



Szintaxis: elemzések

Bevezetés

- **Múlt óra: mondatrészek**
- **Mai óra:**
 - **Szintaktikai reprezentációs elméletek**
 - **Ágrajzok**
 - **Problémás jelenségek**





Számítógépes szintaxis

- **Szabályalapú elemzés**
 - Szakértők kézzel írnak szabályokat
- **Statisztikai elemzés**
 - Nagyméretű adatbázisok (treebankek)
 - Elemzők (parserek)
 - Treebankekből automatikusan kinyert szabályok alapján zajlik az elemzés

Statisztikai szintaktikai elemzés

- Angol nyelvre kifejlesztett technológiák
- Konstituens- (összetevős) és függőségi (dependencia) alapú elemzők
- Kötött szórend vs. szabad szórend



Konstituenselemzés

- Terminálisok: szavak
- Nemterminálisok: konstituensek
- Szabályok: baloldal egyetlen nemterminális

NP $\rightarrow$ *Det Nominal*

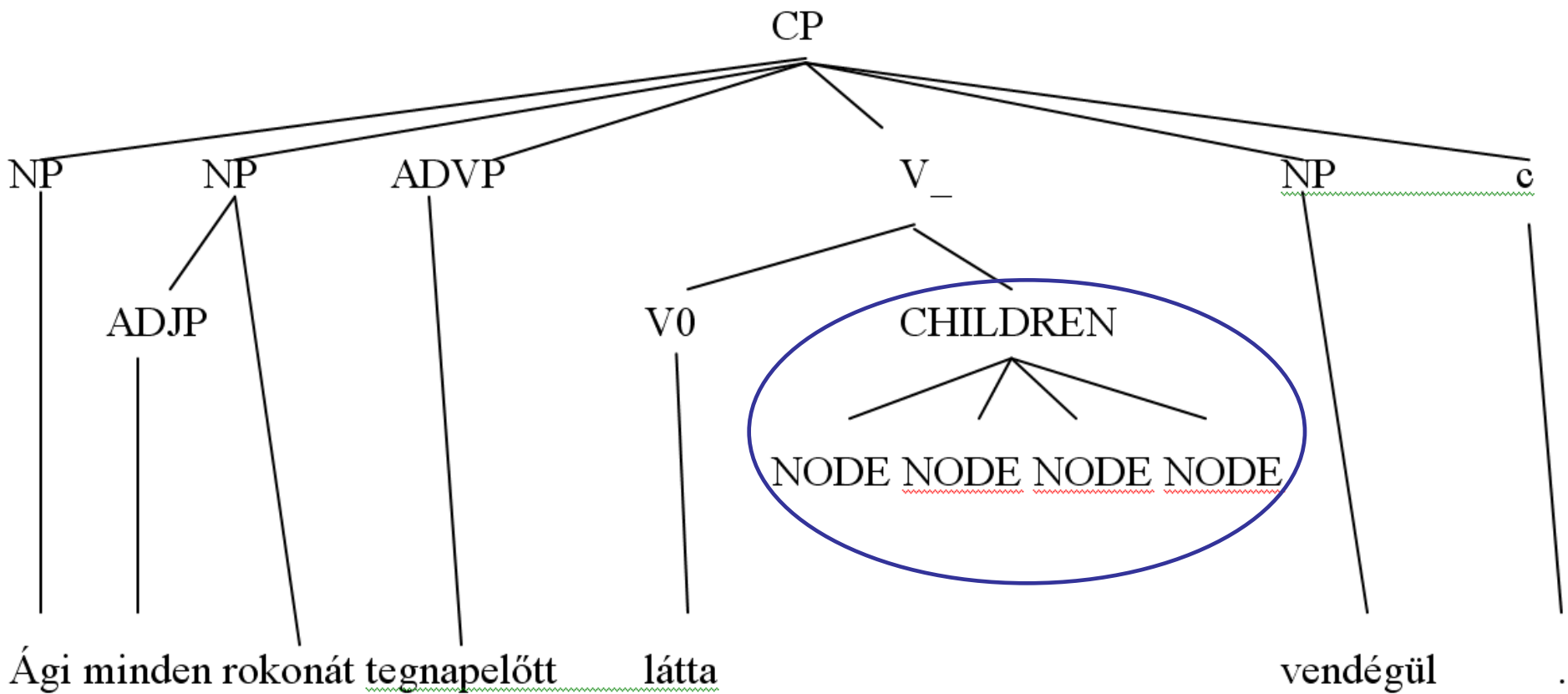
NP $\rightarrow$ *ProperNoun*

Nominal $\rightarrow$ *Noun* | *Nominal Noun*

Konstituensfák a Szeged Treebankben

- **Generatív szintaxison alapul (É. Kiss et al. 1999)**
- **A magyar nyelv szintaktikai sajátosságai szintén figyelembe vannak véve (nem „hardcore” Chomsky-féle nyelvtan)**
- **Az ige-argumentum viszonyokat címkék kódolják**
- **Nagyon részletes információk: minden esetraghoz külön nyelvtani szerep tartozik**
- **Szemantikai információ is megjelenik (idő- és helyhatározók)**







```

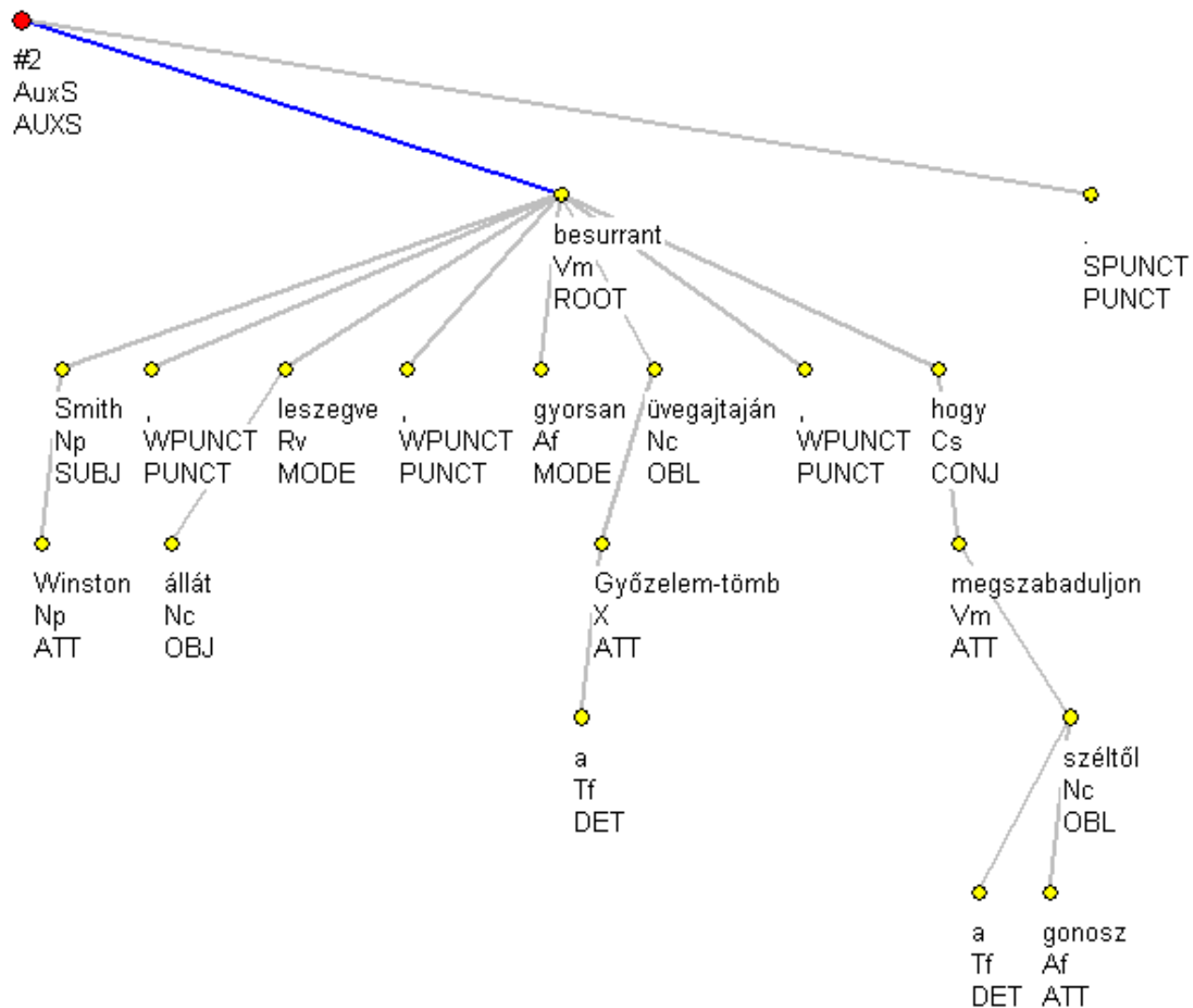
<CP id="fajl.1.1">
  <NP id="fajl.1.2">
    Ági
  </NP>
  <NP id="fajl.1.3">
    <ADJP>
      minden
    </ADJP>
    rokonát
  </NP>
  <ADVP id="fajl.1.4">
    tegnapelőtt
  </ADVP>
  <V_ id="fajl.1.5">
    <V0>
      látta
    </V0>
    <CHILDREN>
      <NODE idref="fajl.1.2" type="NP" role="NOM">
      </NODE>
      <NODE idref="fajl.1.3" type="NP" role="ACC">
      </NODE>
      <NODE idref="fajl.1.4" type="ADVP" role="TLOCY">
      </NODE>
      <NODE idref="fajl.1.6" type="NP" role="ESS">
      </NODE>
    </CHILDREN>
  </V_>
  <NP id="fajl.1.6">
    vendégül
  </NP>
  <c>
    .
  </c>
</CP>

```



Függőségi fák a Szeged Dependencia Treebankben

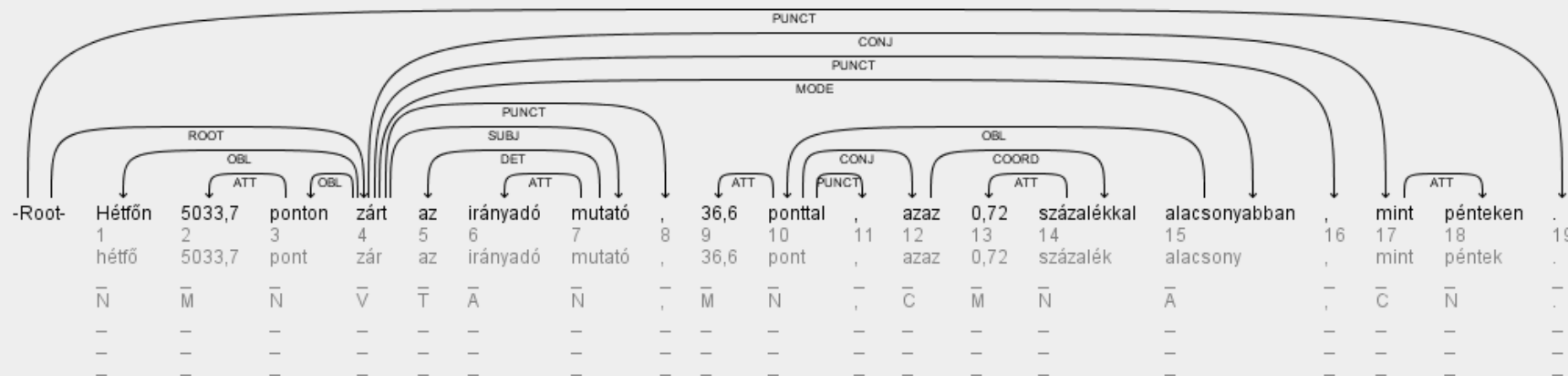
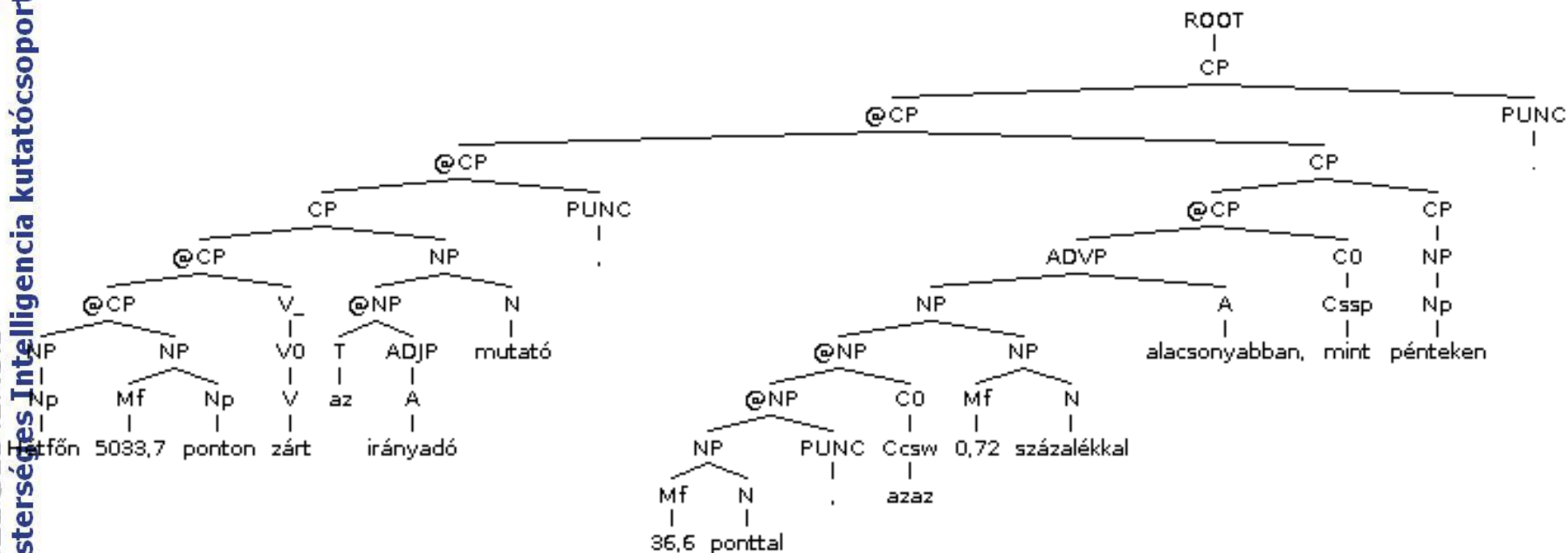
- A szavak a fa csomópontjai
- A főnévi argumentumok sokkal kevesebb szereppel rendelkeznek: SUBJ, OBJ, DAT, OBL, ATT -> morfológia kódolja már az esetragot
- Szemantikai információ megőrződik
- Lehetnek keresztező élek (*A fiúnak látta a kutyáját.*)





Függőségi vs. konstituenselemzés

- Minden csomópont egy szónak felel meg -> nincsenek művi csomópontok (CP, I'...) a függőségi fákban
- Konstituens-nyelvtanok általában kötött szórendű nyelvekre jók
- Mi határozza meg a szintaktikai viszonyokat?
 - a fában való elhelyezkedés (konstituens)
 - függőségi viszonyok (címkézett élek) (dependencia)





Problémás jelenségek

- Virtuális csomópontok
- Pro-drop
- Álalanyok, áltárgyak, ál részes esetek
- Mellérendelés
- Többértelműség

Virtuális csomópontok

- Kijelentő mód, jelen idő, 3. személyben nincs létige

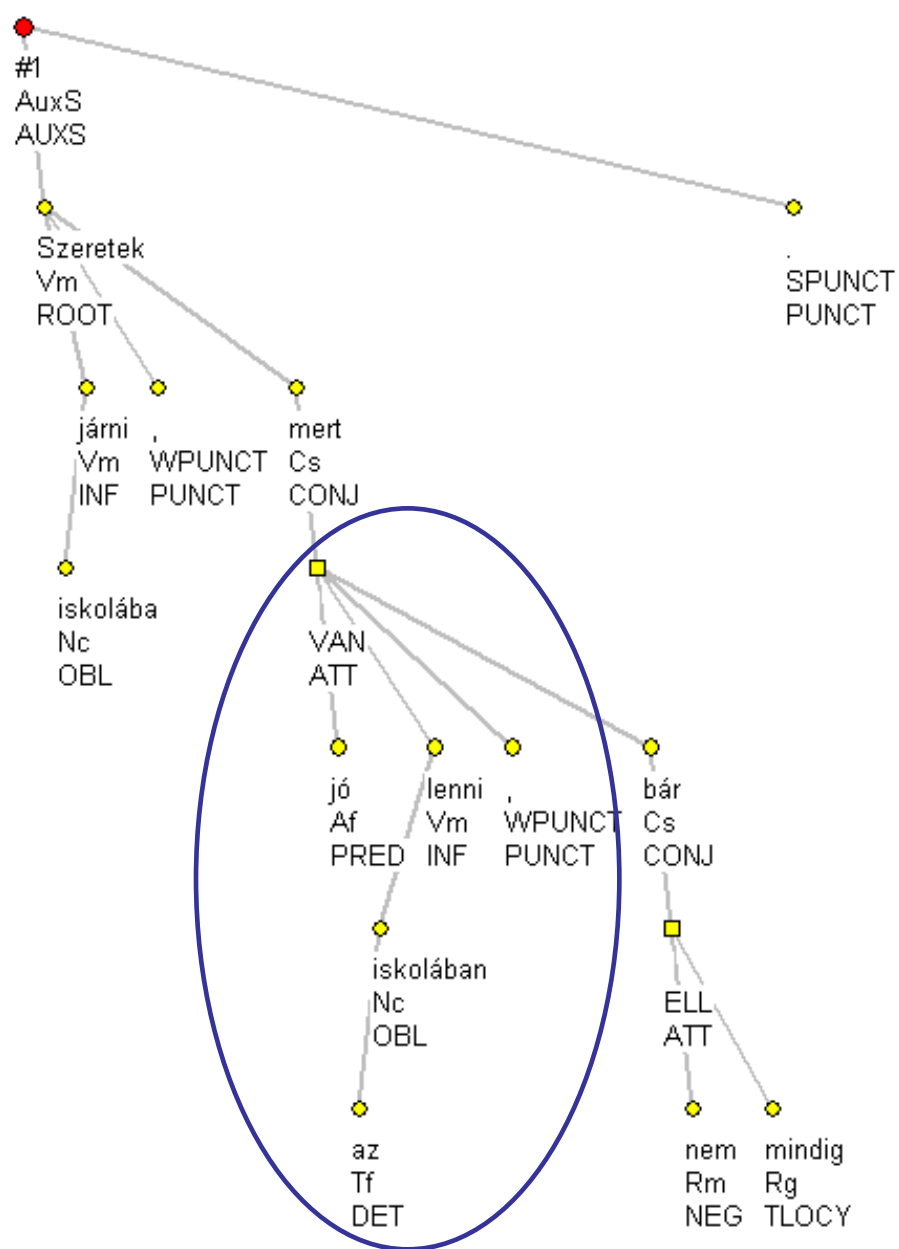
*András katona (*van).*

András legyen katona!

András katona lesz.

- A SzT-ben nincs jelölve ezeknek a mondatoknak a grammatikai szerkezete
- SzDT-ben virtuális csomópontok





Pro-drop

- Alanyi/tárgyi névmások nem mindig jelennek meg a mondatban

Láttad. – Te láttad őt/azt.

- Nyelvek közti eltérések: angol, francia... vs. magyar, olasz, spanyol...
- Attól, hogy nincs az igének látható tárgya, még lehet tárgyas... (egyeztetés)
- Nem összekeverendő az alanytalan igékkel:

Esik.

Havazik.



Alanyesetű, nem alanyi szerepet betöltő főnevek

- Birtokos: *a kisfiú labdája*
 - Predikatív névszó: *István juhász maradt.*
 - Tárgy: *A kutyám kergeti a macska.*
(kerti ösvény mondat)
- A fiam szereti a lányod.*



Lehetséges megoldások

- **Birtokos:**
 - SzT: egy NP tartalmazza a birtokost és a birtokot *((a kisfiú) labdája)*
 - SzDT: ATT reláció
- **Predikatív névszó: PRED reláció**
 - SzDT: virtuális csomópont
- **Tárgy: OBJ reláció**
 - Néha nem elég a szövegkörnyezet...
-> többértelműség



Áltárgyak

Formailag tárgynak látszó határozók:

Futottam egy jót.

Nagyot aludtam.

Tárgyatlan ige -> nem lehet tárgy -> MODE
reláció



Részes esetnek látszó esetek

Nem minden szemantikai alany áll alanyesetben:

- Részes esetű alany:

Sándornak kell elrendeznie az ügyeket.

- Mindkét treebankben DAT
- Bizonyos segédigék részes esetű alannal állnak (kivételek)
- DAT-NOM párhuzam: birtokos szerkezet



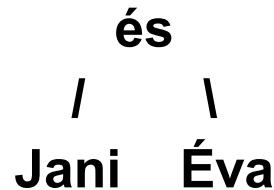
Mellérendelés

- frázisstruktúra-nyelvtanokban a csomópontok kettőzése a szokásos eljárás: NP $\rightarrow$ NP és NP
- dependencianyelvtanokban ez nem járható út, mivel nincsenek művi csomópontok – a szavakat kell összekötni valamilyen módon
- elvi lehetőségek:
 - vagy a kötőszót,
 - vagy az egyik mellérendelt tagot emeljük ki (szerepeltetjük fejként),
 - vagy egyiket sem (egyenrangúak)



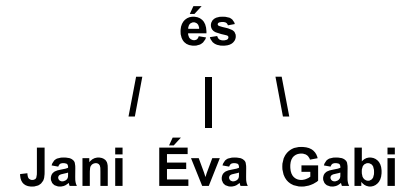
A kötőszó mint fej

- A szerkezet:



- DE: mi van, ha nincs kötőszó? => „virtuális és”

- más probléma: nincs különbség a *Jani, Éva és Gabi*, illetve a *Jani és Éva és Gabi* között



- „virtuális és” *Jani és Éva* közé, de: *Éva* két elemhez is kapcsolódna – ez tilos!



Nincs külön fej

- minden egyenrangú, egy szinten, egymás mellett szerepelnek a mellérendelés tagjai

mézeskalácsháza

/ | \

Jancsi és Juliska

- gond a viszonyokkal: *mézeskalácsháza* + *és*, *Jancsi* + *Juliska* „összetartozna”



Az egyik mellérendelt tag mint fej

- az első elem a fej (mert az tud az egész szerkezet helyett állni)

Elmentem a boltba Józsival és Katival.

Elmentem a boltba Józsival.

**Elmentem a boltba Józsival és.*

**Elmentem a boltba és Katival.*

- három részből áll össze a mellérendelés (Mel'čuk 1988, 2003):

Józsival
| CONJ
és
| COORD
Katival

Előnyös, mert:

- ha nincs kötőszó, ugyanígy ábrázolható:

Józsival
| COORD
Katival

- nincs szükség virtuális csomópontra
- hasonlít az alárendelésre, ami a számítógép szempontjából előnyös



Többértelműség

- morfológiai többértelműség:

szemét – szem+é+t

- szerkezeti többértelműség:

Reggel lelőttem egy elefántot a pizsamámban.

– Kin van a pizsama?

- lexikális többértelműség:

„megveted még hideg ágyamat”

– Visszautasítod vagy megágyazol?

- szemantikai többértelműség:

Minden férfi szeret egy nőt.

– Ugyanazt a nőt vagy mindenki másikat?



Szintaktikai többértelműség

- bővítmények csatolása:

Láttam a lányt a távcsővel.

– Kinél van a távcső?

- mellérendelés:

((Sikoltozó (gyerekek és anyukák)) szaladgáltak a játszótéren.

((Sikoltozó gyerekek) és (anyukák)) szaladgáltak a játszótéren.

- szintaktikai egyértelműsítés: az összes lehetséges fa előállítás a mondathoz és a legjobb kiválasztása

- lokális többértelműség: a mondat egy része többértelmű (több elemzés rendelhető hozzá), de a mondat maga nem (*a fiú kutyája* – az „a” hova csatlakozik?)



Többértelműség az angolban

Time flies like an arrow.

VB VBZ VB DT NN

NN NNS IN VB

NNP NN

RB

CC





Time flies like an arrow.

- Az időlegyek szeretnek egy nyilat.
- Úgy repül az idő, mint egy nyílvessző.
- A Time magazin úgy száll, mint egy nyílvessző.
- Az idő úgy menekül, mint egy nyílvessző.
- A Time magazin kiadója úgy száll, mint egy nyílvessző.
- Mérd a legyek sebességét úgy, mint egy nyílet.
- Mérd a legyek sebességét úgy, mint egy nyíl.
- Mérd meg nyílsebessen a legyek sebességét.
- Mérd meg azoknak a legyeknek a sebességét, amelyek egy nyílra hasonlítanak.

Egyeztetés

- A mondatban két/több összetevő morfoszintaktikai jegyei megegyeznek
- Szintaktikai kapcsolat jele

SUBJ (Per, Num) -> V (Per, Num)

A fiú fut - a fiúk futnak

SUBJ (Per, Num) -> V (Per, Num) lemma: *van, lesz, marad, múlik*

SUBJ (Per, Num) -> PRED (Num)

A lány szép - a lányok szépek

A rózsá piros maradt – a rózsák pirosak maradtak

A férfi harminc múlt – a férfiak harmincak múltak



Egyeztetés - 2

- OBJ (Def) -> V (Def)
Látom a gyereket.
Látok egy gyereket.
- Birtokos (Num, Per) -> Birtok (NumP, PerP)
az én könyvem
a te könyved
az ő könyve
- Kivétel:
az ő könyvük
a fiúk könyve



Egyeztetés - 3

Noun (Cas) -> DET (Cas)

Ez a lány – ezzel a lánnyal

DAT (Num, Per) -> INF (Num, Per)

*A fiúnak nem szabad futnia. - A fiúknak
nem szabad futniuk.*



Házi feladat

Nyomok várnak, s rejtek mélyén vész les.

- A mondat legalább 4 értelmezésének függőségi ábrázolása
- Határidő: november 27.
- Segítség:

<http://www.inf.u-szeged.hu/rgai/dependency>

<http://www.inf.u-szeged.hu/rgai/magyarlanc-service/>

