

User manual for *SyllabO+*

Version August 4th 2016



LABORATOIRE DES NEUROSCIENCES
DE LA PAROLE ET DE L'AUDITION

SPEECH AND HEARING
NEUROSCIENCE LABORATORY



UNIVERSITÉ
LAVAL

User manual

Before downloading the full database, it is recommended to have an idea of what data is needed for your project (given its size 177 MB). The databases contain a large amount of, and it is possible that only a portion of it will be useful for a particular purpose.

Instructions

1. Go to <http://syllabo.speechneurolab.ca/> and after filling in a few questions, you will access the **SyllabO+ database**. All the information will remain confidential and will only be used to compute statistics and report to funding agencies.

Accéder à / Go to SyllabO+

2. You may then **choose** either the **complete database** (information extracted from all the audio recordings), whether **syllables** or **phones, or a specific database**. The complete database contains statistics about the use of phones and syllables across communication contexts, age and gender. It represents the average use of oral language in native adult speakers of Quebec French, in terms of syllables, syllable sequences and phones.

The example below shows the complete syllable database.

Syllable database

When you download the entire syllable database, four files are included in a .ZIP folder: the single syllables, the sequences of two syllables (pairs), sequences of three syllables (triads) and sequences of four syllables (tetrads).

Alternative downloads are also available below in the form of .XLSX files and .CVS files transcribed in SAMPA (that is, without special characters).

[Download complete syllable database](#)

Alternatively, you can also download a **specific database**, which will contain statistics computed from a subset of the audio recordings, by using the following parameters, individually or combined: context of communication (formal, informal), age (range), and sex of the speakers. This means that you could obtain the statistics, for example, of syllable use for only the young adults, or for all women in formal contexts, to name a few of the many possibilities. Note that this option is only available for the syllable database at the moment. We expect to provide this option for the phone database during the Fall 2016. Note also that online viewing of the results is currently not available, but we are working on this!

The image below shows the form to download a specific database.

Search in syllable database

Please identify your specific research parameters. The only mandatory field is the type of unit.

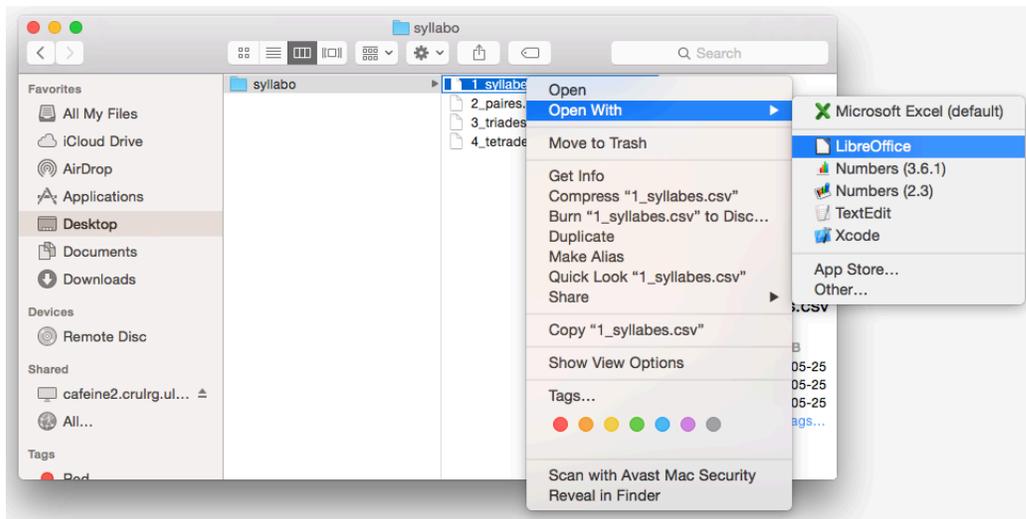
Type*	Syllabes – Syllables
Contexte d'élocution – <i>Communication context</i>	Tous – All
Genre des locuteurs – <i>Speaker gender</i>	Tous – All
Âge minimum – <i>Minimum age</i>	Optionnel – Optional
Âge maximum – <i>Maximum age</i>	Optionnel – Optional
Code de sécurité – <i>Security code*</i>	S Z B + X

Télécharger la base de données spécifique de SYLLABES – Download specific syllable database

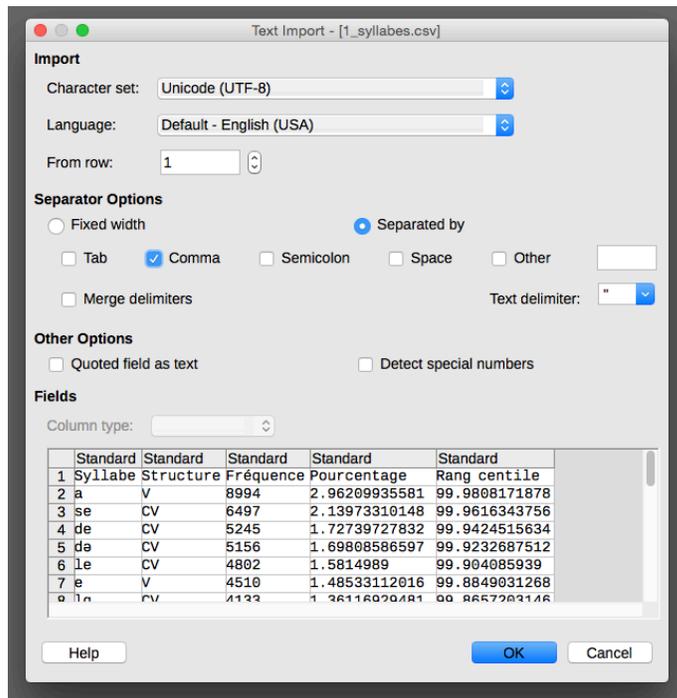
Click to download the desired database.

- If you have **downloaded a complete database**, you will obtain four different files compressed in ZIP format for convenience. For the syllables, there will be one file for the individual syllables and their statistics, one for the pair of syllables and their statistics, one for the groups of 3 syllables (triads) and their statistics, and a last one for the groups of 4 syllables (tetrads) and their statistics. The size of the uncompressed files is 177 MB. For the phones, you will also obtain 4 files, for the individual phones and their statistics, one for the pair of phones and their statistics, one for the groups of 3 phones (triads) and their statistics, and a last one for the groups of 4 phones (tetrads) and their statistics. The size of the uncompressed phone database is 22 MB.
- If you **downloaded a specific database**, you will only obtain the selected file (syllables, pairs of syllables, or triads).

- All files are in **CSV format** (Comma Separated Values), which is a simple way of storing tabular data in plain text files. Alternative downloads are also available in the form of .XLSX files (with special IPA characters already embedded) and .CVS files transcribed in SAMPA – Speech Assessment Methods Phonetic Alphabet (without any special characters). However, these formats are at the moment only available to download the entire database.
- To open the CSV (API) files we recommend using the software **LibreOffice**. We recommend it because it handles all phonetic (IPA) characters correctly. Important: Microsoft Excel does not handle these special characters properly. This is why we provide instructions to guide you through the process of having the characters recognized, and this is also why we provide alternative downloads. Note: LibreOffice is a free and open source software available for download at <http://www.libreoffice.org/> .

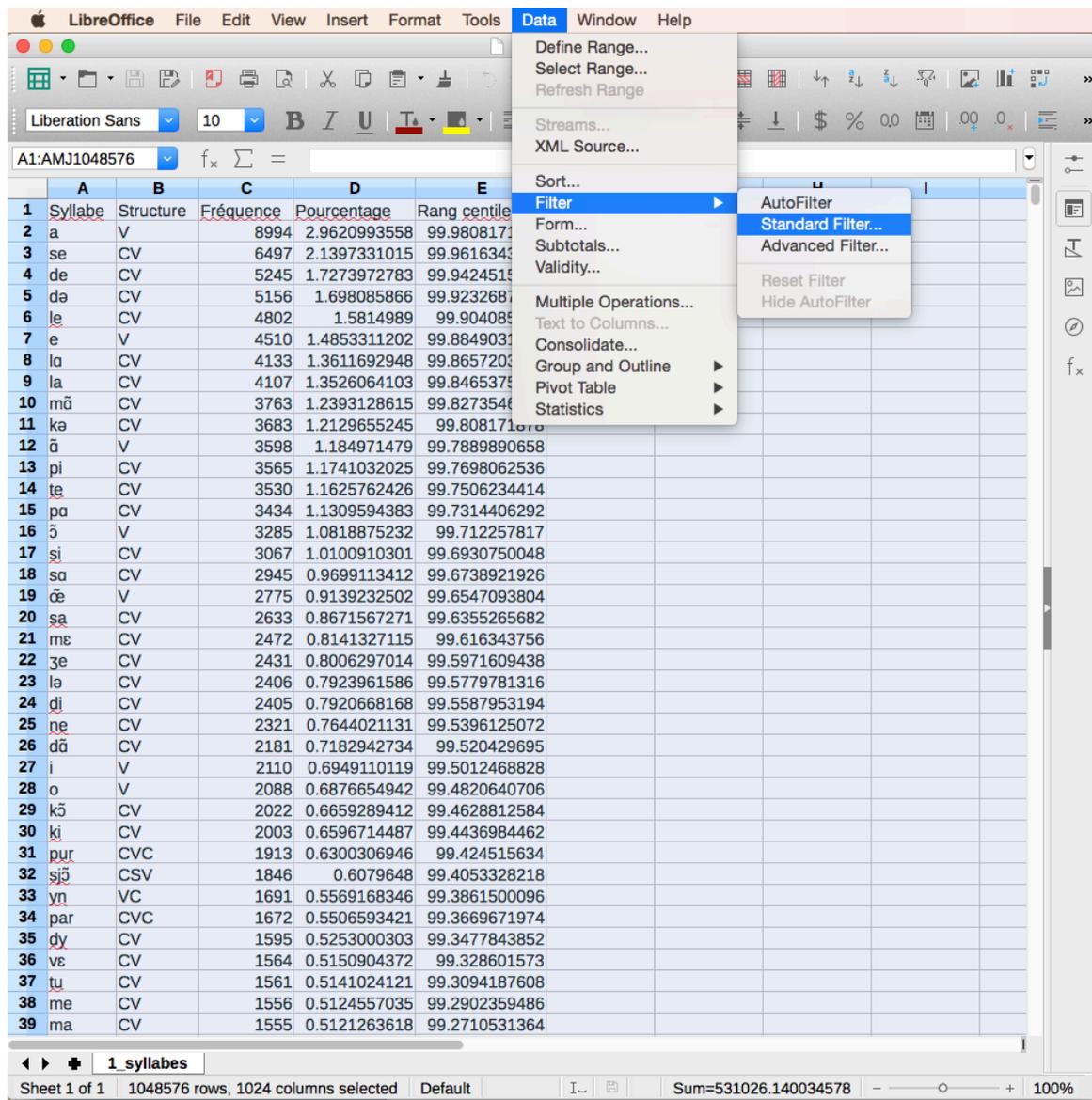


- Select **Unicode (UTF-8)** as the character set. Select **Comma** as the separator option. Click **OK** to open.

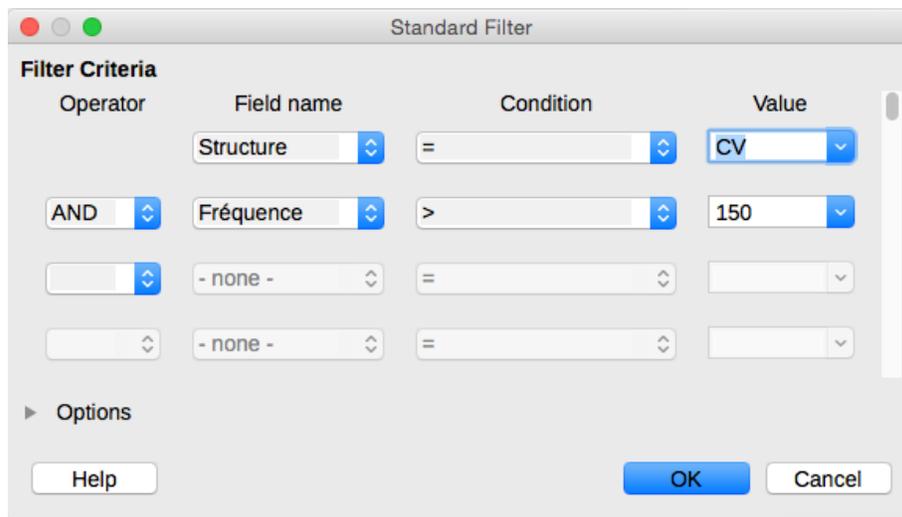


8. Once the file is opened, you can select only the columns relevant to your project (*Hide* or *Delete* unwanted columns).

9. Filter the data according to your needs: select all data, go to the menu, and click on ***Data – Filter – Standard Filter***.



10. You may then use any **filter** you wish to obtain the desired information. Select which column is to be filtered, according to which parameter. For example:



11. You obtain your **results!**

Example of stimuli creation

Here is an example of how specific stimuli can be created from *SyllabO+*. Using the table containing groups of 3 syllables, these specific columns were selected:

- Group of 3 syllables (IPA transcription)
- Structure 1st syllable
- Structure 2nd syllable
- Structure 3rd syllable
- Percentile of score 1st syllable
- Percentile of score 2nd syllable
- Percentile of score 3rd syllable

The filter “ = CV” was applied to all “structure” columns and the filter “ >= 95” was applied to all “percentile of score” columns. The resulting table (*Table 1*) is thus a selection of syllable triads, where each syllable has a simple “consonant-vowel” structure and has a high frequency since it is found

between the 95th and 100th percentile. From this table, the groups of 3 syllables can be easily used as stimuli in an experimental study.

Table 1. Stimuli: triads of simple-structure (CV), high-frequency syllables

Triad	Structure 1	Percentile 1	Structure 2	Percentile 2	Structure 3	Percentile 3
kə mā se	CV	99.0216765778	CV	99.8273546902	CV	99.9616343756
vu za ve	CV	98.849031268	CV	98.6188375216	CV	98.273546902
mā dɔ ne	CV	99.8273546902	CV	97.602148475	CV	99.5396125072
le zā fā	CV	99.904085939	CV	98.4270093996	CV	97.0650297334
dɪ fe rā	CV	99.5587953194	CV	96.451179743	CV	98.0625359678
...

That's it! Hope you enjoy it, and stay tuned for future developments on the phone database, online searches and the lexical database! Please do not forget to cite our article in all your publications and reports derived from the use of SyllabO+ (Bédard et al., Révisions soumises):

Bédard, P., Audet, A.-M., Drouin, P., Roy, J. P., Rivard, J., & Tremblay, P. (Révisions soumises). SyllabO+: a new tool to investigate sublexical phenomena in Québec spoken French. *Behavioral Research Methods*.