

MTA Számítástechnikai és Automatizálási Kutató Intézet Budapest

81639
259

MAGYAR TUDOMÁNYOS AKADÉMIA
SZÁMÍTÁSTECHNIKAI ÉS AUTOMATIZÁLÁSI KUTATÓ INTÉZETE

A magyar nyelv
elemzése
számítógéppel

(Tervek egy természetes nyelvű interfészhez)

Bach Iván
Farkas Ernő
Naszódi Mátyás

A kiadásért felelős:

Dr. KEVICZKY LÁSZLÓ

Főosztályvezető:

CSABA LÁSZLÓ

A tanulmány az OMFB 911-13-1 szakfeladat 61-42-044/86 szerződése alapján készült.

ISBN 963 311 230 3

ISSN 0324-2951

MTA SztAKI 1987. aug. 1.

SZÁMALK

Tartalom

Bevezetés	5
A természetes nyelv feldolgozásának alapproblémája	5
A magyar nyelv szabályai	7
A tervezett program felépítése	8
Az eddig elért eredmények.	10
A magyar szavak morfológiai elemzése	12
A szó.	12
Szóelemzés	12
Morfológiai szeletelő.	13
Elvi modell.	13
Gyakorlati modell.	14
Kivitelezés.	20
Tapasztalatok.	20
A szótár	21
Morfológiai jellemzők.	21
Szintaktikai jellemzők	22
Szemantikai jellemzők.	23
A magyar szófajok morfológiai - szintaktikai felosztása.	24
A szintaxis alapjai.	27
Igei szerkezetek	28
Formális definíció.	30
Névszói kifejezés.	34
Formális leírás	35
Egyeztetés	38
Az alany, a tárgy és az ige	38
A birtokos szerkezet egyeztetése.	41
Vonzatok	43
Argumentum és vonzat.	44
Kötelező és opcionális argumentumok, defaultok.	46
A szintaxis analízis megvalósítása	47
Hivatkozások	50
Szemantikus reprezentáció és szemantikus jegyek.	53

Mellékletek

Melléklet a kísérleti morfológiai elemző toldalékairól	55
Kísérleti szótár	59

A magyar nyelv elemzése számítógéppel

(Tervek egy természetes nyelvű interfészhez)

Bevezetés

Az előző „A magyar nyelv alkalmazása a számítástechnikában” című előtanulmányunkban áttekintést adtunk arról, hogy elsősorban angol és orosz minták alapján, hogyan képzelhetjük el egy nyelvészeti alapokon nyugvó természetes nyelvi interfész felépítését.

Mi korábban programnyelvek fordító programjaival programnyelvi és természetes nyelvű szövegszerkesztő programokkal foglalkoztunk. Ezen tapasztalatok birtokában állíthatjuk, hogy egy természetes nyelvű interfész program létrehozása nem látszik technikailag különösebben nehéz feladatnak. Az áttekintett programok kapcsán sehol nem talákoztunk számunkra teljesen új és meglepő programtechnikai megoldásokkal. Annál nehezebbnek látszik a feladat tartalmi oldala, a magyar nyelv szabályainak lefordítása olyan formális szabályokká, amelyeket azután gépre lehet vinni.

A természetes nyelv feldolgozásának alapproblémája

Amikor Noam Chomsky felállította a nyelvek leírására szolgáló első 1957-es modelljét egy olyan formalizmust állított fel amely alapján a jelsorozatok egyértelműen két csoportba sorolhatók: vannak olyanok, amelyek mondatok, és így a nyelvhez tartoznak, és vannak olyanok, amik nem, Mi, akik hosszú évekig dolgoztunk programozási nyelveket fordító programokon, tudjuk, ez a felosztás így megvalósíthatatlan. Még a programozási nyelvekben is, amelyek pedig gondosan megtervezett emberi alkotások, vannak olyan konstrukciók, amelyeken a szakemberek vitáznak, hogy vajon legálisak és értelmesek-e, Hatványozottan így van ez a természetes nyelvek esetén. A mondatok, szavak, szóalakok, kifejezések döntő többségéről megállapítható ugyan, hogy helyes-e vagy helytelen, de igen széles sáv marad még közöttük, amelyről vitatkozni lehet.

Azt hihetnénk esetleg, hogy ez a bizonytalanság talán nyelvi képzetlenségünkből fakad, és a szakemberek számára kevés az ilyen. probléma. Az esetek többségében ennek szinte az ellenkezője igaz, Mint ismeretes a nyelvészeti munkákban a helytelen alakot az előtte álló *-gal jelölik meg, a vitathatót pedig, a szerző indulatától függően, egy vagy több kérdőjellel. Tapasztalatunk szerint minden nyelvészeti munkában jónéhány kérdőjel akad. Ennek a bizonytalanságnak több forrása van, az egyik a különböző környezet, neveltetés, stb., a másik a nyelv állandó változása.

A nyelv ilyen laza felfogásával élesen ellentétben áll a számítógép alkalmazása, ha azt mondatok analizálására vagy szintézisére akarjuk felhasználni. Ebben az esetben maga a program élesen eldönti, hogy melyik mondatot tekinti helyesnek és értelmezhetőnek, és melyiket hibásnak. A számítógép számára vitatható mondat nincs.

A fentiekből következik, hogy a számítógépes program nem tudja és nem is akarja leírni a nyelvet olyan mélységben (fejlődésében és ellentmondásaiban), ahogy azt a nyelvészet teszi, hanem egy modellt készítünk, amelyről eleve tudjuk, hogy a magyar nyelvet csak részben fedí le. Már most biztosak vagyunk benne, bizonyos nyelvi konstrukciókat (mint pl. a mondatátszövés, régies, de még használt igealakok, stb.) nem fogunk kezelni.

Nyelvi modellünk nemcsak a teljes nyelv átfogása tekintetében lesz korlátozott, hanem feltehetően meglehetősen durva is lesz, és így a helyes mondatok közé valószínűleg becsúsznak olyan mondatok is, amelyek noha hibásak, a program mégis jó mondatként értelmezi őket. A programozási nyelvek körében ortogonalitásnak nevezzük azt a tulajdonságot, ha egy szabály kivételek és más szabályokra való tekintet nélkül alkalmazható. Ortogonális szabályrendszerre nyilvánvalóan könnyebb programot írni. A természetes nyelv viszont (mint azt már a morfológiáról szóló fejezetben is láthatjuk) egyáltalán nem ortogonális. A fontosabb kivételeket és a szabályok egymásra hatását természetesen kezelni kell, de az összes kivétel és egymásra hatás kiderítése és a programba való beépítése megvalósíthatatlannak látszik.

Felmerülhet a kérdés, hogy egyáltalán megoldható-e a feladat, hiszen ha a modellt túl bőre szabjuk, sok nem korrekt mondat is bele fog kerülni, ha viszont túl szűk sok helyes magyar mondat kimarad belőle. Erre azt a választ adhatjuk, hogy mi a gyakorlat számára akarunk programot készíteni, és ebből a szempontból a számítógépes kommunikáció két alfeladatra bomlik: a bemenő szöveg analízise és megértése, és a válaszok generálása. Ha abból indulunk ki, hogy a természetes nyelvű interfészt azok akarják használni, akiknek ez a nyelv a természetes, akkor feltételezhetjük, hogy a felhasználók ritkán fognak nyelvtani hibákat elkövetni a géphez forduláskor, viszont feltehetőleg nem tetszene nekik, ha a gép rossz magyarsággal válaszolna. Így hát világos, hogy két különböző modellt kell felállítanunk a két feladatra, egy tág lehetőségeket adót a bemenetre, és egy szűkebb választékát, de nyelvtanilag pontosabbat a válaszokra.

Mi jelen munkánkban csak az analízissel foglalkozunk, és feltételezzük, hogy a bemenő szöveg nyelvtanilag alapjában véve helyes. Ezért bizonyos vizsgálatokat a modellünkből teljesen kihagytunk. Tapasztalatunk szerint teljesen fölösleges például a szavakat magánhangzó-illeszkedés szempontjából vizsgálni, noha ennek szabályai és a kivételek ismertek. Ilyen hiba a beírt szövegekben praktikusán nem fordul elő, ezért a ráfordított emberi és gépi munka teljesen felesleges lenne.

Általában is azt az elvet követjük, hogy a nyelvtani szabályokból csak annyit veszünk figyelembe, amennyi a megértéshez szükséges, és elhanyagoljuk azokat, amelyek az analízis számára nem adnak újabb információt.

Lehet persze, hogy néhány ponton feltételezésünk csal, és a későbbiek folyamán bizonyos nyelvi jelenségek vizsgálatára mégis rákényszerülünk, amelyeket korábban feleslegesnek hittünk. Igyekszünk modellünket úgy megcsinálni, hogy a későbbiekben könnyen módosítható és fejleszthető legyen. Ahol lehet arra törekszünk, hogy a nyelv modell és a hozzákapcsolódó program ne folyjon össze. Ha a táblázatokban leírt nyelvmodellt módosítjuk, szűkítjük vagy bővítjük, akkor a programon ne kelljen változtatni, vagy csak újabb jól definiált interfészű függvényeket kelljen hozzávenni.

A magyar nyelv szabályai

Ha lett volna a magyar nyelvnek olyan viszonylag teljes formális leírása, amelyből kiindulhattunk volna, akkor a létrehozandó eszközöket bizonyára ehhez igazítottuk volna. Így azonban, azt az utat jártuk, hogy megpróbáltuk összegyűjteni a magyar nyelv szabályait, és megkíséreltük végiggondolni, hogyan lehet az összefüggéseket a gépben ábrázolni, és a szabályokat géppel ellenőrizni, végül megkezdtük a szabályok formalizálását. Erre olyan formalizmust vezettünk be, amely alkalmasnak látszik a magyar nyelv géptől és programtól független absztrakt leírására, és ugyanakkor ebből a formális leírásból generálható az a program, amely elvégzi a természetes nyelvű szöveg analízisét.

A leírás alapjául a következő forrásokat használtuk:

Bencédy József - Fábrián Pál - Rácz Endre - Velcsov Mártonné:
A mai magyar nyelv, Hatodik kiadás 1985.

A magyar helyesírás szabályai, Tizenegyedik kiadás 1986,

É. Kiss Katalin:

A magyar mondatszerkezet generatív leírása, 1985.
Configurationality in Hungarian, 1987.

Hadrovios László:
A funkcionális magyar mondatan alapjai, 1969.

Prószéky Gábor, Kálmán László, Kornai András és társaik számos
cikke

Antal László számos műve

Kiefer Ferenc:
Az előfeltevések elmélete, 1985.

Pléh Csaba számos munkája az anafórikus kapcsolatokról

Felhasználtuk továbbá:

a nagy Magyar Értelmező szótárat, a Gyakorisági szótárat

Papp Ferenc: Szóvégmutató szótár

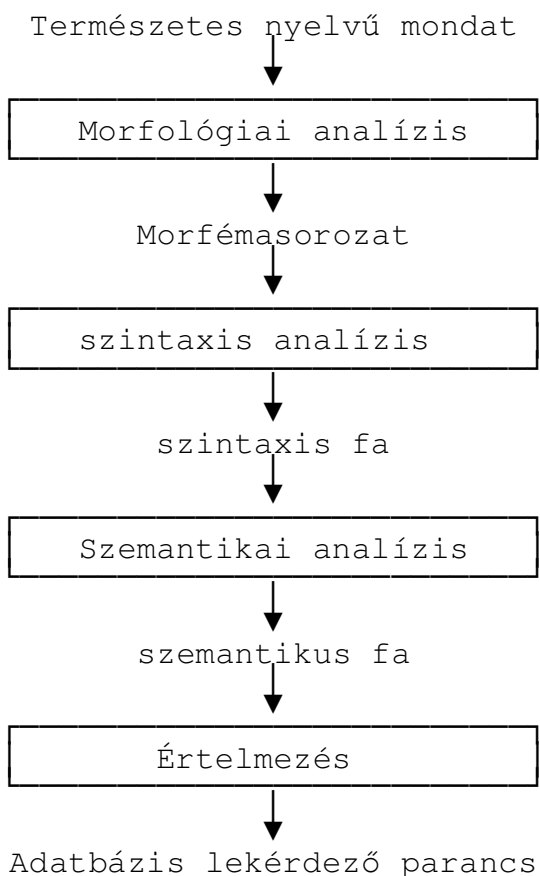
Apreszjan Ju. D. - Páll Erna:

Orosz ige - magyar ige vonzatok és kapcsolódások, 1982.

Továbbá köszönetet kell mondanunk mindazon nyelvészeknek, akik segítségünkre voltak: elsősorban Füredi Mihálynak, Prószéky Gábornak. valamint Varga Dénesnek.

A tervezett program felépítése

Munkánkban a nyelv különböző síkjain különböző mértékben haladtunk előre. Hogy erről részletesebben beszélhessünk, vissza kell idéznünk az előtervében lefektetett terv diagramját, amely programunk vázlatos felépítését mutatja:



A morfológiai analízis a bemenő szöveg szóalakjait morfémák (szótövek és végződés) sorozatára bontja. Az egyes morfémákhoz egy-egy kódszámot rendel hozzá, amely megmutatja az adott morféma grammatikai szerepét. (Pl. ige, főnév, melléknév, stb. illetve kijelentő módú, jelen idejű egyes szám 5. személyű igerag, határozói igenév képzője, névszói többes szám jel, melléknévi fokozás, stb.) A felbontás során természetesen foglalkozni kell a teljes hasonulással és más hasonló jelenségekkel is.

A szintaxis analízis egy mondat morfémasorozatát egy fa alakú gráffá alakítja. A fa levelein a szótövek illetve a végzések állnak. A nyelvtanilag összetartozó elemek egy csomópontban kapcsolódnak össze, ezek még nagyobb nyelvtani egységeket (szintagmákat) alkotnak, és azok még nagyobbakat, amíg végül is létre jön a mondat. A szintaktikus fa a mondat mondattani szerkezetét írja le,

A következő fázis ezt a fát egy olyan fává alakítja át, ami már a mondatban szereplő tartalmi kapcsolatokat írja le. Az átalakítás során az egyik feladat, hogy fel kell oldanunk olyan hivatkozásokat (például a névmásokat), amelyekkel egyszerűsített formában hivatkozunk bizonyos korábbi dolgokra. A másik feladat az, hogy bizonyos egyszerűsítéseket, standardizálásokat hajtsunk végre a mondat ábrázolásán. Például, véleményünk szerint, ha azt mondjuk, hogy „*neki el lehet mennie*” vagy azt mondjuk, hogy „*elmehet*” a két kijelentés azonos, noha külső formáját tekintve meglehetősen sok eltérést vehetünk észre. Kézenfekvő, hogy a szemantika szintjén a két kijelentést azonosan ábrázoljuk. Ezenkívül, ellenőrizni kell, hogy a mondat nem csak nyelvtanilag helyes, hanem a benne leírtak végre is hajthatók. Ezeket a tevékenységeket együtt nevezzük a továbbiakban szemantikus analízisnek.

(A szemantikáról többféle értelemben szokás beszélni. Nem vonjuk kétségbe, hogy valamilyen szinten meg lehet adni a teljes magyar nyelv szemantikus leírását, de ez a leírás igen általános lenne, és utána egy külön lépésben kellene kiválasztani a helyes magyar mondatok közül azokat a mondatokat, amelyek az adott feladatnál értelmesen felhasználhatók. Mi eleve csak azokat a mondatokat tekintjük helyesnek, amelyek az adott helyzetben használhatók, és egy, az adott témához és feladathoz kötött, részletes szemantikus ábrázolásra gondolunk.)

A program utolsó lépése a mondat szemantikus reprezentációját adatbázist lekérdező parancsokká fordítja.

Az eddig elért eredmények

Mint azt fentebb említettük, a különböző területeken a munkában különböző mértékben jutottunk előre. Ennek több oka is van. Az első ok, hogy a munka fázisai, mint azt az ábra is mutatja, szekvenciálisan egymásra épülnek, és amíg a megelőző fázis nem állít elő elfogadható eredményt, addig a következő fázis kipróbálása nem kezdődhet meg.

Ez fokozottan így van esetünkben, azért is, mert a magyar nyelvre vonatkozó számítógépes tapasztalatok hiányában nem tudjuk jól megítélni az egyes problémák súlyát.

A másik ok az, hogy még mindig nem döntöttük el, mi lesz az a konkrét adatbázis, amelyre az interfészt készítjük. Így mindazoknál a pontoknál ahol a szemantika szóba jön, csak igen általános elképzeléseink vannak, mivel terveink szerint a szemantikus jegyeket a téma és a várható feladatok függvényében akarjuk megválasztani.

Eredményeinket a következőkben foglalhatjuk össze:

A morfológiai analízis megvalósítására elkészült két kísérleti program, amellyel sok kísérletet hajtottunk végre, és meglehetősen jó eredményeket értünk el. Ezen tapasztalatok alapján készült el a most már véglegesnek szánt morfológiai elemző rendszerterve, és valószínűleg ez év szeptemberére készen lesz a végleges program is. Ez a program a korábbiaknál több végződést (azon belül képzőt) lesz képes kezelni (így a szótár mérete csökkenhet), és képes lesz kezelni egy sor olyan jelenséget, mint a teljes hasonulás, a törövidülés, a harmadik azonos mássalhangzó kiesése, stb. (így a szótárban csökkenhet az egy szóhoz tartozó tőalakok száma).

Elkészült egy kísérleti szótár, amely körülbelül 2500 tőalakot tartalmaz és ezek a később ismertetett elvek alapján szófajokra vannak osztva. Ez a szótár szükséges a morfológiai elemzéshez, és az itt kapott szófaji kódokból indul ki a szintaktikai elemzés. A szótárnak kb. 700 eleme esik egybe a gyakorisági szótár leggyakoribb szavaival.

Elkészült a magyar nyelv szintaxisának formális leírása kissé módosított attribútumnyelvtan segítségével. A nyelvész kollégák ezt a leírást elolvasták, és elfogadhatónak tartják. Ennek ellenére biztosak vagyunk benne, hogy a leírás tele van hiányosságokkal, de hogy mik ezek, és melyiknek milyen súlya van, majd csak szeptember táján az első próbafuttatások során állapíthatjuk meg. A próbához meg kell írni az elemző program három szükséges alkotórészét: a kötött szórendű szabályok kezelését, a szabad szórendű szabályok kezelését és a vonzatok kezelését, valamint össze kell állítani a vonzatszótárat. Eddig csupán a kötött szórendű szabályok ellenőrzésére történtek kísérletek.

Pontos (bár nem formálisan leirt) terveink vannak az egyeztetések elvégzésével kapcsolatban.

A vonzatok és a hivatkozásokkal kapcsolatos terveink általános jellegűek. Ezek pontosabb leírásához szükség lenne ugyanis a szemantikus jegyek halmazának definiálására.

A következő fejezetekben összefoglaljuk a felmerült problémákat, az ezekkel kapcsolatos döntéseket és terveket. Hangsúlyozni kívánjuk, hogy a magyar nyelv itt megadott leírása csak egy a számos lehetséges felfogás közül, és a gyakorlat fogja megmutatni, hogy mennyire használható. Előbb azonban ki kell küszöbölni a benne található hibákat és hiányosságokat.

A magyar szavak morfológiai elemzése

A következőkben a feladat meghatározása után egy elvi (alapvetően generatív) modellt vázolunk fel, majd az általunk kidolgozott gyakorlati analitikus modellt ismertetjük. Végül néhány tapasztalati tanulságot vonunk le.

A szó:

Szónak (a nyelvészek szóalaknak hívják) nevezzük az írott szöveg egybeírt (csak betűből és kötőjelből álló, más írásjelet, szóközt nem tartalmazó) darabját, mely vagy maga egy szótő, vagy egy szótőből a későbbiekben leírandó szabályokkal létrehozott egység. (A számjeggyel és más írásjellel jelzett egységeket nem tekintjük szónak.) A magyar nyelv ún. agglutináló nyelv, ami azt jelenti, hogy a szótőhöz végződéses tapadhatnak, és néhány előtag is kapcsolódhat a szóhoz. Ezenkívül a szóösszetétel is lényeges szóalkotási mód. A szavak különböző szempontokból osztályokba sorolhatók (lásd később: szófajok, ragozási típusok). A toldalékok (előtagok) szemantikai módosításon kívül a szavakat egyik szóosztályból a másikba viszik. Különböző szavaknak lehet azonos alakjuk. A szavak szintaktikai és szemantikai tulajdonságaival későbbi fejezetben foglalkozunk.

Szóelemzés:

A szóelemzés célja, hogy megállapítsa azt, milyen szótőből és milyen toldalékolással (és előtagokkal), esetleg milyen szóösszetétellel jött létre az adott szó (pl. legelemibb: *leg-elem-i-bb*). Egyes nyelvekben, mint például az angolban, ahol a toldalékolás és előtagok alkalmazása nem létezik vagy szegényes, a szóelemzés gyakorlatilag egy szótár alkalmas kiépítésével megoldható. Nem így a magyarban, ahol a toldalékolt szó is kaphat toldalékot. (Ilyen esetekben relatív szótőről beszélhetünk.) A toldalékok szekvenciája eléggé kötött. A toldalékok (előtagok) is osztályokba sorolhatók aszerint, hogy milyen szóosztályokra alkalmazhatók, és milyen szóosztályba kerül az új szó. A különböző toldalékoknak lehet azonos az alakjuk (pl. *kutyá-nak lát-nak*). Egyes toldalékok módosíthatják a relatív szótövet (tőváltozatok, hasonulások). Ennek alapján négy feladatot kell megoldani:

- A toldalékok (előtagok) levágása.
- A toldalékolás következtében végbement tőváltozások „visszacsinálása”.
- A toldalékok és szóosztályok egyeztetésének ellenőrzése.
- A szótövek és a toldalékok és azok osztályának meghatározása.

Az első hárommal a morfológiai szeletelő foglalkozik, míg a negyedik alapvetően szótár segítségével történik. Bár a négy funkció szétválasztható, a megoldás során összefonódnak, egymást kiegészítik.

Morfológiai szeletelő:

A morfológiai elemző modul a természetes nyelvi elemző rendszer szerves része. Célja, az írott szavak morféimákra bontása a további (szintaktikai, szemantikai) elemzést megelőzően. Mivel a teljes nyelvi rendszer interaktív ember-gép kapcsolat részére készül, a válaszüidők rövidsége érdekében lényeges az elemző nagy sebessége. A rendszer az általunk köznyelvnek tartott alakokkal foglalkozik, és nem foglalkozik régies, vagy tájnyelvi alakokkal, de alkalmas bővítésére, szűkítésre módosítására is, ha ilyenre lenne szükség. A kidolgozott módszer alkalmas morfológiai szabályok leírása alapján gyors elemző előállítására.

Elvi modell:

A morfológia a következő hat fogalommal operál: ABC, Szóosztályok, Tőváltozások, Tőtár, Előtagtár, Toldaléktár.

ABC: a nyelvre jellemző. (Az ABC rendezettsége számunkra nem lényeges.)

Szóosztályok: a szavak csoportosítása szófaj, hangrend, ragozástípus alapján.

Tőtár: tartalmazza az összes használható szó tövét, jelölve, mely szóosztályba tartozik.

Toldaléktár: tartalmazza az összes használható toldaléket jelölve, hogy milyen osztályú szóra (szófajra, hangrendre stb.) alkalmazható, milyen osztályú szófajt állít elő, és milyen jellegű tőváltozást idézhet elő.

Előtagtár: tartalmazza az összes használható előtagot jelölve, hogy milyen osztályú szóra (szófajra, hangrendre stb.) alkalmazható, és milyen osztályú szófajt állít elő. (A magyar nyelv előtagjai nem okoznak tőváltozást.)

Tőváltozások: bizonyos szótövek bizonyos toldalékok esetén megváltoznak, Ilyen változástípusok a magánhangzó kiesések, betoldások, rövidülések, nyúlások, hasonulások (ide sorolható az is, amikor nem a tő, hanem a toldalék módosul) stb. Ezek a változások magyar nyelvben mindig a (relatív) szótő utolsó egy-két betűjét érinti.

Toldalékolás: egy adott (relatív) szótőre alkalmazható egy toldalék, ha a (relatív) szótő szóaszálya megegyezik a toldalék alkalmazhatósági szóosztályával, Ekkor először végrehajtjuk a tövön a toldalékhoz tartozó tőváltozást, majd az így kapott módosított tő után fűzzük a toldaléket, és eredményül a toldalék által előírt szóosztályba tartozó szót kapunk.

Az előtag alkalmazása: a fentihez hasonlóan definiálható.

Elemzés: Az elemzés feladata, hogy egy adott szóról megállapítsa, milyen szótőből, milyen toldalékolási (előtag) szekvencia, esetleg milyen szóösszetétel segítségével jöhet létre. Könnyen belátható, hogy ha az üres toldalék nem okoz rekurzivitást, valamint a szótőváltozások a toldalékkal mindig növelik a szó hosszát, márpedig ez a magyar nyelvre igaz, akkor az elemzés **egy véges automatával véges lépésben elvégezhető.**

Gyakorlati modell:

A fenti megállapítás nem jelenti azt, hogy az elemzés minden részét a gyakorlatban is végesautomata-modell alapján kell kivitelezni. Ha az összes szóba jövő szótövet is véges automatával szeretnénk felismerni, minimális szótár esetén is egy többtízezer állapotú véges automatára lenne szükség, amelynek mind a létrehozása, mind a karbantartása teljesíthetetlen feladat volna.

Hasonlóan rossz megoldás lenne, ha az elemző mindent keresgetéssel, próbálgatással kezelne (ilyen elemzők léteznek, és elsősorban pontos nyelvhelyességi illetve demonstratív céllal készültek), hisz a hatékonyság nem lenne kielégítő.

Az elemzés történhetne a szótótól a toldalékok egyenkénti levágásán keresztül balról jobbra, de mások tapasztalatával megegyezően úgy véljük, hogy ez elég gyakori szótárhoz való fordulással és több hibás részelemzéshez vezetne.

Problémát okozhat a tövátváltozások kezelése is, Az Apreszjan és csoportja által létrehozott orosz elemzőben nincs algoritmikus tövátváltozás-kezelés, hanem a szavak tövátváltozataival együtt szerepelnek a szótárban. Ez a módszer a magyar nyelv esetén azért sem követhető, mert toldalékolt tő is szenvedhet tövátváltozást.

Legmegfelelőbbnek az a megoldás tűnik, hogy a toldalékolási szabályokkal ellentétes irányban, a szóalakok végéről egy megfelelő véges automata leválaszt egy végződést, amely egy (párszáz elemű) toldalékosztály egyik lehetséges eleme, majd végrehajtja az esetleges tövátváltozás inverzét; ha előtag is lehet, azt leválasztja a szó bal oldaláról, végül a lehetséges tövet, illetve töveket a több ezer elemű szótárból keresi ki a morfológiai-lexikai elemző. Ily módon a szótári keresés minimálisra csökkenthető.

Lényeges módosítások vezethetők be az algoritmus gyorsítása érdekében annak tudatában, hogy elemzőt készítünk, és feltételezhetjük, hogy nyelvtanilag helyes szavakat kell vizsgálni. További egyszerűsítésekre is találunk lehetőséget a magyar nyelv esetében (pl. az előtagok nem okoznak tövátváltozást).

A szavak és toldalékok egyeztetésénél mi alapvetően csak a szófaji egyeztetést vesszük figyelembe, és néhány a szó végét érintő hangtani egyeztetést (pl. a múlt idő milyen betű után milyen alakú lehet). A hangrendi és egyéb ragozási szabályok egyébként sem egészen egyértelműek (*mászom-mászok, mondta-mondotta, tőszavak-tőszók, adapterok-adapterek ...*).

Sok helyen érdemes volt eltérni a nyelvészetileg szokásos nyelvtantól és létrehozni egy pragmatikus szóképzési modellt annak tudatában, hogy a morfológia csupán egy előfázisa a „megértésnek”, és mi egy nyelvhelyességi elemző programot kívánunk írni. Ily módon a modell a következőképpen módosul:

Szóosztályok: Az osztályozás alapvetően szófaji, illetve toldalékolási tulajdonságok szerint történik.

A szótári elemzés szempontjából három fő osztályba sorolhatók a szavak: ige, névszó, egyéb.

A morfológiai elemzés során további alosztályokat használunk:

Az ige lehet ragozott és ragozatlan, igekötős és igekötő nélküli (a ragozott igék közé soroljuk a főnévi és határozói igeneveket is, mivel ezek után már nem lehet újabb toldalék).

A névszók felosztása aszerint történik, hogy milyen típusú toldalékot kapott (tehát milyen toldalék szelhető le), így névszói tő, különböző mértékben fokozott, többes számú, illetve birtokjellel ellátott, birtokos jellel ellátott és esetraggal ellátott.

Az egyéb kategóriába taroznak a nem toldalékolható tövek (pl. névelő, kötőszó),

Ezek részletesebb felhasználását lásd a toldalékolás szekvenciáinál.

Ettől függetlenül, a szavak ragozási és hangtani szóosztályokba sorolhatók, Ez utóbbiak közül legismertebb a hangrendi osztályozás, mely alapvetően meghatározó az illeszkedő toldalék kiválasztásánál, de más nehezebben elemezhető hangtani illeszkedések is szerepet játszanak. A hangrendi és ragozástípus szerinti felosztás elemzőnkben nem játszik szerepet, helyette elegendőnek bizonyult a szóalakok végét vizsgálni. (Lásd toldaléktár.)

Szótár: nemcsak szótöveket tartalmaz. Érdemes külön elemként felvenni az olyan szóösszetételeket, képzett szavakat is, melyek értelme nem vezethető le a szóalkotás módjából. A szótóhossz így sem haladja meg a 25 karaktert. Tartalmaz ezen kívül olyan ragozott alakokat is, melyek teljesen rendhagyó módon képződnek (pl. *az, annak*), és ezt a speciális ragozást nem érdemes bekódolni a toldaléktárba. Mindezen bővítések miatt a szótár várhatóan 30 - 50 százalékkal nő. A szóalakon kívül a szótár tartalmazza a szó fajtát, ragozástípusát. Praktikus okokból a szótár tartalmaz egy olyan információt is, hogy érdemes-e tovább darabolni az adott szóalakot. Részletesebb leírást lásd a Szótár című fejezetben.

Toldalék és előtagtár: tartalmazza az összes levágható toldalékot aszerint osztályozva, hogy hova tartozik az adott toldalékkal ellátott szó. Az esetleges kötőhangokat a toldalékok részeinek tekintjük. A toldaléktárban az is jelölve van, hogy milyen tőváltozást idézhet elő, és milyen típusú szó marad a toldalék levágása után. Ilyen értelemben a következő toldalék-típusok vannak:

Igeragok: ragozott ígéről vághatók le, és levágása után ragozatlan ige marad. Ez tulajdonképpen egy összevont toldalék, amely magába foglalja az időt, módot, személyt és számot (pl. *mond-ott-át-ok* helyett egyszerűen *mond-ottátok*). Ide soroljuk a főnévi igenévi képzőket, és a határozói igeneveket is (*mond-anom, mond-va*), mivel ezeket sem követhet további toldalék.

Ige-ige képzők: ragozatlan ígéről választható le, és a maradék is ragozatlan ige; csak a *-hat, -het, -gat, -get, -tat, -tat, -at, -et*.

Ige-névszó képző: névszói töről választható le, és a maradék ragozatlan ige (pl. *dob-ás*).

Esetragok: ragozott névszóról választható le, és a maradék ragozatlan névszó (pl. *tojás-t, ló-ként*).

Birtokos jel: ragozatlan névszóról választható le, és a maradék birtokos jel nélküli névszó (pl. *fül-é*)

Birtokjel-többszámjel: egy kategóriába vettük, mivel egymást kizáróan ugyanolyan pozícióban szerepelhetnek. Birtokos jel nélküli névszóról választható le, és a maradék fokozott tő (pl. *ház-ak, piros-aim*).

Középfok jele: fokozott töről választható le, és a maradék névszói tő (pl. *kék-ebb*). Vele együtt vágandó le a *leg-*, *legesleg* előtag.

Névszó-névszó képző: névszói töről választható le, és a maradék névszói tő (pl. *nyak-ú, kék-es*).

Névszó-igei képző: igtőről választható le, és a maradék fokozott tő (pl. *kék-ít*).

Igekötő: Előtag, mely ragozatlan igeről választható le, és a maradék igerő (pl. *le-húz*).

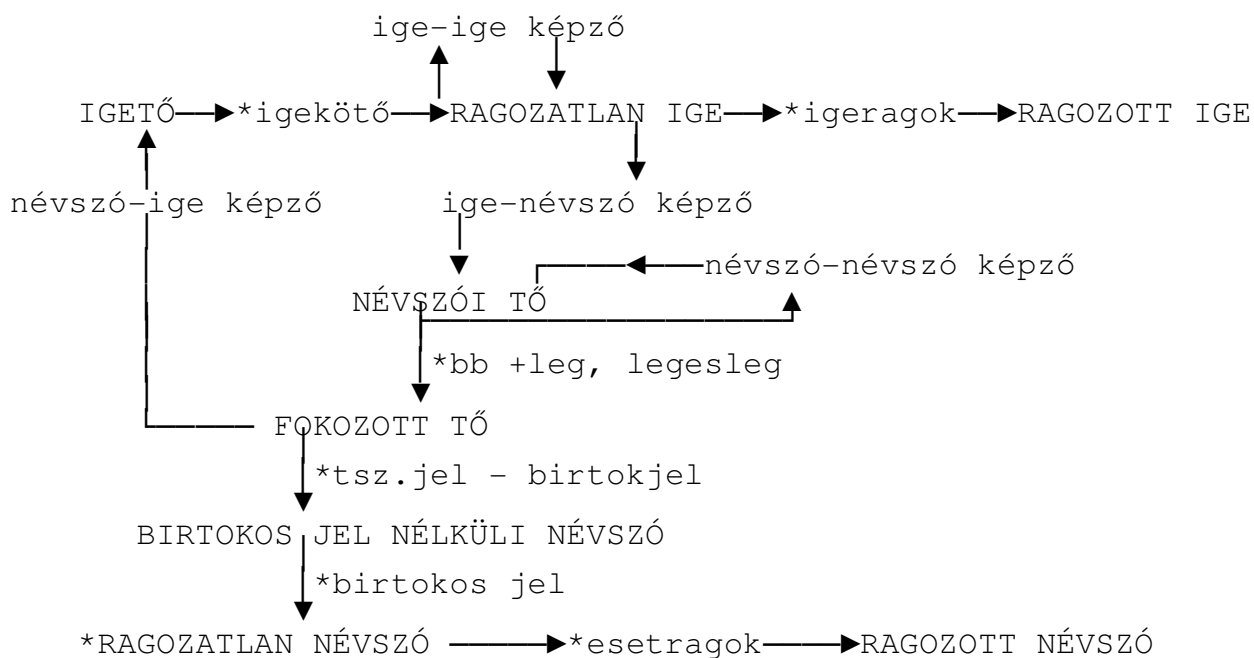
A toldaléktárban az is jelölve van a toldalékoknál, hogy a relatív szótó (az esetleges tőváltozás után) milyen betűre végződhet (pl. az igei múlt egy *t*-je csak bizonyos mássalhangzó után állhat). Ez redundáns információ a ragozási csoportok megjelölése mellett, de gyakorlatilag elégséges az elemzéshez a ragozási csoportok megjelölése nélkül is. Tapasztalatunk szerint elemzésnél a hangrendi osztályozás sem játszik szerepet. A gyakorlatban rendkívül ritkán fordul elő olyan szóalak, mely végéről formálisan úgy vágható le toldalék, hogy az így kapott, nem valódi relatív vagy abszolút szótó hangrendileg ne illeszkednek a toldalékhoz (mint például *balek bal-ek*). Ráadásul éppen az egyik leggyakoribb vita és félreírás abból származik, hogy a szavak hangrendje nem jól meghatározott (pl. *adapter-ok* vagy *adapter-ek*). Lényeges információ az is, hogy egy adott toldalékosztály előtt milyen szótóváltozás lehetséges. A toldaléktár teljesen gyakorlati céllal készült, ezért nem feltétlen követi a hagyományos magyar nyelvi toldalékolást. Külön toldaléknak tekintjük a kötőhangos és a kötőhang nélküli toldalékokat (*-t, -ot, -et, -öt*). Nem szerepelnek benne a régies, ritkán használt toldalékok. Kihagytuk az olyan toldalékokat is melyek szemantikailag nem elemezhetőek (*-ság, -ség, -odalom, -edelem* stb.), viszont szerepelhetnek benne mesterséges ragok, melyeket az egyszerűbb kezelés érdekében vezetünk be (pl. a *lás-sál* szóban a *-sál* külön igerag). Bizonyos toldalék-összevonásokat is alkalmazunk hasonló célból (pl. a *-nom* főnévi igenév egyes szám első személyű alakja szétválaszthatatlan toldalék).

Tőváltozások: A tőváltozások helyett jobb azok inverzeit kódolni, tehát azt, hogy pl. *irodalm* alakból *irodalom* szót kell visszatranszformálni bizonyos toldaléklevágások esetén). A korábbiak alapján nem kell minden tőváltozást felvenni, de azokat feltétlen, melyek a toldalékolt relatív szótóvek esetén is előfordulnak. Ha csak az általunk kezelendő toldalékokat tekintem, akkor ilyen (relatív) szótóváltozás három típusú van a magyarban:

a szóvégi magánhangzónyúlás (*fűzet-e, fűzet-é-ből*), az igei (*üt-het-j=üthess*) és a névszói (*jobb-val=jobbal*) hasonulások következményei. Természetesen más, könnyen algoritmizálható, gyakran előforduló töváltozásokat is kezelünk, ha ez egyszerűbb, mint bevinni a szótövtálozatoakat a szótárba. Ritkábban előforduló, nehezen algoritmizálható töváltozások kódolása helyett jobb, ha a szó tövét más formában is szerepeltetjük a szótárban (pl. *teher, terh-*). Bizonyos hasonulások könnyebben kezelhetők formálisan újabb toldalékok bevezetésével (pl. *lás-s*). Lényeges információ, hogy mely töváltozások inverzei kötelezőek, és melyek opcionálisak (pl. az imént említett felszólító mód előtti *s-t* változás opcionális, hisz *keres-s*, de a *kutyá-t a-á* változás kötelező. (Az *á*-ra végződő névszók olyan ritka kivételek, hogy egyszerűbb ezek ragozott alakját szótárba venni mintsem az *a*-ra végződőeket töváltozataival.)

A következő töváltozásokat kezeljük: szóvégi magánhangzónyúlás, szóvégi magánhangzó-rövidülés, belső magánhangzó-rövidülés, magánhangzó-kiesés, igei hasonulás, igei *t-s* változás, névszói hasonulás. A szótövtálozások inverzeit kódoljuk, jelölve, hogy az inverz művelet toldaléklevágásnál az adott helyzetben kötelező vagy nem. Külön figyelmet érdemel a hasonulásoknál és általában a toldalékolásnál a kettős és a kétjegyű betűk hasonulása, toldalékolása (pl.: *jobb-vá=jobbá*).

Toldalékok szekvenciái: Mint említettük, a toldalékok nem követhetik egymást tetszőleges sorrendben. Helyes szekvenciájuk abból határozható meg, hogy milyen szóosztályba kerül egy adott szó egy adott toldalékolás után. Az elemzőnkben egy egyszerűsített szófaji felosztást használunk, amelyet a következő generatív gráffal ábrázolhatunk.

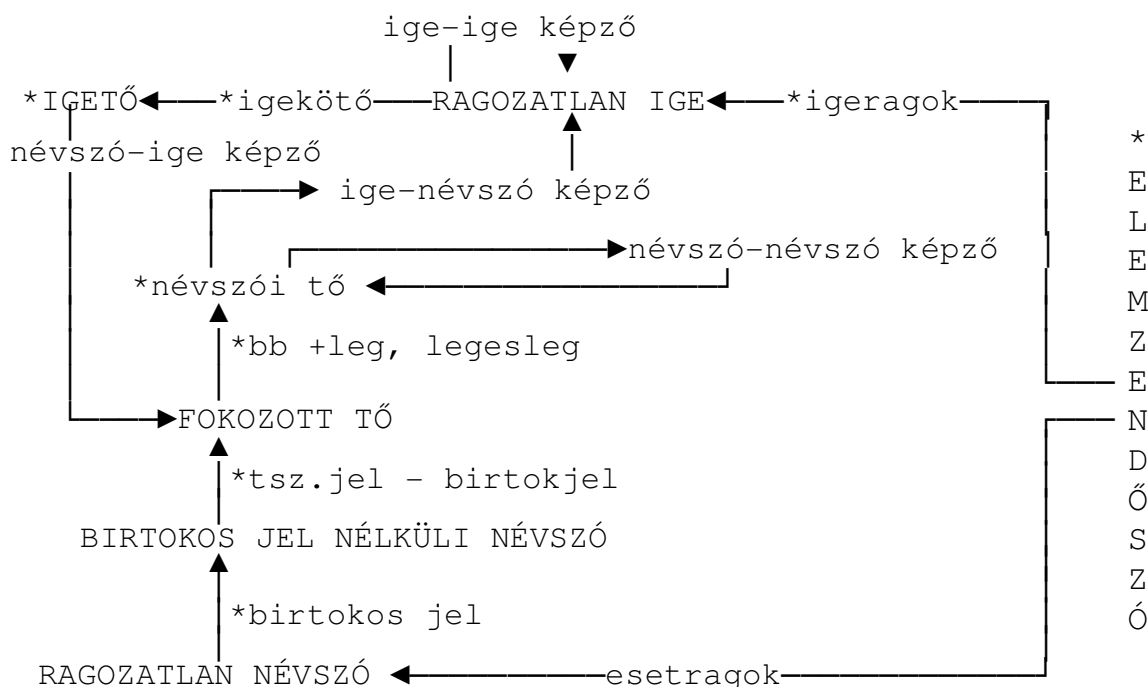


A csomópontok (NAGYBETŰVEL) a szóosztályoknak, az élek (kisbetűvel) a toldalékosztályoknak (előtagosztályoknak) felelnek meg. A *-gal jelölt toldalékok lehetnek „üres” toldalékok is. Bár a felírt gráf alapján a toldalékok száma nem korlátozott, a gyakorlatban 3-4 toldaléknál többet nem használnak a magyar nyelvben, sőt bármiféle rekurzió mesterkéltnél (káposztás-ít-ott-talan-ít-ottátok), de az elemzés során nem zárható ki a ciklus. Pl. *kékíthetetlenül*:

elem alak	kategória	szóosztály	szóalak
<i>kék</i>	szótő	névszói tő	<i>kék</i>
---	fokozó	fokozott tő	<i>kék</i>
- <i>ít</i>	névszó-ige képző	igető	<i>kékít</i>
---	igekötő	ragozatlan ige	<i>kékít</i>
- <i>het</i>	ige-ige képző	ragozatlan ige	<i>kékíthet</i>
- <i>etlen</i>	ige-névszó képző	névszói tő	<i>kékíthetetlen</i>
---	fokozó	fokozott tő	<i>kékíthetetlen</i>
---	számjel	birtokos jel nélküli tő	<i>kékíthetetlen</i>
---	birtokos jel	ragozatlan névszó	<i>kékíthetetlen</i>
- <i>ül</i>	esetrag	ragozott névszó	<i>kékíthetetlenül</i>

Ezzel a szekvenciamodellel jól elemezhetőek a magyar szavak, de elvileg hibás alakok, illetve hibás elemzések is keletkezhetnek. Pontosabban, mivel szófaji felosztásunk rendkívül durva, létezhetnek olyan hibás szóalakok, melyeket elemzőnk értelmez, pedig nincsen értelme. Ez azt jelenti, hogy a valós nyelvénél jóval bővebbet fogad el az elemző. A gyakorlatban ennek kicsi a valószínűsége.

Az inverz (elemző) gráfból is jól látszik, hogy nem fordul elő ürestoldalék-ciklus.



Ez a gráf az elemzés vázát adja meg, pontosabban megadja, hogy az egyes toldalékosztályoknak megfeleltetett automaták hogyan vannak felfűzve. Az elemzés az ELEMZENDŐ SZÓ csomópontnál indul el. A *-gal jelölt csomópontoknál kell ellenőrizni, hogy a szó szerepel-e a szótárban. Így háromféle szót kereshet a szótár rutin:

Teljes alakú (esetleges toldalék levágása nélküli) **szó**: lehet bármilyen okból a szótárba felvett ragozott vagy ragozatlan alakú névszó vagy ige, de lehet az egyéb szófajú szóalak is.

Névszói tő: a gyakorlatban ez lehet ragozott tő is.

Igető: ez is lehet ragozatlan, és a rendhagyó ragozás miatt ragozott ige.

A gráf alapján elemezhető a magyarban használt formák döntő többsége. Az ily módon nem elemezhető formák olyan ritkán használtak (mint pl. *al-hat-nék-om-ban*), hogy feltételezhetjük, hogy az alkalmazási témakörben nem lesz rájuk szükség. Ha mégis, a gráf egyszerű bővítésével építhetők be ezek a formák a morfológiai elemzőbe. A gráf alapján természetesen elemezhetők olyan mesterségesen kitalált szavak is, melyek a magyar nyelvben elő sem fordulhatnak (pl. *fék-ebb-ít*), de a gyakorlatban ez a struktúra helyes szavaknál nem vezet alternatív rossz felbontáshoz. A helyes megoldás minden esetben előáll.

Elemzés: Az adott szó végéről levágja a lehetséges végződéseket a toldalékolás sorrendi szabályainak megfelelően, az elejéről az előtagokat, miközben figyelembe veszi a lehetséges tőváltozások inverzét is. Az így kapott szótövet esetleg felbontja szóösszetételként, és az így kapott szótöveket azonosítja a szótárral. Szóosztályegyezés esetén eredményül a szótövet, az előtagot és a toldalékokat adja vissza, Egy szó több megoldást is adhat, egyrészt mert a tő és a végződés is lehet többértelmű. másrészt mert a szó többféleképpen vágható szét tőre és toldalékokra. Az elemző minden lehetséges megoldást felsorol.

Kivitelezés.

A toldaléktár, illetve előtagtár felhasználásával egy programgenerátor az előfázisban toldalékosztályonként egy-egy véges automatát generál, és ezeket az automatákat összefűzi a szóosztályok közötti átmeneteknek megfelelően. (Így lényegében egy BTN jellegű rendszert kapunk.) Ily módon a toldalékok halmaza és osztályozása az elemzőbe „be van drótozva” míg a szótár adat jellegű információ. Az elemző bemenete egy szó, kimenete egy lista melynek elemei a lehetséges szótókkódból, és toldalékolási szekvenciából állnak. A szótókkód tartalmazza a szótárban talált szó kódját, mely a szintaktikai-szemantikai elemzéshez szükséges (lásd szótár). A toldalékolási szekvencia az alkalmazott toldalékok kódját tartalmazza.

Tapasztalat:

A kezdeti próbálkozások alapján hatékony és rugalmasan generálható módszert dolgoztunk ki. Felmerül a kérdés, mennyire fedi a valós szóelemzést algoritmusunk. A mind szótani, mind hangtani illeszkedési szempontból elnagyolt szabályok miatt a helyes szóalakok helyes elemzését mindenképpen megtalálja az elemző. A kérdés inkább az, hogy keletkeznek-e félreelemzések, vagy nem. Mivel feltételezzük, hogy csak helyes szavakat kell elemezni, a többértelműség problémáját kell vizsgálni. A látszólagos felületetes szóosztályozás (lényegében csak névszói és igei csoportok kezelése) és a ragozási osztályok „semmibevétele” ellenére pontos morfológiai felosztásokat kapunk. Az elemző elvileg helytelenül is elemezhet. A többértelműségnek több oka van:

1. Maga a szótó többértelmű: *lép* mint a *méh viaszháza*, és mint ige. Ez a probléma általában nem jelentkezik, ha a szó toldalékol, de a példánkban szereplő szónál ez sem segít, hiszen a *-nek* lehet igerag is és esetrag is. A szövegben a leggyakoribb többértelműség az az határozott, és az egy határozatlan névelő formailag megegyezik az az utalószóval, illetve az egy számnévvel.

2. A toldalék többértelmű: *-ják* mint kijelentő módú igerag és mint felszólító módú igerag.

3. Opcionális tőváltozás következménye: *gyümölcs-é-t*, *gyümölcs-e-'t*, mármint a *gyümölcs ízét*, vagy *Péter gyümölcsét*.

4. Képzett alak sajátos jelentéssel: *leves*, *lé-s=leves* (*leves gyümölcs*, *gyümölcsleves*).

5. Tévesen alkalmazott tőváltozás, pl.: a *szén-szenet* mintájára *kenet=kén-et*. mint a *kén tárgyesete*.

6. A felületetes modell miatti (hangrendi illeszkedés semmibevétele, egyszerűsített szóosztályozás) téves levágás. pl.: *balek=bal-ek* a *bal* többes száma.

A kísérleti próbálkozások alapján (2000-5000-es szótár, 70%-os toldalékkészlet) az elemző által többértelműen felbontott szavak túlnyomó többsége valóban többértelmű (az első négy csoportba tartozik). A százalékban ki nem fejezhető tévedések (utolsó két csoport) várhatóan jól tisztázhatóak a szintaktikai-szemantikai elemzésnél.

A szótár

Az adott munkához szükséges szavakat egy szótárban gyűjtjük össze. Pillanatnyilag az a feltevésünk, hogy ha a téma megfelelően korlátos, akkor 2000-3000 szó elegendő a dialógushoz.

Kérdés persze, hogy mit értünk szó alatt. Szótári szó alatt egy olyan valóban létező vagy mesterségesen konstruált szótövet fogunk érteni, amelyhez toldalékok járulhatnak. Az előző fejezet, amely a morfémák szétválasztásával foglalkozott, részletesen ismertette, hogy milyen toldalékokat akarunk leválasztani, valamint ismertette azt is, hogy milyen tőtorzulásokat kezelünk. A szótárba tehát bele kerülnek mindazok a képzett alakok, ahol a képzőt nem vágtuk le, az összetett szavak és azok a torzult szótövek, amelyeket az előző fázis algoritmikusan nem kezelte. (Pl. a *tesz* tő mellé belekerül a *ten* tő is a *tenni*, *tenném* stb. alakok miatt, és a *tet* tő is a *tettem*, *tetted* stb. alakok miatt. Az igekötős igéknél az igekötős tő mint főalak és az igekötő és igekötő nélküli tő mint segédalakok, amiből később össze kell állítani a teljes alakot.)

A szótár minden egyes szóhoz egy sereg információt rendel hozzá. A hozzárendelt információt három csoportba oszthatjuk: morfológiai, szintaktikai és szemantikai információkra.

Morfológiai jellemzők

A morfológiai információ arra a célra szolgál, hogy segítségével megállapítsuk, hogy hol lehet a szóról a végződést leválasztani. Erre nézve több különböző módszer is kialakult. A legtökéletesebb módszer persze az lenne ha minden szóhoz felsorolhatnánk, hogy milyen végzódéseket kaphatnak, és azokat milyen nyelvtani értelemben. Ez azonban mérhetetlen nagy tömegű információ volna. Ezért az egyes szavak nyilván ragozási osztályokra fognak utalni. Sajnos, elég finom felosztás mellett a ragozási osztályok száma is igen nagy, ezért ezt különböző kivételleírási módszerekkel kombinálják. Ezekben azt írják le, hogy a szó bizonyos eseteiben az egyik, más esetekben viszont a másik táblázat érvényes, míg további esetekben explicite megadják a kivételes alakot. Ilyen módszert alkalmazott Ju. Apreszján az ETAP fordító programoknál.

Egy másik lehetséges megoldás, hogy a szótő tulajdonságait egy tulajdonság vektorral írjuk le (pl. szófaj, magas illetve mély hangrend, ajakkerekítésesség stb.) és egy hasonló vektorral jellemezzük a végződést. A két vektor kölcsönhatásából azután megállapítható, alkalmazható-e a szóra a végződés és milyen morfológiai tulajdonságokkal rendelkezik a keletkezett szó. Ilyen módszert alkalmazott a Kiss Zoltán a KFKI-ban készített gépelést ellenőrző programban.

A mi módszerünk ezeknél sokkal durvább. A szavakat egyszerűen csoportokba osztjuk aszerint, hogy milyen végződéseket kaphatnak. Igeragot igék kaphatnak, birtokos és esetragot névszók, fokozni melléknévszerű szavakat lehet stb. Minket is meglepett, hogy ezzel az egyszerű módszerrel statisztikusan igen jó analízist kaptunk. Természetesen a szó végéről a potenciális végződéseket mindig levágtuk, de azokat a szétvágásokat elvetettük, ahol a végződés és a tő osztályba sorolása között ellentmondás volt. Kevesebb, mint fél százalék volt az olyan eset ahol nem odaillő végződést vágtunk le és az algoritmus ezt nem mutatta ki. Ilyenek voltak: a *szinte* szóban a *szint* szót is megtalálta -e birtokos raggal (*szintje* helyett). A *másik* szóban a *más* szót vélte megtalálni -ik birtokraggal (a *hintóik* szó analógiájára). Valószínű, hogy a hibák egy részét a raghoz kapcsolt szűrő, transzformáló algoritmusok pontosításával meg lehet szüntetni.

Ez a szám igen kicsi, ahhoz képest, hogy egy szövegben kb. 8% morfológiailag kétértelmű szó van. Ennek nagy része természetesen abból fakad, hogy maga a szótő kétértelmű, például: az névelő vagy mutató névmás, *mi* személyes vagy kérdő névmás, egy határozatlan névelő vagy számnév. Ez teszi ki körülbelül a kétértelműségek felét. Továbbá az igéknél gyakran nem állapítható meg, hogy múlt idő egyes szám 3. személyről, vagy múlt idejű melléknévi igenévről van-e szó. Az esetrag előtti é-ről nem állapítható meg, hogy birtokjel vagy birtokos jel. Pl., a *kését* szónál, nem tudjuk, hogy *valakinek a kését* vagy *a késnek a valamijét* emlegetjük-e. Mindezek mellett természetesen vannak olyan kétértelműségek is, mint pl. a *minden* az *összes* és *mindegyiken* értelemben, vagy *érték* főnévként és ragozott igeként stb. Mindezeket a kétértelműségeket később, a szintaktikus és szemantikus elemzés során kell feloldani, és minden arra mutat, hogy ekkor a durva morfológiai elemzésből származó félreértelmezések is megszűnnek.

Szintaktikai jellemzők

A szintaktikai jellemzők lényegében azt írják le, hogy a szó a mondatban milyen szerepet tölthet be, hol állhat a mondatban és milyen alakokban. A betölthető szerepeket természetesen alapvetően meghatározza, hogy a szó milyen szófajhoz tartozik, és milyen végződéseket vehet fel. Úgy tűnik azonban, hogy a hagyományos nyelvtanokban bevezetett szófaji osztályozás nekünk nem elég finom, mert a mondatban betöltött szerep és a szó toldalékolhatósági típusa nem feltétlenül esik egybe. Ezért, a hagyományos szófaji osztályokat többé-kevésbé megtartottuk, de tovább finomítottuk azokat. Pl. a *muszáj* szót a szótárak tradicionálisan igeként vagy segédigeként minősítik, és ez utóbbi teljesen megfelel a mondatbeli szerepének. Ezzel szemben toldalékolás szempontjából semmilyen igei tulajdonságot nem mutat. Miután számos ilyen szót találtunk, ezekre a „nem igei segédige” elnevezést vezettük be.

Próbaképpen, mi körülbelül 120 Kbyte (kb. 50 nyomtatott oldal) szövegből indultunk ki, amelyet az elmúlt 1 év folyamán írtunk, és a természetes nyelvű ember-számítógép kapcsolattal foglalkozott. Ezen próbálgattuk morfológiai elemzőinket és ebből készítettünk egy szótárt. A szótár körülbelül 2500 szótövet tartalmaz (a fenti értelemben). A következő fejezet részletesen ismerteti az általunk bevezetett szófajták meghatározását, a függelékben pedig megadjuk a fenti 2500 szó ezen elvek szerinti besorolását.

Ez a szófaji besorolás tehát kettős arcú; egyrészt megadja, hogy a szó milyen jellegű toldalékokat vehet fel, másrészt megadja, hogy milyen szerepet tölthet be a szintaxisban. Bennünk is felmerült a kérdés meg lehet-e adni ezt a két vonást egymástól függetlenül. Egyelőre úgy látjuk, hogy a két besorolás nem független, és egy ilyen jellegű szófaji felosztás a szintaktikai analízis számára megfelelő alapot biztosít.

Tudjuk, hogy ez a felosztás még nem végleges, csak nagyobb tömegű szöveg szintaktikai elemzése igazolhatja az egyes bevezetett szóosztályok létjogosultságát, illetve szükségessé teheti azok megváltoztatását.

Szemantikai jellemzők

A szó harmadik jellemzője a szemantikai tartalom. Mi itt, ebben a munkában szemantikának azt nevezzük, amire az illető szót használni lehet az adott feladat megoldásában. Lesznek olyan szók, elsősorban igék, amelyek valamilyen tevékenységet indítanak el a számítógépben. A főnevek táblázatok (adatállományokat) jelölnek, a főnevek és a melléknevek a táblázatok oszlopait illetve sorait jelölik, ki stb. Feltételezzük tehát, hogy a szemantikus információ meglehetősen témakörhöz kötött. Nem áll szándékunkban egy univerzális, minden témakörre alkalmazható általános magyar nyelvi szemantikát kidolgozni, és azután valamilyen áttételen keresztül az adott témakörre alkalmazni. Ilyenformán univerzális, témakörtől független szemantikus leírása valószínűleg csak a kötőszavaknak, módosítóknak, segédigéknek stb. lesz,

A fentieknek az a következménye, hogy a szemantikus információt a különböző feladatoknál ki kell cserélni még azoknak a szavaknak a jó részénél is, amelyek közösek két különböző feladat esetén. Ez indokolhatja azt, hogy a szemantikus információkat tartsuk külön szótárban. Ennek további előnye lenne, hogy a munka két különböző fázisában két, egyenként kisebb szótárral dolgozhatnánk. Hátránya viszont, hogy párhuzamosan két szótárt kell karbantartani, és könnyen inkonzisztens állapot állhat elő.

A magyar szavak morfológiai-szintaktikai felosztása

0---- Ragozhatatlanok

01--- Határozószók

011-- határozói igenevek

Ide csak a régies *-ván, -vén* képzős alakok kerültek. A *-va, -ve* képzőt a morfológiai elemző levágja.

012-- határozói névmások

Ide kerültek azok a ragozott névmások is, ahol a tő erősen torzult és ezért elemzésük problematikus.

02--- Igekötők

Igekötő minden olyan szócska amelyet, ha az ige előtt áll, egybeírunk vele, ha utána, külön.

03--- Névelők

038-- határozott névelő

039-- határozatlan névelő

04--- Kötőszók

További osztályozása szükséges aszerint, hogy szavakat vagy mondatokat köt össze; ha mondatokat alá- vagy mellérendelő mondatokat. Megjelölendők továbbá a kételemű kötőszavak.

05--- Módosító szavak

051-- Igei módosító szavak

Az ige értelmét módosítják: *volna* feltételes mód (hátsó), *hadd* és *hagy* óhajtás, *ne* és *se* tagadás (elől)

052-- Melléknévi (+igei) módosító szavak

A melléknév, melléknév jellegű számnév, melléknévből képzett határozószó és ige jelentését erősíti, gyengíti, tagadja stb. A szó előtt áll.

053-- Számnévi módosító szavak

A konkrét számok értékét bizonytalanítja; a szám előtt áll.

054-- Névutószerű szavak

A névutók kaphatnak formális birtokos ragot, ezek nem.

060-- Igeszerű szavak

061-- Nem igei segédigék

Soha nem állnak jelzőként, mindig főnévi igenévvel. A mondat állítmánya a segédige+főnévi igenév. Az alany *-nak* ragot kap. A főnévi igenév ragozódhat, különösen, ha nincs ágens kitéve. Múlt ideje a *volt* szóval, jövő ideje a *lesz* szóval, feltételes módja a *lenne* vagy *volna* szóval képződik. Pl. *Pistának tanácsos odamenni, nem volna szabad odamenned,* stb.

062-- Ragozhatatlan főnévi névmások

A személyes névmások alany és tárgyesetei. A többi esetet lásd a névutók ragozásánál.

07--- Egyéb szavak

1---- Ige

Ami igeragokat kaphat.

11--- segédigék

A segédige ragozódik, az ige (ragozatlan) főnévi igenévvel áll.

12--- Féligői segédigék

Az állítmány ugyanúgy képződik, mint a nem igei segédigék esetén, de a múlt idő, jövő idő, feltételes és felszólító módja a segédige egyes szám 3. személyű megfelelő alakjával képződik. Pl. nekem el kellene mennem.

19--- Főnévi igenév (* van ragozása, amely az igeragozás része)

2---- Név szó

201-- nevek

A nevek személynevek, cégek, hónapok stb. nevei. A névszói szerkezetben (amikor nem önmaga a fő főnév) a főnév, illetve az azt megelőző 211 számmal jelölt főnévi jelző előtt, a tulajdonságjelzők után szoktak állni. Pl. *Jan Kowalsky lenyel vendégmunkás, Kiss Péter lakatos, Pista bácsi, Vörös Október gépgyár, március hónap* stb.

202-- mértékegységek

21--- főnév

211-- egyszerre főnévi és melléknévi szó

Idetartoznak az anyagnevek, foglalkozások, nemzetiségek, vallások stb. Egyformán gyakran állnak főnévként és közvetlenül a főnév előtt álló jelzőként. Nem fokozhatók. Pl. *arany virágcserep, lengyel vendégmunkás, csőszerelő szakmunkás, katolikus pap, elnök elvtárs* stb.

212-- teljesen főnév jellegű névmás

Teljesen úgy ragozódnak, mint a főnevek; birtokos és esetragot is kaphatnak.

213-- névutóhoz hasonlóan ragozódó névmások

A névmás különböző személyű alakjait egy konstruált tőből személyragokkal képezhetjük, utána birtokos jel és esetrag is állhat. A személyragok formailag a birtokos ragozás ragjaival esnek egybe. Pl. *magam, magad, maga, magáé, magunktól* stb.

214-- birtokosan nem ragozódó névmások

Birtokjelet nem kaphatnak, de birtokos jelet, és esetragot kaphatnak. Pl. *egymás, egymásét, egymást* stb.

22--- Melléknév

Állhat tulajdonságjelzőként névszói szerkezetben, nem igei segédigeként, nem igei állítmányként. Megkülönböztetendők azok, amelyek vonzattal rendelkeznek. Fokozhatók.

221-- melléknévi igenév

222-- melléknév jellegű névmás

223-- melléknévi névutó

23--- Számnév

230-- konkrét szám

Egytől tízig betűvel és számmal, továbbá *száz, ezer, millió* betűvel is, a többi csak számmal, kötőjellel toldalékolva.

231-- melléknév jellegű számnév

Melléknévként viselkedik, fokozható, melléknévi módosítót kaphat, de a névszói szerkezetben számnév helyén áll a tulajdonság jelzők előtt. A *-szor* raggal számhatározó képezhető belőle.

232-- névmás jellegű számnév

Nem mutat melléknévi tulajdonságokat, a számjelző helyén állhat. Számhatározó képezhető belőle. Számnévi módosítót kaphat. Nagyrészüik számnévi névmás.

233-- kvantor jellegű

A névszói szerkezetben a számjelző előtt az *un.* kvantor pozícióban áll. Jöhet utána számjelző is. Számhatározó nem képezhető belőle.

24--- Névutó

A személyes névmások ragozott illetve névutóval ellátott alakját, a névutó+személyrag alakban képezzük. A bizonyos ragoknak egy mesterséges tő felel meg, pl. a *-val -vel* ragnak a *vel-*: *velem, veled, vele*; az *-on -en -ön* ragnak a *rajt-*: *rajtam, rajtad, rajta* stb. Megkülönböztetendőek azok a névutók, amelyek ragot vonzanak.

Egyéb jelölések:

----1 nem szótári alak. csak ragozott formában szerepelhet

----2 ragozott alak (a tő torzulása miatt került a szótárba)

----4 nem kell tovább bontani, csak ez a forma szerepel a szótárban

---1- kötelező vonzata van

---2- opcionális vonzata van

A szintaxis alapjai

A szintaxis a mondatok felépítésével foglalkozik. A mondatban a szavak bizonyos egységekké kapcsolódnak össze, melyek azután ismét nagyobb egységeket alkotnak, és ezekből még tovább; amíg a teljes mondat fel nem épül. A magyar nyelv, mint az közismert, ún. szabad szórendű nyelv. Ez természetesen nem azt jelenti, hogy a magyar mondatban a szavak tetszőleges sorrendben állhatnak. A magyar mondatban a szintagmákat (mondattani egységeket) két kategóriába lehet sorolni. A felsőbb szintű mondattani egységekre az a jellemző, hogy mondatbeli sorrendjük szabad, és az egyes egységek grammatikai szerepét az egység utolsó tagjaként álló morféma (rendszerint egy rag vagy egy névutó) szabja meg. Az alsóbb szinten viszont a mondatrész alkotórészeinek egymáshoz kapcsolódását pozicionális, nevezetesen sorrendi szabályok írják elő.

A szabad szórend nem jelenti azt sem, hogy a felsőbb szinten a szintagmák sorrendje tetszőleges. Ezen a szinten a sorrendet nem a mondattani szabályok írják elő, hanem a beszélő szándéka. A kérdő, tagadó, felszólító mondatokban a sorrend valamivel kötöttebb, a kijelentő mondatokban pedig nagyrészt attól függ, hogy mit kívánunk kihangsúlyozni.

Ezt a jelenséget É. Kiss Katalin igen részletesen leírta, mi ebben a fejezetben ezzel a sorrendi kérdéssel nem akarunk foglalkozni, de később még visszatérünk rá.

Mint más nyelvekben is, a magyarban is a mondat fő alkotó eleme az ige, pontosabban az igei kifejezés; ez akkor is igaz, ha a mondatban az ige nincs is explicite kitéve. Az ige jelentése szabja meg, hogy milyen egyéb mondatrészek szerepelhetnek még a mondatban. Ezeket az igéhez kapcsolódó mondatrészeket az ige argumentumainak fogjuk nevezni. Lehet az igének logikai alanya, aki (vagy *ami*) a cselekvést végrehajtja, vagy akivel valami történt (pl. *alszik+valaki, elesik+valaki*); lehet ezenkívül tárgya (pl. *olvas+valaki+valamit*) vagy részeshatározója (kedvezményezettje) (pl. *ad+valaki+valamit+valakinek*). Ezen argumentumokat többnyire névszói kifejezések adják meg, és az egyes argumentumokat a kifejezés végén álló rag vagy névutó jelöli ki.

Az az absztrakt szemléletmód, hogy a mondat egyes alkotórészeit funkciójuk, és nem nyelvtani esetük szerint különböztetjük meg, fontosnak bizonyult a gépi fordítás esetén. Ugyanis különböző nyelvekben az egyes funkciók különböző névszói szerkezetekben jelennek meg. De azt is látni fogjuk, hogy magán a magyar nyelven belül is részben az igei kifejezés struktúrájától függ, hogy egy argumentum milyen formában kerül a mondatba. A legegyszerűbb esetben az logikai alany a mondat alanya, de a magyar nyelvben sok olyan konstrukció is van a mikor az logikai alany -NAK ragot kap.

A szintaxison belül először az igei szerkezetek felépítésével foglalkozunk. A következő részben az argumentumként szolgáló névszói szerkezetek felépítését ismertetjük. Ezek után a két legfontosabb argumentumnak az alanynak és a tárgynak a felismerésével foglalkozunk az ige-alany és az ige-tárgy egyeztetések kapcsán. Végül a „Vonzatok” című részben arról lesz szó, hogyan ismerjük fel az ige további argumentumait az esetragok és névutók segítségével.

Igei szerkezetek

Az egyszerű (nem összetett) magyar mondatok az ige szempontjából a következő csoportokba sorolhatók:

1. Ragozott ige, a logikai alany alanyesetben van, az igének néhány kötelező és néhány opcionális argumentuma van. Pl. *Pista olvas.*

2. Igei ragozású segédige, mellette az ige főnévi igeneves formában, a kettőnek azonos logikai alanya van alanyesetben. Pl. *Pista olvasni fog.*

3. Nem igei segédige, mellette az ige infinitívusban vagy infinitívuszi ragozásban, a logikai alany -NAK ragot kap. Pl. *Pistának olvasnia kell.*

4. A VAN ige „létezik” értelemben. Az egy argumentuma alanyesetben van. Pl. *Hideg van.*

5. A VAN ige „birtokol” értelemben. A logikai alany -NAK ragot kap, amit birtokol, alanyesetben áll birtok jellel. Pl. *Pistának van mai újsága.*

6. Szenvedő szerkezet. A tárgy van alanyesetben, az ige VAN+határozói igenév alakban áll, a logikai alany, ha egyáltalán meg van adva, által névutót kap. Pl. *A feladat meg lesz oldva.*

7. Valami valamilyen, vagy valami valahány. A jelen idejű VAN nincs kitéve. Mindkét argumentum alanyesetű, Az első egy névelős névszói kifejezés, (személynevek elé nem kerül névelők a másik egy melléknévi vagy számnévi kifejezés (nem lehet névelője). A sorrend tetszőleges. Pl. *A rózsza piros volt. Sok volt az eső. Az út 3 km volt.*

8. Valami valami. ("is-a" kapcsolat) A jelen idejű VAN nincs kitéve. Mindkét argumentum alanyesetű főnévi kifejezés. A sorrend lényeges és a mondat típusától (kijelentő, kérdő, tagadó) függ. Pl. *Pista katona volt. A katona Pista volt.*

A mondat központjában álló igei kifejezés meghatározza (az igein kívül) az ige módját, idejét, logikai alanyát és egy sor más vonást. Az igei kifejezés a ragozott igein kívül, segédigéből, igei módosítóból, szabad határozókból áll, és explicite megadott logikai alanya is lehet. Ezek együtt adják meg a fenti jellemzőket.

Az igei kifejezés bizonyos mértékig mindig meghatározza az ige alanyát is. 1. és 2. személyben nem tekinthető hiányosnak a magyar mondat, ha az alany csak az igeraggal van megadva. Ha a névmás is ki van téve, a forma redundáns, és az ismételt utalás csak a hangsúlyozás miatt van. A 3. személy esetén a névszói kifejezéssel megadott alany hiányát utalásnak kell tekinteni (általában hátra, néha azonban előre). Az igerag által indikált grammatikai alanynak és a mondatban explicite megadott alanyunk bizonyos vonásokban meg kell egyezniük. Ezt az „Egyeztetések” című fejezet definiálja pontosabban.

A grammatikai alanyt a személyrag adja meg. Ezt az 1., 4., 7. és 8. esetben az ige kapja, a 2. esetben a segédige. A 5. esetben az infinitívusz vagy a nek ragból képzett névmás vagy mindkettő; az 5. esetben ugyanezen a névmás illetve a birokjel adja meg.

A 6. esetben, ha nincs kitéve, hogy ki által, és a 2. esetben, ha se az infinitívusz nem ragozódik, se névmás nincsen, általános alanyról beszélhetünk. Pl. *A probléma meg lesz oldva. Tilos bemenni.*

Az ige módján általánosabb értelemben azt értjük, hogy a cselekvés feltételes, lehetséges, szükségszerű, rendszeres, stb. Ez a különböző igei kifejezésekben, különböző formákban jelenik meg. Az ige általánosabb értelemben vett módját megadhatja részben az igerag, részben igei képző (pl. a *-hat, -het*), részben segédige, részben az igei módosítók (pl. *volna*). Pl. *olvasna, tudna olvasni, olvashat, olvashatna, kellene olvasnia, muszáj lenne olvasnia, hideg volna, kell lennie mai újságjának, katona volna* stb.

Hasonló a helyzet az igeidővel is. Az időt megadhatja rag, segédige, segédige ragja, igei módosító. Pl. *olvasta, olvasni fogja, kellett olvasnia, muszáj lesz elmennie* stb. De megadhatja az ige idejét úgynevezett szabad határozó is. Pl. *Holnap elolvassa.*

Összefoglalva: a mondat központjában az ige áll. Az igenek attribútumai vannak, amelyek megadják az ige lefolyásával kapcsolatos tudnivalókat időt, helyet, alanyt, befejezettséget, feltételeességet, lehetőségességet, ismétlődést, stb. Ezeket az attribútumokat, megadhatják igeragok, segédigék, igei módosítók, képzők és a mondat szabad határozói. Azt, hogy a rendszer ezen tulajdonságok közül melyeket és milyen osztályozásban tartja számon, a rendszer céljaitól függ.

Célszerűnek látszik a következő attribútumokat felvenni: alany, idő, hely, mód. Az attribútumok értéküket egy adott értékészletről veszik. Az attribútumok értéke lehet definiálatlan is; ez esetben a baloldalon az attribútum default értéket kap.

Formális definíció

A fentiek szerint a mondat felépítése a következő attribútum nyelvtannal megadva:

(Jelölések:

```
 ::=      definiáló egyenlőség a jobb oldal komponenseinek
         sorrendje közömbös
+        ezen két komponens között a sorrend kötött
mondat   grammatikai fogalom
(...)    a fogalom által definiált attribútumok
VAN, NAK explicite megadott morféma
         (a mondatban helyette a létige valamelyik alakja vagy
         -nak, illetve -nek rag áll.)
[...]    opcionális )
```

(Megjegyzés: a formális definícióban helytakarékoság céljából nincs leírva a

határozói igenév ::= ige+VA

és más hasonló magától értetődő szabályok.)

mondat ::= igei kifejezés(alany, hely, idő, mód)

igei kifejezési ::= [főnévi kifejezés(alany)]
 ige+IGEI KÉPZŐ(mód)+IGERAG(alany, idő, mód)
 [hol(hely)] [mikor(idő)]

Pl. *Pista tegnap a kertben olvas-hat-ott.*
 (alany= (idő= (hely= (mód= (alany=
 e.sz.3.sz) múlt) adott) lehet.) e.sz.3.sz,
 idő=
 múlt,
 mód=0)

igei kifejezés(alany=e.sz.3.sz.\adott, hely=adott, idő=múlt,
 mód=lehetséges)

Holnap elolvas-om.
 (idő= (alany=
 jövő) e.sz.1.sz.
 idő=0,
 mód=0)

igei kifejezés(alany=e.sz.1.sz., hely=0, idő=jövő, mód=kijelentő)

Nem mondat ezzel szemben a következő:

Pl. **Pista iskolába olvasott.*
**Pista holnap olvasott.*

igei kifejezés2 ::= [főnévi kifejezés(alany)]
 segédige+IGERAG(alany, idő, mód)
 főnévi igenév
 [hol(hely)] [mikor(idő)]

Pl. *Holnap fog olvasni.*
 (idő= (alany=
 jövő) e.sz.3.sz.,
 idő=
 jövő,
 mód=0)

igei kifejezés(alany=e.sz.3.sz., hely=0, idő=jövő, mód=kijelentő)

igei kifejezés?) ::= [főnévi kifejezés(alany)+NAK]
 segédige+IGERAG(idő, mód)
 főnévi igenév+[IGERAG(alany)]
 [hol(hely)] [mikor(idő)]

Pl. *Nekem kell-ett elmen-nem.*
 (alany= (idő= (alany=
 e.sz.1.sz.) múlt, e.sz.1.sz.)
 mód=0)

igei kifejezés(alany=e.sz.1.sz., hely=0, idő=múlt, mód=szükséges)

El *kell-ene men-nünk.*
 (idő=0. (alany=
 mód= t.sz.1.sz.)
 feltv)

igei kifejezés(alany=t.sz.1.sz., hely=0, idő=jelen,
 mód=feltételes/szükséges)

igei kifejezés3) ::= [főnévi kifejezés(alany)+NAK]
 segédige VAN(idő, mód)
 főnévi igenév+[IGERAG(alany)]
 [hol(hely)] [mikor-(idő)]

Pl. *Tilos volt bemenni.*
 (mód= (idő=
 tagadó) múlt)

igei kifejezés(alany=általános, hely=0, idő=múlt,
 mód=kijelentő\tagadó)

Pl. *Muszáj bemen-nem.*
 (mód= (alany=
 szükséges) e.sz.1.sz.)

igei kifejezés(alany=e.sz.1.sz., hely=0, idő=jelen, mód=szükséges)

igei kifejezés4 ::= [főnévi kifejezés(alany)]
 VAN(alany=3.sz., idő, mód)
 [hol(hely)] [mikor(idő)]

Pl. Irak-ban háború van.
 (hely= (alany= (alany=
 adott) e.sz.3.sz) e.sz.3.sz,
 idő=0,
 mód=0)

igei kifejezés(alany=e.sz.3.sz.\adott, hely=adott, idő=jelen.
 mód=kijelentő)

Holnap le-het -nek zivatarok.
 (idő= (mód= (alany= (alany=
 jövő) lehet.) t.sz.3.sz, t.sz.3.sz)
 idő=0.
 mód=0)

igei kifejezés(alany=e.sz.3.sz.\adott, hely=0, idő=jövő,
 mód=lehetséges)

igei kifejezés5 ::= [főnévi kifejezés(alany)+NAK]
 VAN(idő, mód)
 névelőtlen főnévi kifejezés+BIRTOKRAG(alany)
 [hol(hely)] [mikor(idő)]

Pl. Új bicikli-d van?
 (alany= (idő=
 e.sz.2.sz.) jelen,
 mód=
 kijelentő)

igei kifejezés(alany=e.sz.2.sz.\adott, hely=0, idő=jelen,
 mód=kijelentő)

igei kifejezés6 ::= [főnévi kifejezés(object)]
 VAN(idő, mód)
 határozói igenév
 [hol(hely)] [mikor(idő)]

Pl. Itt van a kutya elásva.
 (hely= (idő= (object=
 adott jelen. adott\határozott)
 mód=0)

igei kifejezés(alany=általános, hely=adott, idő=jelen,
 mód=kijelentő)

(Megjegyzés: Az igék között vannak tárgyas és tárgyatlan igék. Esetenként tárgyatlan igékből is képezhető ilyen „szenvető” szerkezet. Az ilyen igéknél a szerkezet nyelvtani alanya azonos a logikai alannyal, hasonlóan ahhoz, hogy a múlt idejű és jövő idejű melléknévi igenév után álló jelzett szó sem a jelző tárgya, hanem az alanya. Pl. A csésze össze lett törve. A ruha el van szakadva. Meg van halva. Szépen meg van fészülködve. stb.)

Vannak azonban olyan tárgyatlan igék is, ahol ilyen szerkezet nem képezhető sőt. esetenként a múlt idejű és jövő idejű melléknévi igenév sem. Pl. törekszik, rájön, függ, árt.)

```
igei kifejezés7 ::= [főnévi kifejezés(alany)]  
                  melléknévi kifejezés+[VAN(alany, idő.mód)]  
                  [hol(hely)] [mikor(idő)]
```

```
Pl. Vasárnap szép lesz az idő.  
   (idő=          (alany=          (alany=  
   jövő|múlt)    e.sz.3.sz.      e.sz.3.sz)  
                 idő=  
                 jövő,  
                 mód=0)
```

```
igei kifejezés(alany=e.sz.3.sz.\adott, hely=0, idő=jövő,  
               mód=kijelentő)
```

```
igei kifejezés7 ::= [főnévi kifejezés(alany)]  
                  számnévi kifejezés+[VAN(alany, idő, mód)]  
                  [hol(hely)] [mikor(idő)]
```

```
Pl. 125 kg          vagyok.  
   (alany=  
   e.sz.1.sz.,  
   idő=  
   jelen,  
   mód=0)
```

```
igei kifejezés(alany=e.sz.1.sz., hely=0, idő=jelen, mód=kijelentő)
```

```
igei kifejezés8 ::= [főnévi kifejezés(alany)]  
                  főnévi kifejezés+[VAN(alany, idő, mód)]  
                  [hol(hely)] [mikor(idő)]
```

```
Pl. Az igazgató      vagyok.  
   (alany=  
   e.sz.1.sz.,  
   idő=0,  
   mód=0)
```

```
igei kifejezés(alany=e.s2.1.sz., hely=0, idő=jelen, mód=kijelentő)
```

(Megjegyzés: ha ki van téve az ige, akkor az előtte levő szó tartozik az állítmányba. Ha nincs kitéve, és az egyik határozott, a másik nem, akkor határozott az alany, máskülönben az a feltételezés, hogy az első az alany, a második az állítmány.)

```
Pl. Pista katona. (Pista határozott, mert személynév.)  
   ócskavas ez az autó.  
   A lengyel a katona.
```

A névszói kifejezés

Az igei kifejezésben álló igéhez mondatrészek kapcsolódnak. Ezek a mondatrészek leírhatják az ige eszközt, tárgyát, helyét, irányát, stb. Alapjában véve úgy képzelhetjük, hogy az ige egy függvény, amelynek meghatározott számú argumentuma van. Természetesen az argumentumok száma és típusa igénként más és más. Pl. *esik* (nincs argumentum), *alszík+ki*, *lát+ki+mit*, *összeköt+ki+mit+mivel*, *odaér+ki+hova*, *megkülönböztet+ki+miket*, *megjelöl+ki+mit+mivel* stb. Az igéknek ezeket az argumentumait az jelöli ki, hogy a hozzátartozó névszói szócsoport (szintagma) milyen végződéseket vagy névutókat kaphat, illetve hogy milyen határozószók és határozói névmások tartoznak bele). Egyes argumentumok főnévi igenevet vagy hogy kötőszós mellékmondatot igényelnek vagy ezek közül valamelyiket.

A névszói argumentum fő eleme általában egy főnév, amelyhez különböző jelzők és esetleg egyéb bővítmények, kapcsolódhatnak. Ha a főnév hiányzik, ami a magyar mondatokban igen gyakori, a rag (illetve névutó) az előtte álló szóhoz kapcsolódik.

A jelzők kétféleképpen kapcsolódhatnak a főnévhez. Van olyan eset, amikor a kapcsolódásra csak az utal, hogy a jelző a főnév előtt áll. Semmilyen nemben, számban, esetben történő egyeztetés nincs, nem úgy, mint más nyelvekben. Ilyenkor a jelzők sorrendje kötött: birtokos jelző, (névelő,) számjelző, tulajdonságjelző, főnévi jelzők, főnév; természetesen bármelyikük el is maradhat. Pl. *a megye három élenjáró Arany Kalász szövetkezete*.

Ha a főnévnek van birtokos jelzője, akkor a fenti esetben a birtokos nem kap semmilyen végződést, (és a jelzett szó vagy kifejezés nem kaphat határozott névelőt,) a birtokolt főnév pedig birtokos ragozásbeli birtokjelet kap.

Van a birtokviszonynak egy másik alakja is, ekkor a birtokos *-nak -nek* ragot kap, a birtok és jelzői elé viszont (többnyire határozott) névelő kerül. Ez esetben viszont a birtokos és a birtok el is szakadhat egymástól a mondatban. Pl. *Pistának látom az árnyékát*.

Nemcsak a birtokos jelző, hanem bármelyik jelző elszakadhat a főnév mellől, és ún. hátravetett jelző lehet belőle. Ilyenkor viszont ugyanolyan ragot kap, mint a jelzett szó. A (-NAK ragtalan) birtokos jelzőt is hátra lehet vetni, és ekkor a főnév többnyire nem kap birtokragot, viszont a birtokos megkapja a birtoklás -é vagy -éi jelét. Pl. *Kalapot vettem tegnap a vásárban, pirosat. Elvesztettem a levelet, Jenőét*.

Ha a birtokos (határozott névelős) személyes névmás, az ezzel egyenértékű hátravetett jelző a birtokos névmás lesz. Pl. *az én kalapomat..., a kalapot, az enyémet ...* A birtokos névmás csak olyan helyen állhat, ahol a valódi argumentumként szereplő birtokot nem tettük ki.

Ha az argumentumként álló főnév vagy annak valamilyen jelzője igei jelentésű, akkor ahhoz is kapcsolódhatnak tárgy, határozó és más igei jellegű bővítmények. Pl. *iskolába járás során*. Néhány más szónak, pl. melléknévnek is lehet argumentuma. Pl. *valamilyen+színű*.

A jelzők maguk is lehetnek jelzős kifejezések. Pl. *a téesz bikájának a szarva, múlt idejű ige*.

Argumentumként több kifejezés is állhat *és, vagy, vagy... vagy... nem csak hanem (is)* és más hasonló szavakkal összekapcsolva.

Ha névmás áll az argumentum pozíciójában, akkor a jelző csak hátravetett lehet. Pl. *Ő, a katona elaludt. Az enyémet, a kéket nem adom*.

Formális leírás

argumentum ::= névszói kifejezés+[ESETRAG]+[NÉVUTÓ]

(Megjegyzés: vannak névutók, amelyek esetraggal járnak. A birtokrag után viszont nem feltétlenül tesszük ki a tárgyragot.)

Megjegyezzük továbbá, hogy az ige argumentumainak vonzatát nem befolyásolja az, hogy milyen igei kifejezéshez kapcsolódnak, van-e segédige vagy se, kivéve a szenvedő szerkezetet, amikor a szerkezet alanya valójában tárgy. Segédigének pont azokat az igéket nevezzük, amelyek szükségszerűen infinitívussal járnak és az infinitívusban álló ige vonzatait nem változtatják meg.)

argumentum2 ::= határozószó

argumentum3 ::= határozói névmás

argumentum4 ::= névmás+[TÁRGYRAG]

(Megjegyzés: az ige egyetlen logikai argumentumához, mint azt fentebb vázoltuk, több argumentum jellegű szócsoporthoz kapcsolódhat; ezek összetartozásának megállapítása nem könnyű, és nem is mindig megoldható feladat, de a gyakorlatban a kétértelmű mondatok nem túl gyakoriak. A következő feltételezésekkel élünk:

1. Ha sem az igének, sem a segédigének nincs NAK ragos vonzata és van egy NAK ragos kifejezésünk, és egy másik, birtokos nélküli, birtokjellel ellátott névszói kifejezés, akkor felételezhetjük, hogy az előbbi az utóbbi birtokosa.

2. Ha van egy névszói kifejezés, és a mondatban hátrébb egy ugyanolyan esetraggal ellátott, birtokos jelet viselő, tehát hiányzó birtokú névszói kifejezés, feltételezhetjük, hogy az előbbi az utóbbi birtoka.

3. Ha van két azonos raggal vagy névutóval ellátott argumentum, és az utóbbiból hiányzik a főnév, akkor feltételezhetjük, hogy az utóbbi az előbbinek hátravetett jelzője.

Az előbb felsoroltak mindegyikére lehetne valószínűleg ellenpéldát konstruálni, így ezeket is érdemes szemantikai jegyek alapján ellenőrizni. Ellenkező esetben több lehetséges szintaktikai struktúrát építhetünk fel, amelyek közül válogatnunk kell szemantikai jegyek alapján. Bizonyos esetekben azonban az argumentum hovatarozása nem lesz megállapítható. Különösen a NAK rag okoz problémát amelyet számtalan különböző célra használunk. Pl. *Péternek kell elnevezni.* (Az elnevező vagy az elnevezett Péter?)

névszói kifejezés1 ::= birtokos szerkezet

névszói kifejezés2 ::= névelős kifejezés

névelős kifejezés1 ::= AZ+névelőtlen kifejezés

névelős kifejezés2 ::= határozatlan kifejezés

határozatlan kifejezés ::= [EGY] névelőtlen kifejezés

birtokos szerkezet1 ::= nak ragos birtokviszony

birtokos szerkezet2 ::= egybefüggő birtokviszony

birtokos szerkezet3 ::= hiányos birtokviszony

nak ragos birtokviszony ::= [névelős kifejezés(tulaj.)+NAK]
névelős kifejezés+BIRTOKJEL(tulaj.)

egybefüggő birtokviszony ::=névelős kifejezés(tulaj.)
határozatlan kifejezés+BIRTOKJEL(tulaj.)

hiányos birtokviszony ::= névelős kifejezés+BIRTOKOS JEL

névelőtlen kifejezés1 ::= [számjelző]+[tulajdonság jelző]+
+[főnévi jelző]+[főnév]

(Megjegyzés: természetesen a 4 rész mindegyike nem lehet üres.)

névelőtlen kifejezés2 ::= főnévi névmás

(Megjegyzés: általános szabály, hogy egy adott szinten lévő névmást nem előzhet meg semmi. Tehát, ha a főnév helyén áll névmás nem lehet jelzője, a tulajdonságjelzők közül mindig az első, ha számnév nem lehet kvantora. Pl. **piros ez, olyan nagyfogú fésű, *nagyfogú olyan fésű, annyi méterre, minden 5 méterre, *minden annyi méterre.* De a különböző szintek nem zavarják egymást, pl. *3 olyan fésű.*)

(Megjegyzés: a meghatározás után álló (xxx--) alakú szám, a szó szóosztályára utal.)

főnévi jelző ::= [tulajdon név (201--)]+
+[egyszerre főnév és melléknév (211--)]

számjelző1 ::= [kvantor (233--)]+
+[számnévi módosító (053--)]+
+[konkrét szám (230--)]+
+[mértékegység]

számjelző2 ::= [számnévi névmás (233--)]+[mértékegység]

számjelző3 ::= [melléknévi módosító (052--)]+
+[melléknév jellegű számnév (232--)]+
+[mértékegység]

tulajdonság jelző1 ::= melléknévi névmás (222--)

tulajdonság jelző2 ::= [tulajdonság jelző]+melléknévi szerkezet

tulajdonság jelző3) ::= [tulajdonság jelző]+
+melléknévi igeneves szerkezet

tulajdonság jelző4 ::= [tulajdonság jelző]+
+melléknévi névutós szerkezet

melléknévi szerkezet ::= [melléknévi módosító]+melléknév

melléknévi igeneves szerkezet ::= [melléknévi módosító]+
+[argumentumok és határozók]+
+melléknévi igenév

melléknévi névutós szerkezet ::= névszói kifejezés+[ESETRAG]+
+MELLÉKNÉVI NÉVUTÓ(223--)

argumentumok és határozók ::= [argumentum] ... [argumentum]
[hol][mikor] [módhatározó]

(Tapasztalatunk szerint a melléknévi igenév mellett leggyakrabban módhatározó áll).

Egyeztetés

Az alany, a tárgy és az ige

A magyar nyelvben az igének az igerag mindig megadja a grammatikai alanyát, és a tárgyas igék esetében a tárgyát is. Ez a következő lehet:

egyes szám 1. személy: a beszélő;

egyes szám 2. személyi a hallgatója, vagy egy a hallgatók közül;

egyes szám 3. személy: valaki más, vagy valamilyen más tárgy
vagy fogalom;

többes szám 1. személy: a beszélő és a hallgató vagy hallgatók;

többes szám 2. személy: a hallgatók;

többes szám 3. személy: más valakik vagy más tárgyak vagy fogalmak.

A grammatikai alanyon és tárgyon kívül a mondatban lehet explicit névszói kifejezéssel vagy névmással megadott alany és tárgy is. Ha van explicite megadott alany vagy tárgy, annak meg kell egyeznie a grammatikai alannyal. Az alábbi szabályok azt foglalják össze, hogy mikor kell kirakni az explicit alanyt vagy tárgyat. Ha kellene, de még sincs kitéve, azt utalásnak fogjuk tekinteni, amiről meg kell állapítanunk, hogy mire utal.

Az első és a második személy mindig jól meghatározott, anélkül, hogy bármi egyebet tudnunk kellene. Nem ez a helyzet a harmadik személlyel, akinek vagy aminek a pontosabb meghatározásához további információra van szükség.

A magyar nyelv általában kerüli a redundanciát és ezért az első és második személyben a névmási alanyt nem tesszük ki, hacsak nem akarjuk hangsúlyozni.

A harmadik személyben viszont általában kitesszük az alanyt a pontosítás miatt. Ha a harmadik személyű alany nincs kitéve, akkor ez rendszerint hivatkozás valamire vagy valakire, akit korábban a diskurzus fókuszába helyeztünk. Nem kell megadni az alanyt, ha az magától értetődő, pl. *Villámlik*.

Az igéknek van tárgyas és tárgyatlan ragozása. Nagyon kevés olyan ige van, amelyik alkalmilag sem áll tárgyas ragozásban. Pl. *lefutja a napi 10 kilométert, az igazak álmát aludta*, stb. Az ige tárgyatlansága nem a tárgy hiányából látszik, sokkal inkább abból, hogy múlt idejű melléknévi igeneve cselekvő.

A magyar igeragozásban a tárgyra a következő lehetőségek vannak: meghatározatlan, 1. személy, 2. személy, 3. személy. Látható, hogy a tárgy egyes vagy többes száma az igeragozást nem befolyásolja.

tárgy	határozatlan	1. személy	2. személy	3. személy
alany				
e.sz.	A:látok	T:látom	?:látlak	T:látom őt
1.sz.	valakit	magamat	téged	
e.sz.	A:látsz	A:látsz	T:látod	T:látod őt
2.sz.	valakit	engem	magadat	
e.sz.	A:lát	A:lát	A:lát	T:látja
3.sz.	valakit	engem	téged	őt/magát
t.sz.	A:látunk	T:látjuk	A:látunk	T:látjuk őt
1.sz.	valakit	magunkat	téged	
t.sz.	A:láttok	A:láttok	T:látjátok	T:látjátok
2.sz.	valakit	engem	magatokat	őt
t.sz.	A:látnak	A:látnak	A:látnak	T:látják
3.sz.	valakit	engem	téged	őt/magukat

Tradicionalisan az első oszlopot nevezik alanyi, az utolsót tárgyias igeragozásnak. Mivel az ige alanyának és tárgyának egybeesése viszonylag ritka, ezt explicite meg kell adni. Ezeket az eseteket elhagyva a következő szabályok alakulnak ki:

1. A tárgyias ragozás mindig harmadik személyű tárgyat vonz. Mivel a harmadik személyű tárgy nincs pontosan meghatározva, ezért megadása szükséges, illetve, ha hiányzik hivatkozásnak tekintjük.

2. Ha az alanyi ragozás egyes szám első személyű alakjánál hiányzik a tárgy és az igének kötelező vonzata a tárgy, az utalás valami határozatlanra.

3. Ha az alanyi ragozás egyes szám második vagy harmadik személyű, illetve többes szám második vagy harmadik személyű alakjánál hiányzik a tárgy, feltételezzük az *engem* vagy *minket* névmást. Pl. *Szeretsz?*

4. Ha az alanyi ragozás többes szám első személyű alakjánál hiányzik a tárgy, feltételezzük a *téged* vagy *titeket* névmást. Pl. *Megkeresünk.*

5. Ha az alany egyes szám első személyű, és a tárgy második személyű, a *téged* szót nem szokás kirakni, és néha a *titeket* se tesz ki. Viszont a főnévvel megadatott tárgy nem kap ragot. Pl. *Szeretlek, Éva.*

Ha a tárgy explicite is meg van adva, annak egyeznie kell tárgyasság tekintetében az igeraggal. Határozott a tárgy, ha van határozott névelője, ha tulajdonnév, ha birtokos vagy birtokjelet visel, illetve ha maga egy olyan névmás, amely tárgyias ragozást vonz, és nem tűr maga előtt határozott névelőt, vagy olyan névszói kifejezés, amelynek élén ilyen névmás áll.

(Megjegyzés: az igereszerűen ragozott segédigék esetén a segédige kerül tárgyias ragozásba. Pl. *Meg akarom a feladatot oldani.*)

Pl. *Nem olvastad a könyvet?*

Pistának nem olvastad a könyvét?

Nem olvastad a könyvét?

Pistának még nem olvastad könyvét? („Nem tudom van-e neki.”)

Melyik könyvet olvastad?

Mindkét könyvet olvastad?

Azt a könyvet olvastad?

Azt olvastad?

Határozatlan a tárgy, ha nem határozott. A valamennyi névmás mindkét csoportba beletartozik, de tárgyas ragozás esetén az „összeset” jelenti, tárgyatlan esetben „néhányat”. Pl. *Olvastad valamennyit? Olvastál valamennyit?*

Egyes szám harmadik személy kivételével a birtokjel után nem kötelező kitenni a tárgyragot. Pl. *Vettem a kalapom.*

Az alany, illetve a tárgy nemcsak egyetlen valami lehet, hanem felsorolás vagy „és”-kapcsolat is. Ebben az esetben a személy és szám a következőképpen alakul:

<i>én+te=én+ő=</i>	<i>én vagy mi</i>	<i>Én és Pista elmegyek oda. Én és Pista elmegyünk oda.</i>
<i>én+ti=én+ők=</i>	<i>mi</i>	<i>Én és Pistáék elmegyünk oda.</i>
<i>te+ő=</i>	<i>te vagy ti</i>	<i>Te és Pista elmész oda. Te és Pista elmentek oda.</i>
<i>te+ők=</i>	<i>ti</i>	<i>Te és Pistáék elmentek oda.</i>
<i>ő+ő=</i>	<i>ő vagy ók</i>	<i>Ő és Pista elmegy oda. Ő és Pista elmennek oda.</i>
<i>mi+te=mi+ő=mi+ti=mi+ők=</i>	<i>mi</i>	<i>Mi és Pista elmegyünk oda.</i>
<i>ti+ő=ti+ők=</i>	<i>ti</i>	<i>Ti és Pista elmentek oda.</i>

A számbeli egyeztetés kapcsán meg kell jegyeznünk, hogy számnevek után mindig egyes számot használunk, de csak a mondaton belül. Ha következő mondat ugyanezekre hivatkozik (ki nem tett explicit alannyal), akkor rendszerint már többes számot használunk. Pl. *A két gyerek elment hazulról. A mozi előtt találkoztak.* Hasonló a helyzet a csoportos jelentésű főnevekkel. Pl. *Osztályunk elment moziba. A 6. és 7. sorban ültünk.* Tehát megkülönböztetendő a mondatrész grammatikai száma (és személye) és a logikai száma (és személye) amivel később hivatkozhatunk rá.

(Megjegyzés: az *aki* névmásnak nincs személye, minden személyraghoz alkalmazható. Pl. *Én voltam az, aki megtudtam a dolgot.*)

A fentiekből látható, hogy az ige egyeztetése az alannyal és a tárggyal közepes méretű feladatnak látszik. Felmerül a kérdés, hogy szükséges-e ezt a feladatot az elemzés során elvégezni. Mivel nem az a célunk, hogy szöveg nyelvtani helyességét ellenőrizzük, és a magyar anyanyelvűek ezeket az egyeztetéseket aligha fogják elrontani, ennek csak akkor van értelme, ha az egyeztetéssel sikerül további információkat nyerni. Ez nyilván olyan szituációkban

fordulhat elő, amikor több alanyesetű (esetrag nélküli) mondatrész van a mondatban és eldöntendő, hogy mi az alany. Van olyan szerencsés szóhasználat, amikor eldönthető, hogy mi micsoda; máskor viszont nem.

Pl. *Elvesztettem a kalapom.* (A „kalapom” tárgy, mert van birtokjele és nem egyezik a grammatikai alannyal.)

Magad is látod. (Eldönthetetlen a „magad” lehet alany is tárgy is.)

Magad is látszol. (A „magad” nem lehet tárgy, mert tárgyas ragozást vonzana.)

Péter kenyerét eszi. (Eldönthetetlen, hogy Péter eszi-e, vagy Péterét.)

Péter kenyerét eszik. (Egyértelműen az utóbbi és az is látható, hogy az *ik* nem e.sz.3.sz., hanem t.sz.3.sz.)

A fentiekből látható, hogy az egyeztetés elvégzésére néha szükségünk van a kétértelműségek feloldásához. Viszont tapasztalatunk szerint a szöveg 80-90%-a e nélkül is egyértelmű. Elképzelésünk szerint vagy úgy kell implementálni, hogy az egyértelmű esetekben rendkívül gyors legyen, vagy csak a kétértelmű esetekben fogjuk ezt a vizsgálatot elvégezni.

A birtokos szerkezet egyeztetése

A nem hiányos birtokos szerkezet birtokosból és birtokból áll, és a birtok birtokjelet visel. Ez megadja a birtokos személyét és számát és a birtok számát is. Ha a birtokos első vagy második személyű, a birtokrag teljesen meghatározza, és így nem szükséges kirakni. Ha a birtokos ki van téve, egyeztetni kell a birtokjellel. A birtok száma alapján megkülönböztetjük a következő eseteket:

egyes számú birtokos - egyes számú birtok
egyes számú birtokos - többes számú birtok
többes számú birtokos - egyes számú birtok
többes számú birtokos - többes számú birtok.

Sajnos többes számú birtokos és egyes számú birtok esetén nem egyértelmű, hogy hány dologról van szó. Pl. *Elkérték a könyvünket.* (Lehet, hogy egy közöset, lehet, hogy mindenkitől egyet-egyet.) Ha ilyen kifejezés az alany, az állítmány egyes számú lesz. Viszont később többes számú hivatkozás is lehet rá (ha a második értelmezésről van szó). *Bemenetkor elkérték a könyvüket, és kilépéskor lepecsételve kapták vissza azokat.* Itt is megkülönböztetendő tehát a grammatikai és logikai szám, az utóbbi ilyen esetben bizonytalan.

Ha a birtokos számneves főnév, csoportos jelentésű vagy felsorolás, a birtokjelet a birtokos grammatikai számával egyeztetjük. Ilyen esetben, noha a birtokjel egy birtokost és egy birtokot mutat, mégis lehet logikailag többes számú. Pl. *Beszedte a három fiú könyvét, és azok elvesztek.*

Ha a birtokos felsorolás, amely több különböző személyből áll, az utolsóval egyeztetünk. Pl. *Csak a te vagy a tanár kérésére engedem el.*

(Megjegyzések:

- A t.sz.3.sz.-ben a birtokos szerepében nem az *ők*, hanem az *ő* személyes névmás áll. Pl. *az a könyvük.*

- A maga névmás e.sz.3.sz.-ű birtokjelet követel. Pl. *a magam dolga.*)

A birtokos szerkezet egyeztetésére különösen az egymástól elszakadt birtokos és birtok összekapcsolásánál lehet szükség.

A vonzatok

A szintaxisleírás igei és névszói szerkezetekkel foglalkozó része szóosztályokon operálva általános sémákban írta le a mondat szerkezetét. Vannak ezzel szemben a helyes mondatnak olyan szabályai is, amelyeket nem annyira az általános szabályok határoznak meg, hanem a benne szereplő egyes szavak. Például az *ad* ige lehetséges argumentumai: *hogymit ad, kinek ad, miért cserébe, milyen okból* vagy *milyen célból*. A *megy* ige lehetséges argumentumai ezzel szemben: *honnnan, hova megy, milyen útvonalon, milyen okból* vagy *milyen célból*. Az argumentumok egy részét az jelöli ki, hogy az illető argumentum valamilyen ragot vagy névutót vonz. Az igének ezek a vonzatai annyira kötődnek az igéhez, hogy ugyanezek lesznek a vonzatai az igeneveknek és az igéből képzett más szavaknak is. Nemcsak az igéknek, hanem más szavaknak, például mellékneveknek is lehet vonzata. Például, a *hasonló* szónál meg kell mondani, hogy *mihez*, a *generációs* szónál meg kell mondani *hanyadik*, a *hivatott* szónál, hogy *mire*.

A vonzatok egy sereg problémát vetnek fel. Az első és legfontosabb, hogy mi tartozik a vonzatok közé és mi nem. Vonzata-e az igének az alany, a tárgy és az összes határozó. A mi leírásunkban nem. Azokat a mondatrészeket, amelyek minden igéhez hozzákapcsolhatók nem tekintjük vonzatnak. (És hasonlóan a melléknévnek sem vonzata az *alig, kissé, majdnem* és hasonló szavak, mert ezek minden melléknévhez hozzá kapcsolhatók.) Mi csak azokat a szintaktikus egységeket tekintjük vonzatnak, amelyek arra az egyedi szóra jellemzők. Így hát nem vonzat az alany, a hely és az idő, ahol az ige végbemegy, a mód-, cél-, eredet-, eredmény- stb. határozók, amelyek minden igéhez hozzákapcsolhatók.

Mi a vonzatokat egy szótárban kívánjuk tárolni. Amikor a szintaktikai elemzés során megállapítottuk, hogy az adott szintaktikai szerkezetbe valamilyen konkrét szó került, és ennek vonzatai lehetnek, akkor ehhez a szótárhoz fordulunk, és ennek alapján keressük meg a kapcsolódó komponenseket.

Természetesen nem ez az egyetlen lehetséges megközelítése a szintaxis leírásának. Elképzelhető lenne egy olyan leírás is, ahol a szavak olyan részletesen lennének osztályozva, hogy egy osztályba csak az éppen azonos vonzatokkal rendelkező szavak kerülnének. Ekkor a szintaxisleírásba konkrétan bele lehetne írni a vonzatokat. Ezzel szemben a szavak felosztása igen sok csoportra történne, és egy-egy szó igen sok csoportba beletartozna. (Mert ugyanannak a szónak több különböző vonzatprofilja is lehet.) Ezzel együtt a szintaxisleírás is jelentősen megnagyobbodna.

A másik véglet az lenne, hogy a szótárban minden egyes szónál megfelelő részletességgel le lenne írva, hogy milyen más szavakhoz kapcsolódhat. Ez esetben külön szintaxisleírásra nem lenne szükség.

Tisztán áttekinthetőség, kezelhetőség és hasonló gyakorlati szempontok alapján választottuk ezt a leírást. Úgy véljük, megadható a mondatoknak egy általános leírása, és ettől elválasztva az adott témakör szavai, kiegészítve azokkal az információkkal, hogy az adott téma esetén, hogyan lehet ezeket a szavakat felhasználni (azaz milyenek a vonzataik).

A következő kérdés, hogy milyen vonzatai lehetnek egy szónak. Az igék esetében a vonzat leggyakrabban egy ragos vagy névutós névszói csoport. Gyakori az olyan ige is, amelyik egy másik igtét vonz, főnévi igenévi alakban. Pl. *Elfelejtett bevásárolni. Megtanult traktort vezetni.* Továbbá lehet a vonzat teljes mellékmondat is. Pl. *Elhiszi, hogy ...* Melléknevek esetén lehet a vonzat számnév, pl. *hány méteres,* sorszámnév, pl. *hanyad osztályú,* melléknév, pl. *milyen minőségű,* tulajdonnév, pl. *milyen nevű* stb. De itt is gyakori a ragos névszói kifejezés pl. *mire képes.*

Kezelhetőségi szempontok alapján érdemes bizonyos vonzatokat egyetlen vonzatosztállyá összevonni, például egyetlen osztálynak venni a honnan kérdésre válaszoló *-BŐL, -TŐL, -RŐL* ragos vagy a *mögül, alól, felől,* stb. névutós névszói kifejezéseket. (Természetesen ideértve a *tőle, róla, mögülem* stb. alakokat is.)

Ezek szerint a vonzat lehet: egy konkrét szó, egy szóosztály, egy szóosztály valamilyen konkrét végződéssel vagy névutóval, főnévi igenév, mellékmondat, vonzatosztály. (A vonzatrendszer leírása körülbelül azonos azzal, amit az Apreszján-Páll-féle orosz-magyar igei vonzatszótár alkalmaz.)

(Megjegyzés: Bizonyos esetekben, pl. a között igekötővel kijelölt argumentumnál szükséges, hogy az logikailag többes számú legyen. Ilyenkor az argumentum vagy többes számú, vagy felsorolás, vagy több argumentum kötőszóval összekapcsolva.)

Argumentum és vonzat

Mint azt már korábban megállapítottuk, az igéknek argumentumaik vannak, amelyek bizonyos szemantikus szerepeket töltenek be. A szintaxisanalízis egyik fő feladata, hogy ezeket az argumentumokat megtalálja. Az argumentumok felismerése a vonzatokon keresztül történik. Az adott argumentumhoz adott vonzat(ok) tartoznak, és általában ezek jelölik ki az argumentumot, de sajnos az argumentum és vonzat között nincs egyértelmű kapcsolat. Egyrészt ugyanazt az argumentumot több különböző vonzattal is meg lehet adni. Például: *A folyadékokat literrel mérjük.* vagy *A folyadékokat literben mérjük.* Másrészt ugyanaz a vonzat több különböző, argumentumhoz is tartozhat, pl. *Gyomorfekéllyel kezelte magát.* *A híres belgyógyással kezelte magát.*

Tovább nehezíti a helyzetet, hogy például a *-NAK* rag nemcsak igei vonzatként léphet fel, hanem az elváló birtokos ragjaként is, és bizonyos segédigék esetében az alany is *-NAK* ragot kap.

A melléknevek és melléknévi igenevek esetén a vonzat mindig a szó előtt áll, de ha több vonzat is van (ami a gyakorlatban igen ritkán fordul elő) ezek sorrendje általában nem meghatározott. Pl. *a főnököket beadványaival ostromló beosztott, a beadványaival a főnököket ostromló beosztott.*

Amikor a szintaxiselemzés során eljutunk egy olyan pontra, hogy egy adott ige argumentumait össze kell szednünk, több olyan szószerkezetet is találhatunk, ami potenciálisan argumentum lehet. Ekkor négy eset lehetséges: vagy több argumentumunk van (felsorolás vagy kötőszavas szerkezet formájában), pl. *Pista vagy Jancsit vagy Jóskát látta a piacon.*, vagy egy argumentumunk van hátravetett jelzővel, pl. *Pistát, régi barátját látogatta meg.*, vagy valamelyik argumentum, a másik pedig nem, pl. *Pistának kellett enni adnia a disznóknak.*, vagy egyik sem argumentum, pl. *Pistának a levelét nekem kellett elvinnem.* (Feltéve, hogy Pista írta és én viszem.).

Az ilyen mondatok szintaktikailag kétértelműek. A kétértelműségek kiküszöbölésére a leghatékonyabb és legbiztosabb út, a szemantikus analízis, amely megállapíthatja, hogy bizonyos mondatfelépítések értelmezhetők mások viszont nem. Például, ha tudjuk, hogy az *ad* ige egyik lehetséges vonzatformája: *ad+személy+élőlény-NEK+(enni/inni)*, akkor a *Pistának a disznóknak kellett enni adnia.* mondatban egyértelműen *Pista* lesz az ágens és a *disznók* a datívusz.

A fenti szemantikai elemzéshez ún. szemantikai jegyek bevezetésére van szükség. Noha bevezethetők lennének olyan általános szemantikai jegyek, amelyek általában témakörötől és nyelvtől függetlenek, mi a gyorsabb és pontosabb elemzés érdekében erősen a témakörhöz kötött jegyek alkalmazására gondolunk. (A szemantikai jegyekre egy későbbi fejezetben visszatérünk.)

A másik kérdés, hogy a fent leírt analízist mikor végezzük; először összeállítjuk-e az összes lehetséges fát, és azután végezzük-e el az analízist, vagy már a vonzatok keresése közben tekintettel vagyunk a szemantikai jegyekre. Hatékonysági megfontolások alapján az utóbbi megoldás látszik célszerűnek.

A kétértelműségek feloldásában a szemantikus jegyeken kívül más ad hoc szabályok alkalmazása is szóba jöhetnek, például azé a szabályé, hogy a birtokjeles szó előtt álló *-NAK* ragos névszó nagy valószínűséggel birtokos lesz. Ilyen és hasonló szabályok alkalmazása azért problematikus, mert a magyar nyelv szabályait nem egy programmal akarjuk definiálni. Hanem egy géptől és programnyelvtől független formalizmussal, amelybe az ad hoc megoldások nehezen illeszthetők bele.

Kötelező és opcionális argumentumok, defaultok

A továbbiakban vonzatprofilnak fogjuk nevezni a vonzatoknak egy olyan maximális készletét, ami egy mondaton belül lehetséges, de további elemekkel nem bővíthető. Egy szónok több vonzatprofilja is lehet. Két vonzatprofil akkor különbözik egymástól, ha az egyikben van olyan vonzat, ami a másikban nincs, vagy ha ugyanaz az a vonzat más szerepet játszik az egyikben, mint a másikban. Pl. *valami valamire szolgál, valaki valakit szolgál; összehasonlít valamit valamivel, összehasonlít (több) valamit; feltölt valamit valamilyen anyaggal (vízzel töltötte fel a medencét), feltölt valamit valamilyen eszközzel (vödörrel töltötte fel a medencét).*

A vonzatprofilon belül vannak kötelező és opcionális argumentumok. Kötelező az az argumentum, amelyik szükséges a mondatok értelmezéséhez. és opcionális, amelyekre nincs feltétlenül szükségünk. (Mint látható, a fenti definíció igen szubjektív, és nagyon függ attól, milyen adatkészleten fogunk dolgozni.) Ha egy opcionális argumentum hiányzik, helyére egy előre meghatározott default elemet teszünk. Van egy univerzális elemünk; a meghatározatlan. Az esetek jelentős részében ez lesz a default, de nem mindig. Pl. *Pista elmegy moziba. elmegy+honnan+hova+mivel* és mindhárom argumentuma opcionális. A *mivel* argumentum meghatározatlan lesz, a *honnan* ezzel szemben az alany aktuális helye lesz.

Ha ezzel szemben egy kötelező argumentum hiányzik, akkor ennek helyébe egy speciális jelzés kerül, ami arra utal, hogy a mondat hiányos, és ez a hiány egy implicit hivatkozás valamire, amit meg kell találnunk. (Ezzel a kérdéssel a hivatkozásokról szóló fejezet foglalkozik.)

(Megjegyzések:

Kötelező argumentum lehet olyan „Hol?” kérdésre válaszoló helyhatározó vagy időhatározó is, amit más esetekben nem tekintünk argumentumnak. Pl. *valami valahol található.*

Az ige összes argumentuma az igeneveknek és az igéből képzett -ás, -és végű főnévnek is argumentuma, sőt a hely és időhatározó is argumentuma lesz ezeknek, de egyik argumentum sem kötelező.

)

A szintaxis analízis megvalósítása

Mint azt az előző fejezetekből láttuk, a magyar nyelv szintaxisának két szintje van. Az alsóbb szinten a névszói szerkezetben a pozicionális meghatározottság majdnem teljes, ezzel szemben a felsőbb szinten az összetartozás nagyrészt ragokkal (és névutókkal) van definiálva. Ez indokolja azt, hogy a két szint analízisét elkülönítsük, és két különböző módszerrel oldjuk meg. Mindkét feladat során fellép viszont egy közös probléma: a vonzatok megtalálása. A kötött szórendű részben a vonzatok a vonzattal rendelkező szó elé kerülnek, a szabad szórendű részben viszont végződésük alapján kell a szóhoz kapcsolni őket.

Elképzelésünk szerint a szintaxisfa felépítése két menetben történne: az egyik menetben összeszednénk pozicionális alapon a névszói kifejezések részeit, a következő menetben pedig „felülről” az ige felől elindulva, összeszednénk az igei kifejezést, az ige vonzatait és a szabad határozókat.

A névszói kifejezés komponenseinek összeszedése egy ATN, vagy valami ehhez hasonló elemző eszköz segítségével történhet.

(Az ATN (Augmented Transition Network) mind a természetes nyelvek, mind a programozási nyelvek szintaxisának jól ismert leíró/elemző eszköze. Az elemek sorrendjét egy végesautomata-szerű sémával írjuk le, ahol azonban az egyes állapotok közötti átmenet vagy egy szimbólum hatására történik vagy az átmenetet magát is egy hasonló automatával írhatjuk le.

Ezenkívül arra is lehetőséget nyújt, hogy globális változókon keresztül információkat továbbítsunk.)

Mivel a magyar nyelvben a pozicionálisan kötött vonzatok mindig a lexikai elemtől balra állnak, célszerűnek látszik, ha az elemzés a szokásostól eltérőleg jobbról balra halad.

Ezek szerint a névszói kifejezések megkeresése vázlatosan a következő lépésekben történik. Beolvassuk a mondatot és elvégezzük a morfológiai analízist. Ezek után, a legutolsó esetragból vagy névutóból indulunk ki. Ez előtt egy főnév állhat, az előtt egy anyag-, nemzet- vagy foglalkozásnév, az előtt egy tulajdonnév, az előtt tulajdonságjelzők, az előtt számjelző, az előtt névelő. Ha valamelyik elmarad, csak az előtte álló következhet. Lehetnek azonban a sorban vonzattal járó szavak is. Ilyenek az igéből képzett főnév (pl. *iskolába járás*) a melléknévi igenevek, a melléknévi névutó (pl. *a ház előtti fa*), vonzattal járó melléknév (pl. *eredményeire büszke vezető*). A vonzat lehet kötelező vagy opcionális. (Pl. a *méretű* melléknév vonzata kötelező, a *büszke* szóé opcionális.) A vonzat előírja, hogy milyen kategóriájú szó vagy kifejezés állhat az adott szó előtt. Az esetek nagy részében a vonzat maga is egy ragos névszói kifejezés. Az ilyen beágyazott szerkezetek elemzése után vissza kell térnünk a magasabb szintű szerkezet elemzésére. Amikor a teljes névszói kifejezést megtaláltuk, ugyanezt az

eljárást megismételve további névszói kifejezéseket keresünk a mondat megmaradt elején.

Ez a módszer két technikai problémát vet fel. Az első az, hogy az alanyesetnek nincs ragja. Ezért az esetrag nélkül álló soron következő névszót úgy kell kezelnünk, mint egy potenciális névszói kifejezés végét.

A második probléma, hogy ha a névszói kifejezés egy opcionális argumentummal kezdődik, akkor ez vajon az utána álló vonzatos szóhoz tartozik-e, vagy esetleg egy magasabb szintű konstrukcióhoz, például az igéhez. Pl. Az *utcán játszó gyerekeket látok*. (Én vagyok az utcán, vagy a gyerekek, vagy mindannyian ott vagyunk.) Ehhez hasonlóan meg kell vizsgálni, hogy ha a soron következő szó többértelmű, az adott pozíció meghatározza-e egyértelműen, hogy az adott esetben melyik értelme jön szóba. Az analízisnek biztosítania kell, hogy különböző összevonások jöhessenek létre és a több lehetséges értelmezés megmaradjon.

Ezek után az analízis úgy mehet végbe, hogy megállapítjuk, milyen igei konstrukció szerepel a mondatban (ragozott ige, ragozott segédige, nem ragozott segédige, explicit VAN ige, ki nem tett VAN ige stb.). Majd megállapítjuk az ige vagy az egyéb vonzattal rendelkező szó vonzatait, és megvizsgáljuk, hogy a többi kifejezés szabad határozónak tekinthető-e.

(Megjegyzés: ez a szintaxisleírás nem foglalkozik a határozói igenévvel általában, és általában a szabad módhatározókkal (állapot-, ok-, cél-, eredmény-, stb.). Ennek az az oka, hogy ezek szemantikus értelmezése problémát jelent.)

A program mindenesetre három komponensből fog állni: egy ATN-szerű pozicionális szabályokat kezelő eszközből, egy vonzatokat kezelő eszközből, és egy a szabad sorrendű szabályokat kezelő eszközből.

Felmerül a kérdés: meg lehet-e a teljes elemzést valamilyen ismert és egységes eszközzel oldani. Szóba jöhet például valamilyen ismert CF-elemző. Ezekkel az a probléma, hogy a programozási nyelvekre létrehozott elemzők a balról jobbra egyértelműen elemezhető nyelvekre készültek. A LISP nyelv alkalmazása csak a programozás stílusát változtatná meg ahhoz képest, mintha egy Pascal vagy MODULA nyelvben implementált listakezelőt használnánk. Felmerül a lehetőség a PROLOG nyelv alkalmazására, meg kell azonban vizsgálnunk, hogyan írhatók le a szabad szórendet kifejező szabályok. Elképzelhető még egy Production System jellegű rendszer alkalmazása is. Az utóbbi kettőnél a nyelv szintaktikai szabályai alkotnák a program egyik fő komponensét. Szándékunkban áll ilyenekkel kísérleteket végezni, de tartunk tőle, hogy a hatékonyság nem lesz megfelelő.

Az utóbbi kettőnél az a fő probléma (amely előny a kísérleti megvalósításoknál), hogy a programozó csak azt írhatja elő, hogy milyen összefüggéseket kell megállapítani, és nem befolyásolhatja, vagy csak igen nehezen, hogy milyen stratégiával, milyen sorrendben kell az összefüggéseket megkeresni. A minden utat végigjáró és a zsákutcákból visszalépő algoritmusok igen általánosak és kevésbé hatékonyak. A hatékonyság érdekében viszont arra kell törekednünk, hogy amit csak lehet, egyetlen menetben állapítsunk meg.

Ha a szintaktikai elemzés valamilyen tradicionális nyelven valósul meg, gondoskodni kell a szintaxisleírás könnyű gyors változtatásáról, javíthatóságáról. olyan megoldást kell választani, hogy vagy az elemző program adatként kapja meg a szintaxisleírást. vagy (és ez a valószínűbb) az elemző programot egy generátor-program állítsa elő a szintaktikus leírásból.

Hivatkozások

A természetes nyelv egyik ismert tulajdonsága, hogy nem beszélünk kerek mondatokban, vagyis nem adunk meg minden információt, amely a mondat megértéséhez szükséges, hanem előzőleg megadott információkra hivatkozunk.

A hivatkozások feloldása igen bonyolult feladat, mert a hivatkozásoknak számos fajtája van, és kezelésük rendkívül szerteágazó problémákat vet fel. A következőben összefoglaljuk a legfontosabb problémákat.

A hivatkozásoknak a magyar nyelvben is (más nyelvekhez hasonlóan) három alapvető formája van. Az első forma a névmás, amely definíciója alapján egy névszói kifejezést pótol (vagy annak egy részét). A második forma a hiány, nyelvészi szakkifejezéssel az ellipszis, amikor az adott mondatrész hiányzik a mondatból, de oda kell érteni. Pl. *Látod azt az embert? Látom.* A harmadik lehetséges forma a határozott névszói kifejezés. Ilyen esetben a határozott névszói kifejezés utal egy másik névszói kifejezésre, amely korábban már szerepelt a szövegben. Pl. *Van-e folyó New Yorkban? Van. Mi a neve a folyónak?*

A harmadik fajta (a határozott névszói) hivatkozás a legproblematikusabb mert, mert nem minden határozott névszói kifejezés hivatkozás. Pl. *Az ember halandó.* mondatban az *ember* nem utal semmire. Sajnos egyelőre semmilyen leprogramozható kritériumot nem tudunk mondani arra, hogy mikor visszautaló egy névszói kifejezés.

A visszautalás természetesen nemcsak ugyanazzal a szóval történhet, mint amit a korábbiakban használtunk, hanem szinonimákat is használhatunk, sőt mindenféle más azonos értelmű kifejezést is. Pl. *Pista és Jóska bement a moziba. A két gyerek a 10. sorban ült le.* A visszautaló névszói kifejezés gyakran maga is hiányos. Pl. *Melyik nyakkendőt vegyem meg? A kéket!*

Ha a számítógéppel dialógust folytatunk, természetesen nemcsak azokra a dolgokra hivatkozhatunk vissza, amit mi kérdeztünk, hanem azokra a dolgokra is, amit a gép felelt.

Ezen túlmenően, nemcsak olyan dolgokra hivatkozhatunk, ami explicite előfordult a korábbi szövegben, hanem hivatkozhatunk egy egész mondatra egyetlen szóval, sőt olyan dolgokra is, amelyekről egyáltalán nem volt szó, de a korábbiakból kikövetkeztethetők. Pl. *Pista repülővel utazott Amerikába. - Milyen hosszú volt az utazás? Mennyibe került a jegy?*

Vannak ugyan olyan természetes nyelvű interfészek, amelyek ezeket a visszahivatkozásokat egyáltalán nem kezelik, de ezek tulajdonképpen igen távol vannak a természetes nyelvtől. Manapság egy interfész tudását nagyrészt aszerint ítélik meg, hogy milyen szabadságot nyújt ezen a téren.

Noha a probléma teljes általánosságban való kezelése rendkívül nagy apparátust igényel, bizonyos eredmények már viszonylag egyszerű eszközökkel is elérhetők. Ezek közé tartoznak a nyelvtani szabályok és a szemantikai egyeztetés.

Az európai nyelvekben a leggyakrabban alkalmazott eszköz a nemben, és számban történő egyeztetés. Ez sajnos a magyarban egyáltalán nem alkalmazható, mert a magyarban nincs nem, a számbeli egyeztetés ingadozik az ún. logikai és grammatikai egyeztetés között. (Mondaton belül az egyeztetés inkább grammatikai, mondatok között általában logikai szám szerint egyeztetünk. Lásd „Egyeztetések“.) Hasonlóan, élettelen dolgokra is gyakran hivatkozunk ő névmással és élőkre az-zal.

A nyelvtani szabályok közül a következőkre lehetünk tekintettel:

Ha úgy képzeljük, hogy a mondat egy ige szerkezetből és annak bővítményeiből áll (alanyból, tárgyból, határozókból) ezek hivatkozhatnak egymásra és a következő mondat is hivatkozhat rájuk. Az ezekben beágyazott igei szerkezetek (almondatok és igeneves szerkezetek) bővítményei hivatkozhatnak, a főszerkezet bővítményeire, és saját társ bővítményeikre, de nem hivatkozhatnak egy másik beágyazott szerkezet belső elemére. A következő mondat sem hivatkozhat az előző mondatba beágyazott szerkezet belsejére.

Ez alól némileg kivételt képez a fölső szintű bővítmény birtokosa, amire mégis szokás hivatkozni.

Arra nézve, hogy egy visszahivatkozásnál mikor használunk hiányt, mikor személyes- és mikor mutató névmást, valamint ezekhez hasonló dolgokra É. Kiss Katalinnál és Pléh Csabánál számos hipotézist találunk. (Pl. Ha két egymást követő mondat alanya azonos, akkor a második mondatban nem tesszük ki. Ha a második mondat alanya azonos valamivel, ami az előző mondatban nem alany, akkor az az névmással hivatkozunk rá.) Nekünk azonban úgy tűnik, hogy ezek a szabályok inkább tendencia jellegűek, és csak szemantikus egyeztetéssel együtt alkalmazhatók.

A szemantikai egyeztetés lényege, hogy minden névszói kifejezés bizonyos szemantikai jegyeket hordoz, az a hely pedig, ahol hivatkoznak rá, bizonyos szemantikai jegyeket kíván meg. A hivatkozottat tehát azok között kell keresni, akik a szükséges szemantikai jegyeket hordozzák.

Fontos megfigyelés még, hogy az említett dolgokra, ha újra nem került rájuk szó, az idő előrehaladtával egyre kevésbé lehet hivatkozni, bár a „felejtés” gyorsasága emberenként változó.

Ezek alapján az első programváltozatban a hivatkozások feloldását a következő lépésekben képzeljük el: amikor a számítógép kap egy mondatot, a főmondat bővítményeit feltesszük egy hivatkozási listára. Az eredményül kapott objektumokat szintén erre a listára tesszük. A következő mondatban szereplő objektumokat szintén erre a listára tesszük. Ha a mondatban hivatkozás van, akkor ezek közül keresünk olyat, ami szemantikailag megfelelő jegyeket hordoz. Ha több ilyen találunk, vagy egyet se, visszakerdezzük a felhasználótól. Ha egyet találunk, azt tekintjük a hivatkozottnak. A következő kérdés előtt a kettővel korábbi kérdést és a rá adott választ töröljük a hivatkozható dolgok listájáról.

Szemantikus reprezentáció és szemantikus jegyek

A szemantikáról igen nehéz beszélni a konkrét témakör ismerete nélkül. Elképzelhető ugyan egy olyan szemantikus definíció, amely szélességében (sok témakörre kiterjedően) eléggé jól átfogja a teljes magyar nyelvet, de ez a definíció nagyon az általánosságok szintjén fog mozogni, azaz mélységében igen sekély lesz. Nyilvánvaló, hogy a természetes nyelv egész mást jelent egy matematikusnak, nyelvésznek vagy jogásznak, amikor saját szakterületéről van szó (ilyenkor talán meg sem értik azt, amiről a másik beszél), és mást, amikor egy közös témakőről, például egy külföldi utazásról van szó.

Mint azt már korábban többször lerögzítettük, mi a programunkban egy témakör- és feladatspecifikus szemantikára gondolunk, és mivel a feladat jelenleg nincs még kiválasztva, nagyon nehezen tudunk róla beszélni. Az alábbiakban példákon keresztül fogjuk megvilágítani, mire is gondolunk, amikor szemantikáról beszélünk.

Tegyük fel, hogy van egy geometriai alapú rajzoló programunk, amelyet természetes nyelvű utasításokkal próbálunk vezérelni, olyan parancsokat képzelhetünk el, mint: *Kösd össze a háromszög AB oldalának felezőpontját a szemben lévő csúccsal, Hosszabbítsd meg az AB szakaszt a három és félszeresére!, Rajzolj egy szaggatott felező vonalat az AB szakasz közepére!* stb.

Nyilvánvaló, hogy mind a három fenti parancs a rajzoló programnak ahhoz a lehetőségéhez kapcsolódik, hogy húzz vonalat valahonnan, valameddig. (Azt, hogy milyen legyen ez a vonal: szaggatott, kék, vastag, stb. az rendszerint egy másik utasítása a rajzoló programnak. Hasonlóan, a felezőpont, a szemben lévő csúcs megkeresése, sőt esetleg az A és a B pont koordinátáinak megkeresése, stb. külön tennivaló, amelyre vagy a rajzoló programot kell utasítani, vagy az interfésznek kell megoldania.) Mind a három mondatot vissza kell tehát vezetnünk egy olyan sémára amelyben adva van egy parancs, hogy „rajzolj” és három argumentum: mit (egyenes szakaszt, körívet, illeszkedő görbét, stb.), mitől, meddig. Az argumentumok meg lehetnek adva közvetlenül felhasználható operandusokkal, vagy újabb parancsok sémáival, amelyek végrehajtása a szükséges argumentumot szolgáltatni fogják.

A szemantikus reprezentáció elkészítése során, azt is ellenőriznünk kell, hogy a megadott parancs értelmes és teljes-e. Nem fogadhatunk el olyan parancsokat, mint: **Húzzál háromszöget az A és B pont közötti,* **Kösd össze az AB szakasz felezőpontját piros ponttal!,* **Felezd meg a háromszöget!* stb. Ennek az ellenőrzésnek két oka is lehet: vagy arról van szó, hogy a rajzoló program bemenete primitív (pl. csak abszolút koordinátákat tud összekötni), ekkor nekünk magunknak kell erre a primitív nyelvre a mondatot lefordítanunk, és ez hibás parancs esetén nem sikerülhet; vagy a rajzoló program bemeneti nyelve bonyolult dolgokra is lehetőséget ad, és látszólag át tudnánk írni rá a hibás parancsot is, de a

rossz parancs rossz működést váltana ki, és ennek következményeit kellene viselnünk (pl. a meghívott program abortál).

A parancs akkor értelmes, ha jók az argumentumai. Ellenőriznünk kell, hogy minden argumentum megvan-e, és olyan-e, mint amilyennek lennie kell. Esetünkben vannak síkidomok, vonalak, pontok, színek, vonalstílusuk (folytonos, szaggatott, pontozott, stb.), vonalvastagságok, távolságok, stb. Ellenőrizni kell, hogy ahova pont szükséges, oda pont kerüljön, ahova szín, oda szín, ahova távolság, oda távolság.

Éppen ezért a felmerülő fogalmakat osztályokba kell sorolni. Az osztályba sorolás hierarchikus is lehet, például: geometriai fogalmak, azon belül síkidom, vonal, pont; a síkidomokon belül négyszög, háromszög, kör; a négyszögon belül paralelogramma, téglá, négyzet, stb. Azt, hogy egy fogalom milyen osztályba illetve osztályokba tartozik, szemantikus jegyeknek nevezzük.

A parancsokat leíró sémák előírják, hogy milyen argumentum pozícióba milyen szemantikus jegyeket viselő argumentum szükséges. Például a rajzolj parancs mit argumentuma csak valamilyen vonal lehet, a másik két argumentuma csak pont.

Természetesen egy másik témakör esetén mások lesznek a parancssémák és mások lesznek a szemantikus jegyek. Például egy térképészeti adatbázis esetén a következő fogalmakkal dolgozhatunk: ország, város, főváros, hegy, víz, folyó, tó, terület, hosszúság, lakosság. Megkérdezhetjük, hogy egy tónak mekkora a területe, de nem kérdezhetjük meg, mekkora a lakossága. Egy folyóról megkérdezhetjük, mekkora a hossza, de nem kérdezhetjük meg, mekkora a területe. stb.

Felmerül a kérdés, hogy a szemantika leírására szolgáló sémákat milyen formában ábrázoljuk a számítógépben. Több megoldás képzelhető el, a logikai formuláktól kezdve, a fogalmi hálókig. Mi pillanatnyilag frame-ekben képzeljük el a megvalósítást.

Ezt az ábrázolást azután le kell fordítanunk annak a programnak az input nyelvére, amelyhez az interfészt készítettük. Egy mondatból az esetek nagy részében több utasítás is lesz, ezeket az utasításokat olyan sorrendben kell kiadni, hogy a későbbi a korábbi eredményét fel tudja használni. Ezt a fordítást egy mechanikus tevékenységnek képzeljük el, melynek mérete azért nem elhanyagolható.

Melléklet a kísérleti morfológiai elemző toldalékairól.

A melléklet három táblázatot tartalmaz. Az első az általunk kezelt tőváltozásokat, a második a tővégi betűhalmazokat, a harmadik pedig magukat a toldalékokat sorolja fel toldalékosztályonként. A toldalékok előtti szám a toldalékot megelőző betűhalmazát (esetleg betűhalmazok metszetet) jelöli, míg az azt követő szám(ok) a lehetséges tőváltozás(oka)t jelöli(k).

Tőváltozások (inverz):

- 1 igei hasonulás *zz ss ssz ddz ggy* után, + igei *s-t* (opcionális) (*fut, de tát fos*)
- 2 névszói *áa ée*, ill *+a +e* (opcionális mert *lé, -é*, de *á-nál* kötelező)
- 5 névszói hasonulás dupla mássalhangzó után (ha nem talál olyan tövet opcionális)
- 4 névszói *aó eő +ó +ő* (*ajtó apró tető erdő*)
- 5 igei *v* előtt *oá öő úú uú úú íí íí +sz* (*ró lő nyű fú rí alsz esz isz*)
- 6 névszói *v* előtt *oó aó öő őő uú üú űű eé +u +ü* (*ló tó mű bő tő fű szú lé falu tetű*)
- 7 *ú ű* kiesés (opcionális) (*borjú fiú lassú hosszú szörnyű*) + névszói belső *eé aú uú űű íí* (*tér nyár úr tűz híd*)
- 8 *o e Ö* kiesés (opcionális) *jlrms* (*irodalom forog késelem képez köröm sör*)

Betűhalmazok:

- 0 -aábcdeéfgghiiijklmnoóöőpqrstuúüúvwxyz
AÁBCDEÉFGHIÍJKLMNOÓÖŐPQRSTUÚÜÚVWXYZ0123456789 (bármilyen)
- 1 -aábcdeéfgghiiijklmnoóöőpqrstuúüúvwxyz
A E Ó Ő (magyar vég)
- 2 - bcd fgh jklmn pqrst vwxyz (mássalhangzók)
- 3 - á é hi j l noó ő rs uúüúv yz
Ó Ő (félmogánhangzók)
- 4 -a e i l o öő r tu üú
A (v-s névszó előtt)
- 5 b de fghiiijklmnoóöőp rstuúüúv yz
E I (igető vég)
- 6 b d fgh íjklmn ó őp r ú úv y (-e-i-s-t-z)
- 7 -aá eé hií oóöő q uúüú y
A E Ó Ő (magánhangzók)
- 8 -aá íijklmnoó rs uú z
A Ó (Mély magánhangzók)
- 9 - eé íijklmn öő rs üü z
E Ő (Magas magánhangzók)
- 10 - ábcd éfghiiijklmnoóöőpqrstuúüúvwxyz
Ó Ő (-a-e)

Toldalékok:

*Birtokos	*névszónévszó	*névszóige	*ígenévszó
0 0	0 i 8	1 ít 8 7	2*5 amdó 8
10 é 2	0 as 8 7	2 az 8 7	2*5 atlan 8
10 éi 2	0 es 8 7	2 ez 8 7	2*5 ás 8
	0 os 8 7	2 oz 8 7	2*5 endő 8
	0 ös 8 7	2 öz 8 7	2+5 etlen 8
*igeige	2 atlan 8 7	7*10 z 2	2*5 ett 8
2*5 eget 8	2 etlen 8 7		2*5 és 8
2*5 ogat	2 talan 0		2*5 ott 8
2*5 öget 8	2 telen 0		2*5 ó 8
5 at 0	7*8*10 tlan 2		2*5 ött 8
5 et 0	7*9*10 tlen 2		2*5 ő 8
5 gat	4 vas 6		2*5 t 0
5 get 0	4 vatlan 6		7*5 tt 0
5 hat 0	4 ves 6		8*5 vandó 5
5 het 0	4 vetlen 6		8*5 vatlan 5
5 tat 0	7*10 s 2		8*5 vás 5
5 tet 0			9*5 vendő 5
5*7*9 veget 5	*fokozó		9*5 vetlen 5
5*7*8 vogot	2 abb 8 7		7*5 vett 5
5*7*9 vöget 5	2 ebb 8 7		9*5 vés 5
	2 b 0		7*5 vott 5
	4 vabb 6		8*5 vó 5
	4 vebb 6		7*5 vött 5
	7*10 bb 2		9*5 vő 5
*eset			
0 0	10 ba 2	10 ban 2	10 be 2
10 ben 2	10 ból 2	10 ből 2	10 ért 2
10 hez 2	10 hoz 2	10 höz 2	10 ig 2
10 ként 0	10 kor 0	10 nak 2	10 nál 2
10 nek 2	10 nél 2	10 ra 2	10 re 2
10 ról 2	10 ről 2	10 tól 2	10 től 2
2 _al 3	2 _á 3	2 _el 3	2 _é 3
2 an 8 7	2 at 8 7	2 en 8 7	2 et 8 7
2 on 8 7	2 ot 8 7	2 ön 8 7	2 öt 8 7
3 t 2	4 vat 6	4 vet 6	4 von 6
4 vön 6	7*10 n 2	7*10 vá 2	7*10 vé 2
	7*8*10 val 2	7*9*10 vel 2	

*igerag

5	0	5 ja	0	5 ják	0	5 játok	0
5 juk	0	5 jük	0	5 lak	0	5 lek	0
5 na	0	5 nek	0	5 ná	0	5 nád	0
5 nák	0	5 nál	0	5 nálak	0	5 nám	0
5 nának	0	5 nánk	0	5 nátok	0	5 ne	0
5 nek	0	5 né	0	5 néd	0	5 nék	0
5 nél	0	5 nélek	0	5 ném	0	5 nének	0
5 nénk	0	5 nétek	0	5 ni	0	5 nia	0
5 nie	0	5 niuk	0	5 niük	0	5 nod	0
5 nom	0	5 notok	0	5 nöd	0	5 nöm	0
5 nötök	0	5 nunk	0	5 nünk	0	5 sz	0
5 tok	0	5 tök	0	5 va	0	5 ván	0
5 ve	0	5 vén	0	5*2 _	1	5*2 _a	1
5*2 _ad	1	5*2 _ak	1	5*2 _alak	1	5*2 _em	1
5*2 _anak	1	5*2 _atok	1	5+2 _ák	1	5*2 _ál	1
5*2 _átok	1	5+2 _d	1	5*2 _e	1	5*2 _ed	1
5*2 _ek	1	5*2 _elek	1	5*2 _em	1	5*2 _en	1
5*2 _enek	1	5*2 _etek	1	5*2 _ék	1	5*2 _él	1
5*2 _étek	1	5*2 _on	1	5*2 _uk	1	5*2 _unk	1
5*2 _ük	1	5*2 _ünk	1	5*2 ana	8	5*2 anok	8
5*2 aná	8	5*2 anád	8	5*2 anák	8	5*2 anál	8
5*2 análak	8	5*2 anám	8	5*2 anának	8	5*2 anánk	8
5*2 anátok	8	5*2 anék	8	5*2 oni	8	5*2 mníu	8
5*2 aniuk	8	5*2 unod	8	5*2 unom	8	5*2 onotok	8
5*2 tanunk	8	5*2 asz	8	5*2 uttalok	8	5*2 d	0
5*2 ed	8	5*2 ek	8	5*2 el	8	5*2 em	8
5*2 ene	8	5*2 ened	8	5*2 enek	8	5*2 enem	8
5*2 enetek	8	5*2 ené	8	5*2 enéd	8	5*2 enék	8
5*2 enél	8	5*2 enélek	8	5*2 eném	8	5*2 enének	8
5*2 enénk	8	5*2 enétek	8	5*2 eni	8	5*2 enie	8
5*2 eniük	8	5*2 enünk	8	5*2 esz	8	5*2 etek	8
5*2 ett	8	5*2 ette	8	5*2 etted	8	5*2 ettek	8
5*2 ettelek	8	5*2 ettem	8	5*2 ettetek	8	5*2 ették	8
5*2 ettél	8	5*2 ettétek	8	5*2 ettük	8	5*2 ettünk	8
5*2 i	8	5*2 ik	8	5*2 ítek	8	5*2 ned	0
5*2 nem	0	5*2 netek	0	5*2 od	8	5*2 ok	8
5*2 01	8	5*2 om	8	5*2 otok	8	5*2 ott	8
5*2 ottcu	8	5*2 ottad	8	5*2 ottok	8	5*2 ottulak	8
5*2 ottam	8	5*2 ottatok	8	5*2 Ottók	8	5*2 ottól	8
5*2 ottútok	8	5*2 ottuk	8	5*2 ottunk	8	5*2 öd	9
5*2 ök	8	5*2 öl	8	5*2 öm	8	5*2 ötök	8
5*2 ött	8	5*2 ötte	8	5*2 ötted	8	5*2 öttek	8
5*2 öttelek	8	5*2 öttem	8	5*2 öttetek	8	5*2 ötték	8
5*2 öttél	8	5*2 öttétek	8	5*2 öttük	8	5*2 öttünk	8
5*2 t	0	5*2 te	0	5*2 tad	0	5*2 tok	0
5*2 tolok	0	5*2 tum	0	5*2 tatok	0	5*2 tók	0
5*2 tél	0	5*2 tótok	0	5*2 te	0	5*2 ted	0
5*2 tek	0	5*2 telek	0	5*2 tem	0	5*2 tetek	0
5*2 ték	0	5*2 tél	0	5*2 tétek	0	5*2 tuk	0
5*2 tunk	0	5*2 tük	0	5*2 tünk	0	5*2 unk	8
5*2 ünk	8	5*6 j	0	5*6 jad	0	5*6 jak	0

5*6 jalak	0	5*6 jam	0	5*6 janak	0	5*6 jatok	0
5*6 jál	0	5*6 je	0	5*6 jed	0	5*6 jek	0
5*6 Jelek	0	5*6 jem	0	5*6 jen	0	5*6 jenek	0
5*6 jetek	0	5*6 jék	0	5*6 jél	0	5*6 jétek	0
5*6 jon	0	5*6 junk	0	5*6 jünk	0	5*7 tt	0
5*7*9 vi	5	5*7*9 vik	5	5*7*9 vitek	5	5*7*8 vod	5
5*7*8 vok	5	5*7*8 vom	5	5*7*9 vöd	5	5*7*9 vök	5
5*7*9 vöm	5	5*7*8 vunk	5	5*7*9 vünk	5	5*7*8 tta	0
5*7*8 ttclld	0	5*7*8 ttok	0	5*7*8 ttolak	0	5*7*8 ttam	0
5*7*8 ttutok	0	5*7*8 ttók	0	5*7*8 ttál	0	5*7*8 ttótok	0
5*7*8 ttuk	0	5*7*8 ttunk	0	5*7*9 tte	0	5*7*9 tted	0
5*7*9 ttek	0	5*7*9 ttelek	0	5*7*9 ttem	0	5*7*9 ttetek	0
5*7*9 tték	0	5*7*9 ttél	0	5*7*9 ttétek	0	5*7*9 ttük	0
5*7*9 ttünk	0						

*többes

0	0	1 ja	2 4	1 jaí	2 4	1 jaíd	2 4
1 jaik	2 4	1 juim	2 4	1 jaínk	2 4	1 jaítok	2 4
1 je	2 4	1 jei	2 4	1 jeid	2 4	1 jeik	2 4
1 jeím	2 4	1 jeink	2 4	1 jeitek	2 4	1 juk	2 4
1 jük	2 4	2 e	8 7	2 ad	8 7	2 ei	8 7
2 aid	8 7	2 aik	8 7	2 aim	8 7	2 uink	8 7
2 oitok	8 7	2 ok	8 7	2 aim	8 7	2 atok	8 7
2 e	8 7	2 ed	8 7	2 ei	8 7	2 eid	8 7
2 eik	8 7	2 eim	8 7	2 eínk	8 7	2 eitek	8 7
2 ek	8 7	2 em	8 7	2 etek	8 7	2 od	0
2 ok	0	2 om	0	2 otok	0	2 öd	8 7
2 ök	8 7	2 öm	8 7	2 ötök	8 7	2 uk	8 7
2 unk	8 7	2 ük	8 7	2 ünk	8 7	4 vu	6
4 vad	6	4 vai	6	4 vaid	6	4 vaík	6
4 vaim	6	4 vaink	6	4 vaitok	6	4 vak	6
4 vam	6	4 vutok	6	4 ve	6	4 ved	6
4 vei	6	4 veid	6	4 veik	6	4 veim	6
4 veink	6	4 veitek	6	4 vek	6	4 vem	6
4 vetek	6	4 vuk	6	4 vunk	6	4 vük	6
4 vünk	6	7*10 d	2	7*10 i	2	7*10 id	2
7*10 ik	2	7*10 im	2	7*10 ink	2	7*10 k	2
7*10 m	2	7*10 nk	2	7 tek	2	7 tok	2
7 tök	2						

Kísérleti szótár

10

Határozószó

aktuálisan	itt	mindjárt
alakilag	jelenleg	míg
alaposan	jelentősen	most
alapvetően	jobban	nagyrészt
automatikusan	jól	nehezen
azonnal	jólláthatóan	nemcsak
azonosan	keletebbre	nevezetesen
aztán	később	néha
általában	kinn	nyelvtanilag
általánosan	kint	nyilván
belül	kizárólag	nyilvánvalóan
benn	kizárólagosan	nyugatabbra
bizonyára	kívül	nyugodtan
egészen	komolyan	olcsóbban
egyaránt	korábban	ott
egyedül	korán	pillanatnyilag
egyelőre	könnyedén	pontosan
egyenként	könnyen	potenciálisan
egyértelműen	közben	pozícionálisan
egyszer	közvetlenül	rendkívül
egyszerre	különbözőképpen	rendszerint
előbb	különféleképpen	rég
elől	különösen	régóta
előre	lassan	részletesebben
elsősorban	látszólag	rögtön
elvileg	legfőképpen	sehol
eredetileg	legközelebb	soha
felfelé	legmesszebb	sohasem
feltehetőleg	legrészletesebben	soká
felül	legtöbbször	szívesen
fenn	legvégén	szükségképpen
fokozatosan	lehetőleg	szükségszerűen
függetlenül	lényegesen	tartalmilag
grammatikailag	ma	tavaly
gyakorlatilag	magyarul	tárgyasan
gyakran	majd	tárgyatlanul
hamar	maximálisan	távlatilag
hasonlóan	már	tegnap
határozottan	máris	teljesen
hatékonyan	máshol	természetesen
hátsul	másképp	tényleg
heurisztikusan	máskor	többé
hiába	megint	többféleképpen
hirtelen	meglehetősen	többnyire
ide	messze	tradicionálisan
idén	még	tulajdonképpen
időnként	mégis	túl
ismét	mindig	úgy

úgyszólván	valahol	végre
újra	valószínűleg	végül
valahogy	végleg	világosan

153 db Határozószó

11 Határozói igenév

belevéve	tévedvén	véve
----------	----------	------

3 db Határozói igenév

12 Határozói névmás

abba	bárhogy	hol
abban	belé	honnán
addig	ebbe	hova
ahhoz	ebben	hozzá
ahogy	ebből	innen
ahol	eddig	így
akkor	ehhez	közé
amerre	ekkor	mellé
amint	elé	mielőtt
amióta	emiatt	mindenképpen
amíg	ennek	miután
amúgy	ennél	mögé
anélkül	erre	neki
annak	erről	oda
annál	ezáltal	onnan
arra	ezenkívül	rá
arról	ezután	semmiképpen
attól	felé	számára
avval	fölé	ugyanahhoz
azóta	hogy	ugyanakkor
azután	hogyan	ugyanannak

63 db Határozói névmás

20 Igekötő

át	fenn	meg
be	hátra	oda
bele	hozzá	össze
együtt	keresztül	rá
el	külön	tovább
ellen	le	túl
elő	létre	vissza
fel		

22 db Igekötő

38 Határozott névelő

a az

2 db Határozott névelő

39 Határozatlan névelő

egy

1 db Határozatlan névelő

40 Kötőszó

akár	illetve	noha
avagy	is	pedig
azaz	különben	például
azonban	legalábbis	pl.
ám	mert	sem
bór	mégsem	sőt
de	mind	tehát
és	mint	továbbá
ha	mintegy	vagy
hanem	mintha	vagyis
hiszen	mivel	viszont
hogyan	míg	

55 db Kötőszó

51 Igei módosító

hadd hagy ne volna

4 db Igei módosító

52 Melléknévi - igei módosító

abszolút	igazan	nem
alig	igen	nemigen
annyira	kevésbé	sem
egyre	kissé	szinte
elég	közismerten	talán
eléggé	majdnem	vajmi
éppen	mind	viszonylag
főleg	nagyon	

23 db Melléknévi - igei módosító

55 Számnévi módosító

kb	legfeljebb	mindössze
legalább	mind	mintegy

6 db Számnévi módosító

54 Névutószerű szó

ellenére	keresztül	következtében
közben	közel	múlva
nyomán	óta	során
számára	szemben	túl

12 db Névutószerű szó

60 Igeszerű szó

nincs	sincs
-------	-------

2 db Igeszerű szó

61 Nem igei segédige

érdemes	muszáj	tilos
képes	szabad	
köteles	tanácsos	

7 db Nem igei segédige

62 Ragozhatatlan vagy ragozott főnévi névmás

engem	ő	te
engemet	ők	téged
én	őt	tégedet
mi	sokan	ti
ön		

13 db Ragozhatatlan vagy ragozott főnévi névmás

70 Egyéb szó

bizony	igen	sajnos
csak	inkább	stb
csupán	leginkább	szintén
egyáltalán	mindazonáltal	ugyan
egyébként	mindegy	ugye
esetleg	netán	vajon
hát	persze	

20 db Egyéb szó

abbahagy	egyeztet	felmerül
ad	egyszerűsít	felold
ajánl	elalud	felsorol
akad	eldönt	felszerel
akar	elemez	feltár
alakít	elemz	feltételez
alakul	eleged	feltölt
alapít	elér	felül
alapsz	elhelyez	felvetőd
alapul	elindul	fennáll
alkalmaz	elkezd	fél
alkot	elképzél	fér
alsz	elkészít	figyel
analizál	elkészül	fog
azonosít	elkülönül	fogalmaz
ábrázol	ellát	foglal
áll	ellenőriz	foglalkoz
állapít	elmúl	foly
állít	előállít	folytat
árt	előáll	fordít
átalakít	előfordul	fordul
átalakul	előkészít	függ
átfog	elront	fűz
bead	elszakad	gátol
beépít	eltekint	generál
beérkez	eltér	gondol
beír	elutaz	gondoskod
beleépít	elvár	gyűjt
beleért	elveszít	hogy
belevesz	elveszt	hajt
beprogramoz	elvégez	halad
beszél	emleget	hall
bevezet	említ	hallgat
biztosít	esz	használ
bont	él	használtat
bonyolít	épít	hat
buk	épül	határol
céloz	ér	határoz
csap	érez	hazaugr
cserél	érkez	hazautaz
csinál	ért	hátravet
csoportosít	értelmez	hiányz
csökken	érvényesül	hisz
csökkent	fejez	hív
definiál	feksz	hoz
derül	felbont	hozzácsatol
dolgoz	feldolgoz	hozzárendel
dönt	felépít	húz
döntet	felépül	igazít
duzzad	felhasznál	iktat
egyesít	felismer	illet

implementál	kitér	megkap
indul	kitisztít	megkeres
irányul	kiutaz	megkérdőjelez
ismer	kiválaszt	megkülönböztet
ismertet	kizár	megnevez
ír	kíván	megnyom
javít	koncentrál	megold
jár	korlátoz	megoszl
jelen	kódol	megőriz
jelent	köt	megőrz
jelez	kötöd	megpróbál
jelöl	következ	megszok
jelz	küld	megszűn
jön	különböz	megtalál
jut	különböztet	megtanul
kap	lassít	megten
kapcsol	lót	megtesz
kapcsolód	látsz	megvalósít
keres	leáll	megvan
kerül	lefordít	megvizsgál
kezel	legy	megy
képez	legyen	men
képz	leír	mond
képzél	leledz	módosít
kér	lelombozód	mutat
kérdez	len	mutatkoz
kés	lesz	működ
késleltet	leválaszt	működtet
készít	lép	nehezít
kéztet	létez	nevez
készül	magábarejt	nő
kiad	magyaráz	nyilvánul
kialakít	marad	nyújt
kialakul	megad	odaér
kiderül	megállapít	okoz
kidolgoz	megcáfol	old
kiegészít	megcéloz	olvas
kielemez	megcsappan	oszt
kielemz	megegyez	összead
kiérkez	megelégsz	összeállít
kigyűjt	megelőz	összefoly
kihagy	megemlít	összegez
kihelyez	megérkez	összegyűjt
kiindul	megért	összegz
kiismer	megfelel	összekapcsol
kiír	meghalad	összeköt
kijavít	meghatároz	összevon
kiküld	megindít	produkál
kiküszöböl	megismerked	programoz
kiokoskod	megismétel	próbál
kirajzolód	megismétl	ragaszt
kisajátít	megjelen	ragoz
kisérel	megjelöl	rajzol
kisérl	megjósol	rak

rakód	tartalmaz	valósít
rájön	tartoz	van
rájöv	támaszkod	vál
reagál	támaszt	válaszol
redukál	támogat	választ
rendel	tán	változ
rendelkez	tárgyal	vár
reprezentál	tárol	ven
rögzít	te	vesz
segít	tekint	veszít
sorol	ten	vet
szab	teremt	vetőd
szakad	tervez	vezet
számít	tesz	vezérel
számol	tevőd	vezérl
szán	tételez	végez
származ	toldalékol	végignéz
szemantizál	torzít	végigolvas
szemléltet	továbbfejleszt	végigpróbál
szenved	továbbvin	végrehajt
szerepel	továbbvisz	végz
szeret	tökéletesít	vél
szerez	töreked	világít
szervez	töreksz	vin
szerz	töröd	visszafordít
szolgál	történ	visszajön
szorít	tud	visszakérdez
szól	tükröz	visszaküld
szület	tűn	visszaszól
talál	tűz	vizsgál
találkoz	un	volt
tanul	utasít	vonatkoz
tanulmányoz	utaz	zavar
tart	vagy	

416 db Ige

110

Segédige

akar	kíván	szok
fog	próbál	talál
kezd	szeret	tud

9 db Segédige

120

Féligői segédige

ill	lehet
kell	sikerül

4 db Féligői segédige

201

Név

Apreszján	Eötvös	Kovács
április	Erna	Kurcsatov
Bach	Ernő	Lóránd
Béla	Eszenszki	május
Budapest	február	március
Chomsky	György	Mátyás
COSY	János	Moszkva
Dezső	József	Naszódi
DOS	Kantor	Páll
Dubna	KFKI	Péter
	KGST	Pista

32 db Név

202

Mértékegység

fő Mbyte méter tucat

4 db Mértékegység

210

Főnév

adat	attribútum	birtokjel
adatbázis	azonosítás	birtoklás
adatkapcsolat	ábra	birtokosjel
adatkészlet	ábrázolás	birtokviszony
akció	ágens	bizonyítás
alak	ágy	bizottság
alany	állapot	bővítmény
alanyeset	álláspont	Bulgária
alap	állítmány	cél
alapforma	általánosság	célgép
alapszó	átalakítás	célkitűzés
aláírás	átalakulás	címszó
alfejezet	átállás	csap
algoritmus	átmenet	Csehszlovákia
alkalmazhatóság	áttekintés	cselekvés
alkalm	átütemezés	csomó
alkalom	átvétel	csoport
alkotórész	baj	csoportvezető
alrendszer	beépítés	csökkenés
alternatíva	beiktatás	csőd
analizálás	bejelentkezés	csütörtök
analízis	betű	darab
anyagcég	bevezetés	definíció
anyag	bevitel	dialógus
anyagnév	bél	diszk
argumentum	bika	dolg
aspektus	birtok	dolog

egyesszám	fejléc	gyökér
egyetemista	feladat	halmaz
egyezmény	felbontás	hang
egyeztetés	feldolgozás	hangsúlyozás
egység	felépítés	használat
egységesítés	felfogás	határozó
elem	felismerés	hatás
elemzés	feloldás	hatékonyság
elírás	felsorolás	hálózat
eljárás	felsőfok	hátrány
elképzelés	felszólítómód	háttér
elkészítés	feltevés	háttértároló
ellenőrzés	feltételesmód	hely
elmélet	féleség	helyesség
előállítás	figyelem	helyhatározó
előírás	figyelembevétel	helyzet
előny	figyelm	hét
előtag	fizika	hiány
eltérés	fizikus	hiánycikk
elv	fogalm	hiba
elválasztás	fogalom	hibajelzés
ember	fogás	hipotézis
eredmény	fok	hivatkozás
eredményül	fokozás	hívás
esemény	folyam	homloktér
eset	fordítás	homloktér
esetenként	forma	homonímia
esetgrammatika	formula	hozzárendelés
esetrag	forrás	hó
esetragozás	fókusz	idő
esz	földarabolás	időhatározó
eszköz	főnév	ige
eszközül	főnév	igealak
élőlény	főnévrag	igeidő
ér	frázis	igekötő
érdek	függőség	igenev
érdekeltség	függvény	igenév
érdekesség	generáció	igerag
érdeklődés	generálás	igeragozás
érdem	generátor	igénként
értelem	gép	igény
értelm	gépírás	illeszkedés
értelmezés	gomb	indikáció
értelmezhetőség	gond	indoklás
érték	grammatika	infinitivusz
érthetőség	gráf	információ
ész	gyakoriság	információközlés
ételrecept	gyakorlat	input
év	gyengeség	intelligencia
évtized	gyerek	interfész
fajta	gyermek	interfészül
fázis	gyermekbetegség	interrupt
fejezet	gyorsaság	intézet
fejletlenség	gyöker	intézmény

irány	kifejezés	ló
irányvonal	kikeresés	magánhangzó
irányzat	kiküldetés	Magyarország
irodalm	kinyerés	manipuláció
irodalom	kiutazás	maradék
iskola	kiút	matematikus
ismeret	kiválasztás	mássalhangzó
írásjel	kivétel	megadás
javítás	Kína	megállapítás
járás	kíséret	megértés
jegy	kísérlet	meghatározás
jel	kombináció	megindoklás
jelenidő	kommunikáció	megjegyzés
jelentés	korlát	megkeresés
jelentőség	korlátozás	megközelítés
jelleg	kód	meglét
jelölés	kölcsönhatás	megnevezés
jelzés	költemény	megnyomás
jelző	könyv	megoldás
jogász	körülmény	megtalálás
jogosultság	kötel	megvalósítás
jogtudomány	kötél	megye
jövő	kötőhang	mellékmondat
jövőidő	kötőjel	melléknév
kabát	követelmény	melléknév
kalandozás	közép	memória
kalap	közkinccs	menet
kapcsán	központ	mennyiség
kapcsolat	köz	mennyiségjelző
kapcsolódás	Kuba	mentesség
karakter	kutatás	menü
kategória	különbség	mező
katona	különlegesség	méret
kedv	kvantor	mérőrendszer
kelet	labor	mérték
kereset	laboratórium	mértékegység
keresés	látóhatár	milyenség
keret	láz	minta
kezdet	lefordítás	mondanivaló
kezelés	legelő	mondat
kép	lehetőség	mondatrész
képesség	leírás	mondattan
képlet	lekérdezés	Mongólia
kérdés	lekszéma	morféma
kés	Lengyelország	morfológia
készítés	leválasztás	mód
készlet	level	módszer
kétértelműség	levél	múlt
kéz	lé	múltidő
kiadás	lényeg	munka
kialakítás	lépés	munkahipotézis
kicserélés	létrehozás	munkatárs
kidolgozás	lista	munkás
kiegészítés	logika	mutatónévmás

működés	projekt	számológép
működőképesség	prompt	szándék
műnyelv	rag	szelekció
nagyságrend	ragozás	szellem
nehézség	reagálás	szemantika
nev	remény	személy
név	rendelkezés	személynév
névelő	rendszer	személyrag
névmás	rendszerterv	szeminárium
névszó	rész	szempont
névutó	részeredmény	szerep
nézet	részhalmoz	szerkezet
növelés	részlet	szerviz
nő	részletesség	szerződés
nyelv	résznyelv	szétvágás
nyelvtan	részvényes	szimbólum
nyelvtudomány	réteg	szinonima
nyugat	robbanás	szint
objektum	Románia	szintagma
ok	rovat	szintaxis
oldal	rugalmasság	szituáció
olvasás	ruha	szokás
ország	sablon	szombat
osztály	sajátosság	Szovjetunió
önállóság	sebesség	szó
összeállítás	segédige	szócska
összeg	segédszó	szócsoport
összekötés	segítség	szófaj
ősz	siker	szókészlet
papír	sikertelenség	szóosztály
paraméter	sokaság	szórend
parancs	szó	szótár
park	sorozat	szótő
párbeszéd	sorrend	szöveg
példa	specifikáció	szövegrész
péntek	státusz	szövetkezet
piac	súly	szükség
pillanat	szabály	születés
pont	szakasz	találkozás
pontpár	szakember	találkozó
pozíció	szakértő	tanulmány
preferencia	szakirodalom	tanulmányozás
prepozíció	szakképzettség	tapasztalat
probléma	szarv	tartalm
processzor	szál	tartalmisság
profil	szám	tartalom
program	számítástechnika	tábla
programozás	számítástechnikus	táblanev
programrendszer	számítástudomány	táblanév
programrész	számítógép	táblázat
	számjel	tájékoztató
	számjelző	támogatás
	számnev	tár
	számnév	tárgy

tárgyasság	tudás	vásár
tárgyeset	tudományegyetem	vektor
tárgyrag	tudomás	veszély
tárolás	tulajdonság	vesztés
távolság	tulajdonságjelző	vezérlés
technika	tündérmese	vég
teherátvitel	tűz	végeredmény
tekintet	univerzalitás	végignézés
termék	utalás	végigpróbálgatás
terminológia	utasítás	végpont
terület	utca	végrehajtás
terv	útibeszámoló	végződés
tevékenység	utód	vélemény
téesz	újdonság	vér
téma	újraalkalmazhatóság	Vietnam
témakör	út	visszahatás
tér	üzenet	visszakeresés
tisztázás	valószínűség	vita
típus	vasárnap	vizsgálat
toldalék	válasz	vonal
tötöbbszám	válaszidő	vonás
többszám	választék	vonzeró
többség	vállalat	zaj
tőtár	változat	zavar
tudat	változó	

642 db Főnév

211

Egyszerre főnév és melléknév

alkotó	egyenes	képző
angol	fa	konzerv
automata	felhasználó	lengyel
beszámoló	folyó	magyar
beteg	francia	modell
birtokos	illető	nyugatnémet
bulgár	japán	olasz
default	kalcium	orosz
dolgozó	kálium	programozó
		szövegszerkesztő

28 db Egyszerre főnév és melléknév

212

Főnév jellegű névmás

aki	ki	semmi
ami	másik	senki
az	mi	többi
efféle	mindaz	ugyanaz
ez	mindenki	ugyanaz
ilyesmi	mindez	valaki
		valami

19 db Főnév jellegű névmás

213 Személyragozott névmás

egyik	mag
jómag	önmag

4 db Személyragozott névmás

214 Birtokosan nem ragozódó névmás

egymás	enyém	mind
--------	-------	------

3 db Birtokosan nem ragozódó névmás

220 Melléknév

absztrakt	egyértelmű	függőségi
absztraktabb	egységes	gazdag
agglutinatív	egyszerű	gazdaságos
ajánlatos	egyszerűbb	generációs
akadémiai	ellenkező	geometriai
aktuális	elméleti	gépi
alacsony	előbbi	gépikódú
alacsonyabb	előnyös	gondos
alakú	előtti	grafikus
alapú	előző	grammatikai
alapvető	első	gyakori
alábbi	elvi	gyakorlati
algoritmikus	eredeti	gyenge
alkalmas	erős	gyors
alkalmi	európai	halvány
aprólékos	explicit	hangtani
aspektusú	érdekes	hasonló
atomkutató	érdektelen	hasonszórú
automatikus	érdemi	használatos
automatizálási	értelmes	hasznos
azonos	értelmesebb	haszontalan
állandó	értelmű	határozatlan
általános	érvényes	hatékony
általánosabb	ésszerű	hatékonyabb
átlagos	felesleges	helyes
belső	felszíni	helytelen
bemeneti	feltételes	heurisztikus
bizonyos	fenti	hiányos
biztos	finom	hibás
célratörő	finomabb	hivatott
célszerű	fogalmi	homályos
divatos	fontos	hosszabb
divatosabb	formális	hosszú
dubnai	fő	ideális
eddig	főnévi	idegen
egész	fős	igaz
egyedi	furcsa	igei
egyenértékű	független	igényű

implicit	lehetséges	pozicionális
indirekt	lexikai	praktikus
infinítívuszi	lexikális	profi
interaktív	lényeges	programfüggetlen
ismeretlen	lényegi	programozástechnikai
jelen	logikai	ragos
jelenidejű	magas	ragtalan
jelentésű	magasabb	reális
jelentős	matematikai	redundáns
jellegű	maximális	releváns
jelzői	márciusi	rettenetes
jelzős	más	részleges
jobb	meglehető	részletes
jó	megszámlálhatatlan	rokonszenves
jólstruktúrált	melléknévi	rossz
kapcsolatos	mesterséges	rózsaszín
karakteres	mély	rövid
kategóriájú	méretű	rövidtávú
káros	mindennapi	sajátos
kellemetlen	minimális	sajnálatos
kereseti	moduláris	sikeres
kezdeti	mondattani	sokféle
kék	morfológiai	sorrendi
képzettségű	múltidejű	speciális
kérdéses	nagy	specifikációs
kész	nagyobb	statikus
kétértelmű	nagyobbméretű	stílusú
kétféle	nehéz	strukturális
kicsi	nehéz	súlyos
kifejezési	nemzetközi	szabad
kijelentőmódú	nevű	szaggatott
kintdolgozó	német	szakmai
kinti	névelőtlen	számítástechnikai
kis	névelős	számítógépes
kisebb	névszói	számú
kiváló	normális	százszoros
kívánatos	nyelvészeti	szelektív
konkrét	nyelvi	szemantikai
korábbi	nyelvtani	szemantikus
korrekt	nyelvű	személyes
könnyebb	nyilvánvaló	személyű
könnyű	objektív	szerű
kötetlenebb	opcionális	szigorú
közbülső	osztatlan	szimmetrikus
közismert	önálló	szintaktikai
köznap	párbeszédes	szintaktikus
közös	piros	szintű
közvetlen	pontatlan	szocialista
külföldi	pontatlanabb	szokásos
külső	pontos	szovjet
laza	pontosabb	szórendű
legegyszerűbb	pontszerű	szöveges
legfőbb	portábilis	szükséges
lehetetlen	potenciális	szükségszerű

születési	terminológiai	utóbbi
tajvani	tetszőleges	ügynevezett
tartós	téves	új
tájékozatlan	típusú	újabb
tárgyas	további	valódi
tárgyatlan	többes	valószerű
tárgyragos	többszámú	változatlan
tárgyú	töbértelmű	váratlan
távolitele	tradicionális	végtelen
teljes	tudományos	világos
tematikus	túlsó	vonalszerű
természetes	univerzális	zavaros
természetű	utolsó	

316 db Melléknév

221

Melléknévi igenév

adható	foglalkozó	kiolvasható
alakítható	fordító	kiváló
alkalmazandó	függő	kötelező
alkalmazható	gyorsítható	következő
azonosítható	gyorsító	kutató
állítható	hallgató	különböző
álló	használandó	látható
bedolgozó	használható	látszó
beszélő	határozó	leíró
bevezethető	hiányzó	lekérdező
bővíthető	hozzáilleszthető	levő
cselekvő	hozzátartozó	létező
cserélhető	idevonatkozó	megelőző
eldöntendő	idéző	megfelelő
elemző	illeszthető	megjelenő
elérhető	javítható	megjósolható
elfogadható	járó	megoldható
elképezhető	jellemző	megtalálható
elkészítendő	jelző	megtehető
ellenőrizhető	jósolható	menő
ellenőrzendő	jövő	módosító
elváló	kapcsolható	működtethető
elvégezhető	kapcsolódó	működtető
elvégzendő	kapható	műveltető
élenjáró	kedvező	nyelvelemző
érhető	kezelő	nyugtalanító
értetődő	képezhető	olvasható
fekvő	kérdő	összetartozó
felbontható	készítendő	összetett
felhasználható	kiadandó	összetevő
felszólító	kidolgozandó	résztevő
feltételezhető	kielégítő	rögzíthető
fentemlített félő	kiérkezendő	segítő
figyelhető	kigyűjtő	sorolható
fogható	kiinduló	szenvedő
foglalható	kijelentő	szereplő

szolgáló	tekinthető	valósítható
szóló	továbbosztható	választható
található	történő	velejáró
találkozó	tudható	vezérlő
tartozó	túlmenő	végrehajtó
tájékoztató	utaló	visszautaló
tároló	való	vonatkozó

130 db Melléknévi igenév

222 Melléknévi névmás

akkora	egyéb	olyan
akkori	ily	saját
amolyan	ilyen	semmilyen
bármilyen	milyen	ugyanolyan
egyes	mindenféle	valamilyen

15 db melléknévi névmás

223 Melléknévi névutó

beli	közti	szerinti
feletti	nélküli	

5 db Melléknévi névutó

230 Szám

egy	félig	két
egyetlen	harmadik	második
egyszer	hat	nulla
először	három	öt
első	hét	ötödik
fél	kettő	

17 db Szám

231 Melléknév jellegű számnév

csekély	sok	több
elegendő	számos	többször
kevés		

7 db Melléknév jellegű számnév

232

Számnévi névmás

ahány	ennyi	semennyi
akárhány	mennyi	számtalan
akármennyi	mindkét	valahány
amennyi	néhány	valahányszor
annyi	némi	valamennyi
bármennyi	pár	
csomó	sehány	

19 db Számnévi névmás

223

Kvantor

akármelyik	amely	amelyik
bármely	bármelyik	egyik
mely	melyik	mindegyik
minden	összes	semelyik
valamelyik		

13 db Kvantor

240

Névutó

alatt	felé	mellől
alá	felől	miatt
alól	felül	mögé
által	fölé	mögött
belé	fölött	mögül
belől	helyett	nál
belül	hozzá	nek
benn	kívül	nélkül
elé	közé	rajt
ellen	között	ra
elől	közt	ról
előtt	közül	szerint
ért	mellett	től
felett	mellé	után
		vel

43 db Névutó