Cog

Cog est un outil de comparaison des langues utilisant les techniques de la lexicostatistique et de la linguistique comparative. Il peut être utilisé pour automatiser une grande partie du processus de comparaison de listes de mots provenant de différentes variétés de langues.

Voici une introduction (en anglais) : https://www.youtube.com/watch?v=UX_q0FNlqUI[1]

Téléchargement

Prendre la toute dernière version ici :

https://software.sil.org/cog/[2]

Survol

L'objectif de Cog est de fournir un cadre pour expérimenter différentes techniques de comparaison de variétés de langues. Il est conçu pour être utilisé de manière itérative : effectuer une comparaison, analyser les résultats, affiner le processus, effectuer à nouveau la comparaison, et ainsi de suite. La plupart des étapes du processus peuvent être adaptées. Il ne prend actuellement en charge que quelques techniques de comparaison, mais nous espérons en inclure beaucoup d'autres à l'avenir.

Analyser rapidement

Cog vous permet de produire rapidement des résultats en automatisant de nombreuses étapes du processus de comparaison de listes de mots :

- Segmentation basée sur l'IPA : divise automatiquement les mots en segments.
- Syllabation : marque automatiquement les syllabes.
- Identification du radical : identifie les préfixes et les suffixes afin de les ignorer lors de la comparaison.
- Alignement des mots : aligne les segments entre les paires de mots
- Identification des correspondances sonores : identifie automatiquement les correspondances sonores et les environnements dans lesquels elles se produisent.
- Identification des cognats : fournit diverses méthodes d'identification des cognats.
- Similitude lexicale/phonétique : calcule la similitude lexicale/phonétique pour plusieurs variétés de langues.

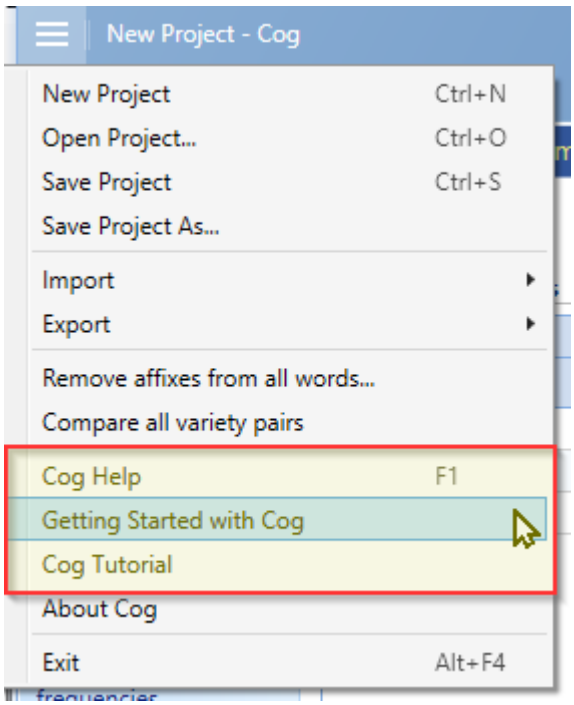
Visualiser

Cog vous permet de visualiser vos résultats sous différents angles. Cog génère automatiquement les types de représentations visuelles suivants :

- Arbres radiaux
- Dendrogrammes
- Graphiques de réseau
- Matrices de similarité
- Cartes géographiques

Ressources en anglais :

L'aide en Cog est au complet en anglais. Si vous lisez l'anglais, les ressources dans Cog sont plus à jour et sont disponibles ici :



En plus, voilà un vidéo qui fait un survol de Cog.

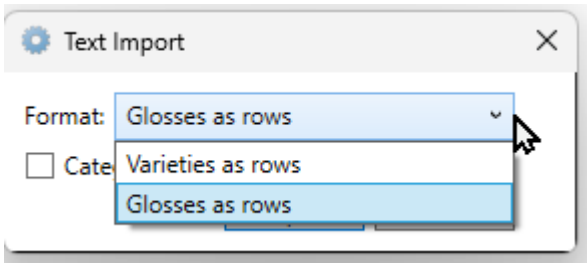
https://www.youtube.com/watch?v=TuDVJrbXcos&t=55s (0m55s - 4m30s)

Mode d'emploi

1. Préparation des données

La façon la plus facile d'importer des données dans Cog est de les aligner dans une feuille de calcul (Excel ou Libreoffice Calc).

Les gloses peuvent être en colonnes (« Varieties as rows ») ou en rangées (« Glosses as rows »). Le choix de l'orientation est à vous.



1.1. Mode « Glosses as rows »

Ici, les gloses sont au début de chaque rangée, et les dialectes sont au début de chaque colonne.

	nsiteu	ndàʔndàʔ	mədʉ̂mbù
lit	kúndɛ̀	kûndzə́	kûndə́
chaise	lʉ́ŋ	ndʒɔ́ŋ	ndʒə́ŋ
lampe	làmbù	lámbù	lámbù

1.2 Mode « Varieties as rows »

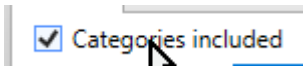
Ici, les dialectes sont au début de chaque rangée, et les gloses sont au début de chaque colonne.

	lit	chaise	lampe
nsiteu	kúndɛ̀	lʉ́ŋ	làmbù
ndàʔndàʔ	kûndzə́	ndʒɔ́ŋ	lámbù
mədʉ̂mbù	kûndə́	ndʒə́ŋ	lámbù

1.3 Categories

Dans les deux modes, vous pouvez ajouter une colonne ou rangée (selon l'orientation) des catégories après la glose, comme verbe ou nom, pour vous aider à cibler les recherches plus tard.

Si c'est ajouté, il faut cocher la case « Categories included » plus tard pendant l'importation.

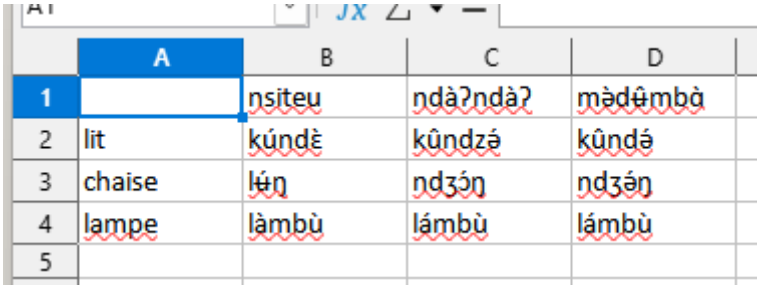


2. Libreoffice Calc

2.1. Exporter les données de LibreOffice Calc

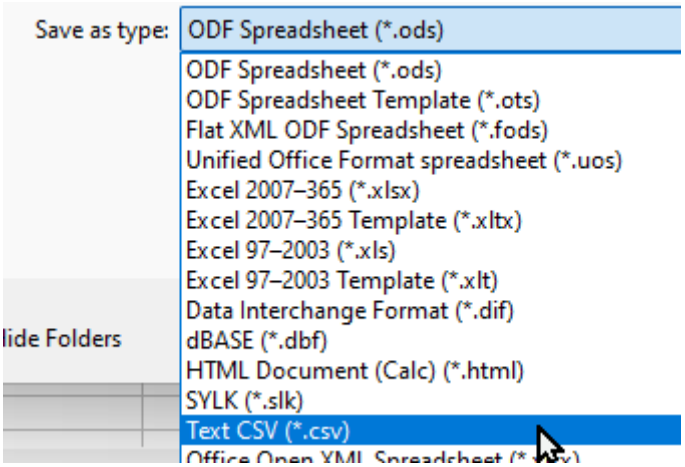
Cog peut importer un fichier « CSV », qui est semblable à une feuille de calcule (spreadsheet). Si vous avez LibreOffice (ou OpenOffice) Calc, c'est facile à exporter vos données en format CSV.

- Ouvrez le fichier.

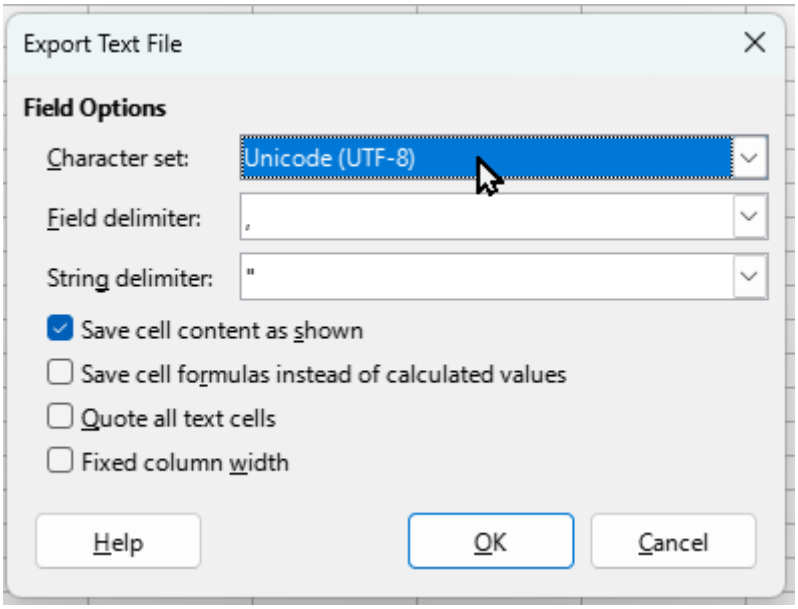


	A	B	C	D
1		nsiteu	ndàʔndàʔ	mədʉ̂mbù
2	lit	kúndɛ̀	kûndzə́	kûndə́
3	chaise	lʉ́ŋ	ndʒɔ́ŋ	ndʒə́ŋ
4	lampe	làmbù	lámbù	lámbù
5				

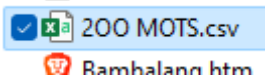
- Ficher>Enregistrer sous :
- Choisissez le dossier cible et saisissez un nom pour le fichier.
- Choisissez « Text CSV »



- Cliquez sur Enregistrer
  - LibreOffice demandera si vous voulez enregistrer en mode CSV, accepter d'exporter en format CSV.
- Assurez'vous que l'encodage UTF-8 est sélectionné.

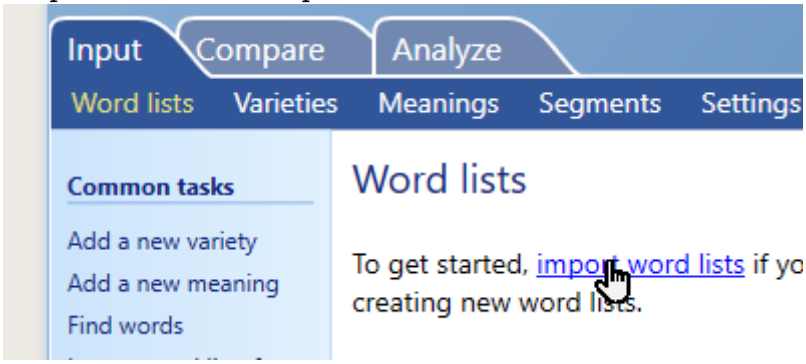


- Cliquez sur OK.
- Vous allez voir le fichier CSV dans votre dossier, prêt à importer.

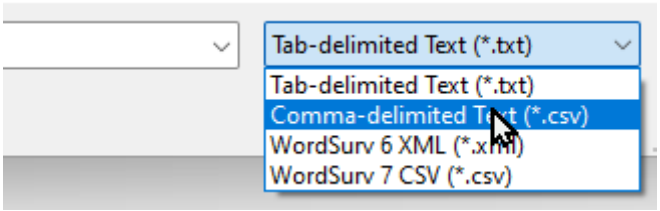


2.1 Importer les données en format CSV dans Cog

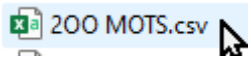
- Ouvrez COG
- Aller à l'onglet « Input, section "Word Lists".
- Cliquez sur le lien "import word lists".



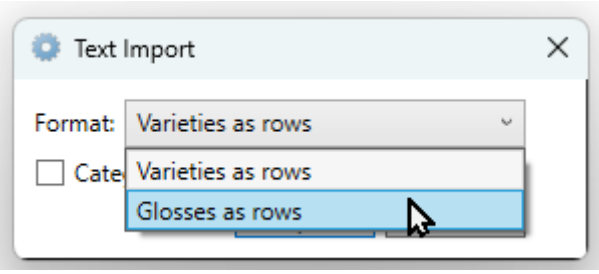
- En bas de la fenêtre, choisissez le CSV.



- Choisissez le ficher CSV et cliquez "Open".



- Choisissez l'option voulue :



- Cliquez “Import”.
- S’il y a des erreurs d’importation, allez à la section 4.

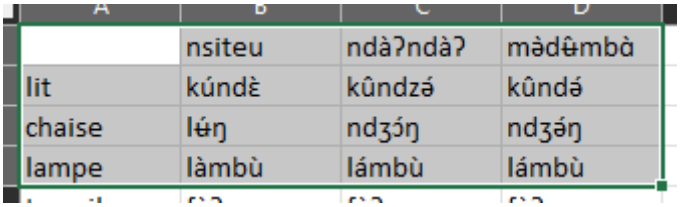
3. Microsoft Excel

3.1. Copier les données d’Excel

Selon ma compréhension, c’est impossible d’exporter un CSV en format Unicode depuis Excel. Si on essaie de le faire, Excel écrasera tous les caractères spéciaux.

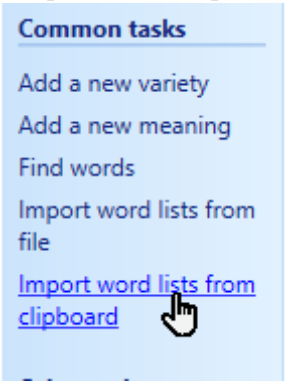
Heureusement, c’est possible de copier la feuille et le coller dans Cog.

- Sélectionnez les cellules :

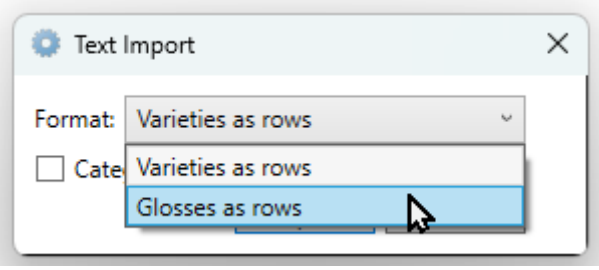


	nsiteu	ndàʔndàʔ	màdʉ̀mbò
lit	kúndɛ̀	kûndzə́	kûndə́
chaise	lʉ́ŋ	ndʒɔ́ŋ	ndʒə́ŋ
lampe	làmbù	lámbù	lámbù

- Copiez
- Cliquez sur “Import words from clipboard”



- Choisissez l’option voulue :



- Cliquez “Import”.
- S’il y a des erreurs, voir la section 4.

3.2. Exporter un fichier TSV d'Excel

Lorsque votre liste de mots est complète et organisée, procédez comme suit :

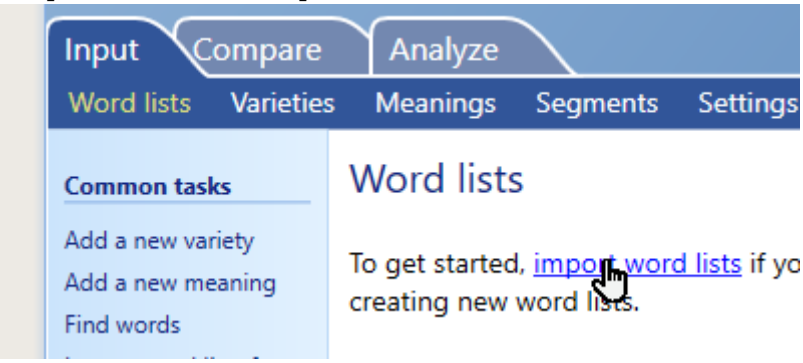
Dans Excel, enregistrez la feuille de listes de mots en tant que Texte Unicode.

Cliquez sur Oui si Excel vous demande si vous souhaitez conserver le classeur dans ce format, car il ne prend pas en charge certains types de données dans ce format car il ne prend pas en charge certaines fonctionnalités.

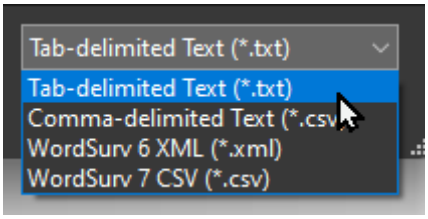
Souvenez-vous de l'endroit où vous avez enregistré le fichier texte.

3.3 Importer les données en format TSV dans Cog

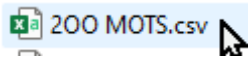
- Ouvrez COG
- Aller à l'onglet « Input, section "Word Lists".
- Cliquez sur le lien "import word lists".



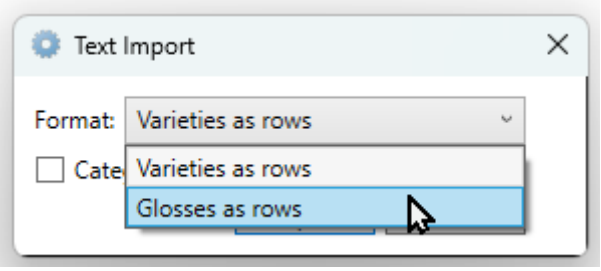
- En bas de la fenêtre, choisissez le "Tab Delimited Text".



- Choisissez le ficher .txt et cliquez "Open".



- Choisissez l'option voulue :



- Cliquez "Import".
- S'il y a des erreurs d'importation, allez à la section 4.

4. Erreurs d'importation

Cog n'accepte pas deux gloses qui sont pareilles. Cog vous dira pendant l'importation s'il y a des doublons. Il faut modifier vos données avant de les réimporter. Je vous suggère d'ajouter des chiffres aux doublons (par exemple "homme" et "homme 2") dans vos données exportées et réimportées.

5. Enregistrement de votre projet

Si vous voulez garder votre travail en Cog, il faut enregistrer votre projet.

Cliquez l'icône Enregistrer ou choisissez "Save project" du menu .

6. Utiliser Cog

Cog a trois onglets, et chaque onglet a plusieurs sections comme ceci :

6.1 Input (Entrée des données)

L'onglet "Input" vous aide à importer ou modifier vos données.

Voir sections 2 (Calc) ou 3 (Excel).

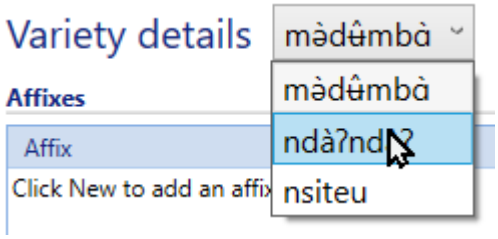
6.1.1 Word Lists (Liste de mots):

Regarder ou modifier vos données.

6.1.2 Varieties (langues ou dialectes) :

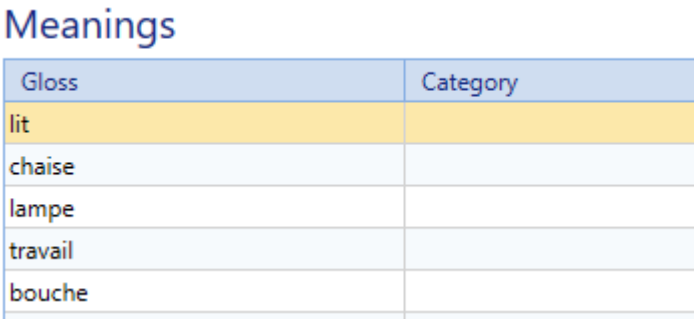
Ici, on voit les mots, affixes (préfixes et suffixes), et les statistiques de caractère pour chaque langue.

Pour voir les statistiques d'une autre variété linguistique, la choisissez de la liste "Variety details".



6.1.3. Meanings

Ici, on peut gérer la liste de mots qui ont été collection.



6.1.4. Segments (lettres, digrammes, et trigrammes)

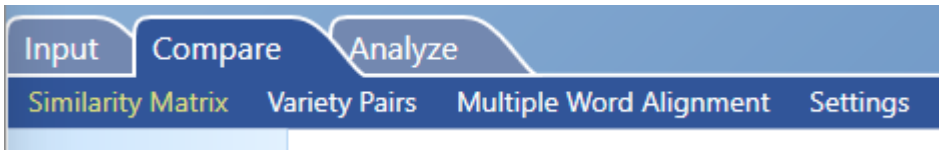
Ici, on peut voir le nombre d'exemplaires de chaque son.

	Close					
	i	iɛ	ya	yə	ɨ	ʉ
nsiteu	41	1	0	0	0	19
ndà?ndà?	26	0	1	0	2	24
màdə̂mbà	27	0	0	1	1	20

6.1.5. Settings (Paramètres)

Voici quelques paramètres pour configurer Cog.

6.2 Compare (Comparer)



Ici, Cog vous aide à comparer des langues.

6.2.1 Similarity Matrix (Lexicale)

Cog peut vous montrer les similarités lexicales (par mot) ou Phonétiques (par son).

Si on veut comparer les mots, choisissez “Lexical” et puis “Compare all variety pairs”.

Cog affichera un tableau des statistiques. Les chiffres plus grands représentent les langues avec les mots le plus semblables.

Similarity Matrix

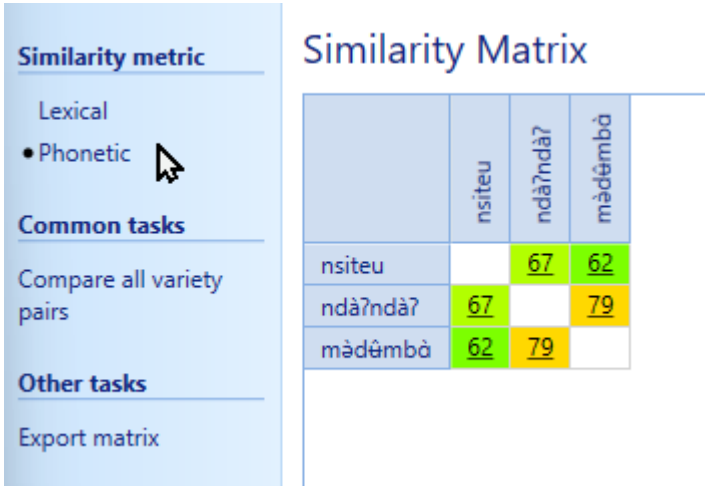
	nsiteu	ndà?ndà?	màdə̂mbà
nsiteu		38	29
ndà?ndà?	38		65
màdə̂mbà	29	65	

Cliquez sur les chiffres pour voir les détails de chaque paire. Cog affichera l’outil “Variety pairs” (Section 6.2.3).

6.2.1 Similarity Matrix (Phonétique)

Si on veut comparer les sons, choisissez “Phonetique” et puis “Compare all variety pairs”.

Cog affichera un tableau des statistiques. Les chiffres plus grands représentent les langues avec les sons le plus semblables.



Cliquez sur les chiffres pour voir les détails de chaque paire. Cog affichera l'outil "Variety pairs" (Section 6.2.3).

6.2.3. Variety Pairs (Pairs linguistiques)

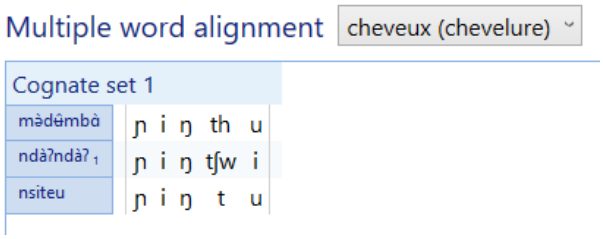
Ici, Cog vous montre les statistiques de similarité lexicale et phonétique selon vos données.

Il propose les mots cognats (lié phonétiquement) et les mots qui ne semblent pas être liés.

En plus Cog montre les changements entre des sons.

6.2.4. Multiple Word Alignment

Ici, vous pouvez choisir un mot et voir les changements entre des variétés linguistiques. Choisissez le mot voulu.



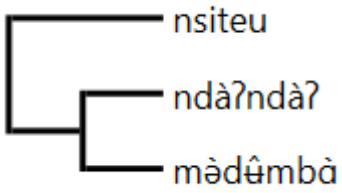
6.3. Analyze (Analyser)

L'onglet "Analyze" (Analyser) vous aide à explorer d'autres dimensions de vos données.

6.3.1. Hierarchical Graph

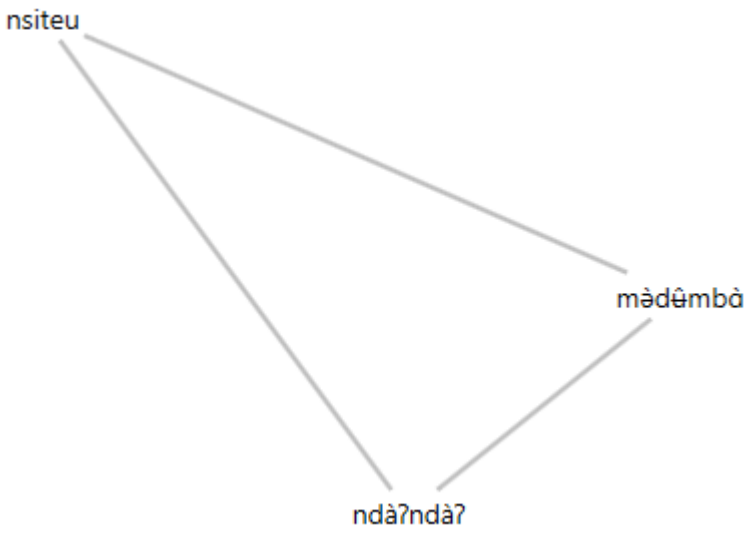
L'outil "Hierarchical Graph" essaie de deviner la généalogie linguistique (quelles langues sont le plus proches ou distantes).

Les options à gauche vous permettent de choisir les paramètres à comparer.



6.3.2 Network Graph

Le graphe du réseau est un autre moyen de voir les variétés semblables rassemblées.



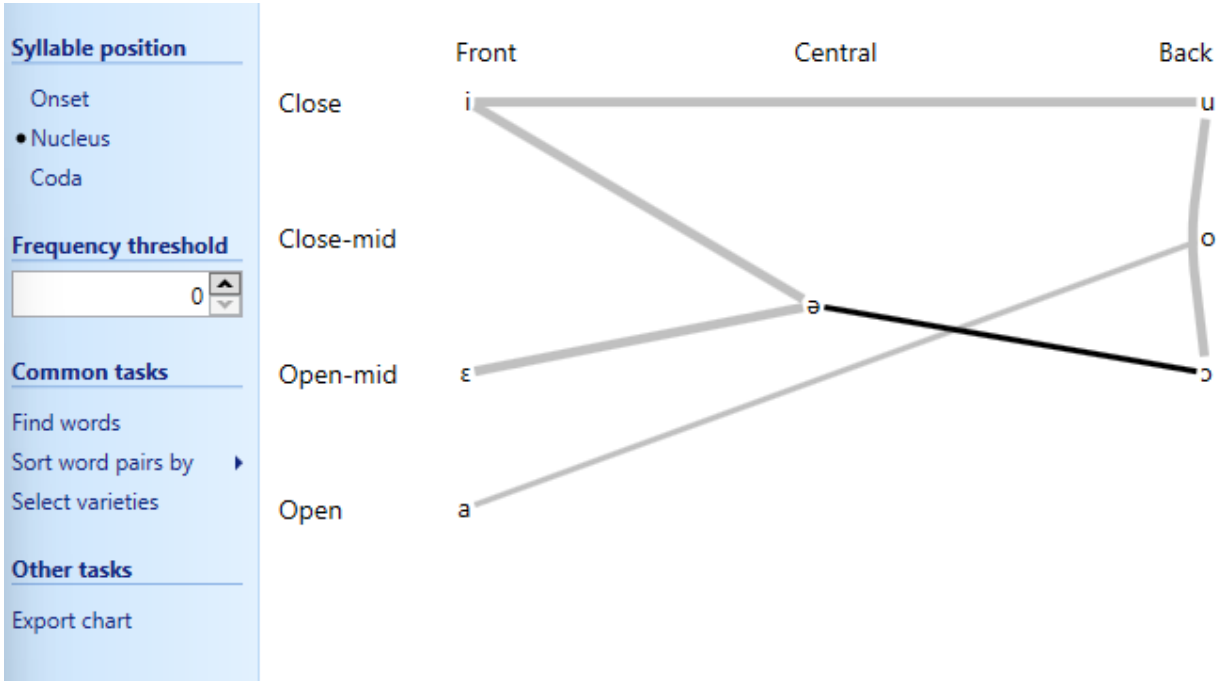
6.3.3 Géographie

Si vous avez les données cartographiques, vous pouvez marquer les régions ou chaque dialecte est parlé sur une carte mondiale.

Dessiner une forme avec l'outil et Cog vous donnera l'option de le lier avec une variété.

6.3.4. Global Correspondences

Dans l'outil "Global Correspondences", vous verrez les sons du debut de syllable (onset), les noyeaux (nucleus), et les terminaisons (coda). Cliquez sur chaque ligne et son pour voir la correspondence.



[1] https://www.youtube.com/watch?v=UX_q0FNlqUI

[2] https://software.sil.org/cog/