MTA SZTAKI A Magyar Tudományos Akadémia
Számítástechnikai és Automatizálási Kutatóintézete

Magyar nyelvű mondatok elemzése természetes nyelvű interface céljából

Tanulmány

(Harmadik utánnyomás)

Farkas Ernő *Naszódi Mátyás*

1990. május, MTA SzTAKI

Ez a tanulmány az OMFB támogatásával, a „Távlati (5. generációs) számítástechnikai rendszerek fejlesztéséhez" című kutatás keretében jött létre.

ISBN 963 311 307 5

Copyright © 1990
by MTA SzTAKI

TARTALOMJEGYZÉK

Bevezetés 7
Munkamódszerünkről 7
A nyelv modelljéről 9
Az elkészült program 11
Az implementáció eszköze 11
A program részei 11
Végződések 15
Ragok 16
Esetragok 16
Igeragok 17
Jelek 18
Képzők 19
Kivételes esetek 20
Magyar szavak morfológiai elemzése 21
I. A magyar nyelv morfológiája 21
Szavak, morfémák és alakjuk 21
Szóelemzés 22
Morfológiai szeletelő 22
Elvi modell 22
II. A kivitelezés gyakorlati kérdései 24
A jelenleg megvalósított modell 24
1. Szótár – szóosztályok 25
2. Toldaléktár - toldalékosztályok 25
3. Morfoszintakszis - a toldalékok szekvenciái .26
4. Tőváltozások 29
5. Fonológiai illeszkedési kódok 29
A modell implementálása 30
A gyakorlati tapasztalat 31
III. Egy tervezett modell 33
Egy lehetséges kivitelezés 35
Az automata implementációs modellje 38
Kivételkezelés 38
Az új módszer előnyei 39
Függelék 40
I. A végződések kódolása 40
II. A toldalékok fonológiai osztályai 44
III. A szavak morfoszintaktikai felosztása 45
IV. Példa egy morfológiai elemzésre 48
A szófajok 49
Igék 50
A létigék 50
A segédigék I. 51
Segédige II. 52
Igenevek 52

Névszók 53
Főnevek 53
Nevek 54
Mesterség-beosztás, nemzetiség, pártállás, vallás .55
Anyagnevek 55
Főnév jellegű névmások 56
Melléknév 57
Melléknévi igenév 58
Pozícionáló melléknév 58
Melléknévi névutó 58
Melléknévi névmások 59
Mértékegységek 59
Konkrét számnév 60
Sorszámnév 60
Határozatlan számnév 60
Számnévi névmások 60
Kvantorok 61
Szelektorok 61
Névutók 61
Ragozhatatlan szavak 62
Határozószók 62
Igekötők 62
Névelők 63
Kötőszók 63
Módosítószavak 64
Nem igei segédigék 64
Ragozhatatlan személyes névmások 64
Névutószerű szavak 65
Egyéb szavak 65
A névszói kifejezések 67
A névszói szerkezetek informális leírása 67
Az egyszerű névszói kifejezés 69
A kiemelés 71
A névutós szerkezetek 73
A névszói szerkezetek formális leírása 74
A kötőszavak 79
A névszói kifejezés elemzésének stratégiája 80
Az elemzés taktikája 81
A névszói kifejezés elemzése 82
Problémák és hiányosságok 83
Az egyszerű mondatok szerkezete 85
Az állítmány 85
Ragozott ige 85
Segédige ragozatlan főnévi igenévvel 86
Segédige ragozott főnévi igenévvel 87
Melléknév + főnévi igenév, mint állítmány 88

Egyebek 89
Jelen idejű (folyamatos) melléknévi igenév mint állítmány 90
"Szenvedő" mondatok 91
-ható/-hető és *-andó/-endő* állítmányú mondatok . . .92
Az igei állítmányhoz kapcsolódó mondatrészek 93
Az alany, a tárgy és a részeshatározó 93
Ragos és névutós egyszeres vonzatok 95
Hely- és időhatározós vonzatok 96
Az igekötők és kulcsszavas vonzatok 96
Alárendelt mondatok és a főnévi igenévi vonzat .97
Névszói mondatok 98
Egyszerű egzisztencia mondatok 100
A szabad határozók. 100
Az argumentumok megtalálása 103
Másodlagos vonzatok 104
Az összetett mondatok 107
Egy mondat - több mondat 107
Az összetett mondat felbontása tagmondatokra . . . 110
A tagmondatok viszonya 112
Az almondatok 113
A mellérendelés 114
A vonzatokról 117
Igei vonzatok 118
Névszói vonzatok 120
A főnevek vonzatai 120
A melléknevek vonzatai 120
Névutók vonzatai 121
A vonzatok súlya és kiválasztása 122
A névutókról 125
Ragokhoz tartozó névutók 128
Összefoglalás 129
Igenév 131
A főnévi igenév 133
Melléknévi igenév 135
Határozói igenév 137
Az igei főnév alakja 138
Implementációs megjegyzések 140
A határozott és a határozatlan névszói kifejezések . 141
Az elnyelődés 141
Összefoglalás 147
Igei és névszói relációk 149
Az igei és névszói relációk megjelenési formái a mondatban 152
A legfelső igei reláció 152
Alsóbb szintű igei reláció 154

A legfelső szintű névszói reláció 159
Alsóbb szintű névszói relációk 162
A kommunikáció egyszintű modellje 165
Sémák és konkrétumok 170
Ellentmondó információk 171
Az eltárolt információ teljessége 172
A kommunikáció kétszintű modellje 173
Bekezdések 173
A tennivaló megértése 174
A hivatkozások feloldása 174
Összefoglalás 175
INDEX 177

BEVEZETÉS

1988-ban két egymást követő változatban kidolgoztunk egy olyan programot, amely magyar nyelvű szövegek morfológiai és szintaktikai elemzését tudja elvégezni. Ez a program alapja lehet olyan számítógépes munkáknak, amelyek magyar nyelvvel kapcsolatosak: így elsősorban természetes nyelvű interfésznek, esetleg gépi fordítónak, gépelési hibákat kiszűrő programnak, olyan szöveges visszakereső programnak, amely a szavakat ragozott alakban is megtalálja stb.

Ezzel az írással az a célunk, hogy összefoglaljuk a magyar nyelvről szerzett ismereteinket, és leírjuk azt az utat, ahogy egy program eljuthat a magyar mondattól, mint karaktersorozattól, annak tartalmának absztrakt adatstruktúrában történő reprezentálásáig.

Ezen írás készítése közben újra végiggondoltuk mindazt, amit a programírás során tettünk és rendszereztük a leszűrt tapasztalatokat, ezért az itt leírtak nem az elkészült programot tükrözik, hanem egy annál remélhetőleg pontosabb modellt.

Munkamódszerünkről

Mivel munkánk megkezdésekor a magyar nyelvnek nem volt sem formális modellje, sem olyan félformális leírása, ami a nyelv igen nagy részét lefedte volna, munkánkat úgy végeztük, hogy kezdetben hipotézisek alapján egy programmagot készítettünk, és ezt ráengedtük a rendelkezésünkre álló szövegekre, illetve abból kigyűjtött egyszerű mondatokra, és vizsgáltuk az eredményt: mit nem tud elemezni a program, mit elemez hibásan a program, miért elemzi hibásan a program. Ezeknek a dolgoknak, ha lehetett, utánanéztünk a nyelvészeti irodalomban is, és ennek alapján fejlesztettük tovább a programot.

A felhasznált szövegek a következők voltak: kéziratok elsősorban a programozásról, programokról, mesterséges intelligenciáról, magyar nyelvről és magyar nyelvtanról; MTI-től származó hírek, gyermekelhelyezési perek kivonatai, hivatalos és magánlevek.

Egy későbbi időpontban végigelemeztünk jó néhányat Rácz Endre és Szemere Gyula "Mondattani elemzések" című könyvének példamondatai közül. A mondatok egy kis részét különböző okok miatt nem tudjuk elemezni. Egy meglehetősen nagy részénél viszont az általunk adott elemzés nem azonos a könyvbeli elemzéssel. Hogy miért nem, az a későbbiekből kiderül.

A természetes nyelv elemzése nem könnyű feladat, hiszen nagyon nehéz eldönteni, hogy melyik mondat helyes mondat, melyik mondat elfogadható, és melyik mondat hibás. Hasonlóan, vitatható időnként, hogy melyik mondat mit is jelent. Az, hogy ki milyennek látja a nyelvet, egyénenként változik. Lehet, hogy az általunk vizsgált anyag, ami főleg a saját szövegeink elemzésen alapul, a mi modellünket is torzította.

Mi úgy gondoljuk, hogy „a nyelv" az, amit az emberek használnak, és egy természetes nyelvű interfésznek az a dolga, hogy mindent megértsen, ami megérthető. Nem célszerű, ha a számítógép oktatja ki az embert arról, hogy hogyan kell valamit helyesen mondani. Ezért tehát mi egy olyan tág modellt kívántunk létrehozni, ami minél teljesebben lefedi a helyes és elfogadható mondatokat, és viszonylag kevesebb gondot fordítottunk a helytelen mondatok kizárására. Feltételezzük ugyanis, hogy aki a programot használja, tud magyarul, és az a szándéka, hogy eredményt érjen el, és nem ad szándékosan hibás vagy értelmetlen mondatokat a gépnek.

Az általunk adott leírás többnyire olyan, hogy először elég részletesen leírjuk, hogy mi hogyan látjuk az illető nyelvi jelenséget, majd annak leírása következik, hogy a programnak mit kell tennie. Számos helyen megfigyelhető lesz, hogy a program a jelenség bizonyos tulajdonságaival nem törődik, mert az számára nem fontos. Például: nem törődünk a magas és mély hangrend egyeztetésével a szó és végződés között. Azért nem törődünk ezzel, mert ebből rendkívül ritkán kaphatunk a mondat értelmezésével kapcsolatban valamilyen használható információt. Vagy például: az igekötős igéknél nincs feljegyezve, hogy igekötős és mi az igekötője, viszont minden igéhez oda van írva, hogy milyen igekötőket kaphat. Természetesen az igekötős igéhez az van feljegyezve, hogy nem kaphat semmilyen igekötőt. Ez azért van így, mert a modellünkben nem arra volt szükség, hogy egy igekötős igét felbontsunk, hanem az elvált részeket kellett összekapcsolni.

A program azért készült el két változatban, és azért került sor menetközben számos módosításra, mert időnként a korábbi megoldásokat ki kellett egészíteni, és amikor a kiegészítések már nem segítették a megoldást, felül kellett bírálni. Így végül egy olyan magyar nyelvmodell alakult ki, ami igen távol áll attól a képtől, amit az iskolában 30 éve tanultunk, és ahhoz is csak részben hasonlít, amit a nyelvészektől mostanában tanultunk.

Sajnos a számítógép igen kegyetlen, nemigen teszi lehetővé a homályos, kétértelmű, „filozofikus” fogalmak kezelését. Nem is tudunk mit kezdeni olyan fogalmakkal, mint „mód-” vagy „állapothatározó”. Mi csak olyat tudunk mondani, hogy van olyan vonzat vagy szabad határozó, ami a következő végződésekkel és névutókkal járhat: ...

Ez nem jelenti, hogy a hagyományos nyelvtanok minden fogalmát elvetettük. Az alany, tárgy, főnév, melléknév stb. szavakat mi is használjuk, és az olvasó nemigen téved akkor, amikor ezeket a fogalmak korábban tanult jelentésével azonosítja. Viszont nekünk vannak olyan formális kritériumaink is (milyen helyeken állhat a mondatban, milyen végződéseket kaphat, stb.), amelyek alapján eldöntjük, hogy valami odatartozik-e vagy nem.

Természetesen lesz egy sor olyan új kategória, amivel az olvasó itt találkozik először. Nem biztos, hogy ezek elnevezése szerencsés volt. Mi mindenesetre arra törekszünk, hogy világosan megmagyarázzuk, hogy mit minek nevezünk, mit milyennek látunk, mit miért csinálunk.

Biztosak vagyunk benne, hogy nyelvleírásunk hiányos, lehet, hogy vaskos tévedéseink is vannak. Egyetlen mentségünk, hogy a program működik, és az esetek 90%-ában azt csinálja, amit szükségesnek tartunk.

A nyelv modelljéről

Nekünk az a véleményünk: lehet, hogy a magyar nyelv a Jóisten adománya, de az azt leíró nyelvtan biztos emberi csinálmány. Ugyanahhoz a nyelvhez számos különböző modell képzelhető el. Néhány helyen mi is jelezni fogjuk, hogy a nyelvi tényeknek a mienktől elérő modellje is lehet.

A nyelv modellje akkor jó, ha megfelel a célnak, amire használni akarjuk. A mi céljaink a következők voltak:

- Egy adott témakör nyelvének minél teljesebb lefedése. Így véljük, hogy a szókészletet limitálni lehet. Egy szűk témakört elég jól le lehet fedni néhányezer szóval. Viszont a lehetséges nyelvi konstrukciókat nemigen lehet megszorítani.
- A modell nem lehet túl nagy, túl részletes. Mind az emberi kezelhetőség, mind a gépi adottságok korlátokat szabnak nekünk. Nincsenek sem szupernagy, sem szupergyors gépeink. És a programot se pusztán tudományos kutatási célzattal, hanem elsősorban gyakorlati célra, mindennapi gépekre szánjuk. (Ennek a célkitűzésnek az elérhetősége korábban előttünk is kérdéses volt, de a tapasztalatok igen pozitívak. A mi modellünk egy mai személyi számítógépben is kényelmesen elfér, és a gyorsasága is meggyőző, egy-két másodperc alatt képes kielemezni egy mondatot.)

AZ ELKÉSZÜLT PROGRAM

Az implementáció eszköze

Sokak szerint a program azért ilyen gyors, mert nem a szokásos eszközökkel LISP-ben vagy PROLOG-ban íródott. Ez lehetséges. Az első változat MODULA-2-ben készült a nyelvre jellemző megoldásokkal, az adatokat tömbökben és rekordokban tárolva. Ez a tárolási mód egy későbbi fázisban merev szerkezete miatt a további fejlesztések gátja lett. Ekkor felmerült, hogy mégis a LISP nyelvet kellene használnunk.

Sajnos, azt tapasztaltuk, hogy Magyarországon igazán professzionális LISP fordító nem szerezhető be.

Így a program második verziója is MODULA-2-ben készült el, de úgy, hogy először egy listakezelő modult készíttettünk, és a program összes adatát listákban ábrázoltuk, méghozzá kétfélében: egy állandó listában azokat az adatokat, amelyek nem változnak a program élete során, és egy időleges listában azokat, amelyek csak egy mondat idejére élnek. Ez a megoldás eddig minden további bővítést lehetővé tett.

Viszont a második megoldásban a listakezelés miatt jó néhány függvény komplikáltabb lett, és hossza lényegesen megnőtt. Ha egy harmadik implementációt is MODULA-2-ben vagy hasonló nyelven csinálnánk, akkor valószínűleg egy hibrid megoldást választanánk, a morfológiai elemzés eredményét tömbökben és rekordokban tárolnánk, de a magasabb struktúrákat lista formában ábrázolnánk.

A program részei

A program két nagy részből áll: a morfológiai analízisből, és az azt követő szintaktikai analízisből. Közbeékelődik még egy kis rész, a posztmorfológiai szűrés. Ez tulajdonképpen egy teljesen önálló feladat, ami megindokolhatóan hozzácsapható akár az előtte, akár az utána jövő fázishoz.

A morfológiai elemző beolvas egy szöveget, amely karakterek (betűk, számok, szóközök, írásjelek stb.) sorozata, és felbontja szavakra és írásjelekre, és a szavakat felbontja szótövekre és végződésekre (és előtagokra). A program végeredménye egy lista, amelynek minden eleme egy allista. Esetünkben az allista első eleme a szó eredeti alakja, ezt csak a nyomon követhetőség érdekében tesszük el. A további elemei viszont olyan listák, amelyek a szó lehetséges felbontásait adják meg. Pl.:

```
((A, (a:határozott névelő)),
   (válás, (vál[ik]:ige, ás:képző)),
   (8, (8:számnév)),
   (éve, (esz[ik]:ige, ve:határozói igenév képző),
         (év:főnév, e:birtokjel),
         (éve:időhatározószó)),
   (történt,(történ[ik]:ige, t:igerag),
         (történ[ik]:ige, t: melléknévi igenév képző)),
         (.:írásjel))
```

Természetesen az eredmény nem ilyen szöveges formában, hanem kódolva van megadva. A szófajokat, ragokat, képzőket, jeleket számok adják meg, csak a szó eredeti alakja és a szótő rekonstruált alakja van karakteresen ábrázolva. Az előbbi mondat így néz ki:

```
A:
 a|3804
válás:
 vál|10005,8004
8:
 8|23004
éve:
 éve|1410
 év|21004,2013
 esz|10005,1530
történt:
 történ|10005,8002
 történ|10005,1103
.:30046
```

A számok persze sokkal részletesebb információt tartalmaznak, mint a fenti szöveg. A *történ*-nél az 10000 jelzi, hogy ige, míg az 5, hogy nem áll meg ragtalanul (ikes ige). Az éve szónál az 1000 jelenti, hogy határozószó; a 400, hogy időhatározó; a 10, hogy kötelező vonzata van; a végén álló 0, hogy más felbontása is lehetséges. A szófajra utaló szám 5 jegyű. A felső négyet későbbi munka során használjuk; az ötödik, ami morfológiai tulajdonságokra utal, csak ellenőrzés kedvéért marad benn.

A következő lépés a posztmorfológiai szűrés. Láttuk, hogy a mondatban sok kétértelmű szó van. Az egytagú szavak (*az*, *mi*, *nő* stb.) igen nagy része kétértelmű, és közülük számos olyan van, ami gyakran fordul elő a szövegekben. Ezen kétértelműségek egy részét (kb. 20 %-át) egyszerű ökölszabályokkal megszüntethetjük. Pl. az *az* szó nem lehet névelő, ha ige előtt áll, az a névutó, amelyik nem vonz ragot, csak névszó után állhat, stb. Így *Az írta alá, aki*... mondatban az *Az* csak névmás és az alá csak igekötő lehet. A posztmorfológiai szűrés során csak egymást követő szavakból álló párokat vizsgálunk, és így szüntetjük meg az első vagy második szó kétértelműségét.

A szintaktikai analízis a morfológiai analízis eredményéből indul ki. Az irodalomban találkozunk olyan elképzelésekkel, ami a magyar nyelv morfémáit az angol nyelv szavaival állítja párhuzamba. Egy egyszerű morfémaszekvenciából indul ki, és függőségi nyelvtan esetén a morfémákat hozza egymással összefüggésbe, például az igeragra akasztja az igét. Ennél a felfogásnál minden harmadik, negyedik mondat, ami többértelműséget tartalmaz, 2, 3, 4, 6, 8, 9, esetleg ennél is több lehetséges morféma sorozatot ad. Ezeket mind elemezni kell, és az elemzések közül kell kiválasztani a legjobbat. Sajnos a választás nem könnyű. Ha valaminek nem találunk helyet a mondatban, akkor az elemzés valószínűleg nem jó, de legalábbis nem teljes. Ha azonban mindent sikerült elhelyezni, az még nem garancia arra, hogy az elemzés jó. Azt tapasztaltuk, hogy a mondatoknak vannak valószínű és roppant valószínűtlen elemzései. Például annak a mondatnak, hogy *Az ajtóból látta Pétert*. elképzelhető egy olyan elemzése, az *Az* az ige alanya, de roppant valószínűtlen.

Nekünk az a stratégiánk, hogy mindig az általunk nagy valószínűségűnek vélt irányba (vagy irányokba) folytatjuk az elemzést, és a valószínűtlen ágakat elhanyagoljuk. Nem vonjuk kétségbe azt, hogy a mondatot a fent vázolt felfogással is lehet elemezni, de mi nem ezt az utat követjük.

Nálunk az elemzés egy olyan sorozatból indul ki, ami az eredeti mondat teljes szavaiból áll. A szavakat (illetve ezek összes lehetséges morfológiai magyarázatát) egy-egy összetett adatnak tekintjük, és az elemzés során ezeket kapcsoljuk össze egy a mondat szerkezetét leíró adatstruktúrává, és közben a szó morfológiai alternatívái közül csak az odaillőt, illetve a legvalószínűbbet tartjuk meg. Nálunk tehát a szó egy olyan komplex egység, ahol a szótőt (a képzett alakot) tekintjük a jelentést hordozó és nyelvtani szerepű magnak, és a ragok és jelek mint attribútumok jelennek meg.

Az elemzett mondat szerkezetét végül a következő adatstruktúra adja meg:

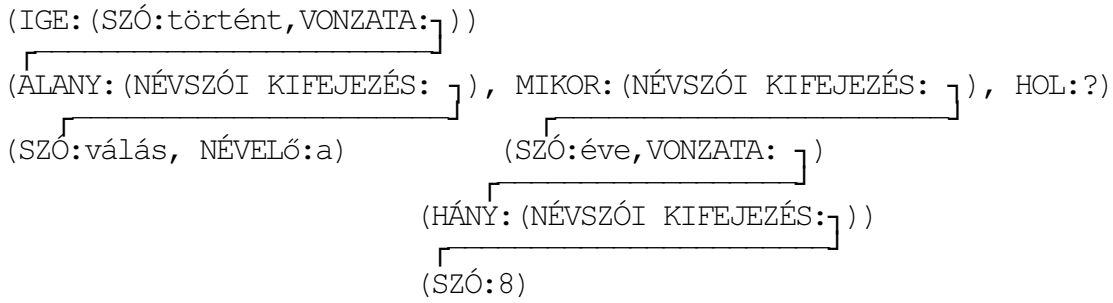
Ha a mondat mellérendelő, akkor a legfelső szint egy kötőszólista, ami mondatok összekapcsolását írja le. Az egyes kötőszavakat egy-egy adatstruktúra követi, ami a megfelelő tagmondat szerkezetét írja le. (Az alárendelt mondatok be vannak kötve a főmondat szerkezetébe.)

A mondat szerkezete listákból áll. A listának mindig van egy tartóeleme, ami a szerkezetben az általunk legfontosabbnak tekintett mondatrész, és ehhez egy szó van kötve. Ezt követik a listában azok a szavak, és nyelvtani szerkezetek, amelyek általános érvényű nyelvtani szabályok alapján kapcsolhatók a tartóelemhez.

A szabály általános, ha csak a szófajokra, ragokra, jelekre hivatkozik, és nem függ a benne szereplő konkrét szavaktól. Ilyenek például azok a szabályok, amelyek különböző igei kifejezéseket írnak le. Pl. *nem kellett volna odamenned* ennek tartalma: (IGE:*odamenned*, SEGÉDIGE:*kellett*, MÓDOSÍTÓ1:*volna*, MÓDOSÍTÓ2:*nem*). A szabály itt azt mondja, hogy van olyan igei kifejezés, amely egy személyragozott igenévből, egy egyes szám harmadik személyű segédigéből áll, és a segédige után egy, előtte pedig egy vagy két módosító szó állhat.

Egy ilyen általános konfigurációs szabályban több szó is szerepel, és nyilván nagyon szubjektív dolog, hogy melyiket tekintjük a legfontosabbnak a tartóelemek közül. Nálunk az igei állítmányú mondatok esetén az ige. Névszói mondatok esetén többnyire az alany. A névszói szerkezetben a szerkezet utolsó szava (többnyire főnév). Hogy ez miért van így, az az igei és névszói relációkról szóló fejezetből derül ki.

Vannak a nyelvnek olyan szabályai is, amelyek azt írják le, hogy ha egy bizonyos szó megjelenik a mondatban, akkor milyen további mondatrészek jelenhetnek meg. Ezeket vonzatoknak nevezzük. A vonzatokat leíró adatstruktúrát az után a szó után kötjük, amihez tartoznak. Pl.:



Ezt az adatszerkezetet pillanatnyilag a következő formában listázzuk ki:

```
*ragozottIGE*:
  ([történt:+10005+ 1103],
   *Vonzata*:
    (*Vonzattípus*:(ALANY), *Vonzott*:
      (*Névszói*:
        ([válás:+21020],
         [A:+3804])),
     *Vonzattípus*:( 1904,), *Vonzott*:
      (*Névszói*:
         (*Vonzattípus*:( 3300,), *Vonzott*:
           (*Névszói*:
            ([8:+23004])))))),
     *Vonzattípus*:( 1901,)))
```

Természetesen a számok itt is sokkal többet jelentenek, mint amit a első vázlatos ábra szövege mond.

VÉGZŐDÉSEK

A magyar nyelv tipikusan agglutináló nyelv. Ez azt jelenti, hogy a szavakhoz több rétegben rakódnak végződések. Nyelvünkben néhány fajta előtag is szerepet játszik. Ezek az igekötők, melyek egyes esetekben különválhatnak a szótőtől, és a melléknevek felső és túlzófokát kifejező *leg-* illetve *legesleg-* előtag. Az előtagok alkalmazása több szempontból olyan, mint a szóösszetétel - pl. szóelválasztás szempontjából - míg a toldalékok helyes használata sok gondot okoz úgy a magyar anyanyelvűeknek, mint a magyarul megtanulni szándékozóknak.

A morfológiai elemzés feladata, hogy a szavakat funkcionális részekre vágja. A magyar nyelvnél gyakran alkalmazott módszer, hogy a szavakat jelentéssel bíró morfémákra, és jelentés nélküli fonetikai töltelékre (kötőhangokra) bontják. Kérdés persze, hogy mik a morfémák, és mik a kötőhangok. Mi ezt a kérdést megkerüljük, és a toldalékok részének tekintjük az összes hozzátartozó fonémát. Ily módon egy morfémának számos megjelenési formája van. Azzal a kérdéssel, hogy egy végződés mikor milyen alakot ölt a felszínen, a morfológiai elemzés kapcsán részletesebben foglalkozunk, de hogy miért, arról nem kívánunk elméleti meggondolásokba mélyedni. Jelen fejezetben a toldalékok funkcionális tulajdonságairól szeretnénk képet rajzolni.

A toldalékok hagyományos felosztása: ragok, jelek, képzők. A szokásos (jelentéstani) magyarázat szerint: a képzők megváltoztatják a szó jelentését, a jelek módosítják azt, míg a ragok a szónak a mondatbeli viszonyát határozzák meg.

Antal László "Egy új magyar nyelvtan felé" című művében ennél pontosabb kritériumot ad a felosztáshoz. A kritériumok alapvetően formaiak, és másodsorban szemantikaiak. Névszók esetén szerinte:

1. A jelek egyértelműen felsorolhatók: a többes szám jele, a birtokragok és a birtokos jel.
2. A rag olyan kötött toldalék, mely megjelenhet mindenféle jel (és képző) után, nincsenek tiltott szemantikai sávjai, és a ragot nem követheti semmilyen más toldalék.
3. A képzők kötött toldalékok, melyek társulhatnak más képzőkkel, nem léphetnek fel jelek után, és vannak szemantikai tiltó sávjaik.

A mi kiindulási alapunk hasonló, bár alapvetően szintaktikai alapú. Egy toldalékot aszerint sorolunk be, hogy:

- milyen szerepe van a mondatban,
- hol a helye a szóban,
- milyen szavakra alkalmazható.

Ennek alapján a következőképpen osztályozzuk a toldalékokat:

Ragok

A ragok közös tulajdonsága, hogy mindig a szó utolsó toldalékai, tehát a később említendő kivételeket nem tekintve a ragot már semmilyen toldalék nem követheti.

Esetragok

Névszó végen álló toldalék, amely a névszóból határozót csinál (ideértve a tárgyat és alanyt is). Az esetragok meghatározzák egy egész névszói csoportnak a szerepét egy mondatban, vagy egy részkifejezésben. (ld. vonzatok, névszói kifejezések).

A különböző nyelvtani munkákban különböző számú esetragot tartanak nyilván. Mi a következő 27-et tartjuk számon:

t: *pirosat*,
ban: *pirosban*,
ba: *pirosba*,
ból: *pirosból*,
an:* *pirosan*,
on: *piroson*,
ra: *pirosra*, adszorra:* *nyolcadszorra*
ról: *pirosról*,
vá: *pirossá*,
nál: *pirosnál*,
hoz: *piroshoz*,
tól: *pirostól*,
ként: *pirosként*,
ért: *pirosért*,
nként:* *pirosanként*,
kor:* *piroskor*,
ig: *pirosig*,
szor:* *nyolcszor*, szorta:* *nyolcszorta*
val: *pirossal*,
nak: *pirosnak*,
ul:* *magyarul*,
stul:* *pirosostul*,
képp(en):* *eredményképpen*,
lag:* *színleg*,
onta:* *naponta*,
tt:* *Pécsett*, szerte:* *országszerte*
szám(ra):* *zsákszám*, rét:* *négyrét*.

Ez a lista több szempontból is eltér Antal Lászlóétól. Mint a fentiekből is látszik, mi ide soroltuk mindazokat a végződéseket is, amit Antal László határozószói képzőnek tart (a *-gal jelölteket), olyan alapon, hogy nem alkalmazhatóak minden szóra, vagy nem állhat előttük jel. Ebből a szempontból nem azt tartjuk mérvadónak, hogy jelek megelőzhetik-e, hanem azt, hogy más toldalék már nem követheti, és határozót képeznek a szóból.

A számítógépes elemzés, során egyébként sok olyan furcsasággal találkoztunk, ami nem nekünk tűnt fel, hanem a számítógépnek. Az *-ul/-ül* rag például számos esetben állt birtokjel után (pl.: *fiául fogadta, társául fogadta*, stb.), sőt számos olyan ige van, aminek *-ul/-ül* vonzata van (pl.: *fogad, választ, beszél* stb.). Tehát a mi megítélésünk szerint rag. De találkoztunk olyan mondattal is, hogy: *Pontjainkként egy-egy merőlegest állítva...*

Az sem látszik teljesen igaznak, hogy a tisztán ragnak tekintettek bármikor alkalmazhatóak szemantikai ismérvek nélkül (pl.: ?*eléghez*, **tegnapban*, ***itthonon*, holott *eleget*, *tegnapra*, *itthonról* alak van). Inkább azt a véleményt osztjuk, hogy a ragok használata is esetleges, bár kevésbé, mint a képzőké. Időnként még határozószók is kapnak ragot, de ezt kivételnek tekintjük. Viszont például a *-szor* szinte kizárólag számnévhez fűzhető, a *-szám* csak mértékegységgel (esetleg számmal) használható, az *-an* pedig többnyire melléknévvel.

Ebből következően az esetragokat tovább lehetne csoportosítani aszerint, hogy milyen szófajú névszót követhet.

Elvileg minden raghoz hozzárendelhetnénk egy-egy latinul megnevezett esetet, de ezt azért nem tesszük, mert egy-egy raghoz több eset is tartozhat. Így pl. a *-nak* esetrag kifejezhet részes esetet, vagy birtokviszonyt. A *-t* rag általában a tárgyeset jele, de gyakran fejez ki mértéket.

Kihagytuk azokat a ragokat, melyek már egyáltalán nincsenek használatban, egy-két egyedi szót kivéve, mint pl.: imént. Így a gyakorlatilag 27 (az alanyesettel együtt 28) esetet tartunk számon. Ezek közül 11 olyan van, melyet ritkán, vagy egyáltalán nem előzhet meg jel.

A névmások ragozása is rendhagyó, de erről az alkalmas fejezet részletesen ír.

Igeragok

Az ige végén álló toldalék, amely meghatározza az ige módját, tárgyasságát, az ige alanyának számát és személyét. A valóságban az igeragban egy - a finn nyelv morfológiájához hasonló - összetett toldalék nyomai fedezhetők fel:

igemód: kijelentő, felszólító, múlt, feltételes, igenévi
tárgyasság: elkorcsosult utalás a tárgy kilétére
az alany személye: 1., 2., 3. személy
az alany száma: egyes és többes szám

Például:

látnálak=lát+ na + ál+ ε + ak
tő +feltételes mód+2.sz tárgy+1.sz+e.sz alany

láttunk=lát+ t + ε + un + k
tő +múlt mód+határozatlan tárgy+1.sz+t.sz alany

Az ilyen jellegű elemzés rendkívül bonyolult és felesleges, ezért az igeragot mi is egy oszthatatlan többfunkciós toldaléknak tekintjük. Persze nem minden esetben különböztethető meg a tárgyas és a tárgyatlan ragozás, és más esetekben is lehetséges, hogy különböző részjelek felszíni formája megegyezik:

nézzük = felszólító/kijelentő mód
láttam = tárgyas/tárgyatlan ragozás

Igeragnak tekintjük a határozói igenév képzőjét is (*-va*/*-ve*, *-ván*/*-vén*), mivel ez is az ige végén szerepel, bár funkciója inkább az esteragokénak felel meg. Ily módon mintegy 60-ra becsüljük a különböző igeragok számát.

Jelek

A jelek kötött sorrendben előzik meg az esetragot. Funkciójuk alapján három csoportjuk van; egymáshoz képest előfordulási sorrendjük:

Hátul áll a **birtokos jel**: *Pistáéi*

Egy összetett toldalék, melyben az *i*-jel többes számot fejez ki.

Elöl áll a **birtok-** vagy **többesszámjel**: *Pistája*/*Pisták*
Összetett toldalék, mely (birtokos személyrag esetén) kifejezi a tulajdonos számát és személyét, de ezt kiegészítheti egy *-i* többes számjel, mely a tulajdon számára utal: *Kutyáink*, *fiai*.

A kettő között a családi többes (pluralis familiaris) állhat, ami nem minden névszóra alkalmazható: *Pistáék*

További megszorítás, hogy mind az előtte, mind az utána álló jel csak egyes számú lehet: *fiaméké*
de nem jó: **fiaimék*
**Pistáékéi*

Mivel az egyes toldalékok további jelekre való felbontása nehézkes (ezeknél a jeleknél az *-i* többes számot jelöl), ezért az igeragokhoz hasonlóan ezeket sem bontjuk tovább, hanem többfunkciós toldalékként tartjuk számon.

Megjegyzendő, hogy egyes nyelvészek szerint több birtokos jel is szerepelhet egy szóban (pl.: *fiáéé*), de a gyakorlatban ilyen formával még soha nem találkoztunk. Szerintünk egyfajta jel csak egyszer szerepelhet egy szóban.

Képzők

A képzők azok a toldalékok, melyek a többi toldalékot megelőzik. Az a meghatározás, hogy a képzők azok a toldalékok, melyek megváltoztatják a szó jelentését (szemben a jelekkel, amelyek csak módosítják) elég homályos. Az igaz, hogy a képzők az esetek többségében megváltoztatják a szó szófaját. Főnévből melléknevet, igéből főnevet stb. képeznek. Azokat a toldalékokat is a képzők közé soroljuk, melyek egy szófajból önmagára képeznek pl. *-hat* ige-ige képző. Ellentétben a korábbi végződésekkel az így kapott szó teljes értékű marad és újabb képzők alkalmazhatók rá.

A képzők lehetséges sorrendjét tulajdonképpen az határozza meg, hogy konzisztens-e az adott morfémasorozat a szófaj-transzformációval. Persze ezen kívül sok más korlátozó tényező is szerepet játszik. Nem szerepelhet például egy toldalék kétszer egy szóban. Az egy szóban szereplő toldalékok száma is erősen korlátozott. Probléma merül fel a magyar szavak szófaji felosztásánál is, hisz a szófaji határok átfedőek a névszókon belül. Gyakran használunk főnevet melléknévi szerepben, és viszont. Ebből következően bizonyos toldalékok nem sorolhatóak egyértelműen az elvi 16 csoportba (ha 4 szófajt veszünk figyelembe). Némelyik több csoportba is beilleszthető, míg más csoportok üresek maradnak.

Mindezek alapján a képzőknek a következő csoportjaik lehetnek:

ige-ige:	*-hat, -gat, -tat ...*
ige-főnév:	*-ás, -alom, -dalom, -at, -mány...*
ige-melléknév:	*-ó, -ott, -atlan, -andó ...*
névszó-ige:	*-al, -ez, -kodik ...*
melléknév-ige:	*-ít, -oz ...*
főnév-főnév:	*-ság, -ka, -né*
névszó-melléknév:	*-s,-talan,-szerű,-féle,-ú,-i,-beli ...*
melléknév-melléknév:	*-bb*
számnév-melléknév:	*-ad, -adik,* stb.

A képzőkre sokkal nagyobb mértékben jellemző, mint a ragokra és jelekre, hogy a képző az egyik szóra alkalmazható a másik szóra nem. (Pl.: *jár-kál, buj-kál, mász-kál* stb. de nincs **ház-kál*, vagy **zár-kál*.) Az is jellemző, hogy egyes képzők különböző szavak jelentését különbözőképpen módosítják. (Pl. *katonaság*, *asszonyság*, *hercegség*, stb.) Ha egy képzőről nem tudjuk megmondani valamilyen rövid és tömör formában, hogy melyik szóra alkalmazható, hogyan változtatja meg annak az értelmét, milyen lesz annak a szónak a ragozása, milyenek lesznek a keletkezett szó vonzatai, akkor semmi okosabbat nem tehetünk, mint hogy az összes képzett szót berakjuk szótárba, és ott adjuk meg a minden egyes képzett szó szükséges jellemzőit. Van viszont egy sor olyan képző, aminél jól megadhatók a szón végbement változások; az ilyen képzőket

hívjuk reguláris képzőknek. Általában is egy végződést akkor tekintünk regulárisnak, ha meg tudjuk mondani milyen szavakhoz kapcsolható, és a keletkezett szó hogyan alkalmazható.

A ragokat és jeleket eleve reguláris végződéseknek tekintjük, noha mint láttuk, számos megszorítás lehet az alkalmazhatóságukra. Egy képzőt is akkor tekintünk regulárisnak, ha a megszorítások mértéke hasonló.

Mi jelenleg a következő csoportokat kezeljük:

ige-ige képzők: *-hat, -gat, -tat*,
ige-főnév: *-ás*,
ige-melléknév: *-ó,-ott,-atlan,-andó,-ható,-hatatlan*,
melléknév-melléknév: *-bb, leg...bb, legesleg...bb*,
számnév-melléknév: *-adik*,

Valószínűleg ennél sokkal többet lehetne kezelni, csak alaposan végig kellene gondolni a dolgot elsősorban a vonzattranszformáció szempontjából. (Ilyenek például az: *-ít, -s, -ú, -talan, -beli, -féle, -fajta, -szerű* stb.) (A vonzattranszformációt lásd az **igenevek**nél.)

Mint jól látható, a középfok jelét mi képzőnek tekintjük, mivel akár más képző is követheti (pl.: *jobbító*). Azt, hogy milyen képzőt milyen másik követhet, alapvetően az határozza meg, hogy a képző milyen szófaj-transzformációt hajt végre. Egy biztos. Az egy szóban található képzők számát a megérthetőség korlátozza. Így egy képző csak akkor ismétlődhet egy szóban, ha a korábbi képzett alak sajátos jelentéssel bír, így önálló képzetlen szónak tekinthető. Pl.: *biztonság-os-ság*.

Mindezek alapján érthető, hogyha figyelembe vesszük a képzett alakokat is, egy átlagos szótő több ezer formában szerepelhet a szövegekben.

Kivételes esetek

Azt mondtuk, hogy a határozószó és rag után már semmilyen végződés nem alkalmazható. Ezzel szemben bizonyos hol kérdésre válaszoló határozók, bizonyos névutók, a *-nként* és a *-szor* rag után alkalmazható az *-i* melléknévképző.

Hasonló a *-lag* és a *-szor* rag is, ami után az *-os* képző előfordul, így ragozott alakból melléknevet képezhetünk ezekben az esetekben.

Ennek a kezelését úgy képzeljük el, hogy az *-i* képzős határozószók, és névutók szerepelnek melléknévként a szótárban, a *-nkénti, -szori, -lagos, -szoros* végződéseket pedig egyetlen képzőnek tekintjük.

További kivételek - melyet elkészült elemzőnkben még nem veszünk figyelembe - azok az igeragos szerkezetek, melyeket jelzőként használunk a mondatban. Ezek közül a múlt idejű és feltételes módú tárgyas ragozás után használható a *-ban/-ben* esetrag (pl. *jártamban*)

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. Az analitikus modell gyakorlati kivitelezésének leírása után néhány tapasztalati tanulságot vonunk le, s végül ennek megfelelően egy javított módszer megvalósítását javasoljuk.

I. A magyar nyelv morfológiája

Szavak, morfémák és alakjuk

Szónak nevezzük a nyelv azon elemét, melyet írott szövegben egybeírunk (az ABC betűit, számjegyeket és esetleg kötőjelet felhasználva). Szóalaknak nevezzük ezek írásban megjelenő karakterlánc-formáját.

A magyar nyelv úgynevezett agglutináló nyelv, ami azt jelenti, hogy a szótőhöz több rétegben végződések tapadhatnak, és néhány előtag is kapcsolódhat a szóhoz. Ezenkívül a szóösszetétel is lényeges szóalkotási mód.

A szó funkcionális részekből áll, melyeket morfémáknak nevezünk (szótövek, képzők, ragok, jelek). Egy adott funkciót megvalósító morfémának több különböző alakja lehet. A magyar nyelvben egy szó tehát nem más, mint egy morfémasorozat (de a morfémák nem kapcsolódhatnak tetszőlegesen egymáshoz).

A szavak különböző szempontokból osztályokba sorolhatók (lásd később: **szófajok**, **ragozási típusok**). A toldalékok (előtagok) egyrészt a szavak jelentését változtatják meg, másrészt a szavakat egyik szóosztályból a másikba viszik át. A morfológia feladata leírni azt, hogy egy szó milyen morfémasorozattal állítható elő. A generatív leírás célja megmutatni azt, hogyan kell egy szóalakot előállítani. Az analitikus modell feladata: hogyan lehet egy szóalakból a benne szereplő morfémákat megtalálni. Mivel több morféma sorozat ugyanazt a szóalakot eredményezheti, ezeket mind meg kell találnunk.

Mivel szoros kapcsolat van a szavak osztályai és a lehetséges morfémasorozatok között, ezért a morfológia a szavak osztályozásával is foglakozik, de csak a morfológia szempontjából. A szóosztályok szintaktikai és szemantikai tulajdonságait más fejezetben tárgyaljuk.

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 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 szófajának és a toldalékok szerepének meghatározása.

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 morfémákra való bontása a további (szintaktikai, szemantikai) elemzést megelőzően. Mivel a teljes nyelvi rendszer interaktív ember-gép kapcsolat részére készül, a válaszidők rövidsége érdekében lényeges az elemző nagy sebessége. A rendszer az általunk köznyelvinek tartott alakokkal foglalkozik, és nem foglalkozik régies, vagy tájnyelvi alakokkal, de alkalmas bővítésre, szűkítésre, módosításra is, ha ilyenre lenne szükség. A kidolgozott módszer alkalmas gyors elemző előállítására morfológiai szabályok alapján, akár más aglutinatív nyelvek esetén is.

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 toldalékot, 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, és minden előtagnak egy alakja van.)

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 a magyar nyelvben mindig a (relatív) szótő utolsó egy-két betűjét érinti a szomszédos morfémák függvényében.

Toldalékolás: egy adott (relatív) szótőre alkalmazható egy toldalék, ha a (relatív) szótő szóosztálya megegyezik a toldalék alkalmazhatósági szóosztályával. Ekkor először végrehajtjuk a tövön a toldalékhoz tartozó tőváltozást, majd az így kapott módosított tő után fűzzük a toldalékot, és eredményül a toldalék által előírt szóosztályba tartozó szót kapunk. Tehát például a pörög igére alkalmazható az -ő jelen idejű melléknévi igenévi képző (és nem alkalmazható az -ó), mert az -ő magas hangrendű igékre alkalmazható. Mielőtt az igetőhöz hozzáfűznénk a toldalékot, végre kell hajtani a magánhangzó-kiesés tőtranszformációt: *pörg-ő*. Az így kapott szó egy magas hangrendű melléknév lesz.

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 mivel a magyar nyelvben a toldalékolás többnyire növeli a szó hosszát, és csak néhány olyan lépés jöhet egymás után, ami a szó hosszát változatlanul hagyja (üres toldalék vagy hangzókiesés miatt), így egy szóalaknak nem lehet végtelen sok felbontása morféma sorozatokra. Ha tehát egy lehetséges szótőből kiindulva sorban előállítjuk az összes olyan morféma sorozatot, ami nem hosszabb, mint a szóalak, akkor ezek között ott lesz a szó összes lehetséges felbontása. Hasonlóan, ha egy szóról levagdossuk a lehetséges toldalékokat, és közben visszacsináljuk a tőmódosulásokat, előbb-utóbb megtaláljuk a szó lehetséges abszolút tövét vagy töveit. Mivel az egyes toldalékolási lépesek véges automatával leírhatók, így az elemzés egy véges automatával véges lépesben elvégezhető.

II. A kivitelezés gyakorlati kérdései

A fenti megállapítás nem jelenti azt, hogy az elemzés minden részét a gyakorlatban egy klasszikusan felépített véges-állapotú automata átmenetmátrixával kell reprezentálni. Ha például az összes szóba jövő szótő azonosítását ily módon valósítanánk meg, minimális szótár esetén is egy többtízezer állapotú véges automatára lenne szükség, amelynek mind a létrehozása, mind a karbantartása csak valamilyen automatizmussal kivitelezhető, nem is beszélve a memóriaigényről.

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 szótőtől a toldalékok egyenkénti levágásán keresztül balról jobbra, de mások tapasztalatával megegyezően úgy véljük, hogy ez a magyar nyelvnél több hibás részelemzéshez vezetne, így kevésbé lenne hatékony.

Problémát okozhat a tőváltozások kezelése is. Az Apresz-ján és csoportja által létrehozott orosz elemzőben nincs algoritmikus tőváltozás-kezelés, hanem a szavak tőváltozataikkal együtt szerepelnek a szótárban. Ez a módszer a magyar nyelv esetén közvetlenül nem alkalmazható, mert toldalékolt tő is szenvedhet tőváltozást.

A finneknél alkalmazott kétszintű véges automata (Koskenniemi és társai) formalizmusa nehézkesnek tűnik a magyar tőváltozások leírásához és azok fonológiai illeszkedési szabályai esetén. (A finn nyelvben a szótőnek van egy állandó és egy változó része, és a változatok aránylag jól paradigmába gyűjthetők. A toldalékoknál is jellemző a felszíni alakok változatai, míg a magyarban nagyobb súlyt kapnak az alakok transzformációi, melyek ráadásul gyakran opcionálisak.)

Egy másik finn reprezentáció (Jäppinen és társai) a morfémák szomszédsági tulajdonságaiból indul ki, ami nem természetes a morfoszintakszis leírásánál.

A jelenleg megvalósított modell

Mivel erőnk és nyelvtani tudásunk végességével tisztában voltunk, egy egyszerűsített, de belátható időn belül kivitelezhető modellt alkottunk. Modellünk a következőkből áll:

1. Szótár - szóosztályok:
2. Toldaléktár - toldalékosztályok
3. Morfoszintakszis
4. Tőváltozások
5. Fonológiai illeszkedési szabályok

1. Szótár - szóosztályok

A szótárban a szavak tövei szerepelnek, egy kódszámmal, mely a szó szófaját meghatározó, és a morfológiai elemzést elősegítő egyéb kódot tartalmaz. A szótár csak rendhagyó tőváltozatokat, illetve kivételes toldalékolt alakokat tartalmaz. Ez utóbbi esetekben a szótári tételben jelölve van az alapesetre való visszavezetés: *több=sok+abb*, *terh=teher*.

Szófaji osztályok: 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 (a ragozott igék közé soroljuk a 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 tartoznak a nem toldalékolható tövek (pl. névelő, kötőszó).

2. Toldaléktár - toldalékosztályok

A toldalékok tára tartalmazza a toldalékok azon alakjait, melyek szó végén előfordulhatnak. A kötőhang része a toldalékalaknak. Egyrészt a morfoszintakszis szempontjából vannak csoportosítva, másrészt fonológiai kóddal láttuk el őket (lásd **fonológiai illeszkedési szabályok**).

Toldalék és előtagosztályok: tartalmazzák az összes levágható toldalékot a morfoszintakszis szempontjából:

Igeragok: ragozott igéről vághatók le, és levágása után ragozatlan ige marad. Az igerag tulajdonképpen egy összevont toldalék, amely magába foglalja az időt, módot, tárgyasságot, 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 igéről választhatók le, és a maradék is ragozatlan ige: csak a *-hat/-het*, *-gat/-get*, -tat/*-tet/-at/-et* képzőkkel foglalkozunk.

Ige-névszó képző: névszói tőről választhatók 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 választhatók le, és a maradék birtokos jel nélküli névszó (pl. *fül-é*)

Birtokjel-többes számjel: 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*).

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

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

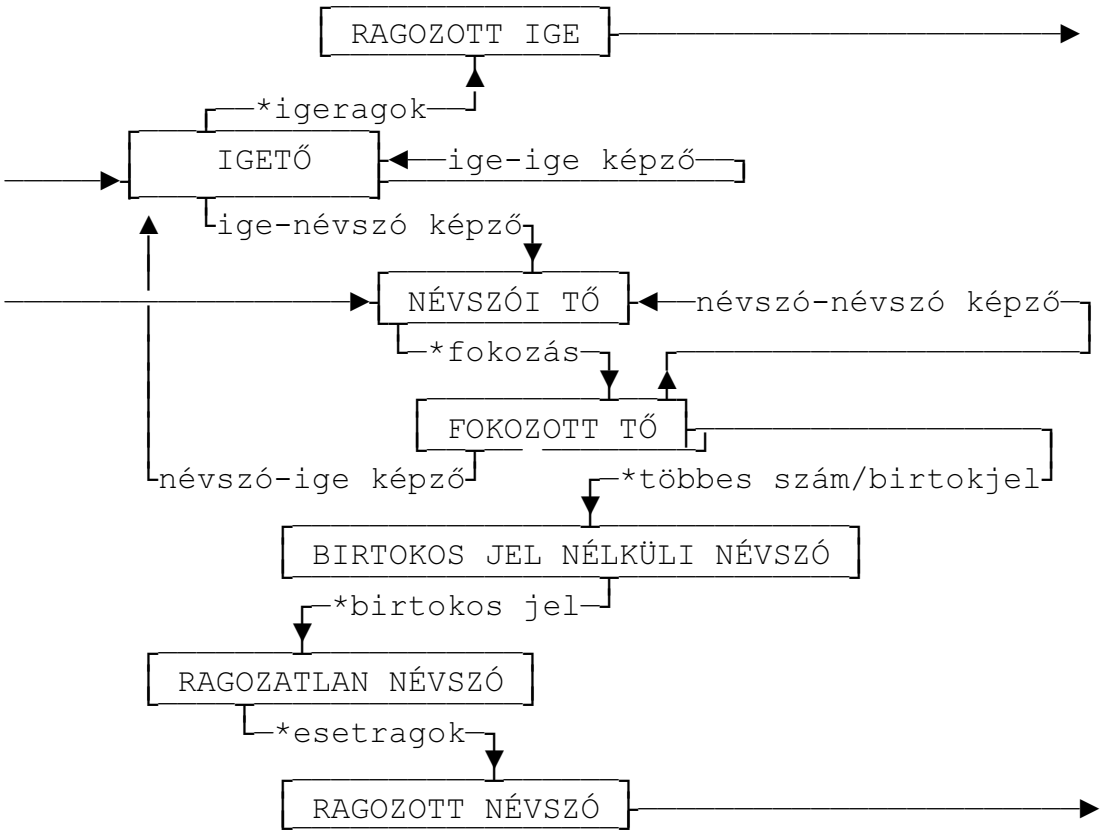
Névszó-igei képző: igetőről választható le, és a maradék fokozott tő (pl. *kék-ít*).

Felsőfok: Előtag, mely a középfok jelének levágásával egy időben választható le. (pl. *leg-első*, *legeslegesleg-jo-bb*).

Az igekötőket jelenleg nem vettük be az előtagok közé, és levágásukat sem az elemzőbe, mivel használatuk általában olyan szemantikai módosulást von maga után, mely nem vezethető le az ige és az igekötő jelentéséből. A későbbiekben ez az álláspontunk módosulhat.

3. Morfoszintakszis, a 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) felel meg. A *-gal jelölt toldalékok lehetnek "üres" toldalékok is. Bár a felírt gráf alapján a toldalékok száma nem korlátozott, a gyakorlatban egyszerre 3-4 toldaléknál többet ritkán használnak a magyar nyelvben, sőt bármiféle ismétlés mesterkélt (*káposztás-ít-ott-talan-ít-ottátok*), vagyis bármely nem üres toldalék csak egyszer szerepel egy szóban, de az elemzés során nem zárható ki a ciklus, vagyis azonos toldalékosztályból több is előfordulhat. Pl.: *kékíthetetlenül* elem alak kategória szóosztály szóalak

kék	szótő	névszói tő	*kék*
---	fokozó	fokozott tő	*kék*
-ít	névszó-ige képző	igei tő	*kékít*
-het	ige-ige képző	igei tő	*kékíthet*
-etlen	ige-névszó képző	névszói tő	*kékíthetetlen*
---	fokozó	fokozott tő	*kékíthetetlen*
---	számjel	birtokos jel nélk.tő	*kékíthetetlen*
---	birtokos jel	ragozatlan névszó	*kékíthetetlen*
-ül	esetrag	ragozott névszó	*kékíthetetlenül*

Ezzel a szekvenciamodellel úgy generálhatunk egy szóalakot, hogy kiindulva a névszói vagy igei tőből a nyilak mentén haladva a szótőhöz újabb és újabb végződéseket fűzünk (amelyek között üresek is lehetnek), míg végül a ragozott igénél, vagy a ragozott névszónál véget ér a szekvencia.

Nem ilyen egyszerű a helyzet az elemzésnél. Itt a gráfot a nyilakkal ellentétes irányban kell bejárnunk, de nem elegendő egyetlen bejárást kiválasztani, hanem minden lehetséges bejárást meg kell keresnünk. Ez technikailag úgy oldható meg, hogy amikor egy bejárást befejeztünk, vissza kell térnünk egy korábbi csomópontra, ahol elágazás volt, és onnan kezdve a többi lehetséges folytatást is végig kell járnunk. Ha a legutolsó elágazás minden lehetséges alternatíváját bejártuk, eggyel korábbi elágazásra kell visszatérni, és így tovább.

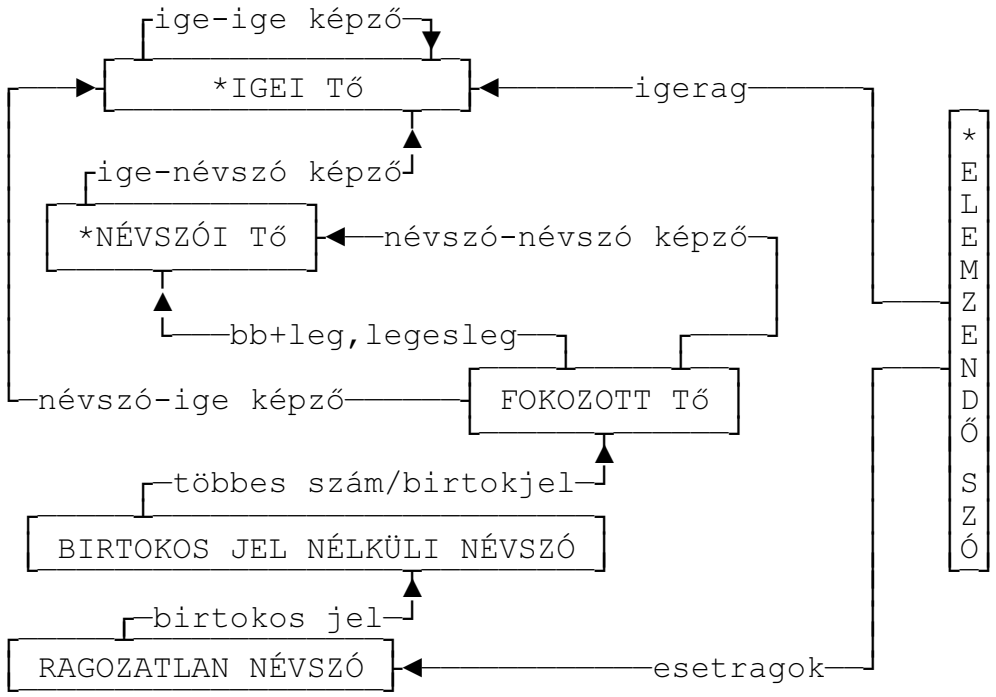
Azt, hogy minden bejárás előbb-utóbb véget ér, az garantálja, hogy a gráfban nincs olyan kör, ami csupa üres toldalékból áll, azaz a szó előbb-utóbb elfogy.

Amikor a névszói vagy igei tőhöz érünk, meg kell állapítanunk, hogy a szó maradéka valóban igei vagy névszói tő-e. Hasonlóan, már az elemzés megkezdése előtt érdemes megnézni, hogy az elemzendő szó nem toldalékolhatatlan tő-e.

(Bizonyos tőtorzulások és más kivételes alakok kezelése érdekében a szótárban nem csak szótövek, hanem összetett alakú vagy jelentésű szavak is szerepelnek. Ezeknél a szótárban a szó felbontása meg van adva. Ennek beillesztése az elemzésbe további feladat.)

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 ezek a formák is beépíthetők a morfológiai elemzőbe. A gráf alapján természetesen elemezhetők olyan mesterségesen kitalált szavak is, melyek a magyar nyelvben nem fordulhatnak elő (pl. *fék-ebb-ít*). A gyakorlatban ez a helyes szavaknál ritkán vezeti rossz felbontáshoz az elemzőt. A jó megoldás ezekben az esetekben is előáll, de alternatívaként nem létező felbontást is kaphatunk. Ezek számát különböző kivételkezelő eljárásokkal csökkenthetjük.

Tehát az elemző gráf a következőképpen néz ki:



Ez a gráf az elemzés vázát adja meg, pontosabban megadja, hogy az egyes morfémaosztályokba tartozó morfémák hogyan vannak felfűzve. Az elemzés az ELEMZENDŐ SZÓ csomópontnál indul el. A *-gal jelölt csomópontoknál szótári keresést is végre kell hajtani. 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.

4. Tőváltozások

A jelenleg kezelt tőváltozások a következők:

Névszóhasonulás: *do**b** - do**bb**al*
Igei hasonulás: *hú**z** - hú**zz***
Ige t-s átalakulás: *lát - lá**ss**uk*
Szóvégi magánhangzónyúlás: *kuty**a** - kuty**á**t*
Szóvégi v betű: *h**ó** - h**av**as, dar**u** - dar**v**ak, n**ő** - n**öv**ő*
Szóvégi magánhangzó-rövidülés: *ajt**ó** - ajt**a**ja*
Szóvégi magánhangzó-kiesés: *ap**a** - apja*
Belső magánhangzó-kiesés: *pör**ö**g - pörgő*
Belső magánhangzó-rövidülés: *szam**á**r - szam**a**rak*
Igei sz-kiesés: *e**sz**ik - ettek*
Harmadik azonos mássalhangzó-kiesés: *jo**bb** - jo**bb**ól*

Ezeknek a tőváltozásoknak számos alváltozata létezik, s egyesek egyszerre is előfordulhatnak: magánhangzó-kiesés+ magánhangzó-rövidülés: *három – harmadik* névszóhasonulás+ harmadik mássalhangzó kiesése: *jobb – jobbal,* ige *sz*-kiesés+ *t-s* átalakulás: *te**tsz**ik-te**s**sék*

A tőtorzulási szabályoknak tulajdonképpen az inverzeit kódoljuk. A jelenlegi algoritmusban ahhoz, hogy egy tőváltozás inverzét végrehajthassuk, minimális ellenőrzést végzünk el. Ellenőrizzük, hogy a toldalékot megelőző torzult szótő vége lehet-e egy torzult szótő. Például a belső magánhangzó-kiesés inverzét, a magánhangzó-beszúrást csak akkor érdemes megkísérelni, ha a maradék szóalak végén két mássalhangzó van. Másik feltétele a tőváltozás visszacsinálásának az, hogy az éppen levágott toldalékalak előidézheti a tőtorzulást.

5. Fonológiai illeszkedési kódok

Ezek a kódok két célt szolgálnak. Egyik funkciójuk, hogy ellenőrizzük, hogy az adott toldalékalakot megelőzheti-e a maradék szóalak utolsó betűje. Másik funkciója, hogy megtudjuk, milyen tőváltozás mehetett végbe az adott toldalék következtében. Például az *-ok* többes számjel csak mássalhangzó után következhet, mely előtti szóalak lehet változatlan vagy belső magánhangzó-kieséses: *korok*, *bokrok*. Ezzel szemben az *-ak* többes számjel magánhangzót is követhet, de előidézhet belső magánhangzó-rövidülést is: *forróak*, *irodalmak*, *szamarak*.

Mivel jelenleg nem áll rendelkezésünkre sem a szavak toldalékolási, tőváltozási szótára, sem az osztályozás pontos elve, itt is csupán egy durva megközelítést alkalmaztunk. A szóalakból következtetünk a lehetséges alakváltozatokra. Ilyen értelemben a toldalékokat 32 fonológiai illeszkedési osztályba soroltuk. Minden egyes osztályhoz megadunk egy betűhalmazt és egy tőváltozáshalmazt, melyek megelőzhetik az adott osztályba sorolt

toldalékalakokat. A morfémák hangtani illeszkedése szempontjából csak bizonyos tiltó szabályokat veszünk figyelembe. Tehát pl. a *-t* tárgyeseti ragalak előtt nem állhat *e*-re *a*-ra végződő morfémaalak, csak ezek nyújtott végű változatai, s ezekben az esetekben kötelezően vissza kell rövidíteni az elemzés során. A jelenleg használt osztályozást a melléklet tartalmazza.

A modell implementálása

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

A szóelemzés két menetben történik. Először megkeressük a kérdéses szó lehetséges feldarabolásait, majd ezeket átkódoljuk egy tömör kóddá, mely csupán a morfémától, s nem annak alakjától függ.

A hatékonyság érdekében a végződéseket véges automatával vágjuk le. A morfémaosztályoknak megfelelő morfémahalmazokból végesállapotú automatákat generálunk. A végződéseket jobbról balra próbáljuk levágni a szótőről az automatával. Az automata végállapotba (egy esetleg helyes levágáshoz) akkor jut, ha a levágott toldaléknak megfelelő fonológiai kódnak megfelel a maradék szóalak vége. Ilyenkor a fonológiai kód alapján megkíséreljük rekonstruálni az esetleges tőváltozások következményében torzult szóalakot.

A részautomatákat a morfoszintakszis gráfja fűzi össze egy nagyobb automatává. Ezen két szint alatt létezik a harmadik szint, mely a levágott morfémaalaknak megfelelő fonológiai egyeztetést, illetve a lehetséges tőváltozások inverzét végzi el. Végül a morfoszintakszisról szóló részben említett pontoknál egy hash-eléssel gyorsított szótári szótőkeresést hajtunk végre.

Az elemzés több helyen alternatív ágakra bomolhat. Ennek egyrészt oka a morfoszintakszis inverz gráf többirányúsága, másrészt az egyes toldalékokat megelőző tőváltozások inverzeinek száma.

A külön ágakat backtrack módszerrel járjuk be. Egy elemzési ág a következő esetekben ér véget:

a./ Elfogy a karakter a szóalakból.

b./ Az automata az adott karakterre már nem lép tovább.

c./ A maradék szóalak megfelelő szófajjal szerepel a szótárban, s a szóhoz rendelt kódban szerepel a "tovább ne elemezz" információ. (Ezt azoknál a szótári tételeknél alkalmazzuk, melyek további elemzése vagy felesleges, vagy félrevezető lenne.)

Egyébként a bejárt út során több megoldást is találhat az elemző sikeres szótári keresés esetén.

A morfémák kódolását csak akkor végezzük el, ha sikeres felbontást kapunk. Ennek két oka van. Egyik oka, hogy a sikeres feldarabolások számához képest a sikertelen ágak száma elég nagy. A másik szempont, hogy helyesen feldarabolt morfémasorozatnál pontosabban tudjuk meghatározni, melyik morfémaalak mit jelent. (pl. *-ja* lehet felszólító módú és kijelentő módú igerag, de *t* végű ige után csak kijelentő módú lehet.)

Sikertelen elemzés esetén megpróbáljuk a szóalakot összetett szóként értelmezni. A magyar nyelvben a szóösszetételek döntő többségében az első tag egy névszói szótő, míg a szó toldalékolási osztályát az összetétel második tagja határozza meg.

A gyakorlati tapasztalat

A kezdeti próbálkozások alapján hatékony és rugalmasan generálható módszert dolgoztunk ki. A végződéseket levágó automatát egy külön fázisban, az elemzést megelőzően állítjuk elő a toldalékok halmazából, így annak ellenére, hogy a végződések készlete rugalmasan változtatható, a felhasználáskor egy gyorsan működő optimált automata kerül a programba.

Felmerül a kérdés, mennyire működik helyesen az algoritmusunk. 100 százalékos algoritmust nem lehet írni, hiszen egyes alakok helyességéről vagy helytelenségéről még a nyelvészek véleménye is megoszlik. Noha a szabályok mind szótani, mind hangtani illeszkedési szempontból elnagyoltak, a helyes szóalakok helyes elemzését mindenképpen megtalálja az elemző. A kérdés inkább az, hogy milyen mértékben keletkeznek félreelemzések is. Ahhoz, hogy a félreelemzések okát kiderítsük, a többértelműségeket kell megvizsgálni. A többértelműségnek több oka lehet:

1. Maga a szótő többértelmű: *lép*, mint a méh viaszháza, és mint ige. Egyes toldalékok megszüntethetik a kétértelműséget pl.: *lépjek*, más toldalékok viszont nem szüntetik meg, pl.: *lépnek*, hiszen a *-nek* lehet igerag is, és esetrag is. A szövegben a leggyakoribb többértelműség a névelőkből adódik, mert az *az* határozott, és az *egy* határozatlan névelő formailag megegyezik az *az* utalószóval, illetve az egy számnévvel. Félreelemezés azért keletkezhet, mert a toldalékolási (fonológiai?) modellünk nem elég finom, pl.: adott ragozott ige is lehet és melléknévi igenév is, de az *adottak* egyértelműen többes számú melléknévi igenév, mert az *ad* nem kaphat *-ottak* igeragot, noha ilyen rag van például az *oltottak* szóban.

2. A toldalék többértelmű: *-ják* mint kijelentő módú igerag és mint felszólító módú igerag, pl.: *csinálják*. Az a tény, hogy egy végződés többértelmű, nem mindig jelenti azt, hogy a szó maga is többértelmű lesz. Pl.: az *iszik* és a *nézik* szó egyértelmű, noha az *-ik* rag egyik esetben egyes szám harmadik személy, másik esetben többes szám harmadik személy. Ahhoz, hogy eldöntsük, hogy az adott esetben melyikről van szó, ismerni kellene, melyik szó melyik ragozási (fonológiai?) osztályba tartozik. Ez a jelenlegi modellünkből hiányzik. (Szerencsére, ez a mondat értelmezéséhez nagyon ritkán szükséges.)

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

4. A kétértelműség annak a következménye, hogy egy toldalékolt alak egybeesik egy toldalékolatlan szóval. Pl.: *leves, lé-s*=leves (*leves gyümölcs, gyümölcsleves*), vagy *adó* és *ad-ó*, vagy *élet* és *él-et*.

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

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

Félreértelmezés az utóbbi két esetben fordul elő, hisz a korábbiak a szóalakok valódi többértelműségének következménye. Igaz, hogy például a hangrendi illeszkedést figyelmen kívül hagyjuk, mivel ennek az elemzés szempontjából általában nincs jelentősége, valamint a szavak hangrendje nem mindig határozható meg egyértelműen a szó írott alakjából. (A *derék* vagy az *iszik* szó mély hangrendű.) A fonológiai szabályok durva volta már zavaróan sok félreelemzéshez vezetne, ha nem alkalmaznánk a leállító kódot. A leállító kód a szótárban a szótő mellé van írva, és azt jelzi, hogy a maradékot nem kell tovább elemezni. Ha azonban nem szerepel a szótárban a tévesen felbontható szó (példánkban a *balek*), akkor előfordulhat az 5. és a 6. típusú hibás elemzés.

A kísérleti próbálkozások alapján (5000-es szótár, 160 toldalék, 500 toldalékforma) az elemző által többértelműen felbontott szavak túlnyomó többsége valóban többértelmű (leggyakrabban múlt idejű ige és múlt idejű melléknévi igenév gyakori azonos alakja miatt). Mint a fentiekből is látható, az egy százalék körüli félreértelmezéseknek nem a morfoszintakszis elnagyolása az oka, hanem a fonológiai modell gyengesége.

II. Egy tervezett modell

A jelenleg működő morfológiai elemzőnk a morfoszintakszist meglehetősen nagyvonalúan kezelte, például egy csoportba rakott minden igeragot. Ez a felfogás nagyon jól megfelelt annak a felfogásnak, hogy a mondatbeli szerepét tekintve nagyjából minden igerag azonos szerepet játszik, így például az alanyállítmány-egyeztetést többé-kevésbé azonos módon kell elvégezni, a jelen időben, a múlt időben, a feltételes módban, a felszólító módban, sőt a személyragozott főnévi igenévnél is. A morfémáknak ezt a felosztását tehát a szintakszis szempontjából meg kell tartani.

Nem biztos azonban, hogy a morfémáknak ugyanez a felosztása célszerű a morfológiai analízis szempontjából is. Ha például megnézzük a *mászik, néz* és *lát* igét a ragozás szempontjából, azt találjuk, hogy ezek az igék meglehetősen különbözően ragozódnak. Például a *-lek* végződés csak a néz igéhez kapcsolható, a másik kettőhöz nem, az *-om* végződés nem egészen ugyan azt jelenti a mászik és a lát esetén, az *-ik* végződés egészen mást jelent a *mászik* és a *néz* esetén.

Célszerű tehát a végződéseket tovább osztani különböző csoportokba, és ugyanekkor fel kell osztanunk a szótöveket is, és nem azt mondani, hogy igető után igerag jöhet, hanem azt mondani, hogy ebbe a csoportba tartozó igék után, ebbe a csoportba tartozó igeragok jöhetnek. Ezt a csoportosítást oldaná meg a pontosabb fonológiai osztályozás, amely általában megfelel a hagyományosan paradigma-osztályoknak nevezett csoportosításnak.

A magyar nyelv szavainak jelentős hányadára Elekfi professzor készített egy ilyen osztályozást, ami sajnálatos módon mind a mai napig nincs kiadva, és ezért mi csak kevéssé ismerjük.

Sajnos a szavak ilyen felosztásával a dolog még csak félig van megoldva, mert a magyar nyelvben nem csak ragok, hanem képzők és jelek is vannak, amelyek után a toldalékolás folytatódhat. Ezért ezeknél a végződéseknél nem csak azt kell megmondani, hogy egy adott szóra például a birtokjelek melyik alcsoportja alkalmazható, hanem azt is, hogy ezek után a keletkezett szó milyen toldalékokat kaphat.

Például vegyük az *alak-ja* és az *alak-já-t* morfémasorozatot. Az alak szóhoz mind a *-ja*, mind a *-já* végződést hozzá lehet kapcsolni, a -ja végződés után azonban csak az üres, a -kor és a -ként rag kapcsolható, viszont a *-já* végződés után egy sereg más végződés jöhet, de a fentiek nem. (A korábbi modellben ezt úgy írtuk le, legtöbb rag megnyújtja a szóvégi a hangot, de a *-kor* és a *-ként* rag nem.)

Mindezek alapján világos, hogy a morfémák fonológiai tulajdonságát jobban fel kell használni az elemzésnél. Másrészről közismert, hogy a szavak fonológiai tulajdonságai nem mindig határozhatók meg megjelenési formájukból. Ezért szükséges a szótárban ezek kódolása.

A morfémák és felszíni alakjaik előfordulását két szabály befolyásolja:

1. Morfoszintakszis
2. Fonológiai illeszkedés

A morfoszintakszis azt határozza meg, hogy milyen szófajú (relatív) tőhöz milyen típusú toldalék csatlakoztathat. Ezt egy gráffal lehet legkényelmesebben megadni, hasonlóan, mint a már működő elemzőnknél.

A fonológiai illeszkedési szabályok ezzel párhuzamosan a lehetséges morfémaalakok hangtani illeszkedését vezérlik. A magyar nyelvnél a morfémák alakjai nagyobbrészt a szomszédos morfémaalakoknak, kisebb részt a szótő hangrendi besorolásának függvénye. Így ezen összefüggések nagyobb része szomszédsági relációkkal írható le jól, míg néhány szabály szintakszisgráffal adható meg természetesebben. Mindezeket közvetve vagy közvetlenül jobb és baloldali attribútummá kódolhatjuk. A morfémák két hangtani attribútumának - balról jobbra való elemzés esetén - más funkciója van. A jobboldali magára a morfémaalakra vonatkozik, míg a baloldali arra vonatkozik, hogy milyen fonológiai tulajdonságú morfémához csatlakozhat. Ezeket az attribútumokat a szavak, illetve a morfémák kellő osztályozása alapján generálnánk egyszerű szabályok alapján. Pl. *bolond-já-t* morfémasorozatban a *-já* jobboldali fonológiai attribútuma megfelel a *-t* baloldali attribútumának. Ez abból adódik, hogy a *-já* morfémaalak fonológiailag abba a csoportba tartozik, amelyik magánhangzó nyúlást szenvedett, a -t pedig ezt előidézi, valamint a *-t* független a tőle balra álló morfémák hangrendjétől.

A szótövek és más morfémák összes előfordulási alakját generatív úton elő lehet állítani a szavak és toldalékok osztályozása alapján. Az előállított alakok tartalmazhatják a fonológiai illeszkedés kódját is. (A kézi előállítás biztos hibaforrást jelent.) Az osztályozás nem történhet teljesen automatikusan, mivel a magyar szavak alakjából nem lehet biztosan következtetni azok fonológiai, ragozási típusára.

A felszíni alakok generálásával egy időben a szótövek és egyéb morfémaalakok karakterláncát megtoldjuk a morfoszintakszis és a fonológiai illeszkedési kódokkal, tehát a morfoszintakszis szabályait is szomszédsági relációkra vezetjük vissza. Az így kapott szószedet véges sok szótári tétele jól elemezhető véges automatával. Az automata

egy menetben egy morfémát vágna le a szóalakról. Ha van még maradék a morfémaalak levágása után, akkor az utoljára levágott morféma által meghatározott osztályozási kódot hozzáfűzve a maradék alakra (a relatív tő után), az így generált karakterláncot elemezzük tovább.

Jelenleg nem áll rendelkezésünkre sem a szavak toldalékolási paradigmaszótára, sem a fonológiai osztályozás pontos elve, de nyelvészeknél létezik ilyen jellegű feldolgozás (Elekfi kézirata, debreceni és budapesti nyelvészek munkái). A jelenleg működő elemzőnk többértelmű elemzéseknek közel fele téves szóalak-generálásra vezethető vissza. Becslésünk szerint a tervezett módszerrel a jelenlegi rossz elemzéseknek 80-90 százaléka kiküszöbölhető, valamint sok esetben egyértelműsítene többértelmű toldalékalakokat. (A százszázalékos biztonság a magyar morfológia összetettsége és esetleges szabályai miatt amúgy is illuzórikus követelmény. A hangrendi és egyéb ragozási szabályok egyébként sem egészen egyértelműek: *mászom-mászok*, *mondta-mondotta*, *tősza-vak-tőszók*, *adapter-ok, adapter-ek* ...)

Egy lehetséges kivitelezés

Ahhoz, hogy a fenti modellt számítógépen megvalósíthassuk, többirányú munkát kell elvégezni. Ezek egy része elméleti nyelvészeti, másik része viszont programtervezési és megvalósítási jellegű. A tennivalók a következők:

1./a. Megadjuk a morfémaosztályokat (szófaji felosztás).

1./b. Megadjuk a morfoszintakszis gráfját.

2./a. Leírjuk a tőváltozási szabályokat.

2./b. Leírjuk, hogy milyen fonológiai osztályokhoz milyen tőváltozások és milyen illeszkedések tartoznak.

3./a. Egy segédprogrammal elkészítjük a bázisszótárat, mely tartalmazza a szavak alapalakját (igéknél a szó alanyi ragozás egyes szám harmadik személyű, névszóknál jeltelen alanyesetű alakját), szófaját és hangtani osztályozási kódját, esetleges rendhagyó, nem szabályosan torzult alakjait. Szerepelhetne benne ezen kívül szintaktikai és szemantikai információ is, amit nem a morfológia, hanem egy későbbi elemzési fázis használna fel (pl. a szó vonzatai). Ezeket a morfológia szempontjából másodlagos információnak nevezzük. A segédprogram feladata a szavak fonológiai osztályának meghatározása. Ezt csak interaktív módon lehet megtenni, mivel a szavak alakjából csupán valószínűsíteni lehet azok fonológiai tulajdonságát, de biztosra meghatározni nem lehet.

3./b. Hasonlóan módon készítjük el a toldaléktárat, mely tartalmazza a toldalékok illeszkedési szabályait, fonológiai osztályozását.

A toldaléktár tartalmazná az összes toldalékot, megadva a toldalék karakteres alakját, azt, hogy a toldalék milyen osztályba vagy osztályokba esik, melyik osztályban mi a szerepe/jelentése (pl. milyen módú, idejű stb. igerag) és az így kapott szó milyen toldalékolási osztályba kerül.

Bízunk benne, hogy ennek a munkának egy jelentős részét Elekfi professzor már elvégezte, de számos nyitott kérdés is van. Kérdés például, hogy a jelek, de különösen a képzők nem hoznak-e létre új toldalékolási osztályokat.

Egy másik, programozástechnikai kérdés, hogy hány és mennyire eltérő ragozási osztály van, milyen az eltérések jellege. Például érdemes-e egy ige összes ragját egy osztálynak venni, vagy érdemes szétválasztani alanyi és tárgyas, jelen és múlt idejű ragosztályokra, mert így kevesebb osztályt kapunk. Vajon, ha egy szó többféleképpen is ragozódik (ami nem ritka), akkor olyan osztályt kell-e hozzárendelni, amiben alternatívák vannak, vagy több osztályt kell hozzárendelni. Ezek mind olyan technikai kérdések, amit látatlanban nem tudunk megítélni.

Az Apreszján által leírt (orosz) morfológia a nyelvészeknél szokásos variánsos leírásra támaszkodik. Nála az 1b. igeragozás azt jelenti, hogy az ige az 1. csoportban leírt ragokat kapja, kivéve bizonyos alakokat, amikor valami mást, amit a b pontban adunk meg. Amennyire ismerjük, az Elekfi-féle leírás is hasonló. Kérdés, hogy ezt az ábrázolást tudjuk-e hatékonyan implementálni, vagy valamikor át kell térni egy, terjengősebb, de egyöntetűbb ábrázolásra az elemzés meggyorsítás érdekében.

A toldaléktárnak és a toldalékolási osztályoknak a létrehozása nagy méretű feladat, elvégzéséhez fel kell használni a létező nyelvészeti eredményeket, különösen a számítógépes szótárakat, de ezen kívül segédprogramokra is szükség van, amelyek szövegek feldolgozásával segítik az osztályok megállapítását és a szavak osztályba sorolását.

4. A bázisszótárak alapján állítjuk elő a morfológiai nyers-szótárat, mely minden lehetséges morfémaalakot tartalmaz, kiegészítve az említett jobb- és baloldali morfoszintaktikai és fonológiai illeszkedési kódokkal. Ez a szótár az eredetinél háromszor-ötször több tételt tartalmazna az alakváltozatok miatt.

5. A morfológiai nyers szótárat felhasználva állítjuk elő az egy morfémát levágó véges automata optimális alakját. (Az automata interpretációját később kifejtjük.)

6. Megalkotjuk a morfémákat dekódoló algoritmust.

7. A dekódolt - morfémánként különböző – kódokhoz hozzárendelem a szófaji és a másodlagos információkat tartalmazó másodlagos szótár tételeit.

A morfológiai elemző az így létrehozott véges automata segítségével gyorsan megtalálja a szóalak lehetséges felbontásait, majd az elemzés menetéből adódó hash-elés segítségével kikódolja a másodlagos szótár megfelelő elemeit.

A módszer előnye, hogy a felhasználás vagy a javítások céljából végrehajtandó módosításokat aránylag jól ellenőrizhető anyagokon és gépi segédlettel kell végrehajtani. Várhatóan a szótár és a toldaléktár szorul leggyakrabban módosításra.

Tehát a **bázisszótár** nemcsak szótöveket tartalmaz, hanem azok szófaji és hangtani besorolását is. Érdemes külön elemként felvenni a szótárba az olyan szóösszetételeket, képzett szavakat is, melyek értelme nem vezethető le a szóalkotás módjából. A teljesen rendhagyó módon toldalékolt alakoknak is szerepelnie kell a szótárban. Például a következő formában:

szóalak, szófaj, fonológiai osztály, rendhagyó alak, egyéb ..
pehely, főnév, 12,4: *pelyh*, ...

A 12-es fonológiai kód azt jelentheti, hogy magas hangrendű hangzókieséses, az azt követő rész pedig azt jelenti, hogy a hangzókieséses alakja nem *pehly*, hanem *pelyh*.

Az ebből generálandó nyersszótári tételek a következők:

bal illeszkedési kód, alak, tételszám, illeszkedési osztály
2, 0, *pelyh*, 56724, 53,14
2, 0, *pehely*, 56724, 53,7

Mindkét tétel bal illeszkedési kódja megengedi, hogy megelőzze egy tőváltozás nélküli névszó, függetlenül annak hangrendjétől. A jobboldalt szereplő illeszkedési osztály arra utal, hogy magas hangrendű névszóról van szó, mely egyike hangzókieséses, a másikat pedig nem követheti hangzókiesést okozó toldalék. A középső tételszám mindkettőnél a másodlagos információkat tartalmazó szótár egyazon tételének száma.

A **toldaléktárnál**, mivel a toldalékok száma kisebb, esetleg megengedhető az összes felszíni alak felsorolása az illeszkedési kódokkal. Ez egyrészt azt jelenti, hogy az őt megelőző relatív szótő milyen esetleges tőváltozást szenvedhetett, milyen hangtani osztályba sorolható a relatív szótő, másrészt milyen változást szenvedett a toldalék, s a toldalékolt szóalak milyen hangtani osztályba került. Elképzelhető az a megoldás is, hogy a toldalékalakokat is osztályozás segítségével állítjuk elő alapalakjukból.

Az automata implementációs modellje:

A generálandó automata egy morféma levágására képes. Az elemzendő szót kiegészítjük a morfoszintakszis és a hangtani besorolás két kezdeti attribútumkódjával (ragos szó tőváltozás nélkül ...), majd az elemző levágja a szélső morfémát. A vizsgálat az automata úgynevezett végállapotaiban történik. Ha az automata nem képes levágni egy morfémát sem, vagy úgy vág le toldalékot, hogy nem marad a szótőből semmi, akkor az adott elemzési ág téves. Ha az utoljára levágott morféma egy szótő, és a szóalakból nem marad semmi, akkor egy helyes felbontását kapjuk a szónak. Más esetekben az elemzést úgy kell folytatni, hogy a maradék szóalakot kiegészítjük a végállapotban kódolt baloldali hangtani és morfoszintaktikai kódokkal, s így próbálkozunk újabb morféma levágásával.

Lényeges még az egy morfémát levágó automata implementációja is. Egyrészt az automatát optimalizálni kell, másrészt a véges automaták azon modelljét kell használni, mely egy állapotból a másikba nemcsak egy karakter hatására, hanem egy (nem üres) karakterlánc beolvasására is átmegy. (Egy állapotból kiinduló karakterláncok kezdetei nem lehetnek azonosak.) Az automata ilyen jellegű ábrázolásával és pár technikai trükkel várhatóan a morfémaelemzés több tízezer szótő esetén is elfogadható méretűre csökkenthető. További helyspóroláshoz vezet az is, ha a morfémalevágó automatát több automata metszeteként (párhuzamos futtatásával) valósítjuk meg. Ezekből egy a morfémaalakoknak, egy a morfoszintakszisnak és egy vagy több részautomata a fonológiai illeszkedésnek felel meg.

A morfémák dekódolására csak akkor van szükség, ha a levágó egy helyes megoldást talált. Az elemzés alatt a holtvágányok száma nagyobb, mint a találatoké. Ilyenkor az automata elágazási kódjaiból megfelelő hash-elés készíthető, aminek eredménye a másodlagos szótár elérési pontja.

Kivételkezelés

Léteznek olyan szóalakok, melyek a fent vázolt módszerrel közvetlenül nem elemezhetőek, illetve nem érdemes a rendhagyó képzést szabályba foglalni. Ezek egyik része a kivételes szótőalakot tartalmazók, másik része a teljesen rendhagyó alakúak. Az előbbi jól kezelhető, ha a bázisszótárban feltüntetjük ezeket a töveket jelölve azt is, hogy milyen esetben fordulhatnak elő: pl. *ló - lov*. Ily módon a legenerált nyers szótárban már a megfelelő tételek szerepelhetnek. A második esetben a szóalak egy külön szótári tételt alkotna a bázisszótárban, ahol a morfológiailag kifejtett formát is le kell kódolni: pl. *gyere = jön + j*, *több = sok + abb*.

Az új módszer előnyei

Mivel a jelenlegi morfológiai elemző viszonylag jó hatásfokkal működik, felmerül a kérdés, miért van szükség egy új elemezőre, amelynek elkészítése meglehetősen nagy feladat. Az alapvető indok tudományos. Így érezzük, hogy lehetőség van egy olyan morfológiai elemző létrehozására, amely elvileg tiszta, tudományosan megalapozott; és nem többé-kevésbé beváló heurisztikákra épül. Lenne azonban a munkának gyakorlati haszna is, melyet a következő pontokban foglaljuk össze:

1. A többértelműségek és hibás elemzések csökkenése.
2. A szavak és egyéb toldalékok egységes kezelése.
3. Az elemzési sebesség növelése egyrészt kevesebb "holt ág", másrészt a szótári keresés automata elvű implementálása következtében.
4. A gyakorlatban alkalmazható szótári méret nagyságrendi növekedése.
5. Más, morfológiával összefüggő feladat megoldására is mód nyílik: nyelvhelyességi ellenőrző, generatív modell implementálása ...

Megjegyzés

Morfológiai elemző programból már korábban is elég sok készült hazánkban. Ezekről jó áttekintést ad Prószéky Gábor: Számítógépes Nyelvészet című könyvének 15. fejezete. Jelenleg is legalább 4 helyen dolgoznak ilyen munkán (ezekről kevesebbet lehet tudni, mert üzleti céllal készülnek).

Az elméleti munkák közül Papp Ferenc, Elekfi László, Kornai András eredményei látszanak hasznosíthatónak.

FÜGGELÉK

I. A végződések kódolása

A morfológiai elemzés eredményeként született szövegfájl a következő szintaktikájú kódot adja ki:

```
ELEMZÉS => {ELEM}
ELEM => TALÁLT / IRÁSJEL / NEMTALÁLT
NEMTALÁLT => forrásszó <:*>
IRÁSJEL => forrásjel <:> tőkód
TALÁLT => forrásszó <cr sp> MEGOLDÁS {<cr sp> MEGOLDÁS}
MEGOLDÁS => TŐKÓDLISTA TOLDALÉKLISTA
TŐKÓDLISTA => tő <|> tőkód {<|> tőkód}
TOLDALÉKLISTA => ő <:> TOLDALÉK {<_>TOLDALÉK}]
TOLDALÉK => toldalékkód
```

A forrásszó és forrásjel a szó illetve írásjel eredeti formája.

A tő a szótárban megtalált szótőforma.

A tőkód a szótőhöz rendelt kód (10000 - 29999 lásd **szótár**), illetve írásjel esetén egy 30000 feletti szám (pontosabban 30000 + a jel kódja).

A toldalékkód egy 10000-nél kisebb szám, mely ezresei a toldalékosztályt, az ezernél kisebb része az osztályon belüli felosztást jelentik:

```
    0- 999 esetrag
1000-1999 igerag
2000-2999 többes szám+birtokjel
3000-3999 birtokos jel
4000-4999 melléknévfokozó
5000-5999 névszóból igét képző
6000-6999 névszóból névszót képző
7000-7999 igéből igét képző
8000-8999 igéből névszót képző
9000-     -e kérdőszócska
```

Az osztályon belüli további jegyeket a melléklet tartalmazza, de a jegyek értelmezése egyes osztályokban jegyenként meghatározhatók. (Jelenleg a toldalékkódot megelőzi a toldalék eredeti formája nyomkövetési okokból.)

Néhány toldaléknál algoritmikusan dekódolhatóak a további számjegyek, míg a többi kód a további táblázatból jól követhető.

A végződésalakokat megelőző kódszám a fonológiai besorolását jelenti. A sorokat bezáró kódszám a morfémák korábban említett kódszámát tartalmazza.

igeragok:

Az utolsó három jegy értelmezése:

```
                  MTS
      ┌─────────────┘│└──────────────┐
     Mód        tárgyasság        szám,személy
  0=kijentő      0=alanyi          0=indefinit
  1=múlt         1=tárgyas     1-3=egyes szám  1-3. szem.
  2=felszólító   2=sptárgy     4-6=többes szám 1-3. szem.
  3=feltételes    (1->2)
  4=ni...       3=indefinit
  5=-va,-ve
  6=-ván,-vén
  7=kijel.+felsz.
```

Az ettől eltérő kódok egyéb többértelműséget jelentenek:

```
1808=tek 1005 vagy 1106            1039=ik 1003 vagy 1016
1336=nék 1301 vagy 1316
23:ek,          11:ok,ök,          15:vok, 14:vök :001
                11:ol,el,öl,sz,asz,esz             :002
26:-                                               :003
          23:unk,ünk,              15:vunk,14:vünk:004
12:tek                                  :808=005 vagy 106
24:tok,12:tök,26:etek,    11:otok,ötök             :005
24:nak,nek,23:anak,enek                            :006
          23:em,          11:om,öm, 15:vom, 14:vöm :031
24:lak,lek,23:alak,26:elek                         :021
          23:ed,          11:od,öd, 15:vod, 14:vöd :012
          11:i,                     14:vi          :013
28:juk,jük,17:suk,sük  :714,014,214 a szótővégtől függően
          11:itek,        14:vitek                 :015
          11:ik,          14:vik              :039,016,003
24:tál, 12:tél, 30:ttál, 31:ttél, 11:ettél, ottál,öttél        :102
22:t,          29:tt,     11:ett,   ott,ött,  15:vott,14:vett :103
24:tunk, 12:tünk, 30:ttunk, 31:ttünk, 11:ettünk, ottunk, öttünk :104
24:tatok,12:tetek,30:ttatok,31:ttetek,11:ettetek,ottatok,öttetek:105
24:tak,          30:ttak,  31:ttek,  11:ettek,  ottak,  öttek  :106
24:tam,  12:tem,  30:ttam,  31:ttem,  11:ettem,  ottam,  öttem  :131
24:tad,  12:ted,  30:ttad,  31:tted,  11:etted,  ottad,  ötted  :112
24:talak,12:telek,30:ttalak,31:ttelek,11:ettelek,ottalak,öttelek:121
24:ta,   12:te,   30:tta,   31:tte,   11:ette,   otta,   ötte   :113
24:tuk,  12:tük,  30:ttuk,  31:ttük,  11:ettük,  ottuk,  öttük  :114
24:tátok,12:tétek,30:ttátok,31:ttétek,11:ettétek,ottátok,öttétek:115
24:ták,  12:ték,  30:ttág,  31:ttég,  11:ettég,  ottág,  öttég  :116
27:ak, ek,    18:jak, jek,    17:sak,sek,     20:gyak,gyek      :201
27:ál, él,    18:jál, jél,j,  17:sál, sél,s,  20:gyál, gyél,gy  :202
27:on,en,ön,  18:jon,jen,jön, 17:son,sen,sön, 20:gyon,gyen      :203
              18:junk, jünk,  17:sunk, sünk,  20:gyunk, gyünk   :204
27:atok,etek, 18:jatok,jetek, 17:satok,setek, 20:gyatok,gyetek  :205
              18:janak,jenek, 17:sanak,senek, 20:gyanak,gyenek  :206
27:am, em,    18:jam, jem,    17:sam, sem,    20:gyam, gyem     :231
27:alak,elek, 18:jalak,jelek, 17:salak,selek, 20:gyalak,gyelek  :221
27:ad,ed, 21:d, 18:jad,jed,   17:sad,sed,16:sd 20:gyad,gyed,dd  :212
              28:ja,          17:sa                             :713
27:a, e,      18: je,         17: se,         20:gya, gye       :213
27:uk, ük,                                    20:gyuk, gyük     :214
```

```
            28:játok,       17:sátok                          :715
27:átok,étek, 18:jétek,     17:sétek,      20:gyátok,gyétek   :215
            28:ják,         17:sák                            :716
27:ák, ék,    18:jék,       17:sék,        20:gyák, gyék      :216
                           11:anék           :301
  24:nék,                 11:enék           :336=301 vagy 316
  24:nál, nél,            11:anál,enél      :302
  24:na, ne,              11:ana,ene        :303
  24:nánk, nénk,          11:ánk,enénk      :334
  24:nátok,nétek,         11:anátok,enétek  :335
  24:nának,nének,         11:ának,enének    :306
  24:nám, ném,            11:anám,eném      :331
  24:nálak,nélek,         11:análak,enélek  :321
  24:nád, néd,            11:anád,enéd      :312
  24:ná, né,              11:aná,ené        :313
  24:nák,                 11:anák           :316
  24:ni,                  11:ani,eni        :430
  24:nom, nem,nöm,        11:anom,enem      :431
  24:nod, ned,nöd,        11:anod,ened      :432
  24:nia, nie,            11:ania,enie      :433
  24:nunk, nünk,          11:anunk,enünk    :434
  24:notok,netek,nötök,11:anotok,enetek     :435
  24:niuk, niük,          11:aniuk,eniük    :436
  24:va,ve   :530
  24:ván,vén :630
```

esetrag:

```
        1:-    :0
        1:ként :50
        1:kor  :60
      2:ban, ben       :20
      2:ba, be         :21
      2:ból, ből       :22
      2:ra, re         :31
      2:ról, ről       :32
      2:nál, nél       :40
      2:hoz, hez,höz   :41
      2:tól, től       :42
      2:ért            :51
      2:ig             :61
      2:szor,ször,szer :64
      2:nak, nek       :71
      2:ul, ül         :72
 13:t,             25:at,et, 5:ot,öt, 9:vat,vet :11
    10:n,          25:en                :38=30 vagy 34
                   25:an :34
                              5:on,ön,9:von,vön  :30
    19:á,é, 10:vá,vé                           :35
            10:nként, 5:anként,enként,onként,önként :54
    19:al,el, 7:val,vel                        :70
```

ige-ige képző:

```
 24:hat,het :1
( nem kezeljük: tat,tet,vat,vet,at,et :2 )
 28:gat,get,11:ogat,öget,eget,15:vogat,14:vöget,veget :3
```

birtokos jel:

```
1=egy birtok 2=több birtok
                  1:-  :0
                  2:é  :1
                  2:éi :2
```

névszó-névszó képző:

```
 10:s, 25:as,es, 5: os, ös, 9:vas,ves :2
2:tlan,tlen,6:talan,telen,5:atlan,etlen,9:vatlan,vetlen:3
                          5:adik, edik, odik,ödik      :4
```

fokozó:

```
 1:-                                    :0
 10:bb,5:abb,ebb,obb,öbb,9:vebb,vabb :1
```

felső- illetve túlzófok esetén a kód eggyel több, mint a leg-ek száma.

névszó-ige képző:

```
 5:ít                              :1
 10:z,5:oz,öz,25:az,ez,9:vaz,vez :2
```

ige-névszó képző:

```
           11:ó, ő, 15:vó, 14:vő                  :1
 24:t,29:tt,11:ott, ett,ött,15:vott,  14:vett,vött :2
           11:atlan,etlen, 15:vatlan,14:vetlen    :3
           11:andó, endő,  15:vandó, 14:vendő     :5
           11:ás, és,      15:vás, 14:vés         :4
```

többes szám, birtokjel:

Többes szám + birtokjel: Az utolsó jegy kódolása megegyezik az igerag utolsó jegyével (a birtokosra vonatkozó szám és személy) míg az utolsó előtti számjegy a birtok (birtokos hiányában a megnevezett névszó) számát jelöli: 1=egyes szám
2=többes szám

```
1:- :0
10:k,  25:ak,   5:ek, ok, ök,          9:vak, vek     :20
10:m,  25:am,   5:em, om, öm,          9:vam, vem     :11
10:d,  25:ad,   5:ed, od, öd,          9:vad, ved     :12
       25:a,    5:e 32:ja, je,         9:va, ve       :13
10:nk, 25:unk,  5:ünk,                 9:vunk, vünk   :14
10:tok,tek,tök,25:atok,5:etek,otok,ötök,9:vatok,vetek :15
       25:uk,   5:ük,  32:juk,jük,     9:vuk, vük     :16
10:im, 25:aim,  5:eim, 32:jaim, jeim,  9:vaim, veim   :21
10:id, 25:aid,  5:eid, 32:jaid, jeid,  9:vaid, veid   :22
10:i,  25:ai,   5:ei,  32:jai,  jei,   9:vai, vei     :23
       25:aink, 5:eink,32:jaink,jeink, 9:vaink,veink  :24
10:itek,25:aitok,5:eitek,32:jaitok,jeitek,9:vaitok,veitek:25
10:ik, 25:aik,  5:eik, 32:jaik, jeik,  9:vaik, veik   :26
```

II. A toldalékok 32 fonológiai osztálya:

A következő táblázat a fonológiai osztályokhoz rendelt illeszkedési tőveget és tőváltozást tartalmazza.

Az egy szóban felhasználható karakterkészlet nem más, mint a PC-n elérhető latin betűk, a számjegyek, valamint néhány írásjel: aáâàäbcçdeéêèëfghiíîìjklmnñoóöőpqrstuúüűvwxyz
AÁÂÀÄBCÇDEÉÊÈËFGHIÍÎÌJKLMNÑOÓÖŐPQRSTUÚÜŰVWXYZ
0123456789'".-

														Nem szótővégi magánhangzó-kiesés
														Nem szótővégi magánhangzó-rövidülés
													Szóvégi u, ü kiesés	
												Igető végi sz kiesés		
											Szótővégi magánhangzó-rövidülés			
										Szótővégi ó-a, ő-e rövidülés				
									Szótővégi a, e kiesés					
								Szótővégi a, e nyúlás						
							Névszói hasonulás							
						Igei t-s átalakulás								
					Igei hasonulás									
				Változatlan										
	A toldalékot megelőző betű													
1	A teljes ABC	2x												
2	ábcdéfghiíjklmnoóöőpqrstuúüűvwxyz'-	2x				x								
3	A teljes ABC	2x												x
4	bdfghjklmnpqrvy-	2x												
5	bcdfghjklmnpqrstvwxyz'-	2x											x	x
6	bcdfghjklmnpqrstvwxyz'.-	2x												
7	áéiíóőúű'-	x				x								
8	éiíóőúű'-	x				x						4x		
9	aábcdeéfghiíjklmnoóöőpqrstuúüűvwxyz'.-					x			x					
10	áéhjlóőquúüűy'-	x				x								
11	bdfghjklmnprstvyz	2x												x
12	bdfghjklmnprstvyz	2x								x				
13	ábcdéfghiíjklmnoóöőpqrsuúüűvwxyz'-	2x				x								
14	eégiíklöőüűE	2x							x	3x				
15	eégiíklöőüű	2x							x	3x				
16	st	2x												
17	st	2x		x										
18	bcfghiíjklmnoóöőpruúüűvyE	2x												
19	bcdfghjklmnpqrstvwxyz'.-				2x									
20	eéhíjoóöőuúüűyE									3x				
21	bdeéfghiíjklmnoóöőprstuúüűvyzE-	2x		x										
22	bdfghjklmnprsvyz'	2x												
23	bdeéfghiíjklmnoóöőprstuúüűvyzE	2x	x											x
24	bdeéfghiíjklmnoóöőprstuúüűvyzE	2x								3x				
25	A teljes ABC	2x											x	x
26	bdeéfghiíjklmnoóöőprstuúüűvyzE	2x	x											
27	yz		x											
28	bdéfghiíjklmnoóöőprstuúüűyE	2x												
29	eéhíjóöőuúüűyE	2x							x	3x				
30	éiíóőúű	x							x	3x				
31	áéiíóőúűE	x							x	3x				
32	A teljes ABC	2x				x	x	x						
	A tőváltozás feltétele		zz- ddz- ssz- ggy-	s-	hosszú mással- hangzó	á- é-	mással- hangzó	a- e-	rövid magán- hangzó	nem t-t	mással- hangzó	hosszú magán- hangzó	mással- hangzó+ m,z,g, r,k,m	

1x esetleg kombinálva igei sz kieséssel
2x esetleg kombinálva harmadik mássalhangzó-kieséssel
3x esetleg kombinálva magánhangzónyúlással
4x esetleg hosszú ú, ű kiesés

III. A magyar szavak morfológiai-szintaktikai felosztása

A szótárban szereplő kódszám első három jegye a szófajt határozza meg, a tízesek a szó vonzatára utalnak, az egyesek pedig a morfológiai elemzést megkönnyítő kód.

0xxxx Ragozhatatlanok

01xxx Határozószók
010xx Vegyes határozók (ami nem sorolható a továbbiakba)
011xx Hol kérdésre válaszoló helyhatározók
012xx Hova kérdésre válaszoló helyhatározók
013xx Honnan kérdésre válaszoló helyhatározók
014xx Időhatározók
015xx Módféle határozók

02xxx Igekötők - Igekötő minden olyan szócska amelyet, ha az ige előtt áll, egybeírunk vele, ha utána, külön.
03xxx Névelők
038xx határozott névelő
039xx határozatlan névelő

04xxx Kötőszók További osztályozása szükséges aszerint, hogy szavakat vagy mondatokat köt össze; ha mondatokat alá vagy mellérendelő mondatokat. Megjelölendők továbbá a kételemű kötőszavak.
040xx Mondat elején álló mondatkötőszók: *bár*

041xx A vesszőtől függ, hogy mondatrész- vagy mondatkötőszók: *és, vagy, meg*
042xx Nem feltétlenül a mondat elején álló mondatkötőszó: *azonban*
043xx Mondat-, mondatrészkötőszó-pár első fele: *nemcsak*
044xx Mondat és mondatrész kötőszó: *tehát, illetve*
045xx Mondat-, mondatrészkötőszó-pár második fele: *hanem*
047xx Ismétlő mondatrész- és mondatkötőszó: *akár*, *vagy*, *se*, *sem*
049xx Utáni kötőszó: *is, se, sem*

05xxx Módosító szavak
051xx Igei módosító szavak. Az ige értelmét módosítják: *hadd* és *hagy* óhajtás, *ne, se, nem* és *sem* tagadás, *nemigen* és *alig* gyengítés. A szó előtt áll.
052xx Melléknévi módosító szavak. A melléknév, melléknév jellegű számnév, melléknévből képzett határozószó jelentését erősíti, gyengíti, tagadja stb. A szó előtt áll.

053xx Számnévi módosító szavak. A konkrét számok értékét bizonytalanítja; a szám előtt áll.

060xx Igeszerű szavak
061xx Nem igei segédigék. Nemigen 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.
062xx Ragozhatatlan főnévi névmások. A személyes névmások alany és tárgyesetei. A többi esetet lásd a **névutók ragozásá**nál.

07xxx Egyéb szavak

08xxx Névutószerű szavak. A névutók kaphatnak formális birtokos ragot, ezek nem.
080xx Vegyes névutószerű szavak
081xx Hol kérdésre válaszoló szavak
082xx Hova kérdésre válaszoló szavak
083xx Honnan kérésre válaszoló szavak
084xx Mikor kérdésre válaszoló szavak
085xx Hogyan kérdésre válaszoló szavak

1xxxx Ige - ami igeragokat kaphat.
101xx Létige a *van, lesz, megvan* alakjai. Igeragozásuk speciális. A mondatban két alanyesetű névszói kifejezést hozhatnak relációba.
11xxx Segédigék. A segédige ragozódik, az ige (ragozatlan) főnévi igenévvel áll.
12xxx Féligei 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*.
190xx Főnévi igenév, ragozatlan
191xx Főnévi igenév, személyragozott
192xx Határozói igenév

2xxxx Névszó
201xx 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*** *lengyel vendégmunkás,* ***Kiss Péter*** *lakatos,* ***Pista*** *bácsi,* ***Vörös Október*** *gépgyár,* ***március*** *hónap* stb.
202xx Mértékegységek

21xxx Főnév

211xx 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 virágcserép, lengyel vendégmunkás, csőszerelő szakmunkás, katolikus pap, elnök elvtárs* stb.

212xx Teljesen főnév jellegű névmás. Teljesen úgy ragozódnak, mint a főnevek; birtokos és esetragot is kaphatnak.

213xx Névutóhoz hasonlóan ragozódó + nyomatékosító 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. Ezenkívül az *ez, az, ugyanez* stb. névmások, és ragozott alakjai.

214xx Birtokosan nem ragozódó névmások. Birtokjelet nem kaphatnak, de birtokos jelet, és esetragot kaphatnak. Pl. *egymás, egymását, egymást* stb.

22xxx Melléknév. Állhat tulajdonságjelzőként névszói szerkezetben, nem-igei segédigeként, nem igei állítmányként. Megkülönböztetendők azok, amelyek vonzattal rendelkeznek. Fokozhatók.

221xx Melléknévi igenév

222xx Melléknév jellegű névmás

223xx Melléknévi névutó

224xx Pozícionáló melléknév. (A felső foka *-bb* nélkül képzendő.)

23xxx Számnév

230xx Konkrét szám. Egytől tízig betűvel és számmal, továbbá *száz, ezer, millió* betűvel is, a többi csak számmal, kötőjellel toldalékolva.

231xx 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.

232xx Névmás jellegű számnév. Nem mutat melléknévi tulajdonságokat, a számjelző helyén állhat. Számhatározó képezhető belőle. Számnévi módosítót kaphat. Nagyrészük számnévi névmás.

233xx Kvantor jellegű. A névszói szerkezetben a számjelző előtt az un. kvantor pozícióban áll. Jöhet utána számjelző is. Számhatározó nem képezhető belőle.

234xx Sorszámnév

235xx Szelektor. Olyan kvantor, ami előtt kell a határozott névelő.

24xxx Névutó. A személyes névmások ragozott illetve névutóval ellátott alakját, a névutó+személyrag alakban képezzük. 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. Megkülönböztetendők azok a névutók, amelyek ragot vonzanak.
240xx Vegyes névutók
241xx Hol kérdésre válaszoló névutók
242xx Hova kérdésre válaszoló névutók
243xx Honnan kérésre válaszoló névutók
244xx Mikor kérdésre válaszoló névutók
245xx Hogyan kérdésre válaszoló névutók

Egyéb jelölések:

xxxx1 Nem szótári alak, csak ragozott formában szerepelhet
xxxx2 Ragozott alak. A tő torzulása miatt került a szótárba
xxxx4 Nem kell tovább bontani, csak ez a forma szerepel a szótárban.
xxx1x Kötelező vonzata van
xxx2x Opcionális vonzata van
xxx9x Úgynevezett külső vonzata van. A vonzat a szó után áll, vagy töltelék melléknévi igenévvel illeszthető a névszói kifejezésbe. Pl.: *házasság Péterrel*, *Péterrel való házasság*, *Péterrel kötött házasság*

IV. Példa egy elemzésre:

```
A: A|3804
végződések: végződ|10005:8és8004_2ek2020
kódolása: kódol|10004:8ás8004_2a2013
.: 30046
A: A|3804
morfológiai: morfológiai|22004
elemzés: elemez|10004:8és8004
eredményeként: eredmény|21004:2e2013_0ként50
született: szület|10005:8ett8002
          szület|10005:1ett1103
text:*
fájl:*
a: a|3804
következő: következ|10005:8ő8001
kódokat: kód|21004:2ok2020_0at11
adja: ad|10004:1ja1713
ki: ki|21204
:: 30058
```

A SZÓFAJOK

A nyelv szavait több különböző szempont szerint lehet osztályozni. A három legfontosabb felosztás:

- hogyan viselkedik a szó a toldalékolás szempontjából,
- a szó (ragozott és ragozatlan) alakjai hol állhatnak a mondatban,
- mit jelent a szó.

Noha ez a három dolog szorosan összefügg, még sincs egyértelmű (algoritmizálható) kapcsolat a három különböző felosztás között. Ezért mi úgy képzeljük, hogy egy olyan szótárat kell létrehozni, ahol minden egyes szóhoz 3 mező kapcsolódik: az első a szó toldalékolási tulajdonságait írja le, a második a szó mondatbeli szerepét és kapcsolatait, a harmadik szemantikus jegyekkel, és valamilyen adottnak tekintett fogalmakra való visszavezetéssel írja le a szó jelentését.

A korábbi munkánkban ez a dolog nem így volt felosztva. Volt szófaji felosztás, amelyben morfológiai és szintaktikai meggondolások keveredtek, és volt egy vonzatszótár, amelyben szintaktikai és szemantikai tulajdonságok keveredtek. A szemantikus jegyekkel való jellemzés és az adott fogalmakra való visszavezetés terv maradt, kivitelezésre nem került.

A toldalékoláshoz meg kell adni a szó ragozási osztályát, amely megmondja, hogy a különböző alakú jelek és ragok közül melyeket vehet fel az adott szó. Egyelőre bízunk abban, hogy a ragozási osztályból az is következik majd, hogy ha a szóra egy reguláris képzőt alkalmazunk, akkor a képzett szó milyen ragozási osztályba kerül. (Ha ez nem így lenne, az beláthatatlan komplikációkra vezetne, mert nemcsak minden szó minden képzőjére kell megadni a ragozási osztályt, hanem a kettő, három hosszúságú képzőláncokra is.)

A szó mondatbeli szerepének meghatározásához meg kell adni a szófaját, és meg kell adni a vonzatait. Mi ebben a fejezetben a szófajjal fogunk foglalkozni. Szófaj alatt azt értjük, hogy milyen pozíciókban állhat a szó a mondatban. Mivel itt nem csak a puszta szóra gondolunk, hanem annak ragozott formáira is, ezért kismértékben ki fogunk térni a végződésekre is. Hasonlóan néha szemantikai jegyek is szóba fognak kerülni. (A vonzatokkal több más fejezet foglalkozik.)

A szavakat alapvetően 3 csoportra osztjuk fel: igékre, névszókra és ragozhatatlanokra.

Igék

Az igék olyan szavak, amelyek egy igei reláció magját képezik, és argumentumok kapcsolódhatnak hozzájuk (alany, tárgy, határozók). Ha a mondat fő állítása ez az igei reláció (ami a mondatok kb. 60%-ára igaz), akkor az ige igeragot kap. Más szerepben (ha nem ez a mondat fő állítása, vagy csak egy része a főállításnak) az ige más végződéseket kap. Lehet főnévi igenév, ragozott főnévi igenév, határozói igenév, melléknévi igenév, *-ás/-és* végű főnév stb. Az ige argumentumai ilyenkor is megmaradnak, de kijelölésük módja megváltozik. (Lásd az **igenevek**ről szóló fejezetet.)

Tágabb értelemben az igékhez tartoznak a segédigék és a létigék is.

A létigék

Létigének nevezzük a *van, lesz* és *megvan* igéket. (A *megvan* a *van* ritkán használt befejezett párja.) Vannak olyan mondatok, amelyek lényegében egy névszói kifejezésből és egy létigéből állnak. Itt a mondat azt állítja, hogy az illető dolog létezik. (És a létigét mindig ki kell tenni.) Ezen kívül számos más konstrukció van, amiben létige állhat, de ezekben a létige egy segédszó, ami az adott reláció idejét és módját fejezi ki (néha az alanyát is). Jelen időben és kijelentő módban egyes szám harmadik személyben el is maradhat, vagy el kell maradnia.

Van egy piros almám.
Ennek a helyzetnek megvannak a maga előnyei és hátrányai.
(Egzisztencia-mondat)

A Tisza be van fagyva.
Meg voltam fázva.
Holnapra legyen megoldva a feladat!
(Határozói igeneves szenvedő mondat.)

Fáradt voltam.
Holnapra piros lesz az eper.
(Névszói állítmány idejét, módját, alanyát adja meg.)

A piacon voltam vásárolni.
(Ragozatlan igenévi állítmány idejét, módját, alanyát adja meg.)

Tilos volt bemennie.
Ha láttam volna az esetet, ...
(Időt illetve módot kifejező módosító szó. Személyragos ige vagy igenév mellett csak egyes szám 3. személyű lehet a létige.)

A létigék ragozása hiányos, a nem létező alakokban a *van* és *lesz* ige alakja egymást pótolják. Ahol mindkét alak megvan, ott nem feltétlenül helyettesíthetők egymással.

van (nincs)	-
-	*lesz*
volt	*lett*
volna	*lenne*
-	*legyen*
-	*lehet*
?vanni	*lenni*
-	*?lévén*
való	*levő*
???*volt*	???*lett* mint múlt idejű melléknévi igenév
-	*?leendő*
-	*?lehető* bár inkább *lehetséges*
-	*lehetetlen*

A létigének nincs vonzatkerete, nem tőle függ, hogy milyen határozók és egyéb mondatrészek kerülnek a mondatba. Nem tekintjük létigének a *van*-t és a *lesz*-t, amikor saját vonzataik vannak (ott+)*van*+valahol, *van*+valamiből, *lesz*+valamivé. Ezekben az esetekben ugyanolyan ige, mint az (ott+)*áll*+valahol, *áll*+valamiből, *válik*+valamivé. Csak akkor tekintjük létigének, amikor létezést fejez ki, vagy amikor nyelvtani segédszóként szerepel.

A létige tagadó alakja a *nincs, nincsen, nincsenek, sincs, sincsen, sincsenek* forma, ami csak az egzisztencia mondatokban és szenvedő mondatokban állhat. Ezeket olyan kivételes ragozhatatlan szavaknak tekintjük, amelyhez már a szótárba be van írva a megfejtés: *nincsen=nem_van*, *nincsenek=nem_van+nak* stb.

A segédigék I

A magyar nyelvben van néhány ige (felsorolásukat lásd **az egyszerű mondatok szerkezeté**ről szóló fejezetben), amelyek majdnem mindig egy (ragozatlan) főnévi igenévvel együtt állnak a mondatban, más saját vonzatuk nincs, a mondatot a főnévi igenév vonzatai alkotják, a két ige alanya azonos. Ezeket az igéket mi segédigének tartjuk, és úgy gondoljuk, hogy az igének valamilyen általánosabb értelemben vett módját adják meg. Erre utal az is, hogy ez a szerkezet gyakran helyettesíthető valamilyen végződéssel (*-hat/-het*, *-ható/-hető*, *-andó/-endő* stb.).

A segédigék ragozása hiányos, illetve más nyelvtani időben és módban más általánosított módot jelöl. Jelentését az is befolyásolja, ha kérdő vagy tagadó mondatban áll. Pl.:

Meg tudom mondani...
(Képesség.)
Meg tudná mondani?
(Udvarias felszólítás, de csak kérdő mondatban.)
Nem tudná megmondani?
(Udvarias felszólítás, de csak kérdő mondatban.)

(Megjegyzés: a feltételes mód kérdő mondatban általában is óhajt (udvarias felszólítást) fejez ki, de az alany képességéről itt már szó sincs.)

(Lásd **Az egyszerű mondatok szerkezete** című fejezetet is.)

Segédige II

Találtunk 4 olyan igét (*illik*, *kell*, *lehet*, *sikerül*), amelyek egészen másképp viselkednek, mint a többi ige, amikor ragozott vagy ragozatlan főnévi igenév áll mellette. Ezek az igék ilyenkor mindig egyes szám harmadik személyben állnak a mondatban a személyragot a főnévi igenév kapja, a logikai alany *–nak* ragot visel és ezzel kell a főnévi igenév személyragját egyeztetni.

Ezeket a szavak is olyan segédigének tekintjük, amelyek az ige általánosított módját fejezik ki. Pl.:

Sikerült megoldania a feladatot.

Neki sikerült megoldania a feladatot.

Neki sikerült megoldani a feladatot.

Nem tekintjük ezeket az igéket segédigének, amikor nem főnévi igenév mellett állnak. Pl.:

Péter illik Marihoz.

Lehettem volna katona is.

Igenevek

A mi szótárunkban igenevek csak akkor szerepelnek, ha az igenév képzése nem szabályos, és így kivételképpen bekerül a szótárba. A szintakszis természetesen ismeri az igeneveket mint szófajt, a morfológia viszont szétbontja a szót igére és képzőre. Ahhoz, hogy a szintaksziselemzés működni tudjon a morfológiailag elemzett szövegen, egy közbülső fázis, a posztmorfológia összerakja azt, amit a morfológia szétszedett.

A szavak összerakása tőből és képzőkből két lépésből áll: egyrészt meg kell határozni az összetett szó szófaját, másrészt meg kell állapítani a keletkező szó vonzatait, (továbbá, ha a szemantikával is foglakoznánk, a szemantikai jegyeit és az alapfogalmakra való visszavezetést is).

Ezt természetesen nemcsak az igeneveknél, hanem minden képzett szónál el kell végezni, többszörösen képzett szónál több lépcsőben.

Főnévi igenév

A főnévi igenév bizonyos mondatmintákban a fő reláció egyik összetevőjeként szokott szerepelni, más mondatokban az igei állítmány egyik, semmi mással össze nem cserélhető vonzata szokott lenni. (Azt, hogy milyen mondatmintákban szerepel főnévi igenév a fő reláció összetevőjeként, **Az egyszerű mondatok szerkezete** című fejezet írja le legrészletesebben.)

Határozói igenév

Vagy a szenvedő mondat állítmányának egyik összetevője, vagy hogyan kérdésre válaszoló szabad határozóként kerülhet a mondatba.

Melléknévi igenév

A melléknévi igenév teljes értékű melléknév. Tovább képezhető. Mint a melléknevek általában vagy jelzőként áll egy névszói kifejezésben vagy önállóan a mondat állítmánya lehet.

***-ás/-és* végű folyamatot leíró főnév**

Teljes értékű főnév. Az igenevek és az *-ás/-és* végű főnév részletes leírását lásd az **Igenevek** című fejezetben.

Névszók

A névszók azok a szófajok, amelyek névszói végződéseket kapnak. A névszói végződéseket többnyire azért kapják, mert egy névszói kifejezésben szerepelnek. A névszói kifejezésnek bármely része elmaradhat. Ilyenkor a ragok és jelek a megelőző komponenshez kapcsolódnak. Vannak azonban olyan esetek is szép számmal, amikor a szóban forgó szó formailag birtokjelet kap, ezzel azonban a személyes névmás valamilyen speciális formáját fejezzük ki. Pl.:

magam, magad, maga, ...
egymagam, egymagad, egymaga, ...
alám, alád, alá, ...
rajtam, rajtad, rajta ...

Főnevek

Az igei relációk argumentumaiba többnyire névszói kifejezések kerülnek. A névszói kifejezés utolsó eleme általában egy főnév. Ha nem főnév az utolsó elem, akkor a névszói kifejezés hiányos. Mi azonban nem mindent tartunk főnévnek, amit a hagyományos nyelvtanok főnévnek neveznek. Mint majd látjuk, elég sok olyan szófaj van, ami nem főnév ugyan, de majdnem az. Ilyen például a foglakozásnév. A *cipész mester*, *cipész úr*, *cipész tanuló*, *cipész üzlet*, *cipész szakma* stb. névszói kifejezésekben a *cipész* szó

(főnévi) jelző és nem főnév. Mi tehát, ha azt mondjuk, hogy *rosszul kereső cipész*, akkor ezt egy hiányos névszói kifejezésnek tartjuk. Ez a kifejezés természetesen nem annyira hiányos, mintha azt mondanánk, hogy *három nagy, piros*, mert a foglakozásnévnek van egy defaultja (a *szakmunkás*, vagy valami ilyesmi), ami itt automatikusan belép, és nem kell hivatkozásként keresni.

A főneveknek különféle jelzői lehetnek. Ezek a névszói kifejezésben meghatározott sorrendben a főnév előtt állnak. De nem csak jelzője lehet egy főnévnek, hanem különféle határozó jellegű vonzatai is. Pl.:

idő+valamire
játék+valakivel/valakivel szemben/valaki ellen
ítélet+valaki ellen

A főnév határozója leggyakrabban közvetlenül a főnév mögött áll, néha azonban elég messze el is szakadhat tőle. Amikor a főnév határozója a főnév elé kerül, egy töltelék melléknévi igenév iktatódik közbe. Gyakran használjuk erre a célra a való szót. Ha vonzat olyan névutóval van jelölve, amiből melléknévi névutó képezhető, akkor ezen az úton is csinálhatunk jelzőt a határozóból. Az ilyen vonzatokat mi külső vonzatnak nevezzük, mert a főnév vonzatát a névszói kifejezésen kívül is kell keresni a szintaktikus elemzés során. (Lásd még az **Igenevek**nél is!) Pl.:

A játék fárasztó volt Péterrel.
Fárasztó volt a Péterrel folytatott játék.
Fárasztó volt a Péter elleni játék.

Nevek

A dolgok neve is olyan szófaj, ami átmenet a főnév és a nem főnév között. A magyar helyesírási szabályzat igen változatos képet mutat abban a tekintetben, hogy hogyan kell a dolgot és a nevét leírni:

január hónap
csütörtök este
Bodri kutya
Olasz Köztársaság
Mátyás-templom
Pál apostol templom
Csepel-sziget
Holt-Tisza
Magyarország

Mint a fentiekből is látható, név alatt nem a nyelvtankönyvi tulajdonnevet értjük. Ha valami mindig a névszói kifejezés utolsó eleme, akkor az egy főnév. Ha a szó valaminek a neve, de nem feltétlenül áll a névszói kifejezés végén, akkor az egy név, a név és a főnév között más szavak is állhatnak:

Roszik Gábor evangélikus lelkész,
Erős Bálint Kossuth-dijas tudós,
Bodri házőrző kutya
stb.

Nincs túl sok olyan eset, amikor a nevet a mögötte álló főnévvel egybeírjuk. Ezek így kerülnek bele a szótárba. Problémásabb a helyzet, amikor a főnevet és a nevet kötőjellel írjuk. Mivel a morfológia könnyedén le tudja választani az előtagot, be lehetne vezetni olyan speciális vonzatokat pl. a sziget, templom, díj stb. szavakhoz, ami azt jelezné, hogy a nevét kötőjellel kell eléírni. Így, ha a név-rész nem szerepelne a szótárban, a szótár kisebb lehetne. Másrészt biztosak vagyunk benne, hogy 10 emberből 8 külön írja, kötőjel nélkül azt, hogy:

Arany János-idézet
Mátyás-templom
Csepel-sziget
stb.

Mesterség - Beosztás, Nemzetiség, Pártállás - Vallás

Ezek a szavak átmenetet képeznek a főnév és a melléknév között. Általában a név és a főnév közt állnak, de egyedül is előfordulnak. Hallgatólagosan odaértendő, hogy ember. A foglalkozásnév kicsit főnevebb, mint a másik kettő, hátrébb szokott állni a névszói kifejezésben és gyakrabban marad el mögötte a főnév. Pl.:

Jesek Kowalsky lengyel csőszerelő szakmunkás
Pietro Longo olasz kommunista képviselő
– Pietro Longo kommunista olasz képviselő

Anyagnevek

Az anyagnév lehet a névszói kifejezés utolsó eleme és lehet az utolsó előtti elem is. Pl.:

3 tonna öntöttvas
3 tonna öntöttvas radiátor

A magyar helyesírás igen kellemetlen szabálya, hogy hol egybeírjuk a főnévvel, hol nem. Pl.:

acélszeg
kovácsoltvas szeg

Mivel nekünk az a fő törekvésünk, a szövegbe foglalt információt egységes, jól kezelhető formára hozzuk, a szemantikai ábrázolásban a főnevet és az anyagnevet két egymással kapcsolatban álló fogalomként írjuk le, függetlenül attól, hogy egybe volt-e írva vagy nem.

Az egybe- és különírás megint egy olyan területe a helyesírásnak, amiben az emberek meglehetősen bizonytalanok. Így a gyakorlatban célszerűnek látszik, ha ezeket a szabályokat lazán kezeljük, például az anyagnevet elfogadjuk egybeírva is, különírva is. Elvégre a programnak nem az a célja, hogy a felhasználót kioktassa.

Az anyagnevek megkülönböztetése a főnevektől azért is célszerű, mert bizonyos képzőket az anyagnevekre alkalmazva a hatásuk megjósolható, míg általában főnévre alkalmazva nem jósolható meg.

Főnév jellegű névmások

A főnév jellegű névmásokra az a jellemző, hogy önmagukban egy teljes, névszói kifejezés helyett állnak, jelzőt nem kaphatnak. (Nem minden névmás hivatkozás, például a kérdő névmás biztosan nem az. A hivatkozások hatóköre is különböző, az erre vonatkozókat lásd **A határozott és határozatlan névszói kifejezés** című fejezetben.) Toldalékolásuk és mondatbeli szerepük alapján 4 alcsoportba sorolhatók.

Főnév jellegű névmások I

A *ki*, *mi*, *mely*(*ik*) végű névmások (pl.: *amely*, *ilyesmi*, *semmi*, *bárki*, ...) mindenféle névszói ragot és jelet megkaphatnak, és a mondatban minden szerepet betölthetnek, amit egy névszói kifejezés szokott.

Főnév jellegű névmások II/a

Az *ez*, *az* és a *mindez*, *ugyanez* névmások birtokos jelet kaphatnak és esetragot. Minden mondatbeli szerepet betölthetnek, amit egy névszói kifejezés szokott. Ezen kívül gyakran állnak nyomatékosító szerepben határozott névszói kifejezés előtt. Pl.:

Ugyanezt az előadást mondta el akkor is, amit most hallottunk.

Az *ez*, *az* névmás szokott lenni az utalószó, amire a vonatokozó névmással kezdődő almondat hivatkozik.

Főnév jellegű névmások II/b

A *magam*, *magad*, *maga*, ... és az ilyen végű névmások (pl.: *jómagam*, *önmagam*, ...) birtokos jelet kaphatnak és esetragot. Minden mondatbeli szerepet betölthetnek, amit egy névszói kifejezés betölthet. Amikor a névmás és az ige személyragja megegyezik, visszahatást fejez ki, azaz azt, hogy az illető mondatrész azonos az alannyal. Ha a névmás 3. személyű, akkor egy nyomatékos személyes névmásnak tekinthető, amely egy azonos szerepű névszói kifejezés helyett vagy azzal együtt áll. Pl.:

Pétert magát nem láttam.

Hasonlóan, alanyi szerepben, mivel visszahatásról szó sem lehet, szintén nyomatékos névmást fejez ki.

Az *egymagam*, *egymagad*, ... jelentése *egymagam=én_egyedül*, *egymagad=te_egyedül* ... Ragozása hiányos. Az alanyesetű és a *nak*-ragos alanyt, valamint az *egymagamban*, *egymagadban*, ... alakokat egyeztetni kell az állítmánnyal. Pl.:

- *Egymagadnak kell megoldanod a feladatot.*
- *Egymagadnak kell megoldani a feladatot.*
- *Egymagadban kell megoldani a feladatot.*
- *Egymagadban kell megoldanod a feladatot.*

Főnév jellegű névmások III

Az *egymás* névmás teljesen külön kategória. Birtokos jelet és esetragot kaphat. Tárgy, határozó és birtokos lehet.

Főnév jellegű névmások IV

Az *enyém, tied, övé*, ... névmások esetragot kaphatnak, tulajdonképpen a birtokos jeles személyes névmásnak felelnek meg. (*enyém=én+;*, *tied=te+;* stb. formára alakítjuk át.)

Itt jegyezzük meg, hogy az *én*, *te*, *ő*, ... személyes névmások ragozhatatlanok. Hasonló ezek tárgyesete az *engem*, *téged*, *őt*, ... A többi esetet a névutóknál tárgyaljuk.

(engem=én+t, téged=te+t, ...
utánam=én_után, utánad=te_után, ...
velem=én+vel, veled=te+vel, ...
stb.)

Melléknév

A melléknév a főnév, az azt megelőző főnévi jelzők és a név előtt szokott állni a névszói kifejezésben. Elég gyakran áll névszói állítmányként is a mondatban. Lehet vonzata, de az a jelzői pozícióban mindig a melléknév előtt áll. Ellentétben az átmeneti típusokkal (foglakozásnév, vallás stb.) nincs defaultja.

Fokozható. A fokozott melléknév vonzatai állhatnak a névszói kifejezésen belül is meg kívül is. De a továbbképzett szó vonzatai mindig belül vannak. Pl.:

Egy a réginél 3 méterrel nagyobb asztalt vettem.
3 méterrel nagyobb asztalt vettem a réginél.
Nagyobb asztal vettem 3 méterrel.
Egy 3 méterrel megnagyobbított asztalt csináltattam.

Szorosan a melléknév előtt (a vonzatok után) egy határozószószerű melléknévi módosító állhat. (Pl.: *alig, igen, kissé, látszólag, nagyrészt, nemigen* stb.)

Ehhez hasonlóan egy melléknévszerű melléknévi módosító is állhat ebben a pozícióban. (Pl.: *jó nagy, borzasztó rossz, szörnyű hideg* stb.) Megjegyzendő, hogy tudományos, technikai és hivatalos szövegekben ilyennel soha nem találkoztunk.

A színek előtt a *halvány*, *sötét* vagy *világos* szavak állhatnak, amit a magyar helyesírás ismert szabályai szerint, aszerint írunk egybe vagy külön a színnel, hogy a szín összetett szó-e vagy sem.

Az *-s/-os/-es/-*ös és az *-ú/-ű/-jú/-jű* képzős melléknevek tulajdonképpen melléknévi formában megadott birtokviszonyt fejeznek ki. A névszói kifejezésben a főnév, a nominális mondatban (amikor a melléknév állítmány) az alany lesz a birtokos, és a melléknév a birtok. Az -s képzőjű birtoknak lehet saját jelzője, az *-ú/-ű* képzőjű birtoknak biztosan van. (De lehet, hogy egybeírjuk vele.)

Melléknévi igenév

A melléknévi igenév minden fajtája teljes értékű melléknév. (Lásd **Igenevek**!) A melléknévi igenév gyakran szerepel egy-két igei argumentumával együtt a névszói kifejezésben. Ilyenkor az argumentumok előtte állnak.

A melléknévi igenév (és az argumentumai) vagy a tulajdonságjelzői pozícióban állnak a mennyiség jelző után és a főnévi jelzők előtt, vagy az úgynevezett szelektor pozícióban a mennyiség jelző előtt. Ebben a pozícióban egyszerű melléknév nemigen szokott állani. (Lásd a **névszói kifejezés**t!)

Pozícionáló melléknév

A pozícionáló melléknévnek nincs középfoka, és felső foka, ha van, *-bb* nélkül képződik. (Pl.: *alsó, közbülső, utóbbi* stb.) Vagy a tulajdonságjelzői vagy a szelektor pozícióban áll a névszói kifejezésben.

Melléknévi névutó

A melléknévi névutót *-i* képzővel képezzük a névutókból. Az ilyen névutót akkor alkalmazzuk, amikor a főnév névutós vonzatát jelzővé alakítjuk. Sajnos nem minden névutóból lehet melléknévi névutót képezni. Nem lehet például azokból, amelyek hova kérdésre válaszolnak. Ennek ellenére a morfológiai analízisnek az *-i* végződést le kell választani, a vonzat felismerése érdekében.

(A dolog másik oldala, hogy az emberek nemcsak ott használják ezt a képzőt, ahol lehet. Illetve, különböző vélemények lehetnek a képző használhatósági köréről.)

Melléknévi névmások

Van egy sereg névmás, ami melléknév helyén állhat. Ha egy főnévnek több melléknévi jelzője van, akkor a névmás mindig legelöl áll. Pl.:

Olyan piros, ropogós cseresznyét vettem, ami ...

Ezek közül bizonyosak, például a kérdő vagy a vonatkozó névmás, mindig a szelektor-pozícióban állnak. Pl.:

Milyen 3 könyvet választottál?

Mások a tulajdonságjelző helyén szeretnek állni, bár időnként a szelektor-pozícióban is előfordulnak. (Attól függ, mit hangsúlyozunk ki.) Pl.:

3 másfajta könyvet választottam.
Másfajta 3 könyvet választottam.

Egy névszói kifejezésben két névmás is előfordulhat, Feltételezzük, hogy azért, mert az egyik tulajdonságjelzői, a másik szelektor-pozícióban áll.

Miféle olyan könyvet ismersz, amelyik ...
Milyen másféle angol nyelvű cserebogarakról szóló könyvet ismersz?

Mértékegységek

A mértékegységek közé tartoznak az SI mértékegységek, a korábbi elavult mértékegységek, és egy sor természetes mértékegység is. Pl.:

3 liter bor
3 akó bor
3 icce bor
3 hordó bor
3 palack bor

A számnév és az azt követő mértékegység a tulajdonságjelző előtt, de a szelektor után áll. Ha egy névszói kifejezésben csak számnév szerepel, akkor defaultként *fajtá*-t, *darab*-ot vagy *fő*-t kell feltételezni. (Azoknál kell fajtát feltételezni, amelyek többes száma nem nagyobb mennyiséget, hanem több fajtát jelenet. Pl.: *füvek*, *borok* stb.)

A mértékegység nem főnév. Persze a szó lehet egyszerre mértékegység is és főnév is. Pl.:

Tudom, hogy szereted a bort, hoztam egy üveget.
(Itt mértékegység.)

De összetörtem az üveget.
(Itt főnév.)

A mértékegységnek lehet saját tulajdonságjelzője. Pl.:

3 nagy hordó bor

Konkrét számnév

A konkrét számnév egy szám, amely megadható betűvel, vagy számmal. *-szor/-ször* raggal számhatározót csinálhatunk belőle. (Lásd **Az egyszerű mondat szerkezete** című fejezetben!) *-dik* raggal sorszámnevet képezhetünk belőle.

Sorszámnév

A sorszámnév kiválaszt egyet a szóban forgó dolgok közül, ezért a szelektor pozícióban szeret állni a névszói kifejezésben. Szoros rokonságban van a pozícionáló melléknévvel jelölt szófajjal - egyes esetekben mindkettőbe tartozik a szó: *első*.

Viszonylag sok szónak van sorszámnév vonzata. Pl.: *éves, osztályos, helyezett, generációs, szintű* stb.

Határozatlan számnév

Vannak olyan melléknévszerű szavak, amelyek mennyiséget fejeznek ki, bár nem mindig tudni, milyen mértékben mérve. Állhatnak mértékegységgel és mértékegység nélkül is. Ilyenek a *kevés, kevesebb, legkevesebb, sok, több, legtöbb, rengeteg*. Pl.:

Sok finom ételt megevett.
Sok pohár finom bort ivott hozzá.

A több és a kevesebb összehasonlításra használható, mint a középfok általában. (A *-val/-vel*-es és *-nál/-nél*-es vonzat lehet a névszói kifejezésen belül és azon kívül is.)

A *számtalan, jónéhány, pár, ...* szó után általában áll mértékegység. Illetve feltételezhetjük a *fajta, darab, fő* defaultot. Pl.:

Számos bort megkóstolt.
Számos palack bort megkóstolt.

Számhatározó mindegyikből képezhető. Állhatnak számnévi vonzat helyén.

A határozatlan számnév előtt nem állhat kvantor, ezért elég gyakori, különösen az első csoportba soroltak esetében, hogy összetévesztik a kvantorokkal és a szelektorokkal.

Számnévi névmások

A *mennyi* végű számnévi névmások (*valamennyi, semennyi, bármennyi, akármennyi*) a teljes mennyiségjelző helyett állnak, mértékegység nem alkalmazható rájuk.

A *hány* végű számnévi névmások (*ahány, valahány, sehány*) csak a számnév helyett állnak, a mértékegységet ki kell tenni.

A *hányadik* végű névmások sorszámok helyén, a *hányszor* számhatározó helyén használható.

Az *annyi, ennyi* névmást (*ugyanannyi, ugyanennyi*), bár az első csoportba tartozónak tűnik, elég gyakran használják mértékegységgel is. Pl.:

Ugyanannyi méter szövetet vegyél, mint legutóbb.

Kvantorok A *minden*, *bármelyik*, *egy-egy*, *semelyik*, ... Ezek a szavak megszabják azon objektumok körét, amire az állítás érvényes. Bizonyos értelemben ezt teszi a mennyiség, a sorszámnév, a szelektor, felsőfokú jelző és a névelő is.

A kvantorokra az jellemző, hogy a mennyiségjelző előtt állnak, nem állhat előttük névelő. Pl.:

Bármelyik 3 könyvet elviheted.

Szelektorok Körülbelül ugyanolyanok, mint a kvantorok, de mindig határozott névelő áll előttük. (Ennek következtében például a birtokos mindig *-nak* ragot kap.) Pl.:

Az egyik 3 könyvet elviheted.

Az *egyik*, az *összes* és a többi szavak tartoznak ide.

Névutók

A névutókkal külön fejezet foglalkozik. A névutókat két osztályba soroljuk aszerint, hogy személyragozhatók-e vagy sem. A személyragozhatókat nevezzük névutóknak - a ragozhatatlanokat névutószerű szavaknak. Pl.:

közöttem, közötted, közötte, ... (névutó)
közben (névutószerű szó)

A névutókat aszerint osztályozzuk, milyen vonzat szerepét tölti be a névutós kifejezés. Eddig a következő alosztályokat különböztettük meg:

Vegyes névutók.
"Hol" kérdésre válaszolók.
"Hova" kérdésre válaszolók.
"Honnan" kérdésre válaszolók. Időhatározói névutók.
"Hogyan" kérdésre válaszoló névutók.

Egy névutó természetesen több csoportba is beletartozhat. Például minden időhatározói névutó egyben helyhatározói is.

A ragozhatatlan szavak

A ragozhatatlan szavak a mi felfogásunk szerint semmilyen végződést nem kapnak. (Se ragot, se képzőt, se jelet. A határozószavak képzőiről, más ehhez hasonló kérdésekről a **Végződések** című fejezet **Kivételes esetek** című alfejezetében írunk. Mindenesetre, ilyen szavak csak igen kivételesen kaphatnak képzőket, és ilyenkor a képzett szót is betesszük a szótárba.)

Határozószók

A határozószók olyan szavak, amelyek egy ragos vagy névutós névszói kifejezés helyén állhatnak. A határozó szavakat aszerint osztjuk fel, hogy milyen vonzat helyén állhatnak. Jelenleg a következő alosztályokat különböztettük meg:

Vegyes határozószók.
"Hol" kérdésre válaszolók.
"Hova" kérdésre válaszolók.
"Honnan" kérdésre válaszolók.
Időhatározói határozószók.
"Hogyan" kérdésre válaszoló határozószók.

Egy másik megközelítés szerint vannak konkrét határozószók és vannak névmási határozószók. Az előbbiek viszonylag konkrétan megadják a helyet, időt stb., az utóbbiaknál meg kell az antecedenst keresni.

Mi az *amitől, amerről, amettől, valamikor, bármeddig, akármibe*, ... szavakat ragos névmásoknak tekintjük. (Jó néhányat közülük rendhagyó alakként teljes felbontásával együtt beteszünk a szótárba. *valamikor=valamikor+kor*, *bármeddig=bármikor+ig*, *amerről=ami_felől*, *amettől=amikor+tól* stb.) Az *ahol, amerre, ahova* alakokat pedig fiktív, gyűjtő jellegű HOL, HOVA, HONNAN ragokkal ellátva vezetjük vissza a névmásokra. (*ahol=ami*+HOL, *ahova=ami*+HOVA, *ott=ez*+HOL, *ide= ez*+HOVA, stb.) Az utóbbiakat úgy tekintjük, hogy például a hol vonzatnak megfelel, de se a *-ban/-ben*, se az *-on/-en/-ön*, se a *-nál/-nél* vonzatnak nem.

Igekötők

Az igekötő és az ige nagyon szorosan összetartozik. Az igekötős ige jelentésének gyakran igen kevés köze van az eredeti ige jelentéséhez, és egész más vonzatai vannak. Számos olyan igekötő van viszont, ami alakilag egy határozószóval vagy egy névutóval esik egybe. Ezért a mondat elemzésének sebessége miatt elég lényeges kérdés, annak eldöntése, hogy egy szó igekötője-e az igének vagy sem.

A főnévi igenév igekötője vagy az igenév előtt áll vagy segédige előtt. A határozói igenév és az *-andó/-endő*, *-ható/-hető* végű melléknévi igenév igekötője pedig vagy

az igenév előtt áll, vagy az azt megelőző létige előtt. (A többi melléknévi igenév igekötője nem válik el.)

A ragozott ige igekötője viszont bárhol állhat a mondatban. Ilyenkor csak az segíthet a döntésben, hogy melyik ige vonzatkeretét elégíti ki a mondat. Az esetek jelentős részében, ahol az igekötő csak kissé módosította a jelentést, majdnem mindegy, hogy igekötőnek tekintjük a szót, vagy másnak. Pl.:

Futott a mezőn keresztül.
Néha azonban nem mindegy:
Itt a Mikulás, járuljatok hozzá!

Mi minden szót igekötőnek tekintünk, amit egybe írunk az igével, ha előtte áll és egyébként külön. (A helyesírási szabályzat a *tönkre* szónál jelzi, hogy igekötő, a *létre* szónál nem.)

Névelők

Az a és az nemcsak határozott névelő, de névmás is. Mégsincs velük különösebb probléma, mert névmásként nemigen szoktak határozatlan névszói kifejezés előtt állni.

(Kivéve talán az egyszerű nominális mondatokat. Pl.: Az egy szamárság.)

Sajnos a határozatlan névelőt soha nem tudjuk megkülönböztetni az egy számnévtől. Ez tisztán hangsúlyozás kérdése.

Kötőszók

A kötőszók két nagy csoportba oszthatók. Az első csoportbeliek csak mondatokat kapcsolnak össze, a második csoportba tartozók mondatrészeket és mondatokat is összekapcsolhatnak. (Olyat nem találtunk, ami csak mondatrészeket kapcsolt volna össze.) Jelenlegi felosztásunk a következő:

Mindig mondat elején álló mondatkötőszavak. Pl.: *de, ha, hiszen, csakhogy* stb.

Nem mindig mondat elején álló mondatkötőszavak. Pl.: *noha, bár, azonban* stb.

Olyan kötőszavak, ahol a vesszőtől függ, hogy mondatot vagy mondatrészt kötnek össze: *és, vagy, meg, s*.

Mondatokat vagy mondatrészeket összekapcsoló szavak. Pl.: *tehát, illetve, vagyis, azaz* stb.

Ismétlő kötőszavak: *akár, mind, se, sem, vagy*.

Kétrészes kötőszó első fele: *nemcsak, egyrészt*.

Kétrészes kötőszó második fele: *hanem, másrészt*.

Utáni kötőszó: *is, se, sem*.

Módosítószavak

Tagadószó: *nem*.

Néha gondot okoz, hogy a névszói kifejezésre vonatkozik-e, vagy annak első szavára. Pl.:

Nem piros fotelt vettünk (hanem piros kanapét).
Nem piros fotelt vettünk (hanem kéket).

Igei módosító szavak: *hadd* – óhajtás,
ne, se, nem, sem - tagadás,
nemigen, alig - gyengítés.

Mindig az ige előtt állnak, de a *hadd ne* egyszerre is szerepelhet.

Melléknévi módosító szavak: *nem, nemigen, alig, nagyon, teljesen, formailag, állítólag, bizonyára* stb.

Ha melléknév vagy határozatlan számnév előtt áll, annak jelentését erősít, gyöngíti, tagadja. Ha nem ott áll, hogyan kérdésre válaszoló szabad határozónak tekintjük.

Számnévi módosító szavak: *kb., mintegy, legalább, legfeljebb, mindössze*, stb.

Konkrét számok előtt áll és bizonytalan értéket csinál belőle.

Nem-igei segédigék

Azok a melléknevek, amelyeket segédigei szerepben szoktunk használni. (Lásd **Az egyszerű mondatok szerkezete** című fejezetben a **Melléknév+főnévi igenév, mint állítmány** című alfejezetet.)

Ragozatlan főnévi igenév mellett: *alkalmas, alkalmatlan, képes, kész, köteles*.

Ragozott főnévi igenév mellet: *ajánlatos, célszerű, érdemes, felesleges, lehetetlen, szabad, tanácsos, tilos*.

Csak főnévi igenév mellett tekintjük ezeket a szavakat segédigének. Ha névszói kifejezés része vagy nominális mondatban áll, akkor melléknév.

Ragozhatatlan személyes névmások

Az *én, te, ő, ... és az engem(et), téged(et), őt, ...* soroljuk ide. (Lásd **Főnévi névmások IV** című pontot.)

Névutószerű szavak

Azokat a névutókat soroltuk ide, amiket nem birtokos személyragozva alkalmazunk a személyes névmásokra. (Lásd a **Névmások** című fejezetet.)

Egyéb szavak

Azokat a szavakat soroltuk ide, amelyek elhagyhatók a mondatból, és a mondatba foglalt tények nem változnak:

Bizony, csak, csupán, egyébként, egyáltalán, hát, ...

A NÉVSZÓI KIFEJEZÉSEK

A magyar nyelv szintaktikájában lényeges szerepet tölt be a névszói kifejezés. A névszói kifejezés egy olyan mondatrész, amelyet legegyszerűbb esetben egy (ragozott vagy névutóval ellátott) főnév tölthet be. Egy névszói kifejezés önmagában nem alkot mondatot, hanem mindig egy másik mondatrészhez kapcsolódik (ami akár igei akár főnévi jellegű lehet). Ez a kapcsolat a névszói kifejezés egy vagy több elemén megjelenik rag vagy névutó formájában, pl. az ige tárgyának a *-t* ragja, a helyhatározók ragjai és névutói, a birtokos *-nak* ragja stb. Speciális és nehezen kezelhető eset, amikor a kapcsolódó névszói kifejezés alanyesetű és így nem kap ragot. A kapcsolat bizonyos esetekben nemcsak az esetraggal illetve névutóval van megadva, hanem azon a mondatrészen is jelölve van, amihez a névszói kifejezés kapcsolódik, például az igeraggal vagy birtokjellel. Esetenként ezt is figyelembe kell venni a megfeleltetésnél, amely szerint megállapítjuk a mondatban szereplő kifejezések és részkifejezések hovatartozását.

Ha a magyar nyelvre igaz is lenne az az állítás, hogy szabad szórendű, akkor ezen belül a névszói kifejezésre ez biztosan nem áll. A névszói kifejezések a mondatban egy-egy szét nem váló, kötött szórendű csoportot alkotnak. Vannak ugyan olyan logikai szerkezetek, amelyeket egyaránt leírhatunk névszói kifejezéssel is és olyan nyelvtani konstrukcióval is, ahol egyes részek szétválnak egymástól, de az utóbbiakat nem soroljuk a névszói kifejezések közé. Nem tekintjük például egy névszói szerkezetnek a *-nak*-os birtokos és a birtok együttesét, hisz ezek elválhatnak egymástól. PL: *Péternek láttam a kabátját*.

Célunk az, hogy a mondatelemzés részeként felismerjük ezeket a szét nem váló részeket, és feltárjuk belső szerkezetüket.

A névszói szerkezetek informális leírása

A névszói szerkezet a mondatnak egy olyan része, amely kötődik valamely más mondatrészhez, és amely funkcionálisan egyetlen egységes szerepet tölt be. Lehet egy ige vagy igenév tárgya, alanya, bármilyen határozója, egy nominális mondat két alanyesetű fő összetevőjének bármelyike és bármiféle határozója, egy névutót vagy melléknévi névutót megelőző szerkezet, egy névszó birtokosa, stb. A névszói szerkezet hovatartozására általában a szerkezet utolsó szavához függesztett rag vagy névutó utal, és ezen keresztül lehet megtalálni, abból a mondatrészből kiindulva, amelyik egy ilyen vonzatot igényel illetve lehetővé tesz.

A névszói kifejezésben a benne szereplő alkotórészek helye kötött, bár szinte minden alkotórész elhagyható. A névszói kifejezés felismerése tehát nagyjából úgy történik, hogy

ellenőrizzük, hogy a soron következő pozícióban ott van-e az ott lehetséges elemek valamelyike. Egy névszói kifejezésen belül állhat egy másik beágyazott névszói kifejezés, de ennek a helye is meghatározott. Ilyenkor a névszói kifejezést felismerő algoritmus rekurzíve hívja önmagát.

Ez az eljárás gyökeresen különbözik attól az eljárástól, amit máskor használunk a mondat elemzésére során. Például a ragozott ige, a melléknévi igenévi állítmány, a vonzatos melléknévi állítmány esetén a hozzá kapcsolódó bővítményeket úgy találjuk meg, hogy bárhol a mondatban olyan elemeket (többnyire bizonyos raggal ellátott névszói kifejezéseket) keresünk, amelyek megfelelnek az adott szerepre. (Lásd részletesebben a **mondatelemzés**nél.)

Amikor ragos vagy névutós névszói kifejezést keresünk, akkor természetesen csak a legkülső szintű névszói kifejezések jönnek szóba, a beágyazottak nem. (Kétséges viszont az az eset, amikor a névszói kifejezés elején egy olyan beágyazott névszói kifejezés áll, ami egy külsőbb szintnek, például az igének is egy lehetséges argumentuma. Mivel a magyar mondatok általában hiányosak, ilyenkor nehéz eldönteni, hogy az ige vonzatkerete hiányos-e, vagy a névszói kifejezés hiányos, esetleg a kifejezés mindkét szerepet betölti. Lásd **elnyelődés.**)

Amikor egy adott szerepet betöltő névszói kifejezést keresünk, mi két különböző típusú megfeleltetést ellenőrzünk:

1. Egy szó (pl. ige) kívánhat bizonyos argumentumokat, melyeket, mint függvényt az argumentumai, alkalmas raggal vagy névutóval jelölt mondatrészek elégíthetnek ki. Ezt nevezzük vonzatnak. Pl.: *Elvitte => a maradék három zsák búzát*

2.a. Lehet birtokviszony, amikor is a birtok végén levő birtokjellel kell egyeztetni a birtokos számát és személyét. Pl.: *Anyám <= tyúkja*.

2.b. Lehet alany vagy tárgy, ahol az igeraggal kell egyeztetni az alany számát személyét illetve a tárgy határozottságát. Pl.: *Egy kicsit elkérte az órámat. elkérte => az órámat*

Mivel a névszói kifejezésen belül állhat újabb névszói kifejezés, ezért az 1. és 2.a. egyeztetéseket ezekre is el kell végezni.

Egy névszói kifejezés lehet egyszerű és halmozott. A halmozott névszói kifejezés nem más, mint azonos szerepet betöltő névszói kifejezések vesszővel vagy kötőszavakkal összekötött sorozata. Ilyenkor általában, de nem mindig azonos raggal, vagy névutóval rendelkeznek az összekötött névszói kifejezések. Ennek két funkciója lehet. Vagy különböző objektumokat kötünk össze logikailag, pl. Pétert és gyerekeit, vagy egy objektum pontosabb leírására értelmezőt használunk, pl. *Péter kovácsot, a hosszú hajút*.

A kettőt külső jegyek alapján nemigen lehet megkülönböztetni.

Pl. a *Pista és öccse mindkét foltos kutyakölykét eladták a múlt héten rendezett debreceni vásáron.* mondatban a következő névszói kifejezések találhatók:

Pista és öccse mindkét foltos kutyakölykét
az *eladták* tárgya, legfelső szintű
a múlt héten rendezett debreceni vásáron
az *eladták* helye, legfelső szintű
a múlt héten
a *rendezett* ideje, beágyazott
Pista és öccse
a *kutyakölykök* birtokosa, beágyazott

Ezen belül a legutolsót kivéve mindegyik egyszerű (nem halmozott) névszói kifejezés.

Az egyszerű névszói kifejezés

Az egyszerű névszói kifejezés - továbbiakban névszói kifejezés - szigorúan kötött szórendű, melyben a szavak sorrendjét az összetartozó szavak szófaja határozza meg. A szerkezet lehet például egy ragozott főnév jelzőkkel, névelővel. Ebben az esetben, legelöl áll a névelő, utána a jelzők, majd az utolsó elem a ragos főnév. A probléma az, hogy ennél bonyolultabb szerkezetek is léteznek, és a jelzők között is lehet sorrendi szabályokat felállítani. A könnyebb megértés érdekében felépítünk egy nagyon összetett kifejezést, melynek során jól érzékelhető lesz a részek meghatározásának szabálya.

Az *a lengyel vendégmunkással* egy aránylag egyszerű szerkezet.

Ha neve is van, akkor azt csak a névelő helyére írhatom:
Jan Kowalski lengyel vendégmunkással

Ha viszont szerepel más jellegű jelző, azt a név elé írjuk:
a vörös hajú Kowalski lengyel vendégmunkással

Ha a *vörös hajú*-t a név után akarom írni, akkor ezt már csak értelmezőként lehet, tehát formálisan halmozott névszói kifejezésként:
Kowalskival, a vörös hajú lengyel vendégmunkással

A két jelző között az a különbség, hogy míg a *vörös hajú* egy egyszerű melléknév, addig a lengyel egy úgynevezett főnévi melléknév. Ilyenek pl. a nemzetiséget, foglakozást jelentő szavak (lásd **szófajok**).

Ha azt is meg kell mondani, hogy hány munkásról van szó, akkor azt a tulajdonságjelzők előtt kell megtennem:
három vörös hajú lengyel vendégmunkással

A mennyiség után mértékegység is állhat:

a három hordó pasztőrözött Radeberger sör

Tehát a sorrend: névelő, mennyiség, tulajdonságjelzők, név, főnévi melléknév, főnév, jelek, rag.

Ebből a listából bármi kimaradhat, de a jelek és a rag előtt névelőtől különböző szónak kell szerepelnie. A ragot megelőző elemen kívül minden ragozatlan (formálisan alanyesetben szerepel). Egyes elemek lehetnek halmozottak, pl. vesszővel, kötőszavakkal elválasztott azonos funkciót ellátó jelzők, vonzatos szerkezetek:

a három nagy hordó ecetesedő, de még iható asztali bor

Alanyesettel kapcsolódhat a szerkezethez a birtokos jelző is, és ha van, megelőzi a mennyiségjelzőt is:

a hajógyár három vörös hajú lengyel vendégmunkásával

Ilyenkor természetesen a birtokjelnek egyeznie kell a birtokossal számban, személyben:

A mi fiaink.

Van még egy szófaj, amely a névszói szerkezetet bevezetheti, sőt csak legelöl szerepelhet. Ez a kvantor:

mindegyik vörös hajú lengyel vendégmunkással

Természetesen a kvantor önmagában is lehet önálló névszói szerkezet. Ilyenkor a ragot, jelet ehhez kell ragasztani, s a kötelező nevelő is elmarad:

mindegyikkel

A határozott névelőt csak egy dolog előzheti meg. Ez a nyomatékosító névmás, amely ugyanazzal a raggal, esetleg jelekkel szerepel, mint maga az egész kifejezés:

magukkal a vörös hajú lengyel vendégmunkásokkal
ezzel a három vörös hajú lengyel vendégmunkással

A névelőt és az azt megelőző részt mi vonatkozásnak nevezzük. Tehát a globális sorrendi szerkezet, melyben bármely rész elmaradhat: Vonatkozás, mennyiségjelző, tulajdonságjelzők, név, főnévi melléknév, főnév, jelek, ragok.

A vonatkozás lehet kvantoros, határozott és határozatlan. A kvantor meghatározása nálunk eltér a megszokottól. Mi kvantornak nevezzük azt az elemet, mely pozíciójában a névelő helyén szerepelhet, de azt kizárja. Pl. mindegyik, valamelyik. Sok olyan, hagyományosan kvantornak nevezett szót, melyek a logika nyelvén esetleg valóban kvantornak neveznénk, mi inkább kvantor jellegű számnévnek tartjuk.

Ezeket a kvantor jellegű szavakat - ellentétben a kvantorral - minden esetben egy határozott névelő, vagy egy birtokos előz meg.

Pl.: *Az összes állat*

A határozott vonatkozás részei opcionálisak, kivéve a névelőt: nyomatékosító, határozott névelő, birtokos jelző.

A határozatlan vonatkozás minden része opcionális: birtokos jelző, határozatlan névelő.

A birtokos jelző egy alanyesetű névszói kifejezés, mely számban és személyben meg kell egyeznie az egész kifejezés végén álló birtokos jel számával, személyével. Magának a birtokos jelzőnek, mint névszói kifejezésnek, lehet birtokos jelzője: *Józsi bácsi kutyája harapásának nyoma*.

A példában is jól látszik, hogy hosszabb birtokoslánc esetén az utolsó birtokviszonyt általában *-nak* raggal fejezzük ki. Ezt nem tekintjük egy névszói kifejezésnek. A többi viszont egy egymásba ágyazott névszói szerkezetnek tekinthető.

A mennyiségjelző csak egyes számú szerkezetben szerepelhet. Részei: szám, mértékegység.

A szám lehet egy konkrét szám, vagy egy számjelző, vagy egy kvantor jellegű számnév. Ez utóbbi esetben meg kell hogy előzze vagy egy határozott névelő, vagy egy birtokos. A másik két esetben egy melléknévi vagy számnévi módosítószócska előzheti meg:

csaknem ezer
nagyon sok

Néha maga a mértékegység is kaphat jelzőt, de ezek száma erősen korlátozott:

négy tizenhárom akós hordó bor
néhány amerikai mérföld út

A tulajdonságjelzők típusa elég sok: melléknevek, melléknévi igenevek bővítményeikkel, melléknévi névutós szerkezetek. Természetesen mindegyik elé rakhatunk melléknévi módosítószót.

A kiemelés

A korábbinak ellentmondani látszik a következő példa:

a hajógyár tegnap érkezett utolsó hat lengyel vendégmunkása
birtokos jelző jelző meny. fn.mn. főnév

Ennek az az oka, hogy a fent említett sorrendiségen kívül egy másik szabály is szerepet játszik. Ez a szabály hasonlít a mondatszerkezeteknél felismert topik-fókusz szórend-szabályra: ez a különböző jelzők között tesz különbséget aszerint, hogy mi az, ami csak egy leíró, és mi az, ami a korábban meghatározott halmazból szelektáló (fókusz jellegű). Ráadásul ezek a szerkezetek egymásba is

lehetnek ágyazva. A fenti mondatot úgy kell értelmezni, mintha azt írták volna, hogy:

Azzal a hat lengyel vendégmunkással, akik utolsók voltak, és tegnap érkeztek.

Tehát a példában a következő lehetséges szerkezetet lehet felismerni:

a hajógyár (*tegnap érkezett* (*utolsó* (*hat lengyel vendégmunkásával*)))

Az *utolsó* kiválasztja, hogy melyik *hat lengyel vendégmunkás*-ról van szó, a *tegnap érkezett* pedig megmutatja, hogy melyik *utolsó hat lengyel vendégmunkás*.

A kiemelés általában, de nem mindig, valamilyen kiválasztást jelent abból a halmazból, amelyet az utána álló kifejezés leír. Ezért bizonyos jelzők előszeretettel szerepelnek a kiemelésben. Egyszerű melléknevet igen ritkán emelünk ki. Míg a középfokú, felsőfokú jelzők szinte kizárólag kiemelt jelzők, a bővített melléknévi igenév, a több szóból álló jelzők csak többnyire kiemeltek. A kiválasztás következtében a névszói kifejezés vagy határozott névelős, vagy kvantoros lesz.

A főnév és a név rendkívül ritkán szerepel kiemelt részben. Míg név elé nem szoktunk névelőt tenni, addig - ritkán használt nyelvtani szerkezetekben - a kiemelt név elé teszünk:

Péter – a Péter gyerek

A főnév és a név kiemelésére gyakoribb megoldás a jelzősítés, azaz be kell ágyaznunk egy összetett jelzői kifejezésbe:

a vörös hajú Jan Kowalski lengyel vendégmunkás –
a Kowalski nevű vörös hajú lengyel vendégmunkás

a gyanakvó rendőrőrmester –
az őrmesteri rendfokozatú gyanakvó rendőr

A főnévi melléknév ritkán kerül előre, illetve ha szerepel, akkor az egyedüli kiemelt jelző:

A lengyel vörös hajú munkást hívták. (Nem a jugoszlávot.)

Van egyfajta jelző, amely szinte kizárólag kiemelve fordul elő. Ezek a sorszámnevek, és az általunk pozícionáló mellékneveknek nevezett szavak. Ezek olyan melléknevek melyeknek nincs középfoka, és ha van felsőfoka, akkor azt az alapfokból képezzük: szélső, középső, utolsó stb. Ezeket együttesen pozícionáló jelzőknek nevezzük.

A névutós szerkezetek

A névutós szerkezeteknek két fajtája van: raggal kapcsolódó és alanyesettel kapcsolódó szerkezetek. Az alanyesettel kapcsolódó szerkezetnél a névutót meg kell előznie egy alanyesetű névszói szerkezetnek:

Péter után.

Vannak viszont olyan névutók amelyek ragos névszói kifejezést kívánnak. Ha egy ilyen névutót megelőz egy megfelelő esetraggal ellátott névszói szerkezet, akkor ezeket együtt egy névszói szerkezetnek tekintjük.

Péterrel szemben.

Viszont a ragot vonzó névutók el is szakadhatnak a hozzájuk tartozó ragos névszói szerkezettől:

túl az Óperencián.

A melléknévi névutóktól sohasem válik el a hozzátartozó ragos szerkezet.

az Óperencián túli

Ha alanyesetet vonzó névutós szerkezetek elején egy nyomatékosító névmás áll, akkor utána a névutó megismétlődik.

ez után a tanácskozás után,
** ez a tanácskozás után*

Ragot vonzó névutós szerkezet elején álló névmásnál viszont nem ismétlődik a névutó, hanem csak a rag:

** ezzel szemben az utcával szemben,*
ezzel az utcával szemben.

Bizonyos esetekben, amikor a szó igekötői és névutói szerepe keveredik, a nyelvtani szerkezet felismerése meglehetősen komplikált.

Utánanéztem az ügynek.
Nézz az ügy után!
Ránéztél a gyerekre?
Csak rájuk kell nézni a gyerekekre.

A névutókról külön fejezetben írunk részletesebben.

A névszói szerkezetek formális leírása

A formális leírás azt a célt szolgálja, hogy alapja legyen az elemző algoritmus megadásának. A leírás nem teljes: tartalmaz formálisan nem kifejtett részeket is. Ezen kívül sok helyen tágabb a valós nyelvnél, mert az analízisre használt modellben arra törekedtünk, hogy a nyelvet minél jobban lefedjük. Nem foglalkozunk például azzal a korláttal, hogy egy névszói kifejezés nem tartalmazhat túl sok jelzőt, mert érthetetlenné válik, és a szerkezeteket sem lehet korlátlanul egymásba ágyazni.

A továbbiakban az egyszerű névszói kifejezéseket a NÉVSZÓI_KIFEJEZÉS szimbólummal jelöljük, a halmozott névszói kifejezést pedig a NÉVSZÓI_SZERKEZET szimbólummal.

Egy NÉVSZÓI_SZERKEZET nem más, mint egyazon megfeleltetést kielégítő NÉVSZÓI_KIFEJEZÉS-ek vesszőkkel *és/vagy* kötőszavakkal összefűzött sorozata. Ilyen sorozatok szerkezetének leírásával a későbbiekben foglalkozunk.

A definíció rekurzív, hisz egy névszói kifejezésnek, mint a következő példa is mutatja, a részei is lehetnek újabb névszói szerkezetek, kifejezések.

A névszói kifejezés elemeinek egyszerű sorrendisége, önmagába ágyazódó struktúrája lehetővé teszi, hogy szerkezetét környezetfüggetlen grammatikával közelítsük. Nyelvtanát egy BNF-szerű leírással adjuk meg. A **::=**, **|**, **[]**, **{}** jelek használata a szokásos:

A **::=** jel baloldalán levő nyelvtani egységet a ::= jel jobboldalán fejtjük ki.

A **|** jel a különböző alternatívákat választja el.

A **[]** zárójelek opcionális részt foglalnak magukba.

A **{}** zárójelek akárhányszor ismételhető (esetleg egyszer sem szereplő) nyelvtani egységet jelölnek.

A **^** jelet azokon a helyeken használjuk, ahol az adott szót közvetlenül megelőző pozícióban szerepelhetnek a szó vonzatai, ha vannak neki.

A **()** jelek között arra utalunk, hogy milyen megfeleltetésnek kell eleget tennie az adott mondatrésznek. Csupa nagybetűvel jelöljük a további kifejtésre váró nyelvtani egységeket. A nagybetűvel kezdett elemek azonosítása a morfológia alapján történik. Ahol szükséges (pl. nem CF jellegű kötöttségek vannak), a formális leírást egyéb informális megjegyzéssel egészítettük ki.

A megfeleltetés, mint említettük, vagy birtokbirtokos-illeszkedés, vagy vonzatilleszkedés alapján történik. Mivel egy (rész-)szerkezet modellünkben egyszerre csak egy nyelvtani szerepet tölthet be, kettős megfelelés egyidejűleg nem történhet, tehát egy névszói kifejezésen belül a birtokos mindig alanyesetben áll. (A *-nak*-os birtokost nem tekintjük

egy névszói kifejezés részének, az ilyen párosítás meghatározása a mondatszerkezet elemzésére hárul.)

```
NÉVSZÓI_KIFEJEZÉS(vonzat/birtok)::= BIRTOKOS(birtok) |
                                   VONZAT_SZERKEZET(vonzat)
```

A gyakoribb eset, a vonzat szerinti illeszkedés:

```
VONZAT_SZERKEZET(vonzat)::= NÉVUTÓS_SZERKEZET(vonzat) |
                            RAGOS_SZERKEZET(vonzat) |
                            [Módosítószó] ^Határozószó(vonzat)
```

Pl.: *tegnap, otthon, a Feri kovácson kívül, ettől az éjszakától, nagyon gyorsan*

Egyesek számára meghökkentő lehet, hogy a határozószókat is a névszói kifejezésbe soroljuk. Meg kell állapítanunk viszont, hogy a határozószó is ott áll a mondatban, ahol a ragos vagy névutós főnevek szoktak állni; és a halmozott kifejezésekben a határozószók keverednek a ragos névszói kifejezésekkel. Nincs tehát más lehetőségünk az elemzés során, minthogy a határozószókat is kezeljük, mégpedig ugyanolyan bontásban, mint a ragokat és névutókat. (Például: hol kérdésre válaszolók, hova kérdésre válaszolók, stb.)

A lehetséges ragot, a névutót illetve a határozószó osztályát a vonzat határozza meg.

```
NÉVUTÓS_SZERKEZET(vonzat)::=Birtokragos_névutó(vonzat) |
          [RAGOS_SZERKEZET(névutó vonzata)]
                            Ragot_Vonzó_Névutó(vonzat) |
           RAGOS_SZERKEZET(Alanyeset)
                            Alanyesetet_Vonzó_Névutó(vonzat)
```

Tehát azt, hogy a névutó előtt milyen esetben áll a névszói szerkezet, a névutó vonzata határozza meg.

Pl.: *utánuk, mindezek előtt, ezen a lyukas fazékon kívül, ez után a könyv után*

A ragos kifejezések a következő két csoportra oszlanak:

```
RAGOS_SZERKEZET(vonzat)::= Ragozott_névmás(vonzat)|
 [VONATKOZÁS]
 [SZELEKTOR]
 [MENNYISÉGJELZŐ]
 {TULAJDONSÁGJELZŐ}
 {Név} [^Főnévi_melléknév] [^Főnév]
      +Birtok/Többes számjel [+Birtokos jel] +Rag(vonzat)
```

A + jel azt jelenti, hogy egybe kell írni az előtte álló nyelvtani egységgel. (Az alanyeset üres ragját, illetve az üres birtokos-többes szám-jelet is jelnek tekintjük.) Jól látható, hogy a ragokon és jeleken kívül minden opcionális. Mivel a ragok és jelek önállóan nem szerepelhetnek, az őket megelőző elemek közül legalább egynek szerepelnie kell. A rag és a jel ahhoz a részhez tapad, amelyik őtőle balra a legközelebbi előforduló. A ragozott névmás leírását lásd a **névutók**ról szóló fejezetben.

Az egyes részek sorrendje rendkívül kötött. Pl.:

Ragozott névmás
ezen

Vonatkozás Tulajdonságjelző Főnév+rag
az én kedves feleségemről

Név Név Főnévi_melléknév Főnév+rag
Kovács Péter magyar bányásznak

Vonatkozás Szelektor Mennyiség Tul.ság+birtokos jel+rag: ezét a második hat vagon penészesét

A ragokat, a jeleket és a szófajokat a morfológiai elemző állapítja meg. A név, a főnévi melléknév és a főnév magyarázatát lásd a **morfológiá**nál.

A tulajdonságjelzők a hagyományosan melléknév jellegű mondatrészek:

```
TULAJDONSÁGJELZő::=
          NÉVSZÓI_SZERKEZET(melléknévi névutó vonzata)
          Melléknévi_névutó |
          [Melléknévi_módosító] ^Melléknév |
          [Melléknévi_módosító] ^Melléknévi_igenév |
          [Melléknévi_módosító] Melléknévi_névmás
```

Pl.: *Péter utáni, nagyon piros, utcára néző, ugyanolyan*

```
MENNYISÉGJELZő::= SZÁMJELZő [[EGYSZERś_JELZő] Mértékegység]
SZÁMJELZő::=
     [Melléknévi_módosító] ^Melléknév-jellegű_számnév |
     ^Számnévi_névmás |
     Kvantor_jellegű_számnév |
     [^Számnévi_módosító] {Számnév}
```

Mennyiségjelző után csak egyes számú részkifejezés állhat. Természetesen az utolsó "[Számnévi_módosító] {Számnév}" verzióban legalább egy számnévnek kell szerepelnie. (A számnevek egybe- és különírási szabályával itt nem foglalkoztunk.)

A határozatlan névelő általában megkülönböztethetetlen az egy számnévtől.

Mint már említettük, a kvantor jellegű számnevet meg kell előznie egy kvantortól különböző vonatkozásnak.

Pl.: *számos, néhány teli hordó, legalább négy üveg*

Nem elírás, hogy számnévi módosítót is megelőzhet vonzat. A *mint* szócskát számnévi módosítószónak tekintjük, amelyet megelőz egy több vagy kevesebb szó.

Pl.: *Több mint négy percet késtem.*

A szelektor valamilyen értelemben pontosít, hogy melyik a sok közül, illetve itt szerepelhetnek a kiemelt jelzők:

SZELEKTOR::={KIEMELT_JELZŐ} [SORREND] {KIEMELT_JELZŐ}

Pl.: *akármelyik szélső, második tegnap érkezett, századik*

KIEMELT_JELZŐ::=TULAJDONSÁGJELZŐ |
MELLÉKNÉVI_IGEI_SZERKEZET

A kiemelt jelző elvileg bármilyen jelző, de leggyakrabban többtagú részszerkezet (pl. melléknévi igenév a vonzatával). Van egy fajta szerkezet, amely viszont csak kiemelt lehet. Ez a melléknévi igei szerkezet.

A melléknévi igei szerkezetnek két fajtája van. Az egyik egy olyan múlt idejű tárgyas ragozású egyes szám harmadik személyű ige, melyet mindenképpen megelőz az ige alanya, és az ige tárgya maga a névszói kifejezésben jelölt objektum(ok). (Ez úgy formailag, mint értelmezés szempontjából más, mint a múlt idejű melléknévi igenév.) A másik fajtája ritkábban fordul elő. Az ige egyes számban, leggyakrabban de nem kizárólagosan felszólító módban szerepel, melyet valamely vonzata megelőzhet, esetleg követhet.

MELLÉKNÉVI IGEI SZERKEZET::=
NÉVSZÓI_KIFEJEZÉS(Alanyeset 3.személyű)
IGE(múlt idő egy.sz 3 sz tárgyas ragozás) |
^IGE(egy.sz. 1/3 sz. tárgyas ragozás)^

Pl.: Élve *a demokratikus törvények adta lehetőségekkel...*
Amolyan tedd-ide tedd-oda kislány.

SORREND:=Sorszám {pozícionáló_melléknév} |
pozícionáló_melléknév

A pozicionáló melléknév olyan melléknév, amelynek felsőfoka - ha van - a *-bb* nélkül képződik.

Pl.: *a legszélső három fa*

A vonatkozás vagy egy birtokos, vagy egy ismételt ragozású mutató névmás, vagy egy névelő, pontosabban:

VONATKOZÁS::=[Mutató_névmás+Többes számjel+Birtokos jel+Rag]
[Névutó] Határozott_névelő |
BIRTOKOS(birtok) |
[BIRTOKOS(birtok)] Kvantor

Ha a kifejezés élén egy mutató névmás áll, az utána jövő ragnak, illetve többes szám- és birtokos jelnek meg kell egyeznie az egész kifejezés végén levő raggal, jellel. A névutó csak akkor szerepelhet, ha ugyanez az alanyesetet vonzó névutó állt a névszói szerkezet után és nem volt birtokjele.

Pl.: *azt az embert, azok az emberek, azé az emberé, az után az ember után, azoké az embereké, azokét az emberekét, azok után az emberek után, azoké az embereké után, azokkal az emberekkel szemben stb.*

A második szerkezet csak birtokjeles kifejezéseknél szerepelhet. Természetesen a mutató névmással kezdődő névszói szerkezetnek is lehet birtokosa, de az mindig *nak*-ragos, így számunkra az már egy másik névszói kifejezés.

A mondatban a birtokviszonynak három megjelenési formája van. Mindhárom formában a birtoknak birtokjele van. A birtokos lehet implicit, *nak*-ragos, illetve alanyesetű.

Pl.: *az eszem , Pista <= kutyája,*
Feri édesanyjának <= testvére

Az implicit birtokosra a mondatban csak a birtokvégi birtokjel utal, és semmi szerkezeti változást nem okoz. A *nak*-ragos birtokos esetén a birtokos a mondatban "elmászhat" a birtoktól, ezért a névszói szerkezetek elemzésénél független névszói szerkezetnek tekintjük őket, és később, a mondatszerkezet elemzésénél párosítjuk őket. Csak az alanyesetű birtokviszonnyal foglalkozunk, mert ekkor a birtokos mindig közvetlen megelőzi a birtokot. A birtokos illeszkedés elég jól meghatározható. A vonzatot a birtok birtokjele határozza meg:

```
BIRTOKOS(birtok)::= Határozott_névelő
                        Személyes_névmás(birtok vonzata) |
                    NÉVSZÓI_SZERKEZET(Alanyeset)
```

Ha személyes névmással jelöljük a birtokost, akkor a határozott névelő használata kötelező. Névszói szerkezet csak harmadik személyű birtoknál lehetséges. A birtokkal való illeszkedést a következő táblázat mutatja:

birtok birtokjele :	Esz.1.sz	Esz.2.sz	Esz.3.sz
birtokos :	*én*	*te*	*ő*/bármi más
birtok birtokjele :	Tsz.1.sz	Tsz.2.sz	Tsz.1.sz
birtokos :	mi	ti	ő

A *saját* szó bármelyik névmást pótolja.

Pl.: *Pista és Mari kutyája, az én munkám, az ő házuk, a magam igaza, a saját kalapom* stb.

Az irodalomban sokszor vitatott kérdés, hogy a birtokos előtt álló névelő a birtokhoz vagy a birtokoshoz tartozik-e. A mi véleményünk szerint tartozhat a birtokhoz is, a birtokoshoz is, de mindkettőhöz is. Ennek részletesebb kifejtését lásd az **elnyelődés**ről szóló fejezetben.

Pl. *a kutya farka = a kutyának a farka*

A kvantor szóhasználata, mint említettük, nálunk lényegesen különbözik a logikusok és a nyelvészek által használttól. Mi pragmatikusan azokat a névelőtől és vonatkozó névmástól különböző szavakat soroltuk be ebbe a szófajba, melyek kizárólag a névszói kifejezés elején szerepelhetnek, és amelyek utáni rész formálisan csak egyes számban lehet.

A kötőszavak

A névszói szerkezeteken belül gyakran használunk kötőszavakat, amelyek azonos funkciójú részeket kötnek össze, így például egy ige több tárgyát vagy egy főnév jelzőit. Kötőszavak lehetnek mondatokat, illetve mondatrészeket összekötő kötőszavak. A jelen fejezetben minket csak a mondatrészeket összekötő, tehát az egyszerű mondatokban is előforduló kötőszavak érdekelnek.

A kötőszavakkal nem csak halmozott névszói szerkezeteket lehet előállítani, hanem a fenti szintakszisban szereplő szintagmák többségét halmozhatjuk. Ez természetesnek tűnik a különböző jellegű jelzőknél, de a beszédben előfordul, hogy csak a névutó, esetleg a rag van halmozva.

Pl. *Ebéd előtt és után*,
? *Tanári értekezlet a diákok fejlődéséről és -ért*.

A szintakszis szempontjából ötféle mondatrész-kötőszót különböztetek meg:

1. Felsoroló (ide tartozik a vessző is): *és s meg avagy* ...
   Pl.: *Péter, Pál és Endre*.
   Ezek az összekapcsolandó részek közé kerülnek.
2. Ismétlődő: *vagy akár mind* ...
   Pl.: *Akár Péter, akár Pál. Akár Pál is megteheti*.
   Ezeket mindegyik összekapcsolt mondatrész elé ki kell írni.
3-4. Páros: *nemcsak – hanem* ...
   Pl.: *nemcsak Pétert, hanem Pált is*.
   Itt megkülönböztetendő az először és a másodszor szereplő kötőszó.
5. Követő: *se sem is* ...
   Pl.: *Péter se, Pál se. Színét sem láttam*.
   Ezek vagy mindegyik összekapcsolt mondatrész után állnak, vagy csak az utolsó után.

Egy szó egyszerre több, különböző osztályba tartozó kötőszónak lehet az alakja. Például a vagy alak lehet 1-es és 2-es, a se lehet 2-es és 5-ös típusú is.

Egy felsoroláson belül minden egyes összekötött mondatrészhez ki kell tenni a megfelelő kötőszót. (Természetesen az és-típusúnál a legelső egység elé nem.) Ebben a részben nem foglalkozunk azzal, hogy a kötőszavak milyen mondatbeli szerkezeteket hoznak létre. Ennek az az oka,

hogy nem tártuk fel eléggé a nyelvtanilag elfogadható halmozott szerkezeteket. Ehelyett egy minden bizonnyal túlságosan leegyszerűsített szintakszissal rendelkező szerkezetet fogadunk el.

```
HSZ:={HSZ KövetőKsz VESSZŐ} HSZ KövetőKsz |
     {IsmétlődőKsz HSZ ,} IsmétlődőKsz HSZ |
     Pár1Ksz EgyszerűSzintagma Pár2Ksz EgyszerűSzintagma|
     {EgyszerűSzintagma FelsorolóKsz } EgyszerűSzintagma
VESSZŐ , | FelsorolóKsz
```

Kiegészítésként megjegyezzük, hogy az egy halmozott szintagmában szereplő részkifejezések a teljes kifejezésben ugyanazt a szerepet kell, hogy betöltsék. Ha például a halmozott szintagma egy ige tárgyát jelöli, akkor annak minden része egy-egy tárgyesetű névszói kifejezés. Ez az elemzésnél egy lényeges támaszpont. Ily módon az a kifejezés hogy a pirosat és a kék zászlót egyértelmű, míg a piros és kék zászló kifejezés esetén nem tudjuk, hogy egy kétszínű zászlóról van szó, vagy pedig két darab egyszínűről.

A másik kiegészítés a fent megadott szintakszishoz, hogy a szabály egy sorában megadott kötőszavaknak meg kell felelniük egymásnak. Ily módon a követő és az ismétlődő kötőszavaknak azonosaknak kell lenniük egy halmazatban, a páros kötőszavaknak pedig meg kell felelniük egymással. A felsoroló kötőszavaknál megengedjük a keveredést, mivel nincs egyelőre megoldásunk arra, hogyan lehet ezeket strukturálni:

(*Péter vagy Pál*) *és Feri* <--> *Péter vagy* (*Pál és Feri*)

Szerencsére a gyakorlatban ilyen szerkezetek nem fordultak elő, magunk is mesterkéltnek érezzük.

A névszói kifejezés elemzésének stratégiája

Az alapvető probléma az, hogy egyrészt a morfológia többértelmű analízist szolgáltat a szóalakokról, másrészt a fent említett szerkezet elég bonyolult. Egy ilyen rekurzív, esetleges felbontású szerkezet elemzésénél felléphet a kombinatorikus robbanás, vagyis az elemzés időigénye ugrásszerűen megnőhet. A lehetőségek ész nélküli próbálgatása olyan időigényes lenne, ami nem felelne meg a rendszer végcéljának: kiszolgálni egy interaktív rendszert. A stratégia ezért az, hogy ha egy hipotetikus elemzési út jónak bizonyul, akkor nem próbálkozunk másik lehetőséggel.

Ez a módszer csak akkor hatékony, ha az analízis során aránylag gyorsan kizárja a lehetetlen utakat, és helyenként megjelöli az esetleges tévedési lehetőségeket későbbi korrekciók érdekében. Másrészt az elemzés során az alternatív utakat olyan prioritási sorrendbe rakjuk, melyek - sikeres részelemzés esetén - kizárják más alternatívák valószínűsítését.

Az elemzés taktikája

Az elemzés kiinduló adata egy mondat szavaiból (és írásjeleiből) álló lista, valamint egy vonzatkeret-lista. Mindegyik szóhoz hozzátartozik a szó egy vagy több morfológiai felbontása, melyet a morfológiai elemző állít elő. Egy morfológiai alternatíva tartalmazza a szótő alakját, szófajának kódját, a hozzákapcsolódó jelek és ragok kódját. (Képzett szavak esetén az alapszótő és a hozzákapcsolódó képzők is adva vannak, de ezt csak a szó vonzatainak meghatározásához használjuk fel.) A vonzatkeret-lista alternatívákat tartalmaz. Egy alternatíva olyan vonzatokat sorol fel, amelyek egyszerre előfordulhatnak a mondatban. Ezek ismeretében a következőket tesszük:

1. A mondatot jobbról balra elemezzük. Ennek az az oka, hogy egy kifejezésnek mindig a legjobboldalibb szava határozza meg, hogy milyen vonzatot elégít ki, így valószínűleg egy legfelső szintű névszói kifejezéssel találkozunk először, és ennek elemzése során a beágyazott kifejezéseket automatikusan beolvasztjuk.

2. Ha a soron következő szóformának van olyan elemzése, ami kielégít egy vonzatot (pl. megfelelő raggal rendelkező szó, megfelelő névutó vagy határozószó), akkor ebből a szóból kiindulva megkeressük a hozzátartozó névszói kifejezést. Ezt felfűzzük a vonzatkeret-lista azon alternatíváiba, ahol szerepel a vonzat. A továbbiakban ezeket az alternatívákat vizsgáljuk először.

3. A névszói kifejezésen belül a leírt szintakszis alapján keressük a megfelelő szófajú (és ragtalan) alakokat. Ha a soron következő szónak vannak vonzatai, azt rögtön elkezdjük keresni, és ha találtunk, hozzákapcsoljuk a szóhoz. (Ez az elemzés "falánk, direkt" módszere)

4. Megjelöljük azokat a helyeket, ahol esetleg széteshet a névszói kifejezés két független részre.

5. Halmozott névszói kifejezésnél az összes alkotórésznek ugyanazt a vonzatot kell kielégítenie. Ha az egyik alternatíva vonzata bővebb, a másiké szűkebb (p.: HOVA és *-ra/-re*), akkor a szűkebb vonzatot tartalmazó alternatíva kieshet, ha különböző ragok vagy névutók vannak összekapcsolva.

A névszói kifejezés elemzése

Az elemzés jobbról balra a fent leírt struktúra alapján "falánk, direkt" módszerrel történik.

- A "falánk" azt jelenti, hogy az algoritmus összekapcsol mindent, amit lehet.
- "Direkt" olyan értelemben, hogy a jelen algoritmusban nem használunk sem visszalépést, és az alternatívák keresése is korlátozott.

Ahhoz, hogy egy ilyen algoritmus aránylag jól működjék, szükséges, hogy az alternatívák közül azt próbáljuk ki először, amelyikről hamar kiderülhet, ha helytelen. A fent leírt formalizmusban az alternatívák ennek megfelelő sorrendben vannak meghatározva. Például ha egy szó névutó is lehet és határozószó is, akkor először névutóként vizsgáljuk és megvizsgáljuk, hogy illeszkedik-e hozzá az előtte levő szó. Ha igen, akkor névutónak vesszük, ha nem, határozószóként vizsgáljuk.

Tapasztalatunk szerint az algoritmus falánksága miatt viszonylag gyakran bekebelezi a névszói kifejezések elején álló, nem oda tartozó másik névszói kifejezést. Ez azokban az esetekben fordul elő, ahol a mondat szintaktikusan többértelmű. Ilyenkor az elemzés csak egyet talál meg, és ez lehet, hogy szemantikailag nem is megfelelő. Ezért azokat a helyeket, ahol elvileg bevégződhet a névszói szerkezet megjelöljük (a példában függőleges vonallal). Ilyenek első sorban a szavak opcionális vonzatainál fordulnak elő.

Például többértelmű a következő mondat, melyet az algoritmusunk félreértelmez:

(((*Budapesten*)| *közismert színészekkel*)| *forgatott filmeket*)| *mutattak be*.

Más jellegű tévesztést tapasztaltunk a belső struktúra pontos feltérképezésénél. A következő példában a *három*-at a kombináció mennyiségének elemezte, és nem a birtokosának:

(*a három kombinációja**)|

Mindennek ellenére úgy tűnik, hogy egy pontosabb morfológiai elemzés, valamint a jelzőknél és a vonzatoknál egy tervezett minimális szószemantika felhasználása a problémák nagyobb hányadát megoldja.

A kötőszavas szerkezetek elemzésénél, azon kívül, hogy figyelembe vesszük, mely kötőszavak milyen pozícióban szerepelhetnek, lényeges, hogy azonos vonzatot kielégítő névszói kifejezéseket illetve azonos funkciójú szintaktikai egységet fűzzenek össze. Más megszorítás nincs.

Problémák és hiányosságok

1. Nincs kifejtve az EGYSZERŰ_JELZŐ nyelvtani egység. A mértékegységek előtt szerepelhető jelzők nem gyakoriak, és nem is lehet minden mértékegységnek jelzője.

Pl.: *egy bő liter tej*
két horpadt hordó káposzta

de nem jó * *egy görbe kanál orvosság*.

2. Ugyanúgy nem világos, mikor lehet egy jelző kiemelve. Így tűnik, az egyszerű melléknevek nem mindig kiemelhetőek, míg a múlt idejű, vonzattal rendelkező melléknévi igenevek minden további nélkül.

Pl.: *a tegnap érkezett három hordó sör*,
de ? *a habos három korsó sör*

3. A határozott névelő hatáskörével külön fejezet foglalkozik.

4. Nem megoldott a kötőszavas szerkezetek meghatározása. Két gyakorlati problémával találkoztunk. Az egyik az olyan összetett kötőszavas struktúrák, melyekben többféle felsoroló kötőszó szerepel. Ennek előfordulása rendkívül ritka, de szemantikai elemzés nélkül megoldhatatlan probléma. Gyakoribb előfordulású az az eset, amikor egy alanyesetű névszói szerkezet halmozott jelzővel rendelkezik, illetve az ettől nehezen megkülönböztethető alanyesetű halmozott névszói kifejezés.

5. Az egyes struktúrák a gyakorlatban nem lehetnek tetszőleges mélységbe beágyazva. A gyakorlat azt mutatja, hogy a kifejezések egy bizonyos bonyolultság után szétesnek, helytelenné válnak. A határt nem tudjuk meghúzni, csupán néhány korlátot tudunk adni.

AZ EGYSZERŰ MAGYAR MONDATOK SZERKEZETE

Az egyszerű (nem összetett) magyar mondatok felépítésében kétféle szabály játszik szerepet: a konfigurációs szabályok azt írják elő, hogy egy adott nyelvtani szerkezet alkotórészei milyen sorrendben állnak egymás után a mondatban; a másik szabálytípusban ragok jelölik ki a nyelvtani szerkezet elemeit. A magyar mondatokban a kétféle szabály körülbelül azonos súllyal szerepel, és az esetek túlnyomó részében a kétféle szabály élesen elválik egymástól.

Azokban a nyelvtani szerkezetekben, amelyeket konfigurációs szabályok írnak le, a szavak egyetlen összefüggő sorozatot alkotnak, és nem kapnak olyan végződést, ami összekapcsolná őket. A szó szerepét pusztán a pozíciója határozza meg, de természetesen nem kell a szerkezet minden összetevőjét kitenni.

Azokban a szerkezetekben viszont, amelyek összetartozását a ragok biztosítják, a komponensek szétszóródnak a mondatban, és sorrendjüket nem nyelvtani szabályok szabják meg, hanem az, hogy mit hangsúlyozunk a mondatban. A komponensek kijelölése természetesen nem csak raggal történik: névutós szerkezetek, határozószók, névmások is betölthetik a megfelelő szerepet.

Az állítmány

Ragozott ige

A magyar mondatok több mint felében egy ragozott ige áll a mondat központjában, és ehhez kapcsolódnak a további mondatrészek: az alany, a tárgy és különböző határozók. Ezek kijelölése ragokkal (és névutókkal) történik. Viszont az esetek egy részében az ige nem áll egyedül, hanem előtte egy vagy két módosítószó áll, illetve mögötte egy módosító *volna* szócska állhat. Azaz formálisan valahogy így írhatjuk le:

```
IGEI KIFEJEZÉS -> [hadd] [ne] IGE |
                  [IGEMÓDOSÍTÓ] IGE [volna]
IGEMÓDOSÍTÓ -> nem |nemigen |aligha | ...
```

Például: *Hadd ne menjek oda.*
Oda hadd ne menjek.
Nem Péter ment volna oda.
Oda nemigen ment volna Péter.

Látható, hogy az igei kifejezés elemei mindig együtt maradnak, és sorrendjük meghatározott. Ezzel szemben az igéhez kapcsolódó bővítmények (az alany vagy a hova kérdésre válaszoló helyhatározó) többé-kevésbé akárhol állhat a mondatban.

Ha az embereket megkérdezzük, hogy hogyan néz ki egy magyar mondat szerkezete, rendszerint azzal kezdik, hogy a mondat állítmánya egy ige. Ez nem véletlen, hiszen, noha felméréseink azt mutatják, hogy csak a mondatok 50-60 százalékában van ragozott ige, további 25-30 százalékban az állítmány valamilyen igeneves szerkezet, és ezek is több-kevesebb igei tulajdonságot mutatnak. Az igenevekkel egy külön fejezetben is foglakozunk.

Az igenevek közül a leggyakoribb a főnévi igenév, melynek nevével ellentétben semmilyen névszói tulajdonsága nincs, nem szokott olyan szerepet betölteni a mondatban, mint amit egy főnév szokott. Ezzel szemben igen gyakran tölti be az állítmány szerepét, de sohasem egymagában, hanem mindig egy nagyobb szerkezet részeként.

A főnévi igenévnek két formája van: a személyragozott főnévi igenév és a ragozatlan. Azokban a szerkezetekben, ahol ragozott főnévi igenév állhat, ott állhat ragozatlan is, de vannak olyanok is, amelyben csak ragozatlan állhat.

Segédige ragozatlan főnévi igenévvel

Segédigének nevezünk egy sor olyan igét, amelyet mindig főnévi igenévvel használunk. A segédige és az ige alanya mindig azonos, és a segédigének semmilyen önálló vonzata nincs. Ilyenkor a mondat a főnévi igenév vonzataiból épül fel, és megítélésünk szerint ez az ige lesz a mondat állítmánya, a segédige pedig az igének valamilyen általánosabb értelemben vett módját vagy az idejét adja meg. A segédige szeret az igekötő és az ige között állni, de helyéről semmi biztosat nem mondhatunk.

A segédigés szerkezet is kaphat igemódosítókat. Ezek a segédige körül csoportosulnak olyan formában, ahogy azt az igéknél leírtuk.

Pillanatnyilag a következő szavakat tartjuk segédigének: *akar, fog, kezd, kíván, próbál, szeret, szokik, talál, tetszik, tud*.

Példák:

meg akarom mondani	szándék
el fogom mondani	jövő idő
járni kezd	indulás
el kívánom érni	szándék
meg próbálta oldani	szándék
át szeretnék menni	szándék
fel szoktam öltözni	gyakorítás
be talált menni	véletlen
ide tetszik adni	udvariasság
meg tudjuk oldani	képesség

Felmerül a kérdés, hogy mi szükség van a segédigék bevezetésére. A bevezetés indokait a szemantikában kell keresni. Megítélésünk szerint az igei állítmányú mondatok központjában álló igei kifejezéshez egy olyan általánosított módot lehet hozzárendelni, amelyet az igerag, bizonyos képzők (*-hat/-het*, *-gat/-get*, *-tat/-tet* stb.), az igei módosító szavak és különböző segédigék adnak meg. A felszólító módot nem csak felszólító móddal lehet kifejezni, hanem néhány segédige felszólító módjával is. Például: *Kezdd megoldani a feladatot. Próbáld megoldani a feladatot. Tessék megoldani a feladatot. = Oldd meg a feladatot*. Hasonló a helyzet a szándékkal, képességgel, kötelességgel stb.

A segédige általában nem teljes ragozású, illetve bizonyos módokban vagy időkben mást fejez ki. Például: a *fog* ige nem lehet múlt idejű vagy felszólító módú. A *talál*j segédige nemigen áll meg tagadás nélkül. Például: *Be ne találj menni.*, *Nehogy be találj menni.*, de: **Be találj menni.*, vagy **Találj bemenni*! A *szeret* segédigének csak a feltételes módja fejez ki szándékot. Kijelentő módban valami olyasmit jelent, mint a *szokott*. Annak részletes kidolgozása, hogy melyik segédige, milyen módban, milyen módosítószavakkal, mit fejez ki, még hátra van. Minden ilyen segédigéhez tartozik egy vagy több azonos alakú ige, ami nem segédige, hanem mindenféle vonzatai vannak.

Lásd még az **igenevek**ről szóló részt is. Továbbá Kálmán C. Gy. - Kálmán L. - Nádasdy Á. - Prószéky G.: A magyar segédigék rendszere. Ez utóbbi kissé más kritériumokból indul ki, más szempontokat is vizsgál (pl. a prozódiát), de ugyanezeket a szavakat tartja segédigének.

Segédige ragozott főnévi igenévvel

Vannak olyan mondatok, amelyeket az előző pontban leírtakkal analóg módon szintén segédigés szerkezetnek tekintünk, de felépítésük erősen eltér az előzőektől. Az *illik*, *kell*, *lehet* és *sikerül* igék mellett személyragozott főnévi igenév állhat. A főnévi igenév kifejezi a logikai alany személyét és számát, de nem fejezi ki a cselekvés idejét és módját. Ezt a igemódosítókkal együtt a segédige adja meg, amely mindig alanyi ragozású és egyes szám harmadik személyben áll. Ha a mondatban ki van téve a logikai alany is, akkor ez *-nak/-nek* ragot kap. Mi ezt a szót tekintjük a mondat alanyának, és *-nak*-ragos alanynak hívjuk. Nem ez az egyetlen mondattípus, amiben *-nak*-ragos alany áll. *Nak*-ragos alany mellett biztosan nem áll alanyesetű alany.

Példák: *Pistának kell odamenni.*
Pistának kell odamennie.
Nekem kell megoldanom.
Meg kell oldanom.

Noha csak négy ilyen ige van, nagyobbrészt a kell, kisebb részt a lehet ige teszi ki igeneves állítmányi szerkezetek zömét, legalább 60-70 százalékát.

Melléknév + főnévi igenév, mint állítmány

Hasonlóan a "segédige + főnévi igenév" alakú állítmányhoz a "melléknév + főnévi igenév" alakú állítmánynak is két formája van.

A *valaki valamilyen valamit megcsinálni* mintát követik az alkalmas, képes, köteles stb. melléknevet tartalmazó mondatok.

Például: *Péter képes megoldani a feladatot.*
Képes voltam megoldani a feladatot.
Képesek lennének megoldani a feladatot.
stb.

A *valakinek valamilyen valamit megcsinálni* mintát követik a *muszáj*, *szabad*, *tanácsos*, *tilos* stb. mellénevet tartalmazó mondatok.

Például: *Péternek tanácsos megoldania a feladatot.*
Tanácsos volt megoldanom a feladatot.
Nekik tanácsos lenne megoldani a feladatot.
stb.

Az első mintához tartozó mondatok szerkezetükben nagyon hasonlóak a ragozatlan főnévi igenevet tartalmazó segédigés mondatokhoz. Ezekben a mondatokban is csak ragozatlan főnévi igenév állhat, a főnévi igenév és a melléknévi állítmány alanya azonos. Ha az állítmány nem jelen idejű, kijelentő módú, és nem harmadik személyű alanyra vonatkozik, akkor a melléknév elé vagy mögé a létige megfelelő alakja kerül. Az igemódosító a létige illetve annak hiányában melléknév elé kerül.

A másik mondatmintában a másik segédigés szerkezethez hasonlóan ragozott főnévi igenév és *nak*-ragos alany áll. Az ige személyét és számát az igerag adja meg, idejét és módját pedig a létige egyes szám harmadik személyű alakja: a *volt, lesz, volna, lenne, legyen* szó adja meg. Ez vagy a melléknév előtt, vagy mögötte áll. Ha a kifejezés igemódosítót is tartalmaz az vagy a melléknév, vagy a fenti szavak előtt áll.

Van néhány melléknév, ami sokkal gyakrabban áll állítmányként főnévi igenévvel, mint jelzőként, és a velük alkotott kifejezések jelentése is közel áll bizonyos igei segédigés kifejezésekhez: ezeket a szavakat nem igei alakú segédigéknek nevezzük, és a szemantikus ábrázolásban szintén általánosított móddá alakítjuk át. Ezek a szavak a következők:

alkalmas, alkalmatlan, képes, kész, köteles
ajánlatos, célszerű, érdemes, felesleges, lehetetlen, szabad, tanácsos, tilos.

(Az első csoport ragozatlan, a második ragozott főnévi igenévvel áll.)

Mi azonos jelentésűeknek tekintjük a következő mondatokat:

Nem lehetett elkészíteni. - *Lehetetlen volt elkészíteni.*
Meg tudom oldani. - *Képes vagyok megoldani.*
Pista kész odamenni. - *Pista oda akar menni.*
Nekünk kell észrevenni. - *Muszáj észrevennünk.*
Ajánlatos észrevennünk. - *Célszerű észrevennünk.*

Természetesen nem csak a fenti melléknevekhez lehet főnévi igenevet kapcsolni. Vannak ugyan olyan melléknevek, például az *abszolút*, *bemeneti*, *ellenkező*, *zalai* stb., amelyek nemigen kapcsolhatók össze főnévi igenévvel.

Sok melléknévből lehet *Valami valamilyen valamit csinálni* alakú mondatot alkotni, de ezeket a mondatokat melléknévi állítmányú mondatokként kezeljük, ha a melléknév nem a fent felsorolt szavak valamelyike. Ilyenkor a főnévi igenevet a melléknév vonzatának tekintjük. Pl.:

A cikk túl rövid ezt a témát áttekinteni.
Péter elég magas levenni a könyvet a polcról.

A legtöbb melléknévből készíthető *Valami valamilyen valakinek valamit csinálni* mintát követő mondat. Itt sem az igenév az állítmány, hanem a melléknév. A melléknévnek egy ragozott melléknévi igenév a vonzata, és annak a *nak*-ragos mondatrész az alanya. Pl.:

A polc túl magas Pistának levenni a könyveket.
A barack még túl zöld megennünk.

Sajnos ezt a két mondatmintát igen nehéz megkülönböztetni egymástól. Az *A barack még túl zöld megenni*. mondat látszólag az első típushoz tartozik, de valójában a második mintát követi, csak hiányzik a *nak*-ragos alany, illetve az azt pótló személyrag. Bizonyos melléknevek inkább az első, mások inkább a második mintát szeretik; a leggyakrabban használt melléknevek viszont egyaránt fordulnak elő mindkét típusban. (Mint például a magas melléknév a fenti mondatokban.)

Egyebek

A további két mondat-mintában áll még főnévi igenév. *A Van mit ennie. Nincs hol laknia. Lesz hova mennie. Péternek nem volt kivel beszélgetnie*. stb. fazonú mondatokban a kérdőszóval megjelölt mondatrész léte vagy nem-léte az elsődleges állítás. Viszont a szemantikai reprezentáció olyan ábrázolássá alakítja, ahol az igenév áll a mondat közepén, és a kérdőszóval megadott argumentum valamilyen speciális elemmel van jelölve. A logikai-szemantikai leírásban például ez az argumentum egy olyan változó, ami egzisztenciális kvantorral van lekötve. (Érdekes, hogy a negatív mondatok orosz megfelelőiben, ragozott ige áll a mondat középpontjában és a megfelelő argumentum bizonyos speciális

névmásokkal - pl. *nekuda*, *negde* stb. - van megjelölve.) Mi tehát egy olyon szintakszisfát építünk fel, aminek a középpontjában a ragozott főnévi igenév áll, és egyik argumentuma a "*van*+kérdőszó" illetve "*nincs*+kérdőszó" kötött szórendű kifejezés. Az alanyt *-nak* rag jelöli, mint a ragozott főnévi igenév mellett mindig.

Az utolsó idetartozó mondat-minta: "*valamit csinálni + volt/van/lesz/volna + valaki*". Mindig ragozatlan főnévi igenévvel áll. Nem világos előttünk, hogy a *Péter tegnap a kertben volt sétálni.* és a *Péter tegnap a kertben sétált.* mondatok jelentése között van-e valami különbség, és az sem világos, mikor kell az egyiket, és mikor a másikat használni. Egyelőre úgy tekintjük, hogy az igenév a mondat főigéje, és a létige fejezi ki az időt, a módot, a személyt, a számot.

(Valami homályos kapcsolat van a ragozott melléknévi igenév és a felszólító mód bizonyos használata között. A *Nincs hol laknia.* mondat helyett azt is használják, hogy *Nincs, ahol lakjon.* A *Meg kell látogatnom.* helyett azt is használják, hogy *Meg kell, hogy látogassam.* sőt *Meg kell látogassam.* Mi ezeket az alakokat nem használjuk, de mivel a nyelvet a lehető legtágabban akarjuk érteni, valószínűleg kezelni fogjuk.)

Jelen idejű (folyamatos) melléknévi igenévi állítmány

A jelen idejű melléknévi igenév erősen melléknévi jellegű, és igen kevéssé igei jellegű. Az *-ó/-ő* végű melléknévi igenevek eleve igen ritkán állnak állítmányként, amikor pedig állítmányként áll, az igei argumentumok rendszerint megelőzik, mint a névszói kifejezésben álló jelzők esetén.

Érvelésed rossz feltételezéseken alapuló.
**Érvelésed alapuló rossz feltételezéseken.*
A versenyző erőtől duzzadó volt.
?A versenyző duzzadó volt az erőtől.
A téma sokakat érdeklő lesz.
**A téma érdeklő lesz sokakat.*

Általában nem igaz, hogy az állítmányként szereplő melléknév vonzata csak a melléknév előtt állhat. Ez tapasztalatunk szerint csak ebben az egyetlen esetben igaz, és minél melléknevebb a jelen idejű melléknévi igenév, annál kevésbé igaz rá. Pl.:

A vélemények eltérők egymástól.

Az Alma eladó, Szoba kiadó stb. mondatokban az *eladó*, *kiadó* szavak látszólag szenvedő jelen idejű melléknévi igenevek. Ezek az általunk felállított szabályok szerint nem vezethetők le, sem szintaktikailag, sem szemantikailag nem kezelhetők. Mi úgy tekintjük, hogy ezek nem jelen idejű melléknévi igenevek, hanem olyan szavak, amelyek valódi jelentése *kapható*, illetve *bérelhető*. Már a

morfológiai analízis során bevezetünk egy olyan alternatívát, hogy az *eladó* szó nem más, mint *kap+ható*. Ilyesmit nem csak ebben az esetben teszünk, például az *engem* szót is átírjuk az *én+t* alakra.

"Szenvedő" mondatok

A "szenvedő" mondat, azt fejezi ki, hogy az alanyesetű mondatrész valamilyen cselekvés (vagy történés) következtében az ennek megfelelő állapotba jutott. (Vagy fog jutni, vagy juthat stb.) A mondatnak két formája van. Az egyikben múlt idejű melléknévi igenév, a másikban határozói igenév áll.

A hús kissé fagyott.
A hús kissé meg van fagyva.

A eredmény többszörösen ellenőrzött volt.
A eredmény többszörösen volt ellenőrizve.

A ruhája összepiszkolódott volt.
A ruhája össze volt piszkolódva.

A múlt idejű (befejezett) melléknévi igenév tagadó párja az *-atlan/-etlen*.

Ez a telek még beépítetlen.
Ez a telek még nincs beépítve.

A mondat centrumában egy igenév áll. Ez az igenév csak befejezett szemléletű lehet, ezért a határozói igeneves alak gyakran kap olyan igekötőt, amire a befejezett szemléletű melléknévi igenévnek nincs szüksége. Az állapot bekövetkezésének idejét illetve módját a *van, volt, lesz, legyen, lenne, volna*, stb. létige adja meg. A *van*-t a melléknévi igenév mellett (mint általában a melléknévi állítmány mellett) soha nem tesszük ki, a határozói igenév mellet viszont többnyire kitesszük, bár nem mindig. A létige vagy az igenév előtt, vagy mögötte áll, illetve szeret az igekötő és a határozói igenév között állni. (A múlt idejű melléknévi igenév igekötője nem válik le.)

Ha azt vizsgáljuk, hogy mi vagy ki volt az, aki abba a bizonyos állapotba jutott, akkor különbséget kell tennünk a tárgyas és tárgyatlan igék között. A tárgyas igék esetén mindig az ige tárgya az, a tárgyatlan igék esetén viszont az alany. Tárgyas igék esetén az ige végrehajtóját (a mi felfogásunkban a logikai alanyát) az által névutóval adjuk meg. Az ige többi vonzatát ugyanúgy jelöljük, mint a ragozott ige esetén, és szétszóródnak a mondatban.

Nem világos mikor lehet egy bizonyos igével egyik vagy másik típusú szenvedő mondatot alkotni.

Ez a sör importált.
**Ez a sör importálva van.*
Az ember már halott volt.
**Az ember már meg volt halva.*
**A kiállított tárgyak eladottak.*
A kiállított tárgyak el vannak adva.
**A sebe még nem begyógyult.*
A sebe még nincs begyógyulva.

"*-ható/-hető*" és "*-andó/-endő*" állítmányú mondatok

Míg a jelen idejű vagy múlt idejű melléknévi igenévi állítmányt körülbelül ugyanannyi indokkal lehet igeinek vagy melléknévinek tekinteni, (lásd **igei és névszói relációk**) a *-ható/-hető* és *-andó/-endő* állítmányú mondatokat szemantikai okból biztos, hogy célszerűbb az igei állítmányok közé sorolni. Mi ezeket az alakokat a segédigék kapcsán kifejtett általánosított mód egy lehetséges kifejezési formájának tekintjük.

A feladat sürgősen megoldandó.
A feladatot sürgősen meg kell oldani.
A feladat könnyen megoldható.
A feladatot könnyen meg lehet oldani.

A *-ható/-hető* tagadó párja a *-hatatlan/-hetetlen*.

A feladat megoldhatatlan.
A feladatot nem lehet megoldani.

Ilyen melléknévi igenév csak tárgyas igéből képezhető. Az alanyesetű mondatrész az ige tárgya, ezen kell vagy lehet az igében leírt akciót végrehajtani. Az ige idejét és módját, amit az igeraggal szoktunk megadni, az igenév előtt vagy mögött álló létige adja meg. Az alany az által vagy a számára névutóval, esetleg a *-nak* raggal adható meg. (Bár ezt a mondatformát éppen azért szokták alkalmazni, hogy ne kelljen megmondani, ki hajtja végre az akciót.)

Péter által könnyen megoldható volt a feladat.
Péter számára könnyen megoldható volt a feladat.
Péternek könnyen megoldható volt a feladat.
A feladat sürgősen megoldandó Péter által.
Péter számára a feladat sürgősen megoldandó.
Péternek sürgősen megoldandó a feladat.

Nem világos, mikor melyiket kell használni, és milyen szórendben. (Mivel a nyelvet igyekszünk a lehető legtágabban érteni, és mivel a hivatalos papírok szeretik a passzív szerkezetet, mindegyiket elfogadjuk.)

Az igei állítmányhoz kapcsolódó mondatrészek

Az igéhez (függetlenül attól, hogy állítmány-e vagy sem) alany, tárgy és különböző határozók kapcsolódnak. Mi ezeket többféle szempont szerint vizsgáljuk. Egy az ittenitől kissé eltérő szemléletű osztályozást találhatunk a **Vonzatok** című fejezetben.

Az igéhez kapcsolódó mondatrészeket vonzatokra és szabad határozókra osztjuk. Durván szólva azt mondhatjuk, hogy vonzatnak azokat a mondatrészeket nevezzük, amelyek szorosan az ige jelentéséhez kapcsolódnak, szabad határozóknak pedig azokat, amelyek úgy általában a külső körülményeket írják le. Egy kissé pontosabban szólva igen sok olyan ige van, amelynek nemigen beszélhetünk a jelentéséről anélkül, hogy tudnánk, hogy milyen mondatrészek kapcsolódnak hozzá. Pl.:

áll (*nem megy*)
áll+*valahol* (*ott van*)
áll+*valamiből* (*van valamiből*)
áll+*valahova* (*odamegy*)
áll+*őrt*/*bosszút*/*sarat*/*igéretet*...
.
.

Az *alapul* igét sem nagyon képzelhetjük el anélkül, hogy meg ne mondanánk, hogy *min*; a *bővít* igét sem anélkül, hogy meg ne mondanák, hogy *mit*. Ezzel szemben a *havazik* igéhez csak azt lehet hozzáfűzni, hogy *hol*, *mikor* és *mennyire*; ezeket viszont akármilyen igéhez hozzá lehet kapcsolni. Vonzatnak tehát azokat a bővítményeket nevezzük, amelyek befolyásolják az ige jelentését, és az ige kapcsán kerülnek a mondatba. Az elemzés során ezeket a mondatrészeket a megtalált ige alapján keressük. Szabad határozóknak pedig azokat a bővítményeket nevezzük, amelyek csak a külső körülményeket írják le, a jelentést nem módosítják, és általában minden mondatban keressük őket. Persze igen nagy különbségek vannak azon a téren, hogy mennyire kötelező az egyes vonzatoknak megjelennie a mondatban.

Az alany, a tárgy és a részeshatározó

Az ige vonzatai közül, úgy tűnik, külön osztályt alkot az alany, a tárgy és a részeshatározó. Alanynak nevezzük azt a mondatrészt, amely a ragozott ige mellett alanyesetben áll, a tárgy *-t* ragot kap (de a mérték értékű tárgyat nem nevezzük tárgynak). Részeshatározónak nevezzük azt a mondatrészt, amely vagy *-nak* ragot kap, vagy a részére vagy számára névutót. Nem tarjuk részeshatározónak azt a *nak*-ragos mondatrészt, amelynél a *-nak* rag nem cserélhető fel a névutóval.

Küldtem egy levelet Péternek. (részeshatározó)
Küldtem egy levelet Péter számára. (részeshatározó)
Küldtem egy levelet Péter részére. (részeshatározó)

Pétert fiának ismerte el. (*nak*-ragos vonzat)
Hitt a barátjának. (*nak*-ragos vonzat)
Pétert elnöknek jelölték az egyesület részére.
(egy *-nak* rag vonzat + egy részeshatározó)

Péternek van egy szürke kalapja. (birtokos rag)
Elkérem Péternek a szürke kalapját. (birtokos rag)

Nekem is oda kell mennem. (nak-ragos alany)
Péternek tilos odamennie. (nak-ragos alany)

Idegenek számára tilos odamenni. (????)

Ha ezt a leírást végigolvassuk, többször találkozunk olyan kitétellel, hogy az a bizonyos mondatrész csak vagy az alany, vagy a tárgy, vagy részeshatározó lehet. Például, ha az ige egy másik igenevet vonz, ennek az igenévnek az alanya az első ige valamelyik vonzata. Ez a vonzat csak az alany vagy a tárgy vagy a részeshatározó lehet. (A részeshatározó sokkal ritkábban tölt be ilyen szerepet, mint az alany vagy a tárgy.)

Másrészt, ha egy adott jelentésű mondatot más mondatformákba transzformálunk, akkor a formális grammatikai alany, tárgy és részeshatározó gyakran szerepet cserél. Pl.:

Ezt a könyvet idegenek nem kölcsönözhetik.
Ezt a könyvet idegeneknek nem lehet kölcsönözni.
Ezt a könyvet tilos idegeneknek kölcsönöznie.
Ennek a könyvnek a kölcsönzése idegenek által tilos.
(kölcsön vesz)

Ezt a könyvet idegenek számára nem lehet kölcsönözni.
Ezt a könyvet idegeneknek nem lehet kölcsönözni.
Ezt a könyvet idegenek számára tilos kölcsönözni.
Ez a könyv idegenek számára nem kölcsönözhető.
Ez a könyv idegenek által nem kölcsönözhető.
Ennek a könyvnek a kölcsönzése idegenek számára tilos.
(kölcsön ad)

Mivel a szöveg gépi megértésére törekszünk, ezeket a különböző formájú, de lényegében azonos jelentésű alakokat azonos belső szemantikai reprezentációra akarjuk hozni. A szemantikai reprezentáció azon alapul, hogy az igének különböző meghatározott szerepű argumentumai vannak. Ezeket az argumentumokat a különböző mondatmintákban más és más grammatikai jellemzők alapján találjuk meg.

A logikai alany a mondatok zömében (ahol ragozott ige van) egy alanyesetű mondatrész, más mondatokban viszont *-nak* ragot kap, megint más esetben az által vagy a *számára*/*részére* névutóval lehet megadva, és a tárgyatlan igékből képzett főnevek mellet lehet, hogy birtokosként áll.

A logikai tárgy többnyire -t ragot visel, de egyrészt az ún. mérték értékű tárgy nem tárgy, hanem mértékhatározó, másrészt igen sok mondatban van különböző okokból alanyesetben álló tárgy. Pl. *-m* vagy *-d* birtokrag után nem kötelező kitenni a tárgyragot:

Keresem a kalapom.

A részeshatározót mindig a *–nak* rag vagy a *számára/részére* névutó jelöli ki, csupán az a kérdés, hogy a *–nak* ragot viselő szó valóban részeshatározó-e, mert lehet esetleg alany, birtokos, vagy egyszerű *–nak* ragos vonzat is.

Az alany, a tárgy és a részeshatározó egyszeres mondatrész: azaz egy mondatban csak egy ilyen mondatrész lehet. Az az egy természetesen lehet felsorolás, kötőszavas szerkezet vagy egy névszói kifejezés a hátravetett jelzőjével (értelmezőjével), de ezeknél különböző jelek utalnak az összetartozásra. Ha tehát egy mondatban több alanyesetű, *t*-ragos vagy *nak*-ragos névszói kifejezést találunk, akkor ezek közül csak egy lehet alany vagy tárgy vagy részeshatározó, a másik valami más. Nem ez a helyzet a helyhatározókkal és időhatározókkal. Egy mondatban a legkülönbözőbb helyeken egymástól függetlenül több hol kérdésre válaszoló helyhatározó is állhat. (Pl.: A XII. kerületben lakom Budapesten. ezzel szemben: Péternek kellett odaadnia a levelet Pálnak.)

Ragos és névutós egyszeres vonzatok

Rengeteg olyan ige van, aminek valamelyik argumentumát csak egy bizonyos rag vagy névutó jelölheti ki. Ezeket a vonzatokat a nyelvtankönyvek csoportosítják, és különböző határozónak nevezik el. Mi ilyet nem teszünk, mert vagy rettentő sok csoport születne, vagy elég különböző dolgokat kellene egy csoportba sorolni, és egy-egy rag túl sok csoportba is beletartozna. Annál a mondatnál, hogy *Beszélt az esetről/a fiáról/a mérkőzésről/a feladatról.* a *-ról/-ről* rag semmi mással nem helyettesíthető. Hasonló a helyzet a Találkozott valakivel. esetében is, a *-val/-vel* rag semmi mással nem pótolható. Nem egészen ez a helyzet a *Kanállal ette a levest.* fazonú mondatokban, ahol azt is lehet mondani, hogy Kanál nélkül ette a levest. Itt a *-val/-vel* és a nélkül ugyanazt az argumentumot jelöli meg, még ha negatív értelemben is. Ilyen értelemben eszközhatározóról lehet beszélni, és ez alatt a *-val/-vel*-t és a *nélkül*-t kell érteni. Már a korábbiakban láttuk, hogy különbséget teszünk a *nak*-ragos vonzat és a részeshatározó között.

Néhány ilyen csoport már kialakult munkánk során, de mások is elképzelhetők. A fentiek kivételével még nem neveztük el őket.

A ragos vonzatok is csak egy helyen állhatnak a mondatban.

Hely- és időhatározós vonzatok

Hely- illetve időhatározónak nevezünk egy vonzatot, ha nem egy konkrét ragot vagy névutót vonz az adott ige, hanem egy olyan mondatrészt, amely bármilyen hol, hova, honnan illetve mikor kérdésre válaszoló ragot vagy névutót kaphat. Ilyen mondatrészt szinte minden igei állítmányú, sőt majdnem minden névszói mondatban is elhelyezhetünk, ezért általában ezt a mondatrészt szabad határozónak tartjuk. Mikor tartunk egy ilyen bővítményt mégis vonzatnak? Csak abban az esetben, ha a mondatba való helyezése a mondat értelmét megváltoztatja. Pl.:

Az óra áll.
Az óra az asztalon áll.

A hely- és időhatározó többszörös határozó, több is lehet belőle a mondatban.

A konyhában az óra az asztalon áll a váza mellett.

Annak eldöntésénél, hogy egy helyhatározó értelemmódosító vonzat-e vagy szabad határozó, a mondatbeli helye meghatározó, de sajnos a hangsúlynak is szerepe lehet.

Az asztalon az óra áll.
Az óra az asztalon áll.

Általában elmondhatjuk, hogy az ige meghatározza azt, hogy a mondatban milyen hozzákapcsolódó bővítmények lehetnek, de ez a kapcsolódás lehet szorosabb vagy lazább, amin azt értjük, hogy a bővítmény mennyire határozza meg az ige értelmezését. Az egyik véglet az idiómák, a másik véglet a szabad határozók. Az idiomatikus vonzat teljesen megváltoztatja az ige értelmét (pl. *módjában áll*), a szabad határozó pedig alig befolyásolja. Sok közbeeső fázis van, és ezeknek mind más és más a súlya a mondat elemzésénél.

Az igekötők és kulcsszavas vonzatok

Vannak olyan szavak, amelyek teljesen megváltoztatják az ige jelentését, amikor az ige mellett állnak. Pl.:

figyelmen kívül hagy
munkához lát
nehézséget támaszt
véget vet

Az ilyen szavakat mi idiomatikus vonzatoknak nevezzük. Ha ilyet találunk, úgy tekintjük, hogy nem az eredeti ige szerepel a mondatban, hanem egy másik, az idiómával kifejezett ige, aminek saját vonzatkerete van. Ennek egyik speciális esete az igekötő, ami azzal a tulajdonsággal bír, hogy ha az ige előtt áll, akkor az igével egybeírjuk.

Mivel egy sor igének egy sor igekötőjére igaz, hogy az eredeti vonzatkeretet nem változtatja meg, felmerült a lehetőség, hogy ezt jelezni kellene, és ezáltal a vonzat-szótár egyszerűsödne. Ezzel mi még nem foglalkoztunk.

Az idiómáknak két formája van. Az egyikben, amit mi megkövesültnek nevezünk, az igéhez kapcsoló szó egyetlen lehetséges formában állhat az ige mellett. (Mint a fenti példákban.) Az ige különböző ragokat vehet fel, de a kapcsolódó szó alakja változatlan. A másik esetben az igéhez kapcsolódó szó töve adott, esetragja adott, de különböző birtokjeleket, többes szám-jelet vehet fel a vonzat.

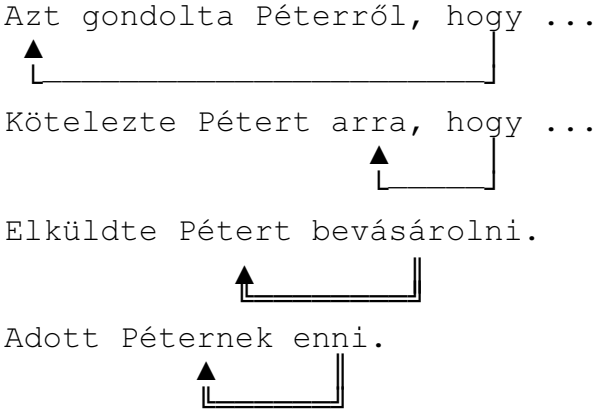
felhívta figyelmünket
rendelkezésére áll
elfoglalta a helyét
feltételeket szabott

Ilyenkor a vonzat megfelelően transzformálódik, ha az igét igenévvé alakítjuk.

a vevő által szabott feltétel

Alárendelt mondatok és a főnévi igenévi vonzat

Az ige nem csak azt szabja meg, hogy milyen végződéssel, névutóval vagy egyéb formában kapcsolódhatnak hozzá további mondatrészek, de azok szemantikus szerepét is meghatározza. Ebből következik, hogy az ige meghatározza, hogy melyik argumentumaként állhat *hogy*-os mellékmondat (milyen utalószóval hivatkozhatunk rá); illetve milyen argumentum helyén állhat főnévi igenév.



Főnévi igenév esetén az ige azt is megszabja, hogy melyik argumentuma lesz a főnévi igenév logikai alanya.

Névszói mondatok

A névszói mondat két alanyesetű névszói kifejezést tartalmaz, és a mondat ezek kapcsolatát fejezi ki.

Míg az igei állítmányú mondat igei relációt fejez ki, a névszói állítmány egy névszói relációt. Az igei relációban az ige a meghatározó elem, az alany, a tárgy, és a többi határozó pedig egyformán fontos argumentumok. (Semmi jele, hogy a magyar mondatban az alany másképp viselkedne, mint a többi argumentum.) Mivel az alanyt az igerag többékevésbé meghatározza, (akkor is, ha az igeragot a segédige vagy a létige hordozza,) az alany gyakran elmarad, és ez a mondat értelmezését nemigen zavarja, legalábbis nem jobban, mint bármelyik másik argumentum hiánya.

A névszói mondat két halmaz vagy egy halmaz és egy elem (egy elemű halmaz) közötti viszonyt fejez ki. Nevezetesen azt, hogy egy elem beletartozik egy halmazba (például rendelkezik valamilyen tulajdonsággal), egy halmaz része egy másiknak, két halmaz egyenlő, két elem azonos.

Pista katona. (*Pista* ∈ {*Katonák*})
A bálna emlős. ({*bálnák*} ⊂ {*Emlősök*})
Kovács Péter a tanácselnök. (*a tanácselnök = Kovács P.*)
Az a és az az a magyar nyelv határozott névelői.
({*a, az*} = {*Határozott névelők*})

A névszói relációnak két egyforma fontos argumentuma van. Ha bármelyik elmarad, a mondat igen hiányossá válik.

Kérdés, hogy mikor fejez ki egy névszói reláció azonosságot, és mikor valamilyen halmazba való beletartozást (azaz valamilyen tulajdonságot). A mi véleményünk szerint két eset van: ha az egyik névszói kifejezés határozott és a másik határozatlan, akkor a határozott kifejezéssel jelölt beletartozik a határozatlan kifejezéssel jelöltek osztályába; ha viszont mindkettő határozott, akkor azonosak. Olyan mondatot eddig nem találtunk, amiben két határozatlan névszói kifejezés szerepelt volna. (Hogy a névszói kifejezés mikor határozott és mikor határozatlan ebben az értelemben, azzal részletesen foglalkozunk a megfelelő fejezetben. Lásd továbbá az **igei és névszói relációk**ról szóló fejezetet is.)

Az igei relációhoz hasonlóan a névszói relációnak is van ideje és módja, ezt a létige megfelelő alakjával (*volt, lesz, volna, lenne, legyen*), módosító szavakkal és segédigével fejezzük ki. Véleményünk szerint az

Okosabbnak kellett volna lennie.

mondat is névszói mondat (*Ő okosabb.*), csak múlt időbe, és "szükséges" módba van téve.

Mi a határozatlant hívjuk állítmánynak, a határozottat pedig alanynak. Ez azzal is összhangban van, hogy múlt időben, felszólító módban stb. a létige megfelelő formája előtt a határozatlan áll (hacsak a hangsúly miatt más szó nem kerül oda). Ennek megfelelően azokban a mondatokban, ahol két határozott névszói kifejezés szerepel, az az állítmány, amelyik mögött a létige áll. (Azt érezzük hangsúlyosabbnak.) Ha a mondatban nincs létige, nem lehet dönteni. (Hangsúlytól függ.) Pl.:

Kovács Péter a tanácselnök.
Kovács Péter volt a tanácselnök.
Kovács Péter a tanácselnök volt.

Egy másik szempontból nézve az állítmányt a következő osztályokba sorolhatjuk:

Ha az állítmány egy magányos főnév (esetleg határozatlan névelővel), akkor az alany beletartozik a főnév által megnevezett dolgok halmazába. Pl.:

Ez a négyszög paralelogramma.

vagy *Ez a négyszög egy paralelogramma.*

Ha az állítmány melléknév, akkor az az alanynak valamilyen tulajdonságát adja meg. Pl.:

A kenyér még meleg.

Ha az állítmány számnév és mértékegység, akkor az alany mennyiségét adja meg. Pl.:

A kenyér két kiló.

Ha az állítmány sorszámnév, felsőfokú melléknév, névmás stb. (lásd a szelektorokat a **névszói kifejezés**ben), akkor az alany azonos azzal az elemmel, amit a sorszámév, a mellék- név vagy a névmás kiválaszt valamilyen halmazból.

Péter második volt.
Ez a pont a legnagyobb.
Ő lehet a gyilkos.

Ha az állítmány egy nagyobb névszói kifejezés, akkor az alany azonos ezzel a kifejezéssel. Ha ez egy határozott névszói kifejezés, akkor azzal a konkrét valamivel, amit a kifejezés megjelöl. Ha pedig határozatlan, akkor egy olyan elemmel, aminek ilyen tulajdonságai vannak. Pl.:

Péter volt a verseny utolsó befutója.
Kedvenc ételem egy nagy szelet friss, puha kenyér.

A fenti két definíció nincs ellentmondásban egymással, ha a határozottság és határozatlanság fogalmát helyesen értelmezzük. (Lásd **határozott és határozatlan névszói kifejezések**.)

Egyszerű egzisztencia-mondatok

Vannak olyan mondatok, amelyek egy alanyesetű névszói kifejezésből és egy létigéből állnak. Az ilyen mondat a dolog létezését fejezi ki. (Ezért a létige nem marad el.) A létigén természetesen megint teljesen általánosan egy létigét tartalmazó kifejezést értünk. Pl.:

Tegnap este ott a sarkon egy baleset volt.

Itt kellene lennie valahol a kulcsnak.

Létigés mondatok tapasztalatunk szerint viszonylag ritkán fordulnak elő a szövegekben, ennek is viszonylag nagy része az az eset, amikor egy birtokos kapcsolat fennállását állítjuk, azaz *Valakinek van valamije* alakú mondatok.

A szabad határozók

Mindhárom fenti mondattípusban (igei állítmányú, névszói állítmányú, létigei állítmányú) szabad határozókat helyezhetünk el, amelyek az állítmány módjához hasonlóan módosítják a mondatban szereplő állítás értelmét. Mi eddig a következő szabad határozókat találtuk: *Hol, Mikor, Hányszor, Mennyit, Ki szerint, Hogyan.*

Hol

A mondatban szinte mindig elhelyezhető "hol" kérdésre válaszoló helyhatározó. Néha viszont tulajdonképpen "hol" értelmű, de "honnan" kérdésre válaszoló helyhatározó szerepel a mondatban. Pl.:

Kórházból írt levelet.

Még nem tudtuk eldönteni, ebben az esetben vonzatról vagy szabad határozóról van-e szó. A jelenségnek megvan a párja is. Azok az igék, amelyeknek részeshatározó vonzatuk van, gyakran kapnak helyette "hova" kérdésre válaszoló bővítményt. Pl.:

Fegyvereket adott el Irakba.

Mikor

Szinte minden mondatban elhelyezhető időhatározó is. Az időhatározó felismerése roppant nehéz feladat, mert majdnem minden raggal és névutóval lehet időhatározót csinálni. Pl.:

Elsejével munkába áll.

A sorompó egy órán túl is zárva tartható.

Néha az is kérdéses, hogy az adott szó hely- vagy időhatározó-e.

Fessünk oda három pontot, és utána egy hosszú vonalat.

Hányszor

Számos mondatba tehető be számnév *-szor/-szer/-ször* raggal. Hasonló a szerepe az *-nként* ragnak is.

Mennyit

Mennyit szóval az úgynevezett mérték értékű tárgyat jelöltük. A mérték értékű tárgy egy tárgyragos határozatlan névszói kifejezés, ami az állítmány mértékét jelöli.

Késett valamennyit.
Az alma már piros egy kicsit.

Ennek fölismerése igen könnyű, ha a mondatban szereplő ige tárgyatlan, vagy ha tárgyasan van ragozva, úgyszólván lehetetlen viszont a tárgyatlanul ragozott tárgyas ige mellett.

Várt egy kicsit. - És egy ilyen nagy jött helyette.
(Prószéky Gábor példája.)

Szerint

Az általunk vizsgált szövegekben igen sokszor fordult elő az, hogy *valaki szerint*. Ez mintegy idézőjelbe teszi a mondat teljes állítását. Hasonló szerepet tölt be az *állítólag* szó is.

Amikor gyermekelhelyezési perek kivonatait elemeztük, gyakran találkoztunk a *valakinek a sérelmére* kifejezéssel, ami nyelvtanilag teljesen a fenti kifejezéshez hasonlóan viselkedett.

Hogyan

A "mód", "ok", "cél", "állapot" és egyéb hasonló határozók igen sötét részét jelentik számunkra a magyar nyelvnek. Eddig nem találtunk olyan igét, aminek ilyen vonzata lenne, viszont minden mondatba betehető valami ilyen bővítmény. Semmilyen olyan definíciót nem találtunk, ami alapján kritériumokat tudnánk felállítani, milyen szavak milyen ragokkal, névutókkal sorolhatók ide, ezeknek a határozóknak hol a helyük, mi a szerepük a mondatban, hogy változtatják meg a mondat jelentését stb.

Van néhány határozószó, egy sor névutó, néhány rag, amit ide soroltunk, mert máshova nem tudtunk:

akárhogy, bárhogy, sehogy, valahogy, semmiképpen,
másképpen, legfőképpen, (határozószó)
így, úgy, ahogy (névmásként viselkednek)
alapján, ellenére, függetlenül, folytán, képest,
következtében, miatt, útján, végett, (névutó)
-an/-en pl.: *Vizsgáljuk meg alaposan a mondatot.*
-ként pl.: *Első lépésként oldjuk meg az egyenletet.*
-ul/-ül pl.: *Váratlanul bejött a szobába.*
-lag/-leg pl.: *Foglaljuk le a helyet fizikailag is.*

(Megjegyzés: az ilyen raggal ellátott szavakat csak akkor soroljuk a "hogyan" kérdésre válaszoló szabad határozók közé, amikor nem állnak melléknév előtt. Melléknév előtt annak módosítójául szolgálnak. Pl. *alaposan elrontott*, *fizikailag gyenge* stb.)

"Hogyan" kérdésre válaszol ezen kívül egy csomó, *hogy*-os mellékmondat és a határozói igeneves kifejezések, sőt ezenkívül rengeteg, mindenféle raggal és névutóval ellátott kifejezés. Néhány extremitás például *-t* raggal vagy *-val/-vel*-lel:

Jót aludt.
Nagy kedvvel dolgozott.

Néha nem lehet tudni, hogy a mondatrész "hogyan" kérdésre válaszol-e vagy időhatározó-e vagy helyhatározó-e.

A társhatározó

Noha formailag a szabad határozók közé tartozik, logikailag vagy az alanyhoz vagy a tárgyhoz kötődik a társhatározó, amelyet vagy a *-val/-vel* raggal, vagy a nélkül névutóval adunk meg. Pl.:

Jóska Pistával ette a kenyeret.
illetve *Jóska szalonnával ette a kenyeret.*

A szemantikus reprezentációban ezt olyan formára alakítjuk át, hogy:

Jóska és Pista ette a kenyeret.
illetve *Jóska ette a kenyeret és a szalonnát.*

Grammatikailag sajnos eldönthetetlen, hogy az alanyhoz vagy a tárgyhoz kell-e kötni a társhatározót, és az eszközhatározótól sem lehet elválasztani.

A "helyett"

Ez is egy olyan mondatrész, amely csak formailag határozó, valóságban valamelyik argumentumhoz kapcsolódik. Sajnos egyáltalán nem lehet tudni, hogy melyikhez, mert bármelyikhez kapcsolódhat. Pl.:

Kanállal ette a tésztát villa helyett.
Hatkor érkezett öt helyett.

(Elképzelhető, hogy mindig a fókuszban álló szóhoz kapcsolódik.)

Az argumentumok megtalálása

A fenti mondatokban beszéltünk alanyról, tárgyról, különféle határozókról. Ezeket közös néven a mondat gerincét képező reláció argumentumainak nevezzünk. Ebben a részben azzal a kérdéssel foglalkozunk, mi állhat az argumentum helyén.

Mind igének, mind névszónak lehet olyan argumentuma, amelyet csak (ragozatlan) főnévi igenév tud betölteni.

Az összes többi argumentumot egy sor különböző dolog töltheti be. Elsősorban és leggyakrabban a megfelelő esetragot illetve névutót viselő névszói kifejezés szokta betölteni ezt a szerepet. De természetesen helyettesítheti egy névmás is, vagy állhat itt egy megfelelő jelentésű határozószó is.

A névmások többsége ragozható és kaphat névutót is. Pl.:
az, azt, annak, azzal, azon, az után, az alatt stb.
ami, amit, aminek, amivel, amin, ami után, ami alatt stb.

A személyes névmások nem ragozódnak, az alany- és tárgyesetet külön alak jelöli, a névutók nagy részénél a személyes névmás névutós formáját a személyragozott névutóval adjuk meg, a ragok nagy részéből fiktív névutó képződik, és ezt személyragozzuk. Pl.:

én, engem, nekem, velem, rajtam, utánam, alattam stb.

(Lásd a **névutók**ról szóló fejezetet is.)

Ha tehát egy olyan vonzatot keresünk, ami egy adott rag határoz meg, akkor vagy egy olyan névszói kifejezést kell keresnünk, ami az adott ragot viseli, vagy egy olyan fiktív személyragozott névutót kell keresnünk, ami az adott ragból képződött, vagy egy olyan határozószót, amely noha nem viseli az adott ragot, mégis olyan, mintha az adott ragot viselné. Például az *amíg* szó úgy viselkedik mintha *-ig* ragot viselne.

Amikor pedig egy olyan vonzatot keresünk, amit egy adott névutó határoz meg, akkor vagy egy olyan névszói kifejezést kell keresnünk, ami az adott névutóra végződik, vagy a névutó személyragozott alakját, vagy egy olyan szót, amely olyan, mintha az adott névutót viselné. Például az ... *éve*, ... *napja* stb. szó a "mióta" kérdésre válaszoló szavak közé tartozik.

Amikor egy olyan vonzatot keresünk (például a "hol" kérdésre válaszoló helyhatározót), amely többféle ragot vagy névutót is viselhet, akkor keresnünk kell a lehetséges ragok szerint, a lehetséges névutók szerint, és azok között a szavak között, amelyek még idetartoznak. Például az *otthon* válasz, a "hol" kérdésre pedig a "miben", "min", "mi alatt", "mi mögött" stb. kérdések egyikére sem válaszol.

Összegezve tehát, a vonzatok keresésénél nem elegendő a ragokat és a névutókat figyelni, hanem egy sor ragozhatatlan szó is szóba jön vonzatként. Ezért vagy a szóhoz fel kell jegyezni, hogy milyen vonzatként állhat, vagy a vonzathoz fel kell jegyezni, hogy bizonyos ragokon és névutókon kívül, milyen ragozhatatlan szavakat fogad még el vonzatként.

A vonzat természetesen nem csak egyetlen névszói kifejezés lehet, hanem több is megfelelő kötőszavakkal és vesszővel összekapcsolva. A felsorolásban természetesen különböző dolgok keveredhetnek (ragos névszói kifejezések, névutós névszói kifejezések, határozószók, névmások). Különféle szabályok vannak arra, hogy mikor kell megismétlődnie a ragnak, mikor a névutónak stb. A szabályok részletes leírása a **Névszói kifejezések** című fejezetben található meg.

Másodlagos vonzatok

Nem csak a mondat gerincét alkotó igei relációnak vannak argumentumai, hanem ezen argumentumok maguk is újabb relációk lehetnek. Ha az argumentum egy összetett névszói kifejezés, ez olyan névszói relációkat foglal magában, mint amilyenek a névszói állítmányban már szerepeltek. A névszói kifejezés végén álló főnévnek lehetnek tulajdonságai, lehet mennyisége, lehet birtokosa, ki lehet választva a halmazból valamilyen speciális tulajdonsággal stb. Az értelmező (hátravetett jelző) az esetek nagy részében a kétféleképpen definiált argumentum azonosságát fejezi ki.

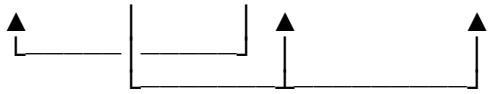
De nemcsak ilyen névszói relációk kapcsolhatók egy argumentumhoz. Ha az argumentum például egy igéből képzett szó, az igének a korábbi vonzatai megmaradnak és hozzákapcsolhatók a szóhoz. Továbbá nem csak igéknek lehetnek ragokkal kijelölt vonzataik, hanem főneveknek és melléknéveknek is.

Ezekkel az argumentumokkal kétféle dolog történhet. A jelzői pozícióban álló melléknevek vonzatai a melléknév elé kerülnek, és ilyen módon az egész reláció beilleszkedik a névszói kifejezésbe. (Ez a **Névszói kifejezés** című fejezetben van pontosan leírva.) Pl.:

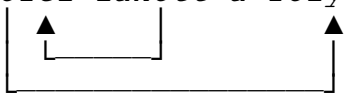
Két régóta keresett könyvet kaptam meg ma a boltban.

Az argumentumokhoz kapcsolódó más mondatrészek viszont szétszóródnak a mondatban. Pl.:

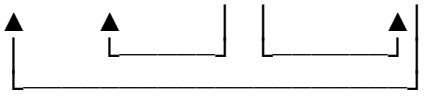
Kenyeret ment vásárolni a boltba.



Közel lakott a folyóhoz.



Kínába soká tartott az út.



A feladat nem oldható meg ezzel a módszerrel.

Ezeket a távoli vonzatokat nevezzük másodlagos vonzatoknak. Ha a mondat gerincét felépítő relációt megtaláltuk, és megtaláltuk a hozzá kapcsolódó argumentumokat, és megtaláltuk az argumentumokhoz kapcsolódó névszói kifejezéseket is, akkor még mindig lehetnek további mondatrészek a mondatban, amelyek valamilyen vonzaton keresztül valamelyik argumentumhoz kapcsolódnak. Ezeket a vonzatokat mindig az adott argumentum legfőbb, legjobboldali eleme jelöli ki.

Ha ez egy névszó, és birtokos jelet visel, lehet elváló *nak*-ragos birtokosa. Pl.:

Péternek keressük a kalapját.

Ha a névszói kifejezés utolsó eleme egy olyan névutó, amely ragot vonz, akkor ez a ragos kifejezés leszakadhat, és máshova kerülhet a mondatban. Pl.:

Túl lakott a folyón.

Ha a szó főnévi igenév vagy határozói igenév, akkor mindig magányosan áll, és argumentumai szétszóródnak a mondatban.

A pohárban egy férget látott úszkálni.

Visszaadta az írást megjelölve a hibákat.

A melléknevek vonzatai általában a melléknevek előtt állnak beilleszkedve a névszói kifejezésbe. Csak akkor szakadnak le a melléknévről, ha az állítmányként kerül a mondatba. Nem vonatkozik ez a közép és felsőfok által okozott vonzatokra, amelyek a névszói kifejezéseken belül is állhatnak, de azon kívül is.

A gyermekeire büszke anya nagyon megsértődött.
Az anya roppant büszke volt tegnap gyermekeire.
Egy kisebbet vettem a tiednél.
Egy a tiednél kisebbet vettem.

Végül, ha a legjobboldali szó főnév, lehetnek ún. külső vonzatai. (Olyan vonzatok, amelyek rendszerint a főnév mögött állnak, mert a névszói kifejezésbe csak egy töltelék igenévvel illeszthetők be. Erről részletesebben lásd **az igenevek**ről szóló fejezetet.) Pl.:

Az út vezet/visz/tart/stb. valahonnan valahova.
Nehezen találta meg az utat haza.
A gyermek van/származik/stb. valahonnan.
Ritkán látogatta gyermekét az első házasságból.

A másodlagos vonzatoknak, csakúgy mint az elsődleges argumentumoknak, szabad a szórendje, és bármelyik fókuszálható. Noha elvileg a szabad szórendű részek bármilyen sorrendben szerepelhetnek a mondatban, az a tapasztalat, hogy az összetartozó mondatrészek többnyire együtt maradnak (a fókuszálástól eltekintve), mert különben a mondat nehezen érthetővé válik.

Elvileg a másodlagos vonzatoknak is lehetnének vonzataik, de ilyennel a gyakorlatban nem találkoztunk. A relációk ritkán vannak háromnál nagyobb mélységben egymásba ágyazva.

AZ ÖSSZETETT MONDATOK

Az összetett mondatok szerkezetének felismerése sokkal bonyolultabb feladat, mint az egyszerű mondatok analízise. Ennek több oka is van:

- Egyrészt kérdéses, hogy mi számít egyszerű mondatnak, és mi több tagmondat összetételének.
- Kérdéses, hogy a szövegben miről tudjuk a tagmondatokat felismerni.
- Milyen kapcsolatok lehetnek a tagmondatok között.

Egy mondat - több mondat

Ha egy mondat nagybetűvel kezdődik; ponttal, felkiáltójellel vagy kérdőjellel végződik, és egy vessző van benne, továbbá a vessző két oldalán egy-egy ragozott ige áll a saját vonzataival; akkor többé-kevésbé világos, hogy itt összetett mondatról van szó, amelynek két tagmondatát a vessző választja el. Nem minden mondat ilyen, sajnos. Pl.:

Egész nap az ágyban evett, ivott, heverészett.

Ennél a mondatnál valószínűleg helyesebb, ha úgy tekintjük, hogy összetett (halmozott) állítmánya van, és az ehhez kapcsolódó hely és idő a benne szereplő összes tevékenységre vonatkozik. Nem biztos, hogy ugyanez a helyzet annál a mondatnál is, hogy:

Egész nap evett, ivott, heverészett az ágyban.

Az egész nap mindháromra vonatkozik, az ágyban pedig lehet, hogy mindháromra, de lehet, hogy csak a heverészés-re. Megint más a helyzet, ha azt mondjuk, hogy:

Egész nap evett, ivott, az ágyban heverészett.

Az ágyban itt egyértelműen a heverészés-re vonatkozik.

Egy lehetséges megoldásnak látszik, ha a mondatot a vesszőknél úgy vágjuk szét, hogy minden egyes részbe egy ragozott ige kerüljön, és az így kapott mondatokat elemezzük. A kapott mondatok természetesen valószínűleg roppant hiányosak lesznek. Nem látszik túl egyszerűnek annak eldöntése, hogy a hiányok közül mit lehet pótolni az előző tagmondatokból öröklődéssel. A megoldásban nyilván fontos szerepe lesz a topiknak és a fókusznak, és a megoldás nem lehet tökéletes, mert a hangsúlyt írott szövegnél nem tudjuk figyelembe venni. Másik oldalról is felmerül a kérdés, honnan ismerjük meg a mondat azon részét, amit önálló tagmondatnak kell tekinteni, milyen mondatrészeknek kell abban szerepelniük.

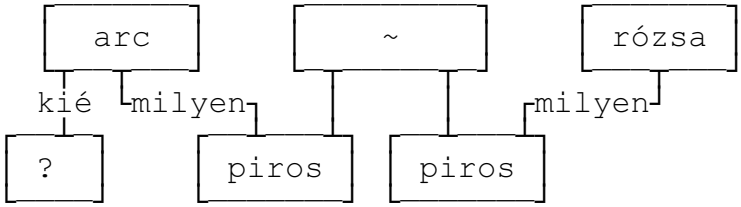
Máskor is gyere hozzám, de este.
Ma is azzal láttam Pistát, akivel tegnap.
Nem csak én mondom, hanem mindenki.
Pestnek jövője van, Budának múltja.
Már nagyocska gyermek voltam, olyan ötödfél esztendős.
Akkor szép az erdő, mikor zöld.
Kicsi a rózsám, de csinos.
Addig hajlítsd a fát, amíg fiatal.

Noha az összetett mondatok 80-90%-a viszonylag normális mondat, van állítmánya, ugyanakkor az sem ritkaság, ha az egyik tagmondat (rendszerint a második) olyan csonka, hogy gyakorlatilag elemezhetetlen. Az igei állítmányú mondat rendszerint azért, mert nincs ott az ige, nincs mihez kötni a vonzatokat; a névszói pedig rendszerint azért, mert egyetlen alanyesetű névszói kifejezés van benne ige nélkül. Az esetek igen nagy részében ezen megpróbálhatunk segíteni. Ha az első mondat igei állítmányú, és a második olyan hiányos, hogy elemezhetetlen, akkor megpróbálhatjuk úgy elemezni, mintha az első mondat igéje állna benne. Ha az első mondat névszói, akkor feltehetjük, hogy a másodikból az első tagmondat alanya hiányzik.

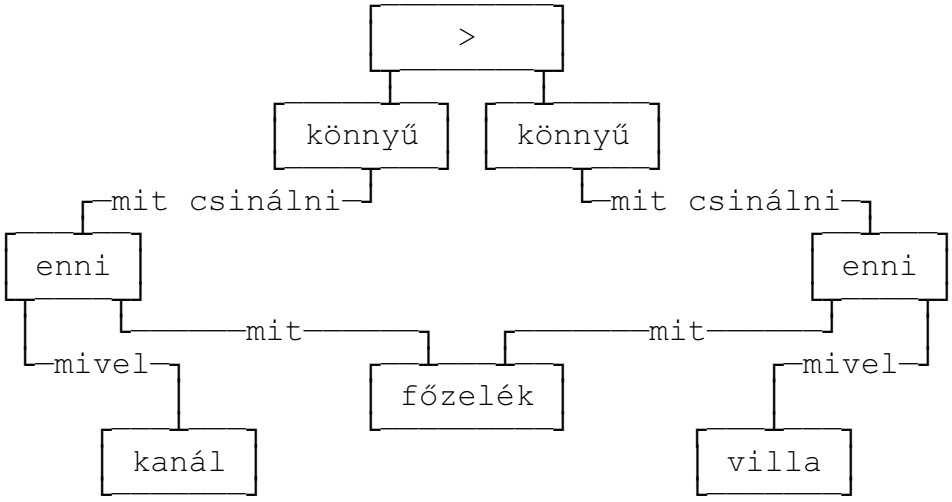
Persze ez az eljárás nem mindig válik be. Például az utolsó példamondatra sem alkalmazható. Az is elképzelhető, hogy félreértés lesz belőle. (Bár ilyenkor rendszerint a kiinduló mondat is kétértelmű volt, legalábbis nyelvtanilag.)

Van egy mondatfajta, az összehasonlítás, ami majdnem mindig hiányos. Legalábbis "A magyar helyesírás szabályai" és "A mai magyar nyelv" egyes fejezetei explicite azt mondják, hogy az összehasonlító mondatokban a *mint* kötőszó utáni rész egy önálló mondat. Mivel azt tapasztaltuk, hogy az esetek 95%-ában egyetlen névszói kifejezés áll a *mint* után, további 4%-ban pedig vonatkozó névmással bevezetett almondat, mi nem így fogjuk fel a dolgot, mivel nem szeretünk hiányos mondatokat elemezni. A mi szintakszisunk szerint a *mint* után vagy egy névszói kifejezés, vagy egy az utalószóhoz illő vonatkozó névmással kezdődő almondat lehet. Az első esetben a főmondat tartalmát egy igei vagy névszói relációban ábrázoljuk. Ezt a relációt még egyszer lemásoljuk, és a megfelelő mondatrészét helyettesítjük a *mint* után álló névszói kifejezéssel, majd az olyan, annyi stb. utalószóval, vagy a középfokkal jelölt mondatrészeket állítjuk a megfelelő relációba (< kevésbé, ¸ ugyanolyan stb.). A másik esetben két független mondatról van szó, és az utalószóval vagy középfokkal jelölt mondatrészt a vonatkozó névmással jelölt mondatrésszel hasonlítjuk össze. Pl.:

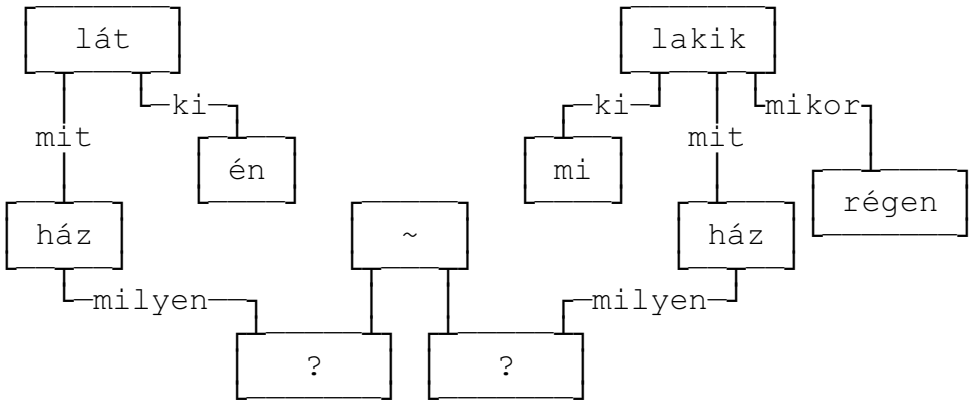
Az arca olyan piros, mint a rózsa.



Kanállal könnyebb enni a főzeléket, mint villával.



Láttam egy olyan házat, mint amiben régen laktunk.



Vannak persze olyan mondatátszövéssel, vagy kihagyással képzett összehasonlítások, amit nem lehet ilyen egyszerűen levezetni. Pl.:

Fülledt a csönd, mint ha üres a kaptár...

Az összetett mondat felbontása tagmondatokra

Ha az ember egy nyelvtankönyvet olvas, azt hiheti, hogy az összetett mondat tagmondatokra való felbontása viszonylag egyszerű feladat, hiszen: "a tagmondatok határán a vessző mindig kiteendő, akár van kötőszó akár nincs". (A magyar helyesírás szabályai 99. oldal.) Sajnos a felbontás nem ilyen egyszerű dolog.

Ennek az a legalapvetőbb oka, hogy a magyar helyesírás túlságosan sokféle és egymástól igen különböző célra használ fel bizonyos írásjeleket. Ezek a túlhasznált írásjelek: a pont, a vessző és a kettőspont. Ha megpróbáljuk szisztematikusan leírni, hogy hova kell (és hova nem kell) pontot vagy vessző írni, minimálisan egy-egy nyomtatott oldalt kapunk, a kettőspont leírása is kb. fél oldal. A Magyar Helyesírás Szabályaiból a következő szabályokat hámozhatjuk ki a vesszők elhelyezéséről:

A vessző egyaránt elválaszt mondatrészeket és mondatokat, sőt néha olyan helyekre is tenni kell, ami mondattanilag nem egészen logikus: a mondathoz lazán kapcsolódó határozók elé (lásd MHSZ 248/e). (Vajon mitől laza egy határozói kapcsolat?)

A vessző tehát önmagában nem jelzi mindig a mondat elejét. Biztosabb támpont, ha hozzá egy kötőszó is társul. A kötőszavak két csoportra oszthatók: egyesek csak mondatokat, mások viszont mondatokat és mondatrészeket is összekapcsolhatnak.

A fentiek miatt első közelítésben vessző és kötőszó kombinációkat keresünk, és ezeknél próbáljuk szétvágni a mondatokat. Ez sem olyan egyszerű azonban.

Az idevonatkozó szabályok a következők: a kötőszavak elé általában vessző kerül, kivéve a következő eseteket:

A. Az *és, s, meg, vagy* szó elé csak akkor kerül vessző, ha:

1. mondatokat választ el. (MHSZ 243/a)
2. hátravetett kötőszavas mondatrész előtt áll. (MHSZ 247/e)
3. páros illetve többszörös kötőszó második, harmadik stb. tagja. (MHSZ 247/f)

B. Ha több kötőszó kerül egymás mellé, csak az első elé teszünk vesszőt. (MHSZ 261)

C. A mondatkötőszó elé csak akkor kerül vessző, ha a mondat élén áll. Vannak ugyanis olyan kötőszavak, amelyek csak a mondat elején állhatnak, pl. *bár, hiszen, noha* stb.; más kötőszavak viszont a mondat második helyén is szeretnek állni: *tehát, ellenben, ugyanis* stb.

(További probléma az, hogy ami az egyik szótár szerint kötőszó, az a másik szerint nem föltétlenül az. Vesd össze pl. az *aztán, mielőtt, ezenkívül* és hasonló szavakat az Értelmező Szótárban, a Gyakorisági Szótárban, Magyar-Angol szótárban stb.)

A kettőspont használata:

A szerző neve és a mű címe között.

Bizonyos értelmezők - hátravetett jelzők - előtt, különösen, ha többet sorolunk fel.

Pl. *S égtek lelkemben kis rőzse-dalok:*

Füstösek, furcsák, búsak, bíborak..

Következtető vagy magyarázó tagmondatok előtt.

Pl. *Közös célunk: egyre jobb eredményeket elérni.*

És végül, a közös figyelemfelhívó mondat után, ha utána felsorolás-szerűen több mondat vagy több bekezdés következik. Ilyenkor már a kettőspontot követő első mondatot is nagybetűvel kell kezdeni. (Lásd például ezt a bekezdést.)

Bizonyára ez is az egyik oka annak, hogy az emberek nemigen ismerik, és nemigen tartják be ezeket a szabályokat. Olyankor, amikor nem egy többszörösen lektorált szöveget kell elemeznünk, nem számíthatunk nagyon arra, hogy a mondatokat elválasztó vesszők helyesen vannak kitéve. De ha még így is lenne, akkor is gondot okoz a mondatelválasztó vesszők és mondatrész-elválasztó vesszők megkülönböztetése.

Továbbá, a magyar nyelv nagyon szereti a közbevetéseket. A vonatkozó névmással kezdődő mondatoknak körülbelül a fele, és a *hogy*-os mondatok jó része is egy másik mondat közepébe van beszúrva (azután a mondatrész után, amelyet részletesebben kifejt). Ezáltal természetesen fizikailag kettévágja a főmondatot.

Így tűnik, a mondat tagmondatokra való szétszedésének legbiztosabb útja az, ha bizonyos mondat elejét jelző kötőszavaktól indulunk el, és innen kiindulva próbáljuk megtalálni a tagmondatot lezáró írásjelet. Az így kapott tagmondat-elemzéseket csak azután próbáljuk meg beilleszteni eggyel magasabb szinten egyetlen egységként a mondatba.

A módszer ellen szól, hogy egyrészt a kötőszó nincs mindig kirakva, másrészt jónéhány olyan kötőszó van, ami szeret a mondatban az első szintaktikai egység után, a második helyen állni. Pl.:

A kettőspont nemcsak mondat belsejében állhat, hanem a mondat végén is.

Összefoglalva tehát azt mondhatjuk, hogy az összetett mondat elemzése azért igen nehéz feladat, mert az összetett mondatot akkor tudjuk viszonylag egyszerűen elemezni, ha a benne szereplő tagmondatokat felismertük; a tagmondatokat pedig akkor lehet jól felismerni, ha a mondat bizonyos összetartozó mondatrészeit már felismertünk. A jövő útja valószínűleg az lesz, hogy az összetett mondat elemzésére egy új, az egyszerű mondatokénál bonyolultabb stratégiát kell kidolgozni. Jelenleg viszont csak azokat az összetett mondatokat tudjuk elemezni, amelyeket a kötőszók és írásjelek segítségével egyszerű mondatokra tudtunk bontani.

A tagmondatok viszonya

"A mai magyar nyelv" 367. oldalán álló meghatározás szerint: "Az alárendelő összetétel két tagmondata közül az egyik a másik valamelyik mondatrészét fejti ki. A mellérendelő összetétel tagmondatai egyenrangúak, egyik sem szerepel a másik valamilyen mondatrészének értékében."

Ebből a meghatározásból az első rész, az alárendelés viszonylag világos, különösen, ha megmondjuk, mit értünk kifejtés alatt. A második rész viszont cseppet sem világos. Vajon miért teszünk több állítást egy mondatba? Mi kapcsolja össze ezeket az állításokat egy mondattá?

A dologban nem az a lényeg, hogy az egyik tagmondat sem része a másiknak, hanem az, hogy ezek a mondatok nem függetlenek, hanem össze vannak kapcsolva egymással, és az egész mondat jelentése több, mint az egyes tagok jelentésének összege. Világosan látható ez bizonyos kötőszavak esetén, például amikor logikai operátorokkal (*és*, *vagy*, *vagy...vagy*, *ha...akkor*, stb.) vannak összekapcsolva.

Amikor az összetett mondatban a tagmondatokat összekapcsoljuk, nemcsak az egyes állítások tényeit kapcsoljuk össze, hanem szándékot, elvárást, előfeltételezéseket is kifejezünk. Pl.:

Az ital hideg volt, és üdítő.
Az ital hideg volt, tehát üdítő.
Az ital hideg volt, de üdítő.

Mi tehát kétféle összetett mondatot különböztetünk meg: az egyikben van egy főmondat és egy almondat, ami ennek egy mondatrészét fejti ki; a másikban két mondatot kapcsoltunk össze, és ez kapcsolat több, mint két állítás összege.

Sajnos a kétféle kapcsolat nem válik el egymástól élesen. Nem véletlen az, hogy a *ha...akkor* kötőszópárt tartalmazó mondatokat "A mai magyar nyelv" című könyv sajátos jelentéstartalmú időhatározónak minősíti, hiszen azt a mondatot, hogy:

Ha fagy, hideg van.

azt úgy is mondhatjuk, hogy;

Amikor fagy, hideg van.

Az almondatok

A mondatrész-kifejtő összetételeket számba véve a következő lehetséges konstrukciókat találtuk:

- Vannak *hogy*-os mellékmondatok. A *hogy*-os mellékmondat a mondat egyik alkotóeleme helyett áll. Ha a mondatnak igei jellegű állítmánya van, és a mellékmondat ehhez az igéhez kapcsolódik, akkor ez az ige eldönti, melyik argumentumaként állhat *hogy*-os mellékmondat. Pl.:

Azt mondta, hogy ... /tárgyi mellékmondat/

De nem minden tárgy helyén állhat *hogy*-os mellékmondat. Pl.:

**Azt ette, hogy ...* /ilyen mondat nincs/

Az adott argumentum helyén a főmondatban utalószó (*az*) vagy utalószóval nyomatékosított névszói kifejezés állhat, de lehet üres is. Pl.:

Látta azt az esetet, hogy ...

Azt látta, hogy ...

Látta, hogy ...

Ha névszói kifejezés egy jelzője helyett áll a mellékmondat, a névszói kifejezésbe többnyire utalószó (*olyan, annyi*, stb.) kerül a jelző helyére. Pl.:

Látott egy olyan esetet, hogy ...

A névszói mondatokban, *hogy*-os mellékmondat egy határozott névszói kifejezés (rendszerint az alany) helyett áll. Pl.;

Világos, hogy...

(*Világos a helyzet, hogy...*)

Az a végeredmény, hogy...

- Vannak vonatkozó névmás-utalószavas mellékmondatok. Itt a mellékmondat-főmondat pár tulajdonképpen két független mondat, amely úgy kapcsolódik össze, hogy a főmondat utalószóval megjelölt mondatrésze azonos a mellékmondat vonatkozó névmással megjelölt mondatrészével. Pl.:

Megnéztem azt a házat, amiből nemrég költöztetek ki.

/azt –amiből/

A vonatkozó névmás helyett a mellékmondatban vonatkozó határozó is állhat. Pl.:

Láttam azt a házat, ahol lakott.

A főmondatból az utalószó el is maradhat. Ilyenkor a vonatkozó névmásoknál és egyes vonatkozó határozószavaknál gondot okoz annak megállapítása, hogy mire is vonatkozhat a mellékmondat. Más vonatkozó határozóknál viszont (*amikor, mielőtt, ahogy, valahányszor*, stb.) nagy valószínűséggel megjósolható, hogy milyen utalószó tartozik hozzá (*akkor - amikor*, *akkor - mielőtt*, *úgy - ahogy*, *mindannyiszor - valahányszor*, stb.) és ha elmarad, úgy tekinthetjük, mintha ott állt volna a főmondatban.

Pl. *Odaért, mielőtt a többiek megérkeztek volna.*

- Az összehasonlító mondatokban rendszerint egy utalószóval ellátott névszói kifejezéshez kapcsolódik egy *mint* kötőszóval kezdődő mellékmondat. Ezek kezelését már korábban leírtuk.

Az alárendelt mondat elejének felismerése viszonylag könnyű feladat. A mondatot egy *hogy* vagy *mint* kötőszó, vagy egy vonatkozó névmás, vagy egy vonatkozó határozószó vezeti be. (A *hogy* sajnos néha elmarad - ilyenkor bajban vagyunk.) Nehezebb a mondat végének felismerése. Ha vége a mondatnak (pl. ponttal zárul az alárendelt tagmondat), nincs probléma. Ha utána újabb alárendelt mondat következik, az az alárendelt mondatnak lesz alávetve. Más a helyzet viszont, ha vesszőt találunk, mert a vessző sok különböző okból kerülhet oda, és csak egy lehetséges cél, hogy bezárja az almondatot. Az is elég kényelmetlen, ha felismerjük, hogy egy mellérendelő mondat következik utána, mert ez egyaránt lehet a főmondat vagy az almondat mellé rendelve.

A mellérendelés

Amikor kötőszavakkal mondatokat kapcsolunk össze, tisztáznunk kell, hogy milyen kötőszavak milyen kapcsolatot hoznak létre.

Ha megmaradunk a tények szintjén, és eltekintünk a szándékoktól, előfeltevésektől és hasonlóktól, és például logikai formulákkal írjuk le a mondat jelentését, azt tapasztaljuk, hogy számos különböző összetétel ugyanazt jelenti. Pl.:

Az ital hideg volt, és üdítő.

Az ital hideg volt, tehát üdítő.

Az ital hideg volt, de üdítő.

Mind azt jelenti, hogy:

Az ital hideg volt. & Az ital üdítő volt.

Ahol az "&" jel a logikai konjunkciót jelöli.

Ez a kisebb gond. A nagyobb gond az, hogy a logikai műveleteket bizonyos kötőszavakkal szokták összefüggésbe hozni, és ez egyesekben valami egészen primitív megfeleltetés képzetét kelti. Ha *és*, akkor konjunkció, ha *vagy*, akkor diszjunkció, stb. Ez sajnos egészen téves elképzelés. Például annál a mondatnál, hogy:

Egész nap evett, ivott, és ágyban heverészett.

Első látásra úgy tűnik, hogy konjunkcióról van szó, de ha egy kicsit pontosítjuk a dolgot, és átfogalmazzuk olyan formába, hogy:

Egész nap (vagy éppen evett, vagy éppen ivott, vagy éppen ágyban heverészett).

akkor kiderül, hogy a három dolog, kizáró vagy-nak nevezett logikai művelettel van összekötve. És ha tovább gondoljuk a dolgot, valószínűleg úgy kell átfogalmaznunk a mondatot, hogy

A nap nagy részében (vagy éppen evett, vagy éppen ivott, vagy éppen ágyban heverészett).

Az angol és orosz irodalomban meglehetős részletességgel foglalkoztak azzal a kérdéssel, hogy milyen kötőszavakkal képzett összetételek milyen logikai kapcsolatokhoz vezetnek. Ez kiindulópont lehet egy hasonló magyar munkához is, de a nyelvek számos dologban különböznek. Elég, ha a kettős tagadásra utalunk.

A dolog másik szépséghibája, hogy az egyszerű logikai szemantikaleírás elég primitív modellt ad a világról. Ha valami bonyolultabb modellben kívánunk dolgozni (időkezelés, intenzionális logika stb.), vagy ha valami nem bonyolultabb, de eltérő leírást akarunk alkalmazni, az összetétellel kapcsolatos megállapításainkat újra kell gondolni, átalakítani, finomítani kell a modellt.

Kérdéses, hogy félretéve a jelenleg használt szemantika-definíciókat nem lehetne-e egy olyan modellt csinálni, amely a létező mondatösszetételekre pontosabban leírná mindazokat a lényeges kapcsolatokat, amik az összetett mondatokban fennállnak; és akkor ebből lehetne származtatni a konkrét szemantika-modellekben a kapcsolatokat.

A VONZATOKRÓL

Noha minden szónak önmagában is van valamilyen jelentése vagy több lehetséges jelentése is, a szó igazi és viszonylag teljes jelentését csak a szövegben szereplő többi szóhoz való kapcsolatában tudjuk megadni.

A mondatban megadott kapcsolatok két csoportba sorolhatók:

Vannak olyan kapcsolatok, amelyek pusztán összegzik, összekapcsolják a szavakat, szerepeket jelölnek ki a bennük szereplő szavaknak. Ilyenkor a kapcsolatban szereplő bármelyik szó más szavakra cserélhető ki, a kapcsolat milyensége nem változik. Például a piros alma jelzői kapcsolatban a piros jelző helyett számos jelző állhatna, és hasonlóan a jelző nemcsak az almára, hanem számos más tárgyra is vonatkozhatna.

Vannak viszont olyan kapcsolatok, ahol van egy domináns elem, és ez határozza meg, hogy milyen más dolgoknak lehet vagy kell belekerülnie a mondatba. Ha a domináns szó helyére egy másik szó kerül, akkor ennek kapcsán másfajta kapcsolatok kerülnek a mondatba. Ezt a fajta kapcsolatot nevezzük vonzatnak.

Tipikusan ilyen szavak az igék. Az ige megszabja, hogy milyen más mondatrészek (alany, tárgy, határozók) kerülhetnek be a mondatba, és az utóbbiakat milyen raggal vagy névutóval kell megadni. Pl.:

árt valamivel valakinek
bead valamit
valahova csatlakozik valamihez
csökkent valamit valamennyivel
definiál valamit valamiként
stb.

De nemcsak igéknek vannak vonzataik, hanem számos más szónak is. Pl.:

főnevek:

példa valamire
fiú valahonnan
játék valakivel

melléknevek:

bűnös valamiben
divatos hol mikor
független valakitől
terhes valakitől

névutó:

később valaminél
belül valamin
közel valamihez

Tapasztalatunk szerint, ha egy szónak van valamilyen vonzata, akkor ez a vonzat gyakran öröklődik a belőle képzett szavakra is. Pl.:

valakinek valamivel ártó
függetlenít valakitől
valaminél későbbi

A morfológiai elemzésnél a szóról azokat a képzőket érdemes levágni, amelyeknél meg tudjuk mondani, hogy ha egy szónak bizonyos vonzatai vannak, akkor milyen vonzatai lesznek a képzett szónak. Az igéknek ebben az értelemben reguláris képzőiről az igenevekről szóló fejezetben van szó. Más képzőkkel ebben a fejezetben foglalkozunk.

Formai oldalról nézve a mondatban szereplő kapcsolatokat két csoportba szoktuk sorolni: olyanokra, amelyeket a szavak sorrendje határoz meg, és olyan kapcsolatokra, amelyeket a ragok jelölnek ki többnyire kötetlen szórendben. A fenti vonzatok ragokkal (és névutókkal) vannak kijelölve. Van néhány pozícionálisan kijelölt viszony is, amit szintén vonzatnak nevezünk, például az *-ú/-ű* végű mellékneveknek szinte mindig van jelzője. Pl.:

milyen című
alakú
hatású
típusú
mekkora igényű
értékű
területű

Itt egyértelműen jelzői kapcsolatról van szó, ahol a jelző kötelező, és a jelzett szó előtt áll. (Az esetek nagyobb részében a melléknévnek nem lehet jelzője, lásd a **szófajok**ról szóló fejezetet.) Az ilyen jelzői vonzat másképpen viselkedik, mint a ragos vonzat, például nem öröklődik.

Igei vonzatok

Az igei vonzatokat elég részletesen leírtuk a szemantikáról szóló Igei és névszói relációk című fejezetben, és a szintakszisról szóló az Egyszerű mondatok szerkezete című fejezetben. Itt most egy rövid összefoglalót adunk. Az igei vonzatokat a következő csoportokba soroljuk.

Kulcsszavas vonzatok

Az ige jelentése megváltozik, ha egy adott szó áll egy adott raggal mellette. Pl.:

állást foglal valakivel szemben

helyet foglal valahol

magában foglal valamit

Névutós vonzatok

Az ige valamelyik argumentumát egy adott névutóval jelöli ki. A névutó dönti el, hogy hogyan kapcsolódik a névszói kifejezéshez (alanyesetben, raggal, birtokviszony-szerűen). Pl.:

elkövet valamit valakivel szemben

elkövet valamit valaki ellen

Ragos vonzatok

Az ige valamelyik argumentumát egy esetraggal jelöli ki. Pl.:

kér valamit valakitől

kér valakit valamire

Csoportos vonzatok

Az ige valamelyik argumentumát több azonos csoportba tartozó (pl. Hol kérdésre válaszoló) ragos vagy névutós kifejezés, illetve az ebbe a csoportba tartozó határozó szó adhatja meg. Pl.:

küld valamit + {részes határozó}

Levelet küldött Péternek.

Levelet küldött Péter számára.

Levelet küldött Péter részére.

Főnévi igeneves vonzatok

Az ige egy bizonyos argumentumaként csak ragozatlan melléknévi igenév állhat. Pl.:

elfelejt valamit csinálni

Hogy-os mellékmondatok

Az ige egy bizonyos argumentumát egy *hogy* kötőszavas mellékmondattal lehet részletesebben kifejteni. Pl.:

látta azt, hogy ...

Névszói vonzatok

A főnevek vonzatai

A főneveknek elég gyakran van ragos vonzata. A főnév ragos vonzata rendszerint az illető főnévvel záródó névszói kifejezés mögött áll. Ha a vonzatot jelzősíteni akarjuk, akkor töltelék melléknévi igenevet kell beiktatnunk, vagy melléknévi névutóvá kell alakítanunk a névutót. Pl.:

eljárás valaki *ellen*

valaki *elleni* eljárás

valaki *ellen foganatosított* eljárás

Részletesebben lásd a **szófajok**nál a **főnevek**ről szóló alfejezetet és az **Igenevek** című fejezetben az ***-ás/-és* végű főnevek**ről szóló fejezetet.

A melléknevek vonzatai

Elég sok melléknévnek van ragos vonzata. Ez rendszerint a melléknév előtt áll, például ha a melléknév egy nagyobb névszói kifejezés része, viszont bárhol állhat a mondatban, ha a melléknév állítmányi szerepű.

Az eddig felsorolt névszói vonzatok lehetnek kulcsszavas, ragos, névutós és csoportos vonzatok is. Főnévi igeneves vonzat is van, ha helyes magyar kifejezésnek tartjuk például azt, hogy:

egy magyar nyelvet tanulni akaró diák.

Amikor a melléknévből határozószót képzünk, fokozzuk stb., a melléknév vonzatai megmaradnak. Pl.:

Az ügyben a polgármester bűnösebb volt a vállalkozóknál.
Az ügyhöz csatlakozólag egy másik eljárás is elindult.

A melléknév előtt az esetek döntő többségében nem áll jelző, különösen, ha a *jó nagy*, *oltári szép*, *tök jó* és hasonló kifejezésektől eltekintünk. Mégis elég sok olyan melléknév van, ami előtt kötelezően vagy opcionálisan kell vagy lehet jelzőnek állnia. Ezt a szintakszis-elemzés során jó tudni, mert ha a jelző kötelező, akkor biztos, hogy a melléknévhez tartozik és nem a főnévhez, ha pedig opcionális, akkor tartozhat ide is, oda is; végül ha nincs vonzata a melléknévnek, akkor a melléknév előtt álló jelzőt a főnév jelzőjének tekintjük. Pl.:

3 elemű halmaz = (3 elemű) halmaz

3 mázsás súlyú szikla = ((3 mázsás) súlyú) szikla
avagy 3 ((mázsás súlyú) szikla)

3 sötét ló = 3 (sötét ló)

A jelzőket csoportokba sorolhatjuk, és ettől talán még pontosabb lesz az elemzés. Mi a következő jelző-csoportokat különböztetjük meg.

Bármilyen névszó. Pl.: *alma alakú*
hosszúkás alakú
háromalakú

Mennyiség (szám és mértékegység).
Pl.: *3 négyzetméter területű*
3 méter hosszú
3 forint értékű

Szám. Pl.: *3 rétegű*
3 ajtós
3 szintű

Sorszám. Pl.: *3. generációs*
3. szintű
1989. évi

Név. A *című*, *nevű*, *gyártmányú* stb. szavak előtt egy tulajdonnév szokott állni, ami több szóból is állhat, de feltehetőleg nagybetűvel kezdődik. Ezt különösen akkor érdemes tudni, ha a tulajdonnevek nincsenek a szótárban.

A melléknevek középfokának is van vonzata. Ez nem a melléknévé, hanem a középfoké:

Valaminél valamivel (jobb, kevesebb, erősebb ...)

Névutók vonzatai

A névutók egy része valamilyen ragot vonz. Pl.:

valamivel szemben
valamihez közel
valamin túl
stb.

Ilyenkor a ragos névszói kifejezés elszakadhat a névutótól, leggyakrabban a névutó mögé szokott kerülni.

Egy sereg névutóból képezhető melléknévi névutó. A melléknévi névutó vonzata azonos az eredeti névutóéval.

Mivel a névutók egy másik jelentős része látszólag birtokos viszony, felmerül a kérdés, vonzat-e a birtokviszony. Számos érvet lehetne felsorolni pro és kontra. Ha vonzat, igen speciális vonzat, ha nem vonzat, vannak vonzatszerű tulajdonságai is. Lásd a **névutókról** szóló fejezetet.

Néhány névutónál, igénél és egyéb szónál felmerült, hogy többes számot vonz. Ez a többes szám azonban logikai és nem grammatikai többes szám, az elemzésnél nagy körültekintéssel kezelendő. Pl.:

Kössük össze a pontokat.
Kössünk össze két pontot.
Kössük össze az A és a B pontot.
Kössük össze az A pontot a B-vel. (Lásd **társhatározó**!)
Kössük össze a fákat.
Kössük össze a fát.

Természetesen mind a névszóknál, mind az igéknél, egy szónak több különböző vonzata lehet.

A vonzatok súlya és kiválasztása

Az, hogy egy vonzat mennyire befolyásolja az eredeti szó jelentését, elég tág határok között mozog: az idiómától a szabad határozóig. Ennek az a következménye, hogy egyes vonzatoknak kötelező ott állni az illető szó mellett, mert különben a szó értelmetlen vagy egész mást jelent. Más vonzatoknál a szó ugyan nem értelmetlen, de az egész kifejezést hiányosnak érezzük, ha nincs ott a vonzat. Vannak olyan vonzatok, amelyek hiányában valamilyen default lép be. Végül vannak olyan vonzatok, amelyek csak egy kicsit pontosítják a szó jelentését és el is hagyhatók.

A szóhoz általában nem egyetlen szó kapcsolódik, hanem több szó, illetve kifejezés egyszerre. Mi vonzatkeretnek vagy vonzatprofilnak nevezünk egy olyan vonzatkészletet, amibe a szó adott értelmezése mellett már több vonzatot nem lehet elhelyezni.

A dolog másik oldala, hogy egy szónak rendszerint nem egyetlen vonzatkerete van, és nem ugyanazt jelenti az egyik vonzatkerettel, mint a másikkal. Tapasztalatunk szerint igen ritka az az eset, amikor egy teljes vonzatkeretet megtalálunk a szó mellett. Rendszerint az egyik keretből többet, a másikból kevesebbet találunk meg. Kérdés ezután, hogy melyiket tekintsük az igazinak. Tapasztalataink azt mutatják, hogy ha a megtalált vonzatokat pozitív és a hiányzó vonzatokat negatív súllyal vesszük figyelembe, a fontosabbakat nagyobb, a lényegtelenebbeket kisebb súllyal, akkor már a szintakszis analízis végén elég jól megjósolható a mondat igazi szerkezete, de legalábbis számos valószínűtlen változatot ki lehet zárni.

Maradt a kérdés, hogy hogyan állapítsuk meg az egyes vonzatok súlyát. Van-e összefüggés a vonzat típusa és a vonzat súlya között? A mi rendszerünkben van. A vonzat súlyát két jellemzőből kalkuláljuk ki: az egyik megadja, hogy az illető vonzat kötelező-e vagy opcionális, a másik a vonzat típusa (kulcsszavas, névutós, ragos stb.).

A kötelező kulcsszavas vonzatnak rendkívül nagy súlya van. Ha a kulcsszó nem áll ott, akkor ez a vonzat szóba sem jön. Ha ott van, a szóban forgó vonzat, kiválasztása biztos, de esetleg több ilyen vonzat is lehet.

Az opcionális kulcsszavas vonzat tulajdonképpen egy jellemző töltelékszó. Ha ott van, az jelentősen megerősít bennünket abban, hogy szóban forgó vonzatkeretről van szó, de a hiánya nem jelent semmit.

Általában az opcionális vonzat hiányáért nem jár büntetés, viszont megléte megerősít bennünket abban, hogy a valódi vonzatkeretet találtuk meg.

A névutós szerkezet súlyosabb, mint a rag. Ezt azért tartjuk igaznak, mert azt tapasztaltuk, hogy az elemzés általában jobb, ha a névutós szerkezetet egyetlen egységként kezeljük, mintha két különálló valaminek vennénk. Pl.:

Velem szemben futott a ló.

Egy lehetséges félreelemzése:

Velem futott a ló. Hol? Szemben.

Két órával ezelőtt érkezett meg.

Hasonlóan: *Két órával érkezett meg. Mikor? Ezelőtt.*

Nem ilyen egyértelmű a helyzet persze, a névszói kifejezés és a névutó nem ilyen sorrendben áll. Pl.:

Velem futott a ló szemben.

Ezelőtt két órával érkezett meg.

Ilyenkor viszont ezt a kapcsolatot csak a másodlagos vonzatok keresése során találjuk meg és ezért a súlyozás is másként alakul. A ragos és névutós vonzatról általában nem mondható el, hogy súlyosabb, mint a csoportos vonzat. Pl.:

A disszertációján dolgozik. (min?)

Az egyetemen dolgozik. (hol?)

Viszonylag egyszerű szemantikai jegyeket figyelembe véve, (élő - élettelen - fogalom, hely - idő, stb.) azt mondhatjuk, hogy ha az illető argumentum szemantikailag illik a vonzathoz, akkor a vonzat mindig előnyben van a szabad határozóval szemben. Pl.:

Segített Pistán.

Senki nem gondolja, hogy rajta állva vagy ülve.

A szabad határozó elhelyezése a mondatban nem jelent semmit a vonzatkeret kiválasztása szempontjából. Ha viszont maradt valami a mondatban, amit nem tudunk sehova elhelyezni, annak nagy negatív súlya van az adott vonzatkeretre nézve. Ezt a súlyozásos mondatelemző stratégiát több nyelvésznek elmondtuk, és a fogadtatás meglehetősen vegyes volt. Voltak, akik teljesen egyetértettek vele, voltak, akik igen veszélyesnek tartották. A gyakorlatban meglehetősen jól bevált. Ahol nem, ott a vonzatoknak egyszerű szemantikai jegyekkel való kiegészítése, és ezek szintaktikai elemzés közbeni ellenőrzése még sokat javíthatna az elemzés hatékonyságán.

A NÉVUTÓKRÓL

Ebben az írásban a névutókról és a névutószerűen viselkedő szavakról lesz szó, továbbá a ragok és névutók viszonyáról. Mindezek alapján a névutók néhány gyakorlati jelentőségű osztályozásáról is szó lesz.

Az Értelmező Kéziszótár szerint:

A névutó névszó után álló, vele együtt határozói szerepet betöltő viszonyszó.

A fenti definíció elég tág, sok minden belefér:

> *az asztal alatt, a fák között, a fa között, az olvasottak szerint, vasárnap óta, a folyón túl, a ház körül, a házhoz közel, velem szemben, az ajtón kívül, az érvek ellenére,* ? *az alperes javára,* ? *Péter kedvéért,* ? *a felperes sérelmére, stb.*

Valószínű, hogy a fenti sorozat első három elemét bárki habozás nélkül elfogadja névutós szerkezetként; de az utolsó kettőt valószínűleg igen kevesen tartják névutós szerkezetnek, holott jogi szövegekben a sérelmére szó használata szinte azonos azzal, ahogy például a tudományos szövegek a szerint szót használják. Azaz van olyan ige (pl. az *elkövet*), aminek ez a vonzata, és szinte minden mondathoz hozzácsapható szabad határozóként, hogy az valakinek a *sérelmére* történt.

Ha a fenti kapcsolatok mindegyikét névutós szerkezetnek tekintjük, akkor a névutókat két különböző kritérium szerint is csoportosíthatjuk:

Az első megkülönböztető elv, hogy hogyan alkalmazzuk a névutót a személyes névmásokra:

> *alattam, közöttünk, szerintem, körülöttem, ellenedre, javamra, sérelmemre* stb.
>
> *rajtam túl, hozzám közel, velem szemben, rajtunk kívül* stb.

Látható, hogy az esetek egy részében a névutó változatlan formájú, és a személyes névmás veszi fel az odaillő alakot, az esetek másik felében viszont a személyes névmás nem jelenik meg, hanem a névutó kap személyragot, végül vannak esetek, amikor a névutó nem járulhat személyes névmáshoz.

Ha azt is megvizsgáljuk, hogy milyen személyragról van szó, azt tapasztaljuk, hogy ezek formailag egybeesnek a birtokos ragozás végződéseivel.

A második megkülönböztető elv, hogy ragos vagy ragtalan névszó után áll a névutó:

az asztal alatt, a fák között, a fa között, az olvasottak szerint, vasárnap óta, az érvek ellenére,

a folyón túl, velem szemben, rajtam kívül, a házhoz közel, ehhez képest,

kinek a részére, az eset kapcsán, a baleset következtében, ? a felperes sérelmére, stb.

Az első csoportra az jellemző, hogy a névutó mindig a ragtalan névszó után áll. A második csoportra az jellemző, hogy a névszó egy a névutótól függő ragot kap. Ez lehetővé teszi, hogy a mondatban a névszó és a névutó elszakadjon egymástól.

Megpillantotta a házat túl a folyón.

Velem lakik szemben.

Kívül lesz az ajtón.

A harmadik csoportba tartozó névutószerű szavak, noha tartalmilag már esetleg jelentős változást szenvedtek, formailag birtok és esetragos főnevek. A kapcsolat formailag birtokviszony, amely akár *nak*-ragos, akár ragtalan lehet. Hogy ez a kapcsolat mégsem birtokviszony, hanem inkább névutói, az az általunk abszurd birtokos szerkezetnek nevezett jelenségben mutatkozik meg. Pl.:

Péter fiának a tanulása alapján...

Ez egy háromszoros vagy négyszeres birtokviszony. Ezt a magyar helyesírás szabályai szerint úgy kell leírni, hogy az utolsó birtokviszony *nak*-ragos, a többi viszont alanyesetű (Péter fia tanulásának a kedvéért). A gyakorlat is azt mutatja, hogy ez a leggyakoribb forma, bár elég gyakran előfordul a tisztán alanyesetű is, illetve az is, amikor egy darabig csupa alanyesetű birtokviszony jön, majd ezeket csupa *nak*-ragos birtokviszony követi. Ilyen forma, ahol egy alanyesetű birtokviszonyt egy *nak*-ragos követ, majd ismét alanyesetű birtokviszony következik, tapasztalatunk szerint csak akkor fordul elő, amikor az utolsó kapcsolat átmenet egy névutós és egy birtokos reláció között.

Mivel a két felosztási elv független, sőt ortogonális, (azaz mindegy, hogy először az egyik, majd azután a másik szerint osztjuk fel a szavakat, vagy fordítva); készítsük el a névutók keresztfelosztását.

	személyragozó	nem személyragozó
ragtalan névszók	*alatt,alól,alá, által,ellen, előtt,elől,elé, felett,fölött,fölül, fölé, felől,felé, helyett, iránt, között,közt,közül,közé, körülötte#,köré, mellett, mellől,mellé, miatt, után, mögött,mögül,mögé, nélkül, szerint, után,*	*hosszat, körül#, közben, közepette, mentén, múlva, múltán, óta, tájban, végett,*
ragos névszók	*kívül,*	*át, belül, együtt, előbb, ezelőtt, felül, fogva,folytán,függetlenül, kezdve,képest,keresztül, közel,külön,messze,túl, szemben,*
birtokos kapcs.	*javára, részére, számára, révén, ?érdekében,?kedvéért, ?sérelmére,* stb.	*alapján,ellenére,folyamán, kapcsán, következtében, nyomán, során,*

(# Megjegyzés: a *körül* névutó *körülöttem*, *körülötted*, stb. formában személyragozódik.)

Kommentárok:

A személyragozó, ragozatlan névszókat követő szavak egyértelműen névutók, mert mindig a névszó után állnak, nem állnak meg önmagukban, csak együtt alkotnak határozói viszonyt.

A nem személyragozó, de ragozatlan névszót követő névutók jelentésüknél fogva nem kapcsolhatók a személyes névmásokhoz. Nagyrészük időhatározói viszonyt fejez ki.

Egyetlen olyan névutót találtunk a *kívül*-t, amelyik személyragozható, és mégis *-on/-en/-ön* ragot vonz. Ez a személyes névmásokra kétféleképpen is alkalmazható: *kívülem*, *kívüled*, ... vagy *rajtam kívül, rajtad kívül*, ...

A nem személyragozó, de ragos névszóval járó névutókra általánosan jellemző, hogy másnak is felfoghatók. Vannak, amelyek igekötőnek is tekinthetők, ebben az esetben az igekötős igének rendszerint van olyan vonzata, amelyik formailag egybeesik a névutó vonzatával. Pl.: *Nézett az üvegen keresztül*. Mások határozószónak tűnnek. Pl.: *Már*

ezelőtt is találkoztak. Már három évvel ezelőtt is találkoztak. stb. Még tovább erősíti a kételyt az iránt, hogy ezek valódi névutók, hogy gyakran fordított sorrendben: névutó, vonzott szó sorrendben állnak. Pl.: *túl a folyón, szemben velem, függetlenül a véleményedtől* stb. Sőt néha még sokkal messzebbre is elszakadnak a mondatban egymástól.

A birtokos kapcsolatra emlékeztető névutók két osztálya a legkevésbé körülhatárolható. Vannak benne szavak, amelyek már egészen elvesztették eredeti jelentésüket pl.: *révén, kapcsán, terén,* stb., mások jelentése szinte változatlan pl.: *segítségével, érdekében, következtében*, stb. Valószínűleg azért tűnnek névutónak, mert nagyon gyakran használjuk őket, és mindig a névszó után állnak, még akkor is, ha az *–nak* ragot kapott. Sokszor fordulnak elő abszurd birtokos szerkezetben. A különböző szak- és csoportnyelvekben valószínűleg más és más ezen szavak osztálya.

Ragokhoz tartozó névutók

Már a körül névutó esetén is láttuk, hogy az nem ebben a formában, hanem körülötte alakban kapja a személyragokat. Hasonló a helyzet az úgynevezett esetragokkal is. Van olyan rag, ami a személyes névmásokra alkalmazva önálló szóként ragozódik, pl. a *-val/-vel velem*, *veled*, ... alakban. Másoknál kis eltérés tapasztalható az esetrag alakja és a hozzátartozó személyragos "névutó" formája között pl.: *-ba/-be, belém*, *beléd*,... Megint másoknál a formák közötti eltérés igen nagy, pl. *-on/-en/-ön* esetén *rajtam, rajtad, rajta*, ... A következő felsorolás mutatja melyik raghoz milyen "névutó" tartozik. Ez a "névutó" csak személyragozott formában él.

-ban/-ben	*bennem,*	*benned,*	*benne*, ...
-ba/-be	*belém,*	*beléd,*	*belé*, ...
-ból/-ből	*belőlem,*	*belőled,*	*belőle*, ...
-on/-en/-ön	*rajtam,*	*rajtad,*	*rajta*, ...
-ra/-re	*rám,*	*rád,*	*rá(ja)*, ...
-ról/-ről	*rólam,*	*rólad,*	*róla*, ...
-hoz/-hez/-höz	*hozzám,*	*hozzád,*	*hozzá*, ...
-tól/-től	*tőlem,*	*tőled,*	*tőle*, ...
-ért	*értem,*	*érted,*	*érte*, ...
-val/-vel	*velem,*	*veled,*	*vele*, ...
-nak/-nek	*nekem,*	*neked,*	*neki*, ...

Nem találtunk a raghoz illő névutót a *-vá/-vé*, *-ként*, *-nként*, *-kor*, *-ig*, *-szor/-szer/-ször*, *-ul/-ül* ragokhoz. Felmerülhet egyesekben a gondolat, ezek egyike-másika nem is rag. Mi azért tartjuk ezeket a végződéseket ragoknak, mert mindegyikre találtunk olyan példát, amikor ez a végződés többes vagy birtokos jel után állt. Pl. *A király*

fiául fogadta a juhászt. És mindegyikhez találtunk olyan igét, aminek ez a rag a vonzata.

Összefoglalás

A fentiek alapján mi a személyragozó, ragos vagy ragtalan névszókat követő szavakat, illetve az esetragok személyragozott párját nevezzük névutóknak, a többi névutói szerepet betölteni képes szót névutószerű szónak nevezzük. A névutók ragozhatók a birtokos személyragokkal, de más végződést nem kaphatnak.

Az elnevezés persze nem lényeges dolog. Minden névutót és névutószerű szót 3 attribútummal jellemezhetünk:

- vonz-e esetragot: Ha nem vonz esetragot, akkor az után a szó után áll, amelyre vonatkozik. Ha esetragot vonz, a ragos szó távol kerülhet tőle, bár leggyakrabban előtte vagy utána áll. (Névszói kifejezésen belül csak előtte állhat.)
- van-e személyragozása: A személyragozott névutók könnyen felismerhetők, előttük nem állhat semmi, a nyomatékosító személyes névmással egybeírjuk őket. A későbbi egységes kezelés kedvéért célszerű, ha átírjuk egy standard mesterséges alakra. (*alattam=én+alatt*, *mögötte =ő+mögött* stb.) Ha nincsenek személyraggal ellátva, akkor meg kell keresni, hogy mire vonatkoznak. Sokkal nehezebb az ő révén, számomra, Péter részére stb. névutószerű szavakból képzett szerkezetek kezelése. Azt feltételezzük, hogy a főnév értelmű *rév*, *szám*, *rész* szavak mellet van *rév*, *szám*, *rész* névutó is, de ez a jelentés csak akkor léphet fel, ha a szóhoz birtokos rag és megfelelő esetrag kapcsolódik. Az ezekhez tartozó argumentumot úgy kell keresni, mintha birtokos lenne. Ezeket is érdemes standard mesterséges alakra hozni.(a te *réveden=te+révén*, *számomra=én+számára* stb.)
- létezik-e személyragozatlan alakban: Ha nem létezik, akkor egy személyes névmás ragos formája, és célszerű átírni ilyen alakra. (*veled=te+vel*, *belőlünk=mi+ből* stb.)

A fenti három jellemző elég arra, hogy a szót vagy a kifejezést beleillesszük a mondat szintaktikai szerkezetébe.

IGENÉV

A Magyar Értelmező Kéziszótár szerint az igenév "az igének bizonyos képzőkkel alkotott névszói vagy határozói szerepű, de igei természetét is megőrző származéka. Fajtái: főnévi igenév *-ni* képzővel képzett főnévé határozói igenév *-va/-ve* vagy *-ván/-vén* képzővel képzett határozószó; melléknévi igenév *-ó/-ő* vagy *-t/-tt* vagy *-andó/-endő* képzővel képzett melléknév".

A fenti definíció a számítógépes elemzés szempontjából teljesen használhatatlan, ezért új meghatározást fogunk adni. Először is megmondjuk, hogy mit értünk az igei természet megőrzése alatt, ennek alapján a képzők szélesebb osztályát választjuk ki, végül megadjuk, hogy ezek a szavak hol szerepelhetnek a mondatban. (Nem ott, ahol a főnév, melléknév, határozószó.) Lehet, hogy helytelen dolog az "igenév" szónak ennyire más jelentést tulajdonítani, de úgy érezzük, hogy az általunk kialakított fogalom jól alkalmazható mindazokra a szavakra, amelyek mind a magyar nyelvben, mind más nyelvekben így nevezhetők.

Az első kérdés, hogy mit értünk az alatt, hogy igei természetét megőrzi a szó. Mi az igét egy függvénynek tartjuk, amelynek argumentumai vannak, amelyeket alanynak, tárgynak és határozóknak hívunk. Az igei természet megőrzése alatt azt értjük, hogy a keletkezett szónak megmaradnak a fenti argumentumai, bár nem biztos, hogy ugyanazokkal az eszközökkel lesznek kijelölve. Ezzel kapcsolatosan további vizsgált tulajdonságok lesznek azok, amelyek azt mutatják, hogy mennyire hasonlít a szó viselkedése a ragozott igére, azaz: személyragozható-e, szétszóródnak-e az argumentumok a mondatban, és leválik-e az igéről az igekötő. Valamint vizsgáljuk azt is, hogy beilleszthető-e egy ilyen szó egy névszói kifejezésbe. Ezek azonban másodlagos tulajdonságok.

A másik kérdés, melyek a szóba jövő képzők. Az igéből igét képzőket itt természetesen figyelmen kívül hagyhatjuk. A többi képző között nem találkoztunk olyannal, ami egy igei jelentésű szót megfosztana valamilyen argumentumától, vagy új argumentumot is rendel hozzá, bár elvileg különösen az utóbbi elképzelhető.

Alapvető kritériumnak a képzők reguláris viselkedését tartjuk. A képző viselkedése akkor reguláris, ha minden igére alkalmazható, és megjósolható módon változtatja meg a szó jelentését és az argumentumok kijelölésének módját. Ez lehetővé teszi, hogy a szótárba csak az alapszó kerüljön bele. A morfológiai elemzés során felismerjük a szót és a képzőt, a szintaktikai elemzés során viszont az alapigéből és a képzőből elő kell állítani a képzett szó összes szükséges jellemzőjét. Ez természetesen nem

elhanyagolható méretű algoritmikus feladat, viszont ennél jelentősebb a szótár csökkenéséből származó előny.

Mi a következő képzőkkel foglalkozunk: a főnévi igenév és a személyragozott főnévi igenév képzői; a határozói igenév képzői; a melléknévi igenevek képzői (azaz: *-ó/-ő* vagy *-t/-tt*, *-atlan/-etlen* vagy *-andó/-endő* vagy *-ható/-hető*, *-hatatlan/-hetetlen*) és az *-ás/-és* képző.

Nem zárható ki, hogy további képzők is vannak, amelyek a fenti módon kezelhetők. Mi eddig ezeket vizsgáltuk meg részletesebben.

A főnévi igenév

A "főnévi igenév" elnevezés igen rossz mert, roppant helytelen asszociációkat vált ki az emberekből. A főnévi igenévnek nemhogy főnévi, de egyáltalán semmilyen névszói tulajdonsága nincsen: nem kaphat névszói ragokat és jeleket; nem tölthet be a mondatban semmilyen olyan szerepet, amely normálisan egy névszó (illetve névszói kifejezés) tölt be. Puszta belemagyarázásnak érezzük mindazokat a mondatelemzéseket, ahol a főnévi igenevet alanynak, tárgynak, esetleg határozónak vélik.

A főnévi igenév helyes elnevezése "másodlagos ige" lenne.

Ez az elnevezés arra utal, hogy a főnévi igenevet mi eleddiglen kizárólag kétféle konstrukcióban láttuk szerepelni a mondatban:

- Az első esetben a főnévi igenév a mondat főigéje, de egy segédigés, illetve egy kitett vagy ki nem tett létigés konstrukcióba beágyazva alkotja a mondat összetett állítmányát. Pl.:

  Meg tudta oldani a feladatot.
  Sétálni voltam.
  El kellett mennem oda.
  Muszáj volt meglátogatnom őt is.
  Jó lett volna meglátogatni őt is.
  Tanácsos [volt/lesz/volna/lenne] meglátogatni őt is.
  Nincs mit ennie.

A példákat két csoportra osztottuk. A második csoportba tartozó konstrukciók esetén a főnévi igenév személyragozott is lehet.

- A másik esetben a főnévi igenév a mondatban szereplő ige egy olyan argumentuma, ami semmilyen más szófajjal nem tölthető be, kizárólag ragozatlan főnévi igenévvel tölthető be. Pl.:

  Elment a boltba almát venni.

Tévesnek ítéljük, ha *venni* szót a *Hova ment el?*, *Miért ment el?*, *Mi célból ment el?* és hasonló kérdések alapján hely-, ok- vagy célhatározónak minősítjük. Egyszerűen arról van szó, hogy az *elmegy* szóhoz kapcsolható egy másik ige, főnévi igenévi alakban, de például az *áll*, vagy *felír* igéhez nem. (A fentiekhez hasonló kérdések ezekre is feltehetők, de a *Mit csinálni ment el?*–nek nincs megfelelője).

Vegyük most szemügyre részletesebben a főnévi igenév tulajdonságait. Az már láttuk, hogy a főnévi igenév személyragozható. Ragjai a következők:

ni			*ani*	*eni*
nom	*nem*	*nöm*	*anom*	*enem*
nod	*ned*	*nöd*	*anod*	*ened*
nia	*nie*		*ania*	*enie*
nunk	*nünk*		*anunk*	*enünk*
notok	*netek*	*nötök*	*anotok*	*enetek*
niuk	*niük*		*aniuk*	*eniük*

Ezek a ragok semmilyen más igeraggal nem esnek egybe. (Tehát jól felismerhetők.)

Nézzük meg, hogy változik meg az ige argumentumainak kijelölési módja. Ebből a szempontból három mondatbeli szerepet kell megkülönböztetnünk:

- Az első eset az, amikor a ragozhatatlan főnévi igenév a *fog*, *tud*, *akar* és hasonló típusú (általunk igei segédigének nevezett) ige mellett áll. Ilyenkor a főnévi igenév alanya azonos a segédige alanyával, azaz alanyesetű, és a többi argumentum kijelölése is változatlan. (A segédigének definíció szerint nincs saját argumentuma.)
- A második eset, amikor a főnévi igenév a *kell*, *lehet*, *sikerül* típusú segédige, vagy a *muszáj*, *tilos*, *tanácsos*, *jó* stb. segédigei szerepű mellékneves kifejezés mellett áll, illetve a *van mit tennie*, *nincs mit tennie* típusú kifejezésekben szerepel. Ilyenkor a főnévi igenév ragozható is. Ezekben a konstrukciókban a logikai alany *-nak/-nek* ragot kap. Ha nincs sem személyrag, sem logikai alany, mi általános alanynak értelmezzük. Pl.:

  El kell oda menni.
  Nincs mit enni.

  (Így általában mindenkinek vagy bárkinek.)
- A harmadik eset, amikor a főnévi igenév egy másik ige argumentuma. Ilyenkor ez a másik ige egyértelműen megszabja meg, melyik argumentumához kapcsolódik a főnévi igenév úgy, hogy ez az argumentum lesz az igenév alanya. Tapasztalataink szerint az argumentum a fő ige alanya, tárgya vagy részeshatározója lehet.

  Pista elment dolgozni.
  Pista dolgozni hívta Pétert.
  Pista enni ad Péternek.

A következő kérdés, hogy elválik-e az igekötő a főnévi igenévtől. Tapasztalataink a következők:

- Ha a főnévi igenév egy másik ige argumentuma, akkor az igekötő nem válik el a főnévi igenévtől.
- Ha a főnévi igenév egy *fog*, *akar* vagy *kell*, *lehet* típusú segédigés szerkezet része, akkor az igekötő szeret elválni az igétől, és a segédige rendszerint közéjük ékelődik.

Meg kell oldani a feladatot.

- Amikor a személyragozott vagy ragozatlan főnévi igenév melléknevet tartalmazó szerkezet mellett áll, akkor a mi megítélésünk szerint helytelen, ha elválik az igétől, de a mindennapi életben találkozunk ilyen szerkezetekkel.

? Meg tanácsos oldani a feladatot.

A következő vizsgálandó kérdés, hol van a helyük a mondatban a főnévi igenév argumentumainak. A mi válaszunk az, hogy szétszóródnak. Mi úgy érezzük, hogy abban az esetben, ha a főnévi igenév egy másik ige argumentuma, a mondat szabatosabb, világosabb, ha a főnévi igenév argumentumai körülötte csoportosulnak. Ilyen nyelvtani szabály viszont már azért sem lehet, mert a másodlagos argumentumok is fókuszálhatók, és a fókusz helye a főige előtt van.

Almát ment venni a piacra.

Végül hadd ismételjük meg azt a megállapításunkat, hogy a névszói kifejezésbe (amit máshol részletesen leírtunk) a főnévi igenév sehova se illeszthető be.

Felmerül ezután a kérdés, hogy ha a főnévi igenév nem névszó, tekinthető-e ragozott igének, és a képzők igeragoknak. A válaszunk az, hogy igen. Morfológiai szempontból semmi akadálya. Az alany és állítmány egyeztetése lényegében azonos a ragozott igénél alkalmazottal. Az alapvető különbség ott van, hogy a ragozott ige az esetek majdnem 100%-ában a mondat igei állítmánya lesz, azaz az elemzés kiindulópontja, ezzel szemben a főnévi igenév soha nem áll egyedül a mondatban, hanem mindig van valami nagyobb kifejezés, aminek része lesz. Ez a körülvevő konstrukció igen sokféle lehet, és ennek felismerése és felépítése kissé összetettebb feladat, mint a ragozott ige megtalálása.

Melléknévi igenév

Ellentétben a főnévi igenévvel, a melléknévi igenév lényegében melléknévként viselkedik. Fokozható, névszói jeleket és ragokat kaphat. A mondatban is melléknévi pozíciókat tölthet be, nevezetesen: névszói kifejezés része lehet, vagy a mondat állítmánya lehet. A két pozícióban a melléknévhez hasonlóan kissé másképpen viselkedik.

A melléknévi igeneveket viselkedésük szerint három csoportba sorolhatjuk:

1. csoport: *ó* *ő* *vó* *vő* képzővel képzett igenevek,

2. csoport: *t tt* *ott* *ett ött* *vott vett vött* *atlan* *etlen* *vatlan vetlen* képzővel képzett igenevek,

3. csoport: *andó* *endő* *vandó* *vendő* *ható* *hető* *hatatlan* *hetetlen* képzővel képzett igenevek.

Minden melléknévi igenévnek közös tulajdonsága, hogy a névszói kifejezésben két lehetséges helyet foglalhat el: az egyik lehetséges hely a tulajdonságjelzők között van, a másik lehetséges hely a névszói kifejezés elején a névelő után van.

a három Péter által megoldott nehéz feladat
a Péter által megoldott három nehéz feladat
a három megoldatlan feladat
a megoldatlan három feladat

A névszói kifejezés elején álló pozícióba közönséges melléknév ritkán kerül. Ez is azt mutatja, hogy az *-atlan/-etlen* képző nem melléknevet, hanem melléknévi igenevet hoz létre.

A melléknévi igenévhez opcionálisan az ige argumentumai és szabad határozók kapcsolódhatnak. A határozók kijelölése ugyanazokkal a ragokkal és névutókkal történik, mint az ige esetén, az alany és tárgy kijelölése azonban igefajtánként különbözik.

A melléknévi igenevet cselekvőnek nevezzük, ha az a szó, amelynek a jelzője, az ige alanya. Szenvedőnek akkor nevezzük, ha az a szó, amelynek az igenév a jelzője, az ige tárgya. Tárgyatlan igéből nem lehet szenvedő melléknévi igenevet képezni. A szenvedő melléknévi igenév alanyát az *által* névutó szokta kijelölni.

Az első csoportba tartozó képzők cselekvő melléknévi igenevet hoznak létre.

az elalvó gyermek
a motorkerékpáron csomagot hozó postás

A második csoportba tartozó képzők tárgyatlan igéből cselekvő, tárgyas igéből szenvedő melléknévi igenevet hoznak létre.

az elaludt gyermek
a postás által motorkerékpáron hozott csomag

A harmadik csoportba tartozó képzők tárgyas igéből szenvedő melléknévi igenevet csinálnak, tárgyatlanra nem alkalmazhatók.

---
a postás által motorkerékpáron hozható csomag

A következő kérdés, hogy hol helyezkednek el az igenév argumentumai. A névszói kifejezésben az alany vagy a tárgy a jelzővel ellátott szó lesz, a többi argumentum a melléknévi igenév előtt áll.

Más a helyzet, amikor a melléknévi igenév a mondat állítmánya. Azokra a mondatokra gondolunk, amelyben két alanyesetű névszói kifejezés van, és ezek közül az egyik olyan melléknévi igenév, amelynek nincs határozott névelője. Ilyenkor a másik alanyesetű névszói kifejezés lesz az igének az az argumentuma, ami a jelzőt kapó szó szokott lenni. Azaz a cselekvő melléknévi igenév esetén az alany, szenvedő melléknévi igenév esetén a tárgy. A többi argumentum az első és második csoportba eső igenév esetén az igenév előtt áll. A harmadik csoportba eső igenevek esetén viszont az argumentumok szétszóródhatnak a mondatban. Ebben az esetben az igekötő is leválhat az igenévről.

Ez a postás csomagot hozó.
Ez a csomag postás által hozott.
Motorkerékpáron csak ez a csomag hozható.
Motorkerékpáron csak ez a csomag hozható el.

Kommentárok:

- A *-ható/-hető* és *-andó/-endő* képzős névszói állítmányú mondatok nagyon hasonlítanak azokra a mondatokra, amelyeknek ragozott ige az állítmánya.
- A ragos vonzatú mellékneveknek is elmehet a vonzata a melléknév elől, ha a melléknév állítmányi szerepű.
  A fiára az anya nagyon büszke.
- A köztudatban erősen él az a tévhit, hogy a határozói igenévvel alkotott szenvedő mondat magyartalan, ezért gyakran átalakítják olyan formára, amikor az állítmány múlt idejű (befejezett) melléknévi igenév. Az ilyen mondatok nagy része viszont a nyelvészek szerint magyartalan. Az ilyen szenvedő mondatok nagyon gyakoriak a matematikában, fizikában és műszaki szövegekben.
  Adott egy háromszög két oldala és a közöttük lévő súlyvonal hossza.

- Fölmerül a kérdés, hogy a *–ható/–hető* képzőt miért tekintjük önálló képzőnek, és miért nem a *–hat/–het* és az *–ó/–ő* képző egymásutánja. A válasz az, hogy a *–hat/–het* nem egészen normális ige-ige képző (ellentétben a *–at/–tet; –gat/–get* stb. képzőkkel). Nem alkalmazható rá a *–t/–tt* képző, a *–ni* és a *–va/–ve* képzők, és ennek következtében számos igei konstrukció. Továbbá az *–ó/–ő*-vel képzett igenév nem cselekvő hanem szenvedő, és az *–andó/–endő*-höz hasonlóan viselkedik.

Határozói igenév

Határozói igenevet a *–va/–ve*, *–ván/–vén* képzőkkel képezhetünk.

A határozói igenév két különböző okból kerülhet a mondatba: lehet a szenvedő mondat állítmánya, és lehet mód-, ok- és állapothatározó. (Az utóbbiakat a továbbiakban módféle határozónak nevezzük röviden.)

Az első konstrukció (a szenvedő szerkezet) felépítése egész világos. Az ige határozói igenévi formában áll a mondatban és személyét, számát, idejét, módját a létige megfelelő alakjával adjuk meg. (A létige *van* alakja időnként elmarad.)

Az ól ajtaja kitárva.

Ebben a konstrukcióban a határozói igenév lehet cselekvő és lehet szenvedő. Ha az ige tárgyatlan, az igenév cselekvő, azaz az alanyesetű mondatrész az ige alanya. Ha az ige tárgyas, az igenév szenvedő, azaz az alanyesetű mondatrész az ige tárgya lesz.

A Balaton be van fagyva.
Zöldre van a rácsos kapu festve.

Az igekötő elszakadhat az igenévtől, az ige argumentumai szétszóródnak a mondatban.

A szenvedő mondat azt az állapotot írja le, amiben az alanyesetű mondatrész az ige végbemenetele során van vagy annak következtében került. Mi azzal a kérdéssel nem foglalkozunk, hogy mikor helyes és mikor helytelen a szenvedő szerkezet használata.

A nehezebb eset az, amikor a határozói igenév valamilyen módféle határozó. Nem találkoztunk egyetlen igével, vagy más szóval, sem olyan nyelvi konstrukcióval, amely vonzatként igényelné a módhatározót vagy a határozói igenevet. A módhatározó tipikusan szabad határozó, bármilyen mondatba betehető, ha akarjuk. Ezen belül a módhatározói szerepű határozói igenév is szinte minden mondatba betehető.

A határozói igenév és argumentumai azt az állapotot írják le, amibe a fő állítás valamelyik szereplője *van* vagy *kerül*, az igenévben szereplő ige végbemenetele során. Az a helyzet tehát, hogy a fő állítás egy igei vagy névszói reláció, és ennek egyik argumentuma azonos a határozói igenév egyik argumentumával. Arra nézve azonban semmilyen nyelvtani szabályt nem sikerült találnunk, hogy ebben a viszonyban melyik argumentum melyik argumentumhoz kapcsolódik.

Tapasztalatunk szerint a szóba jövő argumentum alany vagy tárgy lehet, esetleg részes határozó.

Felöltözve ült az ágyon. (alany-alany)
A facsemeték összekötözve hevertek a földön.
(alany-tárgy)
A facsemetéket összekötözve vittük haza.
(tárgy-tárgy)
?*Neki ágyban fekve szolgálták fel a reggelit.*
(részeshatározó-alany)

A határozói szerepű igenévről az igekötő nem szokott leválni, viszont az argumentumok szétszóródnak a mondatban.

Arról sem tudunk semmit se mondani, hogy az igenévben szereplő ige előbb ment-e végbe, mint a fő állítás.

Kulcsait keresve lépett be a szobába.
Kulcsait keresve kinyitotta a zárat.

Az ige főnévi alakja

Az igéből *-ás/-és* képzővel valódi főnevet képezhetünk, ami az ige által leírt folyamatot fejezi ki. Ez a főnév minden igéből képezhető. Sajnos ez időnként alakilag egybeesik egy olyan szóval, ami nem a folyamatot, hanem az eredményt adja meg, vagy éppen a keletkezett tárgyat.

Elgondolkoztam írás közben.
Olvastam Péter írását.
Összetéptem Péter írását.

A továbbiakban a folyamatot kifejező szóval foglalkozunk. Az igekötő csak tagadószó hatására szakad el az igétől.

a törvény be nem tartása

Az argumentumok kijelölésénél az igenevekhez hasonlóan másképpen viselkednek a tárgyas és a tárgyatlan igék. Ha az ige tárgyatlan, az alany az *-ás/-és* végű főnév birtokosa lesz.

Péter délelőtti felszólalása soká tartott.

Ha az ige tárgyas, két különböző megoldás van az alany és tárgy kijelölésére: Az első esetben a főnév birtokosa a tárgy lesz, az alanyt pedig az által névutó jelöli ki (a határozóknál leírt szabályok szerint). A másik esetben a tárgyrag nélküli tárgyat egybeírjuk az *-ás/-és* végű főnévvel, és az alany az így keletkezett szó birtokosa lesz.

A szövetkezet Péter által történt alapítása során sok papírt kellett beszereznem.

Péter szövetkezetalapítása során sok papírt kért tőlem.

Mi a fentiekben csak azt írtuk le, hogy az alanyt vagy a tárgyat a birtokossal lehet megadni. Azt azonban nem mondtuk, hogy a birtokos mindig az alany vagy a tárgy.

A vita utolsó felszólalása alatt mindenki aludt.

Az utóbbi napok szőlőben való kapálásai nagyon kifárasztottak.

Ha egy igének van tárgyas és tárgyatlan vonzatkerete is, a birtokos az egyik esetben az alany, a másik esetben a tárgy.

Péter hallgat (nem beszél) – Péter hallgatása
Péter zenét hallgat – a zene hallgatása
(Péter által), Péter zenehallgatása

Az ige határozói (amelyek szabad határozók és vonzatok lehetnek) általában nem állnak az *-ás/-és* végű főnév előtt. Elhelyezésükre két út van: az egyik a melléknevesítés. A melléknevesítés során az argumentumot beillesztjük a névszói kifejezésbe. Erre két lehetőség van. Ha a határozó névutós és *-i* képzővel melléknévi névutóvá alakítható, így csinálhatunk belőle melléknévi kifejezést. Egyébként egy odaillő töltelék melléknévi igenév elé kerülnek. Ilyen töltelék melléknévi igenév a *való, folyó, szóló, történő, folytatott, tartott, történt, végzett* stb. A töltelék igenév megadja a cselekvés folyamatosságát, befejezettségét, jövőbeniségét, lehetségességét, de egyébként nem befolyásolja az ige jelentését.

a bérekről való tárgyalás
a bérekről folyó tárgyalás
a bérekről szóló tárgyalás
a bérekről történő tárgyalás
a bérekről folytatott tárgyalás
a bérekről tartott tárgyalás
a bérekről történt tárgyalás
a bérekről folytatandó tárgyalás
a bérekről folytatható tárgyalás

Az argumentumok másik lehetséges helye a névszói kifejezésen kívül van, többnyire az *-ás/-és* végű főnév mögött.

Kicsi volt az eltérés a várt értékektől.
Hivatkozás volt benne több szerzőre is.

Nemcsak az *-ás/-és* végű főneveknek vannak ilyen speciális vonzatai, hanem egy sor más főnévnek is.

a Pistával való házassága, a Pistával kötött házassága – a házassága Pistával

a Kínába vezető út – az út Kínába

Egyes igék bizonyos argumentumait nem kell melléknevesíteni, hanem egyszerűen az *-ás/-és* végű főnév elé kerülnek.

a rács pirosra festése
János laposra verése
János gazembernek nevezése
János Pestre érkezése
János Pesten maradása

Implementációs megjegyzések

A fentiek alapján úgy tűnhet esetleg, hogy az igenevekkel kapcsolatos elemzés implementálása borzasztóan komplikált feladat. Erről szó sincs, mindössze két pontba foglalható össze: az egyik a vonzattranszformáció, a másik az átértékelés.

Vonzattranszformáción a következőt értjük: Ha az igenév olyan pozícióban áll, ahol szenvedővé válik, ott a vonzatok lekérdezése során nem az eredeti ige vonzatait kell keresnünk, hanem a vonzatkeretet transzformálnunk kell. A transzformáció szabályai a következők:

- Ha a keretben nincs tárgyrag, a vonzatok nem változnak.
- Ha a keretben van tárgyrag, és az igenév egybe van írva egy esetrag nélküli névszóval, ez lesz a tárgy. A továbbiakban az ige tárgyatlan igeként viselkedik.
- Ha a keretben van tárgyrag, ezt ki kell hagynunk, és helyette fel kell vennünk opcionális vonzatként az alanyt kijelölő által névutót.
- Az *-ás/-és* végű főnéven be kell jelölni, hogy "külső vonzatos" azaz a vonzatait a névszói kifejezésen kívül kell keresni.

Ide tartozik az is, hogy a külső vonzatos főnevek előtt álló töltelék melléknévi igenevet kihagyjuk, és a mögötte álló külső vonzatos főnév vonzatait keressük.

Átértékelésnek azt nevezzük, hogy szenvedő állítmány esetén a látszólagos alanyt átértékeljük tárgynak, az *-ás/-és* végű főnévnél a birtokost átminősítjük alannyá vagy tárggyá, a külső vonzatos főnevek előtt álló melléknévi névutós jelzőket átminősítjük vonzattá. Mindez megtehető a szintaktikai elemzés elvégzése közben, de biztonságosabb egy későbbi fázisban, például a szemantikus elemzés alatt.

Ez a megoldás tapasztalatunk szerint az esetek 90%-ában válik be. Kivételek kezelésére természetesen itt is fel kell készülni.

A HATÁROZOTT ÉS A HATÁROZATLAN NÉVSZÓI KIFEJEZÉS

AZ ELNYELŐDÉS

Egyes más nyelvekhez hasonlóan a magyar nyelv is megkülönbözteti a határozott és a határozatlan névszói kifejezést. Annak, hogy egy névszói kifejezés határozott vagy határozatlan, nyelvtani jelentősége van. Három ilyen dolgot is tudunk mondani:

1. A határozott tárgy mellett tárgyas igeragozást használunk, a határozatlan mellett alanyit. Az úgynevezett mérték értékű tárgy mindig határozatlan.

Pl.: *Olvasom egy kicsit a könyvet.*
Egy könyvet olvasok.

2. A két alanyesetű névszói kifejezést tartalmazó mondatokban a határozott az alany, a határozatlan az állítmány.

Pl.: *A barack még éretlen.*

3. A határozott névszói kifejezés utal valamire, ami korábban elhangzott, a határozatlan névszói kifejezés új információ. Ez a felfogás teljesen összhangban van az előző ponttal, hiszen azt jelenti, hogy egy-két alanyesetű névszói kifejezést tartalmazó mondatban egy ismert dologról valami újat állítunk.

Ezek után csak két kérdés marad:

- Mitől határozott vagy határozatlan egy névszói kifejezés?
- Vajon mind a három esetben ugyanazt értjük-e a határozott illetve határozatlan kifejezésen?

Első közelítésben azt mondhatjuk, hogy határozott névszói kifejezés az, ami előtt határozott névelő áll, határozatlan pedig az, ami előtt nem áll. Ez azonban igen durva közelítés. A tárgyas ragozás tekintetében a teljes szabályrendszer a következő: Tárgyas ragozást használunk, ha a tárgy

1. tulajdonnév;

Pl.: *Látom Pistát.*

2. amikor a névszói kifejezés elején határozott névelő, illetve az előtt mutató névmás áll;

Pl.: *Látom a szomszéd utcában lakó gyerekeket.*
Ezt a rózsát kérem.

3. bármilyen birtokos szerkezet esetén, beleértve a hiányosakat is;

Pl.: *Kitépem egy légynek egy lábát.*
Kitépem egy légy lábát.
Kitépem egy lábát.
Kitépem egy légyét.

4. bizonyos névmások esetén, beleértve a ki nem tett utalószót is. (Más névmások estén viszont alanyi ragozást használunk.)

Pl.: *Megsértetted-e magad?*
Melyiket válasszam?
Azt Szeretném, ha meglátogatnál.

Most nézzük meg, hogy mi a helyzet a hivatkozással. Adunk-e hozzá az összegyűjtött információkhoz, az összegyűjtött objektumokhoz és relációkhoz egy újabb objektumot vagy relációt a névszói kifejezés kapcsán, vagy egy már meglevőt jelöl meg. Technikailag az is fontos kérdés, hogy milyen nehéz megtalálni azt, amit keresünk.

1. A névmások többsége utal valakire vagy valamire, a hivatkozott valami lehet grammatikailag egyértelműen meghatározott, vagy olyan valami, amit a korábban elhangzottakból kell kikeresni. A kérdő névmás azonban nem hivatkozik semmire. Az általános és határozatlan névmás kvantor jellegű, és a szóba jövő dolgok körét kell meghatároznunk.

Pl.: *Láttuk egymást Péterrel.* (meghatározott)
Találkoztam vele. (keresendő)
Ki van ott? (-)
Bármelyiket megoldhatod. (mi közül)
Semelyiket sem látom. (mi közül)

2. A tulajdonnév mindig valami meghatározott dolgot jelöl. Ez a dolog lehet egy új dolog, és lehet egy régi, már ismert dolog.

Pl.: *A Fülöp-szigetek fővárosa Manila.*

Itt Manila lehet egy teljesen újonnan felmerült név, és lehet egy ismert dolog, amiről már sok mindent tudunk. Mindenesetre egy igen konkrét dolgot jelöl, aminek a kiválasztásában semmiféle alternatívánk nincs.

3. Nem ez a helyzet a határozott névelős köznevekkel. Egy határozott névelős köznév vagy egy korábban említett dologra utal, vagy általában beszél az illető dologról.

Pl.: *A rózsa piros volt.* (utalás egy konkrétra)
A rózsa tövises volt. (utalás egy konkrétra)
A rózsa tövises. (általában)

Ha a kifejezés visszautal valamire, elképzelhető, hogy több korábban emlegetett dolog is szóba jöhet, és ezek közül kell választani.

Sajnos, mindeddig semmilyen kritériumot nem találtunk arra, hogy a határozott névszói kifejezés mikor utalás, és mikor általános. Csak tendenciaszerűen igaz, hogy ha a névszói kifejezés elég bonyolult, akkor nem valószínű, hogy általános dolgot jelöl.

4. Ezek után felmerül a kérdés, van-e olyan határozatlan névszói kifejezés, ami visszautaló. A válasz az, hogy van. A számnevet vagy kvantort tartalmazó névszói kifejezés lehet utaló.

Pl.: *Az asztalon egy piros, egy fehér és egy zöld golyó van.*
Felveszem a pirosat. (utaló)
de: *Felveszek két golyót.* (utaló)
Minden golyót felveszek. (utaló)
de: *Nálam is van két golyó.* (nem utaló)

A fentieket a következőkben summázhatjuk: noha a visszautalás és a határozott névelős névszói kifejezés használatában szemmel láthatóan szoros a korreláció (sajnos, erre nézve nincsenek statisztikai mutatóink), az egyes konkrét esetekre mégsem tudunk semmit se mondani.

Kérdés, hogy vajon az alany és az állítmány tekintetében mi a helyzet a határozottsággal és határozatlansággal. Hipotézisünk a következő: ha a mondatban egy határozott és egy határozatlan alanyesetű névszói kifejezés van, akkor a határozott az alany, a határozatlan az állítmány. (Azaz a határozott névszói kifejezéssel megjelölt dolog beletartozik a határozatlan névszói kifejezéssel jelölt dolgok körébe.) Ha két határozott van, akkor a mondat jelentése az, hogy a két megjelölt dolog azonos. Olyan mondat, amelyben két határozatlan alanyesetű névszói kifejezés lenne, nincs. (Ha az állítmány melléknévi igenév, akkor azt igei állítmánynak tekintjük. Lásd: **igenevek**.)

A fenti feltételezés (használva a határozottság tárgyas ragozásnál megadott négy pontos kritériumát) az általunk számítógéppel feldolgozott több ezer mondat mintegy 90%-ában bevált, és a mondat helyes elemzéséhez vezetett. A maradékban azonban elgondolkoztató kivételeket találtunk. (Zárójelben adjuk meg, hogy mi hogy értelmezzük az adott példamondatot: egy ismert dologról mondunk valami újat, vagy két ismert dolog azonosságát állítjuk. Akkor tekintünk egy mondatot azonosságnak, ha két alanyesetű mondatrész szerepe felcserélhető. Pl.:

Kovács János a gyermek apja.
és *Aki a gyermek apja az nem más, mint Kovács János.*

Ezzel szemben:
Kovács János református lelkész.
viszont *Aki református lelkész, az nem feltétlenül azonos Kovács Jánossal.*)

1. A főnév helyett álló névmás, úgy tűnik, mindig határozott névszói kifejezésnek számít, függetlenül attól, hogy tárgyas ragozással jár-e vagy sem.

Pl.: *Mindenki éhes.* (új állítás)
Én magyar vagyok. (új állítás)
Én vagyok a győztes. (azonosság)
Ki az, (aki ...) (azonosság)
Ez minden. (azonosság)

Ezzel szemben a melléknevek és számnevek helyett álló névmások nem számítanak határozott névszói kifejezésnek.

Pl.: *Ez nem annyi.* (új állítás)
A mienk is akkora. (új állítás)

A mutató névmással nyomatékosított névszói kifejezés mindig határozott.

2. A birtokos szerkezetek köréből nézzük meg a következő példákat.

Pl.: *Ez a csont az ősembernek a lábszárcsontja.* (új állítás/?)
Ez a csont egy ősember lábszárcsontja. (új állítás/?)
Ez a csont egy ősembernek a lábszárcsontja. (új állítás/?)
Ez a csont a lábszárcsontja. (azonosság)
Ez az ember a gyermek apja. (azonosság)
Ez az ember a gyermeknek az apja. (azonosság)
Ez az ember egy gyermek apja. (új állítás)
Ez az ember egy gyermeknek az apja. (új állítás)
Ez az ember az apja. (azonosság)
Az anya az Operaház táncosnője. (új állítás)
Az anya az Operaháznak egy táncosnője. (új állítás/?)
Az anya az Operaháznak az a táncosnője, aki ... (azonosság)

Mi a fentiekből a következő következtetést vontuk le:

Egy birtokos szerkezetnek két összetevője van, a birtokos és a birtok. Ezek közül bármelyik jelölhet egy konkrét valamit, és lehet általános valami vagy egy határozatlan valami. A hiányzó birtokos nagyon határozott utalásnak számít. Ha a birtokos szerkezet egy konkrét valaminek egy konkrét valamije, akkor a szerkezet egyértelműen határozott. (Azaz: vagy alany, vagy az azonosság egyik eleme.) Ha a birtokos szerkezet egy határozatlan valami határozatlan valamije (mint ebben a mondatban), akkor a szerkezet egyértelműen határozatlan, és így csak állítmány lehet. Ha viszont a birtokos szerkezetben

keverednek a konkrét, általános és határozatlan elemek, akkor a kifejezés többarcú és így is,

Ez a csont egy ősembernek a lábszárcsontja.

Ez a mondat úgy is magyarázható, hogy a csont azonos egy ősember lábszárcsontjával; és úgy is magyarázható, hogy a csontról azt állítjuk, hogy az ősember lábszárcsontja. Hasonló a helyzet azzal a mondattal, hogy:

Ez a csont az ősembernek a lábszárcsontja.

mert az ősember nem konkrét, hanem általános.

Igen érdekes viszont, hogy általánosságot kifejező határozott névelős egyszerű (nem birtokos) névszói szerkezet soha nem áll állítmányként. Pl.:

A tegnap fogott hal egy bálna volt.

Ha azt mondanánk:

A tegnap fogott hal a bálna volt.

akkor itt a bálna biztosan egy konkrét bálnát jelölne, és a mondat azonosság volna.

3. A számneves és kvantoros kifejezés alanyként is határozottnak számít. Pl.:

Érett-e már a barack?

Csak két barack érett.

A birtokos szerkezetek kapcsán vetődik fel az a kérdés, amellyel a nyelvészeti irodalom oly nagy előszeretettel foglalkozik, és amelyet mi sem kerülhetünk meg, hogy mihez tartozik a birtokos szerkezet névelője, a birtokhoz-e vagy a birtokoshoz. Azaz például

... az ősember lábszárcsontja ...

vajon azt jelenti, hogy

az ősembernek egy lábszárcsontja

vagy azt, hogy

egy ősembernek a lábszárcsontja.

Nekünk erről az a véleményünk, hogy mindkettőt jelentheti, sőt még azt is jelentheti, hogy

az ősembernek a lábszárcsontja .

Mi úgy érezzük, hogy a névelő bizonyos esetben egyértelműen a birtokoshoz tarozik, pl.:

Az anya az Operaház táncosnője.

azaz

Az anya táncosnő az Operában.

Más esetekben a birtokhoz tartozónak ítéljük, pl.:

Kim Ir Szen a népek Nagy Tanítója.

azaz *Kim Ir Szen a Nagy Tanító.*

Végül igen gyakran ide is és oda is tartozik, pl.:

Ő a legjobb barátom legkisebb gyereke.

azaz

Ő a legkisebb gyereke a legjobb barátomnak.

Kérdés, hogy ebben az utóbbi esetben hova lesz a mondatból egy névelő. Véleményünk szerint a névelő elnyelődik, mert a mondatban nem kerülhet egymás mellé két azonos alakú, de különböző nyelvtani szerepű mondatrész, hanem ilyenkor az egyik elnyelődik, és egyetlen mondatrész tölti be mindkét szerepet. Ebben az állításban nyilvánvalóan az is benne van, hogy nem a fenti eset az egyetlen példa az elnyelődésre.

A névelőelnyelődés másik igen gyakori és hasonló mértékben gondot okozó formája a határozóként beágyazott névszói kifejezés névelőjének az elnyelődése. Pl.:

... a hátsó lakásban lakó legkisebb fiú...

azaz

... a legkisebb fiú azok közül, akik a hátsó lakásban laknak ...

Úgy véljük, hogy a határozott névelő itt is mind a két helyre vonatkozik.

A másik tipikus és gondot okozó elnyelődés az alany birtokos elnyelődés. Pl.:

Pista kalapját keresi.

Ennek a mondatnak a szokásos elemzések szerint 2 magyarázata lehetséges:

VALAKI (*Pista kalapját*) *keresi*.

Pista (VALAKI *kalapját*) *keresi*.

Az üres elemnek (amelyet VALAKI-vel jelöltünk) meg kell keresni az antecedensét. Szerintünk a második elemzés téves, hiszen elképzelhetetlen, hogy a VALAKI ne Pista legyen, azaz a birtokjeles névszói kifejezés előtt álló alanyesetű névszói kifejezés ne legyen birtokos; de ezen kívül még az alany szerepét is betöltheti.

Hasonló dolgok határozókkal is előfordulnak, pl.:

A téren játszadozó gyerekekkel találkoztam.

Ennek a mondatnak az alapján válaszolni tudunk arra a kérdésre is, hogy:

Hol találkoztam a gyerekekkel?

és arra is, hogy:

Hol játszottak a gyerekek?

Elképzelhetetlen egy olyan mondat, hogy:

**A téren a téren játszó gyerekekkel találkoztam.*

viszont az a mondat, hogy

A téren a sakk-körben játszó gyerekekkel találkoztam.

teljesen korrekt.

Összefoglalás

A magyar mondatok szintaktikai és szemantikai elemzése során többször áll elő olyan helyzet, amikor egy névszói kifejezés aszerint tölt vagy nem tölt be valamilyen szerepet, hogy határozott-e vagy határozatlan. Viszont a határozottság kritériuma az egyes szerepeknél kicsit más és más.

Ha a mondatban egy tárgyas ragozású igei kifejezés van (ami azt jelenti, hogy vagy az ige van tárgyasan ragozva, vagy a segédige), akkor az ige tárgyi vonzatának a szerepét csak a fentebb felsorolt négy pont valamelyikének megfelelő határozott névszói kifejezés töltheti be. Az ilyen ige mellett talált határozatlan tárgyragos kifejezés az esetek 90 százalékában úgynevezett mértékértékű tárgy. (Igen ritkán az is előfordul, hogy valamilyen igenév vonzata.)

Tárgyatlan ragozású ige mellett álló határozatlan tárgyragos kifejezésről nem tudjuk pusztán külső jegyek alapján megállapítani, hogy valóban tárgya-e az igének vagy mértéket meghatározó szabad határozó.

Ha egy mondatban nincs ragozott ige (de létige lehet), viszont két alanyesetű névszói kifejezés van, akkor ez a mondat minden bizonnyal egy nominális mondat. Meg kell vizsgálnunk a névszói kifejezéseket. Ezek közül mi azt a névszói kifejezést tekintjük határozottnak, amelyik tulajdonnév, főnévi névmás, számneves kifejezés, vagy az elején határozott névelő áll. A legutolsó esetben időnként tévedünk a minősítésben, vagy legalábbis vitatható a döntés, mert a határozott névelő nem biztos, hogy a legutolsó helyen álló főnévhez tartozik, hanem lehet, hogy csak a névszói kifejezés elején álló birtokoshoz, vagy határozóhoz.

A mi elemzési stratégiánk olyan, hogy ahol elnyelődést tételezünk fel (például az összetett névszói kifejezés élén álló határozott névelő esetén igen gyakran), ott a kettős szerepű szót vagy kifejezést mindkét lehetséges helyre odakötjük.

IGEI ÉS NÉVSZÓI RELÁCIÓK

A természetes nyelv olyan bonyolult, mint maga az élet. Ahhoz, hogy a természetes nyelvű szövegek jelentését számítógépen ábrázolni tudjuk, olyan modellre van szükségünk, amely nagymértékben egyszerűsít. Ez a modell természetesen csak azokat vonásokat emeli ki, melyekre szükség van az adott cél érdekében.

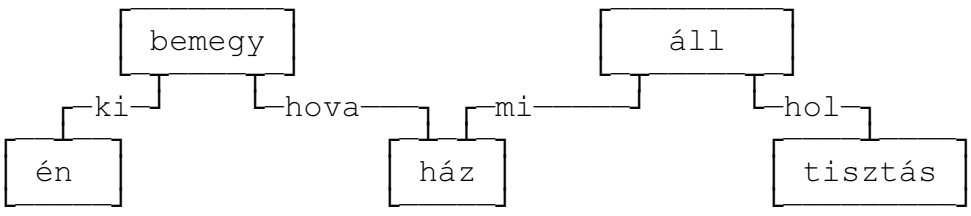
A következőkben mi egy szemantikai modellt vázolunk, de nem konkrét alkalmazási céllal, hanem azért, hogy rávilágítsunk arra, hogy miért olyan a nyelvtani modellünk, amilyen. Ez a modell eléggé általános szinten mozog, de nem áll messze az irodalomból ismert és gyakorlatban létező olyan modellektől, mint például a szemantika logikai kifejezésekkel, például a Horn-klózokkal történő leírása, vagy a fogalmi hálókban (conceptual network) vagy keretekben (frame) való ábrázolás stb.

Ebben a modellben a mondat jelentése igei és névszói relációkból áll. Egy mondatban rendszerint több reláció is van, de ezek között van egy, ami a mondat gerincét képezi, és vannak, amelyek ehhez kapcsolódnak, és ezekhez még továbbiak is kapcsolódhatnak. (Háromnál nagyobb mélység azonban ritkán fordul elő, mert az emberek nem képesek kezelni túl komplikált struktúrákat.)

A relációnak van neve, és vannak argumentumai. A reláció valamilyen kapcsolatot fejez ki az argumentumok között. Ebben az írásban a relációt egy dobozzal ábrázoljuk, amelyre rá van írva a neve. A doboz aljából nyilak indulnak ki, amelyek az argumentumokat tartalmazó dobozok tetejéhez csatlakoznak. Az argumentumoknak is van elnevezése, ami rá van írva a nyílra. Egy adott reláció egy adott argumentumáról előírható, hogy bizonyos szemantikai tulajdonságokkal rendelkezzen.

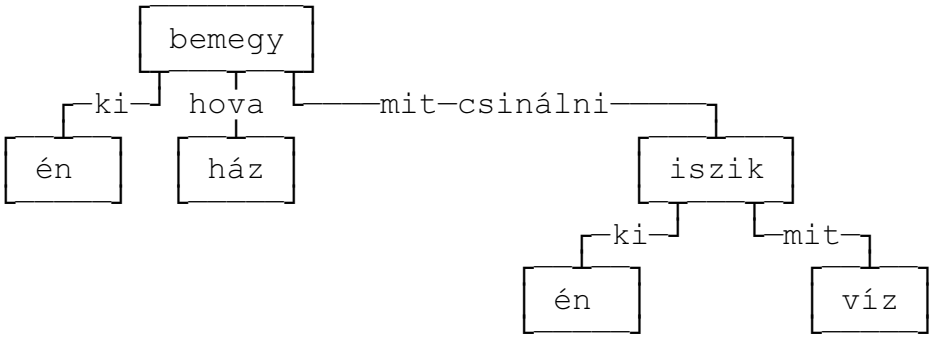
A relációk kétféle módon kapcsolódnak egymáshoz: az első esetben a két relációnak közös argumentuma van. Pl.:

Bementem a házba, ami a tisztáson állt.



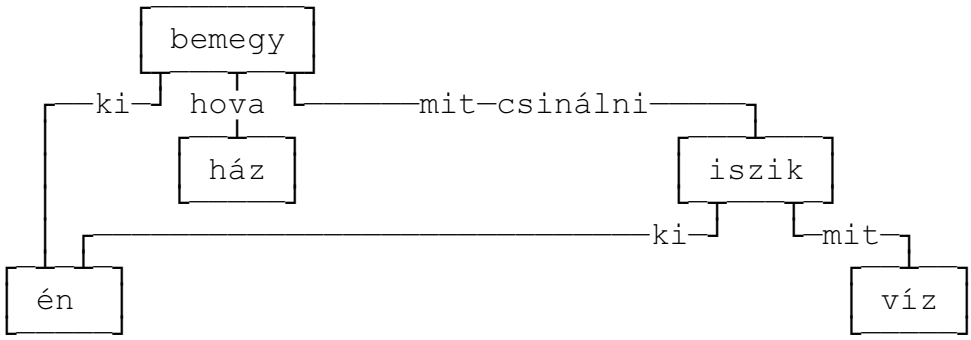
A másik esetben az egyik reláció a másik argumentuma. Pl.:

Bementem a házba, hogy vizet igyak.



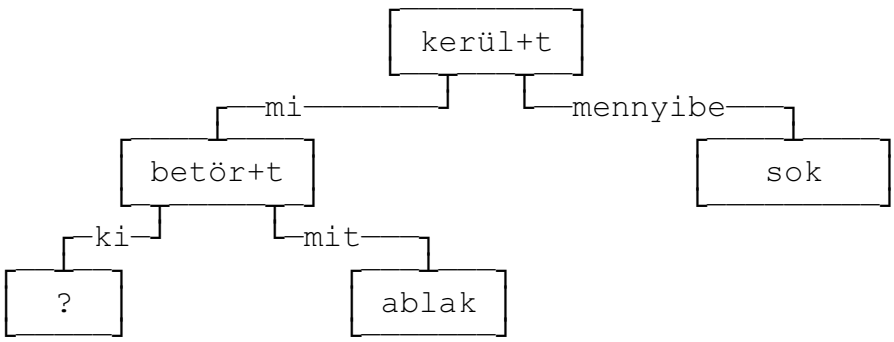
Előfordul természetesen többszörös kötődés is két reláció között. (Kérdés persze, hogy mit tud az adott modell kezelni. Ebbe az ábrázolásba, ami leginkább a fogalmi gráfokra hasonlít, ez is bőven belefér.) Pl.:

Bementem a házba vizet inni.



Minden ige egy olyan reláció, amelynek rá jellemző argumentumai vannak. Ha van ki, mi szóval jelölt argumentuma, ezt alanynak, ha van mit szóval jelölt argumentuma, azt tárgynak, a többi argumentumot határozónak nevezzük, kivéve a *mit csinálni*-val jelöltet, aminek nincs jó neve. Az, hogy egy argumentum alany, vagy tárgy, vagy határozó, meghatározza azt, hogy milyen grammatikai pozícióban, hogyan találjuk meg. Pl.:

Az ablak betörése sokba került.



De ugyanezzel az ábrával írnánk le azt is, hogy:

Sokba került, hogy betörte az ablakot.

vagy *Sokba került betörni az ablakot.*

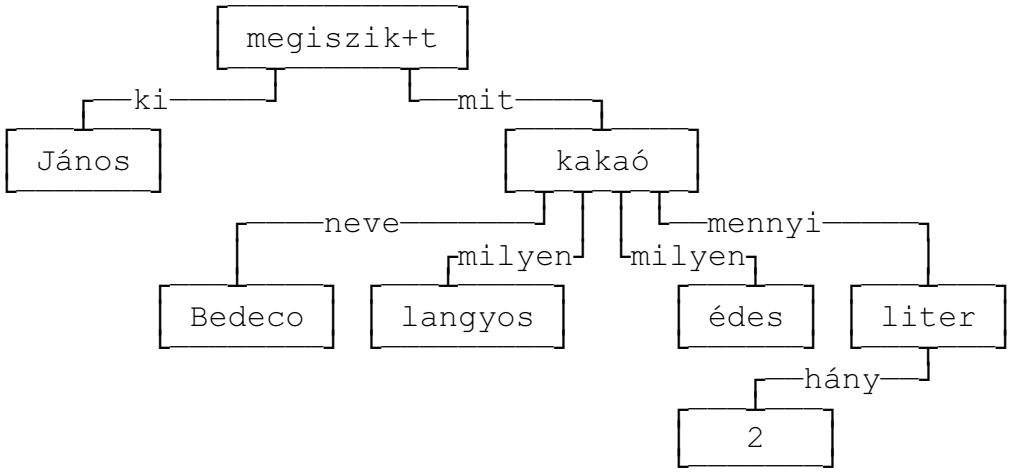
Ez az ábrázolás különböző formájú mondatokat hoz azonos alakra, és ezzel megteremti az egyszerű, egységes kezelés lehetőségét, viszont valószínű, hogy bizonyos közeli jelentéseket össze is mos. Mindig az adott cél dönti el, mit lehet összemosni és mit nem.

Látható, hogy a betörés szó nyelvtani birtokosát a *mit* szóval jelöltük meg. Mi itt is és a továbbiakban majdnem mindig, amikor alanyról, tárgyról, határozóról beszélünk, ilyen "logikai alany, logikai tárgy" stb. értelemben használjuk. A nyelvtani értelemben vett alanyt "nyelvtani alanynak" vagy "alanyesetű mondatrész"-nek nevezzük.

A határozókat nem osztjuk további osztályokra, azaz az argumentumok elnevezése nem egy rögzített előre definiált halmaz, hanem csak elnevezések, hogy beszélni lehessen róluk egy adott igével kapcsolatban. Nekünk nem tűnik célszerűnek a vonzatok osztályokba sorolása. A *kerül* szónak van egy mennyibe szóval jelölt argumentuma, a *csökken* szónak van egy mennyivel szóval jelölt argumentuma. Teljesen felesleges az elsőt "képes helyhatározó"-nak, a másodikat esetleg "képes eszközhatározó"-nak nevezni. Az a lényeg, hogy mindkét igének van egy olyan argumentuma, ami egy szám vagy számokból és mértékegységekből álló kifejezés. Az első igénél ez *-ba/-be* raggal, a másiknál *-val/-vel* raggal van kijelölve. (Ha a került mellett nem (pénz) mennyiség, hanem más névszói kifejezés áll, pl.: Az ablak betörése egy vacsorába került, akkor az argumentumot mibe szóval jelölnénk és egy másik homályosabb értelmű, így önállóan nem is kezelhető relációnak tekintenénk.)

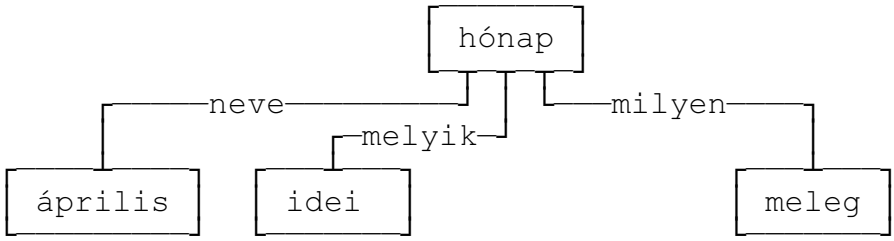
Nemcsak az igéket, hanem a főneveket is relációknak tekintjük. Egy főnévnek a következő argumentumai lehetnek: neve, különböző tulajdonságai, mennyisége, birtokosa. Pl.:

János megivott 2 liter langyos, édes Bedeco kakaót.



Természetesen nemcsak a névszói szerkezeteket ábrázoljuk így, hanem a névszói mondatokat is. Pl.:

Meleg az idei április hónap.



Nemcsak ige és nem csak főnév képezhet relációt, de a továbbiak szempontjából számunkra ezek lesznek a fontosak.

Az igei és névszói relációk megjelenési formái a mondatban

A legfelső szintű igei reláció

Egy mondatban rendszerint egyszerre több reláció is van. Ezek között van egy, ami a mondat nyelvtani gerincét alkotja. Ez egyaránt lehet igei és névszói. Ennek az elsődleges relációnak más a megjelenési formája, mint az alatta levő többinek.

Az elsődleges igei reláció legtipikusabb esete, amikor a mondatnak ragozott ige az állítmánya. Igének egy cselekvést vagy történést kifejező szót nevezünk, de nem nevezzük igének a segédigéket és a létigéket.

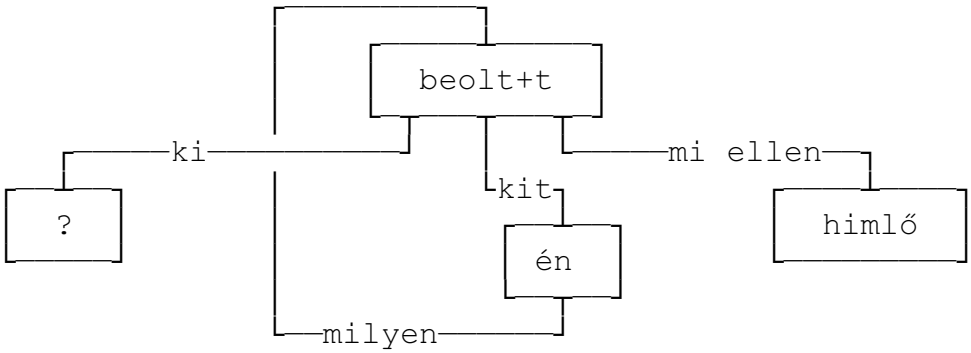
A segédigének olyan igei viselkedésű szavakat nevezünk, amelyek mellett kötelezően egy igenév áll, de semmilyen más vonzatuk nincs. (Az igenév vonzatai alkotják a mondatot.) A segédigéből és igenévből álló szerkezetben az igenév adja meg a relációt, és a segédige csak ezt színezi, módosítja. A *Tessék bemenni a fülkébe!* mondat lényegében a *Menjen be a fülkébe!* mondat udvarias formája, és a bemegy reláció felszólító módban. A *Be kell mennie a fülkébe.* is a bemegy reláció egy olyan "módban", amire nincs megfelelő elnevezésünk.

(A fentiekből látható, hogy a reláció pusztán a dolgok közötti kapcsolatot fejezi ki, de természetesen szemantikailag sem mindegy, hogy ez a kapcsolat fennállt, de most már nem áll, vagy most is fenn áll, vagy most még nem áll fenn, de szükségszerűen fenn fog állni, vagy csak a lehetőség van meg arra, hogy fennálljon a jövőben, stb. Az alkalmazás dönti el, hogy a fenti "mód" milyen értékeket vehet fel, ezeket milyen végződések, segédigék, módosítószók alapján ismerjük fel. A fenti példákban többször találkozhattunk egy a relációnévhez kapcsolt +*t*-vel. Ez azt jelezte, hogy az adott reláció múlt idejű, befejezett változatáról van szó.)

Létigének a *van*, *lesz* és *megvan* szavakat nevezzük, amikor azok nyelvtani segédszóként állnak a mondatban. Ha a mondatban a *van valahol*, *van valamiből* ige szerepel, akkor ez a *van* nem létige, hanem közönséges ige, melynek jelentése közelítőleg azonos az *áll valahol*, *áll valamiből* igékkel. A *van* szó akkor létigei szerepű, ha azért van a mondatban, hogy annak idejét/módját esetleg alanyát vagy tárgyát is megmutassa.

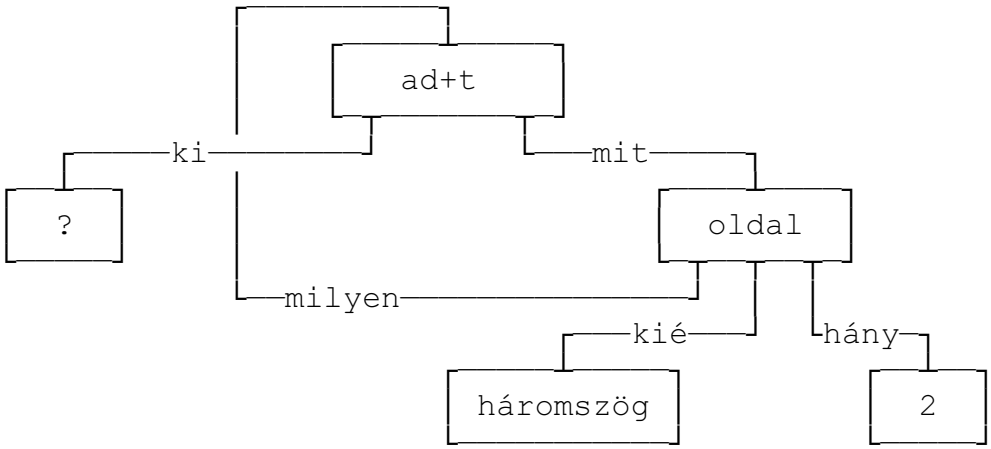
Ilyen eset az úgynevezett szenvedő mondat. Itt az ige határozói igenévi formában áll, idejét, módját, és alanyának vagy tárgyának a számát és személyét a létige adja meg. Pl.:

Be vagyok oltva himlő ellen.



A szenvedő mondat azt fejezi ki, hogy valami végbement és ennek következtében valami (az alany vagy az állítmány) ennek megfelelő állapotba került. Ez a jelentés lényegében azonos a befejezett melléknévi igenév jelentésével, és így nem csoda, hogy a szenvedő mondatot és a múlt idejű melléknévi igenévi állítmányú mondatot egyformán ábrázoljuk. Pl.:

Adva van egy háromszög két oldala.
Adott egy háromszög két oldala.



A legfelső szintű igei reláció utolsó esete az, amikor az állítmány egy melléknévi igenév. (Nem értendők ide azok a mondatok, amelyekben az állítmány egy olyan névszói kifejezés, aminek az utolsó eleme melléknévi igenév, csak azok, amelyek egyetlen melléknévi igenévből állnak.) Ilyenkor egy kölcsönös kapcsolatról van szó. Egyrészt az igenév a nyelvtani alany jelzője, másrészt az alanyesetű mondatrész az ige logikai alanya vagy logikai tárgya. (Aszerint, hogy a melléknévi igenév cselekvő vagy szenvedő.)

Általában is igaz, hogy minden igenév egy ilyen oda-vissza kapcsolatban áll az ige valamelyik argumentumával. Erről szó a következő alfejezet.

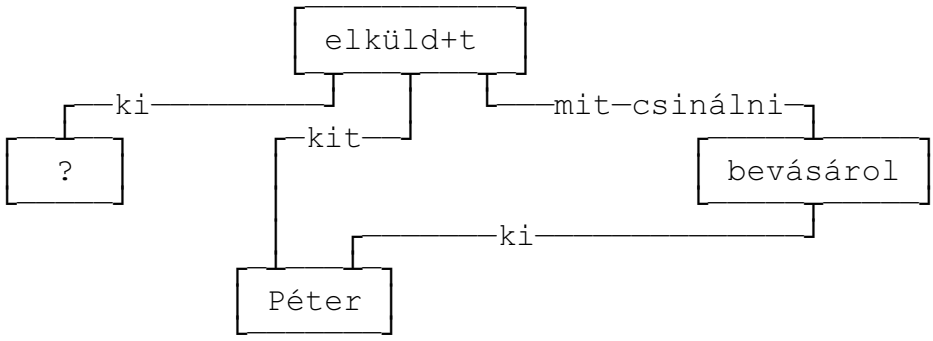
Alsóbb szintű igei relációk

A mondat grammatikailag alsóbb szintjén természetesen nincsen ragozott igei állítmány, sem segédigés kifejezés, sem szenvedő állítmányi szerkezet; van viszont főnévi igenév, melléknévi igenév, határozói igenév, és *-ás/-és* képzővel alkotott folyamatot kifejező főnév.

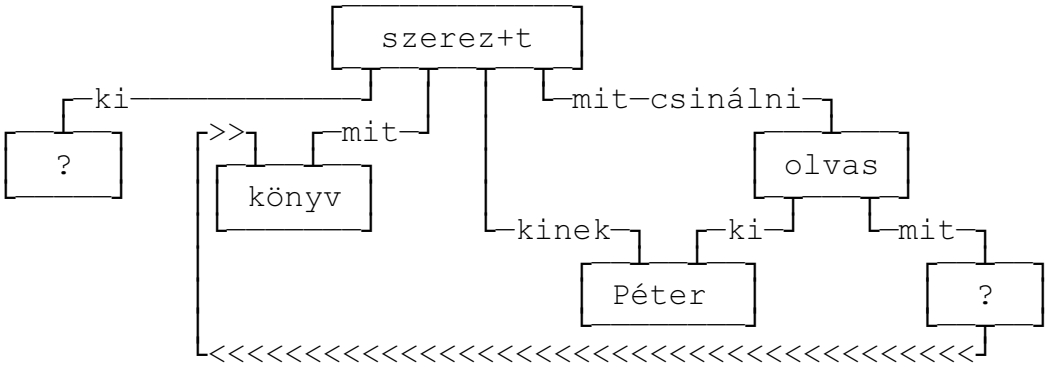
Az ige ezekben az alakjaiban is megőrzi azt a szemantikai tulajdonságát, hogy alanya, tárgya, határozói lehetnek, noha ezeket az argumentumait más grammatikai eszközökkel jelöli ki, mint az állítmányi esetben. Azt, hogy a kijelölés hogyan történik, az **igenevek** című fejezetben részletesen tárgyaljuk. Másrészt viszont az igenévvel adott reláció nemcsak úgy kapcsolódik a felsőbb szintű relációhoz, hogy annak egyik argumentuma lesz, hanem úgy is, hogy a felsőbb szintű reláció egyik alkotórésze az alatta levő igei relációnak is argumentuma lesz.

Főnévi igenév nem állhat akárhol a mondatban. Csak úgy kerülhet be a mondatba, ha egy másik igének a vonzata, vagy egy melléknévnek a vonzata. Ha a főnévi igenév egy másik ige vonzata, a felette álló ige szabja meg, hogy melyik argumentuma (alanya, tárgya, részeshatározója) lesz az általa vonzott másik ige alanya. (Mint egy korábbi példában láttuk a *bemegy+valamit csinálni* igénél, ugyanaz csinál valamit, mint aki bemegy. Más igéknél más a helyzet.) Pl.:

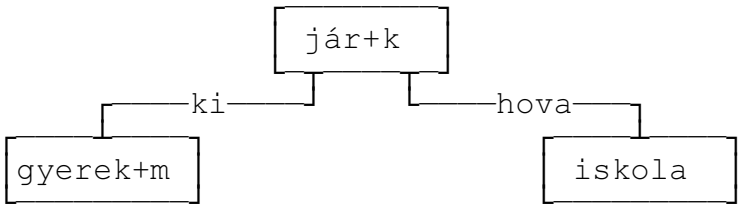
Elküldte Pétert bevásárolni.



Szerzett Péternek egy könyvet olvasni.


szerez+t
ki
?
>>
mit
könyv
mit—csinálni
olvas
kinek
ki
mit
Péter
?
<<

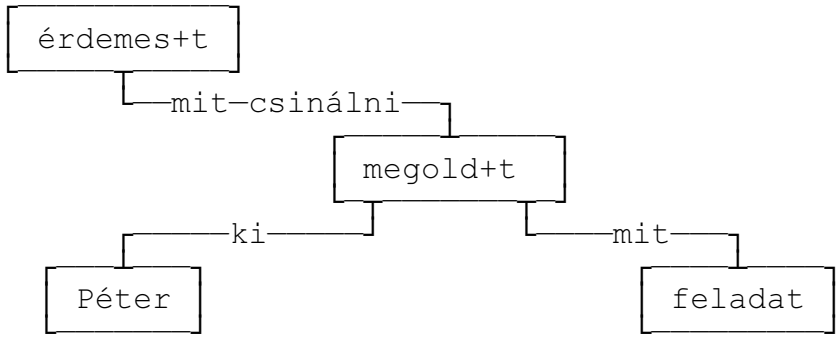
A gyerekeknek iskolába kell járniuk.
A gyerekeknek kötelező iskolába járni.
A gyerekeknek muszáj iskolába járniuk.



(Itt a -k-val jelöltük, hogy a reláció "kötelező módban" van; azaz például: nem biztos, hogy fennáll, de valószínűleg fenn fog állni.)

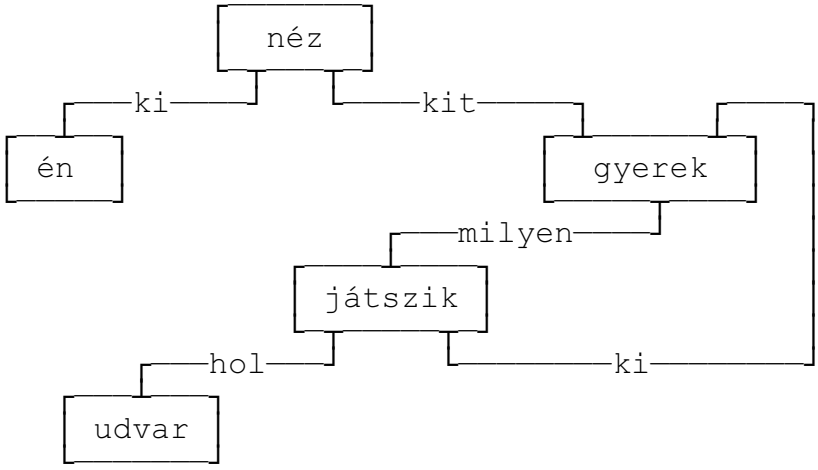
Persze nem minden melléknevet lehet ilyenformán móddá alakítani. Ilyenkor a melléknév argumentuma lesz az igeneves kifejezés. Pl.:

Péternek érdemes volt megoldania a feladatot.

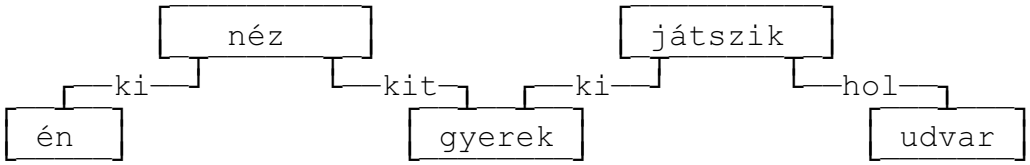


A melléknévi igenév a legfelső szinten állítmányként szerepel a mondatban, az alsóbb szinteken viszont a névszói kifejezésben fordul elő. A melléknévi igenév egyrészt a névszói kifejezésben szereplő főnév jelzője, másrészt ez a főnév az ige egyik argumentuma, mégpedig megmondható, hogy mikor az alanya, mikor a tárgya. Ezt pillanatnyilag egy oda-vissza kapcsolattal ábrázoljuk. Pl.:

Nézem az udvaron játszó gyereket.

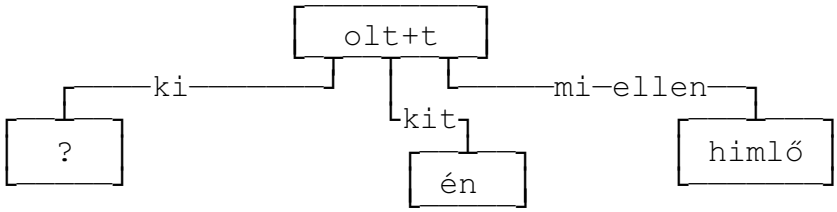


Ez az ábrázolásmód még nagyon visszatükrözi a mondat eredeti formáját. Nem biztos, hogy ez a legcélszerűbb, mert így más ábrához vezetne az a mondat, hogy: *Nézem azt a gyereket, aki az udvaron játszik.*

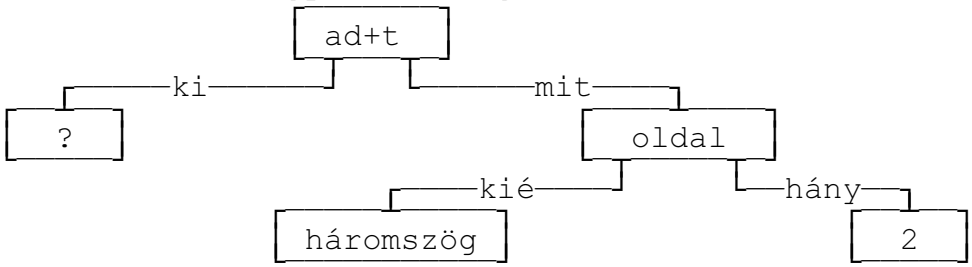


Miután úgy véljük, hogy a két mondat jelentése ugyanaz, célszerűnek látszik a két mondatot azonos alakra hozni. Ez két úton tehető meg, vagy bizonyos kapcsolatokat kihagyunk, vagy bizonyos kapcsolatokat még hozzáadunk az ábrázoláshoz. Az utóbbi út nem látszik célszerűnek, hiszen akkor például abban a mondatban, hogy: *Péter könyvet olvas.*, azt a relációt is be kellene jelölni, hogy a könyv egy olvasott könyv és Péter az olvasó. Mivel ez az ábrából amúgy is leolvasható, a kihagyásos utat választjuk, és a jelzői milyen kapcsolatot elhagyjuk az ábrából. Ezt nem csak az alsóbb szinteken, hanem a legfelső szinten is megtesszük. Pl.:

Be vagyok oltva himlő ellen.



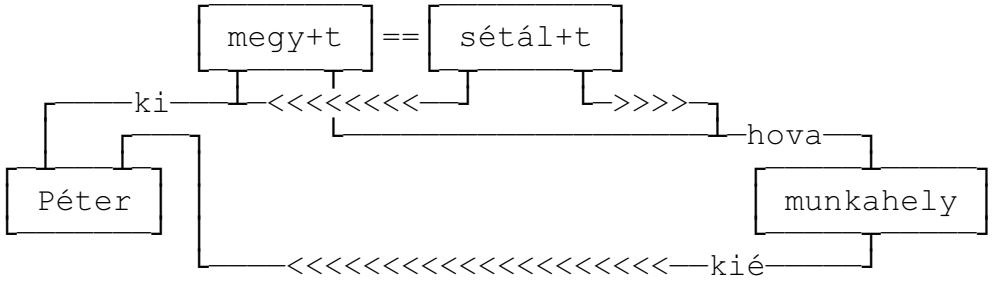
Adott egy háromszög két oldala.



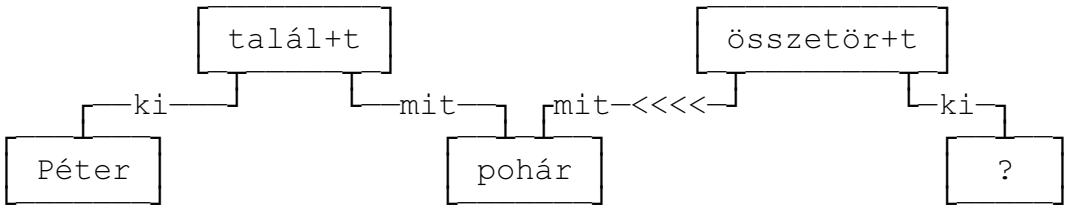
(Vesd össze a mondatok korábbi ábrázolásával!)

A határozói igenév a mondatban mód-, ok-, cél-, állapothatározóként szokott előfordulni. Sajnos ezen határozók jelentése és egymástól való megkülönböztetése nem annyira kemény, hogy számítástechnikailag kezelni tudjuk. Mindössze annyit tudunk mondani, hogy a mondat nyelvtanilag fő állítása mellet, van egy másik igei reláció is, ami az első reláció körülményeit írja le olyan módon, hogy a két reláció egy vagy több argumentuma azonos. Pl.:

Péter sétálva ment a munkahelyére.



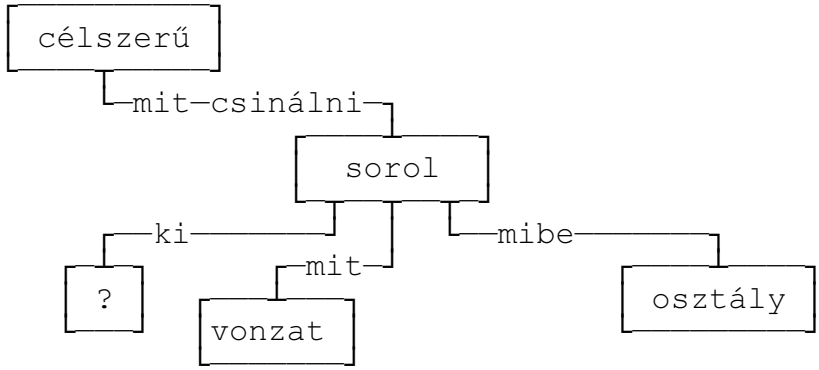
Péter összetörve találta a poharat.



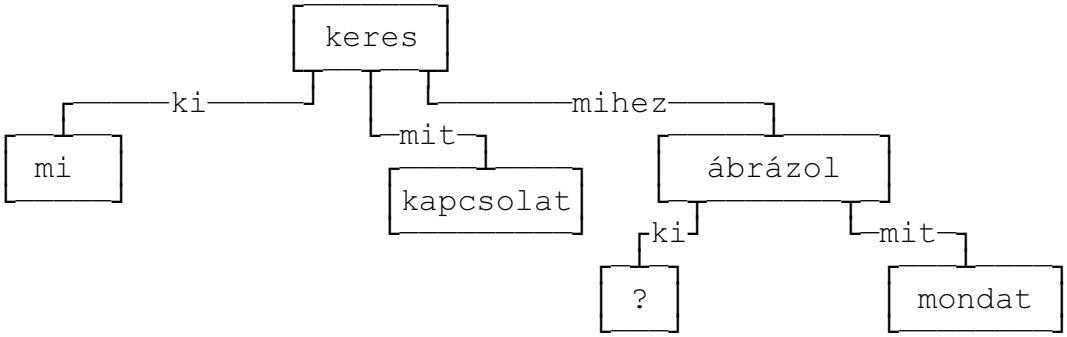
Sajnos pusztán nyelvtani alapon általában semmit nem tudunk mondani sem arról, hogy a két reláció mely argumentumait kell összekapcsolni, sem arról, hogy a határozói igenévvel megadott esemény előbb következett-e be, mint a másik, vagy vele egyidejűleg folyik, vagy netán ugyanazt az eseményt írja le más szempontból.

Az utolsó olyan szófféleség, amelyet egy igei reláció kiindulópontjának tekintünk az olyan igéből képzett -ás/-és végű főnév, ami az ige lefolyását fejezi ki. Ilyen főnév bármilyen főnévi szerepet betölthet a mondatban, lehet alany, tárgy, határozó. Az ige argumentumait viszont többnyire egészen másképpen találjuk meg, mint a ragozott ige esetén. Pl.:

Célszerű a vonzatok osztályokba sorolása.



Kapcsolatokat keresünk a mondat ábrázolásához.



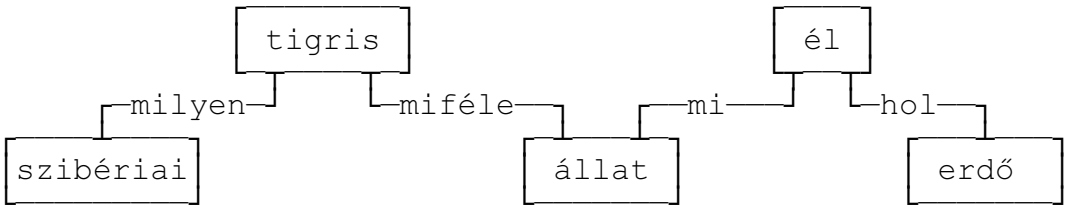
A legfelső szintű névszói reláció

Mi az elsődleges névszói relációnak három esetével találkoztunk. Az első esetben a mondatban két határozott alanyesetű névszói kifejezés van. A másik fajta mondatban egy határozott és egy határozatlan névszói kifejezés áll. A harmadik eset a *valakinek van valamije* formájú mondat.

Határozottnak akkor nevezzük a névszói kifejezést, ha tulajdonnév, vagy ha határozott névelője van, vagy bizonyos névmás. (Az *az, ez, én, te, aki, ami, magam* határozott. A *valaki, bárki, ilyesmi* határozatlan. Mint a fentiekből is látható, nincs egyértelmű kapcsolat a tárgyas ragozás és a határozottság jelen értelmezése között. Azt, hogy hányféle értelemben lehet a névszói kifejezés határozottságáról vagy határozatlanságáról beszélni, külön fejezetben tárgyaljuk.)

Abban az esetben, amikor a mondatban egy határozott és egy határozatlan névszói kifejezés áll, a mondatot úgy értelmezzük, hogy a határozott névszói kifejezés egy halmaznak egy vagy valahány elemét jelöli ki, és ezekről azt állítja, hogy beletartoznak a másik névszói kifejezéssel jelölt halmazba. (Ezt a miféle szóval jelöljük.) Pl.:

A szibériai tigris erdőben élő állat.



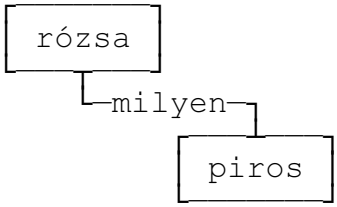
Tudjuk, hogy önkényes dolog, hogy az alanyhoz kötjük az állítmányt, és nem fordítva, de reméljük, mindjárt kiderül, hogy miért ezt tesszük.

Ha a határozatlan névszói kifejezés egy melléknév, akkor tulajdonképpen azt állítjuk, hogy az adott valami rendelkezik az adott tulajdonsággal. Pl.:

Ez a rózsa piros.

Ez szerintünk ugyanazt jelenti, mint:

Ez egy piros rózsa.



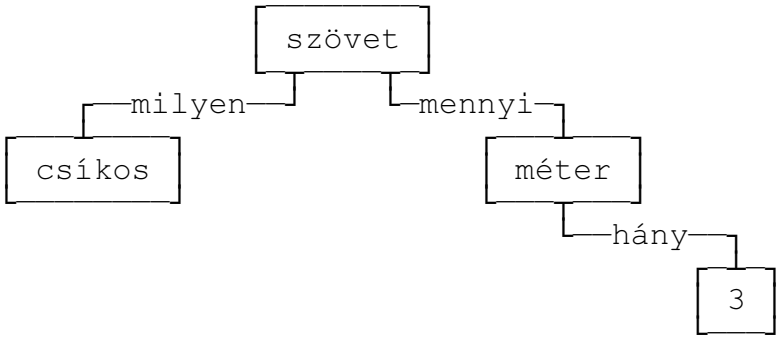
Ha a határozatlan névszói kifejezés egy szám, vagy egy mértékegységes kifejezés, akkor az az illető dolog mennyiségét adja meg. Pl.:

A csíkos szövet 3 méter.

Ez szerintünk ugyannyit jelent, mint a

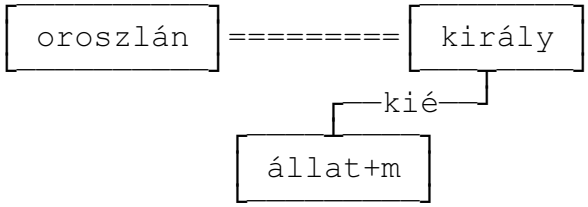
3 méter csíkos szövet

kifejezés.

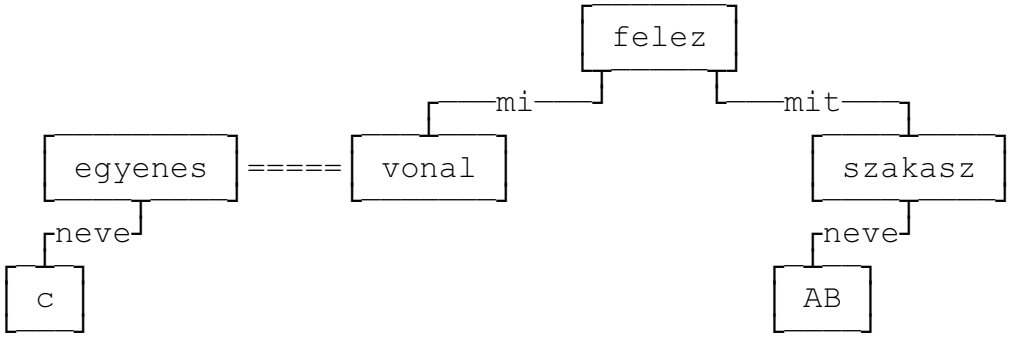


Kicsit más a helyzet akkor, amikor a mondatban két határozott alanyesetű névszói kifejezés szerepel. Ezzel kapcsolatban több felfogás létezik. Mi azt az értelmezést fogadjuk el, hogy ilyenkor azonosság áll fenn a két névszói kifejezéssel kijelölt objektum vagy objektumok között. Az a kérdés tehát, hogy hogyan ábrázoljuk a két objektum azonosságát. Több elképzelés lehetséges: az egyik az, hogy ezt a kapcsolatot egy speciális oda-vissza érvényes nyíllal jelöljük, a másik lehetséges megoldás mindkét reláció egy-egy argumentumával kölcsönösen egymásra mutat, vagy csak az egyik mutat másikra. Hogy melyik megoldásnak mik az előnyei és hátrányai, azt a lehetséges implementációkról szóló részben még tárgyaljuk, mi grafikában a két doboz oldalát fogjuk kettős vonallal összekötni. Pl.:

Az oroszlán az állatok királya.



Legyen a c egyenes az AB szakasz felező vonala.



Érdekes megjegyezni, hogy noha több ezer mondatot megvizsgáltunk számítógéppel, olyan mondatot eddig nem találtunk, amiben két határozatlan alanyesetű névszói kifejezés lett volna, leszámítva azokat, ahol az egyik egy melléknévi igenév volt. Ezeket viszont az igei relációk közé soroltuk.

A *valakinek van valamije* alakú mondatok egy birtokviszonyt adnak meg, jelentésük annyi, mint egy alsóbb szintű birtokos kapcsolatnak. Pl.:

Péternek szakadt volt a kabátja

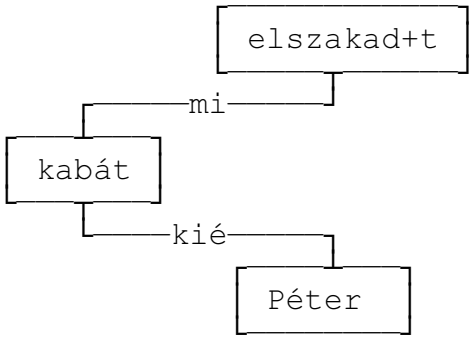
ez ugyanazt jelenti, mint

Péternek el volt szakadva a kabátja.

illetve

Elszakadt Péter kabátja. stb.

Természetesen egy pillanatig sem vonjuk kétségbe, hogy ezek a mondatok csak a tények szintjén jelentik ugyanazt, de az, hogy a beszélő mit akart bennük hangsúlyozni, más és más.

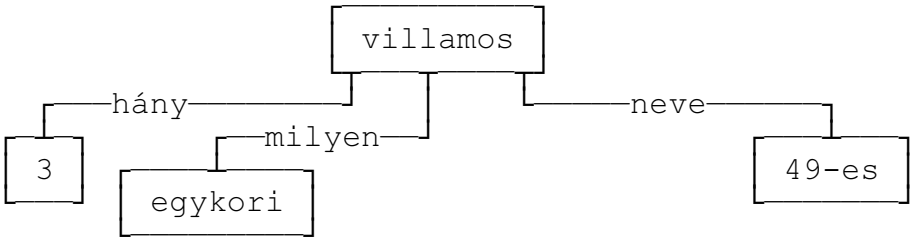


Alsóbb szintű névszói relációk

Az alsó szintű reláció egyik esete a jelzői kapcsolat. Ennek a következő fajtáit különböztetjük meg. A tulajdonságjelzőt a milyen szóval fogjuk jelölni. Melyik szóval fogjuk jelölni azt a jelzőt, ami kiválaszt egyet a több lehetséges közül. Pl.: ... *a 11. emelet* ..., ... *az utolsó nap* ..., ... *a legnagyobb téglalap* ... Név szóval jelöljük a dolgok nevét. A név nem csak tulajdonnév lehet. Pl.: ... *március hónapban* ... Hány illetve Mennyi szóval jelöljük a dolog számát vagy mennyiségét. Kié szóval jelöljük a dolog birtokosát.

Egy névszói kifejezésben nemcsak egy, hanem számos reláció szerepelhet egyszerre. Pl.:

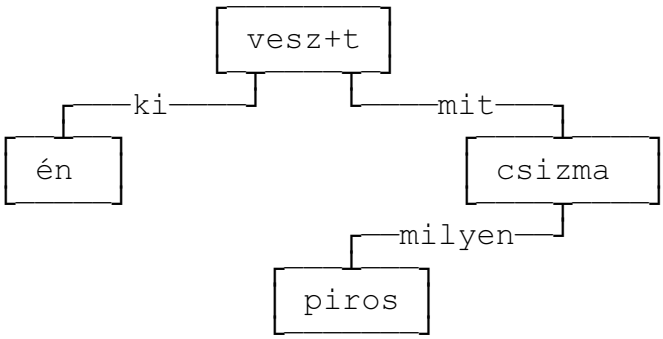
... három egykori 49-es villamos ...



Itt a villamos szóhoz három egymástól független reláció tartozik, de a villamos szó mégis csak egyszer szerepel az ábrában.

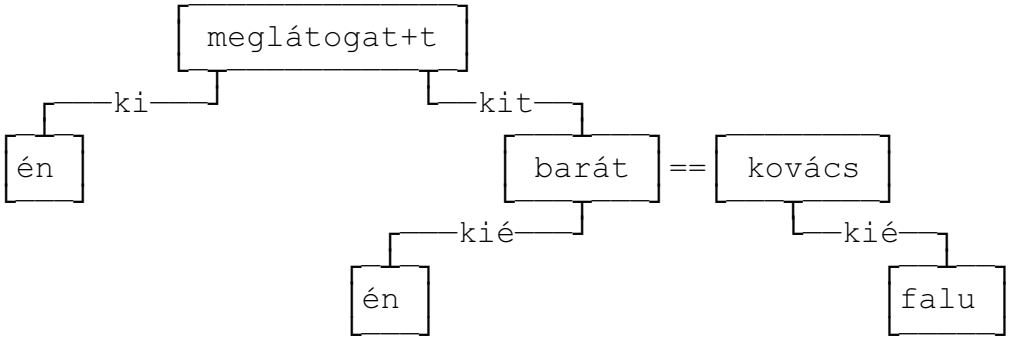
Az értelmezőket két csoportba lehet sorolni. Az egyik csoportban az értelmező tulajdonképpen egy közönséges jelző, amelyik pusztán hangsúlyozási okokból válik el a jelzett szótól. Pl.: *Agyonütöttem a legyeket, mind a hetet*. (*Mind a hét legyet agyon ütöttem*.) vagy

Vettem csizmát, pirosat. (*Piros csizmát vettem*.)



A másik esetben ugyanaz a dolog, két különböző névszói kifejezéssel is le van írva, és a két leírás nyelvtani/szemantikai okokból nem tehető össze egyetlen névszói kifejezéssé. Itt egy ugyanolyan azonosságról van szó, mint amiről a felső szintű relációknál már beszéltünk. Pl.:

Meglátogattam barátomat, a falu kovácsát.



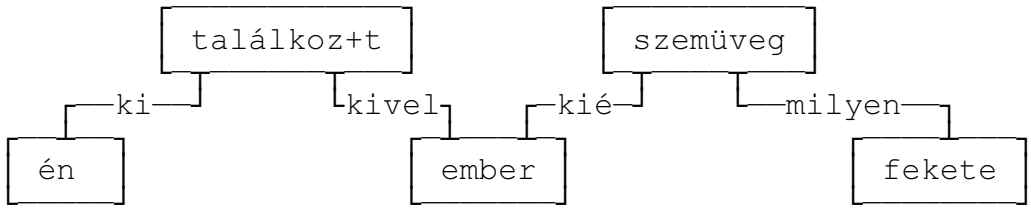
Nem tehető össze például a két névszói kifejezés eggyé, ha mindkettő főnévre végződik. (A fenti példa nem ez az eset, mert a kovács szó úgynevezett főnév/melléknév. Lásd: *Meglátogattam kovács barátomat*.) Vagy, ha olyan birtokos, vagy határozó van a névszói kifejezések legalább az egyikében, ami ezt megakadályozza.

Az alsó szintű névszói relációk között olyanok is vannak, amelyeknek nincs felsőszintű megfelelőjük. Szinte minden főnévből lehet az *-s/-os/-as/-es/-ös* vagy az *-ú/-ű* illetve a *-tlan/-tlen* képzővel melléknevet képezni. Ez tulajdonképpen egy birtokos kapcsolat melléknévi formában kifejezve, ahol a melléknévi formájú főnév a birtok, aminek a jelzője az a birtokos. Az ilyen melléknévnek, azaz tulajdonképpen a birtoknak is szokott jelzője lenni. Pl.: ... *fekete szemüveges ember* ..., ... *igei állítmányú mondat* ... stb.

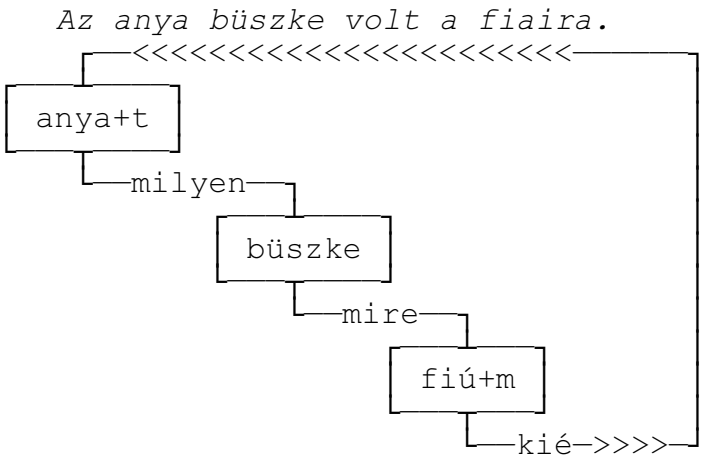
Kérdés, hogy melyik vonatkozást ábrázoljuk: a jelzőit, birtokost, vagy mindkettőt. Mi a birtokos relációt választottuk. Pl.:

Találkoztam a fekete szemüveges emberrel.

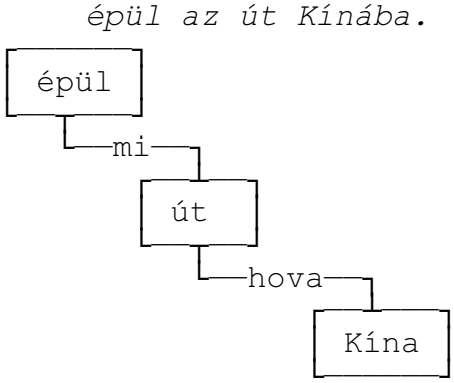
(*Találkoztam az emberrel, akinek fekete szemüvege volt.*)



Továbbá a mellékneveknek is lehetnek határozói, pl.: ... *a fiaira büszke anya* ..., ... *a korábbiaktól független döntés* ... stb. (Ezeket ragjuk alapján találjuk meg. De máshol keressük attól függően, hogy a melléknév jelző volt-e, vagy állítmány.) Pl.:

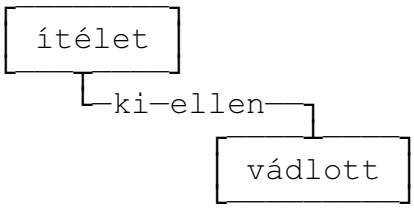


Főneveknek is lehetnek határozói, pl.: ... *ítélet valaki ellen* ..., ... *út Kínába* ... stb. Pl.:



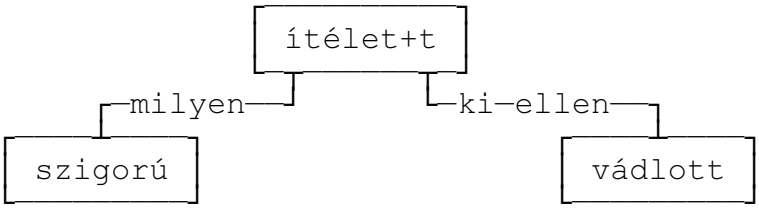
Van olyan vélemény, hogy ezek olyan szerkezetek, amiből hiányzik az ige, pl.: ... *Kínába vezető út* ..., ... *valaki ellen hozott ítélet* ... stb. Nekünk erről éppen ellenkező a véleményünk. Így véljük, a vonzat a főnévhez tartozik és az ige töltelék jellegű. Például a hoz igének nincs ellen névutóval járó vonzata, az út és a hova vonzata közé számos ige betehető lényegében azonos jelentéssel, így pl.: *megy, vezet, visz, tart* stb. Az **igenevek** című fejezetben részletesen leírtuk az általunk "külső vonzat"-nak nevezett vonzatok és a töltelék melléknévi igenév szintaktikai viselkedését. Az alsóbb szintű konstrukcióknál a töltelék igenevet, mint egy esetlegesen megválasztott és szemantikailag tartalmatlan komponenst egyszerűen kihagyjuk az ábrázolásból. Pl.:

... a vádlott ellen hozott ítélet ...
(... a vádlott elleni ítélet ...)
(... az ítélet a vádlott ellen ...)



Mivel az igét itt töltelékszónak tekintjük, így nemcsak a melléknévi igenév esetben, hanem állítmányi esetben is célszerű kiirtani a mondat ábrázolásából. Pl.:

Szigorú ítéletet hoztak a vádlott ellen.
(Szigorú volt a vádlott elleni ítélet.)



A kommunikáció egyszintű modellje.

A korábbiakban azt vázoltuk, hogy a mondat jelentését milyen formában akarjuk ábrázolni. Ebben a részben arról lesz szó, hogyan akarjuk felhasználni ezt az ábrázolást az ember-gép dialógus megvalósításához. Két különböző modellt is bemutatunk.

Az első modellben úgy képzeljük el, hogy az emberek, aki a számítógéphez fordulnak, kétféle tevékenységet csinálhatnak: tényeket közölhetnek a számítógéppel (kijelentő mondatok) és kérdéseket tehetnek fel az eltárolt tényekkel kapcsolatban (kérdő mondatok). Az első esetben a számítógép lefordítja a mondatban szereplő információt az általunk vázolt igei és névszói relációkká és beleilleszti azt a korábbról meglevő információi közé. (Azt a kérdést egyelőre tegyük félre, hogy mi történik olyankor, ha az új információ ellentmondásban van a korábbiakkal.) A feltett kérdés lehet eldöntendő vagy kiegészítendő. Az eldöntendő kérdés lényegében egy ugyanolyan reláció, mint egy állítás. Meg kell vizsgálni, hogy ez az állítás hogyan viszonylik a felhalmozott tudásunkhoz. Három eset lehetséges:

- Az állítás következik a felhalmozott tudásból. A válasz tehát: igen.
- Az állítás ellentétben van a felhalmozott tudással. A válasz tehát: nem.
- Az állítás sem nem következik, sem ellentmondásban nincs a tudással. A válasz: nem tudom.

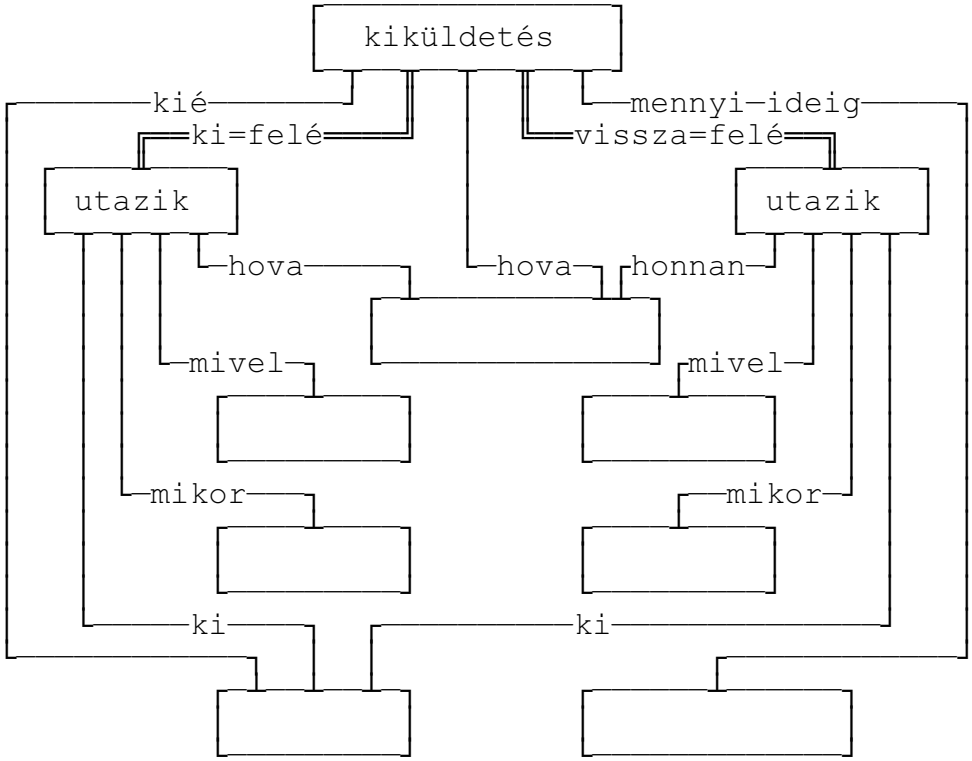
A kiegészítendő kérdést úgy képzelhetjük el, hogy az egy olyan relációvá fordult, amiben a kérdőszó helyén egy üres doboz áll. Feleletként azt kell megadnunk, hogy a felhalmozott tudás alapján mi kerülhet az üres dobozba. A válasz lehet egy vagy több dolog is, vagy az is lehet, hogy Nem találtam ilyent, ha semmi odaillőt nem találtunk.

Már ez az egyszerű modell számos problémát vet fel, amelyek egymással is kapcsolatban vannak, és kapcsolatban vannak a konkrét implementált ábrázolási formával.

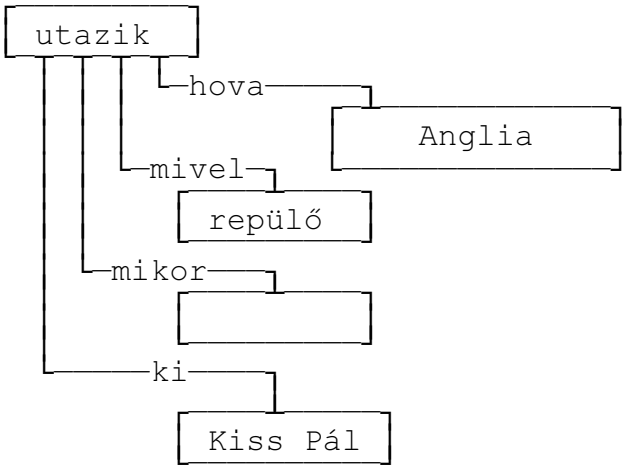
Az első fontos probléma, amit tárgyalnunk kell, az, hogy nem elegendő a kapott mondatok jelentését egymás mellé rakva felhalmozni, hanem a kapott állításokat össze kell egymással kapcsolni, és az állításokból következtetéseket is le kell vonni.

Például, ha egy titkárnőnek azt mondjuk, hogy: *Kiss Pál elvtárs Angliába repül. A kiküldetés 5 napig tart. Január 17-én utazik oda.* akkor elvárható, hogy a titkárnő válaszolni tudjon arra kérdésre, hogy: *Mikor érkezik haza Kiss elvtárs?*. Ugyanez egy számítógéptől nem biztos, hogy elvárható, hiszen ehhez egy modellt kell felépíteni, amely a következőket tartalmazza: *Egy kiküldetés egy kiutazásból, egy kinn tartózkodásból és egy visszautazásból áll. A kiutazásnak és visszaérkezésnek van egy időpontja, a kettő közti különbség a kinn tartózkodás időtartama. A kinn tartózkodásnak van egy helye, ami a kiutazás célpontja és a visszautazás kiinduló helye. Egyéb információ hiányában feltehető, hogy az elutazás itthonról történik, és ide érkezik vissza a kiküldött.*

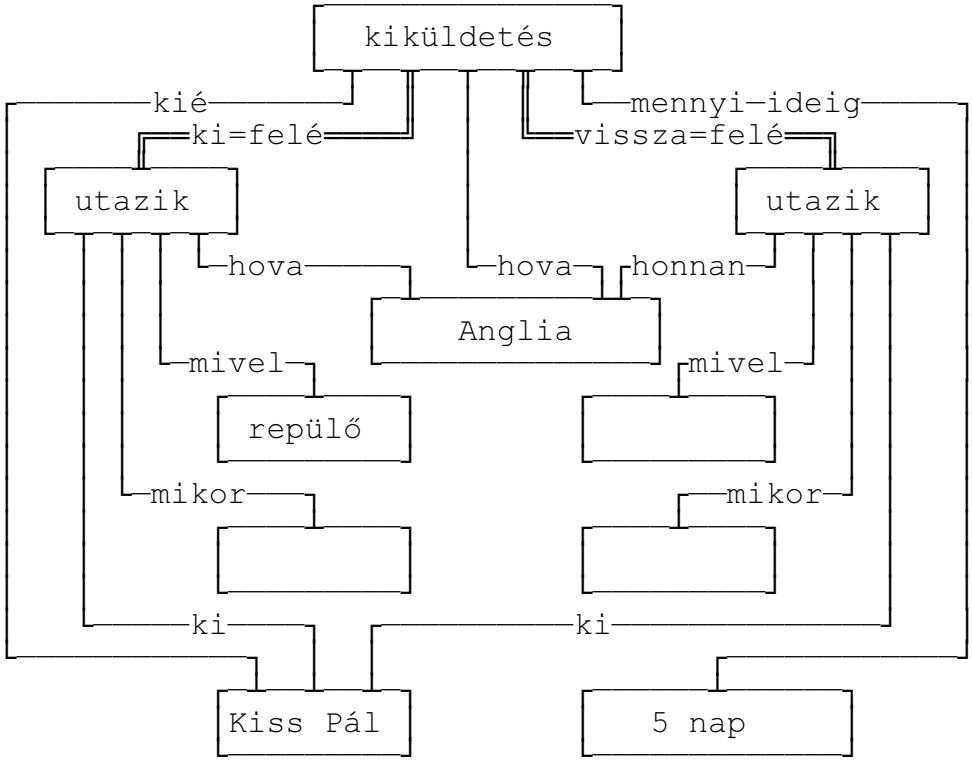
Ezek olyan mindennapi ismeretek (common sense knowledge) amelyek nélkül a fenti szöveg nem érthető meg, ha tehát a számítógép meg akarja érteni a fenti mondatokat, a kiküldetés modelljének eleve benne kell lennie, és ehhez kapcsolva vagy ezt lemásolva kell felépíteni a konkrét eset modelljét.



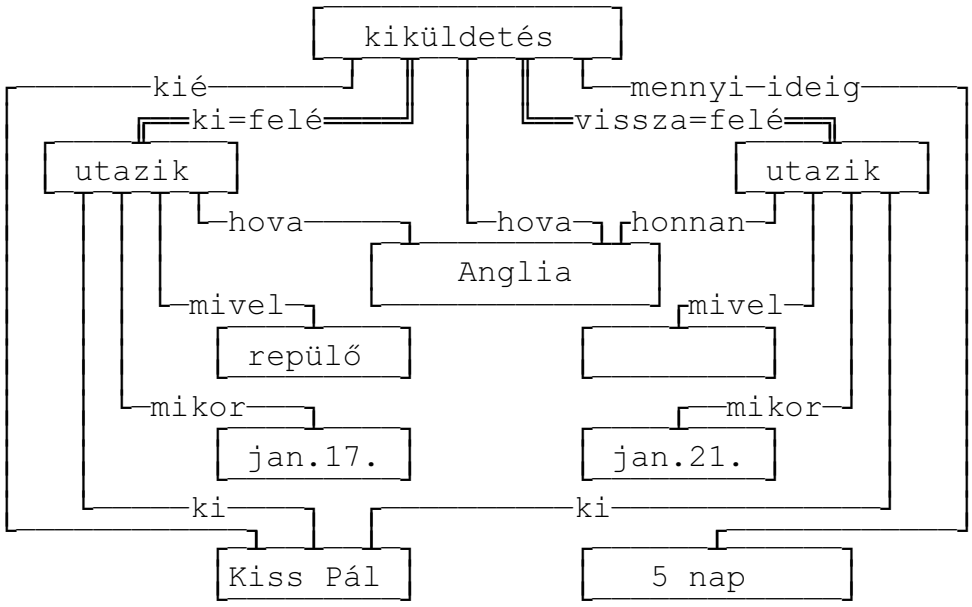
Az első mondatból még nem derül ki, hogy kiküldetésről lesz szó. Csak annyi tudunk, hogy valaki elutazik valahova (repülővel), feltehetőleg innen.



A második mondatból derül ki, hogy egy kiküldetésről van szó. Ekkor kell előhívni azt az ismeretet, hogy a kiküldetés 3 részből áll. Meg kell nézni, hozzákapcsolható-e az előzőleg nyert információkhoz. Kiderül, hogy igen, mert az előzőekben egy kiutazásról volt szó, ami valószínűleg egybeesik a kiküldetés első részével. A kettő egybevetéséből kiderül a küldött személye, a kiküldetés helye, a kiküldetés időtartama, de még mindig nem tudjuk a kiutazás és visszaérkezés időpontját.



A harmadik mondatban ismét egy kiutazásról van szó, aminél nem tudjuk, hogy ki utazik, hova utazik, mivel utazik, csak azt, hogy mikor. Ha ezt azonosnak vesszük a kiküldetés kiutazás részével, megkapjuk a kiutazás időpontját, és ha számolni is tudunk, akkor a visszautazását is.



Ekkor már tudunk felelni a feltett kérdésre, miután találunk egy *visszatérés*-t a Kiss elvtársra vonatkozó információk között.

Mindez még viszonylag kézenfekvőnek látszik, de ha arra gondolunk, hogy a kérdést úgy is fel lehet tenni, hogy: *Mikor lesz itt Kiss Pál?,* akkor további lépéseket is be kell iktatni, nevezetesen azt is hozzá kell fűzni, hogy amikor valaki megérkezik valahova, akkor ott van.

Természetesen ez csak a dolgok logikája, de vannak a feladatnak nyelvészeti vonatkozásai is. A kiküldetés 3. részére a visszatérés-re eddig a következő szavakat használtuk: *visszatér, visszautazik, visszaérkezik, hazaérkezik*; de használhatnánk a *visszajön, hazajön, megjön, megérkezik, visszarepül* stb. igéket is; valamint mindezek főnévi vagy melléknévi igenévi alakját is. Ha azt feltételezzük, hogy a szemantikus ábrázolás már megszüntette azokat a különbségeket, amelyek a tevékenység igei vagy névszói kifejezéséből fakadnak, akkor is szükség van arra, hogy erre a tevékenységre egységes megjelölést vezessünk be. Ez az egységes megjelölés lehet egy a használt megjelölések közül, de lehet valamilyen mesterséges megjelölés is. A kiválasztott igével való megjelölésnek az az előnye, hogy könnyű szavakba önteni azt, hogy mi is van ábrázolva a számítógépben, és könnyebb a válasz szöveges formáját létrehozni. Hátránya viszont, hogy egy olyan szót kell kiválasztani, ami képes az összes szóba jövő ige összes szóba jövő bővítményét viselni. Ilyet nem biztos, hogy könnyű, sőt egyáltalán lehetséges találni.

Ha például az *utazik* igét választjuk ki reprezentánsul, akkor a visszarepül igét át kell alakítanunk *utazik+repülővel+vissza* alakba, a *megjön* igét pedig át kell alakítani *utazik+ide* alakra; miközben a többi bővítményüket értelemszerűen átvesszük.

A mondatban vannak új információk és hivatkozások régi dolgokra. Amikor a rendszerbe információkat viszünk be, meg kell keresni a hivatkozott dolgokat, és össze kell kapcsolni őket az új relációkkal. Az adott és az új szétválasztása meglehetősen nehéz feladat, számos ezzel kapcsolatos nyelvészeti teóriát ismerünk (főleg más nyelvekből), de nincs tapasztalatunk arról, hogy válnak be.

A fogalmi gráfokban a régi és új információ összekapcsolása úgy történik, hogy olyan részeket keresünk a régi információk gráfjában, amelyek minél nagyobb mértékben fedésbe hozhatók az új mondat gráfjának egy-egy részével és ezekhez a régi információkhoz kötjük hozzá az új kapcsolatokat. A fogalmi hálók világában ezt a műveletet minimális projekciónak nevezik, és pontos matematikai definíciója van. Létezik olyan megvalósított információs rendszer, ahol az adatok letárolása és visszakeresése így (fogalmi gráfokon keresztül) történik. Sajnos konkrét tapasztalataink ezen a téren sincsenek. Viszont elképzelhető ennek a módszernek nyelvészeti meggondolásokkal való kombinációja is.

Sémák és konkrétumok

Mint az a fentiekből is látható, a dolgok közti relációk kezelésében két szint különböztethető meg. Egyrészt vannak általános érvényű összefüggések (szabályok vagy sémák), és vannak összefüggések konkrét dolgok között. Vannak olyan rendszerek, amelyben csak konkrét összefüggések vihetők be, tárolhatók el és kérdezhetők vissza. Az általános összefüggések csak segédeszközök a konkrét összefüggések megadásához. Lehet, hogy ezek a sémák egészen más formában vannak ábrázolva a programban, mint a konkrét kapcsolatok.

Az ilyen rendszereket, amelyekben csak konkrét adatokat lehet bevinni, de új összefüggéseket nem, a továbbiakban zárt rendszer néven fogjuk emlegetni. Azokat a rendszereket, amelyekbe viszont új fogalmakat és összefüggéseket is be lehet vinni, nyílt rendszernek fogjuk nevezni.

Az persze, hogy mi a konkrét, és mi az általános, pillanatnyi nézőpontunktól függ. Rosyhoz, a cirkuszban trombitán játszó fehér elefánthoz képest az elefánt fogalma általános fogalom; de az elefánt fogalma egy biológiai rendszerezésben meglehetősen konkrét dolog az élőlény fogalmához képest.

A fogalmak többszintű hierarchiáját az öröklődő tulajdonságokkal és műveletekkel legjobban a szemantikus hálók (semantic network) írják le. Viszont (legalábbis nekünk ez a benyomásunk) az öröklődésen kívül mást nem is igen tudnak leírni. Hasonlóan, a szemantika logikai leírása is lehetőséget ad a fogalmak bármilyen hierarchiájának leírására. Ezekben a rendszerekben abból, hogy Az élőlény halandó, automatikusan következik, hogy Rosy, a fehér elefánt is halandó. Ezzel szemben igen problematikus a kivételek kezelése. Ha tudjuk, hogy az elefántnak két szeme van és szürke, akkor egy félszemű vagy fehér elefánt bevitele gondot okozhat, hiszen egy olyan ellentmondást viszünk be, amely az egész következtetési rendszert felboríthatja.

A frame-ek és fogalmi hálók általában kétszintű modellben gondolkodnak. A felső szint a kitöltetlen vagy általános fogalmakkal kitöltött séma. Az alsó szint, amikor ezt konkrét objektumokkal töltjük ki. Közbülső szintek úgy jöhetnek létre, hogy bizonyos dolgok már konkretizálva vannak, mások pedig még nincsenek. Ezek az eszközök leginkább tényeket tartalmazó zárt rendszerek leírására alkalmasak. És mivel a rendszer nem akar messzemenő következtetéseket levonni, nem jelent különösebb nehézséget, hogy kivételes dolgokat is leírjunk benne, vagy a korábbi tényeket és feltételezéseket megváltoztassuk.

Ellentmondó információk

Ha egy rendszerbe egymásnak ellentmondó információk kerülnek, annak végzetes következményei lehetnek. A kisebbik baj, ha a rendszer egyszer ilyen, egyszer olyan választ ad. Ha például az mondjuk: *Ádám nemzé Sétet, Sét nemzé Enóst, Enós nemzé Kénant, Kénan nemzé Mahalélt, Mahalél nemzé Jeredet, Jered nemzé Énóst, Énók nemzé Methúselahot*,.... Akkor egy kis elírás folytán bármilyen választ kaphatunk arra, hogy *Ki élt előbb: Kénan vagy Jered?*, attól függően, hogy milyen sorrendben kapcsoljuk össze a párokat. Minél messzebbmenő következtetésekre képes a rendszer annál nagyobb lehet a baj, különösen, ha tekintetbe vesszük, hogy a matematikai logikában egy implikáció mindig igaz, ha az előtagja hamis.

Sajnos az ellentmondó információk ellen nincs igazi védekezés, mert nem biztos, hogy az utoljára bevitt információ a hibás, hanem lehet, hogy valamelyik korábbi információnk volt hamis, és máris számtalanszor felhasználtuk egy következtetés kiindulópontjául. Tulajdonképpen csak azok a rendszerek tudnak igazán védekezni, amelyek korlátozott következtetési képességgel rendelkeznek. Egy ilyen rendszer vagy nem érzékeli az ellentmondást, vagy át tudja te-

kinteni egy újonnan bevitt összefüggés összes következményét, és így kideríthetni a fennálló ellentmondást.

Egy olyan rendszerben (például a PROLOG-ban), ahol egy állításnak végtelen sok konzekvenciája lehet, az ellentmondás-mentesség nem deríthető ki teljes bizonyossággal.

A PROLOG-ban írt rendszerek rendszerint abban is különböznek azoktól a rendszerektől, amelyek a felhalmozott tudást adatstruktúrákban tárolják, hogy míg az utóbbiak egy új állítás konzekvenciáit rögtön a bevitel során levonják, és így a kérdezéskor a válasz előállításához kevesebb következtetésre van szükség; addig a PROLOG rendszerek a következtetések zömét válaszadáskor végzik el. Világos, hogy az ellentmondások felderítése szempontjából az a jobb, ha minél előbb látjuk a konzekvenciákat.

Az eltárolt információ teljessége

Képzeljünk el egy olyan rendszert, amelyben egy vállalat személyzeti adatait tároljuk. Tegyük fel a kérdést: *Van-e 25 évnél fiatalabb kazánkovács az üzemben?*. Ha a program nem talál ilyet, nyilván azt válaszolja: *Nincs*, és a válasz korrekt. Ha ezzel szemben a Turbo Prologhoz adott Geobase demonstrációs programnak feltesszük a kérdést: *Van-e 25 kilométernél rövidebb folyó Amerikában?*, azt válaszolja, hogy *Nincs*, és a válasz nyilvánvalóan nem igaz.

A tagadó választ mindkét esetben, azért kaptuk, mert a program a zárt világ feltételezésével (closed world assumption) élt. Feltételezte, hogy amit nem adtak meg neki: az nincs, az nem igaz. Az első esetben ez a feltételezés jogos volt, a másodikban nem. Ahhoz, hogy a rendszer korrekt válaszokat tudjon adni, általában nem elég, a megadott tényeket és összefüggéseket tudni, hanem azt is tudni kell, hogy mit tudunk. Az is primitív megoldás, ha minden olyan kérdésre, amire nem találtunk adatot azt mondanánk: Nem tudom, Nincs adat. Például arra a kérdésre, hogy: *Van-e olyan állam Amerikában, melyben 25000-nél kevesebben laknak*? a válasz joggal lehet: *Nincs*, hiszen ebből a szempontból az információ teljes.

A programnak ezt az önismeretét metatudásnak nevezik. A metatudás körébe nem csak az információ teljességének ismerete tartozik, hanem az is, hogy a program tudja, hogy milyen adatokat kezel és milyen adatokat nem, mire tud válaszolni és mire nem. A jó program tájékoztatja erről a felhasználót, ha az érdeklődik, vagy ha olyan helyes vagy helytelen inputot küld, ami a programot erre indítja. Nekünk az a véleményünk, hogy egy természetes nyelvű rendszer nem viselkedhet kellően értelmesen, ha a metatudást nem kezeljük.

A kommunikáció kétszintű modellje

Egy másik lehetséges modell a következő: adva van a gépben egy program, ami bizonyos feladatokat képes megoldani. Egy olyan interfész programot iktatunk közbe, ami megérti és lefordítja a természetes nyelvű utasításokat a program utasításaivá, átadja a programnak, átveszi a program eredményeit (és hibaüzeneteit), visszafordítja azokat a természetes nyelvre, és úgy adja át a felhasználónak. Ebben a modellben a természetes nyelven megfogalmazott feladat két szinten oldódik meg, az alsó szint csak megérti a feladatot, de a tennivalókat egy másik szint végzi el.

A felhasználó kijelentéseket tehet, utasításokat adhat, és kérdéseket intézhet a programhoz. Ez a modell számos olyan problémát vet fel, amivel korábban nem találkoztunk.

Bekezdések

Az első lényeges különbségnek az látszik, hogy míg az előző program lényegében folyamatosan mondatról mondatra dolgozik, bármennyire hiányos is az a mondat; itt valószínűleg nem ez a helyzet. Csak nagyon primitív programok esetén képzelhető el, hogy a feladatot egyetlen mondatban meg tudjuk fogalmazni. Nem is látszik célszerűnek ezt erőltetni, hiszen egy hosszú komplikált mondat feldolgozása nagyobb feladat, mint amikor ugyanaz több rövidebb mondatban van leírva. Arra kell tehát felkészülni, hogy az input bekezdésekből áll, és a bekezdés végén a felhasználó jelzi (például egy üres sorral), hogy a bevitt információknak vége van, kezdődhet a munka.

Egy sor matematikai feladatot és feladványt átnézve azt tapasztaltuk, hogy a feladat kitűzésének két tipikus formája van. Az első esetben a bekezdés első mondata egy kérdés, hogy mire vagyunk kíváncsiak, vagy egy felszólítás, hogy mit kell kiszámítani; a további mondatok pedig a szükséges paramétereket és korlátozó feltételeket részletezik. A másik esetben először a körülményeket adja meg a bekezdés, és utána jön a kérdő vagy felszólító mondat. Nem készítettünk statisztikát, de úgy tűnik, a két formának körülbelül egyforma esélye van. Fél százaléknál is kisebb az esélye annak, hogy a kérdés vagy felszólítás után még valami kommentár következik. Van viszont némi esélye annak, hogy a kérdés vagy felszólítás elmaradjon. Ilyen esetekben a feladat megadója feltételezi, hogy kitaláljuk, hogy mi a feladat. Pl.

Az f(x), f(f(x)), f(f(f(x))), ... sor divergens, ha x<0. (Bizonyítsd be.),

Nem igaz, hogy ... (Találjunk ellenpéldát.), *Ilyeneket te is készíthetsz.* (Készíts.) stb.

A fentiek alapján úgy tűnik, hogy elég jó közelítés, ha feltételezzük, hogy a bekezdésben pontosan egy felszólító, vagy kérdő mondat van, ez határozza meg, hogy mit kell tennünk. A többi mondat paramétereket és korlátozó feltételeket adja meg.

A tennivaló megértése

Annak a mondatnak a feldolgozása, ami a kérdést vagy felszólítást tartalmazza, meglehetősen különbözik az eddig leírtaktól. Egy dolog ugyanis megállapítani azt, hogy mit szeretne elérni a felhasználó, egy másik viszont annak megállapítása, hogy mit kell ehhez tenni. Az utóbbi nem a nyelvhez, hanem az adott feladathoz kapcsolódik, és feladatonként más és más.

Adatbázis-lekérdezés esetén tulajdonképpen csak a következő négy tevékenység képzelhető el: adatlekérdezés, adatbevitel, adatkitörlés, adatmódosítás. A rendszerek nagy részénél ezek közül is csak az első. Ezek alapján könnyű elképzelni, hogy milyen kérdések és felszólítások hangozhatnak el, és mit mivé kell alakítani.

Egy rajzoló programnál is adott, hogy milyen rajzelemeket ismer a program, hogy lehet ilyen rajzelemeket elhelyezni a rajzon, és milyen kapcsolatokat tud megállapítani közöttük a program.

Egy fizikapéldákat megoldó program esetén viszont meg kell állapítani, hogy mi a keresett mennyiség, és mik az adott mennyiségek, keresni kell olyan összefüggéseket, amelyek elvezetnek az adottaktól az ismeretlenig.

A munka másik fele az, hogy azonosítani kell a megoldásban szereplő objektumokat, és ki kell venni a megoldáshoz szükséges attribútumaikat.

A hivatkozások feloldása

Egy feladatmegoldást akkor lehet elindítani, ha megállapítottuk, hogy mi a teendő, és az összes hozzátartozó adatot összeszedtük. Nem várható, hogy egy mondatban benne legyen a feladat, a hozzátartozó összes objektum és a rájuk vonatkozó összes adat. Ehhez egyrészt össze kell kapcsolni a bekezdésben szereplő objektumokat valahogy úgy, ahogy azt az egyszintű modellben leírtuk. Sőt az esetek nagy részében ez sem elég, hiszen a felhasználó valószínűleg egymáshoz kapcsolódó feladatok sorozatát akarja végrehajtani, és közben visszahivatkozik az előző lépesekben szerepelt objektumokra is. Még tovább menve nem csak azokra a dolgokra fog hivatkozni, amit ő mondott, hanem azokra is, amit a program válaszolt.

Ez utóbbi azt jelenti, hogy a számítógép válaszait is be kell illesztenünk a modellünkbe. Ez különösen problematikus az adatbázis-kezelés esetén, ahol válaszul egy hosszú listát kaphatunk, amit esetleg nem is alakítunk át természetes nyelvi mondatokká.

Ahhoz tehát, hogy a kétszintű modellben a hivatkozásokat feloldjuk, az alsó nyelvi szintre be kell építeni egy ahhoz hasonló modellt, mint amilyen az egyszintű modell maga. Mivel azonban ezeknek az összefüggéseknek a tárolása itt csupán mellékes feladat, arról is gondoskodni kell, hogy az így felgyűlt adatoktól időnként megszabaduljunk. Erre két módszer látszik lehetségesnek:

Az első esetben a számítógéppel folytatott dialógust részekre lehet osztani (amelyeket ülésnek nevezünk). Egy ülésen egy témakört tárgyalunk meg. Az ülés végét valahogy jelezzük, az összegyűlt adatokat töröljük, és tiszta lappal indulunk a következő ülésre.

A másik módszer esetében nem csak megjegyezzük a dolgokat, hanem fokozatosan felejtünk is. Amikor egy mondatot beviszünk a rendszerbe, a benne szereplő új és régi objektumokhoz feljegyezzük a hivatkozás időpontját is. Ugyanakkor töröljük a rendszerből mindazt az információt, amire viszonylag hosszú ideje (mondjuk 5 perce) nem történt hivatkozás.

Valószínűleg az adott feladattól függ, hogy melyik megoldás a célszerűbb.

Összefoglalás

Ebben az írásban megpróbáltuk leírni a természetes nyelvű mondatokban megadott információk útját, attól kezdve, hogy bekerülnek a rendszerbe, addig a pontig, amíg kitörlődnek belőle. Az út állomásai a következők: a nyelvtanilag elemzett mondat előállítása, a nyelvtani összefüggések átalakítása szemantikus összefüggésekké, standardizálás, az új és régi információk összekapcsolása, az elavult információk törlése.

Noha igyekeztünk, hogy egyetlen ismert módszer mellett se kötelezzük el magunkat, a leírás legalábbis az első részben valószínűleg legjobban a fogalmi hálókra hasonlít. Egyes helyesen, ahol erre szükség volt, igyekeztünk objektíve értékelni az egyes ismert módszerek előnyeit és hátrányait.

INDEX

ABC (ábécé) 21 22
alany 9 13 14 15 16 33 35 46 50-52 56-58 67 68 77 85-89 91 92 **93-95** 96 98 99 102 103 117 131-140 141-146 150 151 153 154 156 158 159
  nak-ragos alany 87-90 94 133
alárendelés 13 45, *lásd még* almondat
anyagnév 47 **55-56**
állítmány 14 33 46 47 50 53 57 58 64 68 **85-92** 93 96 98-101 104 106 120 132 134-138 140 141-144 152-154 156 159 163 165
birtokos 56-58 61 67 68 71 74-78 94 100 104 105 125-129 138-140 141 144-147 151 161-163
  nak-ragos birtokos 17 61 67 70 71 74 82 95 105 126 138-140
egybeírás 55 56 58 63 66 96 129 138 140
előtag 15 55
fokozás 15 16 20 57 58 60 61 120
  felsőfok 58 61 72 77 99 106
  középfok 20 58 72 106 121
főnév 9 12 14 19 20 47 50 **53-57** 58 59 67-76 86-88 94 99 104 117 120 126 129 131 132 138-140 143 147 151 152 154 156 158 163 164
hangrend 8
hasonulás 22 23 29
határozó(szó) 8 16 17 20 50 51 53 54 57 60 61 **62** 67 75 81 82 85 93 95 96 98 101-104 117 119 120 125 127 131-133 135 137-139 146 147 150 151 154 157 158 163 164
  helyhatározó 61 62 95 **96** 102 104
  időhatározó 12 61 62 95 **96** 100-102 127
  szabad határozó 8 54 64 93 96 **100-103** 122-124 125 135 137 139 147
hiány 68 89 98 122 123 141 144 164 173
ige 14 17 19 20 **50-53** 56 62 64 67 68 77 79 80 85-98 100 101 103 104 117-119 122 125 127 128 131-135 137-139 147 151-156 158 164 165 169 170
  létige 46 **50-51** 88 90-92 98-100 132 137 147 152 153
  segédige 14 46 **51-52** 64 86-88 92 98 132-134 147 152 154 155
igekötő 8 15 45 **62-63** 73 86 91 96 97 127 132-134
igenév 17 18 49 50 **52-53** 62 64 67 68 86 87 89-92 94 97 106 118 **131-140** 147 152 154-156 164
  főnévi igenév 25 33 46 50 **53** 62-64 86 88-90 97 103 105 120 131 132 **133-134** 135 154 169

határozói igenév 12 18 50 **53** 62 68 91 102 105 131 132 136 **137-138** 153 154 157 158
melléknévi igenév 12 23 31 50 51 **53** 54 58 62 68 71 72 76 77 83 90-92 119 120 131 132 **135-137** 139-140 143 153 154 156 161 164 165 169
szenvedő igenév 135-137 140
írásjelek 11-14 21 47 55 81
jel 15-17 **18** 19 21-47 53 56 57 62 70 81 132 135
birtokjel 12 15 17 18 25-28 33 40 43 46 47 67 68 70 71 75 77-78 97 146
birtokos jel 15 17 18 25-48 40 43 56 57 75 105 128
többes szám jele 15 17 18 25-48 43 75 97
jelző 20 46 54-61 69-72 74 76-77 79 82 83 88 90 95 104 105 111 113 117 118 120-121 135-136 140 154 156 157 162 163
tulajdonság (jelző) 46 47 58-60 69-71 76 98-99 104 135 151 160 162
mennyiség jelző *lásd* mennyiség, mértékegység, szám
birtokos jelző *lásd* birtokos
képző 12 15-18 **19-20** 21 23 25-28 33 36 40 45 49 50 52 56 57 58 62 75 87 118 120 121 131 132 134-139 154 158 163
kötőszó 13 25 45 **63** 68 70 74 **79-80** 95 104
kvantor 47 60 **61** 70 71 72 76 79 119 142 143 145
magánhangzó 23 29 34-37
mássalhangzó 23 29 34-37
melléknév 9 15 17 19 20 45 47 53 55 **57-59** 64 67-72 75-76 83 88-90 99 102 104-106 117 118 120 121 131 132 135 136 140 144 154-156 160 163
pozícionáló melléknév 47 **58** 60 72 77
melléknévi névutó 47 54 **58** 67 71 73 120 121 139 140
mellérendelés 13 45 97 102 **114-115** 118
mennyiség 58-61 70-71 76 82 99 104 121 151 160 162
mértékegység 17 47 **59-60** 61 70 71 76 83 99 121 151 160
mondat 48 49 51-53 56 57 62 63 65 67-68 74 75 78 81-82 **85-115** 117 118 120 122 124 125 126 128 129 131-138 141 143-147 149 151-154 156-161 165-168 170 173-175
almondat 56 **113-114**
egyszerű mondat 9 51 **85-106**
főmondat 13
igei mondat **85-97** 152
kérdő mondat 51 52 165 174
névszói (nominális) mondat 14 57 63 64 **98-106** 136 147 152
szenvedő mondat 49 51 53 **91-92** 136 137 153
tagadó mondat 51
tagmondat 13

mondatrész 13 14 49 56 63 85 89 91-97 102 103 105 106 117 137 146 151 154
módosítószó(cska) 14 45 50 57 58 **64** 71 76 86-88 152
múlt idő 17 20 36 41 46 87 91 98 99
név (tulajdon-) 47 **54-55** 57 69 70 121 141 142 147 151 152 159 161 162
névelő 31 45 61 **63** 69 70 71 135
határozatlan névelő 31 45 63 71 76 99
határozott névelő 12 31 45 61 63 70-72 77 78 136 141-143 145-147 159
névmás 17 46 47 48 53 56 57 59 60-64 70 73 75-79 85 90 99 103-105 125 127-129 141 142 144 159
névszó 12 16-20 25-28 31 35 40 43 46 49 **53-65** 102 121 126-129 134 135
névszói kifejezés 16 50 53-60 62 64 **67-83** 90 95 98-100 103-106 119-123 131 132 134-136 139 140 **141-147** 151 154 156 159 160 162 163 169
határozatlan névszói kifejezés 63 98 99 101 **141-147** 159-161
határozott névszói kifejezés 56 98 99 **141-147** 159
névszói szerkezet 14 **67-79** 145 152
névutó 8 12 20 46 48 54 57 58 **61** 62 65 67 68 73-79 81 82 85 91-97 100-105 117-123 **125-129** 135 138-140 164
összetett szavak 27 31 52 58
rag 15 **16-18** 19 20 21 25-28 31 40 49 51-53 56 60-62 67-81 85-87 93 95 96 100-105 117-123 125-129 132 133 135 136 138 140 142 143 147 151 163
esetrag **16-17** 18-20 25-28 **42** 50 56 57 97 103 119 126 128 129 140
igerag 12 **17-18** 20 25-28 **41-42** 33 67 87 88 92 97 98 131 133 134 141 152 154 158
személyrag 14 17-18 25-28 **43** 50 52 56 61 65 86 87 89 103 104 125-129 131-134
ragozhatatlan 27 45-46 48 51 57 61 **62** 104 133
sorszám 47 **60** 61 72 99 121
szám 11 47 60 121 151 160
számnév 12 17 19 20 45 47 59 **60-61** 63 64 70 71 76 99 101 143-145 147
szelektor 47 58 59 **60** 61 76-77 99
szófaj 12 14 15-17 19-20 21-47 **49-65** 69 70 76 79 81 132
tagadás 30 51 64 87 91 92 117
tárgy 8 30 36 41 46 50 67 68 77 79 80 85 **93-95** 141 147 150 151 153 154 156 158 159
toldalék *lásd* végződés tőváltozatok 22 **23** 24 25 **29** 36

```
vonzat 7 14 16 17 19 20 49 51-55 57 58 60-63 67-70 74-78
       81-83 86 87 89-91 93-97 100 103-106 117-124 125 127-
       129 136 139 140 147 151 152 154 164
      kötelező vonzat 12 122
      opcionális vonzat 122 123 140
végződés 8 9 15-20 21-48 49-51 56 53 58 62 125 128 152
```