

Дегуманізація українців у російському сегменті Телеграма: інструкція з анотування даних в форматі CoNLL04

Що анотуємо

Кожен анотаційний приклад у цьому проєкті — це речення (або неструктурований короткий фрагмент тексту), написане російською мовою. Ці речення отримані з повідомлень популярних російських Телеграм каналів. Повідомлення були зібрані у 2022 році та покривають часовий проміжок з 2015 до 2022 року.

Важливо! Більшість текстів містять образливі сенси. Будь ласка, не продовжуйте виконання анотаційного завдання, якщо ви відчуваєте, що ці тексти здатні вас травмувати.

Де анотуємо

Для анотування текстів ми обрали інструмент [Labelbox](#). Щоб почати анотувати дані, натисніть *Start Labelling*.

Кожен приклад для анотування має такі поля:

- сам текст повідомлення (посередині);
- питання до тексту, на які потрібно відповісти (колонка зліва);
- кнопка *Submit* для збереження ваших відповідей і переходу до наступного прикладу (справа вгорі);
- кнопка *Skip*, яка дозволяє пропустити приклад, якщо текст вам узагалі незрозумілий (справа вгорі);
- кнопки перегляду попередніх/наступних повідомлень, які ви вже проанотували (над текстом).

Навіщо анотуємо

Загальна мета нашого дослідження — створити набір даних для тренування системи, здатної виявити ознаки дегуманізації в тексті в розрізі різних типів дегуманізації.

Головне завдання Частини III (цього проєкту) — навчити систему виявляти дегуманізуючі да дегуманізовані сутності в реченнях російською мовою.

Загальні рекомендації

Будь ласка, читайте речення повністю. Кожне додаткове слово може містити контекст, який може вплинути на відповідь.

Якщо ви бачите в тексті посилання, будь ласка, ігноруйте їх — зміст матеріалів за посиланням не має впливати на ваші відповіді.

Ви можете повернутись до попереднього питання за допомогою навігації стрілочками і виправити свою відповідь, якщо вважаєте за потрібне. Після виправлення натисніть Save і стрілочкою вперед переходьте до наступного питання.

Усі питання сформульовані відносно образу, який формує вжите порівняння, метафора або опис.

Формування образу може здійснюватись за допомогою:

- відвертого прирівняння (наприклад, “*украонацики*” — демонізація через прирівняння до втілення абсолютного зла в російській секулярній секті Победобесія — нацистів)
- словотвору (наприклад, “*укропитеки*” — прирівняння до первісних людей, “*майданутые*” натякає на людину з фізичною або психологічною вадою, “*чубаноиды*” натякає на тільки частково подібних до людей гуманоїдів),
- через лексичні підказки (наприклад, вживання слова “*утилизировать*” щодо українців, яке в мовленні використовується тільки до неістот, або “*воевать до последнего украинца*” — вказання на українців як видатковий матеріал)
- описово, через зміст самого повідомлення (“*украинцы — это воплощение зла*” формує образ диявола, “*украинские хозяева заявили...*” формує образ раба, чиєїсь власності)
- синтаксичних або пунктуаційних підказок щодо відсутності в українців певних унікально людських рис, наприклад, здатності мати свою державність або право на окрему національну ідентичність через використання лапок у виразі “*украинское государство*, “*украинский язык*, прирівнювання до помилки (*страна 404*).
- різними виразними засобами через атрибути українськості — підгрупи українців, держави та її інститутів (наприклад, вживання щодо України терміна “Укрорейх” прирівнює українців до нацистів, вживання щодо українських дітей терміна “укроюгенд” досягає формування аналогічного образу, використання “*кукарехрюк*” як опису мови українців вказує на тваринне походження самих українців, в “*знать хохлов на убой*” українці прирівнюються до свійських тварин).

Будь ласка, зверніть увагу, що дегуманізація українців може здійснюватись за допомогою прирівняння до інших дегуманізованих груп — людей із психологічними або фізичними вадами, представників маргінальних груп (наркоманів, безхатьок), рабів (кріпаків, холопів, маріонеток) та загалом груп людей, не спроможних керувати своєю поведінкою, думати та відчувати як доросла людина.

Окремо варто звернути увагу на демонізацію українців — прирівнювання до синонімів абсолютного зла або людей-прибічників абсолютного зла, які виключаються з кола людей через їх екстремальну духовну та моральну неповноцінність та небезпечність (наприклад, позначення українців як *сатана, сатанисти, шайтаны, фашисти, нацисти, абсолютное зло*).

Для російського суспільства культ Великої Перемоги (або “победобесіє”, як часто його називають) — одна з головних частин ідеології путінського режиму, основа російського державного міфу. В цьому міфі нацисти відіграють роль абсолютного зла, протиставленого абсолютному добру (уособленому в ідеї «російського світу»). **Тому необгрунтоване прирівнювання до нацистів, неонацистів, фашистів має таку ж природу, як і прирівнювання до демонів, і є дегуманізацією.** Зверніть увагу — називання українців бандерівцями не є дегуманізацією, тому що тут йдеться про реальну історичну особу і реальних (хоча і фактично неправильно описуваних) послідовників цієї особи.

[За цим посиланням](#) можна ознайомитись із нашим оглядом літератури, присвяченої різним фреймворкам визначення дегуманізації. Це не є обов’язковим для успішної анотації.

Як розмічати дані

Ukrainian

Сутність дегуманізованої групи розмічається як Ukrainian.

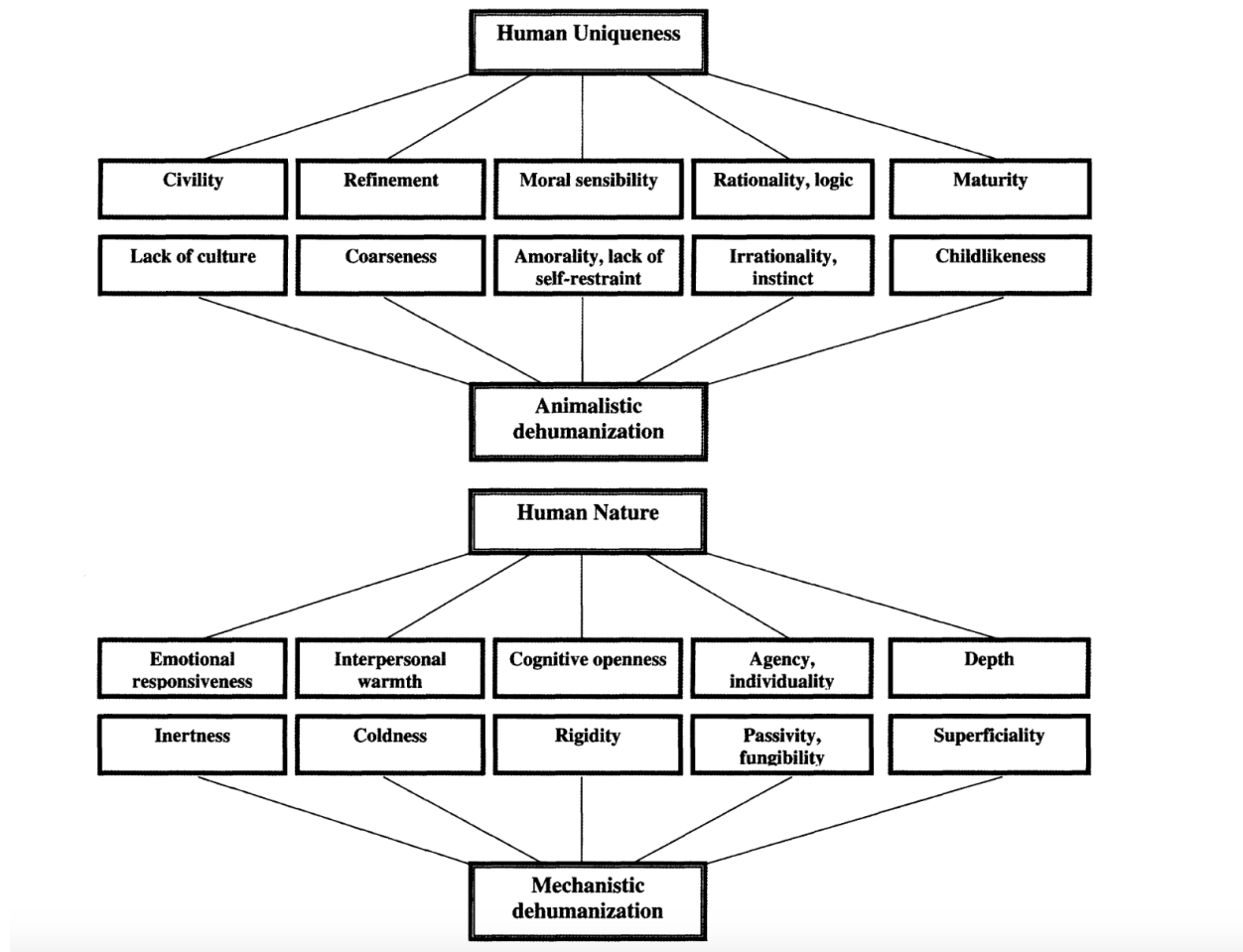
Перше слово в словосполученні позначається як B-Ukrainian, решта слів в цьому словосполученні позначається як I-Ukrainian.

Цим тегом позначаються всі сутності, які називають українців (в тому числі певну підгрупу, наприклад “войска противника”), Україну (в тому числі алегорично чи евфемістично) або є описовою характеристикою приналежності до українців.

Дегуманізуючі послідовності

Сутності дегуманізуючих слів або словосполучень розподіляються відповідно до типу сигналу дегуманізації і має той же принцип позначки відповідно до належності до початку послідовності в межах сутності.

Будь ласка, розмічайте словосполучення, в якому міститься сигнал дегуманізації, в тих межах, в яких воно може бути однозначно інтерпретоване як дегуманізуюче – без надання додаткового контекста з цього ж речення. Інтервали, які ви розмічаєте, неперервні, тобто до прикладу якщо потрібний для однозначної інтерпретації контекст міститься в кінці речення, потрібно виділити всю послідовність з першого слова словосполучення, в якому міститься сигнал, до останнього слова в реченні.



В нашом дослідженні до цієї систематизації інтегруються фрагменти інших популярних сучасних фреймворків. Так бажання, можливість мати емоційний досвід, теплота, творчість це фактори більшого NH.

Факторами більшого UH є агентність, компетенція, здатність до раціонального мислення, переживання вторинних (унікально людських) емоцій, відкритість досвіду, свідомість.

Типи дегуманізуючих послідовностей:

LOW_UH_LOW_NH - дегуманізація за принципом позбавлення частини чи всіх рис, які відрізняють людину від інших живих істот, і також від неживих істот.

HIGH_UH_LOW_NH - дегуманізація за принципом позбавлення частини чи всіх рис, які відрізняють людину від неживих істот, без зміни або з підсиленням рис, які відрізняють людину від інших живих істот.

LOW_UH_HIGH_NH - дегуманізація за принципом позбавлення частини чи всіх рис, які відрізняють людину від інших живих істот, без зміни або з підсиленням рис, які відрізняють людину від неживих істот.

HIGH_UH_HIGH_NH - супергуманізація за принципом підсилення частини чи всіх рис, які відрізняють людину від інших живих істот, і також від неживих істот.

Така послідовність може містити від 1 слова (з або без пунктуацією) до повного співпадіння з реченням.

Повний перетин між сутністю Ukrainian та дегуманізуючою послідовністю також можливий.