**Analyses**

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

For the first issue, we compared condition-first language with person-first language. Unlike Brookes and Baker (2021a) we did not limit the search to the phrase *person/people with obesity*, as a previous study of Australian diabetes coverage (Bednarek and Carr 2021) had shown a range of possible human nouns in (dis)preferred structures (see also Halmari 2011, Price 2022 on mental illness). To identify a broad and relevant range of human nouns for our search syntax, we triangulated three corpus linguistic techniques:

1. A collocation analysis of *obese* (right-hand collocates: R1-R5) and of *with obesity* (left-hand collocates L5-L1) to retrieve co-occurring human nouns (Log Ratio [filtered], minimum frequency 10 for both relevant settings), with follow-up qualitative concordance analysis (randomly ‘thinned’ to 100 instances where necessary) to exclude fully irrelevant human nouns (e.g. *researchers*).
2. A concordance analysis of *obese* (used as adjective, analysis of 500 random concordance lines) and *with obesity* (exact phrase), identifying additional relevant human nouns that occur at position R1 for *obese* and at position L1 for *with obesity*
3. A SketchEngine (Kilgarriff et al 2004) Wordsketch of *obese* (used as adjective; ‘nouns modified by’ *obese*) and for *obesity* (used as noun; focussing on *with obesity*)

Results from this triangulation fed into our search syntax (see online appendix), where for each of the identified human nouns, both singular and plural word forms were typically included (except for *peoples*).[[1]](#footnote-1) The forms *those* and *many* were only included in the search for *with obesity*. The search syntax retrieves alternative phrases such as *obese people*, *obese children*, etc. and *people with obesity*, *children with obesity*, etc. Note that this search also retrieves cases where the relevant word form might be used as adjective, such as *obese* ***Australian*** *adults*. We used this search syntax rather than collocation analysis to maximise precision and recall (based on insights from the collocation analysis during the triangulation). Other syntactic structures for condition- or person-first language were not included.

Table 4 clearly demonstrates that dispreferred condition-first language vastly outnumbers person-first language in the corpus (frequency normalised using Python word counts rather than those generated by CQPweb, as the latter includes counts for punctuation, inflating the word count for longer texts; see 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

For the statistical analysis, 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 resampled 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).

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 (Figure 1). Looking at the resampled data, the mean number of articles using person-first language across all subcorpora is 4.03 articles (per 1000) compared to 122.54 for condition-first language – this difference is statistically significant with a large effect size (see Supplementary Note 2).

Diagram, venn diagram

Description automatically generated

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

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 time is observed. This suggests that there has not been a clear decrease of dispreferred condition-first language over time.[[2]](#footnote-2)

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

For the second issue of interest, we needed to identify important (i.e. frequent) weight-emphasising labels for people and their bodies in the corpus. To do so, two research assistants (RAs) independently surveyed a list of the most frequent adjective and noun lemmas in the corpus (starting point: CQPweb’s tagged lemma list) and retrieved any that they judged to be potentially pejorative weight-emphasising labels. The words retrieved by both RAs in this step were obese, big, overweight, fat, large, heavy, fatty, massive, enormous, giant. For each of these adjectives, a qualitative analysis was undertaken of 100 random concordance lines to test whether they do indeed function in the corpus as weight-emphasizing labels for people or their bodies. Table 5 shows that only obese and overweight are overwhelmingly used as such labels (over 90%). Your Fat Friend (2020) suggests that both words are dispreferred: ‘Do not reach for the sharp cruelty of “obese” or “overweight,” which many fat people find harmful, and some consider derogatory.’ As mentioned above, FAT can be both dispreferred and preferred (in reclaimed usage). Overall, 41% of analysed instances of faT were examples of use as weight-emphasising labels, with much lower proportions for the remaining seven adjectives. Therefore, the concordance analysis leads us to focus on OBESE, OVERWEIGHT and FAT. Note that the high total frequencies of OBESE derive from the corpus design criteria and adjectival frequencies should therefore not be compared.

|  |  |  |
| --- | --- | --- |
| **Tagged lemma** | **Raw frequency** | **Use as weight-emphasising label (based on 100 analysed lines)** |
| obese\_ADJ | 17735 | 94% |
| overweight\_ADJ | 13237 | 91% |
| fat\_ADJ | 9619 | 41% |
| large\_ADJ | 5871 | 10% |
| heavy\_ADJ | 2099 | 9% |
| big\_ADJ | 15350 | 7% |
| giant\_ADJ | 439 | 3% |
| massive\_ADJ | 1310 | 2% |
| enormous\_ADJ | 808 | 2% |
| fatty\_ADJ | 1588 | 0% |

Table 5 Adjective lemmas and their use as weight-emphasising label

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. 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). More specifically, CQPweb’s ‘Categorise’ function was used to identify whether FATwas used as weight-emphasising adjectival label for a person (using a simple categorisation scheme: yes, no, unclear – see 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%). Only instances categorised as YES were then included.

Our statistical analyses showed that OBESE, OVERWEIGHT and FAT are all used more frequently in tabloids than in broadsheets, although the effect is partially attributable to the longer article length in broadsheets. Results for differences over time are reported in supplementary note 7, but are either very subtle or inconclusive from a statistical point of view. 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).

1. Not included: group-based human nouns such as *personnel, police, population, nation, country, family, state, generation, group, troop* as well as *human*, *male*, *female* (potential for use with animals). [↑](#footnote-ref-1)
2. A similar analysis is impossible for person-first language due to lack of data for some years/sources. [↑](#footnote-ref-2)