

# SST annotálási útmutató

## Fluencia-ellenőrzés

Vadász Noémi és Ligeti-Nagy Noémi

2021. november 19.

### 1. Bevezetés

Az annotálási feladatok célja a Stanford Sentiment Treebank (SST) magyar nyelvre történő átültetése. Az angol korpusz kb. 11000, az 1. példában láthatóhoz hasonló mondatot tartalmaz. Minden mondathoz tartozik egy szentimentcímke: pozitív vagy negatív.

- (1) With a cast that includes some of the top actors working in independent film, *Lovely & Amazing* involves us because it is so incisive, so bleakly amusing about how we go about our lives.

Az annotálási feladat második lépésében a feladat a gépi fordító által lefordított, annotátorok által javított fordítás eredményeképp előállt mondatok szép, fluens magyar mondattá alakítása.

### 2. Az annotálás alapelvei

Ebben a lépésben csak a magyar szöveget látjuk, az eredeti angolt nem. A feladat az, hogy a szöveg szép, folyékony magyar szöveggé kerüljön az annotálás következő fázisába. Két lehetőségünk van: i) elfogadjuk a szöveget úgy, ahogy van, mert megfelelő minőségű; ii) egy vagy több helyen javítjuk, hogy gördülékenyebb mondatot kapjunk.

- (2) Merész, megbabonázó és rendkívül nehéz elfelejteni.

A 2. példában a *megbabonázó* kissé magyartalan, ezért így javíthatjuk: *Merész, megbabonáz, és rendkívül nehéz feledni.*

### 3. A felület

Az annotálást az AT1 eszköz segítségével végezzük<sup>1</sup>. A nyitóoldalon bejelentkezés után az `sst_flu` feladat kiválasztásával kezdheted a munkát.

<sup>1</sup><https://at1.nlp.nytud.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 JÓ gomb megnyomása esetén megjelenik egy visszajelző üzenet (*A mondat helyes.*). A ROSSZ gomb megnyomása esetén megjelenik egy szövegdoboz, amibe a helyes mondatot kell írni. Az egész mondatot le kell gépelni (vagy másolni és beilleszteni) még akkor is, ha csak egyetlen szó volt hibás az eredeti mondatban. A javítást az **Elküld** gombbal lehet befejezni. Ekkor megjelenik egy visszajelző üzenet, ami a javított mondatot mutatja, itt lehet ellenőrizni, hogy a kézzel beírt javítás megfelelő-e. Ezen a ponton a mondatra vonatkozó döntés még bárhányszor megváltoztatható, a javítás újra megadható (a JÓ vagy ROSSZ gombok újbóli megnyomásával, a javított mondat begépelésével stb.), de arra figyelj, hogy a ROSSZ gomb megnyomásával a szövegdobozba mindig az eredeti szöveg másolódik be, felülírva az esetleges javításodat. 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.