Apertium - An Open-Source Platform for Automatic Translation

António Silva, Rui Brito

Universidade do Minho

Braga, Março 2013

Index

What it is

- Machine Translation Engine from the OpenTrad project
- Designed to translate closely related languages
  - Recently expanded to support more divergent languages (e.g. es-en)
- Dictionaries come in language pairs
  - Well defined XML format
- Works with lttoolbox
  - Toolbox for lexical processing, morphological analysis and word generation

Install dependencies

autoconf automake expat flex gettext gperf help2man libiconv libtool libxml2 libxslt m4 ncurses p5-locale-gettext pcre perl5 pkgconfig zlib gawk subversion

Get sources

```
$ svn co http://apertium.svn.sourceforge.net/
    svnroot/apertium/trunk/lttoolbox
$ svn co http://apertium.svn.sourceforge.net/
    svnroot/apertium/trunk/apertium
```

Get Dictionary Pairs

```
$ svn co http://apertium.svn.sourceforge.net/
    svnroot/apertium/trunk/apertium-es-pt
```

Full list of dictionary pairs available at: `http://wiki.apertium.org/wiki/List_of_dictionaries`

Setting the environment

```
$ PKG_CONFIG_PATH=/usr/local/lib/pkgconfig
$ export PKG_CONFIG_PATH
```

For each downloaded package do:

Compilation

```
$ ./autogen.sh
$ make
$ sudo make install
```

Sample XML:

```
<pardefs>
    <pardef n="numerals">
      <e>
        <p>
          <l>uno<s n="num"/></
              l>
          <r>um<s n="num"/></r
              >
        </p>
      </e>
      <e>
        <p>
          <l>uno<s n="det"/></
              l>
          <r>um<s n="det"/></r
              >
        </p>
      </e>
```

With lt-expand:

```
estar<vblex>:estar<vblex>
decir<vblex>:dizer<vblex>
tener<vblex>:ter<vblex>
año<n>:ano<n>
país<n>:país<n>
obrar<vblex>:>:fazer<vblex>
.
.
.

primero<adj>:primeiro<adj>
benéfico<adj>:benéfico<adj>
primer<adj>:>:primeiro<adj>
Irak<np>:Iraque<np>
```

Basic Usage

```
$ apertium language-pair file
```

e.g.

```
$ echo "Buenas␣tardes" apertium es-pt
```

Will produce the output:
Boas Tardes

After compilation of the dictionary we can analyse possible lexical forms.

Lexical Analysis

```
$ lt-proc morf-file file|word
```

Example

```
$ echo "prova" | lt-proc pt-es.automorf.bin
```

Will produce the following output:

```
^prova/prova<n><f><sg>/provar<vblex><imp><p2><sg
   >/provar<vblex><pri><p3><sg>/provir<vblex><
   imp><p3><sg>/provir<vblex><prs><p1><sg>/
   provir<vblex><prs><p3><sg>$
```

Apertium - An Open-Source Platform for Automatic Translation

António Silva, Rui Brito

Universidade do Minho

Braga, Março 2013

– ? –