Corpus Comparisons?

What do they do? Given two corpora, outputs a list of words that distinguish the two fields most strongly.

What's so great about them? More than anything else, this tends to be something humanists actually *want*.

What choices do I have to make The algorithm; what

What algorithms are there? Too many, because none of them are especially good.

Dunning Log likelihood gives a probabilistic version that tends to heavily weight common words. The odds ratio gives the probability in one set over another.