**Level 3 partial description**

The first three basewrd lists in this folder consist of families at level 3 partial, the level 3 affixes used being un (not)-, -ly (making adverbs), -er (someone who …; does not include *–or* as in *actor*), -th (only for ordinal numbers) according to Bauer and Nation Level 3 restricted uses. The lists this represents a next step from lemmas, or flemmas (lemmas where a family can contain different parts of speech). The lists are frequency sorted on family frequency using the 14 million corpus made of 14 one million subcorpora. For learners who can handle inflections, these four derivational affixes should not be too big a step and could easily be the focus of a small amount of deliberate teaching and learning.

Basewrds4 to 7 consist of words that are rather different from those counted as target words in a language course Basewrd4 contains a large list of proper names. Basewrd5 contains marginal words such as affixes, swear words, exclamations etc. Basewrd6 contains transparent compounds and basewrd7 contains acronyms. For more detail on all these lists see Nation, I.S.P. (2016). *Making and Using Word Lists for Language Learning and Testing*. Amsterdam: John Benjamins.

**Headwords**

The headwords of the Level 3 partial 1st 1000 change only slightly from the BNC Level 6 1st 1000 headwords with 15 words having their stem form replaced as the headword by a more frequent derived form of the stem (*beautiful, education, government, information, insurance, security, activity, claimant, election, financial, organization, performance, political, analysis, republican*), and with 20 having both the stem and a derived form or in three cases two derived forms as headwords (*state/states/statement, employ/employment/employee, relate/relation/relationship, act/action, blood/bloody, care/careful, centre/central, fun/funny, manage/management, nature/natural, nation/national, out/outside, person/personal, wonder/wonderful, argue/argument, develop/development, economy/economic*).

It is worth noting that the BNC 1st 1000 at Level 6 contains some words that are lower than the frequency required to get into the 1st 1000 but which are part of the survival vocabulary and thus should be learned in the first few hours of English classes, for example *delicious, thirsty*. There are a few items that complete lexical sets - *autumn*

The 2nd 1000 of the Level 3 partial contains over 400 words that are not headwords in the BNC Level 6 2nd 1000 but which are derived forms of BNC Level 6 1st 1000 and 2nd 1000 headwords (*addition, additional, agreement, appearance, arrangement, basically*), or are stem forms that were in the BNC 1st 1000 (*amaze, angry, apart, arrive*). This means that around 400 stem form headwords move out of the BNC 2nd 1000 (e.g. *ace, adventure, alcohol, alter, angel, anxious, atmosphere*). This is a big change.

In the second 1000 Level 3 partial list, there are 75 other words that are derived forms where the stem form or a related form is not in the second 1000 list - *able, after, agree, appear, arrange, consider, cover, difficult, explain, free, fright, good, govern, grandfather, grow, history, hope, hunger, judge, marry, member, move, pay, point, power, real, reason, safe, serve, sex, through, use, when, write, apply, assist, commerce, compare, connect, defence, directed, economy, exist, identify, industry, intent, investigate, minor, operate, percent, president, product, punish, react, represent, require, wed, academy, communicate, confer, distribute, inflate, interpret, legislate, liable, method, oblige, regulate, transact*.

**Text coverage**

When word lists are used for text analysis, it is useful to see how the choice of a different level of word family affects the results. Table 1 (from my book *Making and Using Word Lists*) compares the coverage of 14 million tokens of spoken and written data by different units of counting, from single word types (Bauer & Nation Level 1), Level 2 lemmas where a lemma can include more than one part of speech (flemmas), Level 3 families using inflections plus *-ly, -er, un-, -th* (for ordinal numbers), and Level 6 families.

Table 1: Percentage coverage of 14 million words of spoken and written data by the first 3000 types, Level 2 flemmas, Level 3 partial families and Level 6 families

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Types | Flemmas | Level 3 partial families | Level 6 families |
| 1st 1000 | 76.46 | 80.97 | 81.67 | 82.95 |
| 2nd 1000 | 5.45 | 5.29 | 5.28 | 5.36 |
| 3rd 1000 | 2.88 | 2.41 | 2.32 | 2.07 |
| Total | 84.79 | 87.77 | 89.27 | 90.38 |

Note in Table 1 that the difference in coverage between Level 3 partial and Level 6 is just over 1% for the first 3000.