1. Background:

- There is an ever-increasing interest in exploring the roles of frequency and usage in understanding phonological phenomena (e.g., Bybee 2001, Ernestus 2011, Fritsch 2011, Archangeli & Pulleyblank 2013, Hume et al. to appear).
- Many corpora and analysis tools are focused on dialogue- and sentence-level analysis, and/or the skills needed to efficiently handle large corpora can be daunting to learn.
- Researchers may be using different instantiations of the “same” formulae.
- PCT is an analysis aid for researchers who are specifically interested in investigating the relationships that hold between individual sounds (segments, features) in a linguistic corpus.
- PCT is designed with the phonologist in mind and has an easy-to-use graphical user interface that requires no programming knowledge, though all of the original code (Python 3.3) is freely available for those who would like access to the source.

2. Functionality:

- **Corpus creation**: Ability to turn a (transcribed) text into a frequency-tagged corpus for subsequent analysis.
- **Featural interpretation**: Ability to analyse a transcribed corpus with any set of phonological features, both the transcription system and the feature set may be one of the ones built-in (Hayes 2008; Mielke 2008) or may be user-defined. Phonological tiers may be extracted based on features.
- **Functional load**: Calculation of the functional load of individual pairs of sounds within the corpus, defined at either the segment or feature level (cf. Hockett 1966; Surendran & Niyogi 2003; Wedel, Kaplan, & Jackson 2013).
- **Predictability of distribution**: Calculation of the extent to which any pair of sounds is predictably distributed, given a set of environments that they can occur in (cf. Hall 2009, 2012; Hall & Hall 2013).
- **String similarity**: Calculation of the extent to which pairs of words are similar to each other using either orthographic or phonetic transcription (e.g., for calculating neighbourhood density) (cf. Fritsch et al. 2004, Khorai 2012; Greenberg & Jenkins 1964; Luce & Pisoni 1998; Yao 2011).
- **Frequency of alternation**: Calculation of the frequency with which two sounds alternate with each other, given a measure of similarity (cf. Silverman 2006, Johnson & Babel 2010, Lu 2012).
- **Acoustic similarity**: Calculation of the acoustic similarity between sounds/words, based on alignment of MFCCs (cf. Mielke 2012) or amplitude envelopes (cf. Lewandowski 2012), derived from wav files.

3. Examples:

1. The IPHOD corpus (Vaden et al. 2009), transcribed in CMU (Weide 1994) and read into PCT.


3. Windows for calculating predictability of distribution, for [s] and [f] in the IPHOD corpus (IPA version), across a variety of user-selected, featurally defined environments. PCT allows for calculation on different tiers and can automatically check that chosen environments are exhaustive and unique. Results are shown on screen and can be saved to a tab delimited txt file. (Compare results here for weighted average entropy across a set of environments vs. just a measure of the frequency distribution.)

4. Turning a phonetically transcribed text of Gitksan (courtesy of Barbara Sennott & the UBC Gitksan research group) into a frequency-tagged corpus for analysis.

4. Obtaining PCT:

- **PCT** is available (in beta form) for Windows, Mac, and Linux.
- Go to [http://sourceforge.net/projects/phonologicalcorpustools](http://sourceforge.net/projects/phonologicalcorpustools)
- Click on the “files” tab.
- Click on the link with the highest number (= most recent version).
- As of July 2014, that’s v0.15.
- **Windows**
  - Click on win64.
  - Download the corpustools-0.15-amd64.msi link (click or right-click it).
  - Run the downloaded installer by double-clicking.
  - PCT allows for calculation of the functional load of individual pairs of sounds within the corpus, defined at either the segment or feature level (cf. Hockett 1966; Surendran & Niyogi 2003; Wedel, Kaplan, & Jackson 2013).
  - **Mac**
    - Click on macosx.
    - Download the pclinuxx86.tar.gz (click or right-click it to choose the location).
  - Unzip the downloaded file (double-clicking it should automatically open your local software for unzipping files, e.g., StuffIt Expander or Archive Utility).
  - Double-click pclinuxx86 to start Phonological CorpusTools.
  - You can then move the icon to your toolbar like any other application.
- **Linux / Fallback instructions**
  - Make sure Python 3.3 and Tk are installed first.
  - Download the source code (corpustools-0.15.zip or corpustools-0.15.tar.gz) from the SourceForge page above.
  - Run setup.py, located in the top-level directory.
  - Run Phonological CorpusTools by entering “pct” from any command line.
- Any questions / comments / suggestions: [kathleen.hall@ubc.ca](mailto:kathleen.hall@ubc.ca)

5. References: