```
*** Linguistic Indexes also take up more space than binary indexes


*** Create a simple table with a few values ...

SQL> CREATE TABLE bowies_table (id NUMBER, name VARCHAR2(20));

Table created.

SQL> INSERT INTO bowies_table VALUES (1, 'Bowie');

1 row created.

SQL> INSERT INTO bowies_table VALUES (2, 'BOWIE');

1 row created.

SQL> INSERT INTO bowies_table VALUES (3, 'BoWiE');

1 row created.

SQL> INSERT INTO bowies_table VALUES (4, 'THIN WHITE DUKE');

1 row created.

SQL> COMMIT;

Commit complete.


*** Create a normal binary index. Note database uses ASCII based
WE8MSWIN1252 characterset

SQL> CREATE INDEX binary_bowie_idx ON bowies_table(name);

Index created.


*** Create a GENERIC_M_CI linguistic index

SQL> CREATE INDEX ling_bowie_idx ON bowies_table(NLSSORT(name,
'NLS_SORT=GENERIC_M_CI'));

Index created.


*** Create a BINARY_CI linguistic index

SQL> CREATE INDEX ling_bowie_idx2 ON bowies_table(NLSSORT(name,
'NLS_SORT=BINARY_CI'));

Index created.


*** Perform a Select to dump the raw decimal representation of the ASCII
values and the length of the column

SQL> select dump(name) from bowies_table;
```

```
DUMP(NAME)
--------------------------------

Typ=1 Len=5: 66,111,119,105,101
Typ=1 Len=5: 66,79,87,73,69
Typ=1 Len=5: 66,111,87,105,69
Typ=1 Len=15: 84,72,73,78,32,87,72,73,84,69,32,68,85,75,69


*** Perform a block dump of the Binary Index

SQL> select file_id, block_id from dba_extents where
segment_name='BINARY_BOWIE_IDX';

   FILE_ID   BLOCK_ID
---------- ----------
        18       6153

SQL> alter system dump datafile 18 block 6154;

System altered.


*** Partial dump follows

Leaf block dump
===============
header address 129901148=0x7be225c
kdxcolev 0
KDXCOLEV Flags = - - -
kdxcolok 0
kdxcoopc 0x80: opcode=0: iot flags=--- is converted=Y
kdxconco 2
kdxcosdc 0
kdxconro 4
kdxcofbo 44=0x2c
kdxcofeo 7966=0x1f1e
kdxcoavs 7922
kdxlespl 0
kdxlende 0
kdxlenxt 0=0x0
kdxleprv 0=0x0
kdxledsz 0
kdxlebksz 8036
row#0[8021] flag: ------, lock: 0, len=15
col 0; len 5; (5):  42 4f 57 49 45                    <=== 'BOWIE' index
entry value with a length of 5 as expected
col 1; len 6; (6):  04 80 19 0a 00 01                 <=== 6 byte
corresponding ROWID
row#1[8006] flag: ------, lock: 0, len=15
col 0; len 5; (5):  42 6f 57 69 45                    <=== Note this
'BoWiE' has differing ASCII values to the other 2
col 1; len 6; (6):  04 80 19 0a 00 02
row#2[7991] flag: ------, lock: 0, len=15
col 0; len 5; (5):  42 6f 77 69 65                    <=== Note this
'Bowie' also has differing ASCII values to the other 2
col 1; len 6; (6):  04 80 19 0a 00 00
row#3[7966] flag: ------, lock: 0, len=25
col 0; len 15; (15):  54 48 49 4e 20 57 48 49 54 45 20 44 55 4b 45    <===
"THIN WHITE DUKE" index entry with a length of 15 as expected
```

```
col 1; len 6; (6):  04 80 19 0a 00 03
----- end of leaf block dump -----


*** Note the column values are a HEX representation of the ASCII values
(eg: 66 decimal = 42 hex)

*** Note also as expected that the length of the column index values
correspond exactly with the length of the column values




*** Let's now look at the GENERIC_M_CI Linguistic Index


SQL> select file_id, block_id from dba_extents where
segment_name='LING_BOWIE_IDX';

   FILE_ID   BLOCK_ID
---------- ----------
        18       6281

SQL> alter system dump datafile 18 block 6282;

System altered.

Leaf block dump
===============
header address 129901148=0x7be225c
kdxcolev 0
KDXCOLEV Flags = - - -
kdxcolok 0
kdxcoopc 0x80: opcode=0: iot flags=--- is converted=Y
kdxconco 2
kdxcosdc 0
kdxconro 4
kdxcofbo 44=0x2c
kdxcofeo 7904=0x1ee0
kdxcoavs 7860
kdxlespl 0
kdxlende 0
kdxlenxt 0=0x0
kdxleprv 0=0x0
kdxledsz 0
kdxlebksz 8036
row#0[8009] flag: ------, lock: 0, len=27
col 0; len 17; (17):  01 ef 02 38 02 64 02 1b 01 fe 00 00 02 02 02 02 02
<=== "BOWIE" value is now 17 (not 5) bytes in length
col 1; len 6; (6):  04 80 19 0a 00 00
row#1[7982] flag: ------, lock: 0, len=27
col 0; len 17; (17):  01 ef 02 38 02 64 02 1b 01 fe 00 00 02 02 02 02 02
<=== Note all Bowie columns have the same values
col 1; len 6; (6):  04 80 19 0a 00 01
row#2[7955] flag: ------, lock: 0, len=27
col 0; len 17; (17):  01 ef 02 38 02 64 02 1b 01 fe 00 00 02 02 02 02 02
col 1; len 6; (6):  04 80 19 0a 00 02
row#3[7904] flag: ------, lock: 0, len=51
col 0; len 41; (41):
 02 55 02 13 02 1b 02 32 02 64 02 13 02 1b 02 55 01 fe 01 f7 02 5b 02 23 01
```

```
 fe 00 00 02 02 02 02 02 02 02 02 02 02 02 02 02
<=== "THIN WHITE DUKE" value is now 41 (not 15) bytes in length
col 1; len 6; (6):  04 80 19 0a 00 03
----- end of leaf block dump -----
```

*** Note the column lengths have increased substantially and are basically 3 X binary length of non space characters + 2 bytes

*** Finally, let's look at the BINARY_CI Linguistic Index

```
SQL> select file_id, block_id from dba_extents where
segment_name='LING_BOWIE_IDX2';

   FILE_ID   BLOCK_ID
---------- ----------
        18      28425

SQL> alter system dump datafile 18 block 28426;

System altered.
```

```
Leaf block dump
===============
header address 170599004=0xa2b225c
kdxcolev 0
KDXCOLEV Flags = - - -
kdxcolok 0
kdxcoopc 0x80: opcode=0: iot flags=--- is converted=Y
kdxconco 2
kdxcosdc 0
kdxconro 4
kdxcofbo 44=0x2c
kdxcofeo 7962=0x1f1a
kdxcoavs 7918
kdxlespl 0
kdxlende 0
kdxlenxt 0=0x0
kdxleprv 0=0x0
kdxledsz 0
kdxlebksz 8036
row#0[8020] flag: ------, lock: 0, len=16
col 0; len 6; (6):  62 6f 77 69 65 00
<=== 'BOWIE' value is now 6 (not 5) bytes
col 1; len 6; (6):  04 80 19 0a 00 00
row#1[8004] flag: ------, lock: 0, len=16
col 0; len 6; (6):  62 6f 77 69 65 00
<=== Not all Bowie columns have the same values
col 1; len 6; (6):  04 80 19 0a 00 01
row#2[7988] flag: ------, lock: 0, len=16
col 0; len 6; (6):  62 6f 77 69 65 00
col 1; len 6; (6):  04 80 19 0a 00 02
row#3[7962] flag: ------, lock: 0, len=26
col 0; len 16; (16):  74 68 69 6e 20 77 68 69 74 65 20 64 75 6b 65 00
<=== 'THIN WHITE DUKE' in now 16 (not 15) bytes
col 1; len 6; (6):  04 80 19 0a 00 03
```

```
----- end of leaf block dump -----
End dump data blocks tsn: 21 file#: 18 minblk 28426 maxblk 28426


*** Binary_CI is basically one extra byte per index column value.
```