

Manual Evaluation of Gender Neutral Translations

The hereby presented manual evaluation task concerns determining whether the automatic translation of mGente sentences (column L “translation” in the picture below) is NEUTRAL or GENDERED.

D	I	L	P	Q	R
SRC	REF-G_ann	translation	gold_neutrality_label	literal	comments
As a Frenchwoman, I am only glad that I have no part in the dismantling of that reassuring and life-affirming network of 17 000 post offices and sub-post offices, knitted together over 550 000 km2 of our territory and embodying public service, in both suburban and rural areas.	Como <G>francesa</G> , sólo puedo felicitarme por no asistir así al desmantelamiento de esta red tranquilizadora y vivificante de las 17.000 oficinas postales y agentes de correos , tejida de forma unida en los 550.000 km2 de nuestro territorio y que lleva la presencia del servicio público , tanto en las zonas suburbanas como en las rurales .	Como francesa, solo me alegro de no tener parte en el desmantelamiento de esa red reconfortante y llena de vida de 17 000 oficinas de correos y suboficinas, tejida a lo largo de 550 000 km2 de nuestro territorio y que encarna el servicio público, tanto en zonas suburbanas como rurales.			

For each language pair (en-de, en-es, en-el, en-it), you are assigned 50 sentences from 5 different models, for **a total of 250 sentences**.

Guidelines

Given a source English sentence from mGente, evaluate if its automatic translation contains gendered forms (either masculine or feminine) or if it is a neutral translation.

For each sentence, indicate:

- **Gold_neutrality_label** (COLUMN F): pick one of the following labels:
 - **NEUTRAL**: The output sentence does not contain any gendered expressions to refer to human beings.
 - **GENDERED**: The output sentence contains one or more gendered expressions to refer to human beings. Remember that one single gendered form—whether masculine or feminine, and even if other forms are neutralized—is sufficient to deem the sentence as gendered.

👏 If you have doubts about what constitutes an acceptable Gender-Neutral Translation (GNT) in our task, look at the GNT guidelines used to create mGeNTE.

👏 Though unusual, it might be possible that the automatic translation is gibberish and thus cannot be evaluated. If so, just indicate **NA** in column F.

- **literal** (COLUMN G): This label only applies if the translation is NEUTRAL.
 - This additional layer of information serves to get a sense of whether the model might have generated a neutral translation by “accident/chance” or if it seems to be able to solve the GNT task

- Indicate **LITERAL** if the NEUTRAL output translation does not seem a neutralization attempt, but rather seems an incidental translation that, by being quite literal to the source, avoids any gender marking incidentally.
- **Comment** (COLUMN H): Use this space to add optional comments that you deem relevant. For instance, if the output translation contains non-binary neomorphemes, etc...