Annotation Guidelines

Tore Buscher, Laura Fredholm, Lisa Stuch

March 27th, 2023

1 Introduction

These guidelines are used in the Bachelor Thesis' of the authors (Tore Buscher, Laura Fredholm, Lisa Stuch). They are the basis for the annotation for the test sets. In them, it is outlined which words to mark as "Animated" and which to **not** mark as "Animated". Any data being labeled is first processed by the noun chunker included in the spaCy library.

Any examples and further definitions are in German since the corpora to be labeled are in German.

2 General conventions

2.1 What to mark as "Animate"

- Any span marked as animate must exactly match the span marked as one nominal phrase by the chunker.

- Any noun chunk containing a noun in one of the following categories is always to be marked as "Animate"
    - all humans, including fictional characters
    - all animals
    - all plants
    - all living organisms, including bacteria
    - all political organizations

- Any noun chunk described by an adjective, or any grammatical or semantic subject of a verb, that
    - hints to cognition (feel, hear, see, think, ...)
    - signals social interaction (communicate, tell, hug, ...)
    - suggests animacy (alive, vital, kindly, ...)

    in its context is to be marked as "Animate".

2.2 What not to annotate at all

- indefinite pronouns such as "alle / anderen / solche" (except for "man ", see Section 3)

- gendered pronouns such as "er / sie / es" (he, she, it)

- possessive pronouns such as "meiner / deiner / unser / mich / uns"

3 Further definitions / exceptions

- indefinite pronouns "man / jemand" are always animated

- nominative personal pronouns "ich / wir"(I, us) are always animated

- metonyms are to be marked depending on their context, for instance, countries are animated if they refer to the acting people and are inanimate if they refer to geographic features

- robots are not animated

- viruses (and especially the covid virus) are not animated

4 Examples

The following table shows examples of texts and the associated decision about animacy and the reasons for this decision.

Example [Translation]	Decision about Animacy	Reasoning
"China hat ... beschlossen." ["China has decided."]	"China" is animate	See rule 3 in Section 3: "China" refers to the Chinese government, not a geographic location, and thus is animate.
"Europa ist schön." ["Europe is beautiful."]	"Europa" ["Europe"] is not animate	"Europa" ["Europe"] refers to a continent, which is a geographic feature, therefore it is **not** animate.
"Trag mal alle Tische hier rüber. Die anderen Tische kommen da hin." ["Carry the table over here. The others go there."]	Nothing is animate	Tables are inanimate objects, "die anderen" ["the others"] is not marked according to Section 2.2.
"Das Glas ist zerbrechlich, es geht bestimmt kaputt." ["The glas is fragile, it will likely break."]	Nothing is animate	"Das Glas" ["the glas"] is an inanimate object, "es" ["it"] is not marked according to Section 2.2.
"Die CDU wird morgen gegen diesen Gesetzesentwurf abstimmen." ["CDU will vote against this bill tomorrow."]	"CDU" is animate	Referring to Section 2.1, all political organizations are animated.
"Corona stellt uns vor große Herausforderungen." ["Corona presents us with major challenges."]	Nothing is animate	"Corona" ["Covid"] refers to a virus and therefore is not animate as stated in Section 3
"Seit dem Jahr 2007 leistet die United Nations Hybrid Operation in Darfur, UNAMID, einen wichtigen Beitrag, um ..." ["Since 2007, the United Nations Hybrid Operation in Darfur, UNAMID, has been making a significant contribution to ..."]	"United Nations Hybrid Operation" und "UNAMID" are animate	Both words can refer to an organization since the verb "einen Beitrag leisten" ["to make a contribution"] indicates cognition.