Technology
Arts Sciences
TH Köln

ASSESSING GENDER BIAS IN MACHINE TRANSLATION

A case study with Google Translate

Melanie Fumfack, Kai Gehlen, Pia Störmer

DIS25a NLP – Prof. Dr. Philipp Schaer, Fabian Haak

AGENDA

CONTENT SUMMARY

- HISTORY OF MACHINE TRANSLATION
- MACHINE BIAS & GOOGLE TRANSLATE
- DATA AQUISITION
- STATISTICAL ANALYSIS
- GOOGLE'S APPROACHES

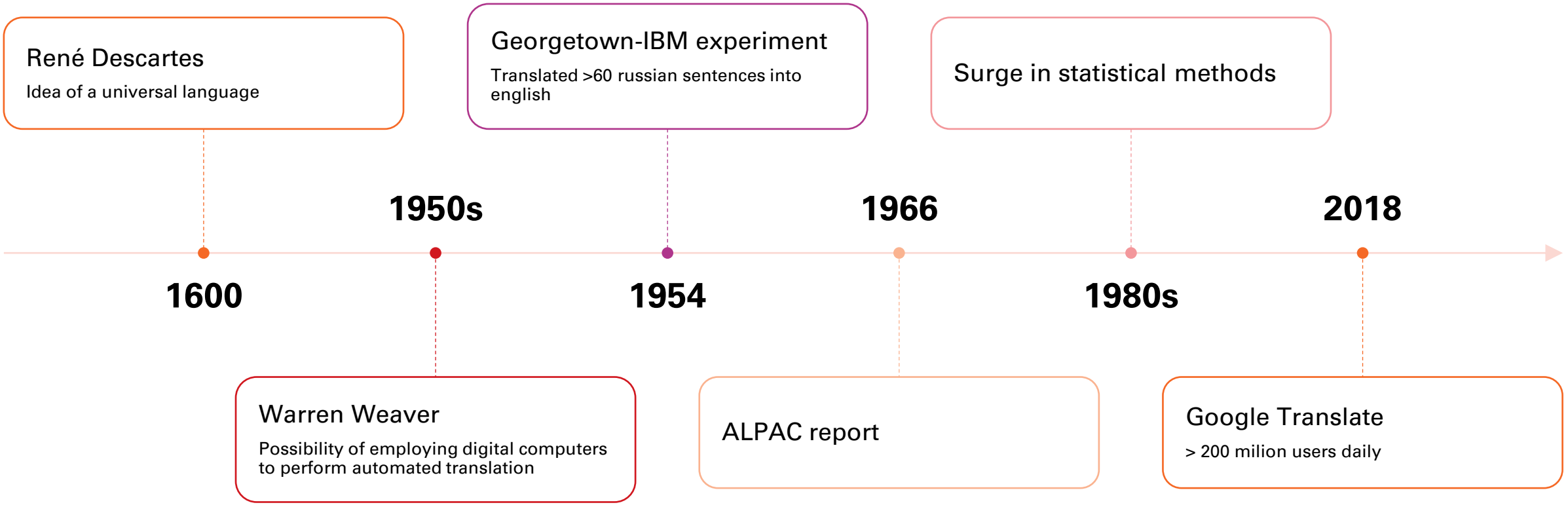
TRANSFER TO ESUPOL DATASETS

- TRANSLATE NATIONALITIES
- GENDER BIAS COMPARISON

CONTENT SUMMARY

HISTORY OF MACHINE TRANSLATION

HISTORY OF MACHINE TRANSLATION



TWO LEADING STANDPOINTS

Noam Chomsky	faith of the MT community in statistical methods is absurd by analogy with a standard scientific field such as physics
Peter Norvig	even standard physical theories such as the Newtonian model of gravitation are, in a sense, trained

CONTENT SUMMARY

MACHINE BIAS AND GOOGLE TRANSLATE

MACHINE BIAS

"trained statistical models unbeknownst to their creators grow to reflect controversial societal asymmetries"

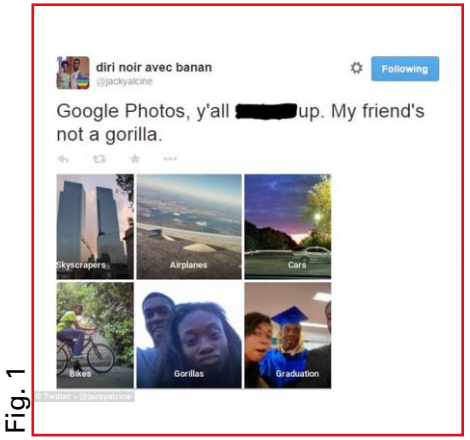


Fig. 1

Fig. 2

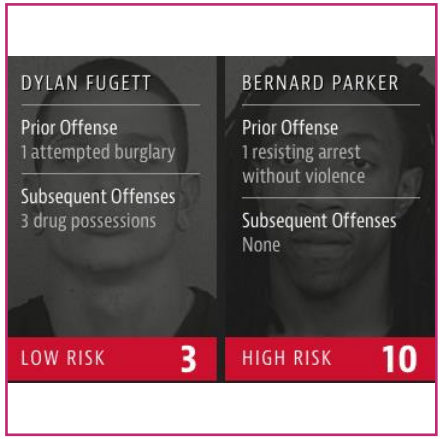


Fig. 3

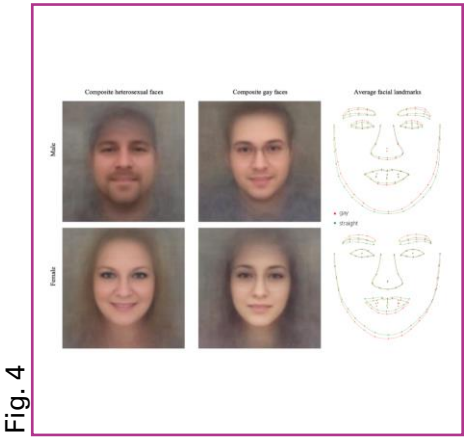


Fig. 4

DEVELOPMENT OF GOOGLE TRANSLATE

Technology Arts Sciences TH Köln

Initially:

relying on UN and EU Parliament transcripts to gather data

since 2014:

User content as input through the Translate Community initiative

Growing concern about gender asymmetrics

- "word embeddings are particularly prone to yielding gender stereotypes"
- Possible solution: simple debiasing algorithm

CONTENT SUMMARY

DATA AQUISITION

LANGUAGES

Table 1 Gender neutral languages supported by Google Translate

Language family	Language	Phrases have male/female markers	Tested
Austronesian	Malay	X	✓
Uralic	Estonian	X	✓
	Finnish	X	✓
	Hungarian	X	✓
Indo-European	Armenian	X	✓
	Bengali	O	✓
	English	✓	X
	Persian	X	✓
	Nepali	O	✓
Japonic	Japanese	X	✓
Koreanic	***Korean***	✓	X
Turkic	Turkish	X	✓
Niger-Congo	Yoruba	X	✓
	Swahili	X	✓
Isolate	Basque	X	✓
Sino-Tibetan	Chinese	X	✓

Languages are grouped according to language families and classified according to whether they enforce any kind of mandatory gender (male/female) demarcation on simple phrases (✓: yes, X: never, O: some). For the purposes of this work, we have decided to work only with languages lacking such demarcation. Languages in bolditalic have been omitted for other reasons. See Sect. 4.1 for further explanation

List of gender neutral languages (formed using WALS and other sources)

Omitted Korean and Nepali due to issues with grammar and no available native speaker

Technology
Arts Sciences
TH Köln

PROFESSIONAL OCCUPATIONS

Table 2 Selected occupations obtained from the U.S. Bureau of Labor Statistics https://www.bls.gov/cps/cpsaat11.htm, grouped by category

Category	Group	#Occupations	Female participation (%)
Education, training, and library	Education	22	73.0
Business and financial operations	Corporate	46	54.0
Office and administrative support	Service	87	72.2
Healthcare support	Healthcare	16	87.1
Management	Corporate	46	39.8
Installation, maintenance, and repair	Service	91	4.0
Healthcare practitioners and technical	Healthcare	43	75.0
Community and social service	Service	14	66.1
Sales and related	Corporate	28	49.1
Production	Production	264	28.9
Architecture and engineering	STEM	29	16.2
Life, physical, and social science	STEM	34	47.4
Transportation and material moving	Service	70	17.3
Arts, design, entertainment, sports, and media	Arts/Entertainment	37	46.9
Legal	Legal	7	52.8
Protective service	Service	28	22.3
Food preparation and serving related	Service	17	53.8
Farming, fishing, and forestry	Farming/Fishing/Forestry	13	23.4
Computer and mathematical	STEM	16	25.5
Personal care and service	Service	33	76.1
Construction and extraction	Construction/Extraction	68	3.0
Building and grounds cleaning and maintenance	Service	10	40.7
Total	–	1019	41.3

We obtained a total of 1019 occupations from 22 distinct categories. We have further grouped them into broader groups (or *super-categories*) to ease analysis and visualization

Table 3 A randomly selected example subset of thirty occupations obtained from our dataset with a total of 1019 different occupations

Insurance sales agent	Editor	Rancher
Ticket taker	Pile-driver operator	Tool maker
Jeweler	Judicial law clerk	Auditing clerk
Physician	Embalmer	Door-to-door salesperson
Packer	Bookkeeping clerk	Community health worker
Sales worker	Floor finisher	Social science technician
Probation officer	Paper goods machine setter	Heating installer
Animal breeder	Instructor	Teacher assistant
Statistical assistant	Shipping clerk	Trapper
Pharmacy aide	Sewing machine operator	Service unit operator

- Comprehensive list of professional occupations from the bureau of labor statistics
- Filtered some professions that were too generic or had gender-specific words
- Grouped into broader categories
- Randomly selected several occupations

ADJECTIVES

Table 4 Curated list of 21 adjectives obtained from the top one thousand most frequent words in this category in the Corpus of Contemporary American English (COCA) https://corpus.byu.edu/coca/

Happy	Sad	Right
Wrong	Afraid	Brave
Smart	Dumb	Proud
Strong	Polite	Cruel
Desirable	Loving	Sympathetic
Modest	Successful	Guilty
Innocent	Mature	Shy

small subset of 21 adjectives

Obtained from the top one thousand most frequent words Corpus of Contemporary American English (COCA)

Manually curated since many of these were not applicable to human subjects

Manual selection of words that would be meaningful to the study

CONTENT SUMMARY

STATISTICAL ANALYSIS

DISTRIBUTION OF GENDER TRANSLATION

Technology
Arts Sciences
TH Köln

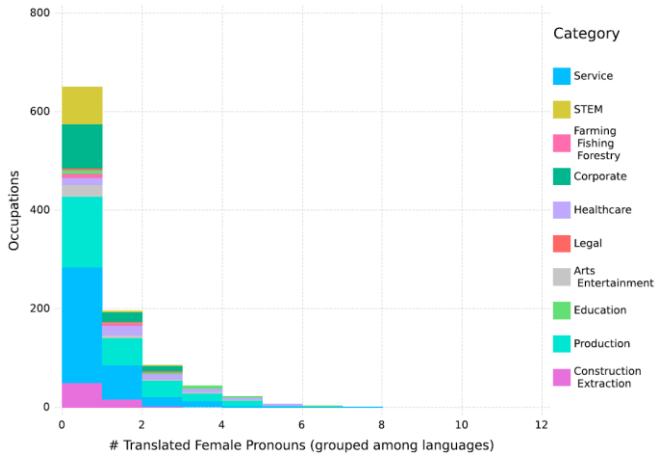


Fig. 5

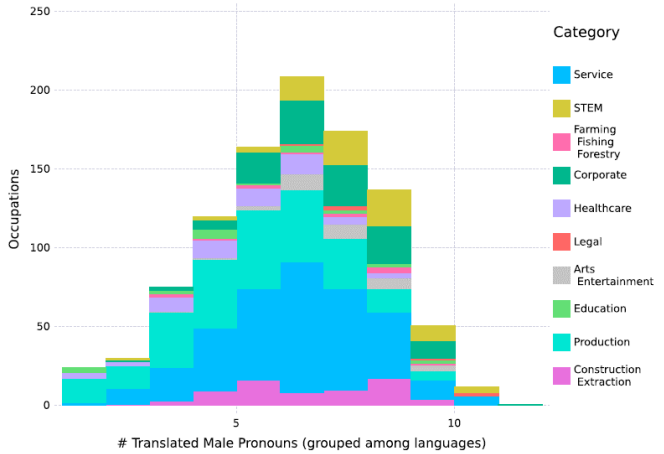


Fig. 6

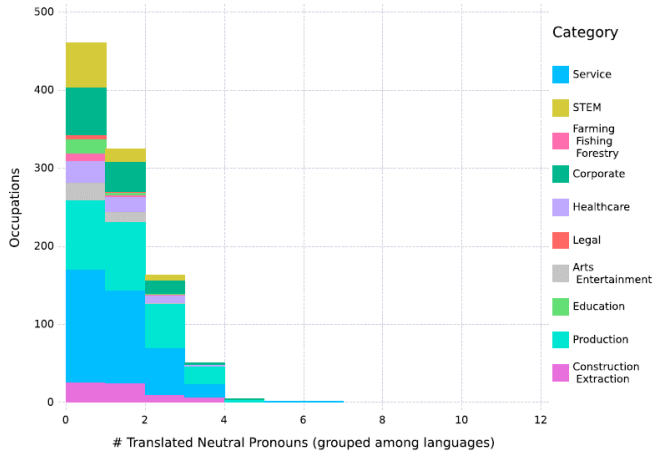


Fig. 7

PROBABILITY FOR GENDER TRANSLATION

Technology
Arts Sciences
TH Köln

FEMALE

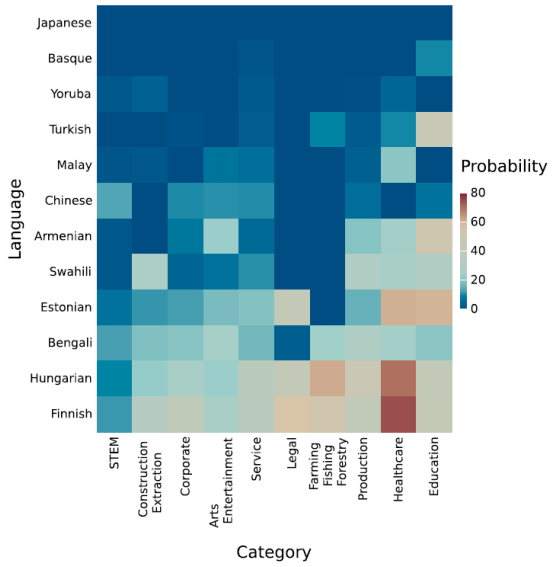


Fig. 12

MALE

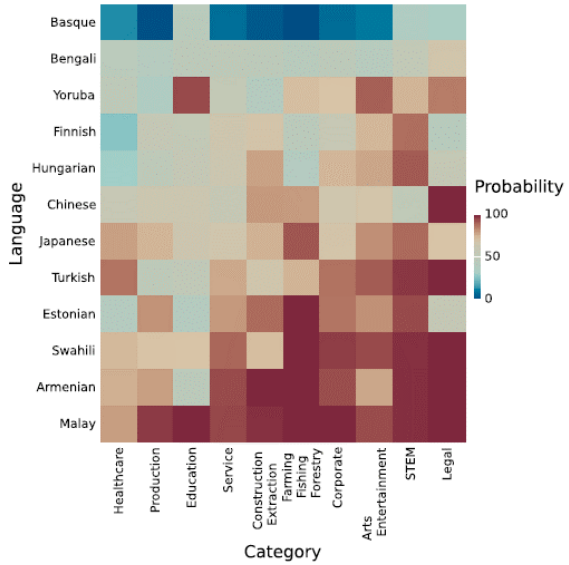


Fig. 13

NEUTRAL

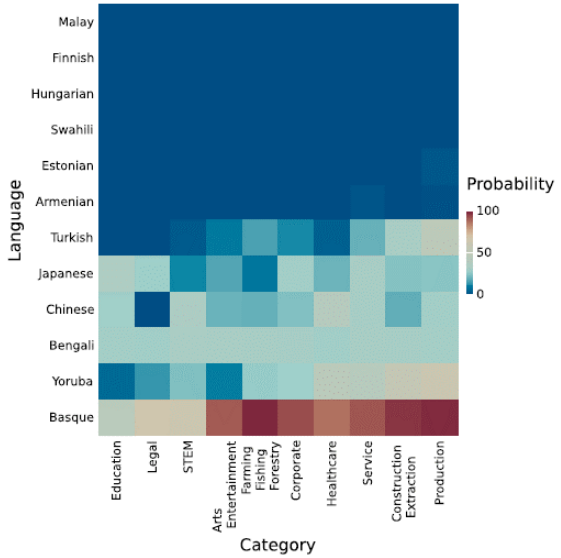


Fig. 14

DISTRIBUTION OF TRANSLATED PRONOUNS FOR VARIED ADJECTIVES

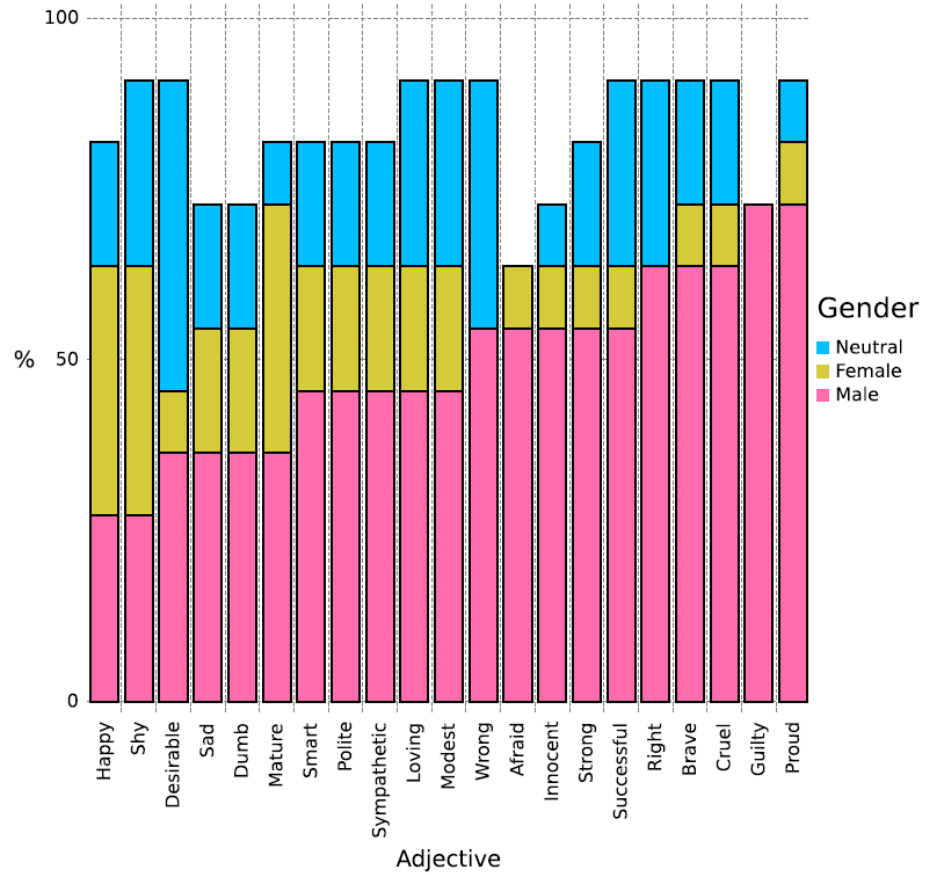


Fig. 15

STATISTICAL METHODS

Table 7 Computed *p*-values relative to the null hypothesis that the number of translated male pronouns is not significantly greater than that of female pronouns, organized for each language and each occupation category

	Mal.	Est.	Fin.	Hun.	Arm.	Ben.	Jap.	Tur.	Yor.	Bas.	Swa.	Chi.	Total
Service	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$
STEM	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$
Farming Fishing Forestry	$<\alpha$	$<\alpha$	**.603**	**.786**	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	*	$<\alpha$	$<\alpha$	$<\alpha$
Corporate	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$
Healthcare	$<\alpha$	**.938**	*1.0*	*.999*	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$
Legal	$<\alpha$	**.368**	**.632**	**.368**	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	**.086**	$<\alpha$	$<\alpha$	$<\alpha$
Arts Entertainment	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	**.08**	$<\alpha$	$<\alpha$	$<\alpha$
Education	$<\alpha$	**.808**	**.333**	**.263**	**.588**	$<\alpha$	$<\alpha$	**.417**	$<\alpha$	**.052**	$<\alpha$	$<\alpha$	$<\alpha$
Production	$<\alpha$	$<\alpha$	$<\alpha$	**.5**	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	**.159**	$<\alpha$	$<\alpha$	$<\alpha$
Construction Extraction	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	**.16**	$<\alpha$	$<\alpha$	$<\alpha$
Total	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$

STATISTICAL METHODS

Table 8 Computed *p*-values relative to the null hypothesis that the number of translated male pronouns is not significantly greater than that of gender neutral pronouns, organized for each language and each occupation category

	Mal.	Est.	Fin.	Hun.	Arm.	Ben.	Jap.	Tur.	Yor.	Bas.	Swa.	Chi.	Total
Service	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	*1.0*	$<\alpha$	$<\alpha$	$<\alpha$
STEM	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	*.984*	$<\alpha$	**.07**	$<\alpha$
Farming Fishing Forestry	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	**.135**	$<\alpha$	$<\alpha$	**.068**	*1.0*	$<\alpha$	$<\alpha$	$<\alpha$
Corporate	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	*1.0*	$<\alpha$	$<\alpha$	$<\alpha$
Healthcare	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	**.39**	*1.0*	$<\alpha$	**.088**	$<\alpha$
Legal	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	**.145**	$<\alpha$	$<\alpha$	**.771**	$<\alpha$	$<\alpha$	$<\alpha$
Arts Entertainment	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	**.07**	$<\alpha$	$<\alpha$	$<\alpha$	*1.0*	$<\alpha$	$<\alpha$	$<\alpha$
Education	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	**.093**	$<\alpha$	$<\alpha$	**.5**	$<\alpha$	**.068**	$<\alpha$
Production	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	**.412**	*1.0*	*1.0*	$<\alpha$	$<\alpha$	$<\alpha$
Construction Extraction	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	**.92**	*1.0*	$<\alpha$	$<\alpha$	$<\alpha$
Total	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	*1.0*	$<\alpha$	$<\alpha$	$<\alpha$

STATISTICAL METHODS

Table 9 Computed p-values relative to the null hypothesis that the number of translated gender neutral pronouns is not significantly greater than that of female pronouns, organized for each language and each occupation category

	Mal.	Est.	Fin.	Hun.	Arm.	Ben.	Jap.	Tur.	Yor.	Bas.	Swa.	Chi.	Total
Service	*1.0*	*1.0*	*1.0*	*1.0*	*.981*	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	*1.0*	$<\alpha$	$<\alpha$
STEM	**.84**	*.978*	*.998*	*.993*	**.84**	$<\alpha$	$<\alpha$	**.079**	$<\alpha$	$<\alpha$	**.84**	$<\alpha$	$<\alpha$
Farming Fishing Forestry	*	*	*.999*	*1.0*	*	**.167**	**.169**	**.292**	$<\alpha$	$<\alpha$	*	**.083**	**.147**
Corporate	*	*1.0*	*1.0*	*1.0*	*.996*	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	*.977*	$<\alpha$	$<\alpha$
Healthcare	*1.0*	*1.0*	*1.0*	*1.0*	*1.0*	**.086**	$<\alpha$	**.87**	$<\alpha$	$<\alpha$	*1.0*	$<\alpha$	*.977*
Legal	*	*.961*	*.985*	*.961*	*	$<\alpha$	**.086**	*	**.178**	$<\alpha$	*	*	**.072**
Arts Entertainment	**.92**	*.994*	*.999*	*.998*	*.998*	**.067**	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	**.92**	**.162**	**.097**
Education	*	*1.0*	*.999*	*.999*	*1.0*	**.058**	$<\alpha$	*1.0*	**.164**	**.052**	*.995*	**.052**	*.992*
Production	*.996*	*1.0*	*1.0*	*1.0*	*1.0*	**.113**	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	*1.0*	$<\alpha$	$<\alpha$
Construction Extraction	**.84**	*.996*	*1.0*	*1.0*	*	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	*1.0*	$<\alpha$	$<\alpha$
Total	*1.0*	*1.0*	*1.0*	*1.0*	*1.0*	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	$<\alpha$	*1.0*	$<\alpha$	$<\alpha$

COMPARISON WITH WOMEN PARTICIPATION DATA ACROSS JOB POSITIONS

Total
GT: 11,76% female pronouns
BLS: 35,94% female workers

H_0 = percentage of female workers is not significantly larger than frequency of female translated pronouns

$\alpha = 0.05$
$p = 6.2 * 10^{-94}$

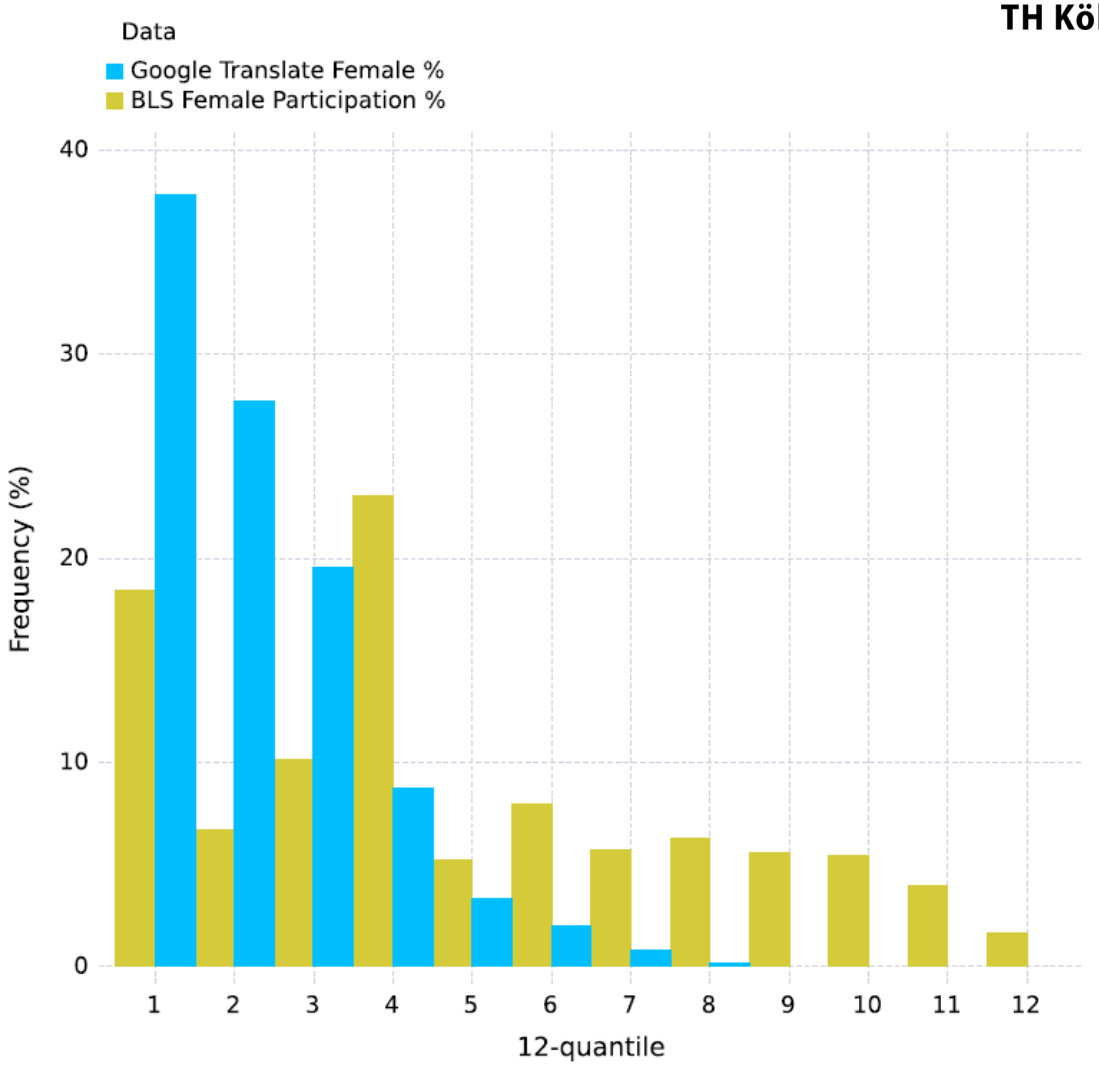


Fig. 16

CONTENT SUMMARY

Conclusions

GOOGLE'S APPROACHES

SOLUTION
Propose both genders.

Only works for translation into English.

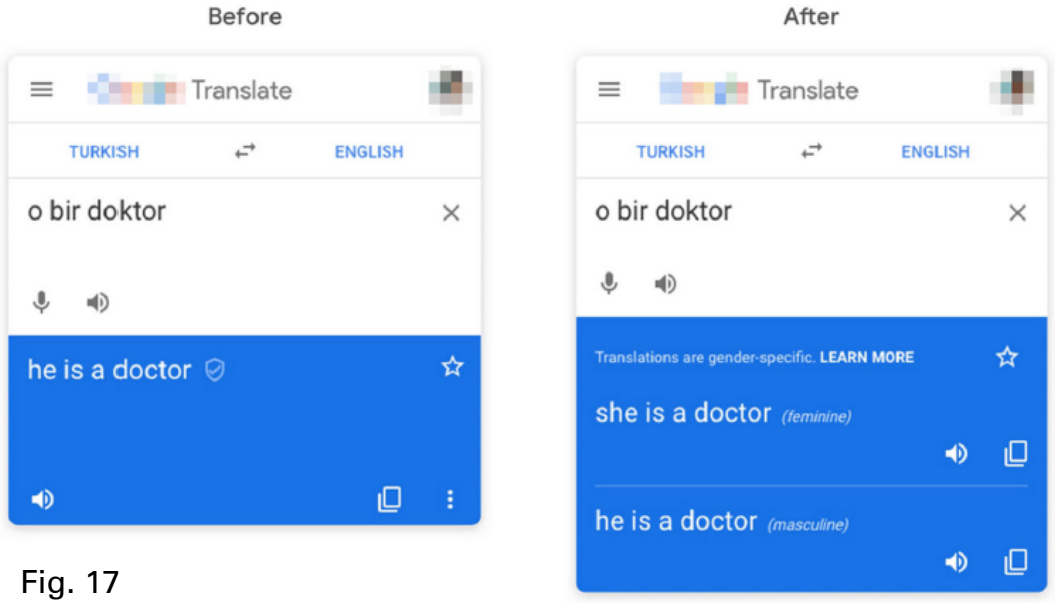


Fig. 17

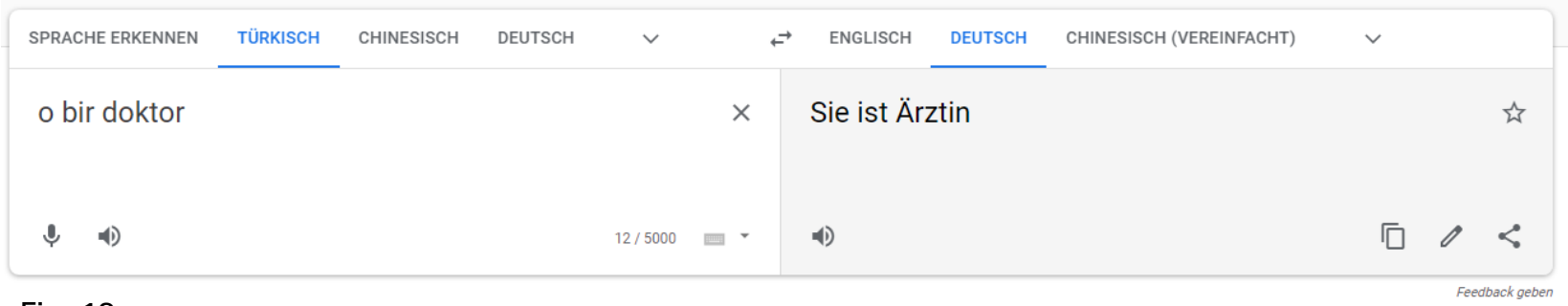


Fig. 18

TRANSFER TO ESUPOL DATASETS

TRANSLATE NATIONALITIES

SUGGESTIONS_MINORITIES DATASET

Fig. 19

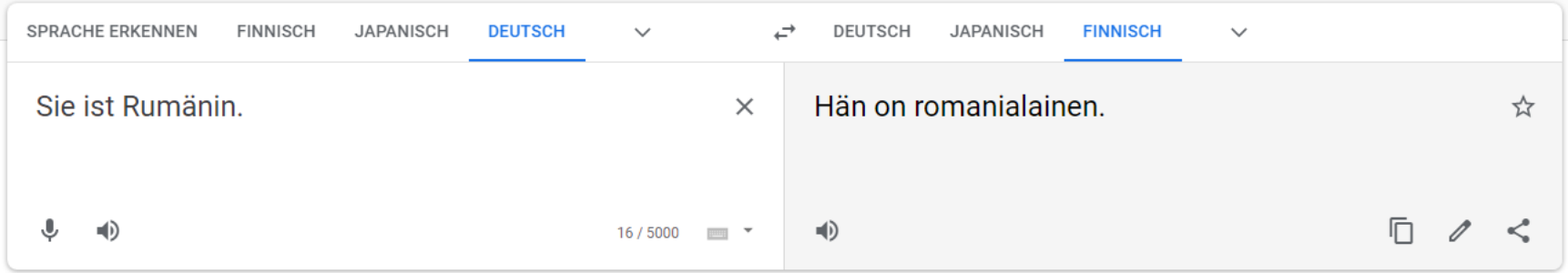
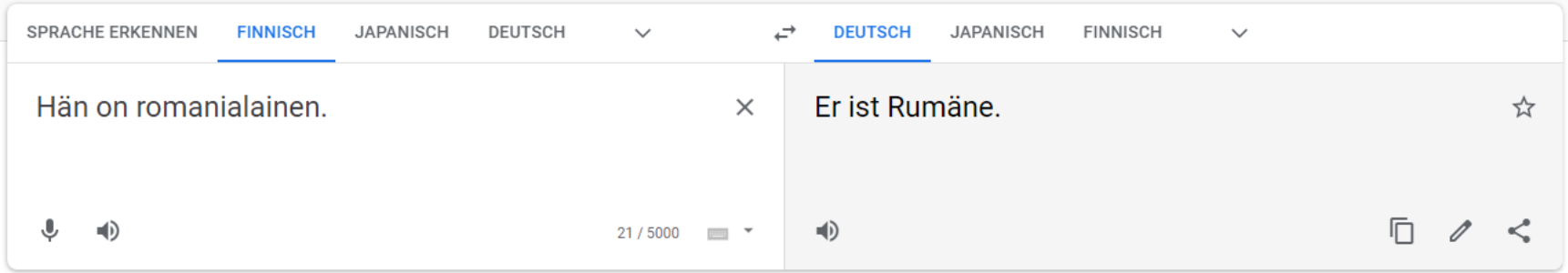


Fig. 20



GENDER BIAS COMPARISON

Are the suggestterms proposed for female queryterms different from those proposed for male queryterms?

Is the gender bias in the suggestions_eu dataset greater than in the btw17 dataset?

gender bias in ethnics based on the suggestterms

gender distribution of politicians per EU country

VIELEN DANK

Melanie Fumfack, Kai Gehlen, Pia Störmer

REFERENCES

- Prates, Marcelo (2019): Assessing gender bias in machine translation: a case study with Google Translate, Neural Computing and Applications, [online] https://doi.org/10.1007/s00521-019-04144-6 [abgerufen am 18.05.2021].
  (All tables used in this presentation are content of this paper)

LIST OF FIGURES