

COPA annotálási útmutató

Fluencia-ellenőrzés

Vadász Noémi és Ligeti-Nagy Noémi

2021. november 16.

1. Bevezetés

Az annotálási feladatok célja az angol Choice of Plausible Alternatives (COPA) korpusz magyar nyelvre történő átültetése. Az angol korpusz 1000, az 1. példában láthatóhoz hasonló egységet tartalmaz. Minden tétel egy premisszából, egy kérdésből, és két lehetséges alternatívából áll. A két alternatíva közül kell kiválasztani a megfelelő okot vagy hatást, attól függően, hogy a kérdés mit tartalmaz.

- (1) Premissza: *I knocked on my neighbour's door.*
Kérdés: *What happened as a result?*
Alternatíva 1: *My neighbour invited me in.*
Alternatíva 2: *My neighbour left his house.*

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 mondatokat kapjunk.

- (2) Premissza: *A férfi eltörte egy lábujját.*
Kérdés: *Mi volt ennek az oka?*
Alternatíva 1: *Lyuk lett a zoknijában.*
Alternatíva 2: *A lábára ejtett egy kalapácsot.*

Itt lehetséges javítás az Alternatíva 1-ben, hogy *Kilyukadt a zoknija.* A javítás során már az egész egység szerkezetét figyelembe kell vennünk, mert az,

hogy az adott premissza okát vagy hatását keressük, segíthet a két alternatíva megformálásában. Fontos, hogy nem kell ragaszkodnunk mindenáron az eredeti szavakhoz; ha kicsit átfogalmazott mondatokkal szebb és értelmesebb eredményt kapunk, akkor nyugodtan módosíthatunk bármin.

Természetesen nem kell mindáron javítani a mondatokat, ha az elolvasás során úgy érezzük, hogy teljesen rendben vannak, szép, jólformázott magyar mondatok, akkor maradhatnak eredeti formájukban.

Ha javításra van szükség, akkor azt az adott cella tartalmát felülírva tegyék meg.