1 TeX distributions

1.1 TeX Live

In TeX Live you can install different language collections `collection-lang<foo>`.
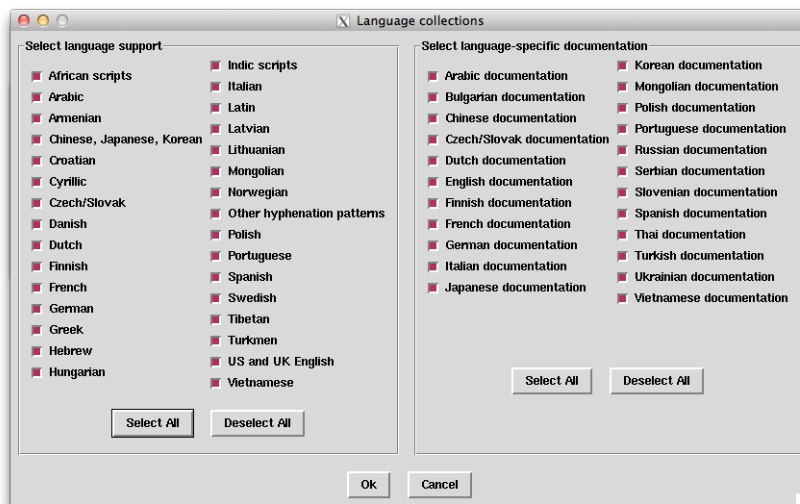

Figure 1

For example, the German collection `collection-langgerman` contains packages `hyphen-german` and `dehyph-exptl` which both contain hyphenation patterns (in addition to many other packages related to German language). The Greek collection `collection-langgreek` contains `hyphen-greek` and `hyphen-ancientgreek` among other packages. The package `dehyph-exptl` is maintained independently, while all packages `hyphen-<foo>` belong to `hyph-utf8`.

Hyphenation packages may contain one or more patterns. For example `hyphen-norwegian` contains two directives for adding two hyphenation patterns:

```
execute AddHyphen \
       name=bokmal synonyms=norwegian,norsk \
       lefthyphenmin=2 \
       righthyphenmin=2 \
       file=loadhyph-nb.tex \
       file_patterns=hyph-nb.pat.txt \
       file_exceptions=hyph-nb.hyp.txt
execute AddHyphen \
       name=nynorsk \
       lefthyphenmin=2 \
       righthyphenmin=2 \
       file=loadhyph-nn.tex \
       file_patterns=hyph-nn.pat.txt \
       file_exceptions=hyph-nn.hyp.txt
```

The keywords **name** and **synonyms** are used as language name to access the patterns in TeX, **file** is used as pattern loader for 8-bit engines and XTeX, while **file_patterns** and **file_exceptions** are plain text files used by LuaTeX.

The information from `AddHyphen` directives end up in three different files: `language.dat` used by LᴬTᴇX, `language.def` used by plain ε-TᴇX and `language.dat.lua` used by LuaTᴇX. All the three are placed in
`$TEXMFSYSVAR/tex/generic/config,`
for example to
`/usr/local/texlive/2011/texmf-var/tex/generic/config,`
while a static copy also exists in
`$TEXMFMAIN/tex/generic/config,`
but that one contains all languages and only makes sense if you have a complete checkout from SVN repository for example.

```
% from hyphen-norwegian:
bokmal loadhyph-nb.tex
=norwegian
=norsk
nynorsk loadhyph-nn.tex
```

Figure 2 `language.dat`

```
% from hyphen-norwegian:
\addlanguage{bokmal}{loadhyph-nb.tex}{}{2}{2}
\addlanguage{norwegian}{loadhyph-nb.tex}{}{2}{2}
\addlanguage{norsk}{loadhyph-nb.tex}{}{2}{2}
\addlanguage{nynorsk}{loadhyph-nn.tex}{}{2}{2}
```

Figure 3 `language.def`

```
-- from hyphen-norwegian:
      ['bokmal'] = {
            loader = 'loadhyph-nb.tex',
            lefthyphenmin = 2,
            righthyphenmin = 2,
            synonyms = { 'norwegian', 'norsk' },
            patterns = 'hyph-nb.pat.txt',
            hyphenation = 'hyph-nb.hyp.txt',
      },
      ['nynorsk'] = {
            loader = 'loadhyph-nn.tex',
            lefthyphenmin = 2,
            righthyphenmin = 2,
            synonyms = {  },
            patterns = 'hyph-nn.pat.txt',
            hyphenation = 'hyph-nn.hyp.txt',
      },
```

Figure 4 `language.dat.lua` (the keyword `loader` is not used anywhere at the moment)

The files `language.dat` and `language.def` are used both by 8-bit and native Unicode engines. On one hand this simplifies things, while on the other it makes very little sense to include sanskrit patterns in pdfTeX – the package doesn't load any patterns in that case any way. It might be that there will be multiple copies of `language.dat` for different engines in future.

In addition to the three mentioned files, TeX Live also provides two additional files for pTeX:

- `$TEXMFDIST/tex/ptexgeneric/config/language.ptx` as an equivalent of `language.dat` which enables all languages that make sense in an 8-bit engine
- `$TEXMFDIST/tex/ptex/config/language.def` which enables just most frequently used languages

The contents are almost the same as in the generated `language.dat` and `language.def`, except that these two files are not generated and might enable a different set of languages, independent of installed language collections. When pTeX was first included into TeX Live it was loading its own files with patterns, so it needed its own `language.ptx`. Nowadays this restriction is gone and one could replace both static files with the generated ones if needed.

TODO: explain where are the sources

1.2   MiKTeX

In MiKTeX application **MiKTeX Options** which you can run from Start menu (*Maintenance → Settings* or *Maintenance (Admin) → Settings (Admin)*) or from command line (`mo.exe` or `mo_admin.exe`) you can select which languages you want to include into TeX formats.
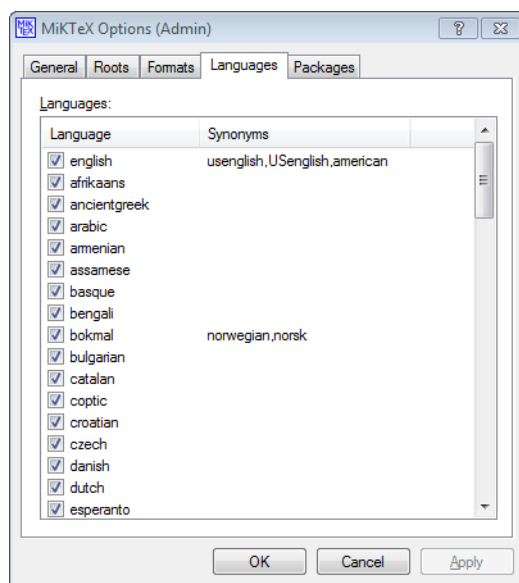

Figure 5   user interface to
select the desired languages

Based on that selection MiKTeX creates three files `language.dat`, `language.def` and `language.dat.lua` in directory

```
C:\ProgramData\MiKTeX\2.9\tex\generic\config
```

The contents of those files are the same as described in TeX Live section, but the needed information about names of languages and required files comes from

```
C:\Program Files\MiKTeX 2.9\MiKTeX\config\languages.ini
```

instead of special `hyphen-<foo>` packages.

(Exact location might depend on where you install MiKTEX and whether you install it for all users or just for yourself. It may also depend on Windows version. The information above holds for MiKTEX 2.9 on Windows 7 with system-wide installation.)

```
[bokmal]
loader=loadhyph-nb.tex
lefthyphenmin=2
righthyphenmin=2
synonyms=norwegian,norsk
patterns=hyph-nb.pat.txt

[nynorsk]
loader=loadhyph-nn.tex
lefthyphenmin=2
righthyphenmin=2
patterns=hyph-nn.pat.txt
```

Figure 6 `languages.ini`

1.3 W32TEX

List of languages in TeX Live (has to be nicer):

```
collection-langafrican
        hyphen-ethiopic
                ethiopic,amharic,geez
collection-langarabic
        hyphen-arabic
                arabic
        hyphen-farsi
                farsi,persian
collection-langcjk
        hyphen-chinese
                pinyin
collection-langcroatian
        hyphen-croatian
                croatian
collection-langcyrillic
        hyphen-bulgarian
                bulgarian
        hyphen-russian
                russian
        hyphen-ukrainian
                ukrainian
collection-langczechslovak
        hyphen-czech
                czech
        hyphen-slovak
                slovak
collection-langdanish
        hyphen-danish
                danish
collection-langdutch
        hyphen-dutch
                dutch
collection-langenglish
        hyphen-english
                ukenglish,british,UKenglish
                usenglishmax
collection-langfinnish
        hyphen-finnish
                finnish
collection-langfrench
        hyphen-basque
                basque
        hyphen-french
                french,patois,francais
collection-langgerman
        hyphen-german
                german
                ngerman
                swissgerman
```

```
collection-langgreek
        hyphen-greek
                monogreek
                greek,polygreek
        hyphen-ancientgreek
                ancientgreek
                ibycus
collection-langhungarian
        hyphen-hungarian
                hungarian
collection-langindic
        hyphen-indic
                assamese
                bengali
                gujarati
                hindi
                kannada
                malayalam
                marathi
                oriya
                panjabi
                tamil
                telugu
        hyphen-sanskrit
                sanskrit
collection-langitalian
        hyphen-italian
                italian
collection-langlatin
        hyphen-latin
                latin
collection-langlatvian
        hyphen-latvian
                latvian
collection-langlithuanian
        hyphen-lithuanian
                lithuanian
collection-langmongolian
        hyphen-mongolian
                mongolian
                mongolianlmc
collection-langnorwegian
        hyphen-norwegian
                bokmal,norwegian,norsk
                nynorsk
collection-langother
        hyphen-afrikaans
                afrikaans
        hyphen-armenian
                armenian
```

```
        hyphen-coptic
                coptic
        hyphen-esperanto
                esperanto
        hyphen-estonian
                estonian
        hyphen-icelandic
                icelandic
        hyphen-indonesian
                indonesian
        hyphen-interlingua
                interlingua
        hyphen-irish
                irish
        hyphen-kurmanji
                kurmanji
        hyphen-lao
                lao
        hyphen-romanian
                romanian
        hyphen-serbian
                serbian
                serbianc
        hyphen-slovenian
                slovenian,slovene
        hyphen-turkish
                turkish
        hyphen-uppersorbian
                uppersorbian
        hyphen-welsh
                welsh
collection-langpolish
        hyphen-polish
                polish
collection-langportuguese
        hyphen-portuguese
                portuguese,portuges
collection-langspanish
        hyphen-spanish
                spanish,espanol
        hyphen-catalan
                catalan
        hyphen-galician
                galician
collection-langswedish
        hyphen-swedish
                swedish
collection-langturkmen
        hyphen-turkmen
                turkmen
```