5.1 Editeur de dictionnaires de mots simples « Delas »

Quand l'utilisateur clique sur Dela -> Edit Delas, une fenêtre avec deux choix apparait, elle permet de choisir entre l'ouverture de tous les dictionnaires qui existent dans le dossier Delas, ou d'ouvrir un dictionnaire spécifique.

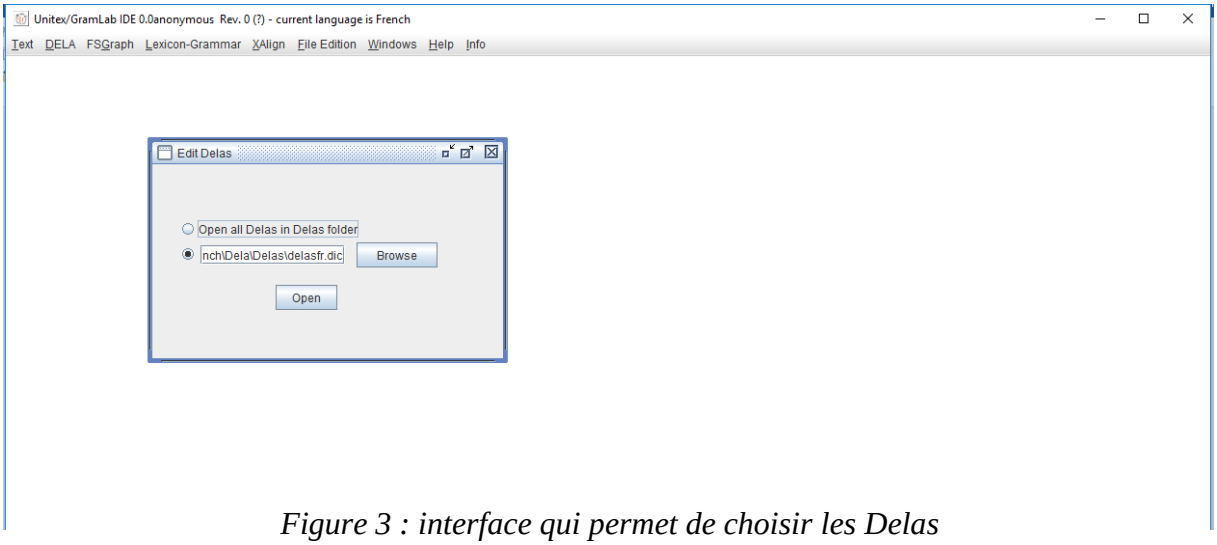


Figure 3 : interface qui permet de choisir les Delas

Ensuite, pour afficher la fenêtre d'édition du dictionnaire des mots simples, il suffit de cliquer sur le bouton Open dans la figure 3 (dans cette démonstration on a choisi d'ouvrir un seul dictionnaire) :

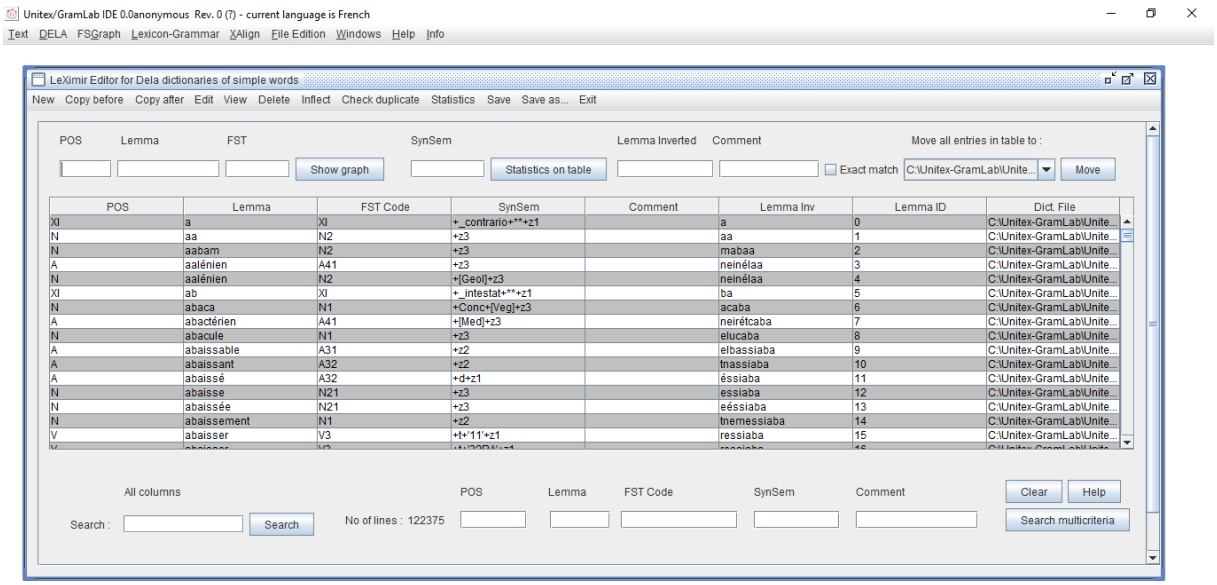


Figure 4 : interface de l'éditeur de Delas

L'importance de cet outil de gestion du mot simple réside dans le fait qu'il permet de faire de nombreux traitements sur le dictionnaire, notamment chercher efficacement un sous-ensemble de lemmes, en fonction du critère de recherche choisi par l'utilisateur. La recherche peut se

faire de plusieurs façons : soit par lemme, soit par catégorie grammaticale, soit par code de flexion, soit par commentaire.
On peut aussi trier les entrées par ordre alphabétique, en cliquant sur l'en-tête des colonnes. Par exemple la figure 5 montre le résultat de la recherche de tous les lemmes qui se terminent par « eur ». Ce type de filtrage de lemme est utile pour trouver les codes de flexion pour un nouveau lemme, étant donné que le code dépend souvent de la fin du mot.

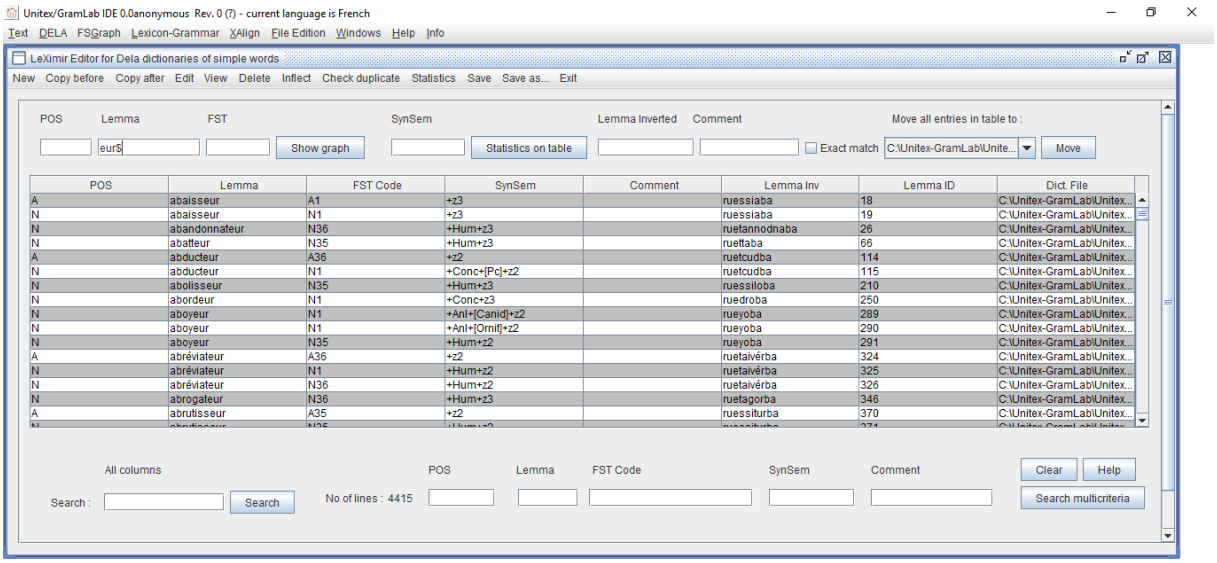


Figure 5 : Liste des lemmes qui se termine par « eur »

Pour réaliser des requêtes complexes on doit passer par la recherche multicritère qui se trouve en bas de la fenêtre. Par exemple, on veut avoir les lemmes qui sont des verbes, se terminent par « oir » et ont pour code de flexion V46. On peut voir figure 6 l'illustration de cette requête.

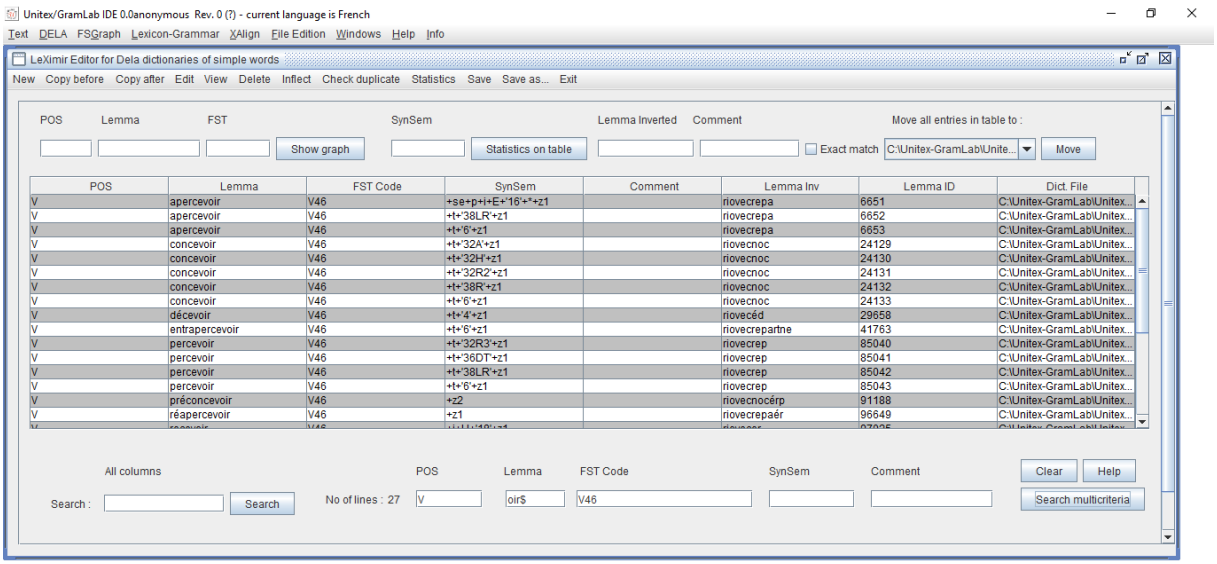


Figure 6 : Liste des lemmes qui sont des verbes et se terminent par « oir »

Une assistance à la construction d'une requête de recherche a été mise en place. Il suffit de cliquer sur le bouton « help ». La figure 7 illustre une capture d'écran sur cette aide :

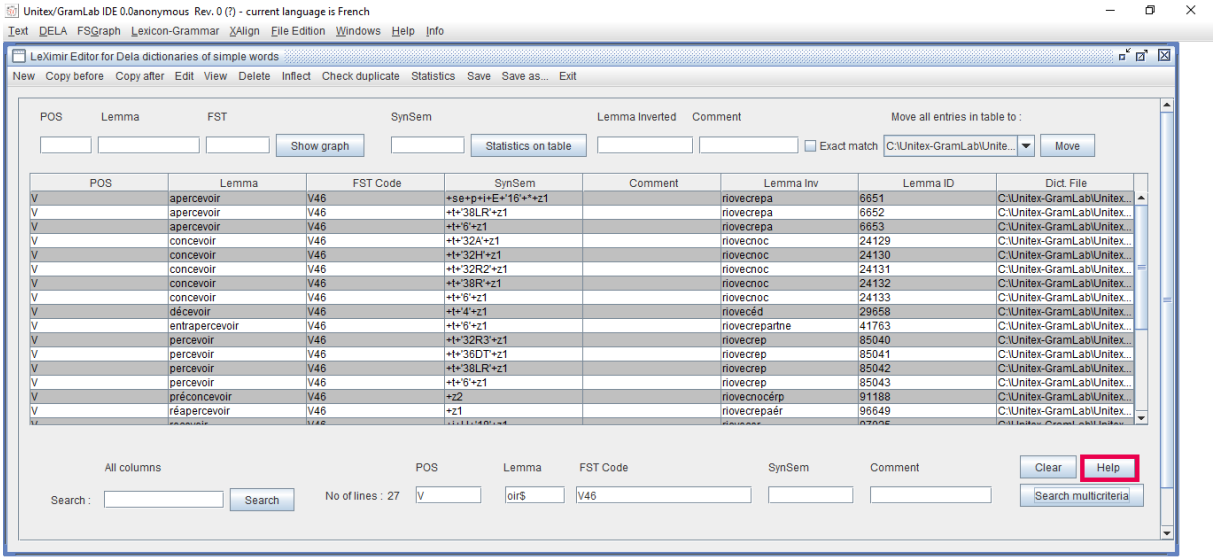


Figure 7 : Bouton pour ouvrir l'assistance à la construction d'une requête

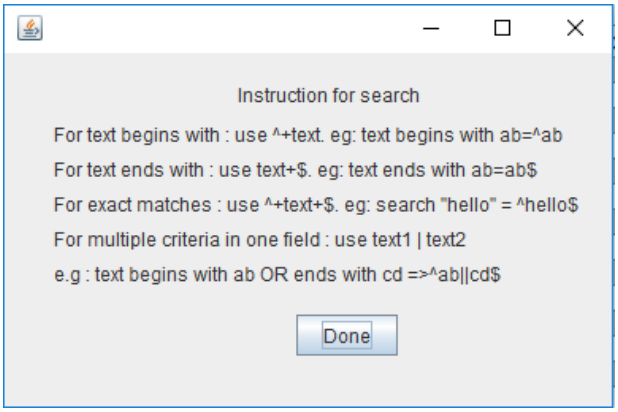


Figure 8 : manuel d'aide

Ajout d'un mot simple

Il y a deux façons d'ajouter un lemme (ligne dans la table). Soit le saisir à partir de zéro, soit en copiant un lemme existant, ce qui peut souvent faciliter et accélérer le travail. Ces fonctionnalités sont accessibles depuis le menu « New->Insert before » ou « New-> Insert after » ou « Copy before » ou « Copy after». Dans la fenêtre d'ajout, on peut aussi fléchir immédiatement le lemme avec le bouton « inflect » pour vérifier si le graphe de flexion convient pour ce lemme.

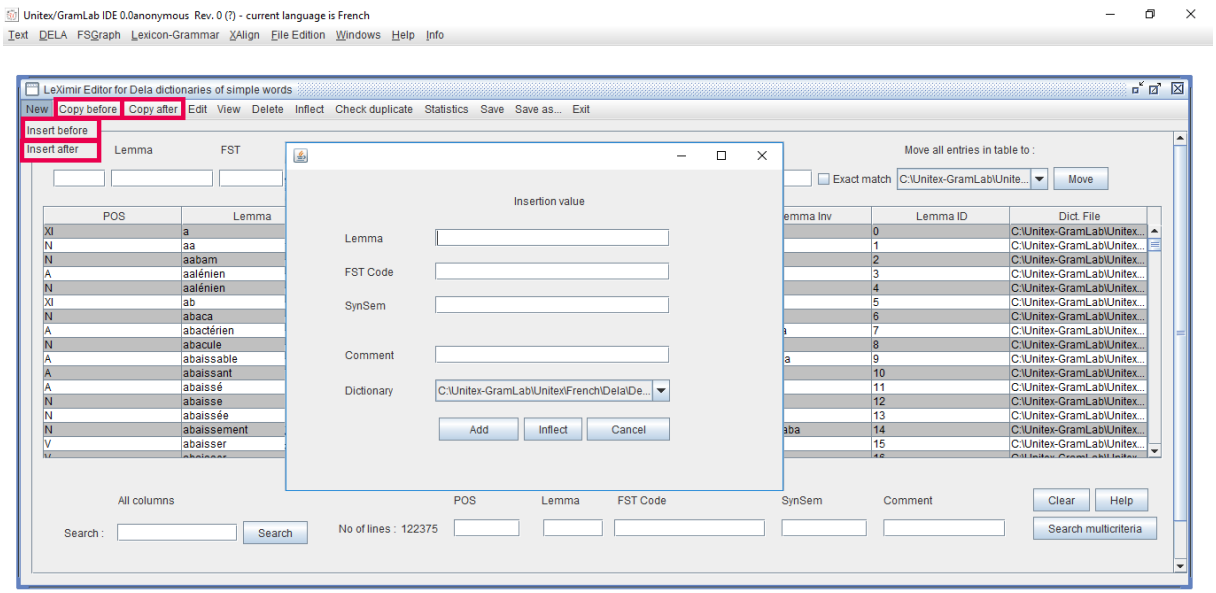


figure 9 : ajout d'un nouveau lemme.

Modification, affichage, suppression d'un mot simple

Pour modifier un lemme, il suffit de cliquer sur la ligne correspondante sur la table et de cliquer sur « Edit ». De même, pour afficher un lemme, il faudra cliquer sur « view » ou pour le supprimer, on clique sur le bouton « delete ».
On peut aussi voir les formes fléchies d'un mot de la table directement en cliquant sur « inflect ».

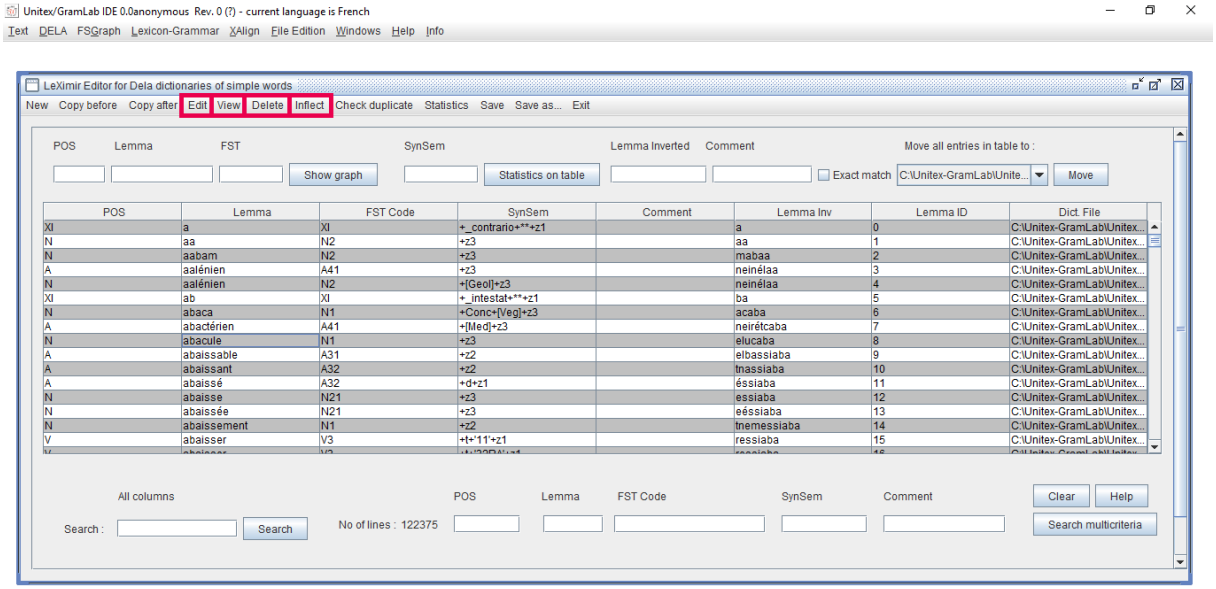


figure 10 : modification, suppression, flexion d'un lemme

Recherche des doublons dans les dictionnaires de mots simples

Le système peut détecter les doublons (même lemme et même code de flexion) même si ces derniers se trouvent dans 2 dictionnaires différents. La figure 12 illustre la fonctionnalité dans le système.

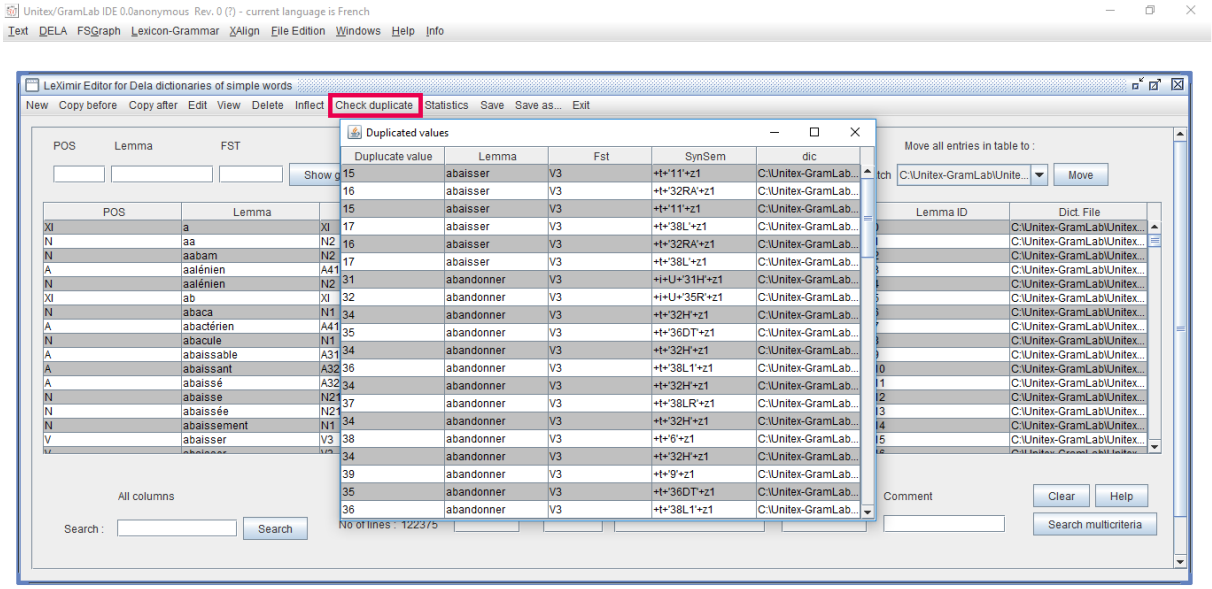


figure 11 : recherche des doublons dans les dictionnaires

Statistiques sur les mots simples

Un aperçu des statistiques des fréquences peut être généré dans un fichier CSV : par dictionnaire et POS. La figure 12 illustre un exemple de cette statistique.

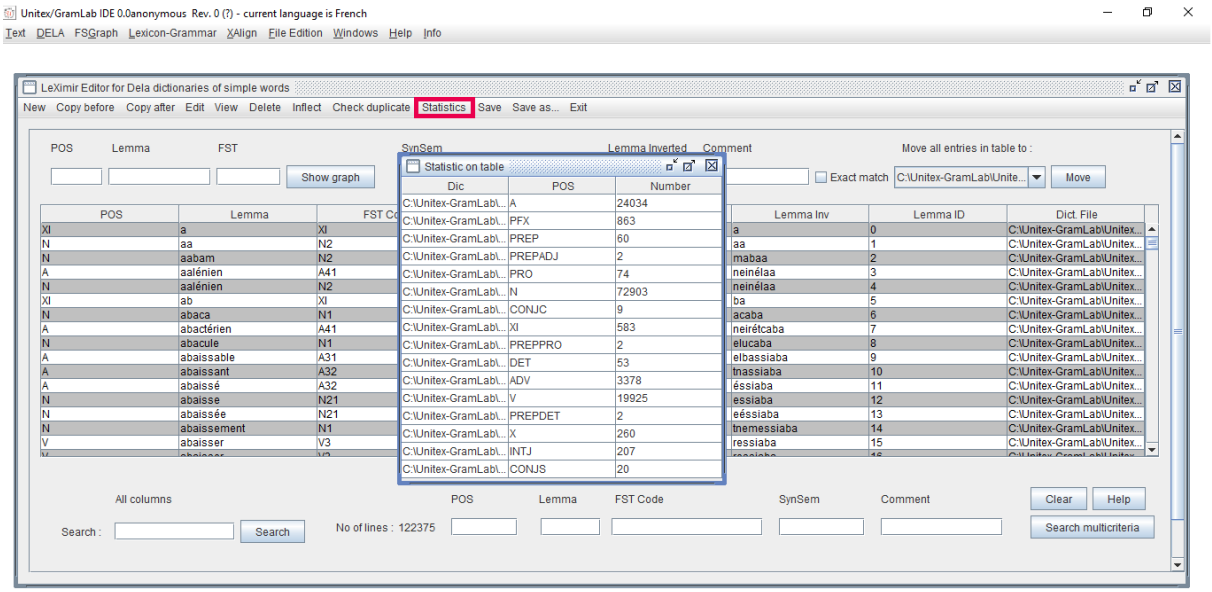


figure 12 : Statistiques sur les mots simples

Enregistrer et enregistrer sous…

Après avoir effectué toutes les manipulations, on peut ensuite sauvegarder les modifications. Il y a 2 façons d'enregistrer les modifications : le bouton « save » pour sauvegarder les modifications faites dans tous les dictionnaires et « save as… » pour sauvegarder dans un nouveau dictionnaire les entrées affichées dans le tableau. Exemple de la deuxième fonctionnalité : on veut afficher tous les lemmes qui se terminent par « er » et qui sont des verbes, ensuite on veut sauvegarder dans un nouveau dictionnaire des verbes du 1er groupe.

Show graph

Le bouton « show graph » est un moyen de visualiser le graphe de flexion de l'entrée sélectionnée dans le tableau. Il permet également de modifier ce graphe.
La figure 13 illustre cette fonctionnalité.

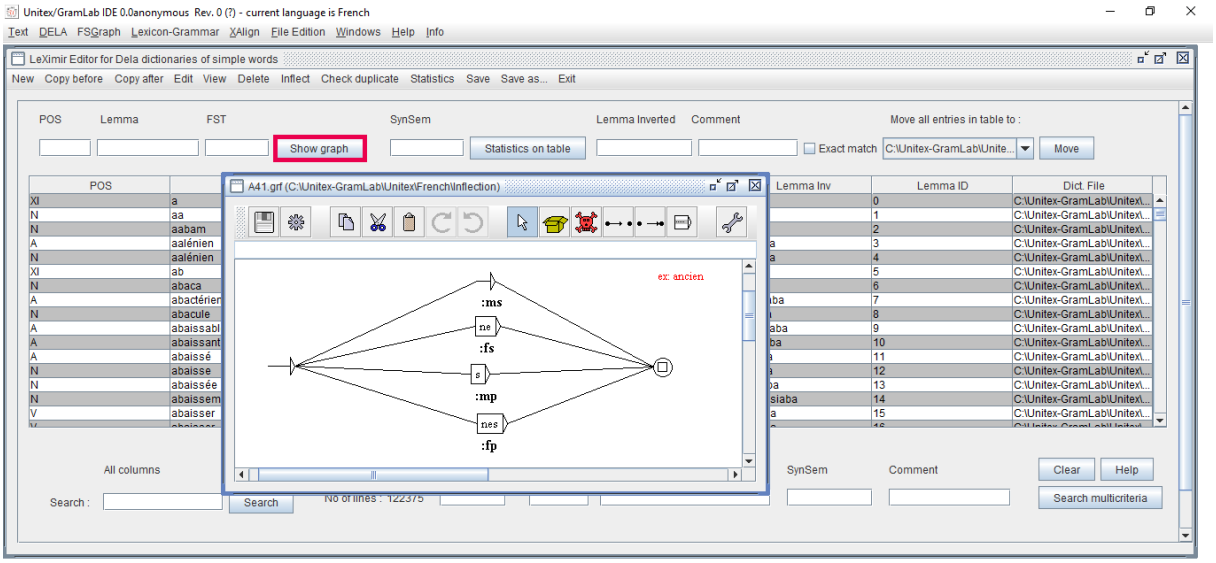


figure 13 : Show graph

Statistics on table

En cliquant sur le bouton " Statistics on table" à côté de SynSem, l'application génère l'inventaire de toutes les marques syntaxiques et sémantiques pour chaque POS. Le fichier généré est automatiquement en format CSV.

Move all entries in table to

Pour fusionner ou diviser des dictionnaires, il est possible de déplacer les lemmes dans un autre dictionnaire. Ainsi l'utilisateur peut filtrer l'ensemble des données en grille puis déplacer toutes les entrées sélectionnées (filtrées) dans le dictionnaire spécifié. La figure 14 affiche le bouton où l'on peut exécuter cette instruction.

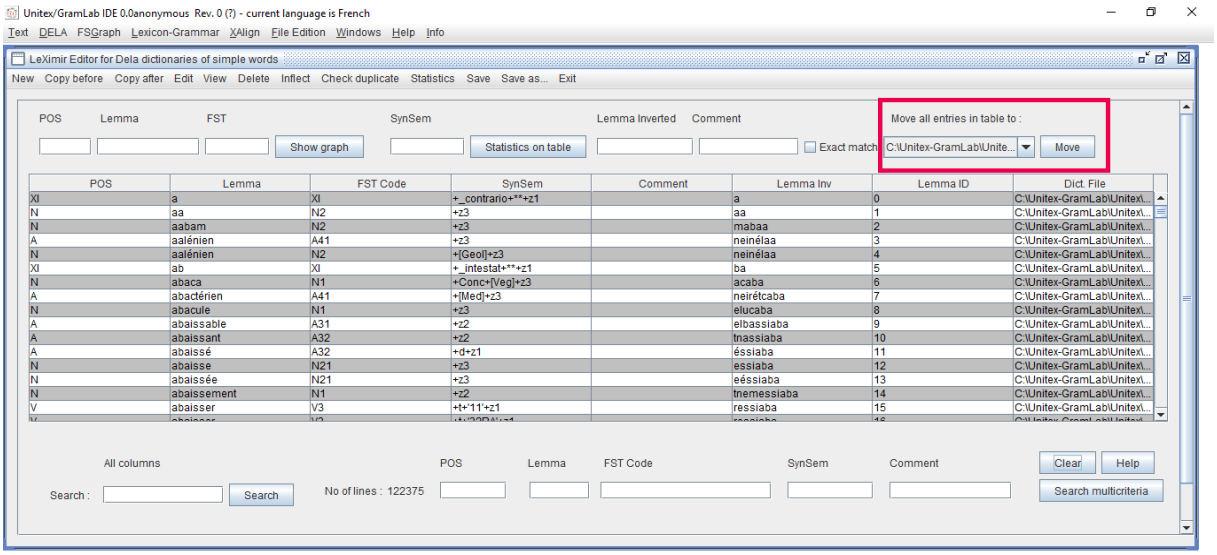


figure 14 : modalité de déplacement des lemmes dans un autre dictionnaire

5.2 Editeur de dictionnaires de mots composés « Delac »

L'affichage de présentation des mots composés est similaire au mot simple. On peut ajouter, modifier ou supprimer une entrée. On peut également voir les formes fléchies du mot. Et enfin, on peut avoir des statistiques de l'ensemble des lemmes dans les dictionnaires.

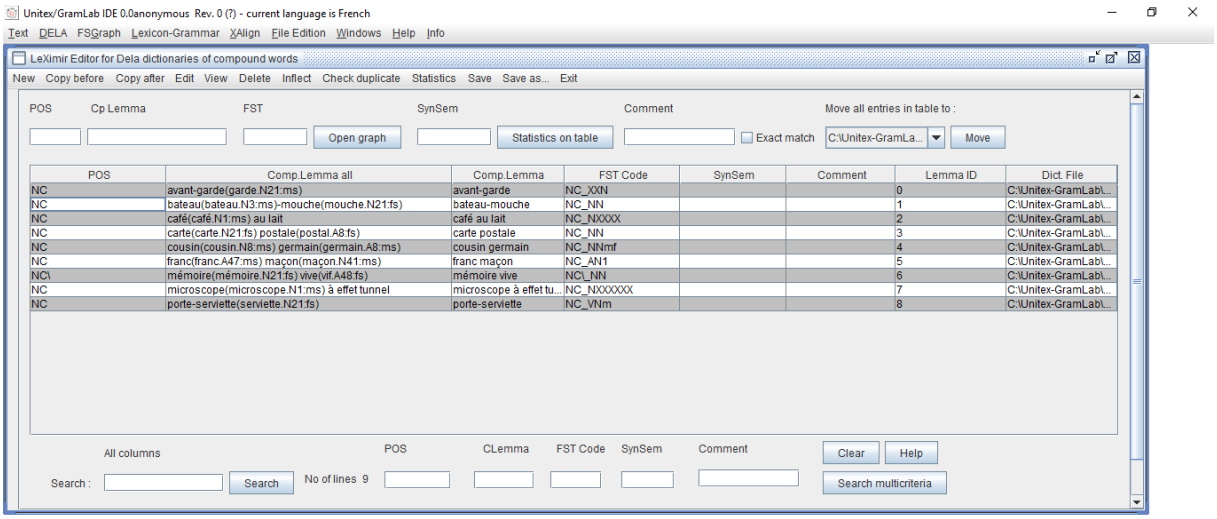


figure 15 : interface de l'éditeur de Delac

Ajout d'un mot composé

Les dictionnaires des mots composés ont une structure un peu plus complexe mais les principes de base de la gestion de la recherche et des données sont conservés. On peut identifier cette importante différence dans la recherche des lemmes. La fin des lemmes n'est pas prise en charge, car dans ce cas, il n'y a pas de signification particulière. Pour le formulaire de l'ajout d'un lemme ou la modification d'un lemme existant, il est un peu plus compliqué parce qu'on devra insérer plus d'informations.
La figure 16 montre la fenêtre d'ajout d'un mot compose. La partie supérieure du formulaire est composée d'informations générales sur le mot composé qui sont: le code de flexion du mot

composé, les catégories syntaxique et sémantique, les commentaires. Par ailleurs, la partie inférieure est composée d'informations sur les formes des mots simples qui constituent le mot composé, le graphe de flexion du mot simple et les traits grammaticaux.

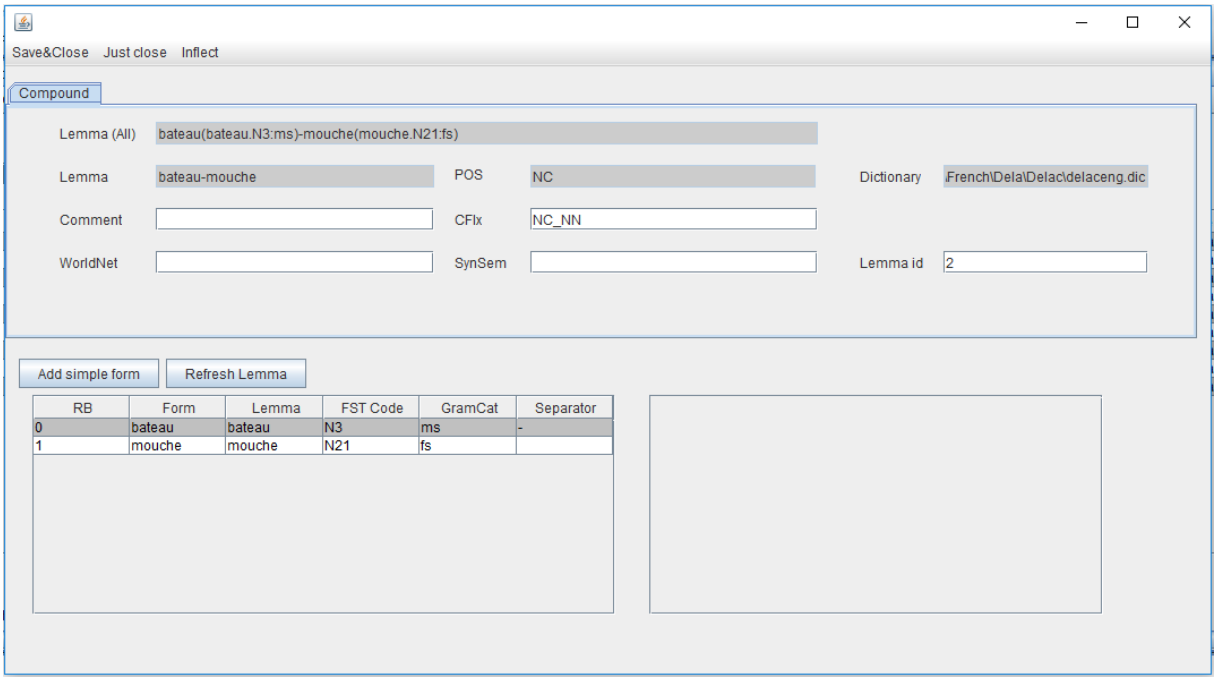


figure 16 : fenêtre d'ajout d'un mot compose.

Pour ajouter un lemme d'un mot composé, on doit ajouter les éléments dans la partie inférieure de la fenêtre. A chaque fois qu'on veut ajouter un élément, il suffit de cliquer sur « Add simple form » et ajouter les informations nécessaires.

Modification, affichage, suppression d'un mot composé

Pour modifier un lemme, il suffit de cliquer sur la ligne correspondante sur la table et cliquer sur « modifier ». Il en est de même pour afficher un lemme, il faudra cliquer sur « view » ou pour le supprimer, on clique sur le bouton « delete ». On peut aussi voir la flexion d'un mot sur la table directement en cliquant sur « inflect ».

Recherche des doublons dans les dictionnaires des mots composés

Le système peut reconnaître les doublons (même lemme et même code de flexion) même si ces derniers se trouvent dans 2 dictionnaires différentes.

Statistique des mots composés

Un aperçu des statistiques des fréquences par dictionnaire et POS peut être généré dans un fichier CSV en cliquant sur le menu « Statistics ».

Enregistrer et enregistrer sous…

Après avoir effectué ces manipulations, on peut ensuite sauvegarder toutes les modifications. Il y a 2 façons d'enregistrer les modifications : utiliser le bouton « save » pour sauvegarder les modifications faites dans tous les dictionnaires ou recourir à « save as… » pour sauvegarder dans un nouveau dictionnaire les entrées affichées dans le tableau.

Show graph

Le bouton « show graph » est un moyen de visualiser le graphe sélectionné dans le tableau, et permet également de modifier ce graphe.

Statistics on table

En cliquant sur le bouton " Statistics on table" à côté de SynSem, l'application génère l'inventaire toutes les marques syntaxiques et sémantiques par POS. Le fichier généré est automatiquement affiché.

Move all entries in table to

Pour fusionner ou diviser des dictionnaires, il est possible de déplacer les lemmes dans un autre dictionnaire. Dans ce cadre, l'utilisateur peut filtrer l'ensemble de données en grille et appliquer l'option pour déplacer toutes les entrées sélectionnées (filtrées) dans le dictionnaire spécifié.

5.3 Aide au codage de dictionnaires de mots composés « Delac »

On peut considérer cette partie comme la partie la plus compliquée de ce projet, elle consiste à produire des dictionnaires de mots composés « Delac » à partir d'un fichier texte qui contient un ensemble de mots composés, et d'un fichier XML qui contient des règles de codage. En effet il faut, pour chaque mot composé dans le fichier texte, trouver la bonne règle qui lui correspond parmi des centaines de règles présentes dans le fichier XML, puis générer un dictionnaire Delac pour ces mots.

Pour ouvrir cette fonctionnalité, il suffit de cliquer sur le menu « DELA>Encode Delac ». Une fenêtre pour choisir les dictionnaires Delaf et Delas qui vont être utilisés pour la production apparait, on peut soit utiliser tous les dictionnaires qui existent ou spécifier des dictionnaires précis.

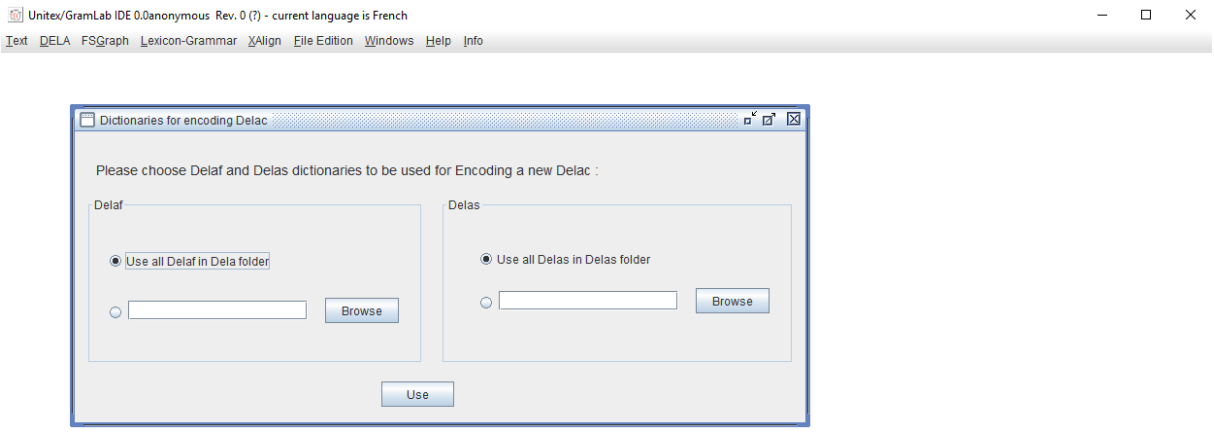


figure 17 : fenêtre pour choisir les dictionnaires Delaf et Delas

Après qu'on a cliqué sur « Use », l'interface principale pour produire les Delac apparait.
La figure 18 montre cette interface :

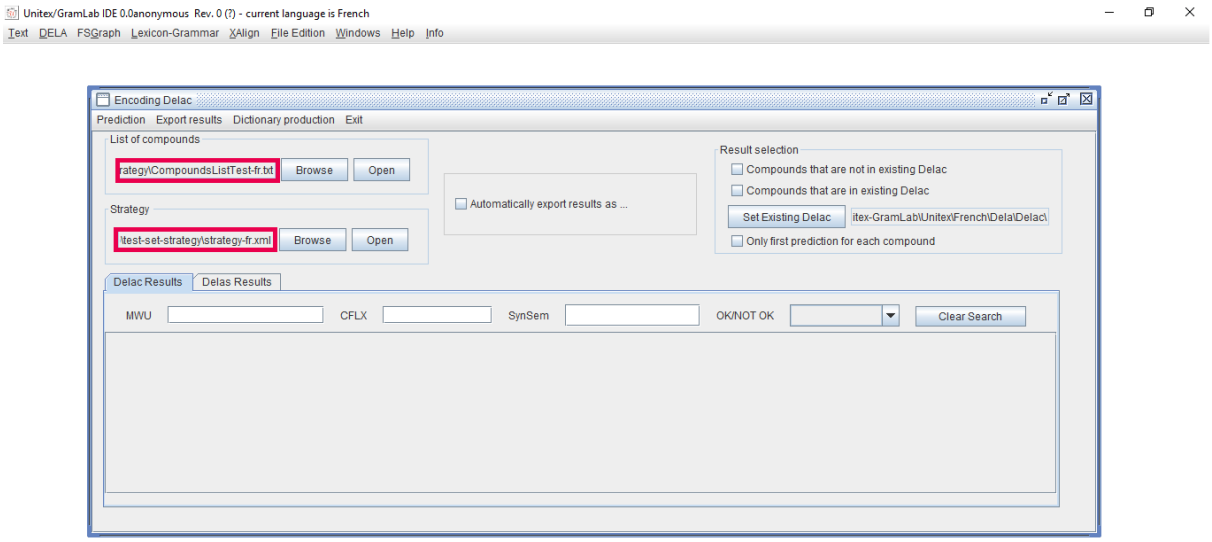


figure 18 : Interface de production de Delac

Comme on peut le constater l'interface permet de renseigner deux fichiers : « List of compounds » qui représente le fichier txt qui contient les mots composés, et « Strategy » qui est le fichier xml qui contient les règles de production.

Prediction

Quand on clique sur « Prediction », le programme parcourt tous les mots composés du fichier texte, et traite chacun d'eux de la manière suivante :

- Recherche des entrées DELAF puis DELAS pour chacun des mots qui constituent le mot composé, en utilisant les dictionnaires qu'on a renseignés dans l'étape précédente
- Affichage des résultats dans deux tableaux dans l'onglet « Delas Results » :

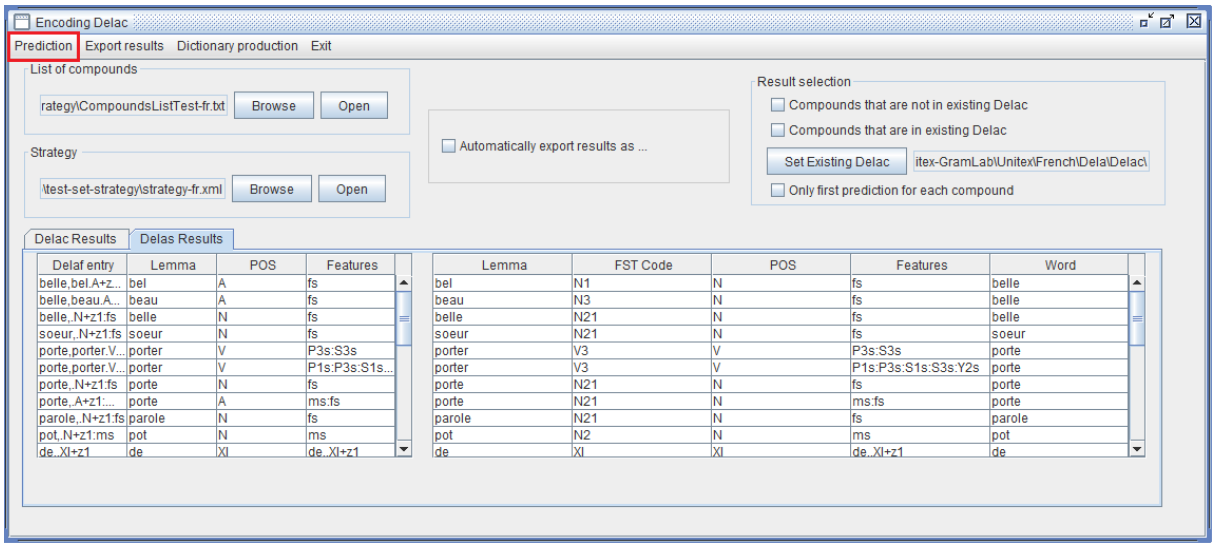


figure 19 : l'onglet Delas Results

- Utilisation des données qu'on a extraites des Delaf et Delas qui sont : le genre, le nombre, code de flexion… de chaque mot, pour trouver la bonne règle dans le fichier XML qui correspond à ce mot composé.
- Affichage des résultats (le mot composé, nombre de mots, code de flexion et les règles qui lui correspondent, codes syntaxiques et sémantiques …) dans un tableau dans l'onglet « Delac Results » :

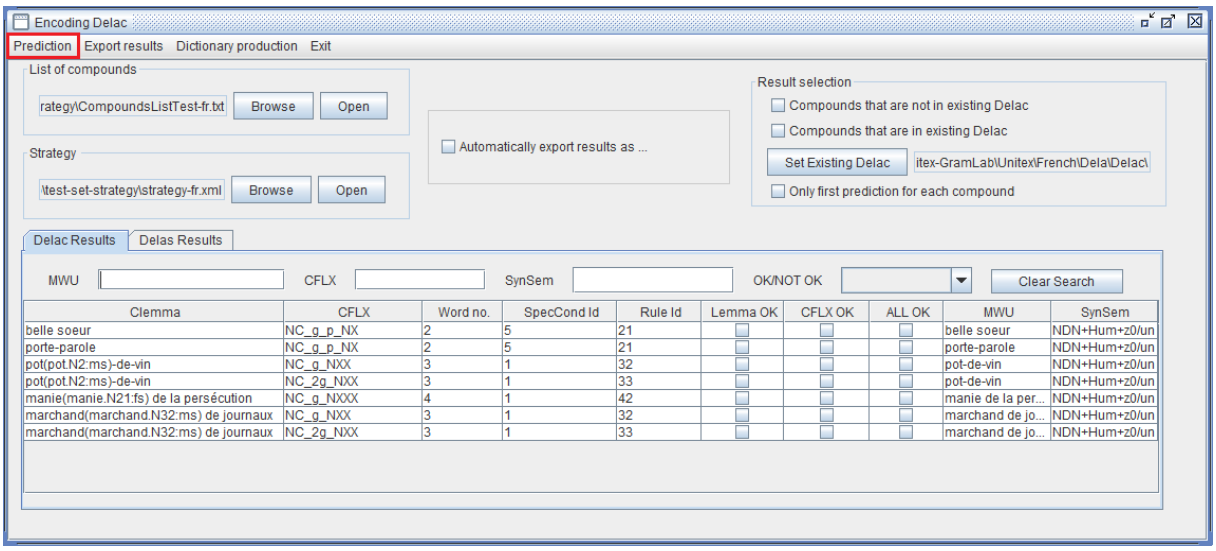


figure 20 : l'onglet Delac Results

Les filtres de recherche

Cette interface -comme les autres- permet aussi de réaliser une recherche dans le tableau de résultats, en utilisant des expressions régulières, dans l'exemple suivant on cherche les mots qui contiennent « de » :

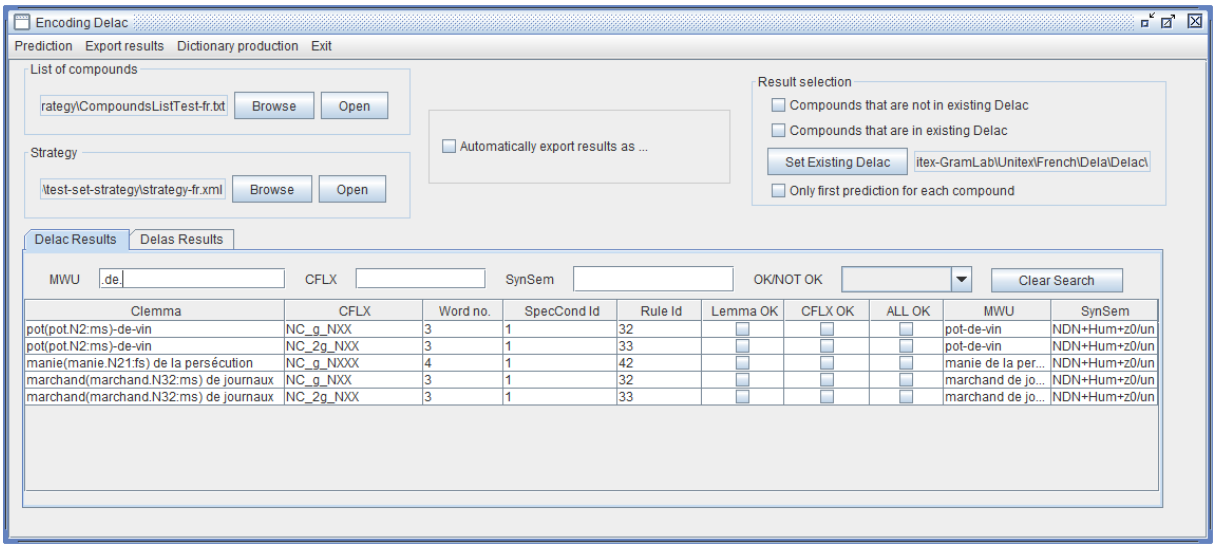


figure 21 : mots qui contiennent « de »

Result selection

Cette section permet d'une part de vérifier si les résultats qu'on a obtenus pour les entrées du nouveau Delac sont déjà dans le Delac existant ou non, d'autre part d'afficher seulement la première prédiction pour chacun des mots composé qui ont plusieurs prédictions.

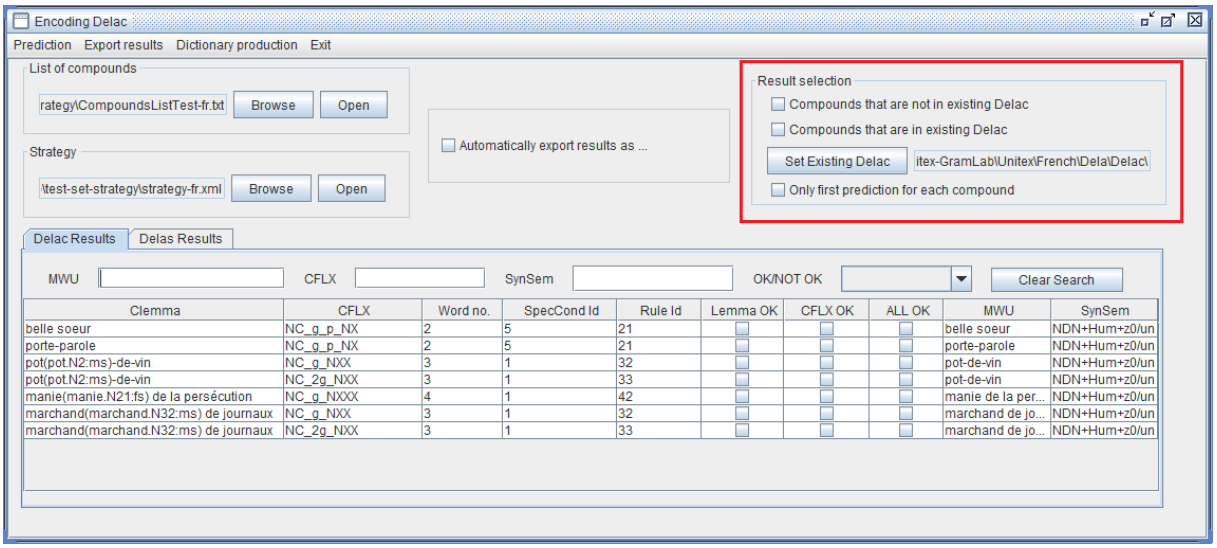


figure 22 : la section « Result selection »

Export results et Automatically export results

Export results permet d'exporter les résultats sous forme d'un fichier texte après qu'on a fini la prédiction, tandis que Automatically export results permet d'exporter les résultats au moment de la prédiction.

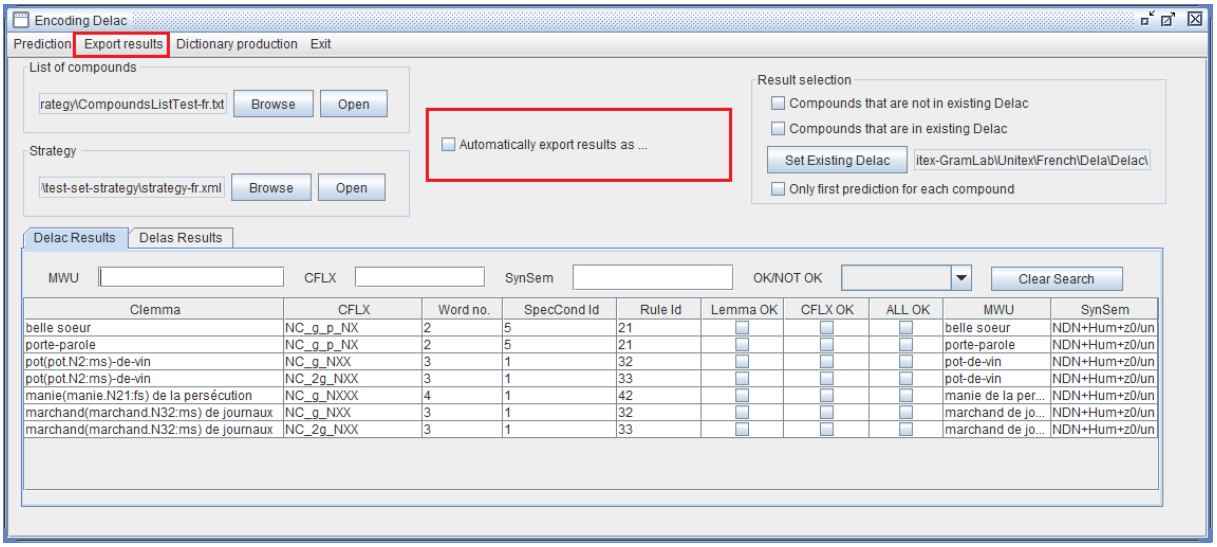
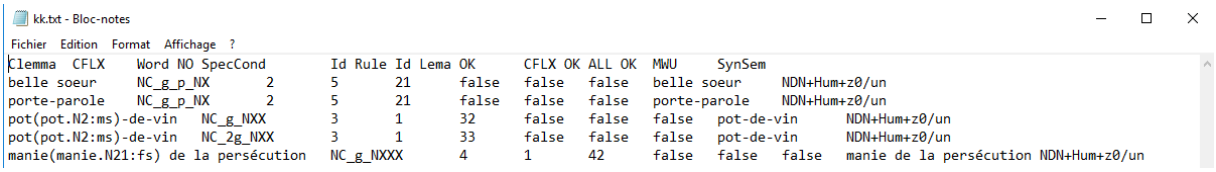


figure 23 : Export results

La figure suivante montre le résultat de l'exportation :



Dictionary production

Quand l'utilisateur a fini le traitement sur les mots composés, il peut produire le nouveau dictionnaire Delac en cliquant sur « Dictionary production ».

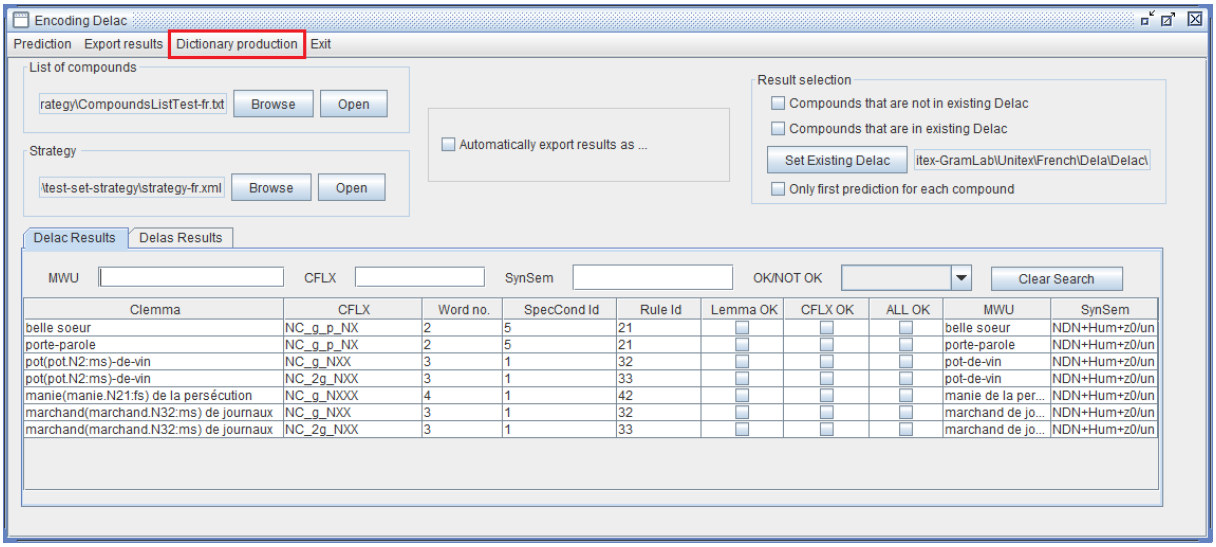


figure 24 : Dictionary production

La figure suivante montre le Delac qu'on a produit :

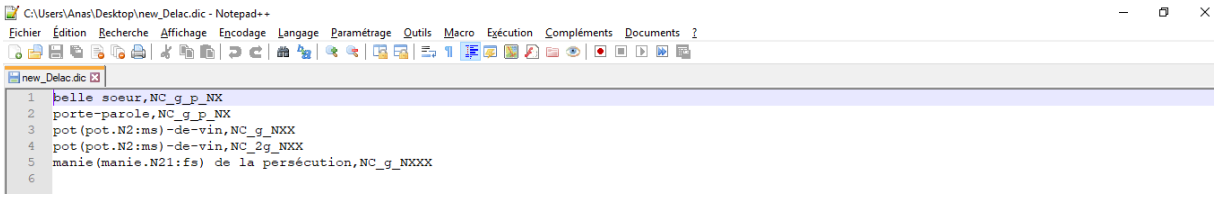


figure 24 : dictionnaire Delac produit

5.4 Les tests

Ces modules ont été tester par le professeur Cvetana Krstev, en l'utilisant sur des données professionnels.