

## Az orosz—magyar gépi fordítás elméleti és gyakorlati kérdései

Sajnos, egyelőre nem vagyunk 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 meglévő fölösleges rejtélyesség alaptalansága nyilvánvalóvá legyen.

### I. 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 meglévő „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 K o z m a L á s z l ó tanszékvezető egyetemi tanár által tervezett és a Budapesti Műszaki Egyetem Vezetékes Híradástechnikai Tanszékén található MESZ—1 típusú jelfogós számológép.<sup>1</sup>

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.<sup>2</sup>

<sup>1</sup> 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.

<sup>2</sup> 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 billentyűzetten 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, kivitel 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

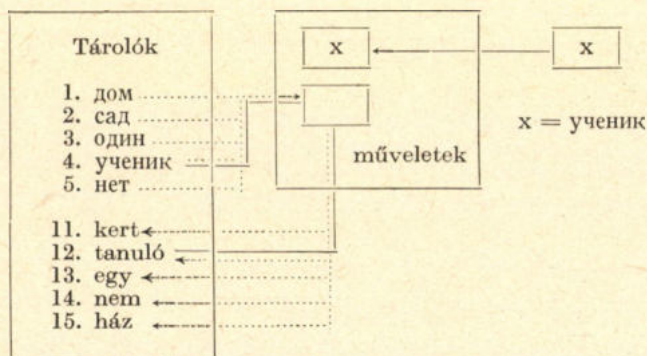
## 2. A fordítás általános menete

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 tömorfémák 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, melyekkel megkapjuk a forrásnyelvnek a mondanivalót hűen tartalmazó szerkezeti egységeit, olyan fordítási szabályokat kell előírni, 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.<sup>3</sup>

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 alpműveleteket, 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.<sup>4</sup> 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” teremtsék, 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

<sup>3</sup> P. Sz. Kuznyecov—A. A. Ljarunov—A. A. Reformackij, Основные проблемы машинного перевода: Вопросы Языкознания 5/5. 107—110 (1956).

<sup>4</sup> 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 mondatstruktúrák megoldásáig. Így a megoldandó feladatok nehézségi fokát is magunk tudtuk megválasztani.

Így készültek algoritmusok toldalékos és előljárós orosz főnevek fordítására, majd egyszerűbb mondatok fordítására is. A első géppel fordított mondat (*Решаем ли машина задачу?*)<sup>5</sup> egy ragozott ige és két toldalékos nőnemű főnév elemzését tartalmazta. A gépi szótár orosz része a *ли* szócska mellett a *реша*, *машин* és *задача* töveket, valamint az *em*, *a* és *y* toldalékokat tartalmazta.

A szavak tárolásához minden betű (cirill és latin) négy kettes számrendszerben írt számjegyből (bit)<sup>6</sup> 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 bite 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

<sup>5</sup> Azért volt célszerű kijelentő mondat helyett kérdő mondatot venni, mert ennek fordítása kevesebb változatot ad.

<sup>6</sup> 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 szótó alapján történt, a program megfelelő rendszerű felépítésével. Ez lényegében azt jelentette, hogy külön ragelemzést kapott az ige ragja, és külön elemzést a nőnemű főnév végződése, ami az adott esetben mindig egyetlen értelmezéshez vezetett. Ez a módszer az indulásnál fölöslegessé tette azt, hogy a szótárban a szótővekre vonatkozóan nyelvtani kiegészítő jelzéseket helyezünk 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.*)

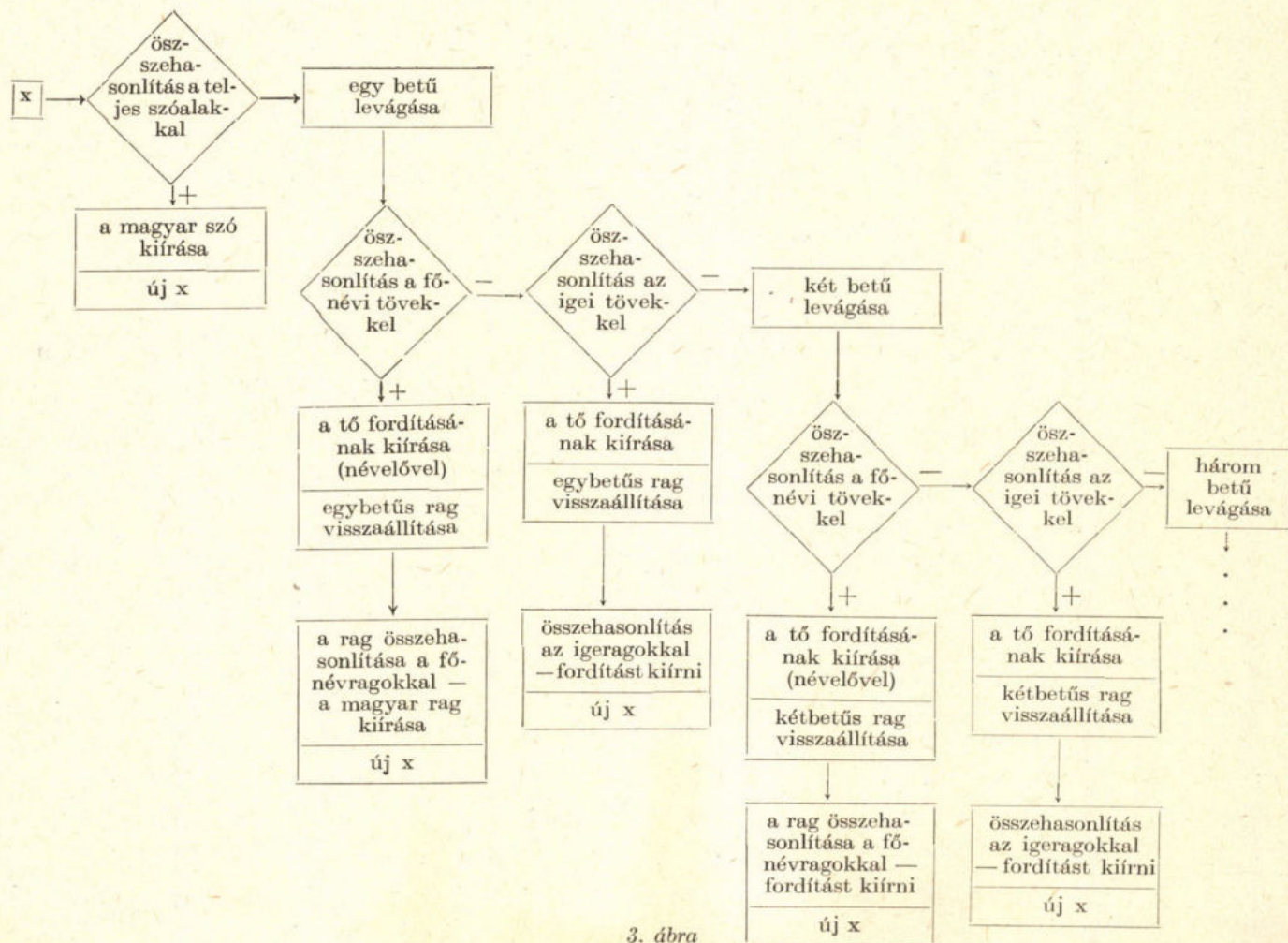
A műveletek egymásutánját a 182. lapon látható utasítási vázlat mutatja.

Látjuk, hogy a szavak morfológiai elemzéséhez a program a szótóból indul ki, és a szótővek 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 mondat szerkezetekben szükség volt arra, hogy egy toldalék-morféma magyar fordítását ne csak a szótó és előljáró, 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



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 információ-sor 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.<sup>7</sup>

## 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 meglehetősen 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.<sup>8</sup> Az elgondolás lényegében 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.<sup>9</sup> Az első kiegészítő elemzésben a főnév a végső határ,

<sup>7</sup> D. Ju. Panov: Автоматический перевод (Moszkva, 1958); E. Delavenay: An introduction to machine translation (London, 1960).

<sup>8</sup> Machine translation of languages (Szerk. W. N. Locke—A. D. Booth, New York, 1955).

<sup>9</sup> 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 előjáró, illetőleg melléknévi, névmási jelzők alapján. A kiegészítő elemzés eredményeképpen a célnyelvi szövegből kihagyják a helytelennek bizonyuló szavakat, előjárókat, 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 ismételt 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.<sup>10</sup> 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.<sup>11</sup>

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 mellékelte külön jelzések, nyelvtani kódok<sup>12</sup> fejezik

<sup>10</sup> Vö. D. Ju. Panovnak a 7. jegyzetben hivatkozott műve 32. lapjával.

<sup>11</sup> I. A. Melesuk, Некоторые вопросы машинного перевода за рубежом: Доклады по машинному переводу (Moszkva, 1961)



ki. Ezeknek a nyelvtani kódoknak kell olyan bőségeknek 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 szótót 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.<sup>13</sup> Míg a nyelvtani alakokat ismerő ember számára elegendő egy szó egyetlen szótövének megadása és a tanult hangtani változások alapján például a *цуч* szótöbben felismeri a *цудемь* ige *цуд* 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 szóalakokkal 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”.<sup>14</sup> 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.<sup>15</sup> Ezzel a szókeresésre fordított idő lecsökken.<sup>16</sup>

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

<sup>12</sup> A kódok tulajdonképpen rendszerbe foglalt (vagy minden rendszer nélkül adott) számok, amelyek mintegy rejtjelezett formában fejeznek ki valamit (a mi esetünkben szintaktikai viszonyokat).

<sup>13</sup> I. K. Belszkaja, О принципах построения словаря для машинного перевода: Вопросы Языкознания 8/3. 98—106 (1959).

<sup>14</sup> O. Sz. Kulagina—I. A. Melesuk, Машинный перевод с французского на русский: Вопросы Языкознания 5/5. 111—120 (1956).

<sup>15</sup> A. G. Oettinger: Automatic language translation (Cambridge, Mass., 1960).

<sup>16</sup> 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 jelleget 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.<sup>17</sup> 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.

### 3. 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.<sup>18</sup> 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.<sup>19</sup>

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,

<sup>17</sup> Vö. I. K. Belszkajának a 13. jegyzetben hivatkozott művével.

<sup>18</sup> 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).

<sup>19</sup> 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.<sup>20</sup>

### 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 elsőnek egy részét, a nem változó morfológiai tulajdonságokat a szótárban készen találjuk. Ilyen a főnév neme, ragozási csoportja (ezen belül az is, hogy élő, vagy életlen 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 szókapcsos-

<sup>20</sup> 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 elsőkben: 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 hangsúlyozták; gyakorlati alkalmazására tulajdonképpen csak az utóbbi években került sor (az 1958-as washingtoni beszámoló<sup>21</sup> például még nem tartalmaz ilyet).

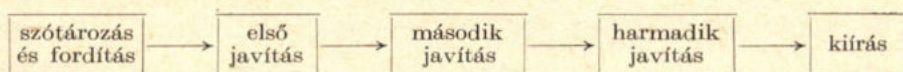
Az orosz—magyar gépi fordításban a mondatelemzés kiindulópontja, mint másutt is, az írásjel. Ha az írásjel után nem jön közbevetés, a már kiszótározott szakaszt mondatelemzésnek vetjük alá. Számon vesszük, van-e állítmány, illetőleg alany a vizsgált szakaszban, és az elemzésbe belevonjuk a szakaszt kezdő 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.

<sup>21</sup> Vö. a 9. jegyzetben hivatkozott művel.

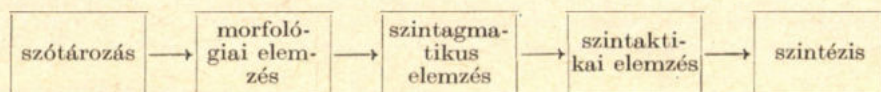
## 10. Szintaktikai elemzés; az információsor tartalma

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.

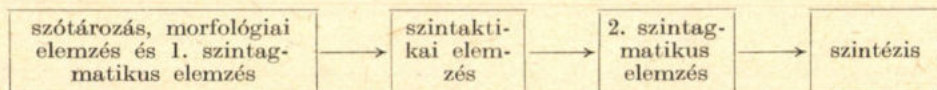
A gépi fordítás általános menete szavankénti fordítás esetén:



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 mondat szerkezetet 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 tárolószáma.

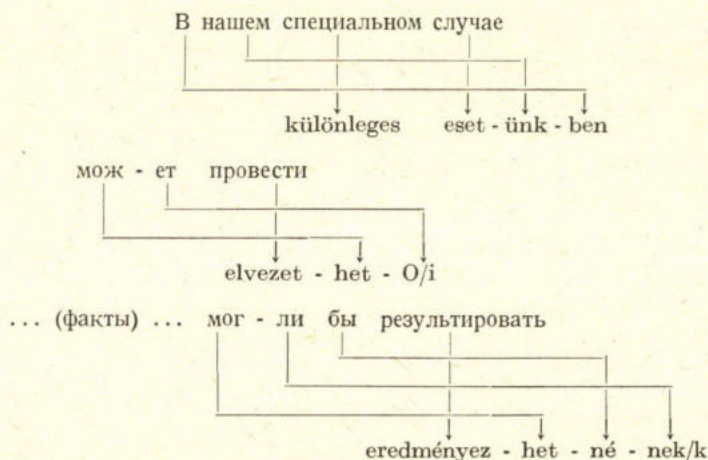
A forrásnyelvi szöveg egyes mondattani szakaszaira vonatkozólag a következő adatokat tartalmazza az információsor:

- (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 szintetikus 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,<sup>22</sup> hogy az automatikus szótárral nyert szóról szóra fordítás csak akkor használható, ha a morféma 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 szókatagóriát toldalékokkal fejez ki. A program ezeket a szavakat „toldalék értékű grammatikai viszony-



4. ábra

<sup>22</sup> Lásd a 9. jegyzetben hivatkozott gyűjteményben.

szók” címen külön utasításrendszer alapján dolgozza fel. Az orosz—magyar gépi fordításban ilyennek számítanak az előjárók, 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 viszonzyszók közül a birtokos névmás a többes szám jelét is megváltoztatja. (Lásd a 4. ábrát.)

A grammatikai viszonzyszóknál 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 igehez tartozó szavak a mondatban bárhol elhelyezkedhetnek, és mindig az igehez 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 sajátos, hiszen más indoeurópai nyelvekhez hasonlóan az oroszban is például egy előjárós vagy esetragos főnév nem mindig a mondat igei állítmányával alkot szó szerkezetet, hanem az őt megelőző főnévvel vagy a megelőző főnevek valamelyikével.<sup>23</sup> Ugyanakkor az ilyen hátravetett, nem egyeztetett jelzőt adó főnév alaktanilag semmiben sem különbözik az igehez 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 ige mondatokban sem.)<sup>24</sup>

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.<sup>25</sup>

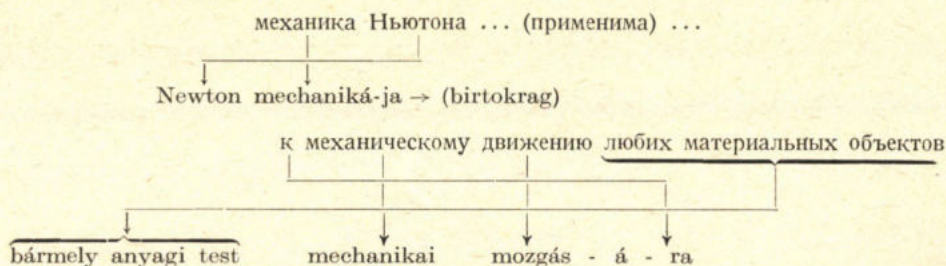
Az ilyen jellegű szórendi változtatásokhoz az orosz—magyar gépi fordításban külön meg kell alkotni a birtokszó, majd a birtokos szó melléknéves, névmásos, illetőleg számnéves jelzős szókapcsolatát, hogy a szórendesét a teljes szókapcsolattal lehessen elvégezni. A forrásnyelvi szó-

<sup>23</sup> Papp Ferenc, Трансформационный анализ русских присубстантивных конструкций с зависимой частью-существительным: *Slavica*: 1. 55—83 (Debrecen, 1961).

<sup>24</sup> Hell György—Sipőczy Győző: Orosz műszaki szövegek fordítása (Műszaki egyetemi jegyzet, Bp., 1961).

<sup>25</sup> Hell György, О нескольких характерных чертах программы МП с русского языка на венгерский: *Slavica* 2. 151—64 (1962).

szervezetek 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 mondatnailag helyes ragokat is megkapják, a nyitószók megválasztása a magyar nyelvi sajátságoknak megfelelően történt, és az előljáró, valamint a többi grammatikai viszonyzó nem számít nyitószónak.

A szószerkezetek nyitó- és zárószavait szófaji kategória, vagy egyedi tulajdonság alapján kapjuk:

Nyitószók:

- (1) melléknév:
  - a) melléknévi jelzős főnevekhez,
  - b) melléknéves szerkezetekhez,
  - c) középfokú hasonlító szerkezetekhez,
  - d) főnévi igeneves szerkezetekhez;
- (2) melléknévi és határozói 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 melléknemekhez 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 előljáró nem nyitószó, mivel egy orosz előljárás főnév a magyar célnyelv szempontjából nem ad szószervezetet.

A nyitószavak rendszere a fordításban előreható tényezőként működik: egy nyitószó megjelenése olyan programozás kezdetét jelenti, mellyel fölkészülünk a nyitószóra jellemző szószervezet fogadására. Ezzel az eljárással kettős eredményt érünk el: egyrészt a szószervezetek kialakítását a sorrendileg elől álló tagtól kezdjük meg, és nincs szükség a később jövő alaptagból kiinduló bonyolult visszakeresésre; másrészt a szószervezet be nem következése esetén értesülünk arról, hogy az adott szónak esetleg különleges szerepe van a mondatban. Az utóbbi információ különösen a melléknemek 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 szószervezetnek még nincs vége, de egy másik szó újabb szószervezetre ad lehetőséget (például: birtokviszony, hátravetett határozói alakú jelzős szerkezet, igeneves szerkezetek stb.). Ilyen esetben a szószervezet 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 szószervezeteket 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 folderített 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.<sup>26</sup> 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 előjárás, illetőleg esetragos főnevet kiterjedt elemzésnek kell alávetni, hogy pontos szintagmatikus kapcsolatát megállapíthassuk. Ez gyakran igen körülményes feladat, amit, ha a főneveknek csak alakítani 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 ...” előjárás főnév az állítmány határozója-e, vagy pedig (a magyar nyelvtani terminológia jelölésével) az előtte álló főnév határozói alakú jelzője.

### 13. A szó szerkezetek 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 viszonyozóktól eltekintve ugyanolyan szófajjal adjuk vissza a fordításban. A szó szerkezetekre vonatkozóan ez az elv azt jelenti, hogy az orosz nyelv szókapcsolatai a szórendtől eltekintve ugyanolyan szókapcsolatokat adnak a magyarban is. A szó szerkezetek egyszerűbb eseteitől eltekintve ez az elv az orosz—magyar gépi fordításban éppúgy nem valósítható meg, mint az egyéb, már meglévő fordítási rendszerekben. Több fordítási központ külföldön a szó szerkezetek fordítására szókapcsolat-mintákat állított össze, melyek a forrásnyelv szó szerkezeteihez hozzákapcsolják a célnyelv megfelelő szókapcsolatait, és a fordítás ezeknek a mintáknak az alapján történik. T. N. Molosnaja az angol—orosz fordításhoz kb. 7 000 szó szerkezet-mintát használ fel, melyek nemcsak a fordítási lehetőségeket adják meg, hanem az angol szöveg szófaji és szó szerkezeti elemzésére is ezek a minták szolgálnak.<sup>27</sup>

Az orosz forrásnyelv gazdag alaktana feleslegessé teszi számunkra a szókapcsolatoknak ilyen módszerrel való tisztázását, mert a szavak alakítani jegyeiből nemcsak azok szófaji hovatartozására, de egymással alkotott szintagmatikus kapcsolataira vonatkozóan is elegendő információt kapunk. Mégis bizonyos szó szerkezeti típusokra, rendszerünkben a nyitó-

<sup>26</sup> Például: *Были исследованы процессы сжигания топлива в двигателях с воспламенением от сжатия*, a határozóragos főnevek lehetséges jelzői szerepének figyelembevételével nélkül: „*Vizsgáltuk az üzemanyag égésfolyamatait a motorokban sűrítéssel gyűjtéssel.*” — A nyelvtanilag helyes fordítás: „*Vizsgáltuk az üzemanyagnak sűrítéssel gyűjtésű motorokban végbemenő égésfolyamatait.*”

<sup>27</sup> 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: *Аббан аз esetében, ha részletesen megvizsgáljuk a merev test pontjainak mozgását . . .* Метод определения понятия количества информации . . . ; Szóösszetétellel és melléknév-képzéssel: *Аз információмөннүйсэг 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: *Аз атомok szerkezetének csak az energetikai szintek feltüntetésével való ábrázolásakor . . .* — A mondat fordítása transzformációval: *На аз атомok szerkezetét csak аз energetikai szintек feltüntetésével áбрázолjuk . . .**

#### 14. Néhány sajátos magyar kérdés

A gépi fordítás már eddig megvalósított algoritmusaihoz hasonlóan a fordítás szintetikus műveleteihez, amikor a gép a célnyelvi szöveget állítja össze, tárolnunk kell a magyar szavaknak mindazon alakotani 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 alakotani ismertető jegyet kell adni. A névszótövek sokféleségének jelzése a magyar célnyelvi fordításnak különleges vonása. Alaktani ismertető jegeket 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 sajátosságai 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 hátravetett 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.