A Appendices

A.1 Impact Statement

This work identifies a weakness of NMT models where they appear to ignore contextual evidence regarding the gender of an occupation noun and apply an incorrect gender marker. It is difficult to measure the adverse effects of biases in NMT, but errors like the ones we highlight reduce trust in the NMT system.

Intended use: We hope that this type of error is further studied by NMT researchers leading to a solution. Furthermore, we expect the speed-optimization aspect of our work provides NMT engineers with an extra point of consideration, as we show gender-bias (errors in our dataset) increases rapidly compared to metrics like BLEU on standard datasets. In this work, we limit ourselves to viewing gender in the linguistic sense. SimpleGEN is not meant to be a replacement for traditional MT evaluation.

Risks: We recognize that socially, gendered language evolves (e.g. in English, "actress" is rarely used anymore). To the best of our knowledge, we selected occupations that are typically gendered (in Spanish and German) at present. Furthermore, we only regard the gender binary as a linguistic construct. It would be incorrect to use this work in the context of gender identity or gender expression etc.

Dataset: The dataset is "synthetic" in that it has been constructed using templates. We did not use crowd-sourcing or private data.

A.2 Full Template and Terms

Keywords	Values
f-n	female, women
m-n	male, men
f-n-pl	women, ladies, females, gals
m-n-pl	men, guys, males, fellows
f-n-sg	gal, woman, lady
m-n-sg	man, guy, fellow
f-obj-prn	her
m-obj-prn	him
f-pos-prn	her
m-pos-prn	his
f-obj-pos-prn	her
m-obj-pos-prn	his
f-sbj-prn	she
m-sbj-prn	he
f-rel	wife, mother, sister, girlfriend
m-rel	husband, father, brother, boyfriend

Table A5: Keywords and the values they can take.

Occupation Keywords	Values
f-occ-sg	clerk, designer, hairdresser, housekeeper, nanny, nurse, secretary
m-occ-sg	director, engineer, truck driver, farmer, laborer, mechanic, physician, president, plumber, carpenter, groundskeeper
f-occ-pl	clerks, designers, hairdressers, housekeepers, nannies, nurses, secretaries
m-occ-pl	directors, engineers, truck drivers, farmers, laborers, mechanics, physicians, presidents, plumbers, carpenters, groundskeepers
f-occ-sg-C	clerk, designer, hairdresser, housekeeper, nanny, nurse, secretary
m-occ-sg-C	director, truck driver, farmer, laborer, mechanic, physician, president, plumber, carpenter, groundskeeper
f-occ-pl-C	clerks, designers, hairdressers, housekeepers, nannies, nurses, secretaries
m-occ-pl-C	directors, truck drivers, farmers, laborers, mechanics, physicians, presidents, plumbers, carpenters, groundskeepers
f-occ-sg-V	
m-occ-sg-V	engineer,
f-occ-pl-V	
m-occ-pl-V	engineers,

Table A6: Occupation keywords and the values they can take. The prefix "m-" and "f-" indicate that according to the U.S Department of Labor these occupations have a higher percentage of male and female works, respectively.

Table A7 shows the template we use to generate our source sentences in SimpleGEN. We can generate sentences in one of the four sub-categories (MOMC, MOFC, FOFC, FOMC) by setting occupation keywords with the prefix m- or f- from our terminology set Table A6). For example, to generate MOFC sentences, we set occupation-keywords with prefix m- and non-occupation keywords with prefix f-.

A.3 Breakdown of scatter plots

Figures A2a and A2b further divides pro-stereotypical into male-occupations in male contexts (MoMc) and female-occupations in female context (FoFc), and anti-stereotypical into male-occupations in female contexts (MoFc) and female-occupations in male contexts (FoMc).

A.4 Evaluation Dictionary

Table A8 shows the dictionary we use for evaluation.

My {f/m}-rel is a {f/m}-occ-sg-C .
My {f/m}-rel is an {f/m}-occ-sg-V .
His {f/m}-rel is a {f/m}-occ-sg-C .
His {f/m}-rel is an {f/m}-occ-sg-V .
Alex's {f/m}-rel became a {f/m}-occ-sg-C .
Alex's {f/m}-rel became an {f/m}-occ-sg-V .
My {f/m}-rel enjoys {f/m}-pos-prn work a {f/m}-occ-sg-C .
My {f/m}-rel enjoys {f/m}-pos-prn work an {f/m}-occ-sg-V .
Those {f/m}-n-pl are {f/m}-occ-pl .
These {f/m}-n-pl work hard {f/m}-occ-pl .
I talked to {f/m}-obj-prn , {f/m}-sbj-prn is a {f/m}-occ-sg-C .
I talked to {f/m}-obj-prn , {f/m}-sbj-prn is an {f/m}-occ-sg-V .
{f/m}-sbj-prn works a {f/m}-occ-sg-C .
{f/m}-sbj-prn works an {f/m}-occ-sg-V .
I worked with {f/m}-obj-prn , {f/m}-sbj-prn is a very qualified {f/m}-occ-sg .
{f/m}-n {f/m}-occ-pl are rare .
{f/m}-n {f/m}-occ-pl are not predisposed for self-promotion .
{f/m}-n {f/m}-occ-pl performed well .
{f/m}-n {f/m}-occ-pl earn less than their counterparts .
{f/m}-sbj-prn enjoys {f/m}-pos-prn work a {f/m}-occ-sg-C .
{f/m}-sbj-prn enjoys {f/m}-pos-prn work an {f/m}-occ-sg-V .
Though {f/m}-sbj-prn argued with {f/m}-pos-prn colleagues, the {f/m}-occ-sg was always respectful .
People laughed at the {f/m}-occ-sg behind {f/m}-obj-pos-prn back .
The {f/m}-occ-sg screamed when {f/m}-sbj-prn saw the mouse .
That {f/m}-occ-sg chooses {f/m}-obj-prn words very carefully .
That {f/m}-occ-sg is a funny {f/m}-n-sg !
Those {f/m}-occ-pl are talented {f/m}-n-pl !
Their {f/m}-occ-sg is a smart {f/m}-n-sg !
The {f/m}-occ-sg is a very dynamic {f/m}-n-sg !
The {f/m}-occ-sg is a {f/m}-n-sg with a lot of experience.
This {f/m}-occ-sg happens to be my {f/m}-rel .
The {f/m}-occ-sg works hard and {f/m}-sbj-prn excels !
The {f/m}-occ-sg drove fast and {f/m}-sbj-prn was on time .
As a {f/m}-occ-sg-C , {f/m}-sbj-prn is very competent .
As an {f/m}-occ-sg-V , {f/m}-sbj-prn is very competent .
When the {f/m}-occ-sg arrived, everyone applauded {f/m}-obj-prn .
The {f/m}-occ-pl with the highest success rate per the polls, were {f/m}-n .
That {f/m}-occ-sg is also my {f/m}-rel .
The {f/m}-occ-sg moved here from Ethiopia where {f/m}-sbj-prn is from .
The {f/m}-occ-sg arrives early for {f/m}-pos-prn shift .
That {f/m}-occ-sg is hella strict, watch out for {f/m}-obj-prn .
That {f/m}-occ-sg retired early, good for {f/m}-obj-prn .

Table A7: Our template set used to generate English source sentences.

English	Spanish-Masc	Spanish-Fem	German-Masc	German-Fem
clerk	funcionario|empleado administrativo	funcionaria|empleada administrativa	Mitarbeiter	Mitarbeiterin
designer	diseñador	diseñadora	Designer	Designerin
hairdresser	peluquero	peluquera	Friseur	Friseurin
housekeeper	amo de llaves|amo de casa	ama de llaves|ama de casa	Haushälter	Haushälterin
nanny	niñero	niñera	Kinderbetreuer	Kindermädchen
nurse	enfermero	enfermera	Krankenpfleger	Krankenschwester
secretary	secretario	secretaria	Sekretär	Sekretärin
director	director	directora	Direktor	Direktorin
engineer	ingeniero	ingeniera	Ingenieur	Ingenieurin
truck driver	camionero	camionera	Lastwagenfahrer	Lastwagenfahrerin
farmer	granjero|campesino	granjera|campesina	Bauer	Bäuerin
laborer	obrero|trabajador	obrera|trabajadora	Arbeiter	Arbeiterin
mechanic	mecánico	mecánica	Mechaniker	Mechanikerin
physician	médico	médica	Arzt	Ärztin
president	presidente	presidenta	Präsident	Präsidentin
plumber	plomero|fontanero	plomera|fontanera	Klempner	Klempnerin
carpenter	carpintero	carpintera	Tischler|Zimmermann	Tischlerin
groundskeeper	jardinero|guardián	jardinera|guardiana	Gärtner	Gärtnerin
clerks	funcionarios	funcionarias	MItarbeiter	Mitarbeiterinnen
designers	diseñadores	diseñadoras	Designer	Designerinnen
hairdressers	peluqueros	peluqueras	Friseure	Friseurinnen
housekeepers	amos de llaves|amos de casa	amas de llaves|amas de casa	Haushälter	Haushälterinnen
nannies	niñeros	niñeras	Kinderbetreuer	Kindermädchen
nurses	enfermeros	enfermeras	Krankenpfleger	Krankenschwestern
secretaries	secretarios	secretarias	Sekretäre	Sekretärinnen
directors	directores	directoras	Direktoren	Direktorinnen
engineers	ingenieros	ingenieras	Ingenieuren	Ingenieurinnin
truck drivers	camioneros	camioneras	Lastwagenfahrerin	Lastwagenfahrerinnen
farmers	granjeros	granjeras	Bauern	Bäuerinnen
laborers	obreros	obreras	Arbeiter	Arbeiterinnen
mechanics	mecánicas	mecánicos	Mechaniker	Mechanikerinnen
physicians	médico	médicas	Ärzte	Ärztinnen
presidents	presidentes	presidentas	Präsidenten	Präsidentinnen
plumbers	plomeros	plomeras	Klempner	Klempnerinnen
carpenters	carpinteros	carpinteras	Tischler	Tischlerinnen
groundskeepers	jardineros|guardianes	jardineras|guardianas	Gärtner	Gärtnerinnen

Table A8: Our dictionary of occupations. Entries with the "|" symbol indicate that we accept either of the references as correct.
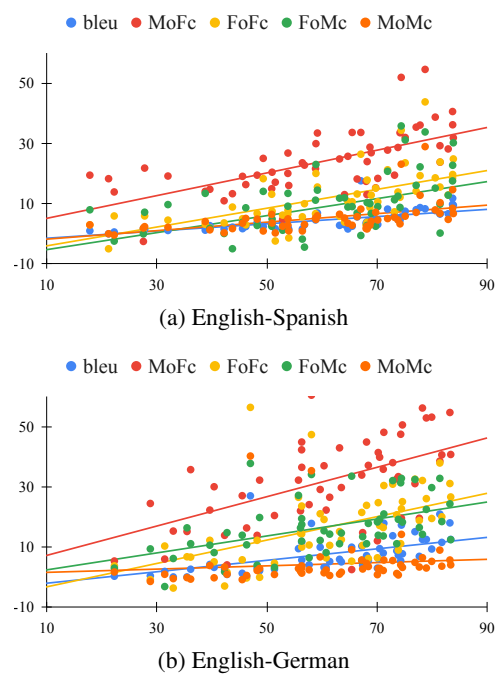
(a) English-Spanish

(b) English-German

Figure A2: Plots showing *relative percentage drop* of BLEU and gender-test metrics on the y-axis and *relative percentage drop* in decoding time in the x-axis.