

SZÖVEGTAN ÉS SZÁMÍTÓGÉPES FORDÍTÁS

HUNYADI LÁSZLÓ

0. Bármilyen értelmezését vegyük is a „szöveg” fogalmának, megkerülhetetlenül hivatkoznunk kell arra, hogy a szöveg valamely nyelvi formáknak azon összessége, amelyek elsődleges funkciója az információátadásban rejlik. Éppen ezért mindazon diszciplínákban fontos a szöveg tanulmányozása, ahol az információ leírásának, értelmezésének, továbbításának, azaz az információ feldolgozásának jelentősége van. Ez olyan széles és közelebbi céljai tekintetében akár eltérő területeket jelent, mint például a verbális művészetek általában, továbbá az irodalomtudomány, a jogtudomány, a közigazgatás, a reklám és általában a médiumok és mint, ami bizonyos mértékben ezeket a területeket kiszolgálja is kész, a nyelvtudomány.

A nyelvtudomány ágas-bogas irányai között alig van olyan, amely valamely módon, igen gyakran meglehetősen intuitívan, ne foglalkozna a nyelvvel, vagy legalábbis ne hivatkozná a nyelvre úgy, mint ami a szöveg építője.

Bármilyen konkrét értelmezését adjuk is a „szöveg” fogalmának, a szöveg aligha választható el a tágabb értelemben vett jelentéstől. Így már a fonológia vizsgálata során is, ha másképp nem, hát látens módon előbukkan a „szöveg”, hiszen a fonéma fogalmának vagy valamely konkrét nyelv fonémarendszerének meghatározásánál a jelölésre és a jelentésre való hivatkozás elkerülhetetlen. Ez még nyilvánvalóbban érvényes a morfológiára, ahol például egy morféma ailomorfiáját éppen azáltal ismerjük fel, hogy az adott morfémával ellátott szót mint *szövegszót* vizsgáljuk. De éppígy lényeges a szövegre való hivatkozás a lexikológiában (ami pedig látszólag izoláltan kezeli a szavakat), hiszen a szavaknak nemcsak a jelentését, de vonzatait és frazeológiai kapcsolódásait is a nagyobb, a szövegegységek felől határozhatjuk meg. Ami pedig a még nagyobb nyelvi egységet, a mondatot illeti, jól tudjuk, hogy vannak olyan viszonyok, amelyek nem vagy nem kielégítően dőlnek el az egyszerű mondatok szintjén, továbbá, hogy jobban vagy kevésbé jobban vannak definiálva a mellérendelt vagy alárendelt kapcsolatokban.

A nyelvészetnek egyik speciális ága a számítógépes nyelvészet, azon belül is a számítógépes fordítás. Ennek során különös hangsúllyal merülnek fel az alulmeghatározottság problémái: ezek a rendszerek akár automatikus, akár interaktív üzemmódban dolgoznak, a mondat szintjénél túl nem jutnak, és így számos esetben nem képesek az adott mondat minden szerkezeti összefüggésének a feltárására. (Ebben a vonatkozásban egyre több konkrét tapasztalattal is rendelkezünk a magyarról angolra történő gépi fordítási rendszerünk kialakítása során.)

A fenti felsorolásból kitűnik, hogy számos tudományterületen a szövegre való hivatkozás elkerülhetetlen. Az is világos ugyanakkor, hogy különböző területek különböző igényekkel lépnek fel. Így például egy jogszabály megalkotása során elengedhetetlen annak a leírása, hogy milyen lehetséges értelmezéseket enged meg maga a

szöveg. Itt tehát a szöveg logikai leírására van szükség, de úgy, hogy eközben az elemzés figyelembe veszi azokat a szociológiai-kulturális viszonyokat is, amelyek a különböző értelmezéseket alátámaszthatják. Egy irodalmi mű irodalomközpontú megítélésére többnyire megkívánja azt is, hogy a szövegből következtetni tudjunk az adott társadalmi-gazdasági viszonyokra éppúgy, mint a szerzőnek bizonyos, a mű születését befolyásoló tulajdonságaira is. A médiumok esetében egy szöveg vizsgálatának egyik fontos célja annak a hatásnak a felmérése, amit az adott szöveg gyakorolhat a megcélzott környezetre.

Folytathatnánk még a sort: további igen szerteágazó igényekkel találkozhatnánk, és mind jogosak lennének a maguk területén. De itt kell szembenéznünk az egyik legfontosabb kérdéssel: mi is hát a szövegnyelvészet feladata? Vajon az-e, hogy magába olvasztva számos jobban vagy kevésbé jobban érintkező tudományterületet, „teljes megoldást” nyújtson, vagy pedig az, hogy a nyelvészet eszközeit maximálisan kihasználva „csak” annyit érjen el, ameddig a nyelvészet terjed. Azaz: meddig terjed(je)nek a nyelvészet határai?

1. Úgy gondolom, hogy a szövegnyelvészet akkor és csak akkor nyújthatja önmaga maximumát, ha megmarad nyelvészetnek, és a nyelvi formákból kiindulva olyan modelleket, olyan szabályokat alkot, amelyek egzakt módon nyomon követhetők magában a szövegben. Azaz a szövegten olyan formális elméletére gondolok, ami támaszkodik a mondattan valamely formális elméletére, integrál magába egy formális (szimbolikus) logikai elméletet, és ezt egy olyan pragmatikai elmélettel egészíti ki, ami kiindulópontként ugyancsak a formákban manifesztálódott nyelvi anyagot tekinti.

Egy ilyen integrált modell megalkotása során fontosnak tűnik még figyelembe venni a következőket. Minél magasabb nyelvi szintek felé közeledünk, annál kevésbé tudunk eltekinteni a nyelvi *funkciótól*, egyúttal annál közelebb kerülünk a nyelvi formát létrehozó *emberhez*. Ugyanakkor az emberhez, egészen pontosan annak *gondolataihoz* mégsem férünk hozzá. Hiú ábrándnak és koncepcionálisan zavarónak tűnhet egy olyan célkitűzés, ami például arra irányulna, hogy egy irodalmi mű mint szöveg elemzése során magának a szerzőnek a gondolatait kívánjuk leírni. Ez ugyanis arra a téves elképzelésre támaszkodna, hogy a nyelv, miközben gondolataink kifejezője, egyben gondolataink *maradékta*lan kifejezője is. Gondolataink talán sohasem lesznek teljesen függetlenek a nyelvtől, de ilyen közvetlen megfelelés nem állhat fenn. Ellenkező esetben nem lenne sem művészet, sem jogértelmezési zavar, sem félreértés a maga hétköznapi értelmében.

Mindennek megfelelően tehát nem várható el, hogy egy szöveg nyelvészeti megközelítése megfeleljen minden, a szöveggel, pontosabban a szöveg kapcsán támasztott követelményeknek. A szöveg nyelvészeti vizsgálata így nem mondhat semmit arról, ami valamilyen jól leírható megjelenési formában nem része a szövegnek, vagy egyéb szövegbeli formákkal nem hozható szabályos összefüggésbe. Ezért nemigen vállalkozhat arra, hogy valamely társadalmi-kulturális hátteret felderítsen, vagy mondjuk a befogadók lehetséges reakcióiról vélekedjen. S bár a szöveg vizsgálata teheti a legtöbbet a jelentés vizsgálatában, a szöveg értelméhez, azaz a *gondolathoz* azonban mégsem juthat el.

A formalizálhatóság mellett van még egy követelmény, ami fontosnak tűnhet: az univerzalitás követelménye. Ez azt jelenti, hogy a kutatásoknak elsősorban azon szövegszerkezeti/szövegszerkesztő összefüggések feltárására kell irányulniuk, amelyek szövegtípusoktól függetlenül jellemzők a szövegekre, és kívánatos módon nem nyelvspecifikusak. Ezt abból a meggyőződésből kell tenni, hogy ami univerzális, az

egyben az egyedi jelenségek (valamely speciális szövegtípus vagy nyelv) leírását, jellemzését is szolgálja.

Az univerzalitás igénye egyben azt is jelenti, hogy nem elegendő a szövegtani jelenségeknek csupán valamely (bármilyen mélységű) *leírását* adni. A leírás ugyanis sohasem lehet teljes, így általános sem. Ehhez egy modell megalkotására van szükség, ami általános elvekből, szabályokból áll. Ezen elveknek és szabályoknak az lehet a céljuk, hogy segítségükkel bármely tetszőleges, helyes szöveget létrehozassunk, természetesen előre jól definiált formális kritériumok alapján. Egy ilyen feladatra egy generatív szemléletű szövegtan alkalmasnak látszik. (Éppen a formálisan jól definiálhatóság kritériuma alapján azonban kitűnhet, hogy aligha lehet célunk az, hogy valamilyen verbális műalkotást generáljunk: az alkotás folyamatában nem kevés olyan tényező játszik szerepet, ami nyelvi formákkal, azok összefüggéseivel nem fogható meg.)

Mindezek alapján, a fenti célokat szem előtt tartva, a „szöveg” fogalmát úgy határozhatjuk meg, hogy az nem más mint (a) amit a szöveggeneráló modellünkben létrehozhatunk, és (b) ami a hétköznapi értelemben vett szövegnek azon része, amit e modell felismer, kezelni tud.

2. Az itt következőkben példákkal szeretném alátámasztani az elmondottakat. Megfigyeléseimben az említett számítógépes fordítási rendszer kialakítása során felbukkanó nehézségeinkre támaszkodom.

Magáról a fordítási rendszerről annyit, hogy ez egy olyan program lesz, ami tetszőleges magyar szövegeket fog angolra fordítani, interaktív üzemmódban. Az interaktív üzemmód azt jelenti, hogy annak érdekében, hogy egzakt fordítást adjunk, a gép nem fog statisztikai gyakoriságon alapuló heurisztikus döntéseket hozni (azaz nem fog az ember helyett dönteni pl. a mondatban meg nem jelenő *tárgy* értelmezésében), hanem ilyenkor a felhasználóhoz fordul, felfüggesztve a fordítási folyamatot. Nyilvánvaló, hogy egy fordítás egyebek mellett akkor lesz hatékony, ha minél kevesebb külső beavatkozásra van eközben szükség. A beavatkozások száma pedig nem tűnik kevésnek, éppen azért, mert a megközelítésünk a leginkább formalizált *generatív* szemléletet követi (vö. CHOMSKY: 1981.), azt, ami éppen a formalizálhatóság szigorú kritériumai miatt mondatcentrikus, azok az elvek, amelyek bizonyos szerkezeti összefüggéseket a nyelvi elemek között meghatároznak, a (tetszőleges bonyolultsági fokú) mondat határain belül érvényesek. Így a hatékonyság növelése érdekében célszerűnek látszik a mondat határain túlra is tekintenünk, távolabbi összefüggéseket is feltárnunk, de anélkül, hogy feladnánk a formális (levezethető) megközelítés alapelvét. A szövegnyelvészet ilyen irányú fejlődése egy interaktív vagy automatikus fordítási rendszer továbbfejlődését is jelentené.

Példáink tehát olyan eseteket mutatnak, amikor a számítógép felfüggeszti a fordítást, és a felhasználóhoz fordul. Ilyenkor a felhasználó a szöveg ismeretében megadja a továbblépéshez szükséges információt.

A számítógépes fordítás szempontjából egy szöveg megértése a következőket jelenti: (a) minden egyes mondat szintaktikai összetevőinek az azonosítása, (b) az egyes mondatok közötti viszony meghatározása, (c) a szavak jelentésének a meghatározása, (d) egy szövegnek a fordításához szükséges és elégséges mértékű megértése.

(a) Minden egyes mondat szintaktikai összetevőinek az azonosítása

A mondat elemzése megköveteli a helyes argumentumszerkezet leírását, az egyes argumentumok azonosítását, függőségi viszonyaik meghatározását. Ehhez szükség van egy morfológiai és egy szintaktikai elemzésre. A morfológiai elemzés az adott nyelv morfológiai típusától függően több vagy kevesebb információt nyújt az adott szó szintaktikai szerepére vonatkozóan. A morfológiában (és a fordítás szempontjából igen előnyös agglutinatív morfológiában) meglehetősen gazdag magyar nyelv esetében is számos alaki egybeesés van különböző szóosztályokhoz tartozó szavak között (vö. *ér, hát*), de egyazon lexéma több alakja között is (vö. *halljuk, olvastam, mentek*). Ezen egybeesések a morfológiánál magasabb szinteken kell hogy eldőljenek. A legtöbb esetben erre a mondat alkalmas is lesz, és így az argumentumszerkezet a mondaton belül tisztázódhat. Erre tulajdonképpen akkor van esély, ha az adott ige minden argumentumára találunk megfelelő jelöltet (és csak egyet) az adott mondatban. Abban a pillanatban, ha a realizált argumentumszerkezet hiányos (a magyarra jellemző névmáshagyás vagy NP-anafora miatt), az azonosítás érdekében ki kell lépni a mondat keretei közül (e problémakör részletes tárgyalását lásd: PETŐFI: 1990.).

Vegyük a következő mondatot.

(1) Felvette a kabátját és elment.

Ebben a mondatban legalább két szerkezeti elem nincs tisztázva: a mondat alanya és a tárgy birtokosa. Ezek közül az alany megtalálása jóval kézenfekvőbb: kell az előzmények között lenni egy alanyjelöltnek. A nehézséget csupán az „előzmény” határainak a kijelölése okozza: ez lehet éppúgy az előző mondat, mint valamely korábbi is. Jelenlegi mondattani és szövegtani eszközeink, melyek jól definiált formákra épülnek, nem alkalmasak ennek a kérdésnek az eldöntésére. A hiányzó birtokos azonosítása még kevésbé biztató a szövegösszefüggésből. Legalábbis kevésbé mondható ki, mint a hiányzó alany esetén, hogy valahol az előzmények között találunk egy hasonló funkcióban álló elemet, amivel e hiányzó elem behelyettesíthető. Ehhez tényleg „érteni” kellene az előzményeket.

Kíváncsún látszik a szövegepítés egy olyan modelljének a kidolgozása, amelyben egyértelműen megfogalmazható az a mechanizmus, ami mentén valójában mi, a szöveg építői vagy befogadói az adott eseményláncolatot magunkban felépítjük. Konkrétabban: feltételezhetünk egy olyan mechanizmust, ami ismereteink folyamatos bővülését, változását biztosítja az elhangzó információk birtokában. Azaz az egyes szavak (szembetűnő módon a névszók) bizonyos szövegbeli változók értékei, úgy, hogy e változók építik tovább a szöveget, egyszer definiált és végig azonos értékkel éppúgy, mint folyamatos vagy többszöri új értékadással. E bonyolult és dinamikus összefüggések leírására egy dinamikus szituációs szemantika alkalmas alaphoz tűnik (vö. pl. KAMP: 1985.).

(b) Az egyes mondatok közötti viszony meghatározása

Ebből a szempontból az elsődleges annak meghatározása, hogy az egymást követő mondatok által leírt események milyen valós sorrendbe állítandók. Van egy viszonylag jól meghatározható, univerzálisnak tekinthető elv, miszerint az egymást követő mondatok egymás utáni eseményeket jelölnek; vö. (2).

(2) Péter hazaérkezett. Kinyitotta az ajtót. Belépett. Levette a kabátját. Elolvasta a leveleit.

Ez az elv azonban bizonyos megszorítással él: abban az esetben, ha ezt valami más elv nem bírálja felül. Így ezek közé tartozik az időviszonyok lexikális jelölése (*miután, azelőtt, hogy, korábban, egyszerre* stb.), továbbá olyan igék alkalmazása, amelyek nem fejezhetnek ki egymásutániságot; vö. (3).

(3) Péter ült a székben. Olvasott. A tévé hangosan szólt. Őt ez nem zavarta.

Bár a magyarban sok esetben a morfológia segítségünkre van (az *ige + igekötő* kombináció és sorrend mellett legtöbbször nem jelölhető egymásutániság), ezt általánosságban mégsem állíthatjuk, vö. (4).

(4) Csengettek. János nyitotta ki az ajtót.

Ugyancsak nem világos az időbeli viszony a következő mondatban.

(5) Kati elkésett a zongoraóráról. Otthon megint megszidták.

Ebből a két mondatból ugyancsak nem derül ki, melyik esemény történt hamarabb. Azaz nem lehet tudni, hogy az elkésés következménye volt-e a szidás, vagy fordítva, a szidás következménye volt-e az elkésés. Nyilvánvaló, hogy a szöveg még hosszabb szakaszának ismerete szükséges az ilyen kapcsolódások egyértelmű meghatározásához (a fenti kijelentések bővebb logikai összefüggéseit vizsgálja a kauzalitás szempontjából BÉKÉSI: 1990.).

Az alábbi szövegrészlet elemzése érdekes összefüggésekre mutat rá.

(6) Péter kinézett az ablakon. Ellopták az autóját.

A két mondat között a felszíni sorrenddel ellentétes időrendi viszony áll fenn. Vajon milyen mechanizmus működése alapján tudjuk ezt, ráadásul általában, univerzálisan? Modellálva a helyzetet, amelyben e két mondat elhangozhat, a következőket látjuk: az *ellopták az autót* azt jelenti, hogy egy adott időben az autó nem volt a rendes helyén. A *kinézett az ablakon* implikálja a „látást”-t, azaz normális esetben elvárható, hogy a nézést valaminek a megpillantása kíséri. (6)-ban az *ellopták az autóját* jelzi, hogy az autó nincs a helyén, azaz azt *nem lehet látni*. Ebből következtetünk arra, hogy az autó ellopása megelőzte az ablakon való kitekintést.

Ezzel az informális leírással azt kívántam bemutatni, hogy a szöveg megértéséhez (amin jelen esetben az események időbeli viszonyának meghatározását értettük) nem elegendő egy szótár és egy mondattani elemző, hanem szükség van egy *pragmatikai tudásbázisra* is, amihez folytonosan fordulunk, amikor a különböző kijelentések közötti *logikai* összefüggéseket kívánjuk feltárni.

A fentebb javasolt szöveggeneráló modellben egy ilyen tudásbázismodulnak kitüntetett helye kell hogy legyen, funkcionálisan hasonlóan, mint egy generatív szintaxismodellben a Logikai Formának.

(c) *A szavak jelentésének a meghatározása*

Ez az a terület, ahol a formák bármilyen mélységű figyelembevétele is a legkevésbé nyújt segítséget, márpedig a fordítás során éppen ez a lépés hagyható ki a legkevésbé. A jelenlegi fordítási rendszerek itt kötődnek leginkább a mondathatárokhoz: az adott szó jelentésének a kiválasztásához feltételezik bizonyos argumentumszerkezetek felszíni megvalósulását, és azt választják, amelyik kielégítő teljességgel manifesztálódik.

Ugyanakkor éppen ez az a terület, ahol a legtöbb meglepetés érheti a fordítót: a háttér ismerete nélkül például igencsak rosszul sikerülhet egy olyan ártatlannak tűnő mondat fordítása is, mint (7).

(7) Ez csak egy kacsza volt.

Ahhoz, hogy kiderüljön, hogy a *kacsza* szó egy állatot vagy egy hamis hírt jelöl-e, egy talán elég tekintélyes szövegelőzmény ismerete és feldolgozása szükséges. Úgy tűnik, hogy a fentebb említett szituációs szemantikai modellek alkalmazása itt is közelebb vihet bennünket a megoldáshoz.

(d) *Egy szövegnek a fordításhoz szükséges és elégséges mértékű megértése*

Ez a következőt jelenti: ha egy szövegben vannak kétértelműségek, alul-meghatározottságok, homályosság, amelyek valamilyen szerkezettel ugyanolyan mértékben átadhatók a célnyelvi fordításban, akkor nem kell törekedni e kétértelműségek stb. feltárására. Így például (8)

(8) A fiúk megették az almát.

nem világos abból a szempontból, hogy vajon egy vagy több almáról van-e szó összességében. Ha a célnyelvben e homály fennmaradhat, akkor nem kell mélyebbre tekintenünk abban a reményben, hogy ott ez jobban meg van határozva. (A „mélyebbre tekintés” egyébként csupán azt jelenti, hogy a nyelvészetben túlra kellene tekintenünk.)

Abban, hogy konkrét esetekben eldönthessük, milyen részletességű elemzés lesz elegendő, komoly segítséget jelenthet az ilyen és hasonló szerkezetek univerzális igényű leírása, azaz egy lista az *univerzálisan homályos* szerkezetekről. E lista pedig legmegbízhatóbban egy, a korábban már említett univerzális modell megalkotásával jöhet létre.

3. A fenti gondolatok csupán töredékesen tükrözik ama gondjainkat, amelyek a számítógépes fordítás során jelentkezhetnek. Az is valószínű, hogy a gépi megközelítés számos praktikus megoldást tud kínálni a konkrét esetekben. Mindez azonban nem helyettesítheti az általános irányban végzett kutatásokat, hogy a szöveg minél univerzálisabb jellemzőit tárhassuk fel. Úgy vélem, hogy bizonyos fogalmi tisztázásra még továbbra is szükség van magának a „szövegnek” az értelmezésében, annak érdekében, hogy a „gondolat” („értelem”) és a szöveg jelentése közötti különbség még markánsabban kitűnjön. Erre azért is van feltétlenül szükség, hogy el tudjuk határolni a szövegnyelvésztől a más diszciplínák segítségével talán az egzaktitás másfajta kritériumaival megvalósítható feladatokat.

Újólá hangsúlyozni szeretném, hogy a szöveg nyelvészeti vizsgálatához egy egzakt, a formákból kiinduló, univerzális elveken alapuló szöveggeneráló modellre van

szükség, ami magába ötvözi a formális mondatmodellek szemléletét és eredményeit éppúgy, mint szimbolikus logikai és szituációs modelleket (ez utóbbira lehet egy példa KAMP: 1985.); végül egy, a mesterséges intelligencia kutatása során létrehozott tudás-reprezentációra alapuló tudásbázisra támaszkodik.

Bár ez egy olyan programnak ígérkezik, ami nem egyik napról a másikra valósulhat meg, azért távolról sem előzmények nélküli, és már menet közben is sok közvetlenül és gyakorlatibb területeken is várható egyéb eredményt hozhat. Ami azonban talán a legfontosabb: hozzájárulhat a nyelvi forma legösszetettebb szerveződésének, a szövegnek a még teljesebb megismeréséhez.

Irodalomjegyzék

BÉKÉSI Imre:

1990. Az ellentétesség és a kauzalitás szerkezetalkotó összefüggésének logikai háttere, in: PETÓFI S. János—BÉKÉSI Imre (szerk.): *Szemiotikai szövegtan I. A szövegtani kutatás néhány alapkérdése*, Szeged, 23—28.

CHOMSKY, N.:

1981. *Lectures on Government and Binding*. Dordrecht.

PETÓFI S. János:

1990. Szemiotikai textológia — Didaktika, in: PETÓFI S. János—BÉKÉSI Imre (szerk.): *Szemiotikai szövegtan I. A szövegtani kutatás néhány alapkérdése*, Szeged, 7—21.