

# A Lírakorpusz annotációs címkéi

## Tokenek címkéi:

**Core** – ige vagy főnévi igenév

**Aux** – segédige

**Prev** – igekötő

**Arg** – az ige bővítménye

**Arg:core** – az ige olyan főnévi bővítménye, amely az igével együtt alkot magmondatot (könnyű igét specifikáló bővítmény)

**Case** – névutó, névutómelléknév (pl. *asztal alatti*) vagy -ként rag funkcióban használt *mint*

**Poss** – jelöletlen vagy -nak/-nek ragos birtokos jelző

**Voc** – vokatívusz (a teljes vokatívuszi funkciót ellátó szerkezet, névelőkkel, jelzőkkel, indulatszavakkal együtt megkapja a címkét)

**Flat** – többtagú kifejezések (általában tulajdonnevek) esetében az utolsó elem előtti elemek

**Core:omitted**, **Arg:omitted**, **Aux:omitted**: elliptikus szerkezetet lehetővé tevő, nem az elliptikus szerkezetben, hanem egy korábbi vagy ritkábban későbbi igei szerkezetben realizálódó elem.

## Tokenek közötti viszonyok címkéi:

**tr** – ige vagy főnévi igenév (Core) kapcsolata a szemantikai alannyal (Arg) vagy a szemantikai alanyra utaló vokatívusszal (Voc)

**Im** – ige vagy főnévi igenév (Core) kapcsolata nem alanyi főnévi bővítménnyel (Arg) vagy nem alanyra utaló vokatívusszal (Voc)

**prev** – ige vagy főnévi igenév (Core) kapcsolata az igekötővel (Prev)

**case** – főnévi bővítmény (Arg) kapcsolata névutóval, névutómelléknével vagy a -ként rag funkciójában használt *mint* szóval (Case)

**aux** – főnévi igenév (Core) kapcsolata segédigével (Aux)

**poss** – főnév kapcsolata jelöletlen vagy -nak/-nek ragos birtokos jelzőjével (Poss)

**cor** – főnévi vagy névmási bővítmény (Arg) kapcsolata a vele koreferens vokatívusszal (Voc)

**flat** – többtagú kifejezések (általában tulajdonnevek) esetében az utolsó elem kapcsolata az előtte lévő elemekkel

## Megjegyzések a címkék használatához:

### Core:

- Ige vagy főnévi igenév kaphatja. Bővítmény nélküli, magában álló igék és főnévi igenevek is megkapják (pl. *Mást [Arg] nem csinál [Core]*, csak *eszik [Core]*, *iszik [Core]*, *alszik [Core]*).
- Főnévi igenév akkor kap Core címkét, ha segédigés szerkezetben szerepel, vagy ha magában áll fölérendelt ige nélkül (pl. *Messzire [Arg] jutni [Core]*, ez *volt a célja*.)

- Határozói igenevek sohasem kapnak címkét. Melléknévi igenevek csak Arg címkét kaphatnak, kizárálag abban az esetben, ha nem jelzői funkcióban szerepelnek, hanem az ige valamilyen bővítményére utalnak (pl. A messzire jutó [Arg] sokat tanul [Core].)

### **Arg:**

- Bővítményként annotáljuk az igék vonzatait, valamint a főnévként, személyes névmásként vagy személyes határozószóként megjelenő szabad határozókat. Az ige esetleges egyéb bővítményeit nem címkézzük. Például a mód-, állapot- vagy időhatározók jellemzően nem főnévként, hanem határozószóként jelennek meg, és nem is vonzatai az igének, így ezeket az esetek nagy részében nem kell annotálni.
- A *melllette*, *előtte*, *előttem*, *alattunk* típusú, személyjelöléssel ellátott személyes határozószavakat egyelőre ADP szófaji címkével annotáltuk. Ugyanakkor ezek a szintaktikai annotálásnál sohasem Case címkét kapnak, mivel Case címkét csak a főnevek és névmások után álló tényleges névutók kaphatnak. A személyes határozószavakat mindig Arg (Im) címkével kötjük a Core-hoz. Ezeket tehát ugyanúgy kezeljük, mint a személyes névmásokat.
- Fontos, hogy ha egy határozószó vagy bármilyen névmás nem szabad bővítményként, hanem vonzatként jelenik meg, akkor minden esetben Arg címkével. Vonzatoknál tehát nem kell a szófaj alapján mérlegelni, ezek automatikusan Arg címkét kapnak. Ugyanakkor nem biztos, hogy sok olyan eset van, amikor az ige vonzata határozószóval fejeződik ki. Ezek nagyrészt szabad határozók.
- Amennyiben nem főnévi szófajúak, a tárgyragos határozók sohasem kapnak Arg címkét. Például az „András sokat olvas.”, az „István egy kicsit fél a vizsgától.” és a „Dóra hármat kopogott az ajtón.” mondatokban a *sokat*, a *kicsit* és a *hármat* nem kap Arg címkét. Ezek ugyanis nem vonzatai az igének és nem is főnevek.
- Speciális esetekben egy ige vonzata melléknév is lehet. Ilyenkor a melléknév ugyanúgy Arg címkét kap. Ezek többnyire Im szerepben vannak (pl. András okosnak tűnik.). De előfordulhatnak tr szerepben is (pl. A szorgalmas sokat dolgozik.)
- Intuitív alapon különböztetjük meg a vonzatokat a szabad határozóktól. A kettő elkülönítésének megfelelő módszerein 100 éve gondolkoznak, úgyhogy ezt nem itt fogjuk megoldani. Annyi mondható, hogy az a bővítmény, amelynek elhagyásával agrammatikus lesz a mondat, biztos, hogy vonzat. Ezek a kötelező vonzatok. Ugyanakkor nem minden vonzat elhagyásával lesz agrammatikus a mondat. Ezek a fakultatív vonzatok. A fakultatív vonzatok olyan bővítmények, amelyeknek a jelenléte az ige sajátos szemantikai szerkezetéből következik. Ezzel szemben a szabad határozók olyan külső körülményekre utalnak, amelyekre szinte bármelyik igénél utalhatnánk.
- Sok olyan eset lesz, amikor az alanyi bővítménye az igének egy korábbi igei szerkezetben szerepel, és az aktuális szerkezetben csak az E/3 igeragozás utal vissza anaforaként a korábbi, fogalmilag kifejtett bővítményre. Ezekben az esetekben sohasem kötjük hozzá az igét a korábbi szerkezetben szereplő alanyi bővítményhez.

### **Trajector és landmark:**

- Szinte minden nominatívuszi bővítményt kötjük tr címkével az igéhez, és a többi bővítményt Im címkével. Ez alól kivételek a következő esetek:

- Az ige argumentumszerkezetében a datívuszi bővítmény áll a figyelem középpontjában / képviseli az aktívabb szereplőt (pl. *Dórának* [tr] *tetszik András* [lm]; *Andrásnak* [tr] *hiányzik Dóra* [lm]; *a leves* [lm] ízlik *Gábornak* [tr]; *Juditnak* [tr] *jól áll a ruha* [lm])
- Birtokos szerkezetek széttörésével létrehozott mondatok (pl. *Andrásnak* [tr] *fáj a feje* [lm]; *Remeg a keze* [lm] *Dórának* [tr]; viszont: *Andrásnak* [Poss] *a feje* [tr] *fáj*; *Remeg Dórának* [Poss] *a keze* [tr])
- *kell*, *szabad*, *lehet*, *muszáj* segédigés szerkezetek datívuszi bővítményei (pl. *Andrásnak* [tr] *találkozna kell Dórával* [lm]; *Zsófinak* [tr] *nem szabad Dórával* [lm] *találkozni*; *A munkát* [lm] *muszáj Bélának* [tr] *befejeznie*)
- Birtoklást kifejező, létigés mondatok datívuszi bővítményei (pl. *Dórának* [tr] *van autója* [lm]; *Nekem* [tr] *nincs könyvem* [lm]). A birtoklást kifejező létigés mondatok esetében akkor is a birtok a landmark, ha nincs jelölve a birtokos (pl. *Van autóm* [lm]; *Nincs autód* [lm]). A fennállást kifejező létigés mondatokban ugyanakkor az alanyesetű főnév a trajector (pl. *Van egy autó* [tr] *az úton* [lm])
- A szórend nem befolyásolja a trajector és a landmark kiosztását (kivéve a birtokos szerkezetek széttörésének esetét). Az *Andrásnak* *tetszik Dóra* és a *Dóra* *tetszik Andrásnak* mondatokban ugyanúgy *Andrásnak* a trajector.

#### **Arg:corePart címke használata:**

- Az Arg címke akkor kapja meg pluszban a corePart jellemzőt, ha a bővítmény könnyű igés szerkezetnek a része. Imrényi András terminológiájában ezek a magmondat részét alkotó igei bővítmények.
- Nem minden esetben egyértelmű, hogy egy igei bővítmény része-e a magmondatnak vagy nem. Ennek meghatározásában segíthet a kérdés-válasz teszt. Ha egy eldöntendő kérdés pozitív megválaszolásában az ige mellett a bővítményt is szükséges szerepeltetni, akkor a bővítmény része a magmondatnak, azaz megkapja a corePart tulajdonságát. Például:
  - Kérdés: András olvasta a könyvet? Válasz: Olvasta
  - Kérdés: András feleségül vette Dórát? Válasz: Feleségül vette (a *feleségül* tehát része a magmondatnak)
  - Kérdés: András letargiába esett? Letargiába esett. (a *letargiába* tehát része a magmondatnak)
- A fenti teszt nem minden könnyű igés szerkezet esetében működik. Amennyiben működik, biztos, hogy a magmondat része a bővítmény. Viszont abból, hogy nem működik a teszt, nem feltétlenül következik, hogy a bővítmény nem része a magmondatnak.

#### **Segédigés szerkezetek és főnévi igenevek annotálása:**

- Segédigés szerkezetek esetén a főnévi igenév kapja a Core címkét, és ennek alárendelt tagja az Aux címkével ellátott segédige. (pl. *András* [Arg] *tegnap nem tudta* [Aux] *elolvastni* [Core] *a tankönyvet* [Arg].)
- Az UD csak néhány szóra alkalmazza az AUX szófaji címkét, a segédigék nagy része VERB címkét kap. Ez utóbbiak a szintaktikai annotálásnál úgyszintén megkapták az Aux címkét.

- Bajzát Timi ezeket a szavakat listázta segédigéként: *akar, bír, fog, kell, kéne, kíván, lehet, lehetne, mer, óhajt, talál* (*talál vmit csinálni*), *tetszik, tud, szándékozik, szeretne* (*a szeret vitatható*), *szokott*
- Úgyszintén AUX címkét kapnak a segédmellékneveknek nevezett szavak, amelyek a következők (Timi alapján ez is): *hajlamos, hajlandó, kénytelen, képes, képtelen, muszáj, szabad, szükséges, tilos*
- Sima ige + főnévi igenév szerkezetek esetén az ige kapja a Core címkét, a főnévi igenév pedig Arg címkét kap lm kapcsolattal. (pl. *Dóra elment [Core] a klubba [Arg] golfozni [Arg] Andrással.*) Ebben az esetben a főnévi igenév esetleges bővítményeit nem annotáljuk.

### **Poss címke és poss viszony használata:**

- A *-nak/-nek* raggal jelölt vagy jelöletlen birtokos jelzők Poss címkét kapnak, és poss viszonnyal kötjük a birtok funkciójú igei bővítmény Arg címkéjéhez (pl. *András [Pos] táskája [Arg] elveszett [Core].*). A szerkezet fölérendelt tagja a birtokra utaló igei bővítmény, azaz a nyíl az Arg címkétől mutat a Poss címkére.
- Igei bővítmények birtokos jelzőinek a birtokos jelzőit nem annotáljuk. Például *A király apródjának kutyája egész nap ugat* mondatban csak az *apródjának* szót címkézzük Poss-al és kötjük a *kutyája* szóhoz.

### **A Case címke használata:**

- Case címkével jelöljük az Arg címkét kapó főnevek és névmások (esetleg melléknevek) után álló névutókat. A case címkével ellátott nyíl az Arg címkéből mutat a Case címkére. A Core-hoz sohasem a Case címkét kötjük, hanem mindenkor az Arg címkét.
- A *-ként* rag funkciójában használt *mint* úgyszintén Case címkét kap, és case viszonnyal kötjük hozzá a *mint* után álló, Arg címkét kapó főnévhez. (pl. *A focicsapat mint [Case] egy megveszekedett oroszlán [Arg] harcolt a győzelemért.*) Ebben az esetben tehát azt feltételezzük, hogy a *mint* szó csak abban különbözik a névutóktól, hogy nem a főnév után, hanem a főnév előtt áll. A főnév nem feltétlenül közvetlenül a *mint* után áll, a főnév előtt lehetnek jelzők és névelő.

### **A Flat címke használata:**

- A Flat címkét elsősorban több tagú tulajdonnevek nem utolsó elemeinél használjuk. Az ilyen esetekben az utolsó elem megkapja az Arg címkét, amely tr vagy lm címkével van hozzáköteve az igéhez. Az utolsó elem előtt álló elemek pedig Flat címkét kapnak, és mindegyiknek a tulajdonnév Arg címkéjű utolsó tagja a fölérendelt tagja. Az Arg címkétől egy flat címkéjű nyíl mutat a Flat címkékre. Arg mellett Poss elem is lehet több tagú tulajdonnév utolsó tagja.
- 1. példa: *Szabó [Flat] Anna [Arg] 1989 óta Budapesten él.* (Ebben az esetben tehát a Szabó Flat címkét kap és egy flat címkéjű nyíllal van hozzáköteve az Arg címkét kapó Anna szóhoz.)

- 2. példa: *András a Dózsa [Flat] György [Flat] utcában [Arg] született.* (Ebben az esetben a *Dózsa* és a *György* szavak is megkapják a *Flat* címkét, és minden kettő fölérendelt tagja az *Arg* címkével ellátott *utcában* szó.)
- Tulajdonnevek mellett egyéb speciális, több elemből álló megszilárdult kifejezések esetében is használható a *Flat* címke.

#### **Az omitted jellemző használata elliptikus szerkezetek annotálásánál:**

- Elliptikus szerkezeteknél, azaz olyan szerkezeteknél, ahol az ige hiányzik, a bővítményeket az elliptikus szerkezetet lehetővé tevő, a szövegben többnyire korábban előforduló ige Core címkéjéhez kötjük. A Core címke ebben az esetben megkapja az omitted tulajdonságot, amely azt jelzi, hogy az ige valójában nem realizálódik a szintaktikai szerkezetben.
- Az elliptikus szerkezetet lehetővé tevő ige minden minimum két Core címkét kap. Az első a saját szintaktikai szerkezetéhez tartozó címke, a második az elliptikus szerkezethez tartozó címke.
- Amennyiben az elliptikus szerkezethez tartoznak olyan bővítmények, amelyek úgyszintén csak a korábbi igei szerkezetben realizálódnak, akkor ezek is kapnak egy plusz Arg címkét az omitted tulajdonsággal. Ezeket az Arg:omitted címkéket a Core:omitted címkéhez kötjük.
- Elliptikus segédigés szerkezetek esetén az Aux:omitted címkét használhatjuk.
- Nemcsak mondaton belül értelmezhető az ellipszis, egy különálló mondat is lehet elliptikus úgy, hogy az előtte lévő mondat igéje kapja a Core:omitted címkét. (pl. *András könyvet olvas. Béla képregényt.*)

#### **Elliptikus szerkezetek és halmozások különbsége:**

- Egy adott szintaktikai funkcióban szereplő bővítmény halmozása nem teszi elliptikussá a mondatot. Például „*A nagy hatalmú király, a bölcs varázsló és a legyőzhetetlen tündér vándorútra indult.*” mondat nem elliptikus, hanem csak három trajector szerepű alanyi bővítménye van az igének. Ugyanígy az „*András evett tortát, hamburgert és pizzát.*” mondat sem elliptikus, hanem csak három landmark szerepű tárgyi bővítménye van az igének.

#### **A Voc címke használata:**

- A megszólítást kifejező vokatívuszok Voc címkét kapnak. Szemben minden más esettel a Voc címkével nem csak egy tokenet láthatunk el. Amennyiben a vokatívuszi szerkezet több tokenből áll (pl. *Ó, drága kismadaram*), akkor az egész szerkezet megkapja a Voc címkét.
- A Voc címkét tr vagy lm viszonttal kötjük az ige Core címkéjéhez, attól függően, hogy a vokatívussal megszólított szereplő az ige trajector vagy landmark szerepű bővítményére utal-e. Amennyiben a szerkezetben megjelenik egy a vokatívussal koreferens bővítmény, akkor a bővítmény Arg címkéjét egy cor viszonttal kötjük a Voc címkéhez. A nyíl az Arg címkéből mutat a Voc címkére. (Ha ez így nem érhető, akkor lásd a cikket, ott példákkal ez be van mutatva.)
- Előfordulhatnak olyan esetek, amikor a vokatívusz nem egy ige kidolgozott vagy implicit bővítményeként megjelenő szereplőre utal. Ilyenkor is felcímkezzük a

vokatívuszt, de azt nem kötjük semmihez. A Voc tehát az egyetlen címke, amit nem muszáj közvetlenül vagy közvetetten hozzákötni egy Core címkéhez.

#### **Névszói állítmányos szerkezetek kopulával vagy kopula nélkül:**

- Névszói állítmányos szerkezeteket nem annotálunk, akkor sem, ha kopulaként szerepel benne létige. A kopulák tehát nem kapnak Core címkét.
- Kopula meghatározásának tesztje: Alakítsuk át a mondatot E/3 kijelentő mód, jelen idejűre. Ha a létige eltűnik, akkor biztos, hogy kopuláról van szó, és ezért nem címkézzük.
- A *van* paradigmatiskus alakjai mellett a *lesz* és a *marad* paradigmatiskus alakjai is viselkedhetnek kopulaként (ezeknél nem működik az átalakítós teszt).

#### **Van + határozói igeneves állítmányok:**

- Van + határozói igeneves állítmányokat tartalmazó szerkezeteket (pl. az ágy *meg van vetve*, a vár össze *volt omolva*) nem annotáljuk. Ezeket tehát hasonlóan kezeljük a kopulás szerkezetekhez.