

Kibbitzer 1

Criteria/data: Uncountable or Countable?

By John Swales (November 15, 2002)

One of Tim Johns' most valuable "talking points" in his Kibbitzer series is his discussion of whether the word data, and subsequently the word criteria are singular or plural. Here below are some updates on Kibbitzer 6, drawing upon two small corpora. One is Ken Hyland's corpus of 80 recent research articles, ten each drawn from eight fields ranging from engineering to philosophy; the other is the Research Sub-corpus of the Michigan Corpus of Academic Spoken English (<http://www.hti.umich.edu/micase>). This latter consists of 36 speech-events, covering such genres as colloquia, research groups meetings, advanced graduate classes, and dissertation defenses, drawn from all of the university's four main divisions.

Consider the case of criterion/criteria first. In Hyland, there were 18 instances of criterion, and 37 instances of criteria (almost exactly double). Of these 37, 25 were plural, 7 were indeterminate as to number, but five were singular. In Kibbitzer 6, Johns noted the occurrence of this singular usage in quality newspapers; it now seems that it has begun to spread to research writing. Here are four examples:

1. Surveys of clothes buying behaviour show that quality is the most important criteria followed by fashion and price.
2. It has been argued that the convenience orientation represents a segmentation criteria which transcends national and cultural boundaries.
3. ...the global firm may be able to exploit an evaluative criteria important to individual consumers...
4. ...mutualistic functioning of these associations should be a defining criteria of the term mycorrhiza.

In the MICASE data, criterion does not occur as a head noun but only in compounds such as "criterion-related validity". And rather surprisingly, only one of the 17 instances of criteria can be identified as singular:

5. i would have to have a very stringent criteria for for a moment in this study

The MICASE research sub-corpus findings for data show a slight preference for the singular over the plural (23 instances over 18, with around 150 hard or impossible to categorize). Here are some examples (with minor editing), the plural ones first:

6. the sequence data are available here
7. the first step would be to plot these data
8. this might help give you confidence if all the data seem to fit the curve.
9. do your data really, allow you to say that?
10. huh may need to back up this data
11. i'll show you that data in a minute

- 12. you need less, less data. uh, you can use averages
- 13. there is very little data collected for these databases
- 14. and we have data that shows that those very shaded plantations

Finally, there are two very interesting cases where we get in effect a mixed message as to whether data is singular or plural. Consider:

- 15. there's very few data that's collected so, the first question is
- 16. because this was identifying data, at least mine were

In the first (15), the verbs are singular but the choice of very few suggests that the speaker believes that data is in fact plural. The second case (16) is even more interesting in that the speaker implies that the general data is singular/uncountable, but his or her own data are plural, perhaps somehow suggesting that the latter is more interesting or more relevant!

The findings from the Hyland corpus for data tend to confirm Johns' findings from Nature. Of those that could be identified as to number, 62 (80%) were plural and 15 (20%) were singular, many of the singular ones occurring-as Johns found-in contexts involving computer science. There remains a single occurrence of that "etymological relic", datum:

- 17. ...a datum vis-à-vis a principle applied to it from without,

Further thoughts on these two words to jmswales@umich.edu