

COLA annotálási útmutató

Vadász Noémi és Ligeti-Nagy Noémi

2021. november 2.

1. Bevezetés

Az annotálás célja, hogy az angol nyelvű Corpus of Linguistic Acceptability (COLA) mintájára létrehozzunk egy nagyjából 10 000 mondatból álló korpuszt, amelyben az egyes mondatok elfogadhatósága van feltüntetve. Első lépésként megtörtént a mondatok gyűjtése, a jelen annotálási feladat ezeknek a mondatoknak az elfogadhatóságának a megítélése.

2. Az annotálás alapelvei

Az annotálás során egyszerre egy mondatot látunk. A feladat annak eldöntése, hogy az adott mondat elfogadható magyar mondat-e, jó mondatnak tartjuk-e, helyesen hangzik-e, elképzelhetőnek tartjuk, hogy mondunk ilyen mondatot. Az 1. példában látunk egy mondatot. Ez a mondat egy elfogadható, grammatikus mondat – függetlenül attól, hogy nehezen tudunk elképzelni piros elefántot, főleg olyat, amelyik ugrál is. Fontos tehát, hogy az annotálás során nem nézzük a mondatok valóságosságát.

(1) A piros elefánt ugrál a kertben.

Ezzel szemben a 2. példában látható mondat, bár sejtjük, hogy miről lehet szó, nem elfogadható. Ha *postást* lenne a vége, elfogadnánk, így azonban nem tudjuk.

(2) A kutya megharapta a postás.

3. A felület

Az annotálást az AT1 eszköz segítségével végezzük¹. A nyitóoldalon bejelentkezés után a `cola` feladat kiválasztásával kezdheted a munkát.

¹<https://at1.nlp.nyttud.hu/login>

Az annotálófelületen a felső sávban látható a bejelentkezett felhasználó neve, valamint a **Kijelentkezés** gomb. A felületen emellett látható az annotálandó mondat, valamint két gomb (**JÓ** és **ROSSZ**). A megfelelő gomb kiválasztása után megjelenik egy visszajelző üzenet (*A mondat helyes.* ill. *A mondat nem helyes.*). Ezen a ponton a mondatra vonatkozó döntés még bárhányszor megváltoztatható (a **JÓ** vagy **ROSSZ** gombok újbóli megnyomásával), a végleges döntés a **Következő mondat** gomb megnyomása után kerül rögzítésre.

A korábban mentésre került válaszok már nem módosíthatók, nincs mód a korábban annotált mondatokhoz történő visszatérésre, ezért kérek, minden esetben jól gondold meg a választ. A böngésző **Újratöltés**, **Vissza** ill. **Előre** gombjait kérek, ne használd.

Az egyes mondatokhoz rendelt válasz tehát mindig a következő mondatra történő továbblépéskor rögzül, így a munkát bármikor félbe lehet hagyni. Ennek megfelelő módja a böngészőlap bezárása. Újbóli bejelentkezéskor az utoljára látott mondat jelenik meg, így ugyanonnan lehet folytatni a munkát.

Tájékoztató üzenet jelzi, ha elfogyott az annotálandó mondat (*Elfogyott a mondat!*).

Ha több különböző feladatban is részt veszel és át szeretnél térni másik feladatra, akkor se a böngésző **Vissza** gombjával próbálkozz, hanem lépj ki a **Kilépés** gombbal, majd jelentkezz be újra, hogy ki tudd választani a megfelelő feladatot a legördülő menüből.

Helyesírási hibák, elütések jelzése a megosztott mappa "helyesírasos" nevű dokumentumában történjen, a mondat id-jának és a mondat szövegének feltüntetésével.