

# NLP and Equality

Anders Søgaard

# Outline

**NLP may  
soon work for  
ALL languages!**

---

**Only SOME  
speakers of  
all languages?**

imgflip.com



NLP may soon work for **all** languages

Intelligent  
document  
classification

Text-based  
recommen-  
der systems

Intelligent  
search  
engines

Speech  
recognition

Machine  
translation

NLP may soon work for **all** languages

Text  
prediction

Knowledge  
base  
population

Question  
answering

Information  
extraction

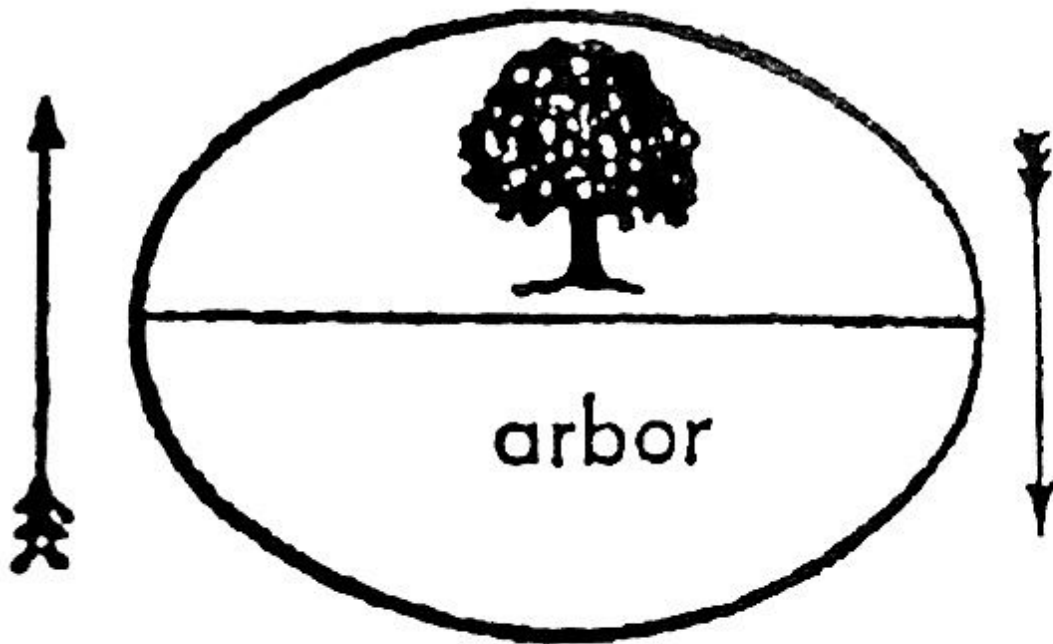
# What is language understanding?

## Standard View

The meaning of a word is what I am pointing to.

## Distributional Hypothesis

The meaning of a word is how it is used.



# Masked language models

MLMs are trained to predict missing (deleted) words in context.

**ANIMAL ANTICS**

EAT RUNS CAN WENT JUMP

THE FROG CAN \_\_\_\_\_

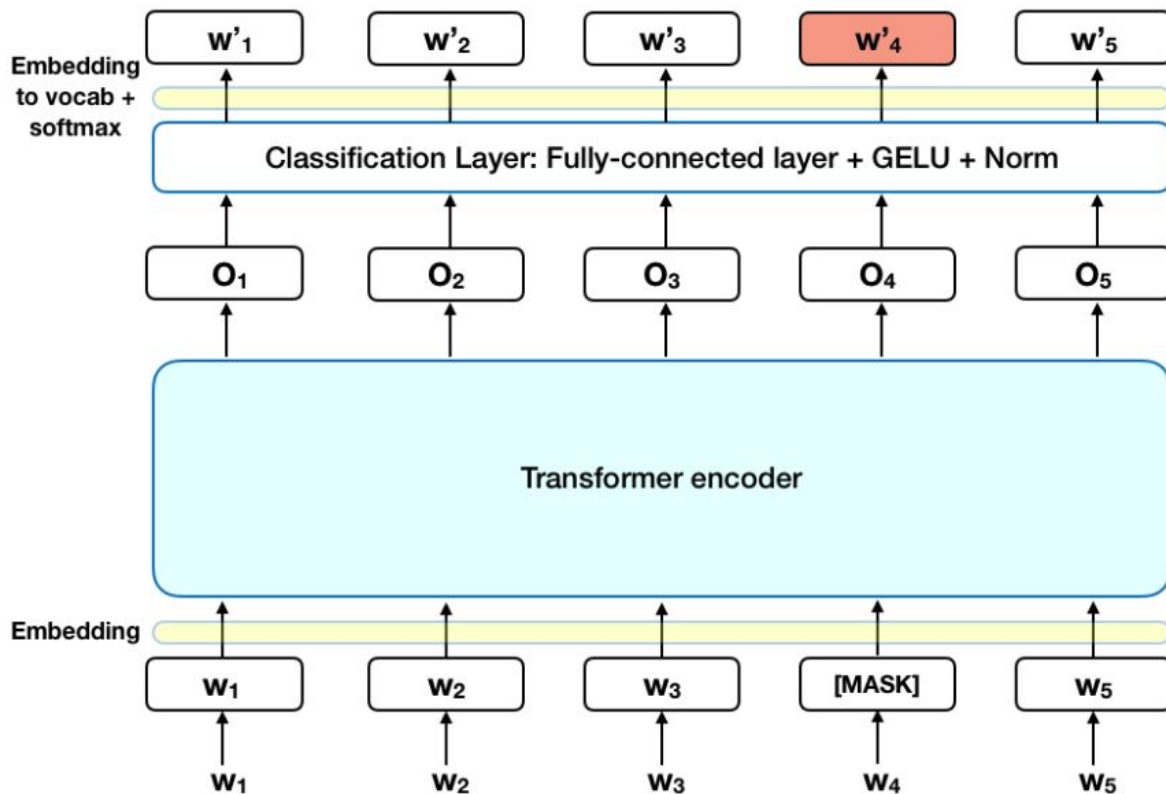


THE DUCK \_\_\_\_\_ SWIM.

THE RABBIT LIKES TO \_\_\_\_\_



FILL IN THE BLANKS WITH MISSING WORDS



# Masked language models

MLMs are trained to predict missing (deleted) words in context.

No linguists needed.



THE FROG CAN\_\_\_\_\_



THE DUCK\_\_\_\_\_SWIM.

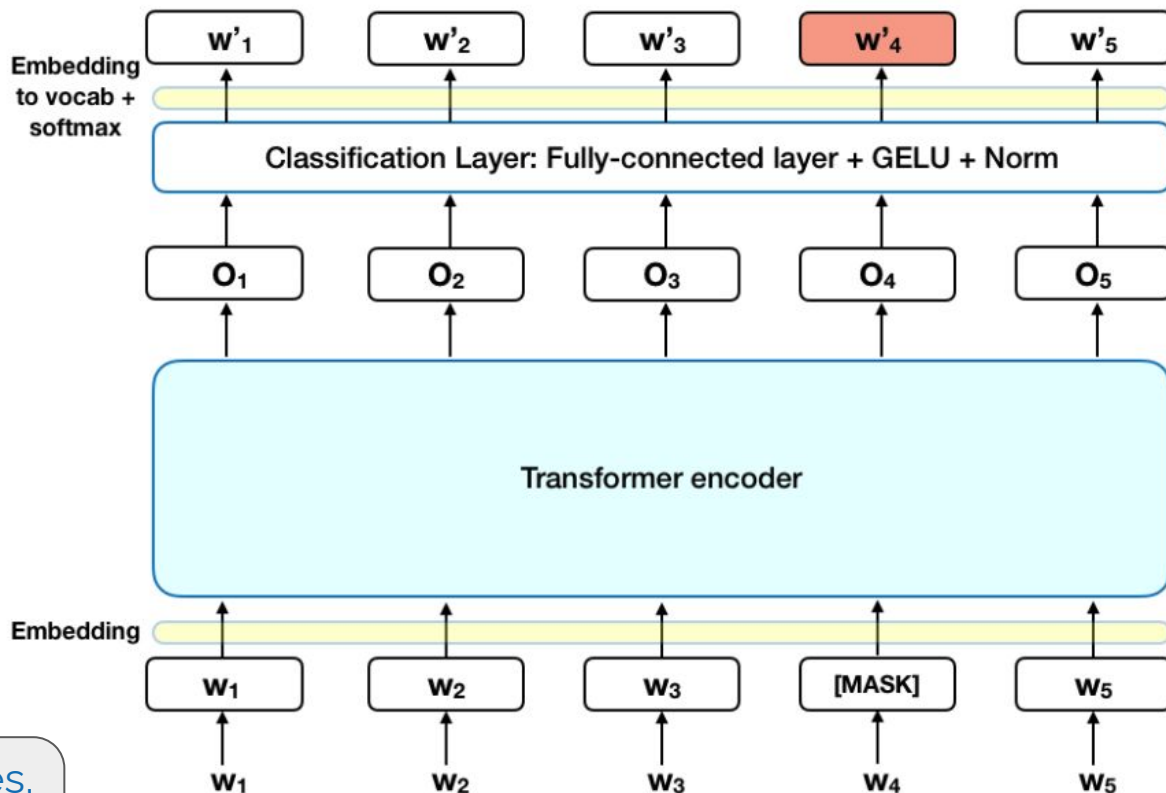


THE RABBIT LIKES TO \_\_\_\_\_



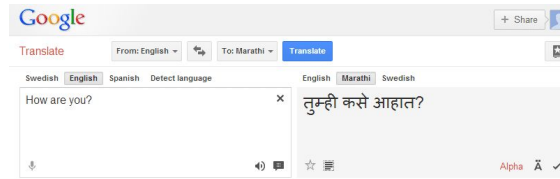
FILL IN THE BLANKS WITH

No dictionaries, grammars, or annotations.



# Still English after all these years

While more than half of the NLP research papers still focus exclusively on English, we now have both resources and models for hundreds of languages, and we see many product begin to support many of these languages. **That's great news for all of us!**



# Sanskrit Added to Google Translate

## 8 Indian languages added in the latest update

Dia.Rekhi@timesgroup.com

**ISAAC CASWELL** Senior Software Engineer, Google Research

**We have significantly closed the gap for at least the scheduled languages**

Chennai: Google has added eight Indian languages including Sanskrit to Google Translate, as the internet firm continues to increase the number of regional languages supported by its online multilingual translation service.

"Sanskrit is the number one, most-requested language at Google Translate, and we are finally adding it," Isaac Caswell, senior software engineer, Google Research, told ET in an exclusive interview.

"We are also adding the first languages from northeast India, which is another rather under-represented place."

Apart from Sanskrit, the Indian languages in the latest iteration of Google Translate are Assamese, Bhojpuri, Dogri, Konkani, Maitthili, Mizo and Meiteilon (Manipuri), taking the

total number of Indian languages supported by the service to 19.

The announcement was made at the annual Google conference I/O that began late on Wednesday night. The latest update does not cover all the 22 scheduled languages of India, as the company was

hoping, but Caswell said: "We have significantly closed the gap for at least the scheduled languages."

All the languages that have been added in the update will only be supported in the text translation feature.

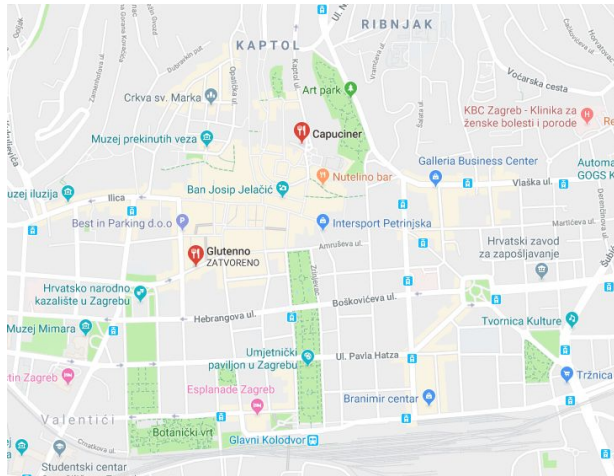




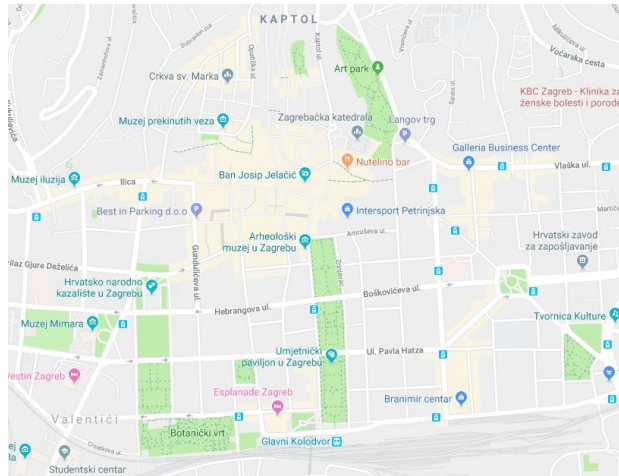
# Caveats

- The good news ignores performance and robustness disparities.
- One model and a few annotated corpora does not make for support.
- Some language classes are systematically ignored:
  - Creole and pidgin languages
  - Polysynthetic languages
  - Sociolects and dialects
  - Sign languages
- Still a long journey ahead.

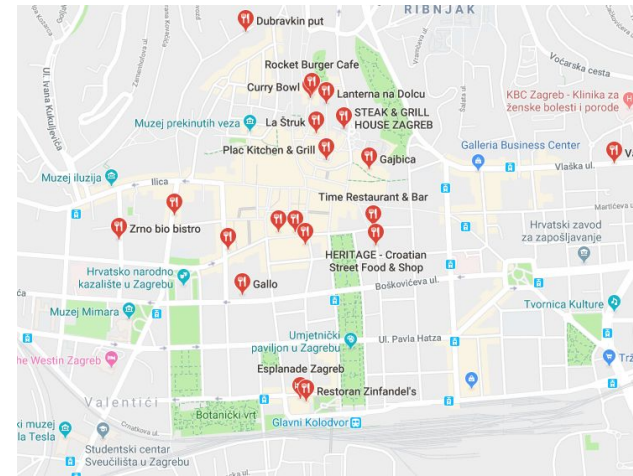
We're in Zagreb searching for...



...éttermek (HU)



...jatekx (EU)



...restaurants (EN)

NLP may soon work for **all** languages

NLP may soon work **for some speakers** for  
**all** languages

# Sociolects

- Gender
- Age
- Region
- Race
- Sexuality
- Religion
- Urbanicity
- Level of education
- Dyslexia
- Subculture



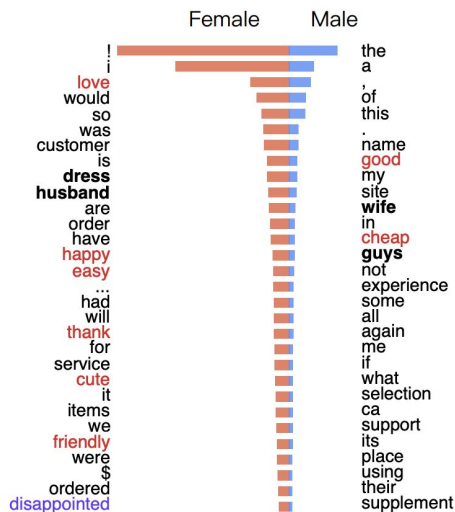
# Fairness

NLP models are not equally good for all people. For example, MLMs align more with the lexical preferences of young (less educated) white men (!).

Models	Demographics Alignment
bert-base-cased	
bert-base-uncased	
bert-base-multilingual-cased	
bert-large-cased	
bert-large-uncased	
distilbert-base-uncased	
albert-base-v2	
albert-large-v2	
albert-xxlarge-v2	
roberta-base	
roberta-large	
google/electra-large-generator	
google/electra-small-generator	
gpt2	
gpt2-medium	
gpt2-large	
gpt2-xl	
<b>Group</b>	
<b>Mean Rank</b>	3.1 3.4 4.0 6.1 6.1 8.1 8.1 9.2 9.8 9.9 10.3 10.3 10.8 11.1 12.0 13.8

# Fairness

NLP models are not equally good for all people. For example, MLMs align more with the lexical preferences of young (less educated) white men (!).



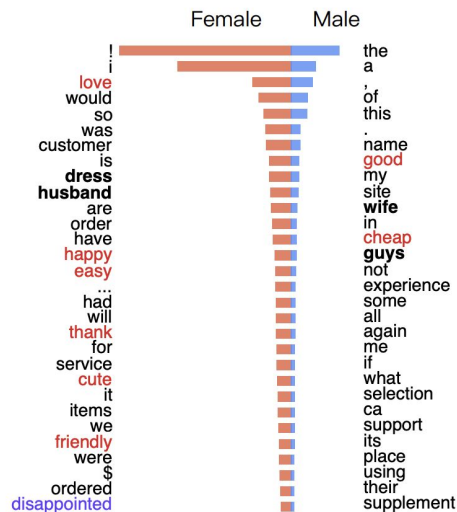
(a) Female v.s. male (US English datasets).



(b) Elder female v.s. young female (US English datasets).

# Worst-case-aware automated curriculum learning

**Idea:** Sample training data to minimize worst case loss. Results on predicting the star-rating of Trustpilot reviews across languages. Min-F1 is performance on minority group.



(a) Female v.s. male (US English datasets).



(b) Elder female v.s. young female (US English datasets).

Trained On	Danish		French		German		UK English		US English	
Transfer Results	$\bar{F}_1$	min $F_1$	$\bar{F}_1$	min $F_1$	$\bar{F}_1$	min $F_1$	$\bar{F}_1$	min $F_1$	$\bar{F}_1$	min $F_1$
EMPIRICALRISK	50.44	46.95	44.77	41.34	41.73	38.61	50.21	46.53	47.22	43.37
ERM-WILDS	50.69	47.18	45.02	41.56	<u>40.55</u>	37.95	50.76	46.98	<b>48.14</b>	43.41
GROUP-DRO	<u>49.09</u>	<u>44.97</u>	43.99	40.91	41.19	<u>37.55</u>	50.23	<u>45.73</u>	<u>45.93</u>	<u>41.13</u>
DEEPCORAL	50.63	46.06	<u>42.00</u>	<u>38.56</u>	41.51	38.00	<u>49.84</u>	46.13	46.97	43.52
OURS	<b>51.19</b>	<b>47.30</b>	<b>45.09</b>	<b>41.83</b>	<b>42.51</b>	<b>39.27</b>	<b>51.13</b>	<b>47.09</b>	47.98	<b>44.15</b>

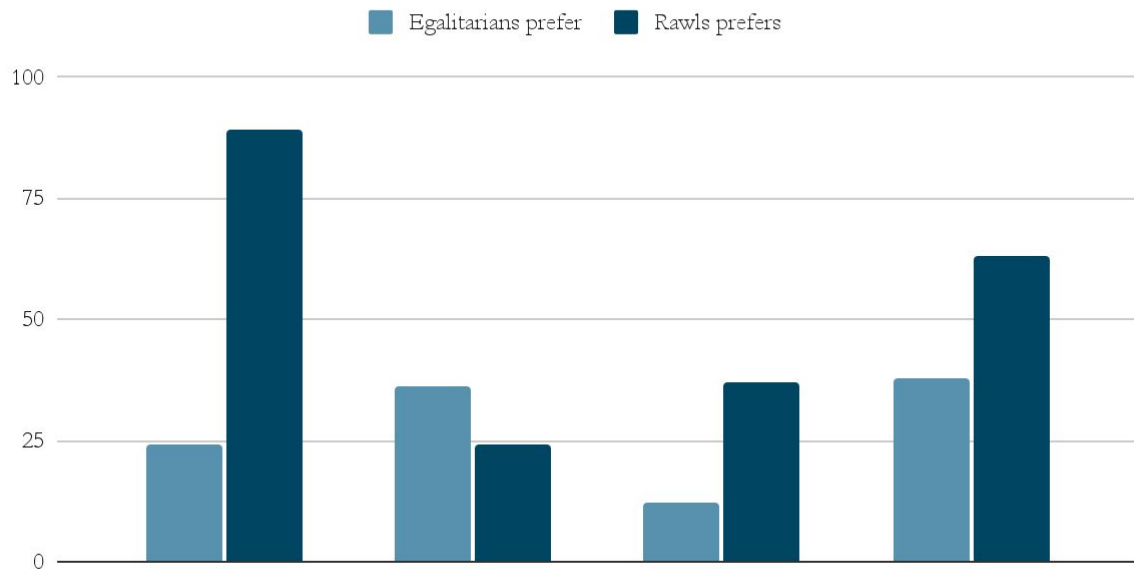


# Rawlsian Fairness

Rawl's difference principle justifies inequality as long as it improves the absolute position for the least advantaged group. This is most widely adopted definition of fairness in NLP/AI, if not the only one.

## Why is it problematic?

Equality illustrated



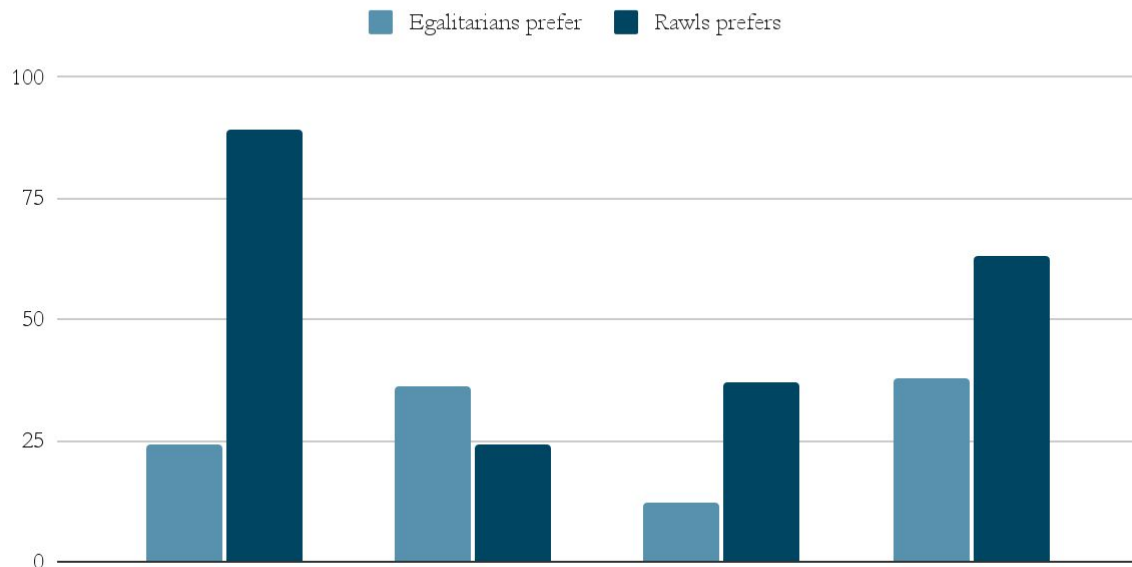
# Rawlsian Fairness

## Challenge

Any intervention, e.g., research or the development of a new technology, has short-term and long-term effects, is evaluated down the line. Doing a proof-of-concept for **English** or for **young white men** may be argued to help others down the line, but what if it doesn't?

**For example:** What if the technology is never scaled up? Or what if being the first with it, is a huge market advantage?

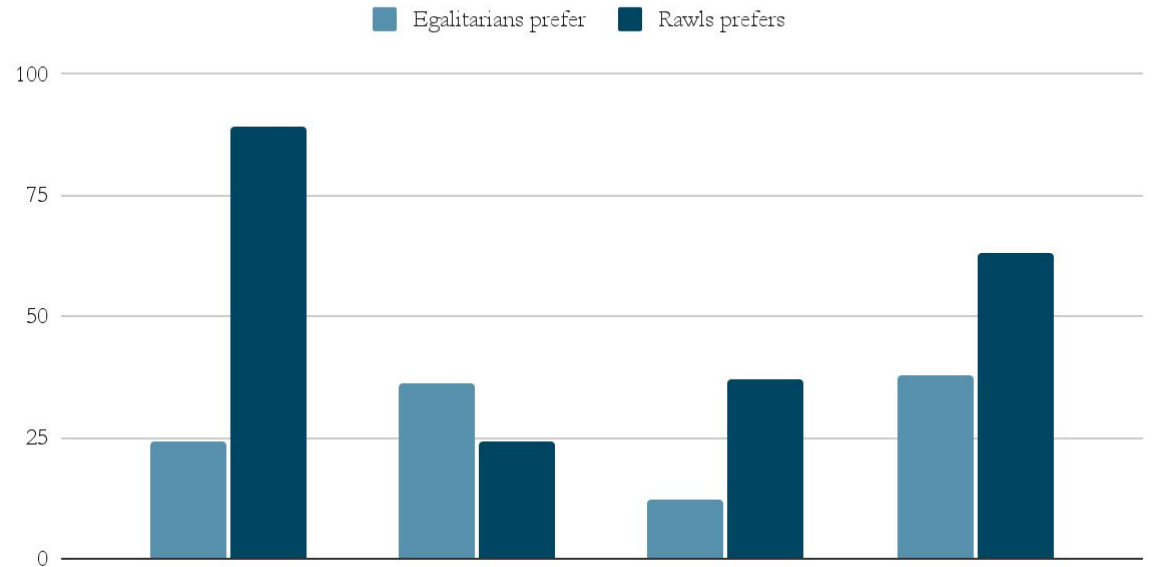
Equality illustrated



# Egalitarian Fairness

Nielsen: 'After provisions are made for common social (community) values, for capital overhead to preserve the society's productive capacity and allowances are made for differing [unmanipulated needs and preferences](#), the income and wealth (the common stock of means) is to be so divided that each person will have a right to an equal share.'

Equality illustrated



# NLP for Special Needs

What is unmanipulated needs, and how does it translate into NLP research?

- Text simplification for dyslexics
- NLP for language learners
- NLP for medical diagnosing
- Morphological analyzers for polysynthetic languages
- NLP in Refugee Camps?



?

<https://anderssoegaard.github.io/>