***Making obesity central to a person’s identity***

…

Table 4 clearly demonstrates that dispreferred condition-first language vastly outnumbers person-first language in the corpus (frequency normalized per 1000 words using Python word counts rather than those generated by CQPweb, as the latter includes counts for punctuation, inflating the word count the longer text; see GitHub ref). We resampled the corpus without replacement, selecting 1000 articles in a batch 10000 times. We then determined the count of articles that used condition-first and person-first language (discussed below) and the frequency per million words of each of these two language types in generated subcorpora. A Welch’s Two Sample t-test testing the difference between the frequency of person-first and condition-first language (mean frequency of person-first = 8.23 words per million, mean frequency of condition-first language = 284.74 words per million) suggests that the effect is negative, statistically significant, and large (difference = -276.51, 95% CI [-277.13, -275.89], t(10666.13) = -871.18, p < .001; Cohen's d = -12.32, 95% CI [-12.49, -11.03])

The difference is statistically significant (add details) and the effect size is big (add details). The breakdown of these aggregate figures per newspaper and year is available at GitHub ref.

|  |  |
| --- | --- |
| **Condition-first language (dispreferred)** | **Person-first language (preferred)** |
| Raw frequency/normalized frequency (per million words) | Raw frequency/normalized frequency (per million words) |
| 4,677/284.56 | 136/8.27 |

Table 4 Condition- vs person-first language in the Australian obesity corpus

This tendency also holds when the number of articles containing condition-first and person-first language are considered. In the corpus, condition-first language is used in 9-14% of articles from all sources (7-14% of articles per year), while person-first language is used in less than 1% of articles (0.17-1.14% of articles per year). Looking at the resampling data, a Welch’s Two Sample t-test testing the difference between the difference in the number of articles (per 1000) which use person-first and condition-first language (mean number of articles using person-first language across all subcorpora = 4.03, mean number of articles using condition-first language across all subcorpora = 122.54) suggests that the effect is negative, statistically significant, and large (difference = -118.51, 95% CI [-118.71, -118.31], t(10748.87) = -1149.40, p < .001; Cohen's d = -16.26, 95% CI [-16.47, -16.04]) Figure 1 shows the proportion of total articles containing at least one mention of condition-first language, person-first language, or both. (add details).

Our analysis demonstrates that among articles that use either condition-first, person-first language or both, the number of articles with only condition-first language is higher in tabloid publications and in right-leaning publications (p < 0.05, with small effect sizes (<0.2), Pearson’s Chi-squared test with Yates’ continuity correction; details at GitHub ref); similar differences are not observed for person-first language. A Welch Two Sample t-test in these articles that use either condition-first, person-first language or both, a difference between the frequency of condition-first and person-first langauge (per 1000 words) is observed: the mean frequency of condition-first language is 4.34 words per 1000, while the mean frequency of person-first language is 2.67 per 1000 words, suggesting that the effect is positive, statistically significant, and small (difference = 1.66, 95% CI [1.16, 2.17], t(131.59) = 6.49, p < .001; Cohen's d = 0.44, 95% CI [0.30, 0.58]). Finally, we used linear modelling to consider whether there are differences in the frequency of condition-first language use across years and individual newspapers (details at GitHub); many years have no instances of person-first language in some sources, so such an analysis is impossible for person-first language. Consistent with the above comparison between tabloids and broadsheets, a difference among sources was observed (Relative to the Advertiser, a tabloid, all considered broadsheets (the Age, Australian, Canberra Times and Sydney Morning Herald) had a lower frequency of condition-first language, as did the Courier Mail); no difference across the time period could be detected. This suggests that there has not been a clear increase of person-first language over time or a clear decrease of condition-first language over time.

Diagram, venn diagram

Description automatically generated

Figure 1 Articles containing condition- and person-first language or both

***Using pejorative weight-emphasising labels for people and their bodies***

…

Given that obese and overweight are overwhelmingly used as weight-emphasising label for people/their bodies, we searched for all instances of these two adjectives based on the search syntax [taglemma=“obese\_ADJ”] and [taglemma=“overweight\_ADJ”] – in other words, the analysis is form-based without any concordance analysis. For fat, where 60% of instances may not be relevant (see Table 5), we cannot rely on such form-based comparison. CB therefore first analysed all concordance lines (a total of 8369 instances, excluding *The Daily Telegraph* and the *Brisbane Times*, given their incomplete coverage over time). The question was: Is *fat* used as weight-emphasising adjectival label for a person? Given the large amount of instances, the categorisation scheme used in CQPweb’s ‘Categorise’ function was simple: yes, no, unclear (REF to coding scheme/notes). Of the 8369 total instances, 2894 were categorised as YES (34.6%), 4907 were categorised as NO (58.6%) and 568 were UNCLEAR (6.8%).

Our statistical analyses (see details in GitHub REF) suggest that the adjective obese is used more frequently in tabloids than in broadsheets (Welch Two Sample t-test; p < .001), with no differences by year. Thus, no change in time can be observed. Topic-based differences in the use of the adjective suggest the need for further research into article content and whether this affects the observed difference between tabloids and broadsheets.

Integrate results from statistical analysis of OVERWEIGHT above

Integrate results from statistical analysis of FAT above