Appendix:

1. Making your own nexus files:

If you want to generate a dataset for your own analysis, you need to create a data file like this somehow. Programs like `Mesquite` (http://mesquiteproject.org/) provide an interface that might be helpful. Or you could use something like `Excel` to layout the matrix part, and then cut and paste into a text file to add the extra stuff. For python fans, I've written a library, `python_nexus` (https://github.com/SimonGreenhill/python-nexus) which provides a simple API to generate nexus files:

```python
from nexus import NexusWriter

nex = NexusWriter()
nex.add('English', 'char1', 1)
nex.add('French', 'char1', 0)
nex.add('German', 'char1', 1)
print(nex.write())
```

```
#NEXUS

BEGIN DATA;
    DIMENSIONS NTAX=3 NCHAR=1;
    FORMAT DATATYPE=STANDARD MISSING=? GAP=-  SYMBOLS="01";

MATRIX
English    1
French     0
German     1
;
END;
```

Ascertainment Correction.

One problem with most linguistic and cultural data is that researchers tend not to collect data that doesn't vary. This is a form of *sampling bias* that is often called *ascertainment bias*. We know that this ascertainment bias is a problem – Lewis ('01) showed that if we don't account for it, then the branch-lengths can be substantially over-estimated as only variable sites are in the data. This over-estimation will have flow-on effects to rate and age estimates, and may influence the tree topology too.

How do we deal with it? `BEAST 2` thankfully has a correction built into the likelihood calculation, and the language templates in `Babel` are set up to use it automatically. **However** you must do one thing to your data: add a single character at the start of the nexus file that is all zero, e.g.:

```
#NEXUS

BEGIN DATA;
    DIMENSIONS NTAX=3 NCHAR=1;
    FORMAT DATATYPE=STANDARD MISSING=? GAP=-  SYMBOLS="01";

MATRIX
English    0(.....etc)
French     0(.....etc)
German     0(.....etc)
;
END;
```

… and `BEAST 2` will correct the likelihood appropriately.