



Morfológia, szófaji egyértelműsítés



Előző órán...

- Morfológiai alapfogalmak
- Szóelemzések
- Ismeretlen szavak elemzése

Mai órán

- Szófajok
- Szófaji egyértelműsítés
- Mini morfológiai elemző készítése



Szófajok

- Szavak kategóriái
- Morfológiai, szintaktikai vagy szemantikai viselkedés alapján csoportok



Főnév (N)

- **Szemantika:** meghatározott fogalmat jelöl
- **Morfológia:** szám, eset, birtokviszonyok
- **Szintaxis:** alany, tárgy....



Ige (V)

- **Szemantika: meghatározott fogalmat jelöl**
- **Morfológia: idő, mód, szám, személy, tárgy típusa**
- **Szintaxis: (tag)mondat fő eleme**





Melléknév (ADJ)

- **Szemantika:** meghatározott fogalmat jelöl
- **Morfológia:** fokozás, szám, eset, birtokviszonyok
- **Szintaxis:**
minőség/mennyiség/birtokos jelző,
predikatív melléknév



Határozószó (ADV)

- **Nem határozó!!! (vö. szintaxis)**
- **Szemantika: meghatározott fogalmat jelöl**
- **Morfológia: fokozás(?)**
- **Szintaxis: határozó**

Számnév (NUM)

- **Szemantika:** meghatározott fogalmat jelöl
- **Morfológia:** sorszámnév, törtszámnév, szám, eset, birtokviszonyok
- **Szintaxis:** mennyiségjelző



Névmás (P)

- **Szemantika: helyettesítő szófaj**
- **Morfológia: helyettesített szófajnak megfelelő**
- **Szintaxis: helyettesített szófajnak megfelelő**



Egyéb szófajok

- **Kötőszó, névelő, névutó, indulatszó**
- **Szemantika: fogalmi tartalom nélkül**
- **Morfológia: változatlan alakok (kivéve névutók)**
- **Szintaxis: szófajfüggő**
 - **Kötőszó: kapcsolóelem**
 - **Névutó: határozó**
 - **Névelő: főnévhez tartozik**





Nyílt és zárt szóosztályok

- **Nyílt szóosztályok:**
 - Szabadon bővíülhetnek
 - Ige, főnév, melléknév (számnév?)
- **Zárt szóosztályok:**
 - Nem bővíülhetnek szabadon
 - „nyelvtani szavak”
 - Névelő, kötőszó, névmás...

Milyen szófajú szó hiányzik?

A ----- kínai állami hírügynökség jelentése közel ----- rendőr, ----- tűzoltó keresi reflektorok fénye ----- éjszaka is a magyar bázisugrót, ----- egy tűhegyes kövekkel teli ----- sziklafal fölött ----- szerencsétlenül ----- dél-kínai Hunan tartományban, a Csiangcsiacsie város közigazgatása alá tartozó Tienmen-hegy térségében. Kováts Viktor az ----- pénteken kezdődő világbajnokságra edzett más ----- együtt.



Milyen szófajú szó hiányzik?

A **Hsziehua** kínai állami hírügynökség jelentése közel **200** rendőr, **illetve** tűzoltó keresi reflektorok fénye **mellett** éjszaka is a magyar bázisugrót, **aki** egy tűhegyes kövekkel teli **meredek** sziklafal fölött **járt** szerencsétlenül **a** dél-kínai Hunan tartományban, a Csiangcsiacsie város közigazgatása alá tartozó Tienmen-hegy térségében. Kováts Viktor az **itt** pénteken kezdődő világbajnokságra edzett más **résztevéőkkel** együtt.



Szófaji egyértelműsítés

- POS-tagging – POS-tagger
- A morfológiai elemző által adott kódok közül kiválasztja az adott környezetbe illőt:

Megcsípett, a szemét!

V DET N-NOM

Hol csípte meg?

ADV V PREV

A szemét.

DET N-ACC



Morfológiailag többértelmű szavak

- A Szeged Korpusz szövegszavainak kb. 50%-a többértelmű
- Szófaji egyértelműsítés fontossága
 - *Várnak*
 - *Népének*
 - *Művére*
 - *Faszék*
 - *Vámpír*



Példamondat

Nyomok várnak, s rejtek mélyén vész les.

Lehetséges elemzés:

N: nyomok, várnak, s, rejtek, mélyén, vész, les

V: nyomok, várnak, rejtek, vész, les

Tényleges elemzés:

N: nyomok, rejtek, mélyén, vész

V: várnak, les



Morfológiai elemző – gyakorlati példa

-val/-vel toldalék

Adatok:

Ház – házzal – zal

Kéz – kézzel – zel

Fa – fával – val

Eke – ekével – vel

Sүн – sünnel – nel

Csoport – csoporttal – tal

Kapu – kapuval – val

Peti – Petivel – vel



Lépések

- **Adatok csoportosítása**
- **Általánosítások**
- **Szabályok felírása**
- **Szabályok rendezése**



1. Adatok csoportosítása

Változik-e a rag v-je?

Ház – házzal

Kéz – kézzel

Sүн – sünnel

Csoport – csoporttal

Fa – fával

Eke – ekével

Kapu – kapuval

Peti – Petivel



1. Adatok csoportosítása

Milyen magánhangzó van
a toldalékban?

Ház – házzal

Fa – fával

Csoport – csoporttal

Kapu – kapuval

őz – őzzel

Eke – ekével

Sün – sünnel

Peti – Petivel



2. Általánosítások

- **Változik-e a rag v-je?**
 - mgh után nem
 - msh után igen
- **Milyen magánhangzó van a toldalékban?**
 - á, a, o, u után a
 - e, i, ü, ő után e
- **Pótló nyúlás: szóvégi a, e -> á, é**



Szabályok

- Alapértelmezett (default érték választása):
 - Legtöbb esetet leírja
 - A defaulttól való eltérést könnyebb megfogalmazni, mint a defaultot 😊
- Default: *val*
- szó végén *msh*
 - *v* -> *msh*
- utolsó *mgh* *e*, *ő*, *ü* vagy *i*
 - *val* -> *vel*
- szóvégi *mgh* *e* vagy *a*
 - *e*# -> *é*#
 - *a*# -> *á*#



Szabályok rendezése

3 szabály, 6 lehetséges sorrend

kéz + val

eke + val

1. msh-szabály

kéz + zal eke + val

2. mgh-szabály

kéz + zel eke + vel

3. pótló nyúlás

kéz + zel eké + vel

1. mgh-szabály

kéz + vel eke + vel

2. msh-szabály

kéz + zel eke + vel

3. pótló nyúlás

kéz + zel eké + vel



Szabályok rendezése

kéz + val

eke + val

1. msh-szabály

kéz + zal eke + val

2. pótló nyúlás

kéz + zal eké + vel

3. mgh-szabály

kéz + zel eké + vel

1. mgh-szabály

kéz + vel eke + vel

2. pótló nyúlás

kéz + vel eké + vel

3. msh-szabály

kéz + zel eké + vel



Szabályok rendezése

kéz + val

eke + val

1. pótló nyúlás

kéz + val eké + val

2. msh-szabály

kéz + zal eké + val

3. mgh-szabály

kéz + zel eké + val

1. pótló nyúlás

kéz + val eké + val

2. mgh-szabály

kéz + vel eké + val

3. msh-szabály

kéz + zel eké + val





Szabályok rendezése

- Msh-szabály és mgh-szabály sorrendje mindegy
- Pótló nyúlás a végére!

Házi feladat

- *hoz/hez/höz* toldalékolás szabályainak megalkotása az alábbi példák alapján

*egér, macska, eke, sün, fű,
kalap, ördög, kapu, Peti*

- Határidő: október 31.

