Are all Spanish doctors male?
Evaluating Gender Bias in German Machine Translation

Michelle Kappl & Philine Kowol
Technische Universität Berlin
michelle.kappl@tu-berlin.de

Communication in multiple languages has become commonplace in both personal and professional settings and Machine Translation (MT) models have emerged as valuable tools to bridge language gaps. However, these translations are not always accurate and gender bias within MT is evident, as an exemplary Translation (1) by Google Translate shows:

(1)   Die Ärztin$_f$ untersuchte den Friseur$_m$, weil sie besorgt war.
              ↓   ↓
       El médico$_m$ examinó a la peluquera$_f$ porque estaba preocupada.

Gender-biased MTs solidify societal assumptions about the abilities and expectations of different genders (Vervecken & Hannover 2015). Therefore, it is crucial to deepen our understanding of gender bias in MT. Prior research on evaluating gender bias has primarily focused on English MT models, with Stanovsky et al. (2019) conducting the first large-scale evaluation on this topic. We introduce a German gender bias evaluation test set, WinoMTDE, which extends the state-of-the-art evaluation method based on coreference resolution developed by Stanovsky et al. (2019) to German, a language exhibiting gender distinctions in its grammatical structure. The dataset consists of German sentences such as (1), casting the subjects into different occupations. These professions are annotated as either pro- or anti-stereotypical using German Department of Labor statistics.

Using WinoMTDE, five German MT models were analyzed specifically focusing on gender bias and occupational stereotyping in respective translations to seven languages displaying grammatical gender. The findings reveal gender bias in German MT systems. All evaluated models perform better on male instances, and a bias towards stereotypical occupations is prevalent in most systems. Moreover, findings suggest that using Hybrid MT models could improve the quality of translations to the Romance language family regarding gender bias to some extent. All evaluation results and the used code are available at https://github.com/michellekappl/mt_gender_german.

References:
• Stanovsky, G., Smith, N. A., & Zettlemoyer, L. (2019). Evaluating Gender Bias in Machine Translation. • Vervecken, D., & Hannover, B. (2015). Yes I Can! Effects of Gender Fair Job Descriptions on Children's Perceptions of Job Status, Job Difficulty, and Vocational Self-Efficacy. In *Social Psychology*, 46, 76–92.