The influence of the corpus on the representation of gender stereotypes in the dictionary
A case study of corpus-based dictionaries of German
Motivation of the study
Have you ever googled woman?

• 2019, the British PR manager Maria Beatrice Giovanardi wrote a blog post titled "Have you ever googled 'womar' primarily complained about the description of women in various dictionaries, including lexicographic works by Oxford University Press
• *filly, biddy or bitch* are listed as synonyms for woman
• petition at change.org, signed by 30,000 people
• Oxford University Press responded by sending Katherine Connor Martin the following statement via The Guardian newspaper: The dictionary editors "are taking the points raised in the petition very seriously ... As ever, our dictionaries strive to reflect, rather than dictate, language so any changes will be made on that basis."
Example phrases taken from the entries on *man*, *woman*, *girl* and *boy* in the Cambridge Dictionary, reproducing stereotypical gender concepts:

- “He plays baseball, drinks a lot of beer and generally acts like one of the boys.”
- “Steve can solve anything - the man’s a genius.”
- “She’s a really nice woman.”
- “Who was that beautiful girl I saw you with last night?”
- “Both girls compete for their father’s attention.”

https://dictionary.cambridge.org/de/
Within the collocation sets for “what is discussed in connection with man or woman?”, man collocates with: car, erectile dysfunction, fire department, soccer, equality, and handball. For woman, it is age, occupation, breast cancer, emancipation, employment, birth, children, sex, and menopause.

The elexiko team expresses critical awareness of these stereotypical representations (Klosa/Storjohann 2011, p. 64).

Newspaper-heavy corpus or language use in general?

https://www.owid.de/docs/elex/start.jsp
Computer-generated collocation profiles

• Typical adjectives for man are young, old, rich, strong, adult, powerful, armed and right, whereas the typical ones for woman are young, old, beautiful, tall, naked, pregnant, gracious and employed.

In contrast, examples edited by lexicographers differ strongly:

• E.g., a young, smart, strong, fascinating, educated, emancipated woman; the woman of today (the modern woman); a Russian woman was the first woman in space

• Much has been done in the Duden editorial office to avoid gender stereotypes and to revise example phrases from older editions; problem here: computer-generated collocations

Gender stereotypes in dictionaries

- Dictionaries are often a reflection of their time, i.e. how they describe the meaning of certain lexical units must always be seen in their respective historical context.
- They are one of the sources to reflect gender roles (Nübling 2010, p. 594, cf. also Hu, Xu & Hao 2019)
- The question arises whether very stereotypical representations of gender in dictionaries are or can be intentional

The representation of gender in dictionaries thus seems to be caught between language use and lexicographic-moral responsibility.

How much must or should the lexicographer intervene in the description of language use?

What is regarded as a basis for the ‘reflection of language’? In the tradition of modern corpus-based lexicography, it is the underlying corpus. But does everything from this corpus always have to be included in the dictionary? Or should it rather be a curated selection?

Should language use find its way into the dictionary, even if it could perpetuate gender stereotypes that, at least in part, no longer fit with contemporary ideas of society? Is it acceptable to reproduce racist and sexist attitudes exactly as they are (still) used?
Case study: Influence of the corpus on collocations sets for Mann und Frau
Newspaper-heavy corpora

"The written standard is to be regarded as the language usage of the national press. [...] It follows that the written standard can in principle be determined statistically by an evaluation of extensive newspaper corpora."


“Since a dictionary has to give an account of all the main meanings and uses of every headword it includes, it follows that a lexicographic corpus needs to include evidence for all these uses. This in turn argues for a corpus whose constituent texts cover the full repertoire of mainstream text-types.”

Case study

- Would the collocation sets for Mann and Frau change significantly if the corpus was not predominantly composed of newspaper texts?
- Three corpora: fiction books, magazines, newspapers
- 'Fiction Books': various works of fiction (20th and 21st century)
- Newspapers (elexiko corpus): daily and weekly newspapers
- Magazines: various periodical magazines (art, BEEF!, brand eins, BRIGITTE, Capital, Chefkoch, Couch, Eltern, Essen und Trinken, Gala, GEO, Living at Home, Nido, NEON, Psychologie Heute and Schöner Wohnen)
## Case study

<table>
<thead>
<tr>
<th></th>
<th>fiction books</th>
<th>magazines</th>
<th>newspapers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Texts</strong></td>
<td>1.320</td>
<td>60.066</td>
<td>15.831.499</td>
</tr>
<tr>
<td><strong>Sentences</strong></td>
<td>1.221.373</td>
<td>2.511.280</td>
<td>263.625.222</td>
</tr>
<tr>
<td><strong>Tokens</strong></td>
<td>22.132.897</td>
<td>37.771.792</td>
<td>4.398.207.319</td>
</tr>
</tbody>
</table>
Fiction Books
• described regarding their external characteristics: blond, pretty or attractive (blond, hübsch, attraktiv), but also in terms of their marital status (married or divorced – verheiratet, geschieden), or even pregnant (schwanger).

Newspapers
• women are also described as pregnant (schwanger) or working (berufstätig), but also as being raped (vergewaltigt).

Magazines
• women are self-confident, employed, attractive, pregnant, independent, or emancipated (selbstbewusst, berufstätig, attraktiv, schwanger, unabhängig, emanzipiert)
• The magazines corpus is therefore the only dataset in which women are characterized by significant collocators that have nothing to do with their appearance, social role, or being victims of crime.
Mann: Adjectives

Fiction Books
• in the fiction books, descriptive adjectives such as gaunt, stout, stocky, bearded or lanky (hager, kräftig, untersetzt, bärtig, schmächtig)

Newspapers
• violent acts are a predominant: armed, masked, alcoholized, previously convicted (bewaffnet, maskiert, alkoholisiert, vorbestraft), but also more general words like unemployed or powerful (arbeitslos, mächtig)

Magazines
• men are described as attractive, married, bearded, naked, gay or *-looking (attraktiv, verheiratet, bärtig, nackt, schwul, aussehend), Surprisingly, a considerable number of terms related to appearance, social role or sexual orientation are found here
Mann: Verbs

- Fiction books: muster, marry, observe, sit opposite, turn to (mustern, heiraten, beobachten, gegenüber sitzen)
- Newspaper texts, the context of violence is again predominant: arrest, assault, threaten, shoot, and rape (festnehmen, überfallen, bedrohen, erschießen, vergewaltigen)
- Magazines: words referring to love life, money or power status are frequent collocators: marry, fall in love, question, earn, cheat, question, or dominate (heiraten, verlieben, befragen, verdienen, betrügen, befragen, dominieren)

<table>
<thead>
<tr>
<th>VV</th>
<th>fiction books</th>
<th>magazines</th>
<th>newspapers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mann zuwenden, beobachten, untersetzen, heiraten, müssen, erwachsen</td>
<td>verdienen, betrügen, befragen, kennenlernen, prüfen, räumen, verheiraten</td>
<td>festnehmen, erschießen, vergewaltigen, verunreinigen, verurteilen, ausrauben, verirren, maskieren, erbauen, erheben</td>
<td>erstechen, fahnden, toten, betreten, mißbrauchen, zusammenschlagen, fluchen, bewaffnen</td>
</tr>
</tbody>
</table>
Mann & Frau: Adjectives

- Both: *sexual, female, attractive*
- Mann: *married, naked, bearded*
- Frau: *self-confident, pregnant, independent*

<table>
<thead>
<tr>
<th>ADJ</th>
<th>Frau</th>
<th>Mann</th>
</tr>
</thead>
<tbody>
<tr>
<td>selbstbewusst</td>
<td>emancipated</td>
<td>bértig</td>
</tr>
<tr>
<td>finanzil</td>
<td>berufstätig</td>
<td>gewalttätig</td>
</tr>
<tr>
<td>männlich</td>
<td>kinderlos</td>
<td>maskulin</td>
</tr>
<tr>
<td>schwanger</td>
<td>sexuell</td>
<td>verstorben</td>
</tr>
<tr>
<td>attraktiv</td>
<td>unabhängig</td>
<td>schwul</td>
</tr>
<tr>
<td>zierlich</td>
<td></td>
<td>verheiratet</td>
</tr>
</tbody>
</table>
Mann & Frau: Adjectives

- Both: *sexual*
- Mann: *armed, masked, alcoholized, previously convicted*
- Frau: *pregnant, raped, working*
Mann Newspapers

Supported by (Joint work with Carole Tiberius)
Discussion
Awareness of gender constructions

• Our results show that in the newspaper texts, the common features of women and men as people who share many characteristics and actions step back in favour of the differences.
• context of violence, for example, is dominant only in the newspaper corpus
• corpus can bring an unnecessarily strong bias towards doing gender into the dictionary (cf. also Nübling 2010: 620)

Dictionaries as social agents

“Bias-free and inclusive paraphrases of meaning reflect real usage and the real consequences of bias and exclusion. Such paraphrases of meaning and illustrations are necessary in order for dictionaries and lexicographical resources as a whole to provide users with information that is trustworthy, culturally aware, and socially responsible. In doing so, such resources also promote a dictionary culture by addressing the unfulfilled needs of several target user groups, which encourages their consultation on a regular basis.“


“Dictionaries influence human behavior and can function as agents of inclusion or exclusion. [...] It is the task of the lexicographer to devise and implement a policy that can avoid a biased approach in the treatment of sensitive lexical items. Where this is done, the dictionary can contribute to processes of harmonization and as a practical instrument it can enhance impartial communication.”

Dictionaries as social agents

“Should data about how language is used control how women are defined? Or should we take a step back and, as humans, promote gender equality through the definitions of women that we choose to accept? […] We talked about how the dictionary is the most basic foundation of language and how it influences conversations. Isn’t it dangerous for women to maintain these definitions — of women as irritants, sex objects and subordinates to men?”

(Subjective) Conclusion

- We have a lot of uncurated data
- Dictionaries should differ from it
- There’s nothing such as a neutral perspective
- So much different perspectives on the world, all of them are subjective in a way
- A good compromise is certainly first to research language use with as much reflection (and self-reflection) as possible and then also - as one does with offensive or vulgar expressions - to find a compromise between language use orientation and the handing-down of outdated role models.
Thank you for your attention!
Mixed Corpus (all three corpora together, unchanged in size)
„Boys will be boys“

- “said to emphasize that people should not be surprised when boys or men act in a rough or noisy way because this is part of the male character” (Cambridge Advanced Learner’s Dictionary)
- “youthful indiscretion or exuberance must be expected and tolerated” (Collins English Dictionary)
- “used to indicate that it is not surprising or unusual when men or boys behave in energetic, rough, or improper ways” (Merriam-Webster Dictionary)

“Both Oxford paraphrases of meaning warn users that this behaviour “should not cause surprise,” the Cambridge paraphrase of meaning had “should not be surprised,” Merriam-Webster has “it is not surprising,” and Longman has “should not be surprised.” [...] Males are “wired” to be this way, so any “childish,” “rough,” “noisy,” “energetic,” “improper,” “foolish,” “aggressive,” “bad,” “indiscreet,” or “exuberant” manifestations are perfectly “normal” and must be expected and therefore tolerated. From these paraphrases of meaning it seems to be clear that it has been that way, continues to be that way, and will always be that way.” (215)