Guide d'utilisation de *SyllabO*+

Version du 4 aout 2016



LABORATOIRE DES NEUROSCIENCES DE LA PAROLE ET DE L'AUDITION

SPEECH AND HEARING NEUROSCIENCE LABORATORY



Guide d'utilisation

Avant de télécharger la base de données entière, il est recommandé d'avoir une bonne idée de l'information dont vous avez besoin, car les possibilités sont nombreuses et certaines de nos bases de données sont assez imposantes (177 MB pour la base de données des syllabes) ! *SyllabO*+ contient en effet une grande quantité de données et il est possible qu'une partie seulement soit pertinente pour vos objectifs.

Instructions

 Allez à <u>http://syllabo.speechneurolab.ca/</u> et après avoir répondu à quelques brèves questions, vous aurez accès aux **bases de données** *SyllabO*+. Vos réponses demeureront strictement confidentielles et ne seront utilisées qu'à des fins de rapports pour les agences de financements et autres partenaires du projet afin de quantifier l'utilisation de nos bases de données.

Accéder à SyllabO

2. Vous pourrez alors choisir entre les bases de données complètes (informations extraites à partir de tous les enregistrements audio), que ce soit pour les syllabes ou pour les phones, ou les bases de données spécifiques. Les bases de données complètes contiennent les statistiques d'utilisation globales, c'est-à-dire pour tous les contextes et tous les locuteurs (peu importe leur sexe et leur âge). Elles représentent ainsi l'usage général des locuteurs adultes du français québécois en terme de syllabes, séquences de syllabes et phones.

L'image ci-dessous illustre la base de données complète des syllabes.

Base de données des syllabes

Lorsque vous téléchargez toute la base de données de syllabes, quatre fichiers sont inclus dans un document compressé (.ZIP): soit les syllabes, les paires, les triades et les tétrades.

Des versions alternatives de la base de données sont également disponibles ci-dessous aux format .XLSX (avec les caractères spéciaux API déjà encodés) et au format .CVS sans caractères spéciaux (transcription avec l'alphabet SAMPA – Speech Assessment Methods Phonetic Alphabet).

Télécharger la base de données complète de SYLLABES

Vous pouvez également choisir de télécharger une **base de données spécifique**, laquelle contiendra alors les statistiques d'utilisation extraites à partir d'un sous-ensemble des enregistrements audio. Les statistiques d'utilisation seront alors compilées en fonction des paramètres suivants, individuellement ou combinés: <u>contexte de communication</u> (formel, informel), <u>âge</u>, et <u>genre</u> des locuteurs. Il est ainsi possible de télécharger les statistiques d'utilisation des syllabes seulement pour les jeunes adultes, ou pour les femmes en contexte formel, etc. À noter que cette option n'est disponible que pour les syllabes pour le moment. À noter également que les résultats ne seront pas affichés en ligne mais téléchargeable. L'option de recherche en ligne devrait être offerte dans le futur (à suivre!). La recherche dans la base de données des phones sera disponible à l'automne 2016.

Pour les téléchargements de ces bases de données spécifiques, la fenêtre suivante est utilisée :

Téléchargement selon des paramètres de recherche

Veuillez identifier les paramètres particuliers.

Туре*	Syllabe	•				
Contexte d'élocution	Tous	•				
Genre des locuteurs	Tous	•				
Âge minimum	Optionnel					
Âge maximum	Optionnel					
Code de sécurité*	EELW					
Télécharger la base de données particulière						

Cliquez pour télécharger la base de données désirée.

- 3. Si vous avez téléchargé une base de données complète, vous obtiendrez quatre fichiers différents. Pour les syllabes, vous obtiendrez un fichier pour les syllabes individuelles et leurs statistiques, un pour les paires de syllabes et leurs statistiques, un pour les groupes de 3 syllabes (triades) et leurs statistiques et un pour les groupes de 4 syllabes (tétrades) et leurs statistiques. La taille pour l'ensemble des fichiers est de 177 MB. Pour les phones, vous obtiendrez un fichier pour les phones individuels et leurs statistiques, un pour les paires de phones (diphones) et leurs statistiques, un pour les groupes de 3 phones (triades) et leurs statistiques et un pour les groupes de 4 phones (diphones) et leurs statistiques, un pour les groupes de 3 phones (triades) et leurs statistiques et un pour les groupes de 4 phones (tétrades) et leurs statistiques. La taille pour l'ensemble des fichiers est de 22 MB.
- Si vous avez téléchargé une base de données spécifique, vous obtiendrez seulement le fichier sélectionné (syllabes, paires de syllabes ou triades).

- 5. Tous les fichiers sont au format CSV en API (Comma Separated Values), qui est une manière simple de stocker des données tabulaires en fichiers textes bruts. À noter que deux autres formats sont maintenant disponibles (avec les caractères spéciaux API déjà encodés) et au format .CVS sans caractères spéciaux (transcription avec l'alphabet SAMPA Speech Assessment Methods Phonetic Alphabet). Ces formats ne sont toutefois disponibles que pour le téléchargement des bases de données complètes.
- 6. Si vous téléchargez les fichiers .CVS API, le plus simple est de les ouvrir avec le logiciel LibreOffice (tableur). Nous recommandons l'utilisation de ce logiciel parce qu'il permet de traiter les caractères phonétiques (API) correctement. Microsoft Excel ne traite pas ces caractères spéciaux correctement. À noter: LibreOffice est un logiciel gratuit et Open source, disponible pour téléchargement à <u>http://fr.libreoffice.org/</u>. Pour la version en français, téléchargez l'installateur principal, puis l'interface en français. (Allez à

http://www.libreoffice.org/ pour la version en anglais).





 Sélectionnez Unicode (UTF-8) comme jeu de caractères. Sélectionnez Virgule comme option de séparateur. Cliquez sur OK pour ouvrir.

• •	Import de texte - [1_syllabes.csv]								
Importer									
Jeu	de caracté	eres : Unic	Unicode (UTE-8)						
	uo ouraon					`			
Lan	igue :	Par	défaut - Fran	çais (Canada)		۵			
Àp	artir de la li	igne : 1	٢						
Optio	ons de sép	arateur							
0	Largeur fix	e		💿 Séparé	par				
	Tabulatio	on 🔽 Virgu	ile 🗌 Poin	t-virgule Espa	ice 🗌 Autre				
	Fusionne	er les sépara	teurs		Séparateur de	e texte : 🛛 💾 🞽			
Autre	es options								
	Champ on	tro quillomotr	oommo tovt		r los nombros cnós	ioux			
	Champ en	tre guillemets	s comme text	e Delecie	r les nombres spec	iaux			
Chan	nps								
Тур	e de colon	ne :	٥						
	Standard	Standard	Standard	Standard	Standard				
1	Syllabe	Structure	Fréquence	Pourcentage	Rang centile				
2	a	v	8994	2.96209935581	99.9808171878				
3	se	CV	6497	2.13973310148	99.9616343756				
4	de	CV	5245	1.72739727832	99.9424515634				
5	də	CV	5156	1.69808586597	99.9232687512				
6	le	CV	4802	1.5814989	99.904085939				
7	е	v	4510	1.48533112016	99.8849031268				
0	10	CV	1133	1 26116020/81	00 86572031/6				
	Alida					Annulan			
	Aide								

- Une fois le fichier ouvert, vous pouvez sélectionner les colonnes pertinentes pour votre projet.
 Masquez ou supprimez les colonnes non désirées.
- Filtrez les données selon vos besoins : sélectionnez toutes les données, allez au menu, puis cliquez sur *Données – Filtre – Filtre Standard*.

.	• 🖪 🖻	! - C	X D = -	। ि र े र <u>Q</u>	Sélectionner la plage Actualiser la plage	2, 3, 5, 🔽 🔟	Ω	-co-	
iberation	Sans 🔽	10 🖌]	BIUT		Flux	% 0.0 🛅 0.0 %			
:AMJ104	8576	f., Σ =	Pourcentage		Source XML				Т
•	B	C	D	F	Trier		1		
Syllabe	Structure	Fréquence	Pourcentage	Rang centile	Filtre 🕨 🕨	AutoFiltre			110
a	V	8994	2 96209935581	99 9808171878	Formulaire	Filtre standard			
SP	CV	6497	2 13973310148	99 9616343756	Sous-totaux	Filtre spécial	1		
de	CV	5245	1 72739727832	99 9424515634	Validité	D.C. Malifalla and a Clines			
da	CV	5156	1 69808586597	99 9232687512		Reinitialiser le filtre			
le	CV	4802	1 5814989	99 904085939	Operations multiples	Masquer l'AutoFiltre	Jt		
0	V	4510	1 48533112016	99.8849031268	Texte en colonnes				
la	CV	4310	1 36116020/81	00 8657203146	Consolider				
la	CV	4103	1 35260641031	99.8037203140	Plan et groupe		_		
mã	CV	3763	1 23031286145	00 8273546002	Table de pilote				
ko	CV	3683	1 21206552451	00 808171878	Statistiques				
ã	V	3003	1 19407147001	00 7990900659					
ni	CV	3550	1 17/10220252	00 7609062526					
to	CV	3530	1 16257624261	00 7506234414					
na	CV	3434	1 12005042927	00 721 4406202					
õ	V	2205	1.00100752222	00 71 225 791 7					н.
J	CV	3200	1.00100102005	99.712257617			-		L
51	CV	3007	0.060011341311	99.0930750048					Ы
su		2943	0.909911341211	99.0730921920					Ł
00	CV	2113	0.913923230207	99.0347093804			-		a.
Sd	CV	2033	0.807130727134	99.0355205082					
31118	CV	2472	0.014132711330	99.010343750					
Je	CV	2431	0.800829701353	99.5971009436					
19	CV	2406	0.792396156556	99.5779781310			_		
	CV	2405	0.792000810846	99.5587953194					
dã	CV	2321	0.704402113056	99.5390125072					
:	V	2181	0.718294273406	99.520429095					
1	V	2110	0.094911011869	99.5012408828					
0	CV (2088	0.00700049421	99.4020040706					
KJ	CV CV	2022	0.000928941232	99.4020812584					
ĸ	CV OVO	2003	0.059671448708	99.4436984462					
pur	CVC	1913	0.630030694648	99.424515634					
SID	CSV	1846	0.607964799958	99.4053328218					
<u>yn</u>	VC	1691	0.556916834631	99.3861500096					
par	CVC	1672	0.550659342107	99.3669671974			_		
dy	CV	1595	0.525300030299	99.3477843852			_		

10. Vous pouvez alors utiliser le(s) filtre(s) de votre choix pour obtenir l'information désirée.Indiquez quelle colonne doit être filtrée selon quel paramètre. Par exemple :

• • •		Filtre standard						
Critères de filtre								
Opérateur	Nom de champ	Condition		Valeur				
	Structure	=	c	cv 🔽				
ET	Fréquence	>	0 1	.50 💙				
	- aucun(e)-	=	•	×				
•	- aucun(e)-	=	٥	~				
 Options 								
Aide			I	OK Annule	er			

11. Vous obtenez vos résultats!

Exemple d'élaboration de stimuli

Voici un exemple qui illustre comment des stimuli spécifiques peuvent être construits à partir de *SyllabO*. À partir du tableau des groupes de 3 syllabes, ces colonnes ont été sélectionnées :

- Triade (transcription en API)
- Structure 1^{ère} syllabe
- Structure 2^e syllabe
- Structure 3^e syllabe
- Rang centile 1^{ère} syllabe
- Rang centile 2^e syllabe
- Rang centile 3^e syllabe

Le filtre « = CV » a été appliqué à toutes les colonnes « structure » et le filtre « >= 95 » a été appliqué à toutes les colonnes « rang centile ». Le tableau résultant de ces filtres *(Tableau 1)* est une sélection de triades de syllabes, où chaque syllabe possède une structure simple « consonne – voyelle » et possède une fréquence élevée puisqu'elle est située entre le 95^e et le 100^e rang centile. Les groupes de 3 syllabes issus de ce tableau peuvent être utilisés comme stimuli dans une étude expérimentale.

Tableau 1. Stimuli: triades de syllabes à structures simples (CV) et fréquence élevée							
Triade	Structure 1	Rang centile 1	Structure 2	Rang centile 2	Structure 3	Rang centile 3	
ko mã se	CV	99.0216765778	CV	99.8273546902	CV	99.9616343756	
vu za ve	CV	98.849031268	CV	98.6188375216	CV	98.273546902	
mã do ne	CV	99.8273546902	CV	97.602148475	CV	99.5396125072	
le zã fã	CV	99.904085939	CV	98.4270093996	CV	97.0650297334	
di fe rã	CV	99.5587953194	CV	96.451179743	CV	98.0625359678	

Voilà ! Bonnes recherches, amusez-vous bien, et restez branchés pour les développement futurs (base de phones, recherches en ligne améliorées et base de données lexicales). Merci de citer notre article dans vos publications et rapports découlant de l'utilisation de 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*.