



Figure 4. The distribution of the 500 most frequents nouns following the five most frequent pre-nominal adjectives in COCA. The noun frequencies have been log transformed, and are plotted against the overall probability of a given noun being preceded by the adjective in the left panels. The right panels show the relationship between the same set of nouns and the definite article "the," plotted in the same way. The bottom two panels show cumulative plots for all 5 adjectives and all 2500 nouns (left) and the definite article "the" and all 2500 nouns (right). Since entropy rises as a power function if the corpus size from which an upcoming item is to be predicted increases linearly, power fits are shown where appropriate. Accordingly, given that log transformation of frequency makes the frequency distribution linear (Zipf, 1949), to the extent the distribution of nouns following an adjective fits a negative power function, it suggests that an adjective will serve to keep the entropy rate of those nouns constant (by providing more specific information about low-frequency nouns). As can be seen, this appears to be the case for all of the adjectives sampled.



red (169 / million) 200 most frequent nouns

nouns that follow red



Figure 5. The left panels show the distribution of the 200 most frequent nouns following the pre-nominal adjectives ranked 50-4th for frequency in COCA The noun frequencies have been log transformed, and are plotted against the overall probability of a given noun being preceded by the adjective. The right panels show the relationship between the same set of nouns and the definite article "the," plotted in the same way. The bottom two panels show cumulative plots for all 5 adjectives and all 1000 nouns (left) and the definite article "the" and all 1000 nouns (right).



just (54 / million) 200 most frequent nouns

nouns following just



ns following comfortable

 $R^2 = 6E-05$

6



a 0.6

0.4

ded bv 0.5

being

I Jo 0.3

0.2

0.1

0

1

2

3 log frequency

 $R^2 = 0.6841$

6

5

(right).

8

probability of being "modified'

5

.

3

2

1

0

1

2

3 log frequency