

P words in EDICT

Source: <http://sci.tech-archive.net/Archive/sci.lang.japan/2006-06/msg00652.html>

- *From:* jwb@xxxxxxxxxxxxxxxxxxxxxx
 - *Date:* Thu, 29 Jun 2006 03:17:43 GMT
-

As many will know, the EDICT file has a "(P)" marker on about 22,000 entries to indicate they are more commonly used in Japanese than the other 85,000 entries (*). The (P) tag has been generated largely by the presence of those words on study lists, newspaper word rankings, etc. There is also a small number of words that have been hand-marked as (P), e.g. Bn, because they are obviously common.

I am interested in filling in gaps in this class of words, and hence invite people to nominate words which are not currently on the (P) list, but which could be considered for inclusion.

As a sanity check, I would like to include only those words which get over 1M hits on Google (with the Japanese language filter set.)

Probably best to email suggestions to me rather than clutter SLJ with them, unless they are worthy of discussion.

(*) I regularly tip Japanese newspaper articles into WWWJDIC's "Translate Words" function, and it is interesting to note that usually 90%+ of the words in most articles are from the (P) list.

—
Jim Breen <http://www.csse.monash.edu.au/~jwb/>
Clayton School of Information Technology,
Monash University, VIC 3800, Australia
jwb@csse.monash.edu.au