyelvek is": például az egyik Eötvös-kollégista a másiknak egyetlen szavából hosszú információsort megért.

Ez a jelenség egyébként az információ különleges természetével függ össze. Információ segítségével hatalmas tömegeket lehet megmozgatni, energiákat lehet aktivizálni. Ha telefonon tüzet jelentek, elenyésző csekély energia fut a dróton; nem ez, hanem az adott információ mozgatja meg a tűzoltókat, hogy a helyszínre robogjanak. Így előzőleg tárolt nagy információmennyiséget is meg lehet kisebbel „mozgatni". Ezt a lehetőséget felhasználjuk az elektronikus számlálógépekben is, amikor például egy, a kettes számrendszerben 40 jegyű számot azáltal viszünk át a memóriából az aritmetikai egységbe, hogy annak a rekesznek a címét adjuk meg, ahol előzőleg tároltuk a számot, s ez a cím esetleg csak 10 jegyű szám a kettes számrendszerben. Ilyen jelenség az idézés is: ha azt akarom, hogy dolgozatom olvasója elolvassa egy másik dolgozat egész információtartalmát, elég, ha az utóbbinak a bibliográfiai adatait adom meg, ami jóval kevesebb információmennyiséget igényel. És ilyen jelenség a célzás is. Azáltal, hogy az ember hozzátartozik egy bizonyos kulturális közösséghez, egy sereg információ van tárolva a memóriájában. A költő egyetlen szava elég ahhoz, hogy ennek az információnak valamely hatalmas részét aktivizálja. Itt tehát olyan jelenségről van szó, amely a számológépektől egészen az irodalomig végigvonul.

9. A vita során Kelemen József érdekes módon foglalta rendszerbe a nyelvtan különböző szintjeit. Bár ez a rendszer nem matematikai modell, irányt mutathat Papp Ferencnek is, hogy miképpen induljon el matematikai mikromodelljeinek makromodellé való szintetizálása felé.

Azonban Kelemen félreértette azt, amit én tegnapi hozzászólásomban mondtam a diakrónia tükröződéséről szinkronikus vetületében. Én nem mondtam azt, hogy a diakrónia felesleges, mert eredményeit szinkronikus módszerekkel is el lehet érni, hanem azt mondtam, hogy nem helyes a kettőt élesen szembeállítani egymással, nem helyes a kettő között mesterségesen konstruálni ellentmondást. Hogy a kettő közötti ellentét mesterségesen konstruált dolog, azt mutatja többek között az is, hogy a diakrónia és a szinkrónia között olyan kapcsolat van, amely lehetővé teszi a

diakrónia vetületének leolvasását a szinkróniáról s ez a leolvasás olyan módszer, amely a természettudományoktól a nyelvészetig egyaránt használható.

Kelemen arról is beszélt, hogy szinkronikus nézőpontból zártnak látszó rendszerek, mint például a járulék-morfémák összessége, diakronikusan szemlélve nyitottnak mutatkoznak. Mit ért azonban Kelemen nyitott rendszeren? Ha jól értettem, olyat, amely függ az időtől, s így még „nyitott" kérdés, milyen lesz a jövőben. Így azonban a matematikusnak nem sokat mond az, hogy amit diakronikusan szemlélünk, arról kiderül, hogy függ az időtől, hiszen diakronikusan szemlélni valamit épp ezt jelenti, mint az idő függvényének tekinteni. Azt hiszem, ez a fogalmazás: amit az idő függvényének tekintek, az függ az időtől, eléggé mutatja, hogy majdnem pleonazmus, amiről szó van.

10. Amit Deme László mondott a vita során a matematikai nyelvészetről és a nyelvészeti matematikáról, az szellemes szójáték. Én azonban másképp látom a helyzetet. Természetesen a nyelvstatisztika nem az egyetlen teendője a matematikai nyelvészetnek, de azt is a matematikai nyelvészethez számítanám. Mint tegnapi előadásomban mondtam, a matematikai nyelvészetnek különböző ágai vannak, amelyekkel különböző mértékben, de már hazánkban is foglalkoznak, amint ez egyébként örvendetesen megmutatkozott ezen a munkaértekezleten is. Azt is hozzátettem azonban, hogy ezek az ágak egymással szorosan összefonódnak. Valójában éppen olyan szerves kapcsolat van ezek között az ágak között, mint egy élő fa ágai között. Például a nyelvstatisztikai vizsgálatok eredményei nagyon fontosak a gépi fordítás számára is. A fordítási algoritmusnak többek között azt is meg kell határoznia, hogy egy-egy szó többféle jelentése esetén ezeket milyen sorrendben vizsgálja végig a gép. Nem mindig célszerű a gép munkája szempontjából az a sorrend, amely az értelmező, vagy a nyelvtörténeti szótárakban található. Hiszen a fordítás szempontjából irreleváns például az a kérdés, melyik volt a kérdéses szónak a legrégibb jelentése, amelyből jelentésváltozás útján a többi jelentése (vagy azok egy része) létrejött, akármennyire fontos is egyéb szempontból a különböző jelentések egymásból való keletkezésének kérdése a nyelvtudomány számára. Hanem az a célszerű a gépi fordítás szempontjából, ha a gép először arra a jelentésre „gyanakszik", amelyet az a szó a mai nyelvben a leggyakrabban szokott jelenteni. Ha aztán kiderül, hogy ez a jelentés valamilyen okból nem illik a mondatba, akkor a második leggyakoribb, legvalószínűbb jelentést kell a gépnek megvizsgálnia, ha az sem fogadható el, a harmadikat, és így tovább. Márpedig azt, hogy milyen gyakran fordulnak elő a mai nyelvben (vagy, ha a gépi fordítás feladatát tudományos értekezések fordítására korlátozzuk, a mai értekező prózai nyelvben) egy-egy szó különböző jelentései, a nyelvstatisztikai vizsgálatok hivatottak megállapítani. Ez csak egy példa arra, hogy a matematikai nyelvészetnek látszólag egymástól távoli ágai is milyen szorosan összefüggenek egymással, s hogy a nyelvstatisztika egyes eredményei hogyan segíthetik a matematikai nyelvészet többi ágain végzett kutatásokat.

A nyelvstatisztikával kapcsolatban azonban óvni szeretném az e téren működő nyelvész kartársakat attól, hogy öncélúan végezzék ezeket a vizsgálatokat. És főleg attól, hogy élvezetnek tekintsék a nyelvstatisztikai munkának azt a részét, ami nem embernek való. Akármennyire élvezi is valaki a számlálgatást, az bizony unalmas, gépies munka. Arra kell törekednünk, hogy az ilyen gépies munkát minél előbb gépek végezzék helyettünk. Engedjük már, hogy a gépek felszabadítsanak bennünket a gépies munkától, hogy igazi alkotó munkára fordíthassuk időnket és

munkaképességünket, olyanra, amelyet a gépek — ha úgy tetszik valakinek, egyelőre még, jóideig — nem tudnak helyettünk elvégezni. (Ha pedig majd valamikor a jövőben sikerül a gépeket valamilyen, ma még alkotó munkának tekintett munkára is „megtanítani", akkor már azt is jogosan tekinthetjük gépies munkának; s addigra majd akad, mindig akad majd olyan még magasabb rendű alkotó munka a nyelvtudós számára, amit a gépek akkor sem tudnak elvégezni helyette.)

Deme László felvetette azt a kérdést is, vajon át lehet-e gépi programba ültetni azt a bonyolult szórendi formulát, amelyet felírt. Igaz, a gépek nagyon „buták", nem értik meg az ilyen formulát, még a sokkal egyszerűbb formulákat sem, egyáltalában, semmit sem értenek meg, csak pontosan követik a program utasításait. A programozásban azonban óriási lehetőségek vannak. Amit a nyelvész pontosan, egyértelműen meg tud fogalmazni, amiről szabatosan meg tudja mondani a programozó számára, hogyan kell elvégezni, azt a programozó több-kevesebb fáradsággal át tudja fogalmazni a gépnek szóló utasítások sorozatává. Ne féljünk tehát az ilyen bonyolult szórendi kérdésektől sem. Például a felírt formulát — ha csak nincs benne valami rejtett mozzanat, amely elkerülte a figyelmemet — könnyűszerrel át lehet ültetni gépi programba.

11. Még azt szeretném kiemelni Papp Ferencnek a vita során elhangzott hozzászólásából, hogy, amint mondta, a gépek igazolták a sokkal korábban elindult strukturalizmust. Hasonló gondolatot fogalmazott meg, még sokkal élesebb formában A. N. Kolmogorov szovjet matematikus a matematikai gépek alkalmazásáról tartott moszkvai össz-szövetségi konferencián. Kolmogorov annak a mély meggyőződésének adott kifejezést, hogy a társadalmi jelenségeknek, így a nyelvnek is, éppúgy megvannak a maguk objektív törvényszerűségei, mint a természet jelenségeinek, és ezeket a törvényszerűségeket tudományos kutatással éppúgy fel is lehet deríteni és szabatosan megfogalmazni. Csak abban kételkedik, hogy a társadalomtudományok szakemberei által eddig megfogalmazott kijelentések valóban mindig objektív törvényszerűségeket tükröznek. A társadalomtudományok területén sokkal ritkábban van mód arra, hogy a gyakorlat kritériumával ellenőrizzük, mely kijelentések tükröznek objektív törvényszerűségeket, melyek nem, mint a természettudományok területén; ezért található sokkal több „mellébeszélés" a társadalomtudományok szakirodalmában az értékes tudományos megállapítások mellett, mint a természettudományos dolgozatokban. Nos, mondja Kolmogorov, a matematikai gépek nyelvtudományi alkalmazásaival jelentősen kiszélesedett az a terület, amelyen a gyakorlat kritériumát alkalmazhatjuk az értékes tudományos megállapítások megkülönböztetésére a mellébeszéléstől. Hic Rhodos, hic salta! Tessék a gépnek megmondani, hogy a nyelvtudomány eddigi megállapításai szerint hogyan kell egyik nyelvről a másikra fordítani. Ha a gép erre helyesen fordít, akkor azok a megállapítások valóban objektív törvényszerűségeket tükröznek; de ha nem, akkor bizony mellébeszélés is volt bennük, amit meg kell keresni, és amitől meg kell tisztítani a nyelvtudományt.

Kolmogorov e gondolataival teljes mértékben egyetértek. Azonban szeretnék a kérdés másik oldalára is rámutatni. Helytelen volna azt kívánni, hogy a nyelvtudomány ezentúl minden kérdést kizárólag a gépek szempontjából vizsgáljon. A nyelvtudós nemcsak a gépek, hanem elsősorban az ember számára dolgozik. A fordításelmélet sem merülhet ki a gépi fordítás elméletével. Még ha egyszer majd a tudományos dolgozatok fordítását kizárólag gépek fogják is végezni, akkor is marad még

az ember számára is fordítanivaló. A műfordítást még nagyon sokáig emberek fogják végezni; egyesek véleménye szerint az idők végezetéig. Márpedig az embernek túlságosan unalmas, gépies munka volna a gépek számára összeállított algoritmusokon átmenni, amikor az egyik nyelvről a másikra fordít. És az ember tökéletesebben tud fordítani, mint a gép, ha az ember számára értelmesen mondják meg neki, hogyan fordítson. Kell maradnia hát a fordításelmélet olyan részének is, amely nem a gépek, hanem az ember számára készül, és általában, a nyelvtudománynak mindig maradni fog a gépek számára készült gépi nyelvészeten kívül az emberhez szóló része is. A kettőt nem szabad összekeverni: éppoly helytelen, ha a géppel azt próbáljuk közölni, ami az ember számára készült, mint ha az embert arra próbáljuk kényszeríteni, ami a gépnek való. Ugyanakkor azonban szoros kapcsolat van a kettő között: ha valamit egyszer a gép számára szabatosan megfogalmazunk, azt az ember számára is szabatosabban tudjuk megfogalmazni. Hiszen az emberhez szóló nyelvtudomány végelemzésben a nyelvi jelenségek ugyanazon objektív törvényszerűségein alapul, mint a gépi nyelvészet, és ha valami a gépi kipróbálás során egyszer mellébeszélésnek bizonyult, abból az ember számára sem lesz soha értékes megállapítás. Szóval, bár a nyelvtudománynak nem lehet az az egyetlen célja, hogy a gépek minél jobb kiszolgálására törekedjék, hiszen a gépekkel való foglalkozáson kívül természetesen nagyon fontos társadalmi funkciója marad az emberrel szemben is, a gépek mégis jó tesztet szolgáltatnak a nyelvtudomány számára, amelynek helyes felhasználásával megtisztíthatja értékes megállapításait a mellébeszélés maradványaitól. És általában, csak hasznára lehet az egész nyelvtudománynak, ha a matematikai módszerekkel együtt a nagyobb szabatosságra való igény is „betör" a nyelvtudománynak — egyelőre — egyes területeire. Meg vagyok arról győződve, hogy ez az igény fokozatosan át fogja majd hatni a társadalomtudományok egészét, s hogy valamennyien gazdagodni fognak általa, mert szabatosabb módszerekkel fel tudják majd tárni a társadalmi jelenségeket mozgató objektív törvényszerűségeket. Büszkék lehetnek a nyelvtudósok, hogy ez a folyamat — a gazdaságtudomány mellett — az ő tudományszakjuk területén indult meg, és hogy ebben ők mutathatnak példát a többi társadalomtudományok szakembereinek!

*Mikó Pálné**

Szépe György referátumában nem szólt a franciaországi gépi fordítási munkálatokról. Ezt a hiányosságot szeretném — részben személyes tapasztalataim alapján — némileg pótolni.

1. Franciaországban később kezdtek foglalkozni a gépi fordítással, mint az Egyesült Államokban és a Szovjetunióban. Ebben a rendkívül gyorsan fejlődő kutatási ágban azonban nem feltétlenül jelent behozhatatlan hátrányt a „késés", hiszen a kutatók bizonyos mértékig kedvezőbb helyzetbe kerülhetnek azáltal, hogy másutt már ellenőrzött eredmények birtokában, azokat saját szempontjaikkal összevetve haladhatnak előre.

A. Sestier tanulmánya[1] részletesen ismerteti a franciaországi gépi fordítói munkálatok történetének korszakait. Ezek a következők:

a) 1958—60. Megalakul a kis létszámú kutatócsoport. Matematikai modelleket alkalmaznak a német, francia és angol szintaxis vizsgálatában — egyelőre automatikus analízis nélkül.

b) Kutatócsoportok jönnek létre Párizsban (Centre d'Études pour la Traduction Automatique — section de Paris: CETAP), illetőleg Grenoble-ban (Centre d'Études pour la Traduction Automatique — section de Grenoble: CETAG). Mindkettő a franciaországi kutatásokat irányító, összehangoló és finanszírozó Centre National de la Recherche Scientifique (CNRS) keretében működik. Nancyban megalakul egy csoport (Centre de Recherches de Linguistique Appliquée et de Traduction Automatique de Nancy: CLATAN). — A csoportok 1960-ban megkezdik az orosz és a japán analízist is. Munkájukban IBM 650-es és BULL gyártmányú, nagy teljesítményű elektronikus számológépeket használnak fel.

c) 1960 végefelé szorosabbá válik a nyelvészek és a matematikusok együttműködése. A kutatók egyre inkább felismerik az automatikus segédeszközök nélkülözhetetlenségét. A tematikában előtérbe kerül a morfológiai rendszer vizsgálata, illetőleg az alapszótár összeállításának a kérdése.

d) 1961-ben elkészül az orosz nyelv morfológiai leírásának rendszere — természetesen elsősorban a franciára való fordítás szempontjából. Ezzel kapesolatban különböző analízis- és szintézisprogramokat és algoritmusokat állítanak össze.

* Írásban benyújtott hozzászólás.

[1] A. Sestier, La traduction automatique des langues: mythes et réalités: Études préliminaires à la Traduction Automatique (Párizs, 1961, CNRS—CETAP).

2. A gépi fordítás francia kutatói erősen kapcsolódnak a modern európai nyelvészet legújabb hagyományaihoz is. Figyelembe veszik a prágai és a koppenhágai strukturális irányzatok eredményeit: honfitársaik közül pedig elsősorban Tesnière, Martinet és Quémada munkásságát. A matematikai és logikai nyelvészet képviselői közül elsősorban Bar-Hillel, illetőleg az amerikai Yngve tanulmányaira támaszkodnak. Szoros kapcsolatban állnak Ceccatónak a milánói egyetemen működő úgynevezett „operatív" [tudniillik operatív filozófián alapuló] csoportjával (Centro di Cibernetica e di Attività Linguistiche). Emellett igen élénk érdeklődéssel követik a szovjet matematikai nyelvész és gépifordító szakemberek eredményeit.

A gépi fordítás francia szakemberei úgy vélik, hogy ezek a modern, illetőleg egzakt nyelvészeti kutatások a gépi fordításnak elengedhetetlen alapjai még akkor is, ha nem kizárólagosan a gépi fordítás előmunkálatainak szánták őket. Ugyanilyen fontosnak tartják azonban azt is, hogy a gépi fordítással kapcsolatos kutatások visszahatnak a nyelvészeti vizsgálatokra: az eddigieknél szigorúbb, pontosabb, félreérthetetlen formális meghatározásokra vezetnek.

3. Az egyre nagyobb intenzitással működő kutatóközpontok elveikről és gyakorlatukról tájékoztatókat és (előzetes) tanulmányokat adnak ki. Ezek során például beszámolnak az orosz morfológiai analízissel és a szintaktikai elemzés automatizálásával kapcsolatos munkásságukról,[2] illetőleg az orosz főnév, ige, melléknév, névmás részletes morfológiai elemzéséről.[3] A „Linguistique Mathématique et Traduction Automatique" című sorozatban megjelenő tanulmányokon kívül kiadásra kerülnek a nyelvészeti analízis és formalizálás módszereit bemutató írások, valamint kidolgozott fordítóprogramok is.

A gépi fordítás és az alkalmazott nyelvészet tanulmányozására és fejlesztésére alakult meg Párizsban 1960-ban az ATALA (Association pour l'Étude et le Développement de la Traduction Automatique et de la Linguistique Appliquée). Céljai között szerepel a francia és a külföldi csoportok munkája összehangolásának előmozdítása és a kölcsönös tájékoztatás is. A társaság rendszeres ülések formájában hozza egymáshoz közel a különböző csoportok kutatóit. 1960-tól kezdve a társaság két havonként megjelenő folyóirata (La Traduction Automatique) a franciaországi és egyre inkább az európai gépi fordítás központi orgánumává válik. Tanulmányok mellett a folyóirat rendszeres és széleskörű bibliográfiai és szemlerovatot és „La TA dans le monde" [A gépi fordítás a világban] című értékes tájékoztató rovatot is közöl. Az „Automatisme" című — elsősorban az ipar és a kereskedelem igényeit kielégíteni kívánó folyóirat — is közöl összefoglaló jellegű cikkeket a gépi fordítás újabb eredményeiről.

A franciaországi gépi fordítással foglalkozó vázlatos ismertetésünkben utoljára de nem utolsósorban E. Delavenaynek, az említett társaság elnökének gépi fordítási bibliográfiájára[4] hívjuk fel a figyelmet, amely a maga műfajában az első és még ma is nélkülözhetetlen.

[2] A. Sestier: L'automatisation de l'analyse syntaxique (Párizs, 1961); L. Dupuis: Un système d'analyse morphologique de la langue russe (Párizs, 1961).

[3] B. Vauquois—J. Veyrons: Présentation de l'analyse morphologigue du russe (Grenoble, 1962); D. Andron: Analyse morphologique du substantif russe (Grenoble, 1962); I. Torre: Analyse morphologique de l'adjectif russe (Grenoble, 1962); A. Gagny: Analyse morphologique du verbe russe (Grenoble, 1962).

[4] E. és K. Delavenay: Bibliographie de la traduction automatique (Hága, 1960) 69 lap.

4. Összefoglalásul elmondhatjuk, hogy a franciaországi kutatók tevékenységét ugyanúgy ismernünk kell, mint a gépi fordítás terén fejlettebb — máris konkrét eredményeket felmutató — országokét. Franciaországban még nem végeztek ugyan nagyobb demonstratív gépi kísérletet (erre ma már nincs is szükség) és nem végeznek rendszeres fordítást sem. Elveik és módszereik, gyakorlati előmunkálataik azonban a magyar szakemberek számára is hasznos útmutatást jelenthetnek különösen az orosz analízis és a — valószínűleg hamarosan napvilágot látó — francia szintézis vonatkozásában. Ez utóbbi különösen akkor lehet számunkra fontos, ha időszerűvé válik a franciáról magyarra történő gépi fordítás is.

Szalai Sándor

1. Előadásom terjedelmi korlátai miatt sajnos nem térhettem ki minden kérdésre, amire szerettem volna — így többek között az általam bemutatott eljárás lehetséges hibaforrásainak s ezzel kapcsolatos finomítási szükségleteinek kérdésére sem. Péter Rózsa hozzászólására válaszolva, teljes igazat kell adnom neki abban, hogy az *és*, illetőleg a *vagy* egy matematikai-logikai munkában nem triviális, hanem igen fontos, sőt szöveg-specifikus szóként is szerepelhet. Így verbális képletekben a konjunkciót, illetőleg a diszjunkciót fogja jelezni, sőt még az is lehetséges, hogy az egész matematikai-logikai tanulmány éppen az *és* és a *vagy* különböző nyelvi jelentéseinek szimbolikus elemzéséről szól. Egyáltalán: n i n c s olyan szó, amely adott szöveg-összefüggésben ne válhatna relevánssá, bármilyen triviális is általában a köznyelvi használatban. De először is az ilyesmi viszonylag ritka s általában a szakszövegek bizonyos típusaihoz vagy csoportjaihoz kötött jelenség, úgyhogy adott esetben megfelelő külön program-utasításokkal (aránylag kis memória-igénybevétellel) az ebből eredő hibák száma csökkenthető. Másodszor a kivonatolásnál használt mondatértékelésünk nem egy-két, hanem sokkal több szó előfordulási gyakoriságának és előfordulási helyeinek figyelembevételén alapszik. Harmadszor a különleges és releváns jelentésben használt triviális szókat a szakszövegekben többnyire már a közönséges olvasásnál is felmerülő félreértési lehetőségek miatt is meg szokták tipográfiailag különböztetni. (A nyelvész például kurziválja az *és*-t és a *vagy*-ot, ha nem triviálisan használja, hanem valamely nyelvészeti vizsgálat tárgyaként emlegeti e két kötőszót; a matematikai logikus külön sorba szokta írni verbális képleteit stb.) Az ilyesmi is „beprogramozható". Negyedszer és utoljára: efféle hibaforrások a gépi kivonatolásnál — mint egyébként az embernél is — végső soron sohasem küszöbölhetők ki teljesen, hanem csak korlátozhatók. Programjaink fejlesztésénél korlátozásukra kell törekednünk — addig a mértékig, ameddig az általában véve mindig külön program-utasításokat igénylő korlátozások „kifizetődők", azaz nem jelentenek túl nagy megterhelést a hiba valószínűségéhez képest.

2. Péter Rózsa hozzászólása kapcsán rámutathatok még arra is, hogy bizonyos t r i v i á l i s szóknak az átlagosnál l é n y e g e s e n g y a k o r i b b előfordulása — stilisztikai vizsgálatot érdemlő módon — bizonyos típusú írásművek stiláris vagy tematikus sajátossága lehet. Például Arany Jánosnak „Vojtina ars poetica"-jában a *mint* szó előfordulási gyakorisága váratlanul magas: a szokásos 0,3 százalék helyett 2,1 százalék, amivel itt a *mint* utoléri az *az* névelőt. Ez pedig azért van így, mert

Arany János költő és gyakran használ hasonlatokat. Hasonló jelenség az *után* szó feldúsulása Verzár Frigyes farmakológiai tanulmányában, amelynek gépi kivonatát előadásomban bemutattam. A Nemes-féle szóstatisztika szerint az *után* nem tartozik a különlegesen gyakori, nagyon triviális szók közé, bár persze azért korántsem szövegspecifikus, hiszen mindenféle magyar szövegben imitt-amott szükségképpen előfordul. Verzár Frigyes tanulmánya azonban gyógyszerhatásokat, így többek között gyógyszerek beadása és bizonyos tünetek fellépése vagy megszűnte közötti időbeli összefüggéseket tárgyal. Tehát gyakoriak benne az olyan megállapítások, hogy ez vagy az a szer hány perc *után* fejti ki hatását, mi *után* lép fel ez vagy az a tünet stb. Az általam bemutatott eljárásnak talán éppen az az egyik nyelvészeti érdekessége, hogy rávilágít az általános nyelvhasználatnak, az egyes témakörökhöz tartozó nyelvhasználatnak, sőt adott esetben az egyéni nyelvhasználatnak bizonyos statisztikai jellemzőire.

3. Minden eljáráshoz, amely valószínűségek alapján dönt el bizonyos kérdéseket, lehet ellenpéldákat keresni és találni, amelyeknél a döntés vagy megoldás hibás lesz, mert a döntés ebben a keretben a valószínűbb (sokkal valószínűbb) eset előnyben részesítését jelenti a kevésbé valószínűvel (esetleg egészen ritkával) szemben, amely azonban — természetesen — mégis bekövetkezhet. A tudományos szövegek gépi kivonatolásának kérdését sem lehet nyelvi valószínűségek alapján úgy megoldani, hogy ne akadjon ellenpélda, ami hibára vezet. Csak az a kérdés, hogy mennyi lesz az ellenpéldák száma! Ugyanez vonatkozik a gépi fordítási programokra is. Bizonyos ésszerű határon túl egyszerűen nem érdemes egészen kivételes hibalehetőségek kiküszöbölése végett túlságosan elmélyíteni a szintaktikus és szemantikus elemzést, túlságosan bonyolítani a programot. Nem érdemes például külön mechanizmust (algoritmust) keresni olyan finomságok megkülönböztetésére, mint amilyenek — mondjuk — egy XIII. századi történeti szöveg fordításánál az „Észtek *lettek lettek*" értelmezésében adódhatnak. Ez jelentheti azt, hogy egyes észtek lettekké váltak, de jelentheti esetleg azt is, hogy észtekké váltak egyes lettek. Más-más értelmi akcentust is kap a mondat aszerint is, hogy a két egymás után előforduló *lettek* szó közül melyik a többes számú főnév (népnév) és melyik a múlt idejű, többes számú és 3. személyű ige. De érdemes-e ilyesminek az eldöntésére berendezkedni, azazhogy gépet programozni? Nyilván nem, bár hangsúlyoznom kell, hogy a szövegkörnyezet gondos számbavételével alighanem még itt is sikeresen elvégezhető a döntés alapjául szolgáló elemzés, csakhogy az ilyesmi aligha kifizetődő. Ne felejtsük el, még olyan kiváló műfordítóval, mint Kosztolányival is előfordult, hogy a francia *cafard* szót olyan összefüggésben, amelyben ez „Blaumontag"-ot, azaz „másnaposságot" jelentett (argot szóhasználat szerint), szótári szabatossággal „svábbogár"-nak fordította. Nyugodjunk bele: nemcsak az ember, hanem a gép is tévedhet . . .

4. De lehet a gépnek is szerencséje, akár az embernek! Köszönettel kell nyugtáznom Kalmár László elismerését arra vonatkozólag, hogy ő is azt a négy mondatot tartja előadása maga készítette sillabuszában a legfontosabbnak, amelyet az általam bemutatott gépi kivonatolási módszer is kiemelt a többiek közül. Számomra természetesen nagyon megnyugtató ez az eredmény. De nem ígérhetem, hogy a módszer minden esetben ilyen meglepően jó teljesítményre vezet. Ismétlem: a gépnek is lehet szerencséje — meg a szerkesztőjének is . . .

Tamás Lajos

Elnöki zárószó

A határozati javaslat előterjesztésétől és megvitatásától eltekintve — erre holnapi ülésünkön kerül sor — munkaértekezletünk érdemi részének végére értünk. Azt hiszem, túlzás nélkül állíthatjuk, hogy munkaértekezletünk jól sikerült, sőt hazai nyelvtudományunk jelenlegi fejlődési szakaszában történelmi jelentőségűnek is nevezhető. Hiszen nyelvészeinknek és matematikusainknak ez az első olyan közös vállalkozása, amely már nemcsak beszámoló és oktató jellegű volt, hanem amelyen már kisebb-nagyobb, de mindenképpen sokat ígérő eredményekről is hallottunk.

Nagy örömmel tölthet el bennünket az az egyetértés, a céloknak az a kölcsönös megbecsülése, amelyet nyelvészek és matematikusok részéről tapasztaltunk. Kitűnő matematikusaink még mentegetőznek, hogy a nyelvtudománnyal nem kötöttek eléggé szoros ismeretséget. Azt hiszem, hogy még mi, nyelvészek sem mélyedtünk el kellőképpen matematikába, a matematikai módszerek tanulmányozásába, de erről nem igen beszélünk. Igen megnyugtató, hogy különösen fiatalabb nyelvészeink nem azt mondják, hogy a matematikát nem ismerik, hanem azt, hogy meg fogják tanulni, s ehhez a munkához sok szerencsét kívánunk. Örömmel vehetjük tudomásul, hogy már csoportok és intézmények foglalkoznak információelméleti, gépi fordítási kérdésekkel, az adatfeldolgozó gépek problémáiról is sok érdekeset hallottunk. Íme az együttműködés első eredményei. A magam részéről szívesen hallottam volna egy előadást arról is, mennyiben és mennyire formalizálhatók a nyelvi jelenségek. A strukturalista nyelvészet eddigi eredményei alapján biztosra vehető, hogy ezen a téren lehetséges a további haladás.

„Matematikai nyelvészet" vagy „nyelvészeti matematika"? Ezt a többektől vitatott kérdést, azt hiszem kielégítően meg lehet válaszolni. Nyilván azt kell néznünk, hogy mi a tárgya a matematikai nyelvészetnek, ide értve a nyelvstatisztikát is. Világosan a nyelv a tárgyuk, a nyelvnek matematikai módszerekkel megragadható aspektusai. Ez dönti el, véleményünk szerint, hogy a most születő új diszciplínák nyelvészeti tudományok maradnak akkor is, ha matematikai módszereket alkalmaznak.

Kalmár elvtárs szóvá tette a nyelv és gondolkodás viszonyát s mint matematikus hajlik annak elfogadására, hogy nyelv nélkül is gondolkodhatunk s juthatunk el összefüggések felismerésére. Ezt a nézetet csak

bizonyos megszorításokkal lehet elfogadni. A 'nyelv' fogalmát ugyanis nem szabad ebben az esetben a beszélt nyelvre vagy valamely konkrét élő nyelvrendszer használatára korlátozni. A matematikus, a vegyész képletekben tudja elgondolni és közölni mondanivalóját, mégpedig szakmai mondanivalóját. Ez a képletnyelv is csak nyelv, bár korántsem azonos a mindannyiunk által használt köznyelvvel. Elképzelhető, hogy a matematikus, a vegyész valami új összefüggésre jön rá s közben nem a köznyelvi nevükön gondolkozik a dolgokon, hanem képletekben, szimbólumokban. De képletek, szimbólumok nélkül képtelenség gondolkozni, mert akkor nem tudom a fogalmakat mihez rögzíteni s azokat egymással vonatkozásba, kapcsolatba hozni. A képletekben, szimbólumokban gondolkodás tehát végeredményben szintén csak valamilyen nyelven való gondolkodás. A NaCl egyfelől és a *só (Salz, sel* stb.) másfelől jelek, ebből a szempontból nincs közöttük különbség, bár egyébként korántsem azonos jellegűek.

Igen meg lehetünk elégedve a munkaértekezletünk iránt érdeklődők számával is, ami nem utolsó sorban a jó szervező munka eredménye, amelyből különösen Szépe György és debreceni segítőtársa, Papp Ferenc vették ki részüket. Emellett az érdemi munkát is eredményesen segítették elő. Legyen szabad remélnünk, hogy munkaértekezletünk eredményessége serkentően fog hatni a nyelvészek és matematikusok további együttműködésére. Remélem továbbá, hogy az illetékes párt- és kormányszervek értekezletünk programját és eredményeit mérlegelve az eddiginél jóval nagyobb mértékben fogják támogatni azokat a célokat, amelyeket holnapi ülésünkön fogunk pontosabban körvonalazni.[1] Mai ülésünket azzal a meggyőződéssel zárom be, hogy nyelvészeink és matematikusaink megfelelő támogatás mellett ezeket a célokat meg fogják valósítani.

[1] „A matematikai nyelvészet és a gépi fordítás kérdései" munkaértekezleten elfogadott határozati javaslat: A Magyar Tudományos Akadémia Nyelv- és Irodalomtudományi Osztályának Közleményei 19. 335—9 (1962).

Telegdi Zsigmond

A konferencia néhány tanulsága

Ha meg akarjuk ítélni a munkaértekezlet eredményét, mindenekelőtt a célt kell tekintenünk, amelyet az maga elé tűzött.

Az elmúlt évtizedek folyamán a nyelvtudomány hatalmasan kitágult és elmélyült: a hagyományos leíró nyelvtan tudománnyá emelkedett, amely a nyelvet mint struktúrát vizsgálja, és ezzel lehetővé teszi, megkívánja a matematika mélyreható alkalmazását. A nyelvtudomány hazai képviselői általában sokáig bizalmatlanul szemlélték ezt a fejlődést, húzódoztak csatlakozni hozzá, értékesíteni eredményeit, meghonosítani módszereit és szellemét. Az utóbbi időben örvendetesen megváltozott a helyzet: felébredt, szemmel láthatóan terjed és mélyül az érdeklődés a nyelv újszerű vizsgálata iránt, megindult, biztatóan, a munka e hosszú ideig elhanyagolt területen.

Így került sor erre az értekezletre; azért rendeztük, hogy az új diszciplína számot adjon, magának és a nyilvánosságnak, elért eredményeiről, arról, hogy mennyire jutott, és felmérje a soron következő feladatokat, a feltételekkel egyetemben, amelyek ezeknek a megoldásához szükségesek.

Nem az a célom, hogy megállapítsam az előző két nap vitáinak tudományos „hozamát", ezt egyébként jelentős mértékben elvégezte már összefoglaló hozzászólásában Kalmár László. Van azonban, úgy látom, értekezletünknek néhány különösen fontos tanulsága; ezekre szeretnék itt befejezésül rámutatni.

Mindenekelőtt kitűnt — s ez alighanem meglepetés a távolabb állóknak —, hogy a matematikai nyelvészet problematikája ma már nálunk is képzett, tehetséges és, részben, tapasztalt kutatók sorát foglalkoztatja, hogy rendelkezünk azokkal az erőkkel, amelyek ezt a tudományágat kibontakoztathatják, ki is fogják kétség kívül bontakoztatni a mi országunkban.

Nyilvánvalóan megmutatkozott továbbá, hogy az új tudományág munkásai készséges segítséget találnak kiváló matematikusoknál, akik maguk is érdeklődnek a matematika nyelvészeti alkalmazásai iránt. Kalmár László, hogy a legjelesebb példát említsem, nemcsak áttekintette egy nagyszabású előadásban a matematika fejlődését, rávilágítva a nyelvészeti alkalmazások sokféle lehetőségére, hanem ezen felül bíráló, kiegészítő, továbbfejlesztő felszólalásaival lényegesen hozzájárult a konferencia sikeréhez általában.

Meggyőződhettünk végül arról — ehhez elég volt végigtekintenünk értekezletünk hallgatóságán —, hogy az érdeklődés a matematikai nyelvészet iránt megtalálható szélesen azoknak a körében is, akik maguk a hagyományos irányban igyekeznek továbbfejleszteni a nyelv tudományát. Ezért is kellett azt, amit eredetileg szűk munkamegbeszélésnek terveztünk, nyilvános értekezletként megrendeznünk: nyelvészközönségünk általában igényt tartott arra, hogy tudomást szerezzen az újszerű kutatások hazai eredményeiről.

A legközelebbi feladatokat és megvalósításuk feltételeit, amennyire felismerhettük őket, kifejtettük a határozati javaslatban. Ennek egyetlen pontját emelném ki ez alkalommal, azt, amely egy újfajta nyelvészképzést, matematikus nyelvészek képzését sürgeti.

Az egyetemi oktatásban a nyelvészet hagyományosan össze van kötve az irodalomtörténettel, a „nyelvszakos" hallgatónak meg kell ismerkednie mindkét tudomány elemeivel. Ez a gyakorlat nem csupán külső kényszerűségből ered. A nyelvtudomány önállósága, igaz, legalább száz év óta vitán felül áll. De ez a tudomány a múlt század folyamán mint történeti tudomány alakult ki, amely vizsgálódásainak anyagát részben (talán mondhatni: jelentős részben) régi irodalmak hagyatékában találta; a nyelvész nem tekinthetett el az ilyen szövegek sajátos jellegétől, viszont kutatásaival értékes segítséget nyújtott az irodalomtörténésznek a műalkotás nyelvi elemzéséhez.

Nyelvészet és irodalomtörténet hagyományos kapcsolatának megértéséhez fontos tekintetbe venni még egy másik körülményt is.

A történeti szemlélet alkalmazása a nyelv vizsgálatában teljességgel újszerű volt ugyan, korszakos haladás; mégis elmondhatjuk, hogy az a tudomány, amely ezt véghez vitte, a történeti nyelvészet, egy bizonyos, éspedig lényeges vonatkozásban nem változtatott a nyelvnek a hagyományos műveltségben közkeletű felfogásán; a nyelvtani leírás sokszázados módszerét a történeti nyelvészet nem tekintette tudományos eljárásnak, de a maga nemében, mint a nyelvtan, a nyelv l e í r á s á n a k módszerét elfogadta, reá épített.

Csak századunk elején kerül sor szakításra a nyelvnek a nyelvtani hagyományban kifejeződő felfogásával. Ez a fordulat elsősorban F. de Saussure nevéhez fűződik; a nyelv új, strukturális felfogásának nála találjuk a legtisztább, legélesebb fogalmazását.

Saussure szerint a nyelv rendszer, amelynek elemeit csupán a rendszer egészében elfoglalt helyük, kölcsönös viszonyuk határozza meg; más szóval: viszonyok hálózata, amelyet a viszonyok hordozóinak természetétől eltekintve, elvonatkoztatva kell szemlélnünk.

Ilyen rendszerekkel elsősorban, sajátosan a matematika foglalkozik, vizsgálatuknak ő szolgáltatja a legváltozatosabb és legfényesebb példáit. Érthető, hogy a nyelvtudomány strukturális irányának képviselői benne, általában a matematizált, egzakt tudományokban látták mintájukat, a nyelv egzakt elemzését írták zászlójukra, azt igyekeztek növekvő sikerrel megvalósítani.

A matematika újabban közvetlenül is behatol a nyelvtudományba. A nyelv társadalmi jelenség, kiderült azonban, hogy egyes aspektusai

csak a matematika, egyes matematikai diszciplínák kategóriáinak és módszereinek segítségével vizsgálhatók eredményesen.

Ez a fejlődés új feladat elé állítja az egyetemet. Tudomány és gyakorlati élet egyaránt igényelnek olyan nyelvészeket, akik a nyelvtudományt egész terjedelmében elsajátították, akik, nevezetesen, képesek a nyelv elemzésének új, egzakt módszereit kutatásban és technikai alkalmazásokban értékesíteni. Ehhez azonban a képzés új módjára van szükség, amely illő helyet ad ezeknek a módszereknek és — ami nélkül ez nem lehetséges — alapos matematikai műveltséget nyújt.

A Szovjetunióban alkalmazott megoldás láthatóan azon a felismerésen alapul, hogy a hagyományos nyelvészképzés puszta kiegészítése, megtoldása néhány újabb tárggyal nem vezet célra. Ehelyett az ország több egyetemén, mindenek előtt Leningrádban és Moszkvában, új szakot állítottak fel a filológiai fakultás keretében, új szakot, amely nem tanárokat, csak nyelvészeket képez, de ezeket a nyelvtudomány minden ágában kiműveli, különös gondot fordítva a nyelv vizsgálatának egzakt, strukturális és matematikai módszereire, és ezek gyakorlati alkalmazására, az alkalmazott nyelvtudományra.

Az egyetemi oktatás bővítését a nyelvtudomány újirányú fejlődése, a nyelv strukturális szemléletén alapuló vizsgálati módszerek kialakulása tette szükségessé, ezeknek a közvetítése áll az új szak munkájának középpontjában, határozza meg felépítését, jellegét. Ezt azonban nem úgy kell érteni, hogy az új szak a nyelvtudományt a nyelv strukturális, formális vizsgálatára redukálja; hallgatói valójában a maga teljességében sajátítják el a nyelvtudományt. Megismerkednek az összehasonlító történeti nyelvészet szemléletével és módszereivel is, éspedig az anyanyelv tudományos stúdiumán keresztül, szabályos russzisztikai iskolát járnak ki (irodalomtörténet nélkül).

Más országok is felismerték egy újfajta nyelvészképzés szükségletét és lépéseket tettek a kielégítésére. Nem kétséges, hogy a magunk körülményeihez képest követnünk kell ezeket a példákat; különben néhány év múlva kutatás és gyakorlat méltán panaszolhatják majd a nyelv egzakt elemzésében járatos káderek hiányát.

Nyelvtudományunk harmonikus fejlődése azonban még egyebet is kíván tőlünk.

A nyelvészet közvetlenül szétágazó, sőt ellenkező irányzatok képét mutatja; igazság szerint, objektíve mégis egységes tudomány, ami határt szab az elkerülhetetlen szakosodásnak. Másrészt szem előtt kell tartanunk, hogy a matematikai tudományok szerepe a nyelvészetben nem szorítkozik a problémák egy sajátos osztályának vizsgálatára; a gondolkodásnak a matematikában kialakult módszerei mélyen behatoltak a nyelvészet, nem utolsó sorban a grammatikai kutatás területére.

Ezért helyénvaló ugyan az elnöki megnyitó figyelmeztetése, amely int a dilettantizmus veszélyétől; de tanúi vagyunk a nyelvtudomány bizonyos mérvű matematizálódásának, s ez az objektív folyamat sürgetően követeli a nyelvészektől általában, hogy magasabbra emeljék matematikai és logikai ismereteik színvonalát.

Értekezletünk csupán kezdeti eredményekről számolhatott be; ennek

ellenére, úgy gondolom, nyugodt lélekkel állíthatjuk, hogy hasznosan, számottevően hozzájárult a matematikai nyelvészet megalapozásához hazánkban, meghonosításához. Jelentősége azonban végül is azon fordul meg, ami utána következik, hogyan, milyen mértékben és mennyi idő alatt sikerül megvalósítanunk a határozatait. Ehhez a munkához kívánok, amikor mostani értekezletünket bezárom, lelkesedést és kitartást.

Содержание

ВОПРОСЫ МАТЕМАТИЧЕСКОЙ ЛИНГВИСТИКИ И МАШИННОГО ПЕРЕВОДА

Доклады конференции, организованной Институтом языкознания АН Венгрии с участием Комиссии по кибернетике при президиуме АН Венгрии и Комиссии по общему языкознанию Отделения литературы и языка АН Венгрии (Будапешт, 8—10 марта 1962 г.)

Предисловие редакторов ..... 5—6

ЧАСТЬ I

Дь. Немет: Вступительное слово председателя ... 7— 9
Л. Кальмар: Математические и языковые структуры 11—74
Ф. Папп: Языковая система, процесс общения и их некоторые математические модели ..... 75—88
Ф. Кифер: Модели теории-множеств и математической логики в языке ..... 89—115
Б. Дёмёльки—И. Фонадь—Т. Сенде: Звуково-статистические исследования разговорного языка .... 117—132
Л. Козма: Устройство для лингвостатистических исследований ..... 138—176

Краткие сообщения .....
Л. Анталь (138), *Й. Ш. Петёфи* (141), *Л. Гальди* (145), *Э. Вертеш* (148), *М. Мико* (154), *Л. Папп* (155), *Л. Дежё* (158), *Дь. Сепе* (160), *Л. Кальмар* (166), *Л. Козма* (173), *И. Фонадь* (174), *К. Киши* (176)

ЧАСТЬ II

Дь. Хелль: Некоторые теоретические и практические вопросы машинного перевода с русского языка на венгерский ..... 177—196
Ш. Кони: Некоторые вопросы машинного перевода с венгерского языка на русский ..... 197—204

Ф. Хадаш: О русско-венгерском грамматическом словаре ..... 205—210
Дь. Сепе: Некоторые вопросы машинного перевода в Венгрии ..... 211—230
Д. Шипёци: Вопросы программирования предложных сочетаний ..... 231—242
Ш. Салаи: К вопросу о теории и практики машинного аннотирования венгерских научных текстов (Тезисы) ..... 243—244
А. Реньи: Теория информации и лингвистика .... 245—251
Е. Ружицки: О некоторых вопросах применения электромеханических машин в лингвистических целях ..... 253—261

Краткие сообщения..... 262—309
Дь. Кёрнер (262), *Р. Петер* (265), *Б. Хайтман* (266), *Б. Дёмёльки* (269), *Л. Дежё* (271), *М. Габор* (275), *Й. Келемен* (277), *Л. Деме* (284), *Дь. Сепе* (290), *Ф. Папп* (292), *Л. Кальмар* (295), *М. Мико* (305), *Ш. Салаи* (308).

*

Л. Тамаш: Заключительное слово председателя ... 317—318
Ж. Телегди: Заключительные замечания ..... 319—320

Contents

Problems of Mathematical Linguistics and Machine Translation

Proceedings of a conference organized by the *Institute of Linguistics* of the Hungarian Academy of Science — with the aid of the *Cybernetic Commission* at the Presidium of the Hungarian Academy of Science and the *Committee for General Linguistics* of the Section of Linguistics and Literature of the Hungarian Academy of Science.
(Budapest, 8 to 10th of March, 1962)

Preface of the editors ..... 5—6

PART I

J. Németh : Presidential opening adress ..... 7— 9
L. Kalmár : Mathematical and linguistics structures 11—74
F. Papp : Linguistic structure, communication process and some of their mathematical models 75—88
F. Kiefer : Set-theoretical and mathematical-logical models in the language ..... 89—115
B. Dömölki—I. Fónagy—T. Szende: Sound-statistical investigations in Colloquial Hungarian ..... 117—132
L. Kozma : A device for linguistic statistics .... 133—137

Minor papers ..... 138—176
L. Antal (138), *J. S. Petőfi* (141), *L. Gáldi* (145), *E. Vértes* (148), *M. Mikó* (154), *L. Papp* (155), *L. Dezső* (158), *Gy. Szépe* (160), *L. Kalmár* (166), *L. Kozma* (173), *I. Fónagy* (174), *K. Kiss* (176)

PART II

Gy. Hell : Some theoretical and practical problems of Russian—Hungarian machine translation 177—196
S. Kónyi : Some problems of the Hungarian—Russian machine translation ..... 197—204

F. Hadas : About the Russian—Hungarian grammatical dictionary ..... 205—210
Gy. Szépe : Some problems of the machine translation in Hungary ..... 211—230
Gy. Sipőczy : Problems in programming the prepositional constructions ..... 231—242
S. Szalai : Theoretical and experimental contributions to the mechanical abstracting of scientific publications in Hungarian (Abstract) ..... 243—244
A. Rényi : Information theory and linguistics ... 245—251
É. Ruzsiczky : On the linguistic application of the electro-mechanical data-processing machine ... 253—261

Minor papers ..... 262—309
Gy. Körner (262), *R. Péter* (265), *B. Hajtman* (266), *B. Dömölki* (269), *L. Dezső* (271), *M. Gábor* (275), *J. Kelemen* (277), *L. Deme* (284), *Gy. Szépe* (290), *F. Papp* (292), *L. Kalmár* (295), *M. Mikó* (305), *S. Szalai* (308) .....

*

L. Tamás : Presidential closing adress ..... 311—312
Zs. Telegdi : Concluding remarks ..... 313—316

Tartalomjegyzék

A szerkesztők előszava ..... 5—6

I. RÉSZ

Németh Gyula : Elnöki megnyitó ..... 7— 9
Kalmár László : Matematikai és nyelvi struktúrák 11—74
Papp Ferenc : Nyelvi rendszer, közlési folyamat és ezek néhány matematikai modellje ..... 75—88
Kiefer Ferenc : Halmazelméleti és matematikai-logikai modellek a nyelvben ..... 89—115
Dömölki Bálint—Fónagy Iván—Szende Tamás : Köznyelvi hangstatisztikai vizsgálatok ..... 117—132
Kozma László : Nyelvstatisztikai berendezés .... 133—137

Kisebb közlemények ..... 138—176
Antal László (138), *Petőfi S. János* (141), *Gáldi László* (145), *Vértes Edit* (148), *Mikó Pálné* (154), *Papp László* (155), *Dezső László* (158), *Szépe György* (160), *Kalmár László* (166), *Kozma László* (173), *Fónagy Iván* (174), *Kiss Károly* (176)

II. RÉSZ

Hell György : Az orosz—magyar gépi fordítás elméleti és gyakorlati kérdései ..... 177—196
Kónyi Sándor : A magyar—orosz gépi fordítás néhány problémája ..... 197—204
Hadas Ferenc : Az orosz—magyar nyelvtani szótárról ..... 205—210
Szépe György : A magyarországi gépi fordítás néhány kérdése ..... 211—230
Sipőczy Győző : Az elöljárós kifejezések programozási kérdései ..... 231—242
Szalai Sándor : Elméleti és kísérleti adalékok a magyar nyelvű tudományos közlemények gépi kivonatolásának kérdéséhez (Kivonat) ..... 243—244
Rényi Alfréd : Információelmélet és nyelvészet... 245—251
Ruzsiczky Éva : Az elektromechanikus adatfeldolgozó gép nyelvtudományi alkalmazásának néhány kérdéséről ..... 253—261

Kisebb közlemények ..... 262—309
Körner Györgyné (262), *Péter Rózsa* (265), *Hajtman Béla* (266), *Dömölki Bálint* (269), *Dezső László* (271), *Gábor Mihály* (275), *Kelemen József* (277), *Deme László* (284), *Szépe György* (290), *Papp Ferenc* (292), *Kalmár László* (295), *Mikó Pálné* (305), *Szalai Sándor* (308)

*

Tamás Lajos : Elnöki zárószó ..... 311—312
Telegdi Zsigmond: A konferencia néhány tanulsága 313—316

*

Angol nyelvű tartalomjegyzék ..... 317—318
Orosz nyelvű tartalomjegyzék ..... 319—320

A kiadásért felel
AZ AKADÉMIAI KIADÓ IGAZGATÓJA

✳

Műszaki szerkesztő
STELCZER MÁRIA

✳

A kézirat beérkezett: 1963. VIII. 10.
Példányszám: 1100.
Terjedelem: 28,35 (A/5) ív

✳

Akadémiai Nyomda, Budapest
Felelős vezető: BERNÁT GYÖRGY

Ára: 55,— Ft



AKADÉMIAI KIADÓ, BUDAPEST