For a different look at the datasets, Table 11 shows the number and proportion of articles with one or more instance of the relevant language type in each subset (some articles may contain more than one type of language).[[1]](#endnote-1) This is also visualised in Figure 1, which shows the proportion of articles with occurrences. Table 11 and Figure 1 demonstrate that a larger proportion of articles appear to be relevant in the Australian corpus for both tabloids and broadsheets.

Table 11 Number and proportion of articles

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Aus tabloids** | **UK tabloids** | **Aus broadsheets** | **UK broadsheets** |
| Person-first | 13 (0.41%) | 29 (0.37%) | 28 (1.27%) | 38 (0.50) |
| Condition-first | 858 (27.08%) | 1519 (19.47%) | 541 (24.48%) | 1279 (16.93) |
| “the obese” (nominal adjectival) | 69 (2.18) | 223 (2.86%) | 62 (2.81%) | 208 (2.75) |

Figure 1 Proportion of articles with occurrences

Chart, bar chart

Description automatically generated

Given the limitations of the log likelihood test, we used three statistical tests to compare the two datasets (Chi-square goodness of fit tests, parametric Welch Two Sample t-tests and non-parametric Fisher-Pitman permutation tests).[[2]](#endnote-2) Results broadly align with Tables 9-10 and are summarized in Table 12, with details available at LINK. Notably, Australian broadsheet articles are longer than UK broadsheet articles (mean 841.70 vs 718.31 words, respectively, p(FP) < 1e-04), but there is no difference in length between tabloid articles (mean 555.04 vs 561.20, respectively).

Table 12 Results from additional statistical tests

|  |  |  |
| --- | --- | --- |
|  | **Tabloid comparison** | **Broadsheet comparison** |
| Person-first language | **no difference** between the number of instances, articles, or frequency | Australian broadsheets have somewhat **higher** total number of instances (p(Chi-sq) < 0.005) and number of articles (p < 0.005) than UK broadsheets. This difference is supported when considering mean frequencies (mean Australian broadsheets = 17.81 ppm, mean UK broadsheets 8.52 ppm) with parametric and non-parametric tests (p <0.05). |
| Condition-first language | There is a **higher** number of instances and articles with at least one instance in Australian tabloids, and a lower number in UK tabloids (p(Chi-sq) < 0.001). The frequency is also higher in Australian tabloids (mean Australian 1315.95 ppm, mean UK 835.37 ppm, p(FP) < 1e-04). This is especially striking given that there is no difference in article length. | There is a **higher** number of instances and articles with at least one instance (p(Chi-sq) < 0.001) of condition-first language in Australian vs UK broadsheets. If we consider the frequency, a very small difference is also detected, with Australian broadsheets having a higher frequency than UK ones (mean Australian 847.96 ppm, mean UK 629.40 ppm, p(FP) < 1e-04). |
| “the obese” (nominal adjectival) | There is a somewhat **lower** number of total instances of use and articles with at least one instance in Australian tabloids, and a higher number in UK tabloids (p(Chi-sq) < 0.05). However, this difference is not supported by frequency analysis (mean Australian 70.46 ppm, mean UK 96.08 ppm), when using both parametric or non-parametric tests. | **No differences** in the number of instances or articles with at least one instance are observed, when considering contingency-table-based analysis. A very small difference is detected (p(t-test) < 0.05, p(FP) < 0.1) when considering frequencies, with Australian broadsheets having a **lower** frequency (mean frequency Australia - 42.65 ppm, UK - 62.96 ppm). |

It thus appears that there is higher use of condition-first language in both Australian broadsheets and tabloids as well as higher use of person-first language in Australian broadsheets. The reasons for this higher use across **both** dispreferred and preferred language practices remain unclear. It could be that there are simply more references to people with obesity in the Australian newspapers or that the UK newspapers use different words (eg *overweight*) or different syntactic structures (e.g. *people who* …) to refer to people with obesity. Future comparison of topic and linguistic differences between the corpora is necessary to investigate this further, as is the potential influence of differences in the news media landscape including use of syndicated material (which results in full or partial duplicates between newspapers).

However, it is clear that regardless of the national context (Australia/UK) or type of newspaper (broadsheet/tabloid), the preference is by far for condition-first language, followed by the nominal adjectival form “‘the obese”, followed by person-first language. These results align with Brookes & Baker (2021)’s finding of ‘little take-up’ (p. 122) of person-first structures in UK newspapers.

1. Total articles in each subset: 2210 Australian broadsheet/3168 tabloid articles; 7556 UK broadsheet/7801 tabloid articles. [↑](#endnote-ref-1)
2. Chi-square tests consider the total sub-corpus word count as the probability of observing a specific feature; t-tests compare the mean frequencies, with texts without the feature considered to have a value of 0; the Fisher-Pitman test (with 10000 re-samples) compares the mean frequencies, and similarly to the t-test texts without the feature are considered to have a frequency of 0. [↑](#endnote-ref-2)