

CLEARPOND: User Guide

Anthony Shook

Northwestern University, Evanston, IL

a-shook@northwestern.edu

Updated February, 2013

Abstract

CLEARPOND provides a user-friendly, web-based interface for obtaining Dutch, English, French, German and Spanish phonological and orthographic neighborhood densities (or, PONDs). The search function allows users to search for POND information in any of the five languages using single word queries or providing full lists of words. CLEARPOND provides a number of important psycholinguistic measures, such as neighborhood density and neighborhood frequency, both for within-language neighbors and foreign-language neighbors. With user-controlled output selection, researchers can choose the output parameters that are most relevant. In addition to allowing users to acquire data for specific words, CLEARPOND can also search by features so that researchers can generate new lists of words that meet precise criteria, such as a specific range of neighborhood sizes or lexical frequency (as provided by the Subtlex databases). Furthermore, multiple filters can be applied simultaneously, providing greater control for stimuli creation. Users also have the option of exporting their results directly to a text file, making it easy to create downloadable documents containing pertinent psycholinguistic measures for all of their stimuli.

1 Example Search Screen

The screenshot shows the CLEARPOND search interface. At the top, there is a navigation bar with links for Home, EnglishPOND, DutchPOND, FrenchPOND, GermanPOND, SpanishPOND, and Additional Tools. The main content area is divided into several sections:

- Words:** A text input field containing "dog", "cat", and "ghost". A blue circle with the number "1" is overlaid on this field.
- Search by:** Radio buttons for "Orthography" (selected) and "Phonology". A "Search CLEARPOND" button is next to it. A blue circle with the number "2" is overlaid on the search button.
- Output Selection:** A section titled "Select the features that you want CLEARPOND to output". It contains several groups of checkboxes:
 - Features:** Neighborhood Size (count) [checked], Mean Neighborhood Frequency [unchecked], Neighbors (list of words) [unchecked].
 - Neighbor Type:** Orthographic [checked], Phonological [checked].
 - Neighbor Metric:** Total [checked], Substitution [unchecked], Addition [unchecked], Deletion [unchecked].
 - Neighbor Frequency:** All Neighbors [selected], Higher Frequency Only [unselected].
 - Cross-Linguistic Neighbors:** A dropdown menu with options "Dutch", "French", "German", and "Spanish". A blue circle with the number "3" is overlaid on this dropdown.
- Optional: Supply ranges to filter your results! (format: e.g., 0-100)**
 - Lexical Data:** Input fields for Lexical Frequency (Subtlex), Word Length (Orthographic), and Word Length (Phonological). A blue circle with the number "4" is overlaid on this section.
 - Language Filters:** Four columns of input fields for "English", "Dutch", "German", and "Spanish". Each column has four fields for "Orthographic Neighborhood Size" and "Phonological Neighborhood Frequency". A blue circle with the number "5" is overlaid on the "French" column.

At the bottom, there is a footer with the text: "CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. Copyright Northwestern Bilingualism and Psycholinguistics Laboratory, 2012. Contact: e-shook (at) northwestern (dot) edu".

1. Search Box: Enter single words or non-words, lists of words or non-words, or leave blank to search the database by features.
2. Search Options: Choose to search by Orthography or Phonology
3. Output Selection: Choose which features you'd like CLEARPOND to output.
4. Lexical Data Filters: Enter single numbers or ranges (e.g., 0-100) to search based on lexical frequency or orthographic and phonological word length.
5. Language Filters: Enter single numbers or ranges (e.g., 0-100) to search based on neighborhood size or frequency (orthographic and phonological) for each individual language.

2 Example Output Screen

The screenshot shows the CLEARPOND web application interface. At the top, there is a navigation bar with links for Home, EnglishPOND, DutchPOND, FrenchPOND, GermanPOND, SpanishPOND, and Additional Tools. The main content area is titled "CLEARPOND Results" and features a search results table. A link "Click here to export your results" is highlighted with a blue circle labeled '1'. The table has columns: Word, PhoWord, Length (Ortho), Length (Phono), Freq (per-million), OTAN, and PTAN. Two rows are shown: 'dog' and 'cat'. A second blue circle labeled '2' points to the table. Below the table, a message says "Calculating Non-words - Please Wait:". A third blue circle labeled '3' points to a small table of non-words, with a message above it stating "The following items were not found in the database, but have been calculated by CLEARPOND". At the bottom, the elapsed time is shown as "1.79 seconds". The footer contains the CLEARPOND logo and contact information.

CLEARPOND

Home EnglishPOND DutchPOND FrenchPOND GermanPOND SpanishPOND Additional Tools

Visit the Northwestern Bilingualism and Psycholinguistics Laboratory

CLEARPOND Results

[Click here to export your results](#) [What do the headings mean?](#)

Word	PhoWord	Length (Ortho)	Length (Phono)	Freq (per-million)	OTAN	PTAN
dog	d.A.g	3	3	192.8431	20	25
cat	k.1.t	3	3	66.3333	33	50

Calculating Non-words - Please Wait:

The following items were not found in the database, but have been calculated by CLEARPOND

NonWord	Length	TotalND
blck	5	10

Elapsed time: 1.79 seconds

CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities
Copyright Northwestern Bilingualism and Psycholinguistics Laboratory, 2012. Contact: a-shook (at) northwestern (dot) edu

1. Export Data Link: Clicking Export Data will save the CLEARPOND output to a tab-delimited text file. This is recommended when the output contains a large number of columns.
2. Output Table: Contains the CLEARPOND output information. By default, CLEARPOND provides each word's length in phonemes and letters, and lexical frequency (frequency-per-million; see Appendix A). For a full list of the column headings, check the Glossary in section 4.
3. Non-word Output: If an item or items are not in the CLEARPOND database, the Neighborhood Information will be calculated on-line. CLEARPOND will calculate orthographic OR phonological neighborhoods for words not in the database, as well as for non-words, dependent on your search parameters.

3 How to Search the Database

3.1 Using a List of Words

The simplest way to search the CLEARPOND database is to type or copy/paste a list of words into the search window, with each word on its own line, then select the desired output values and click "submit." For each word, users can choose to output:

- * Orthographic and Phonological Neighborhood Size
- * Orthographic and Phonological Neighborhood Frequency
- * Neighbors based on Substitution, Addition, and Deletion (or, the sum of all three)
- * Neighbors of higher frequency than the input word only
- * Neighborhood information in other languages

3.2 Calculating Neighborhood Information for Non-Words

CLEARPOND can generate all of the neighborhood information found in the regular search for non-words as well. Simply enter the non-words as you would any other list, and CLEARPOND will calculate the information on the fly. Please note that while the database can calculate both orthographic and phonological neighborhood information for non-words, only one type of neighborhood can be calculated at a time, dependent on your search parameters. If you select additional languages when calculating non-word neighborhood information, CLEARPOND will alert you if any of the non-words are actual words in any selected language.

3.3 Generating Lists of Words

CLEARPOND can be used to generate lists of words that meet specific criteria. For example, consider a researcher who wants to perform a study examining the effect of orthographic neighborhood size on reading. By entering a range of neighborhood size values (e.g., 20-25) without providing a list of words, s/he can quickly generate a list of potential stimuli. Furthermore, multiple filters can be applied simultaneously, which allows the researcher to control for other linguistic features (such as frequency or word length) when creating stimuli. Users can filter by:

- * Lexical Frequency (Obtained from subtitle corpora; see Appendix A)
- * Word Length (by letters or phonemes)
- * Orthographic and Phonological Neighborhood Size and Frequency in other languages

4 Bigram/Biphone Probability Calculators

In addition to providing neighborhood information, CLEARPOND also includes tools to calculate bigram and biphone frequencies for words and non-words based on each of CLEARPOND's five language databases.

The screenshot shows the CLEARPOND website interface. At the top, there is a navigation bar with tabs for Home, EnglishPOND, DutchPOND, FrenchPOND, GermanPOND, SpanishPOND, and Additional Tools. The main content area is divided into two columns. The left column is titled "Bigram / Gram Probability Calculator" and contains a search box with the text "cat", "dog", and "wug" (labeled 1), a "Get Probabilities" button, and a language selection section (labeled 3) with checkboxes for Dutch, English (checked), French, German, and Spanish. Below this is another section for selecting between bigram probabilities, single-character probabilities, or both, with checkboxes for Bigrams (Positional) (checked) and Single-Character (Positional) (checked). The right column is titled "Biphone / Phone Probability Calculator" and contains a search box with the text "m. E. t" and "k. 1. t" (labeled 2), a "Get Probabilities" button, and a language selection section (labeled 3) with checkboxes for Dutch, English (checked), French, German, and Spanish. Below this is another section for selecting between biphone probabilities, single-phoneme probabilities, or both, with checkboxes for Biphones (Positional) (checked) and Single-Phonemes (Positional) (checked). A link "How are the probabilities calculated?" is located at the bottom of the main content area. A footer at the very bottom contains the text: "CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. Copyright Northwestern Bilingualism and Psycholinguistics Laboratory, 2012. Contact: a-shook (at) northwestern (dot) edu".

1. Bigram Search Box: Enter words or non-words to get positional bigram probability information.
2. Biphone Search Box: Enter words or non-words in CPSAMPA format (see Item 4) to get positional biphone probability information.
3. Search Filters: Choose which language(s) to use when calculating bigrams/biphones, and whether you'd like both single grams/phones as well as bigrams/biphones.
4. Additional Tools for Phonological Calculations: Here we provide a link to an IPA-CPSAMPA key in PDF format, which provides IPA symbols, their CPSAMPA equivalents, and examples of those phonemes in each language. In addition, we have provided a tool which converts IPA symbols into CPSAMPA.

5 Glossary: Column Header Definitions

Below is a table that shows each column's four-character label and its definition. In addition, each label may also be preceded by a one-or-two character language identifier: d-Dutch, e-English, f-French, g-German, or s-Spanish. For example, if you query the EnglishPOND for a word's total number of neighbors in all of the available languages, the results will show the columns OTAN, dOTAN, fOTAN, gOTAN, and sOTAN.

Table 5.1: Column Label Definitions

Label	Neighbor Type	Neighbor Metric	Neighborhood Frequency	Value
OTAN	Orthographic	All	All	Size
OTAF	Orthographic	All	All	Frequency(Mean)
OTAW	Orthographic	All	All	Word List
OTHN	Orthographic	All	Higher Frequency Only	Size
OTHF	Orthographic	All	Higher Frequency Only	Frequency(Mean)
OTHW	Orthographic	All	Higher Frequency Only	Word List
OSAN	Orthographic	Substitution	All	Size
OSAF	Orthographic	Substitution	All	Frequency(Mean)
OSAW	Orthographic	Substitution	All	Word List
OSHN	Orthographic	Substitution	Higher Frequency Only	Size
OSHF	Orthographic	Substitution	Higher Frequency Only	Frequency(Mean)
OSHW	Orthographic	Substitution	Higher Frequency Only	Word List
ODAN	Orthographic	Deletion	All	Size
ODAF	Orthographic	Deletion	All	Frequency(Mean)
ODAW	Orthographic	Deletion	All	Word List
ODHN	Orthographic	Deletion	Higher Frequency Only	Size
ODHF	Orthographic	Deletion	Higher Frequency Only	Frequency(Mean)
ODHW	Orthographic	Deletion	Higher Frequency Only	Word List
OAAN	Orthographic	Addition	All	Size
OAAF	Orthographic	Addition	All	Frequency(Mean)
OAAW	Orthographic	Addition	All	Word List
OAHN	Orthographic	Addition	Higher Frequency Only	Size
OAHF	Orthographic	Addition	Higher Frequency Only	Frequency(Mean)
OAHW	Orthographic	Addition	Higher Frequency Only	Word List
PTAN	Phonological	All	All	Size
PTAF	Phonological	All	All	Frequency(Mean)
PTAW	Phonological	All	All	Word List
PTHN	Phonological	All	Higher Frequency Only	Size
PTHF	Phonological	All	Higher Frequency Only	Frequency(Mean)
PTHW	Phonological	All	Higher Frequency Only	Word List
PSAN	Phonological	Substitution	All	Size
PSAF	Phonological	Substitution	All	Frequency(Mean)
PSAW	Phonological	Substitution	All	Word List
PSHN	Phonological	Substitution	Higher Frequency Only	Size
PSHF	Phonological	Substitution	Higher Frequency Only	Frequency(Mean)
PSHW	Phonological	Substitution	Higher Frequency Only	Word List
PDAN	Phonological	Deletion	All	Size
PDAF	Phonological	Deletion	All	Frequency(Mean)
PDAW	Phonological	Deletion	All	Word List
PDHN	Phonological	Deletion	Higher Frequency Only	Size
PDHF	Phonological	Deletion	Higher Frequency Only	Frequency(Mean)
PDHW	Phonological	Deletion	Higher Frequency Only	Word List
PAAN	Phonological	Addition	All	Size
PAAF	Phonological	Addition	All	Frequency(Mean)
PAAW	Phonological	Addition	All	Word List
PAHN	Phonological	Addition	Higher Frequency Only	Size
PAHF	Phonological	Addition	Higher Frequency Only	Frequency(Mean)
PAHW	Phonological	Addition	Higher Frequency Only	Word List

Appendix A: Lexical Frequency References

Dutch - SUBTLEX-NL

Keuleers E, Brysbaert M, New B (2010) SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods* 42: 643-650. doi:10.3758/BRM.42.3.643.

English - SUBTLEX-US

Brysbaert M, New B (2009) Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* 41: 977-990. doi:10.3758/BRM.41.4.977.

French - Lexique

New B, Pallier C, Brysbaert M, Ferrand L (2004) Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc* 36: 516-524.

German - SUBTLEX-DE

Brysbaert M, Buchmeier M, Conrad M, Jacobs AM, Bolte J, et al. (n.d.) The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology*.

Spanish - SUBTLEX-ESP

Cuetos F, Glez-Nosti M, Barbn A, Brysbaert M, Barbon A (2011) SUBTLEX-ESP: Spanish word frequencies based on film subtitles. *Psicologica* 32: 133-143.