



Annotáció, annotációs útmutató

Tervezés

- Mit szeretnénk pontosan annotálni?
- Pontos (nyelvészeti) definíció
- Érthető legyen mások (nem nyelvészek) számára is
- Mit NEM szeretnénk annotálni?
- Határok meghúzása



Cél megfogalmazása

- Miért akarjuk ezt jelölni?
- Mire lesz jó?
- Nyelvészeti felhasználás?
- Elméleti?
- Alkalmazott?
- Számítógépes alkalmazás?
- Mindegyik?



Szövegek kiválasztása

- **Forrás, mennyiség, tematika a felhasználás függvénye**
- **Van-e elegendő hozzáférhető anyag?**





Technikai háttér kiválasztása

- Miben annotálunk?
- Feladatfüggő!
- Nyers szövegben kell szövegrészeket jelölni (pl. tulajdonnévi annotáció): „színező” (TextAnnotator)
- Több lehetőség közül kell választani (pl. szófaji egyértelműsítés): rádiógombos felület (Tagging Assistant)
- Szintaktikai fák készítése: speciális szerkesztőfelület (TrEd)

Próbaannotáció

- Kezdjük el dolgozni a szöveg egy részén még az éles annotáció előtt
 - A kezdeti nehézségek gyorsan előbukkannak...
 - Nem várt nehéz esetek
 - Új kategóriák felbukkannak
 - Technikai gondok
- > ki kell javítani / újra kell definiálni a feladatot





Annotációs útmutató

- **Mi a feladat?**
- **Mi az eszköz, amiben dolgozunk?**
- **Milyen lépései vannak a jelölésnek?**
 - Technikai oldalról
 - Nyelvi oldalról (lehet pl. hierarchikus jelölés: alkategóriákat is jelölünk)
- **Hogyan határozzuk meg, hogy jelölünk-e? (nyelvi tesztek)**
- **Példák**

Példák

- A próbaannotációból sok példát lehet meríteni
- Írjuk bele az útmutatóba
- Segít az annotátoroknak
- Rendszeresen előforduló kétértelmű esetekben foglaljunk egyértelműen állást (birtokos szerkezet névelője?)
- Negatív példák is fontosak: mit NEM kell jelölni (ami esetleg annak tűnik, de mégsem az...)



Annotátorok

- Hány kell?
- Kik annotáljanak?
- Mennyire kell nyelvészeti / nyelvi tudás?
- Nyelvészek
- Laikusok
- Crowdsourcing / Mechanical Turk
- Mennyiség vs. minőség...
- Ha gyorsan kell sok (nem nyelvi) adat, jó lehet...



Annotátorok - 2

- **Elég-e a szimpla annotáció?**
- **Egyetértési arány**
- **Mennyit annotáljanak duplán?**
- **Mi lesz az etalon?**
 - **Eltérések kézi egyértelműsítése**
 - **Többségi jelölés**
 - **Annotációk metszete**
 - **Annotációk uniója**



Annotáció végén

- **Korpuszfájlok összeillesztése**
- **Technikai simítások**
- **Korpuszdokumentáció készítése (cikk, readme)**
- **A korpusz publikálása**
- **Hozzáférhetőség**
 - **Ingyenes (kutatási, oktatási célra)**
 - **Fizetős**
- **Licencek**

