

SST annotálási útmutató

Fordításellenőrzés

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 annotálás első lépéseként a gépi fordító kimenetét kell ellenőrizni és javítani.

2. Az annotálás alapelvei

Az annotálás során egyszerre egy egységet látunk: az eredeti angol mondatot, és a gépi fordító által előállított magyart (2).

- (2) 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.
Egy szereplőgárda, amely magában foglalja néhány, a felső színészek dolgozó független film, *Lovely & Amazing* jár minket, mert annyira metsző, így siváran szórakoztató arról, hogyan megy az életünkben.

Két lehetőségünk van: i) elfogadjuk a fordítást, ii) nem fogadjuk el a fordítást, és új fordítási javaslatot írunk. A 2. példában értelemszerűen javítanunk kell a fordítást.

Néhány alapelv:

- A cél, hogy angol szó ne maradjon a szövegben: az angol filmcímeket fordítani kell. Az nem gond, ha nem is úgy volt hivatalosan fordítva az

adott cím, csak magyarul legyen. A nyelvezet a lényeg, nem az, hogy a filmadatbázisnak megfeleljünk. :) A *Chicago* jellegű, városnevet, személynevet tartalmazó címek ilyen részeit természetesen nem fordítjuk.

- A szokatlan magázást nyugodtan lehet tegezésre cserélni.
 - Kontextus nincs, ez tényleg nehéz, de nem baj, ha az eredeti kontextusba a mi mondatunk már nem lehetne visszailleszthető: a cél, hogy egy kerek magyar mondatot kapjunk.
 - A nyelvtanilag helyes, de "magyartalan" mondatokat javítani kell, hogy minél magyarosabb, szebb mondatok legyenek. Ha valamilyen angol kifejezésre van szebb magyar, mint amit a gépi fordító talált (mert például az csak egy tükörfordítás), akkor javítani kell.
 - Lehet bátor egy mondat átszerkesztése, főleg a hosszabb, bonyolultabb mondatok esetében, ahol egészen magyartalan szórend és kifacsart szerkezetek jöhetnek létre a gépi fordítással. Bátorság a magyartalanság fölött!
 - Fontos, hogy egész mondatokat szeretnénk kapni. Ha az angol mondat feltűnően csonka, hiányzik belőle egy nagy darab, félbeszakad valahol stb., akkor keressétek meg a google segítségével az eredetit, és ha megvan, az alapján fordítsatok. Ha nincs meg az eredeti, akkor jelezzétek a faq dok-siban. Az nem probléma, ha egy mondat kisbetűvel kezdődik, vagy nincs pont a végén. A 3. példában látható eseteket szükséges kiegészíteni.
- (3) Sitting in the third row of the IMAX cinema at Sydney's Darling Harbour, but I sometimes felt as though I was in the tiny two seater plane that carried the giant camera around Australia, sweeping and gliding, banking and hovering over some of the most not

3. A felület

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

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 fordítás helyes..*). A ROSSZ gomb megnyomása esetén megjelenik egy szövegdoboz, amibe a helyes fordítást kell írni. Az egész mondatot le kell gépelni (vagy másolni és beilleszteni) még akkor is, ha csak egyetlen szó volt hibás a fordításban. A javítást az **Elküld** gombbal lehet befejezni. Ekkor megjelenik egy visszajelző üzenet, ami a javított fordítást mutatja, itt lehet ellenőrizni, hogy a kézzel beírt javítás megfelelő-e. Ezen a ponton

¹<https://at1.nlp.nyttud.hu/login>

a mondatra vonatkozó döntés még bárhányszor megváltoztatható, a javított fordítás újra megadható (a JÓ vagy ROSSZ gombok újbóli megnyomásával, a helyes fordítás begépelésével stb.), de arra figyelj, hogy a ROSSZ gomb megnyomásával a szövegdobozba mindig az eredeti fordítás 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 annotálódó 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.