***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). To quantitate use of the two language types, 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. The mean frequency of person-first language across the boostrap-generated subcorpora was 8.23 words per million, while the mean frequency of condition-first language was 284.74 words per million; this difference was statistically significant with a large effect size (see Supplementary Note 1).

|  |  |
| --- | --- |
| **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 full 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). Furthermore, nearly half of articles that use person-first language also use condition-first language as well (Figure 1). Looking at the resampling data, the mean number of articles using person-first language across all subcorpora is 4.03 articles per 1000, while the mean number of articles using condition-first language is 122.54, the difference between which is statistically significant with a large effect size (see Supplementary Note 2).

Our analysis demonstrates that among articles that use either condition-first OR person-first language but not both, the number of articles with only condition-first language is higher in tabloid publications and in right-leaning publications; a similar difference is not observed for person-first language (see Supplementary Note 3). Looking at articles that use either condition-first, person-first language or both reveals that the mean frequency of condition-first language, 4.34 words per 1000, is higher than the mean frequency of person-first language, 2.67 words per 1000 (Supplementary Note 4). Finally, we used linear modelling to consider whether there are differences in the frequency of condition-first language use across years and individual newspapers (Supplementary Note 5). Modelling suggests a difference between newspapers in the frequency of condition-first language and supports the above observation of broadsheets having a lower frequency of use than tabloids; no effect across the time period is observed. This suggests that there has not been a clear decrease of dispreferred condition-first language over time. Many years have no instances of person-first language in some sources, so a similar analysis is impossible for person-first language.

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. Both “obese” (Supplementary note 6) and “overweight” (Supplementary note 7) were used more frequently in tabloids than in broadsheets, with an average of 5.59 and 5.73 instances per 1000 words in tabloids vs 3.16 and 3.55 instances per 1000 words in broadsheets, respectively; these results were also observed when using linear modelling to consider sources individually. These results are attributable in part to tabloids having shorter article lengths in the corpus (Supplementary note 6), irrespective of adjective use, so a similar number of *instances* of use of the adjectives of interest will result in a lower *frequency* in broadsheets than in tabloids. A very subtle decrease in the use of the adjective “obese” was observed over time when using fixed effects linear modelling (beta = -0.01, 95% CI [-0.02, -0.005]), but this was not replicated when using a random effects linear model; a similar very subtle decrease was observed in the use of “overweight” over time (beta = -0.01, 95% CI [-0.02, -0.0066]), supported by both fixed and random effects modelling (Supplementary note 7).

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 number 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%). Tabloids used “fat” as weight-emphasising adjectival label for a person more frequently than broadsheets, as described using both a pairwise comparison between the two classes and considering each source individually, and this effect was, like the case of “overweight” and “obese”, partially attributable to the longer article lengths in broadsheets than tabloids (Supplementary note 8). No differences by year were observed.

Across all the analyses, topic-based differences in the use of the adjectives suggest the need for further research into article content and whether this affects the observed difference between tabloids and broadsheets (see GitHub).

# Supplementary information: Notes on statistical analysis outcomes

## Comparison of condition and person-first language use

### Supplementary Note 1:

A Welch’s Two Sample t-test of 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) in the 10000 instances across the randomly sampled 1000-article subcorpora suggested 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])

### Supplementary Note 2:

A Welch’s Two Sample t-test of the difference between 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])

### Supplementary Note 3:

The Pearson’s Chi-squared test with Yates’ continuity correction contrasting articles from tabloids and broadsheets that use only condition-first vs only person-first language indicate a significant link (X-squared = 4.8274, p-value = 0.02801) between type of publication and number of articles using a specific language type. The effect size is quite small (<0.2), indicating that while the result is statistically significant, the fields are weakly associated.

The Pearson’s Chi-squared test with Yates’ continuity correction contrasting articles from left- and right-leaning publications that use only condition-first vs only person-first language indicate a significant link (X-squared = 4.6405, p-value = 0.03123) between type of publication and number of articles using a specific language type. The effect size is, however, also quite small (<0.2), indicating that while the result is statistically significant, the fields are weakly associated.

### Supplementary Note 4:

A Welch Two Sample t-test testing the difference between the normalised frequency per 1000 words of condition- and person-first language in articles that use either one or both language types (mean of condition-first 4.34 words per 1000, mean of person-first – 2.67 words per 1000) suggests 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]).

### Supplementary Note 5

Consistent with the above binary comparisons of tabloids and broadsheets, simple linear modelling revealed a difference among sources: 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, a tabloid; no difference across the time period could be detected.

More specifically, a linear model of the form log(frequency) ~ scaled\_year + source was fitted and estimated using ML and found to have the lowest AIC among models (see GitHub for additional details of all models considered).

The model’s intercept, corresponding to scaled\_year = 0 and source = Advertiser, is at 1.28 (95% CI [1.21, 1.36], t(3031) = 32.98, p < .001). Within this model:

The effect of scaled year is statistically non-significant and negative (beta = -7.05e-03, 95% CI [-0.02, 2.13e-03], t(3031) = -1.50, p = 0.132; Std. beta = -9.82e-03, 95% CI [-0.02, 3.54e-03])

The effect of source [Age] is statistically significant and negative (beta = -0.45, 95% CI [-0.57, -0.33], t(3031) = -7.43, p < .001; Std. beta = -0.19, 95% CI [-0.25, -0.14])

The effect of source [Australian] is statistically significant and negative (beta = -0.65, 95% CI [-0.79, -0.51], t(3031) = -9.12, p < .001; Std. beta = -0.27, 95% CI [-0.33, -0.21])

The effect of source [CanTimes] is statistically significant and negative (beta = -0.37, 95% CI [-0.50, -0.23], t(3031) = -5.33, p < .001; Std. beta = -0.18, 95% CI [-0.24, -0.12])

The effect of source [CourierMail] is statistically significant and negative (beta = -0.14, 95% CI [-0.25, -0.03], t(3031) = -2.48, p = 0.013; Std. beta = -0.07, 95% CI [-0.12, -0.02])

The effect of source [HeraldSun] is statistically non-significant and negative (beta = -0.06, 95% CI [-0.17, 0.04], t(3031) = -1.18, p = 0.240; Std. beta = -0.05, 95% CI [-0.09, -2.66e-04])

The effect of source [HobMercury] is statistically non-significant and positive (beta = 0.13, 95% CI [-0.02, 0.27], t(3031) = 1.68, p = 0.092; Std. beta = 0.05, 95% CI [-0.01, 0.12])

The effect of source [NorthernT] is statistically non-significant and positive (beta = 0.16, 95% CI [-0.02, 0.35], t(3031) = 1.72, p = 0.086; Std. beta = 0.04, 95% CI [-0.04, 0.12])

The effect of source [SydHerald] is statistically significant and negative (beta = -0.53, 95% CI [-0.64, -0.42], t(3031) = -9.42, p < .001; Std. beta = -0.23, 95% CI [-0.28, -0.19])

The effect of source [WestAus] is statistically non-significant and positive (beta = 0.02, 95% CI [-0.12, 0.15], t(3031) = 0.24, p = 0.808; Std. beta = -0.02, 95% CI [-0.08, 0.04])

Similar results were observed when using more complex mixed effects models: a difference among sources and no effect of year.

## Comparison of use of the adjectives “obese”, “overweight” and “fat”

### Supplementary note 6

The mean frequency per 1000 words of articles where “obese” is used is higher in tabloids than in broadsheets. Based on a Welch’s Two Sample, the mean frequency per 1000 words of articles where “obese” is used was 3.16 in broadsheets and 5.59 in tabloids, which suggests that the effect is negative, statistically significant, and small (difference = -2.43 words per 1000, 95% CI [-2.63, -2.23], t(10062.88) = -24.01, p < .001; Cohen’s d = -0.46, 95% CI [-0.50, -0.42].

This is in part related to the fact that articles that use “obese” are longer in broadsheets than in tabloids. Across articles that used the word “obese”, using a Welch Two Sample t-test to compare the mean word count of articles from broadsheets and tabloids that used “obese” (mean of broadsheet = 778.84, mean of tabloid = 485.42) suggested there was a positive, statistically significant effect, with a medium size (difference = 293.42, 95% CI [273.82, 313.02], t(7348.69) = 29.34, p < .001; Cohen’s d = 0.60, 95% CI [0.56, 0.65]).

We next compared the overall length of all articles in tabloids and broadsheets in the corpus and found that articles in broadsheets were *generally* longer than in tabloids, irrespective of whether any of the adjectives of interest were used. A Welch Two Sample t-test testing the difference between wordcounts in broadsheet and tabloid articles (mean of broadsheets = 810.15 words, mean of tabloids = 502.44 words) suggested that the effect is positive, statistically significant, and medium (difference = 307.71, 95% CI [291.01, 324.42], t(16482.54) = 36.11, p < .001; Cohen’s d = 0.56, 95% CI [0.53, 0.59]).

In order to explore in more depth that the difference in frequency of the word “obese” per 1000 words in tabloids and broadsheets is not fully an artifact of this difference in length, we used bootstrap resampling without replacement (selecting 1000 articles 10000 times) to compare the raw counts of the word “obese” in articles from tabloids and broadsheets, and found that the counts themselves were higher in broadsheets, which had a bimodal distribution of instances of use of the word “obese” – so much of the difference in *frequency* is observed due to this difference in article lengths. The Welch Two Sample t-test testing the difference between resampled broadsheet counts and tabloid counts (mean of broadsheets = 1.71, mean of tabloids = 1.60) suggested that the effect is positive, statistically significant, and large (difference = 0.11, 95% CI [0.11, 0.11], t(18721.64) = 182.01, p < .001; Cohen’s d = 2.57, 95% CI [2.83, 2.83])

Linear modelling by source and year supported the difference between frequency in tabloids and broadsheets, explaining a statistically significant and weak proportion of variance (R2 = 0.10, F(10, 10245) = 109.54, p < .001, adj. R2 = 0.10), with broadsheets considered (the Age, Australian, Canberra Times and Sydney Morning Herald) having lower frequency of use of the word “obese” than the Advertiser (a tabloid), while other tabloids (the Hobart Mercury and Northern Territorian) had a higher frequency of use than the Advertiser. This fixed effects modelling also suggested a slight decrease in the use of the word “obese” with time (beta = -0.01, 95% CI [-0.02, -5.16e-03], t(10245) = -3.96, p < .001; Std. beta = -0.02, 95% CI [-0.02, -8.05e-03]).

Mixed effects modelling supported the observation of higher frequency of counts in tabloids, but the decrease by year was not found to be significant.

### Supplementary note 7

The word “overweight” was used more frequently in tabloids than broadsheets. A Welch’s Two Sample t-test of the difference between the frequency per 1000 words in broadsheets and tabloids across articles that use the word “overweight” (mean of broadsheets = 3.55, mean of tabloids = 5.73 words per 1000) suggests that the effect is negative, statistically significant, and small (difference per 1000 words = -2.18 words, 95% CI [-2.42, -1.95], t(6712.33) = -18.20, p < .001; Cohen’s d = -0.43, 95% CI [-0.48, -0.38]).

Similar to the case of “obese” and the overall corpus statistics (Supplementary note 6), the length of articles that use the word “overweight” is longer in broadsheets than tabloids. The Welch Two Sample t-test testing the difference between the wordcount of articles that use the word “overweight” from broadsheets and tabloids (mean of broadsheets = 732.66, mean of tabloids = 484.51) suggests that the difference is positive, statistically significant, and medium (difference = 248.15, 95% CI [226.60, 269.71], t(4626.50) = 22.57, p < .001; Cohen’s d = 0.58, 95% CI [0.53, 0.63]).

In order to explore in more depth that the difference in frequency of the word “overweight” per 1000 words in tabloids and broadsheets is not fully an artifact of this difference in length, similar to the case of “obese”, we used bootstrap resampling without replacement (selecting 1000 articles 10000 times) to compare the raw counts of the word “overweight” in articles from tabloids and broadsheets, and found that the counts themselves were higher in broadsheets, which had a bimodal distribution of instances of use of the word “overweight” – so much of the difference in *frequency* is observed due to this difference in article lengths. The Welch Two Sample t-test testing the difference between resampled broadsheet counts and tabloid counts (mean of broadsheets = 1.95, mean of tabloids = 1.80) suggests that the effect is positive, statistically significant, and large (difference = 0.15, 95% CI [0.15, 0.15], t(19959.90) = 242.90, p < .001; Cohen’s d = 3.44, 95% CI [3.78, 3.78])

Similar to the analysis of “obese”, linear modelling suggested that the word “overweight” was used less in broadsheet sources (the Age, Australian, Canberra Times and Sydney Morning Herald) than in the Advertiser, and more in the Hobart Mercury and Northern Territorian (both tabloids); a small effect of year was also observed. More specifically, we fitted a linear model (estimated using OLS) to predict frequency with source and scaled\_year (formula: log(frequency) ~ source + scaled\_year). The model explained a statistically significant and weak proportion of variance (R2 = 0.08, F(10, 6710) = 58.73, p < .001, adj. R2 = 0.08). The model’s intercept, corresponding to source = Advertiser and scaled\_year = 0, is at 1.31 (95% CI [1.26, 1.37], t(6710) = 48.02, p < .001). Within this model:

* The effect of source [Age] is statistically significant and negative (beta = -0.44, 95% CI [-0.53, -0.36], t(6710) = -10.23, p < .001; Std. beta = -0.18, 95% CI [-0.22, -0.14])
* The effect of source [Australian] is statistically significant and negative (beta = -0.52, 95% CI [-0.62, -0.43], t(6710) = -10.41, p < .001; Std. beta = -0.21, 95% CI [-0.25, -0.16])
* The effect of source [CanTimes] is statistically significant and negative (beta = -0.29, 95% CI [-0.37, -0.20], t(6710) = -6.52, p < .001; Std. beta = -0.13, 95% CI [-0.17, -0.09])
* The effect of source [CourierMail] is statistically non-significant and positive (beta = 0.04, 95% CI [-0.04, 0.11], t(6710) = 0.95, p = 0.340; Std. beta = 0.01, 95% CI [-0.02, 0.04])
* The effect of source [HeraldSun] is statistically non-significant and positive (beta = 0.03, 95% CI [-0.05, 0.10], t(6710) = 0.72, p = 0.472; Std. beta = 3.54e-03, 95% CI [-0.03, 0.04])
* The effect of source [HobMercury] is statistically significant and positive (beta = 0.17, 95% CI [0.07, 0.26], t(6710) = 3.34, p < .001; Std. beta = 0.06, 95% CI [0.02, 0.10])
* The effect of source [NorthernT] is statistically significant and positive (beta = 0.43, 95% CI [0.31, 0.56], t(6710) = 6.84, p < .001; Std. beta = 0.18, 95% CI [0.13, 0.24])
* The effect of source [SydHerald] is statistically significant and negative (beta = -0.40, 95% CI [-0.48, -0.33], t(6710) = -10.33, p < .001; Std. beta = -0.17, 95% CI [-0.20, -0.13])
* The effect of source [WestAus] is statistically non-significant and positive (beta = 0.03, 95% CI [-0.06, 0.11], t(6710) = 0.62, p = 0.537; Std. beta = -2.62e-03, 95% CI [-0.04, 0.04])
* The effect of scaled year is statistically significant and negative (beta = -0.01, 95% CI [-0.02, -6.59e-03], t(6710) = -4.13, p < .001; Std. beta = -0.02, 95% CI [-0.03, -0.01])

Unlike in the case of “obese”, mixed effects modelling supported both the differences by source type and the slight decrease by year (p < 0.05) (see GitHub).

### Supplementary note 8

A Welch Two Sample t-test testing the difference between the frequency of the use of “fat” as weight-emphasising adjectival label for a person revealed a difference in the frequency per 1000 words of this in broadsheets vs tabloids (mean of broadsheets = 2.62, mean of tabloids = 4.49) suggesting that the effect is negative, statistically significant, and small (difference = -1.87, 95% CI [-2.18, -1.56], t(2251.24) = -11.73, p < .001; Cohen’s d = -0.47, 95% CI [-0.55, -0.39]).

Similar to the case of “obese”, “overweight” and the overall corpus statistics (Supplementary note 6), the length of articles that use the word “fat” in this way is longer in broadsheets than tabloids. The Welch Two Sample t-test testing the difference between wordcount of articles that use “fat” in this way describes a difference between broadsheets and tabloids (mean word count of broadsheets = 922.40, mean of tabloids = 573.96), suggesting that the effect is positive, statistically significant, and medium (difference = 348.45, 95% CI [301.88, 395.01], t(1740.32) = 14.68, p < .001; Cohen's d = 0.70, 95% CI [0.61, 0.80])

Bootstrap re-sampling of the counts revealed that, similar to “obese” and “overweight” raw counts were higher in broadsheets than in tabloids in articles that used “fat” as a weight-emphasizing adjectival label for a person, in part due to more articles in broadsheets having 7+ uses vs only tabloids (23 vs 10 articles, respectively). The Welch Two Sample t-test testing the difference between broadsheet counts and tabloid counts (mean of broadsheets = 1.85, mean of tabloids = 1.48) suggests that the effect is positive, statistically significant, and large (difference = 0.37, 95% CI [0.37, 0.37], t(17733.95) = 491.95, p < .001; Cohen’s d = 6.96, 95% CI [7.65, 7.65]).

Investigating the data by source and year using linear modelling revealed that these variables explained a small amount of variance in the data, with the word fat as weight-emphasising adjectival label for a person being used less frequently in individual broadsheets (Age, Australian, Canberra Times and Sydney Morning Herald) than in the Advertiser (a tabloid), while in the Northern Territorian, another tabloid, the word “fat” was used somewhat more frequently than in the Advertiser. The word “fat” was not used differently in the corpus across the time period.

These results were obtained by fitting a linear model (estimated using OLS) to predict frequency with source and scaled\_year (formula: log(frequency) ~ source + scaled\_year). The model explained a statistically significant and weak proportion of variance (R2 = 0.10, F(10, 2289) = 24.23, p < .001, adj. R2 = 0.09). The model’s intercept, corresponding to source = Advertiser and scaled\_year = 0, is at 1.02 (95% CI [0.92, 1.11], t(2289) = 21.55, p < .001). Within this model:

* The effect of source [Age] is statistically significant and negative (beta = -0.39, 95% CI [-0.53, -0.26], t(2289) = -5.62, p < .001; Std. beta = -0.18, 95% CI [-0.24, -0.12])
* The effect of source [Australian] is statistically significant and negative (beta = -0.52, 95% CI [-0.68, -0.37], t(2289) = -6.66, p < .001; Std. beta = -0.22, 95% CI [-0.29, -0.15])
* The effect of source [CanTimes] is statistically significant and negative (beta = -0.25, 95% CI [-0.41, -0.09], t(2289) = -3.07, p = 0.002; Std. beta = -0.13, 95% CI [-0.20, -0.06])
* The effect of source [CourierMail] is statistically non-significant and positive (beta = 0.05, 95% CI [-0.08, 0.17], t(2289) = 0.70, p = 0.483; Std. beta = -6.33e-03, 95% CI [-0.06, 0.05])
* The effect of source [HeraldSun] is statistically non-significant and positive (beta = 0.12, 95% CI [-7.55e-03, 0.24], t(2289) = 1.84, p = 0.066; Std. beta = 0.03, 95% CI [-0.02, 0.09])
* The effect of source [HobMercury] is statistically non-significant and positive (beta = 0.18, 95% CI [-1.69e-03, 0.35], t(2289) = 1.94, p = 0.052; Std. beta = 0.06, 95% CI [-0.02, 0.14])
* The effect of source [NorthernT] is statistically significant and positive (beta = 0.59, 95% CI [0.40, 0.79], t(2289) = 5.88, p < .001; Std. beta = 0.23, 95% CI [0.15, 0.32])
* The effect of source [SydHerald] is statistically significant and negative (beta = -0.35, 95% CI [-0.48, -0.23], t(2289) = -5.46, p < .001; Std. beta = -0.17, 95% CI [-0.22, -0.11])
* The effect of source [WestAus] is statistically non-significant and positive (beta = 0.06, 95% CI [-0.10, 0.22], t(2289) = 0.69, p = 0.489; Std. beta = -8.61e-03, 95% CI [-0.08, 0.06])
* The effect of scaled year is statistically non-significant and negative (beta = -9.25e-03, 95% CI [-0.02, 1.24e-03], t(2289) = -1.73, p = 0.084; Std. beta = -0.02, 95% CI [-0.03, -9.21e-04])