British, adult, formal, informative language and more English learners in
reason is that the British National Corpus (BNC) is predominantly a corpus of
writings or spoken language in primary or secondary school systems. The
use these treatments as the basis for syllabus design for learners of English,
using the NNC corpus and shows that it is not appropriate to
make use of the British National Corpus and shows that it is not appropriate to
return to the people whose needs are being investigated (Nation 2001).

This paper looks at high frequency words developed from a variety of
classes as well as how those needs are written into the BNC (Nation 2001).

Chapter 1

A study of the most frequent word families in

Abstract

Victoria University of Wellington
Paul Nation

The British National Corpus

A study of the most frequent word families in

Chapter 1
A study of the most frequent word families in the British National Corpus.

Table 1. Two samples of articles from the December, 1989 London Review of Books.

<table>
<thead>
<tr>
<th>Word</th>
<th>Frequency</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>The</td>
<td>100,000</td>
<td>1</td>
</tr>
<tr>
<td>a</td>
<td>50,000</td>
<td>2</td>
</tr>
<tr>
<td>an</td>
<td>25,000</td>
<td>3</td>
</tr>
<tr>
<td>and</td>
<td>10,000</td>
<td>4</td>
</tr>
<tr>
<td>of</td>
<td>1,000</td>
<td>5</td>
</tr>
<tr>
<td>for</td>
<td>500</td>
<td>6</td>
</tr>
<tr>
<td>in</td>
<td>400</td>
<td>7</td>
</tr>
<tr>
<td>to</td>
<td>300</td>
<td>8</td>
</tr>
<tr>
<td>it</td>
<td>200</td>
<td>9</td>
</tr>
<tr>
<td>is</td>
<td>100</td>
<td>10</td>
</tr>
</tbody>
</table>

The British National Corpus consists of 30 million words, of which 5% is the British variety.

The table shows the most common words in the British National Corpus.
A study of the most frequent word families in the British National Corpus.
Table 3. Percentage of word families in the text occurring in various corpora

<table>
<thead>
<tr>
<th>Corpus</th>
<th>PIND</th>
<th>1.2</th>
<th>2.0</th>
<th>4.0</th>
<th>0.9</th>
<th>1.9</th>
<th>3.9</th>
<th>9.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>LINDA</td>
<td>12%</td>
<td>88%</td>
<td>13%</td>
<td>93%</td>
<td>9%</td>
<td>88%</td>
<td>93%</td>
<td>99%</td>
</tr>
<tr>
<td>10%</td>
<td>9%</td>
<td>88%</td>
<td>13%</td>
<td>93%</td>
<td>9%</td>
<td>88%</td>
<td>93%</td>
<td>99%</td>
</tr>
<tr>
<td>1000</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>10000</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>

Table 2. Cumulative percentage coverage of a range of corpora by the text from the text

<table>
<thead>
<tr>
<th>Corpus</th>
<th>PIND</th>
<th>1.2</th>
<th>2.0</th>
<th>4.0</th>
<th>0.9</th>
<th>1.9</th>
<th>3.9</th>
<th>9.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>LINDA</td>
<td>12%</td>
<td>88%</td>
<td>13%</td>
<td>93%</td>
<td>9%</td>
<td>88%</td>
<td>93%</td>
<td>99%</td>
</tr>
<tr>
<td>10%</td>
<td>9%</td>
<td>88%</td>
<td>13%</td>
<td>93%</td>
<td>9%</td>
<td>88%</td>
<td>93%</td>
<td>99%</td>
</tr>
<tr>
<td>1000</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>10000</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>

Table 3. Percentage of word families in the text occurring in various corpora.

6. Does the BNC 3000 contain most of the GSL plus AVM?

Less than half (49%) of the BNC 3000 word families are found in each of the three corpora. The first 3 years show that the BNC 3000 word families are not evenly distributed across the corpora. The second 3 years show that the BNC 3000 word families are not evenly distributed across the corpora. The third 3 years show that the BNC 3000 word families are not evenly distributed across the corpora. The fourth 3 years show that the BNC 3000 word families are not evenly distributed across the corpora.

Because of this, Table 3 focuses on the words that occur in all corpora. Therefore, the information in Table 3 should be used to judge the performance of the corpora. These results indicate that the GSL plus AVM contains the majority of the words that occur in the corpora. Therefore, the results indicate that the GSL plus AVM contains the majority of the words that occur in the corpora.

The results indicate that the GSL plus AVM contains the majority of the words that occur in the corpora. Therefore, the results indicate that the GSL plus AVM contains the majority of the words that occur in the corpora.

8. Study of the most frequent words in the English National Corpus

A study of the most frequent words in the English National Corpus.
The following twenty-four words are all words in the BNC but not in the AVL:

A study of the most frequent word families in the British National Corpus

Table 5. Possible reasons for non-overlapping words in the GSL and BNC

Table 4. Spread of the 10 families of the AVL across the BNC

Not in AVL

(A) 10

BNC

(A) 10

2nd rank

1st rank

BNC

1st rank

2nd rank

AVL

1st rank

2nd rank

Table 3. The BNC 2000 coverage of formal words vs. the BNC 1990

Table 2. The AVL as a test of the BNC

Table 1. The BNC and composition of the BNC

The BNC consists of a large diary of information extracted from print media, which is regularly updated to reflect changes in the language. The BNC contains a large number of examples of the most frequent word families in the English language. The AVL is a list of the most frequent word families in the English language, which is used for various purposes, including language analysis and teaching.

Not in BNC

(A) 10

BNC

(A) 10

2nd rank

1st rank

BNC

1st rank

2nd rank

AVL

1st rank

2nd rank

Table 3. The BNC 2000 coverage of formal words vs. the BNC 1990

Table 2. The AVL as a test of the BNC

Table 1. The BNC and composition of the BNC

The BNC consists of a large diary of information extracted from print media, which is regularly updated to reflect changes in the language. The BNC contains a large number of examples of the most frequent word families in the English language. The AVL is a list of the most frequent word families in the English language, which is used for various purposes, including language analysis and teaching.

Not in BNC

(A) 10

BNC

(A) 10

2nd rank

1st rank

BNC

1st rank

2nd rank

AVL

1st rank

2nd rank

Table 3. The BNC 2000 coverage of formal words vs. the BNC 1990

Table 2. The AVL as a test of the BNC

Table 1. The BNC and composition of the BNC

The BNC consists of a large diary of information extracted from print media, which is regularly updated to reflect changes in the language. The BNC contains a large number of examples of the most frequent word families in the English language. The AVL is a list of the most frequent word families in the English language, which is used for various purposes, including language analysis and teaching.
The main motivation to replace the General Service Test is because of the
large number of words (993) in the BNC 3000 and the CST and AMIC.

Reforms

The number of words covered by the BNC 3000 and the CST and AMIC
does not include

References


University Press. Cambridge.


Lindsey, W. 1995. The British National Corpus: OUP.

