# Documentation IRaMuTeQ
# 0.8 alpha 7
# version 0.1

**Auteurs :**

**Lucie Loubère**

**Pierre Ratinaud**

## Avertissements

**Iramuteq est en cours de développement. Son site est également en refonte complète… Pour retrouver les dernières évolutions** http://www.iramuteq.org/.

24/11/2025

# Table des matières


1 Présentation d'Iramuteq....4
1.1 Présentation générale....4
1.2 Principe de fonctionnement du logiciel....4
2 Analyses de textes....4
2.1 Un peu de vocabulaire....4
2.2 Format des données en entrée....4
2.2.1 Généralités....4
2.2.2 Les règles de formatages....5
2.3 Ouverture et indexation d'un corpus texte....6
2.3.1 Indexation....7
2.3.1.1 Onglet « Général »....7
2.3.1.2 Onglet Nettoyage....9
2.4 Les options sur le corpus....11
2.5 Traitements communs aux analyses (option de lemmatisation)....12
2.5.1 Lemmatisation....12
2.5.2 Paramètres des clés....13
2.5.3 Dictionnaires....14
2.6 Les différentes analyses textuelles....14
2.6.1 Statistiques textuelles....15
2.6.1.1 Paramétrage de l'analyse statistiques textuelles....15
2.6.1.2 Résultats de l'analyse statistiques textuelles....16
2.6.1.2.1 Onglet Résumé....16
2.6.1.2.2 Onglet Total....17
2.6.1.2.3 Onglet formes actives....17
2.6.1.2.4 Onglet Formes supplémentaires....18
2.6.1.2.5 Onglet Hapax....18
2.6.1.2.6 Options Supplémentaires de l'analyse statistiques textuelles....19
2.6.1.2.6.1 Sur chaque forme....19
2.6.1.2.6.2 Sur l'analyse....19
2.6.1.3 Fichiers en sortie....20
2.6.2 Spécificités et AFC....21
2.6.2.1 Paramétrage de l'analyse spécificités et AFC....21
2.6.2.2 Résultats Spécificités et AFC....21
2.6.2.2.1 Menu de réédition des AFC....24
2.6.2.2.2 Options Supplémentaires de l'analyse Spécificités et AFC....26
2.6.2.2.2.1 Sur chaque forme....26
2.6.2.2.2.2 Sur l'analyse....27
2.6.2.3 Fichiers en sortie....29
2.6.3 TGEN....31
2.6.3.1 Editeur de Tgen....31
2.6.3.2 Résultats....32
2.6.3.3 Options Supplémentaires des Tgens....33
2.6.3.4 Fichiers en sortie....34

2.6.4 Distances de Labbé....34
2.6.4.1 Paramètres....34
2.6.4.2 Résultats de l’analyse des distances de Labbé....35
2.6.4.2.1 Représentations de ward des distances de Labbé....36
2.6.4.2.2 Onglet heatmap des distances de Labbé....36
2.6.4.2.3 Onglet matrice des distances de Labbé....37
2.6.4.3 Option supplémentaire des distance de Labbé....38
2.6.4.4 Fichier en sortie....38
2.6.5 Classification Méthode Reinert....39
2.6.5.1 Paramétrages de l'analyse Classification Méthode Reinert....39
2.6.5.2 Résultats de la Classification Méthode Reinert....41
2.6.5.2.1 Onglet CHD....42
2.6.5.2.2 Onglet Profils....43
2.6.5.2.3 Onglet AFC de la méthode Reinert....45
2.6.5.2.4 Options supplémentaires de la Classification Méthode Reinert....46
2.6.5.2.4.1 Les options disponibles pour chaque classe....46
2.6.5.2.4.2 Les options disponibles pour la classification....51
2.6.5.3 Fichiers en sortie pour la classification....54
2.6.6 Analyses de similitude....56
2.6.6.1 Paramétrages de l'Analyse de Similitude sur textes....56
2.6.6.2 Résultats de l'analyse de similitude....60
2.6.6.3 Fichiers en sortie....60
2.6.7 Nuage de mots....62
2.6.7.1 Paramétrage de l'analyse Nuage de mots....62
2.6.7.2 Résultats de l'analyse Nuage de mots....62
2.6.7.3 Fichiers en sortie pour les nuages de mots....62
2.7 Réédition des graphiques dans les analyses textuelles....63
2.7.1 Menu de réédition des dendrogrammes....63
2.7.2 Menu de réédition des AFC....64

# 1 Présentation d'Iramuteq

## 1.1 Présentation générale

Iramuteq est un logiciel d'analyse de textes et de tableaux de données. Il s'appuie sur le logiciel de statistique R (http://www.r-project.org) et sur le langage python (http://www.python.org).

Iramuteq propose un ensemble de traitements et d'outils pour l'aide à la description et à l'analyse de corpus textuels et de matrices de type individus/caractères.

## 1.2 Principe de fonctionnement du logiciel

Iramuteq est une Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires, son fonctionnement consiste à préparer les données et écrire des scripts qui sont ensuite analysés dans le logiciel statistique R. Les résultats sont finalement affichés par l'interface.

Lors de l'ouverture d'un corpus ou d'une matrice, Iramuteq crée un dossier sur le même emplacement que le fichier ouvert (il le nomme NomDuCorpus_X). C'est ici qu'il stockera les résultats.

# 2Analyses de textes

## 2.1 Un peu de vocabulaire...

Pour des raisons de propriété intellectuelle, le vocabulaire de la méthode Reinert a été remplacé :

| Alceste | Méthode Reinert |
|---|---|
| **UCI** : Unité de contexte initiale | **TEXTE** |
| **UCE** : Unité de contexte élémentaire | **ST :** Segment de texte |
| **UC** : Unité de contexte | **RST :** regroupement de segments de texte |

## 2.2 Format des données en entrée

### 2.2.1 Généralités

Les fichiers d'entrée doivent être au format texte brut (.txt), et de préférence contenir les caractères de ponctuation.

Dans ce formatage, l'unité de base est appelée « Texte ». Un texte peut représenter un entretien, un article, un livre ou tout autre type de documents. Un corpus peut contenir un ou plusieurs textes (mais au minimum un).

### 2.2.2 Les règles de formatages



Les **textes** sont introduits par quatre étoiles (****) suivies d'une série de variables étoilées (variables illustratives) séparées par un espace.

```
Un Texte doit obligatoirement avoir au moins une variable étoilée
```

Il est possible de placer des variables étoilées à l'intérieur des textes en les introduisant en début de ligne par un tiret et une étoile (-*). On parle alors de thématiques. La ligne ne doit contenir que cette variable.



Exemple d'un corpus sans thématique :

```
**** *var1_1 *var2_2
texte texte texte texte texte texte texte texte texte texte texte texte
texte texte texte texte texte texte texte texte texte texte texte texte
texte texte

**** *var1_2 *var2_3
texte texte texte texte texte texte texte texte texte texte texte texte
texte texte texte texte texte texte texte texte texte texte texte texte
texte texte texte
```



Exemple d'un corpus avec thématique :

```
**** *var1_1 *var2_2

-*thematique1
texte texte texte texte texte texte texte texte texte texte texte texte
texte texte texte texte texte texte texte texte texte texte texte texte
texte texte texte texte texte texte texte texte texte texte texte texte
texte texte texte texte texte texte texte texte texte texte texte texte
texte texte

-*thematique2
texte texte texte texte texte texte texte texte texte texte texte texte
texte texte texte texte texte texte texte texte texte texte texte texte
texte texte texte texte texte texte texte texte texte texte texte texte
texte texte

**** *var1_2 *var2_3

-*thematique1
texte texte texte texte texte texte texte texte texte texte texte texte
texte texte texte texte texte texte texte texte texte texte texte texte
texte texte texte texte texte texte texte texte texte texte texte texte
texte texte
```

```
-*thematique2
texte texte texte texte texte texte texte texte texte texte texte texte
texte texte texte texte texte texte texte texte texte texte texte texte
texte texte
```



Dans un corpus avec thématique, tous les paragraphes d'un **texte** doivent appartenir à une thématique. La construction suivante n'est donc pas possible :

```
**** *var_1
texteA TexteA texteA texteA texteA
-*thematique
texte texte texte texte texte texte texte
texte texte texte texte texte texte texte
```

**La ligne « texteA texteA... » n'a pas de thématique**



Les variables étoilées et les thématiques introduites dans le corpus ne doivent pas contenir d'espaces ou de caractères spéciaux. Elles ne doivent contenir que des caractères parmi a-z, A-Z, 1-9 et des tirets bas (_).

```
*age 18 ans n'est pas un bon codage
*age_18 est un bon codage
*entretien_d'Emilie n'est pas un bon codage
*ent_emilie est un bon codage
```



Les codages de la forme *variable_modalité doivent être privilégiés pour les variables illustratives. Ils permettent des analyses complémentaires.

Exemple : *sex_h pour les hommes et *sex_f pour les femmes permet de repérer la variable sex et les modalités h et f.



Les analyses effectuées reposant sur le traitement informatique des données, leurs saisies se doivent d'être régulières, en respectant la case (Majuscule/Minuscule).

Exemple :

*sex_h, *sex_homme seront considérés comme deux modalités différentes

*sex_h, *Sex_h, *sexe_h seront considérés comme trois variables différentes

## 2.3 Ouverture et indexation d'un corpus texte

Fichier → Ouvrir un corpus

### 2.3.1 Indexation

Une fois le fichier texte choisi, une boîte de dialogue s'ouvre :

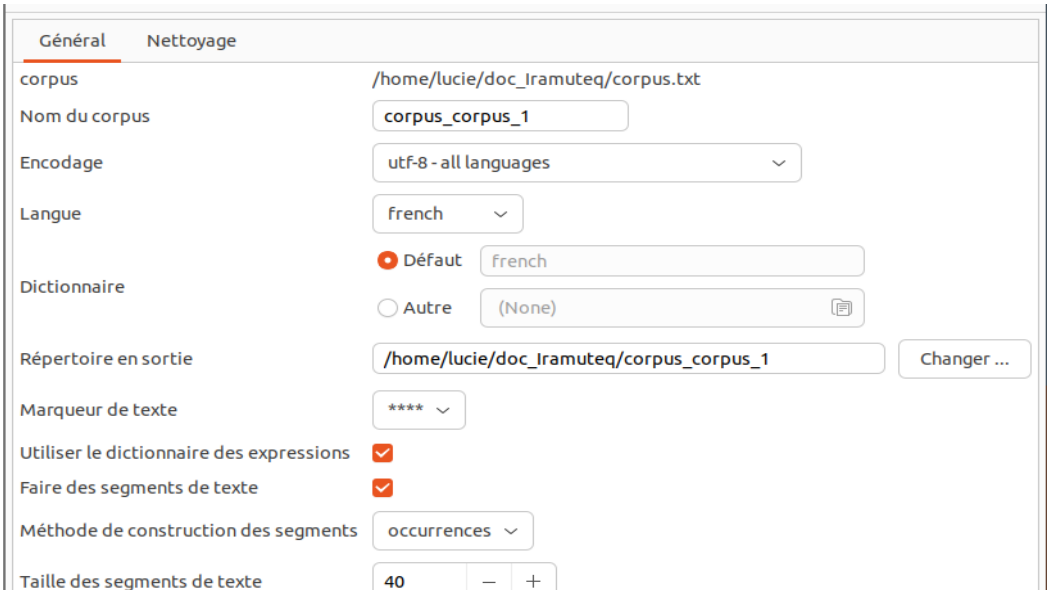



*Illustration 1: Dialogue d'indexation des corpus*

#### 2.3.1.1 Onglet « Général »

Le premier onglet permet de définir les caractéristiques générales du corpus, ainsi que les différentes options de l'indexation.

- **Corpus** : c'est le chemin d'accès du fichier que vous ouvrez (donné à l'étape précédente).
- **Nom du corpus** : permet de personnaliser le nom du corpus affiché dans le volet historique d'Iramuteq.
- **Encodage** : c'est l'encodage que vous avez utilisé lors de la création du corpus, nous vous recommandons de travailler en UTF8.

```
Par défaut l'option UTF-8 sera proposée sur les ordinateurs sous Mac OS X et Linux, et CP 1252 sur ceux sous Windows.
```

```
Faites très attention à l'encodage des corpus lors de leur création. Il est préférable d'utiliser un éditeur de texte capable de gérer l'encodage comme la suite LibreOffice par exemple.
```

- **Langue** : c'est la langue dans laquelle est écrit votre corpus




```
Iramuteq ne peut pas analyser simultanément plusieurs langues dans un même corpus.
```

- **Dictionnaire** : Par défaut, utilise ceux installés par Iramuteq. Cela permet

d'utiliser ses propres dictionnaires. Si vous en créez un, vous devrez respecter le format défini par Iramuteq, à savoir la forme, le lemme et la catégorie grammaticale, dans cet ordre-là et séparées par une tabulation. Les catégories grammaticales sont détaillées en 2.4.2



```
chat	chat	nom
chats	chat	nom
chatons	chaton	nom
```

- **Répertoire en sortie** : Après l'ouverture du corpus, Iramuteq crée un répertoire dans lequel seront stockés le corpus et les résultats. Par défaut il se situe au même emplacement que le corpus, et par défaut se nomme NomFichier_corpus_X. Son emplacement et son nom peuvent être changés ici.
- **Marqueur de texte** : Indique au logiciel comment chaque texte est introduit. Pour des raisons de compatibilités avec les anciens corpus Alceste® il est possible d'utiliser 4 entiers à la place des 4 étoiles.
- **Utiliser le dictionnaire des expressions** : Les dictionnaires des expressions contiennent des expressions ou des mots contenant des tirets (-) des apostrophes (') ou des espaces. Il permet de traiter ces expressions comme un tout, évitant ainsi de casser en deux formes les mots composés.

```
Les dictionnaires des lexiques et des expressions de chacune des langues sont disponibles dans le répertoire .Iramuteq/dictionnaires dans l'environnement de l'utilisateur.
```



| Mot d'origine | Sans le dictionnaire d’expressions | Avec le dictionnaire d’expressions |
|---|---|---|
| vis-à-vis | vis à vis (trois occurrences et deux formes différentes) | vis_à_vis (une occurrence, une forme) |
| aujourd'hui | aujourd hui (2 formes, 2 occurrences) | aujourd_hui (une occurrence, une forme) |

- **Construire des segments de textes** : permet le découpage de chaque texte en segments.

- **Mode de construction des segments de texte** : permet de choisir sur quelle unité se construiront les segments. Au choix :

- caractères : les segments sont déterminés sur un critère de taille lié au nombre de caractères qu'ils contiennent.
- occurrences :les segments sont déterminés sur un critère de taille lié au nombre d'occurrences qu'ils contiennent.



Les segments de texte sont construits à partir d'un critère de taille et de ponctuation. Iramuteq cherche le meilleur ratio taille/ponctuation (par ordre de priorité, les ".", "?" "!" en premier, puis en second ";" et les ":" en troisième la virgule et en dernier l'espace). L'objectif est d'avoir des segments de tailles homogènes en respectant le plus possible la structure du langage.

- paragraphe : Le retour à la ligne est utilisé comme marqueur de fin de segment de texte. La ponctuation n'est plus du tout prise en compte.

- **Taille des segments de textes** : Définit le nombre d'unités (voir plus haut) contenues dans le segment de texte. Cette option n'agit pas sur le découpage en paragraphes.



La modification du mode de construction de segments de texte implique la modification de la taille de ces derniers. L'analyse de segments composés de 40 caractères est rarement pertinente. ;-)

#### 2.3.1.2 Onglet Nettoyage

Le second onglet de cette boîte de dialogue concerne le nettoyage du corpus :

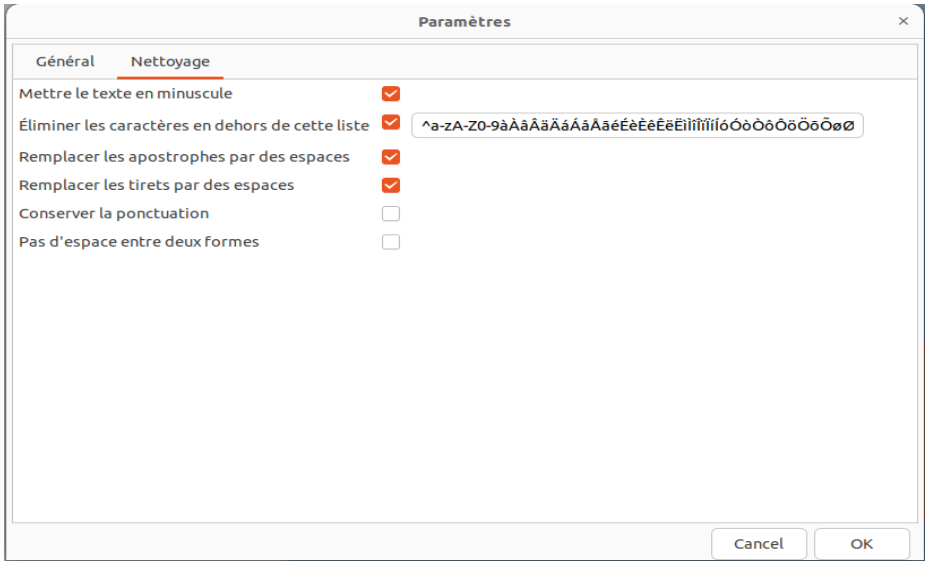



*Illustration 2: Dialogue de nettoyage du corpus*

- **Mettre le texte en minuscule** : Convertit toutes les majuscules en minuscules, si désactivé, Iramuteq considérera « Chat » et « chat » comme deux formes distinctes.
- **Éliminer les caractères en dehors de cette liste** : Par défaut seuls les caractères alphanumériques et accentués sont conservés, si vous ne voulez pas conserver un de ces caractères, il vous suffit de le retirer de la liste. Si vous voulez conserver d'autres caractères, vous pouvez en ajouter à la fin de la liste.

```
Pour les langues n'utilisant pas les caractères alphanumériques il est impératif de modifier cette liste.
```




```
Pour le bon fonctionnement du logiciel, vous ne devez pas forcer la conservation du signe double cote "
```

- **Remplacer les apostrophes par des espaces** : remplace les apostrophes par des espaces
- **Remplacer des tirets par des espaces** : remplace les – par des espaces
- **conserver la ponctuation** : conserve la ponctuation.




```
Si vous conservez la ponctuation, vous devez vérifier l'absence de point-virgule dans votre corpus.
```

- **Pas d'espace entre deux formes :** si cette fonction est activée Iramuteq n'utilisera plus l'espace comme délimiteur de forme.

## 2.4 Les options sur le corpus

Une fois le corpus indexé (voir chapitre 2.3), dans le bandeau gauche d'Iramuteq s'affiche le raccourci NomFichier_corpus_X, un clic droit dessus vous donne accès à un menu contextuel.

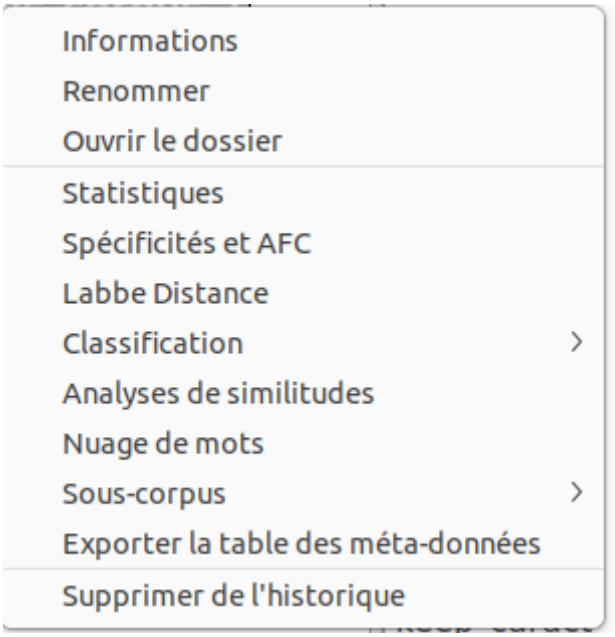



*Illustration 3: menu d'option sur le corpus*

- **Informations** : affiche les informations de l'indexation de ce corpus.
- **Renommer** : permet d'afficher un nom personnalisé dans le volet historique afin de reconnaître les différentes versions d'un corpus indexé différemment (ce nom ne sera utilisé que dans ce volet, non pas dans les dossiers créés).
- **Ouvrir le dossier** : permet d'ouvrir directement le dossier créé par Iramuteq. C'est dans ce dernier que nous pourrons retrouver les fichiers exportés lors des analyses ou exports.
- **Statistiques** : voir 2.6.1.
- **Spécificités** : voir 2.6.2.
- **Distance de Labbé** : voir 2.6.3.
- **Classification** : voir 2.6.4.
- **Analyses de similitudes** : voir 2.6.5.
- **Nuages de mots** : voir 2.6.6.

- **Sous-corpus** : voir 2.6.7.
- **Exporter la table des méta-données** : voir 2.6.8.
- **Supprimer de l'historique** : cette fonction supprime le corpus du volet historique; par contre, le dossier créé lors de l'ouverture du corpus sera conservé.

## 2.5 Traitements communs aux analyses (option de lemmatisation)

À chaque demande d'analyse sur un corpus, le dialogue suivant apparaît :




*Illustration 4: Dialogue de lemmatisation*

- **Lemmatisation** : Cette option choisit si le corpus doit être lemmatisé ou pas.
- **Paramètres des clés** : permet de choisir les clés d'analyse.
- **Dictionnaire** : permet d'utiliser des dictionnaires personnalisés.

### 2.5.1 Lemmatisation

Les verbes sont ramenés à l'infinitif, les noms au singulier et les adjectifs au masculin singulier. Iramuteq réalise la lemmatisation à partir de dictionnaires, sans désambiguïsation. Les dictionnaires pour les différentes langues sont disponibles dans le répertoire .Iramuteq/dictionnaires.

UN EXEMPLE

```
mangé, mangeons, mangera → manger
professionnelles, professionnelle, professionnels →professionnel
```

### 2.5.2 Paramètres des clés

La plupart des analyses sur corpus vont différentier des formes dites « pleines » (ou formes actives) et des mots outils (ou « formes supplémentaires »). Seules les formes « actives » participent à certaines analyses, comme les classifications par exemple. En cliquant sur « préférences », vous accédez à un menu permettant d'affecter les valeurs "active", "supplémentaire" ou "éliminé" pour chaque type grammatical.

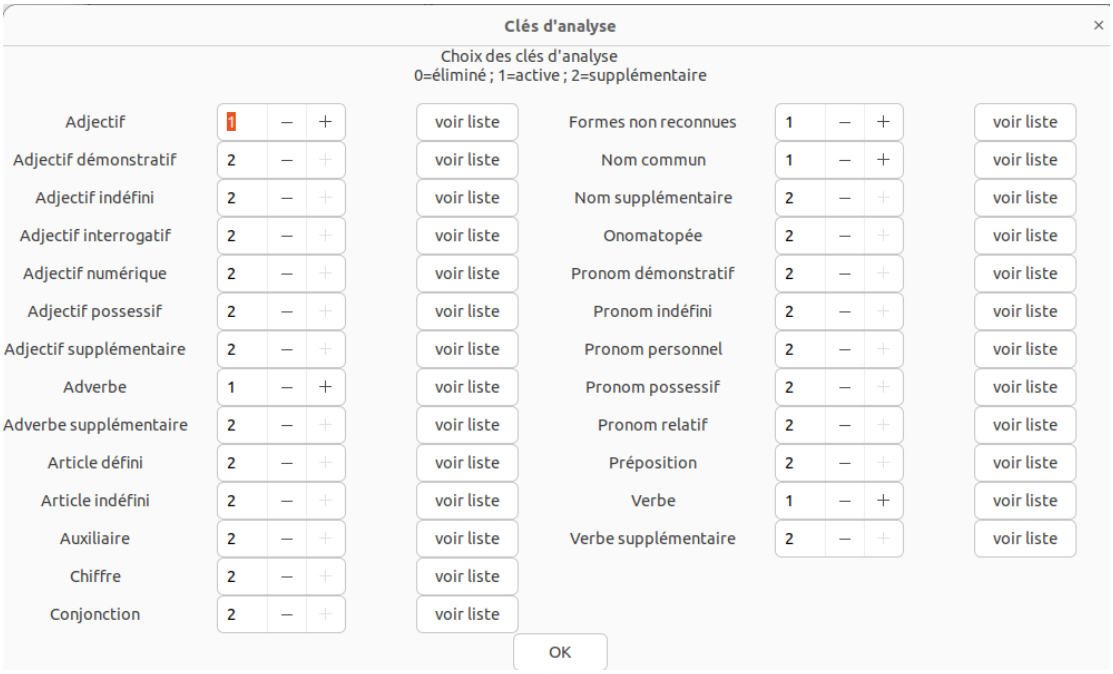



*Illustration 5: Dialogue choix des clés d'analyse*

Ces données sont stockées dans le fichier key.cfg accessible dans le répertoire .lramuteq. Dans ce fichier :

- la valeur 1 signifie que le type grammatical est actif,
- la valeur 2 signifie que le type est supplémentaire,
- la valeur 0 signifie que le type n'est pas pris en compte.

Voici la liste des types grammaticaux différentiés :

| | |
|---|---|
| adj_sup: Adjectif supplémentaire | ver: Verbe |
| art_ind: Article indéfini | adj_num: Adjectif numérique |
| adj_pos: Adjectif possessif | pro_rel: Pronom relatif |
| adv_sup: Adverbe supplémentaire | adj_ind: Adjectif indéfini |
| pro_dem: Pronom démonstratif | pro_ind: Pronom indéfini |
| art_def: Article défini | pro_pos: Pronom possessif |
| con: Conjonction | aux: Auxiliaire |
| pre: Préposition | ver_sup: Verbe supplémentaire |
| ono: Onomatopée | adj: Adjectif |
| adj_dem: Adjectif démonstratif | adj_int: Adjectif interrogatif |
| nom_sup: Nom supplémentaire | nom: Nom commun |
| adv: Adverbe | num : Chiffre |

### 2.5.3 Dictionnaires

Les dictionnaires permettent la lemmatisation et l'identification des formes actives et supplémentaires. Par défaut, utilise le dictionnaire sélectionné lors de l'indexation du corpus. Cependant, certaines analyses peuvent requérir un dictionnaire différent, vous pouvez donc choisir ici un nouveau fichier. Il devra être au même format que ceux d'Iramuteq (voir 2.3.1.1) et conserver les mêmes catégories grammaticales (voir 2.5.2). Ces dictionnaires ponctuels peuvent vous permettre de tester l'ajout d'un lexique spécifique à votre corpus ou encore d'effectuer des regroupements de formes sous des catégories (liste d'usages non exhaustive).




**Exemple d'un regroupement possible**

| | | |
|---|---|---|
| chat | félin | nom |
| chats | félin | nom |
| chatons | félin | nom |

## 2.6 Les différentes analyses textuelles

Les analyses de textes sont accessibles par 3 voies différentes. Une fois le corpus ouvert, le menu analyse de texte s'affiche. En dessous des menus, une barre d'icônes reprend également les différentes analyses. Enfin, un clic droit sur le titre de votre corpus dans le volet historique vous permet d'accéder aux analyses.

| | |
|---|---|
|  | Statistiques |
|  | Spécificité et AFC |
|  | Distance de Labbé |
|  | Classification |
|  | Analyse de similitudes |
|  | Nuages de mots |
|  | Sous corpus par méta-données |
|  | Sous corpus par thématiques |
|  | Export des méta-données |

*Illustration 6: légende du menu d'analyses de texte*

### 2.6.1 Statistiques textuelles

Cette analyse propose des statistiques simples sur les corpus textuels : effectif de toutes les formes, effectif des formes actives et supplémentaires, liste des hapax.

#### 2.6.1.1Paramétrage de l'analyse statistiques textuelles

Il n'y a pas de paramétrage pour cette analyse.

### 2.6.1.2Résultats de l'analyse statistiques textuelles

#### 2.6.1.2.1 Onglet Résumé

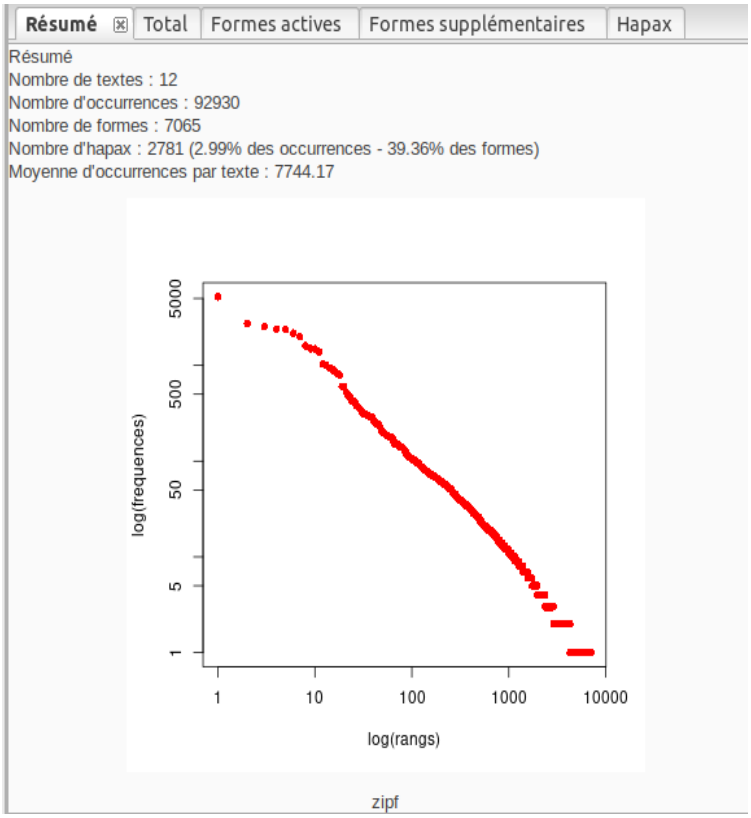


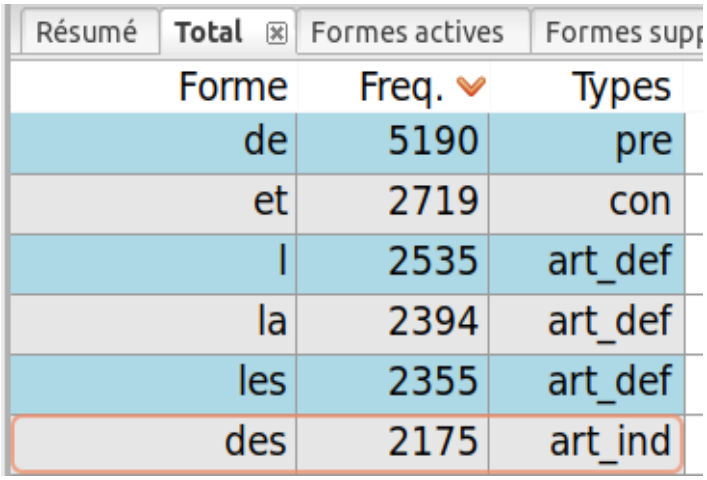

*Illustration 7: Présentation des résultats de l'analyse statistique textuelle*

Sur cet écran vous trouverez plusieurs informations :

- **Nombre de Textes** : c'est le nombre de Textes contenus dans le corpus.
- **Nombre d'occurrences** : c'est le nombre total de mots contenus dans le corpus. Il diffère si le corpus est lemmatisé ou non.
- **Nombre de formes** : c'est le nombre de formes présentes dans le corpus, il diffère si le corpus est lemmatisé ou non.
- **Moyenne d'occurrences par forme** : (nombre d'occurrences) **/** (nombre de formes).
- **Nombre d'Hapax** : c'est le nombre de mots n'apparaissant qu'une seule fois dans tout le corpus.
- **Moyenne d'occurrences par Textes**: (nombre d'occurrences) **/** (nombre de Textes)
- **Le graphique** : présente en abscisse les logarithmes des rangs et en ordonnées les logarithmes des fréquences des formes.

#### 2.6.1.2.2 Onglet Total

Liste de toutes les formes actives et supplémentaires (colonne forme) avec leurs effectifs (colonne nb) et leurs catégories grammaticales (colonne type).

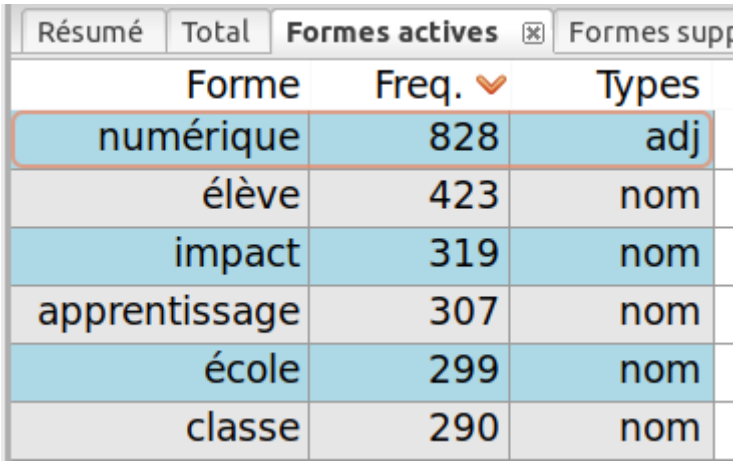



| Forme | Freq. | Types |
|---|---|---|
| de | 5190 | pre |
| et | 2719 | con |
| l | 2535 | art_def |
| la | 2394 | art_def |
| les | 2355 | art_def |
| des | 2175 | art_ind |

*Illustration 8: Présentation des résultats de l'analyse textuelle, onglet Total*

#### 2.6.1.2.3 Onglet formes actives

Liste des formes actives (colonne forme) avec leurs effectifs (colonne nb) et leurs catégories grammaticales (colonne type).

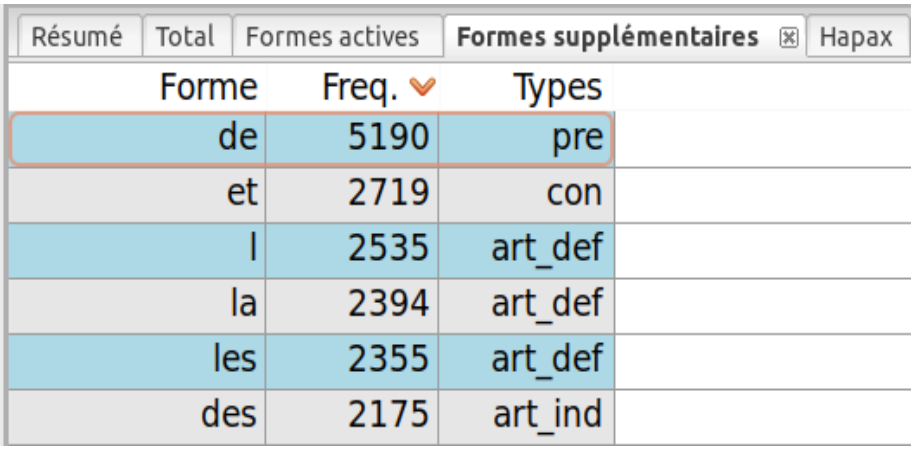



| Forme | Freq. | Types |
|---|---|---|
| numérique | 828 | adj |
| élève | 423 | nom |
| impact | 319 | nom |
| apprentissage | 307 | nom |
| école | 299 | nom |
| classe | 290 | nom |

*Illustration 9: Présentation des résultats de l'analyse textuelle, onglet formes actives*

#### 2.6.1.2.4 Onglet Formes supplémentaires

Liste des formes supplémentaires (colonne forme) avec leurs effectifs (colonne nb) et leurs catégories grammaticales (colonne type).



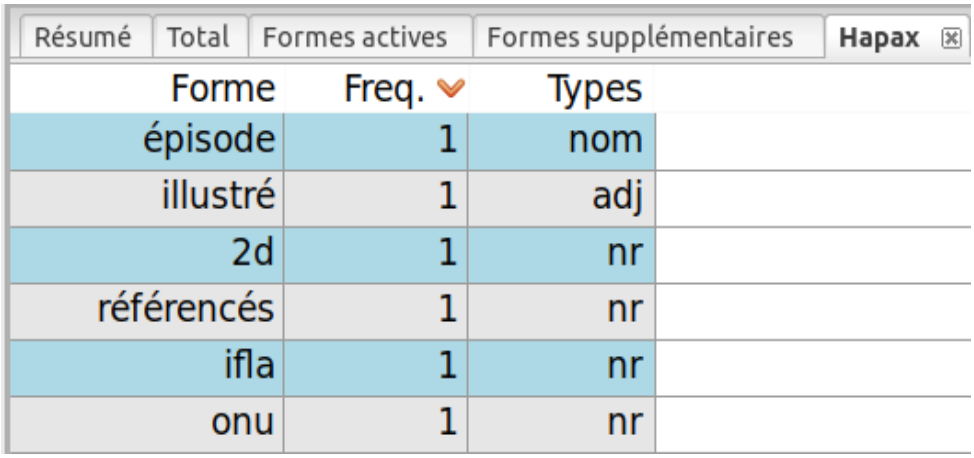
Résumé | Total | Formes actives | **Formes supplémentaires** | Hapax

| Forme | Freq. | Types |
|---|---|---|
| de | 5190 | pre |
| et | 2719 | con |
| l | 2535 | art_def |
| la | 2394 | art_def |
| les | 2355 | art_def |
| des | 2175 | art_ind |

*Illustration 10: Présentation des résultats de l'analyse textuelle, onglet formes supplémentaire*

#### 2.6.1.2.5 Onglet Hapax

Liste des hapax (colonne forme) avec leurs effectifs (colonne nb) et leurs catégories grammaticales (colonne type).



Résumé | Total | Formes actives | Formes supplémentaires | **Hapax**

| Forme | Freq. | Types |
|---|---|---|
| épisode | 1 | nom |
| illustré | 1 | adj |
| 2d | 1 | nr |
| référencés | 1 | nr |
| ifla | 1 | nr |
| onu | 1 | nr |

*Illustration 11: Présentation des résultats de l'analyse textuelle, onglet Hapax*

### 2.6.1.2.6 Options Supplémentaires de l'analyse statistiques textuelles

#### 2.6.1.2.6.1 Sur chaque forme

Un menu d'options supplémentaires est accessible à partir d'un clic droit sur une des formes, il vous permet d'afficher pour chacune d'entre elles les formes associées, et le concordancier des segments de textes où figure la forme.

#### 2.6.1.2.6.2 Sur l'analyse

Dans l'historique du panneau de gauche, un clic droit sur l'analyse ouvre un menu d'options disponibles :

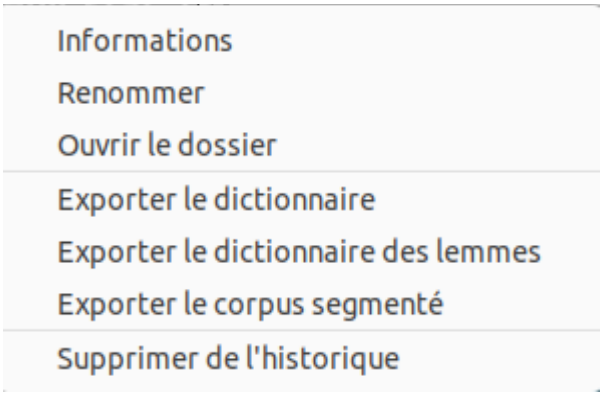



*Illustration 12: Dialogue d'option de l'analyse statistique textuelle*

- **Informations** : permet de récupérer les informations propres à l'analyse.
- **Renommer** : permet de renommer le titre de l'analyse affichée dans le volet historique d'Iramuteq (ne change pas les noms des fichiers attachés).
- **Ouvrir le dossier** : permet d'ouvrir directement le dossier créé par Iramuteq. C'est dans ce dernier que nous pourrons retrouver les fichiers exportés lors des analyses ou exports.
- **Exporter le dictionnaire** : génère un fichier dictionary.csv (Tabulation comme séparateur) reprenant les occurrences. Elles se présentent ainsi :

| Forme | Lemme | Type grammatical | effectif |
|---|---|---|---|
| professeurs | professeur | nom | 25 |

Ce dictionnaire est dans le format utilisé par Iramuteq, il peut donc être directement utilisé pour une analyse.

- **Exporter le dictionnaire des lemmes** : génère un fichier lemme.csv (Tabulation comme séparateur) reprenant le lemme, puis les variantes et leurs effectifs :

| | | | | |
|---|---|---|---|---|
| Lemme<br>Primaire | Forme1<br>Primaire | Effectif forme 1<br>6 | Forme 2<br>Primaires | Effectif forme 2<br>9 |

**Exporter le corpus segmenté :** exporte le corpus complet, mais segmenté

**Supprimer de l'historique** : permet d'effacer cette analyse du volet l'historique d'Iramuteq (n'efface pas les fichiers attachés)

#### 2.6.1.3 Fichiers en sortie

Comme pour chaque analyse, Iramuteq crée un nouveau répertoire dans le dossier créé lors de l'ouverture du corpus. Nommé « NomDuCorpus_Stat_x », il contient les fichiers suivants :




| Répertoire de sortie | NomDuCorpus_Stat_x |
|---|---|
| Fichiers en sortie : | |
| Analyse.ira | Fichier permettant à Iramuteq d'ouvrir l'analyse. |
| Dcitionary.csv | Fichier optionnel, voir 2.6.1.2 |
| Formes_actives.csv | Les formes actives et leur effectif |
| Formes_supplémentaires.csv | Les formes supplémentaires et leur effectif |
| Glob.txt | Informations présentes dans l'onglet résumé |
| Hapax.csv | Liste des hapax |
| Lemmes.csv | Liste des lemmes et de leur(s) déclinaison(s) voir 2.6.1.2 |
| Segmented_corpus.txt | Fichier texte contenant le corpus segmenté |
| Segments_size.png | Histogramme reprenant la fréquence des segments par taille |
| Stsize.csv | Fichier csv le nombre de formes pour chaque segment |
| Total.csv | Toutes les formes et leur effectif |
| Zipf.png | L'image du graphique présent dans « Global » |

### 2.6.2 Spécificités et AFC

Produit une analyse factorielle des correspondances sur un tableau de contingence qui croise les formes actives et les variables.

#### 2.6.2.1 Paramétrage de l'analyse spécificités et AFC

Après le menu des préférences (voir chapitre 2.4), un menu de « choix des variables » s'affiche.

Voici les différentes options proposées :

- **Formes utilisées** : permet d'utiliser les formes actives et/ou supplémentaires pour l'analyse
- **Sélection par** : permet de sélectionner toutes les modalités d'une variable (option « variable »), soit uniquement certaines modalités (option « modalité »).
- **Choix** : vous devez sélectionner une variable
- **Indice** : Deux possibilités de calcul sont possibles, le Chi2 ou la loi hypergéométrique.
- **Fréquence minimale** : seules les formes dont le nombre d'occurrences dans le corpus complet est supérieur à ce seuil seront affichées.




La sélection d'une variable est OBLIGATOIRE, si vous validez sans l'avoir sélectionnée, un message de Bug s'affichera. Cette variable doit avoir au minimum 2 modalités.




Si vous sélectionnez plusieurs variables, une seule sera prise en compte.




Ne jamais sélectionner 2 modalités de variables pouvant être présentes sur la même ligne (deux modalités n'appartenant pas à la même variable), l'analyse n'aurait plus aucun sens.

#### 2.6.2.2 Résultats Spécificités et AFC

Les résultats qui vont suivre sont disponibles pour les formes/lemmes, et les catégories grammaticales.

- L'onglet Statistiques repent le nombre d'occurrences, de formes de segments, d'hapax et le ration hapax/formes pour chaque modalité de la variable étudiée.

| Statistiques | Formes | Formes banales | Types | Fréquences des formes | Fréquences des types |
|---|---|---|---|---|---|
| formes | *espace_DemoCit | *espace_Fiscalite | *espace_Organisation | | |
| occurences | 174987 | 88190 | 157297 | | |
| formes | 12606 | 8399 | 12152 | | |
| segments | 4849 | 2433 | 4400 | | |
| hapax | 3482 | 1490 | 2910 | | |
| hapax/formes | 0.27622 | 0.1774 | 0.23947 | | |

*Illustration 13: présentation des résultats de l'analyse spécificités et AFC onglet Statistiques*

- L'onglet **formes** présente la liste des formes et le score par modalité

| Statistiques | Formes | Formes banales | Types | Fréquences des formes | Fréquences des types |
|---|---|---|---|---|---|
| formes | *espace_DemoCit | *espace_Fiscalite | *espace_Organisation | | |
| vote | 148.2641 | -28.3043 | -55.0401 | | |
| élu | 121.2482 | -24.4519 | -31.5727 | | |
| citoyen | 108.0093 | -20.138 | -28.5179 | | |
| respect | 95.0672 | -21.9465 | -34.0924 | | |
| mandat | 83.1692 | -18.3926 | -25.6 | | |

*Illustration 14: présentation des résultats de l'analyse spécificités et AFC onglet Formes*

- L'onglet **Formes banales** présente la liste des formes non spécifiques à une des modalités et le score par modalité (inférieur à +- 2)

| Statistiques | Formes | Formes banales | Types | Fréquences des formes | Fréquences des types |
|---|---|---|---|---|---|
| formes | eff | *espace_DemoCit | *espace_Fiscalite | *espace_Orga | |
| exemple | 749 | 0.2987 | -0.3416 | -0.4029 | |
| permettre | 744 | 0.5563 | -0.4608 | 0.4316 | |
| niveau | 600 | 0.5284 | 0.3622 | -0.3682 | |
| etc | 399 | 0.4747 | 0.6809 | -0.6156 | |
| petit | 391 | -0.8284 | -1.9103 | 1.789 | |

*Illustration 15: présentation des résultats de l'analyse spécificités et AFC onglet Formesbanales*

- L'onglet **types** présente la liste des catégories grammaticales et le score par modalité

Formes | Formes banales | **Types** | Fréquences des formes | Fréquences des types

| formes | *source_aujourdhui_en_france | *source_la_croix |
|---|---|---|
| pro_per | 6.0066 | 1.1798 |
| aux | 5.3441 | -0.9828 |
| adv_sup | 5.2039 | 0.5387 |
| ver_sup | 2.438 | -0.4565 |
| adj_pos | 2.0583 | -0.5412 |

*Illustration 16: présentation des résultats de l'analyse spécificités et AFC onglet Types*

- Les onglets **Fréquences** : fréquences de chaque forme/lemme (ou catégorie grammaticale) dans les textes présentant la modalité de la variable.

Formes | Formes banales | Types | **Fréquences des formes** | Fréquences des types

| formes | *source_aujourdhui_en_france | *source_la_croix |
|---|---|---|
| de | 182 | 52 |
| être | 114 | 24 |
| la | 96 | 29 |
| avoir | 82 | 27 |
| le | 76 | 29 |

*Illustration 17: présentation des résultats de l'analyse spécificités et AFC onglet Fréquences des Formes*

- L'onglet **Fréquences des types** reprend la fréquence de chaque type grammatical.

Statistiques | Formes | Formes banales | Types | Fréquences des formes | **Fréquences des types**

| formes | *espace_DemoCit | *espace_Fiscalite | *espace_Organisation |
|---|---|---|---|
| nom | 40416 | 20050 | 37230 |
| pre | 24643 | 13108 | 22834 |
| art_def | 20294 | 10724 | 17045 |
| ver | 17019 | 9113 | 13611 |
| adj | 11690 | 5697 | 10329 |

*Illustration 18: présentation des résultats de l'analyse spécificités et AFC onglet Fréquences des types*

L'onglets **Fréquences relatives** des formes donnent les effectifs relatifs en pour mille

(‰) des formes.

| Statistiques | Formes | Formes banales | Types | Fréquences des formes | Fréquences des types | Fr |
|---|---|---|---|---|---|---|

| formes | *espace_DemoCit | *espace_Fiscalite | *espace_Organisation |
|---|---|---|---|
| citoyen | 13 | 3.62 | 4.11 |
| élu | 10.29 | 1.88 | 2.42 |
| public | 7.23 | 9.1 | 12.99 |
| politique | 7.02 | 2.67 | 1.81 |
| vote | 6.68 | 0.35 | 0.32 |

*Illustration 19: présentation des résultats de l'analyse spécificités et AFC onglet Effectifs relatifs des formes*

- L'onglet **AFC** : Présente le graphique de l'AFC (facteurs ½)

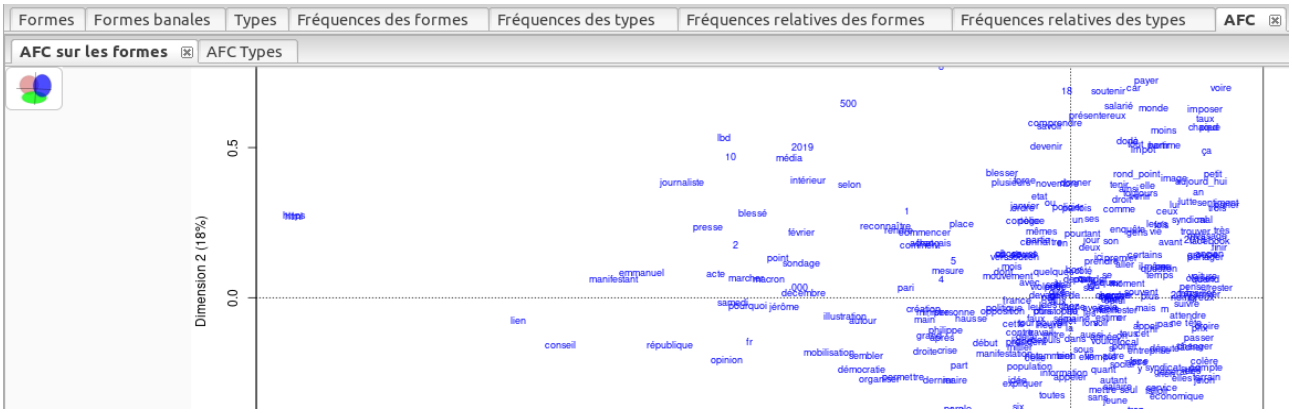



*Illustration 20: Présentation des résultats de l'analyse spécificités et AFC onglet AFC*

L'icône  permet d'afficher les options de réédition du graphique AFC (voir 2.6.2.2.1). Deux sous-onglets existent, pour les formes, et pour les types (voir refaire les AFC)

### 2.6.2.2.1 Menu de réédition des  AFC

Cette boîte de dialogue est accessible par l'icône  à gauche des AFC

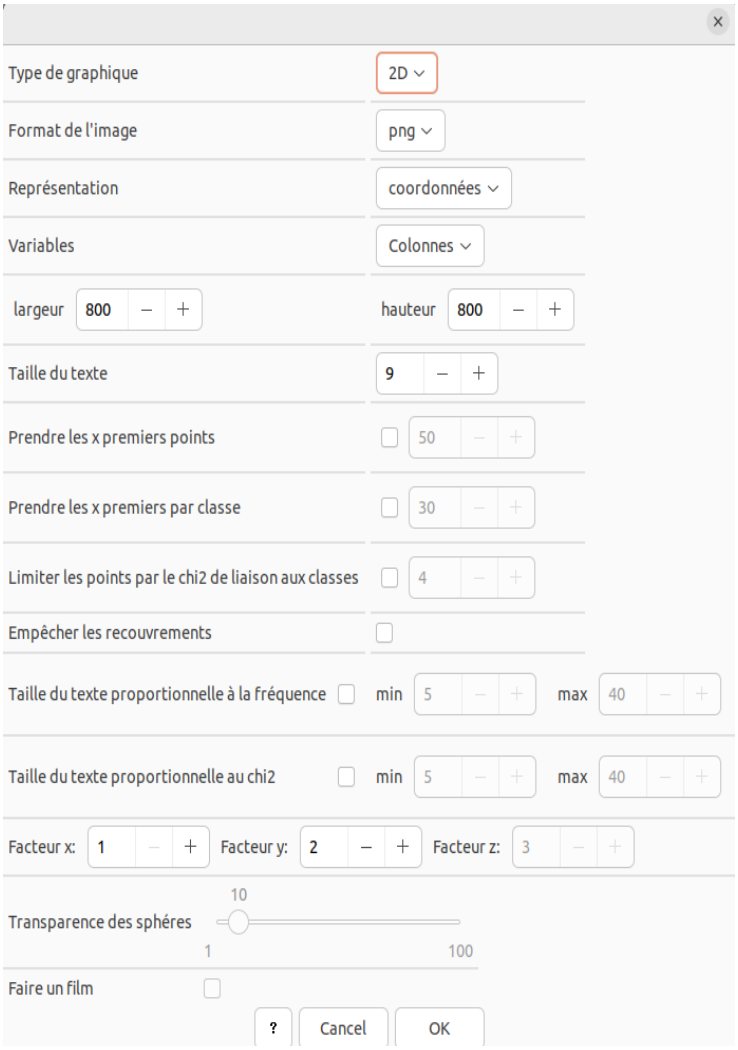



*Illustration 21: Dialoque de réédition des AFC*

- **Type de graphique** : permet de choisir les graphiques en 2D ou 3D
- **Format de l'image** : permet de choisir le format d'enregistrement de l'image
- **Représentation** : choisit le mode de représentation par coordonnées , ou corrélations
- **Variable** : choisit d'afficher les formes (option lignes) ou les modalités de la variable étudiée (option colonnes)
- **Largeur/hauteur** : définit les dimensions du graphique (en px)
- **Taille du texte** : définit la taille du texte
- **Prendre les x premiers points** : Prends les x premiers points qui ont les chi2 les plus forts dans l'ensemble de la classification.
- **Prendre les x premiers points par classe :** Prend les x premiers points qui ont les chi2 les plus forts par classe
- **Limiter les points par le chi2 de liaison par classe** : seuil minimal de

liaison à la classe pour apparaître

- **Empêcher les recouvrements** : cette option évite le recouvrement du label des points, elle produit un biais, mais permet une meilleure lisibilité.
- **Taille du texte proportionnel à la fréquence** : règle la taille du texte en fonction de l'effectif de la forme dans le corpus.
- **Taille du texte proportionnel au chi2** : règle la taille du texte en fonction du chi2 de liaison entre la forme et la modalité
- **facteurX /facteurY/facteurZ** : respectivement facteur en abscisse, facteur en ordonnée, facteur en cote
- **Transparence des sphères** : permet d'ajuster la transparence des sphères dans les graphiques en 3D
- **Faire un film** : disponible uniquement pour les graphiques en 3D, produit un .gif animé qui représente une rotation de 360° du graphique.

#### 2.6.2.2.2 Options Supplémentaires de l'analyse Spécificités et AFC

##### 2.6.2.2.2.1 Sur chaque forme

Un menu d'options supplémentaires est accessible à partir d'un clic droit sur une des formes

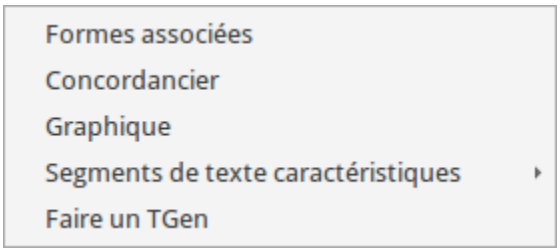



*Illustration 22: Dialogue d'option de l'analyse spécificités et AFC*

**Formes associées** : affiche les déclinaisons du lemme sélectionné

**Concordancier** : affiche le concordancier reprenant les segments de textes contenant la forme

**Graphique** : affiche un histogramme reprenant le score ou la fréquence (selon l'onglet dans lequel est affiché ce menu) pour chaque modalité

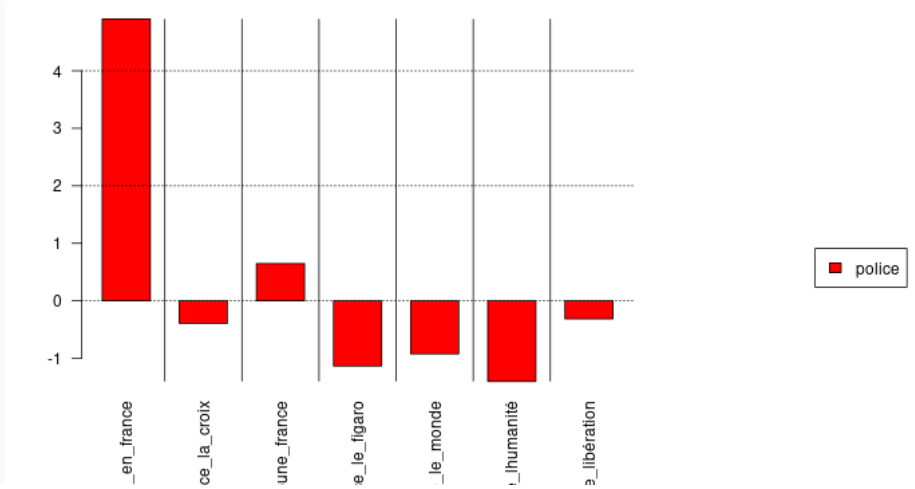



*Illustration 23: graphique des scores de spécificités d'une forme par modalité*

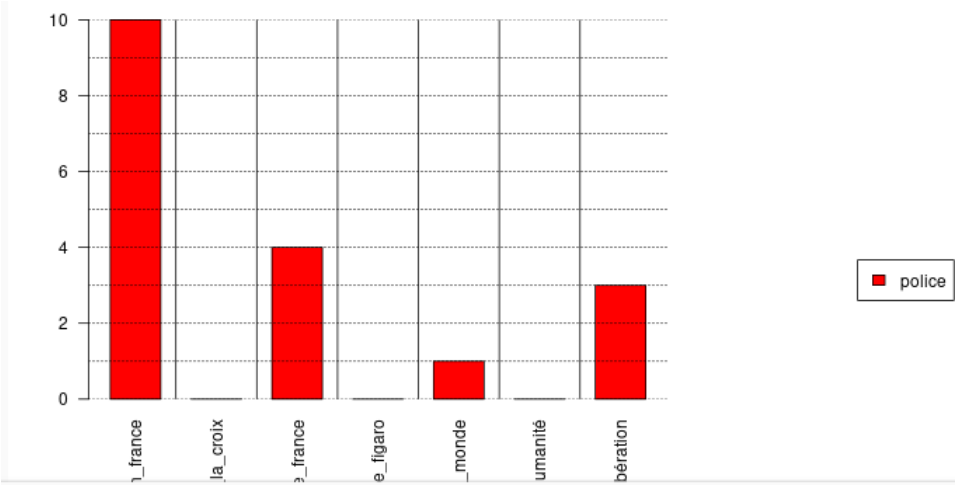



*Illustration 24: graphique des fréquences d'une forme par modalité*

**Segments de texte caractéristiques** : permet d'afficher la liste des segments les plus caractéristiques de chaque modalité.

**Faire un TGEN** : permet d'afficher le menu de création de TGEN

#### 2.6.2.2.2.2 Sur l'analyse

Dans l'historique du panneau de gauche, un clic droit sur l'analyse ouvre un menu d'options disponibles :

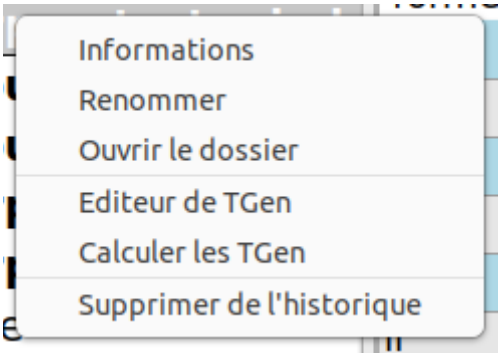



*Illustration 25: Dialogue d'option de l'analyse statistique textuelle (volet historique)*

- **Informations** : permet de récupérer les informations propres à l'analyse.
- **Renommer** : permet de renommer le titre de l'analyse affichée dans le volet historique d'Iramuteq (ne change pas les noms des fichiers attachés).
- **Ouvrir le dossier** : permet d'ouvrir directement le dossier créé par Iramuteq. C'est dans ce dernier que nous pourrons retrouver les fichiers exportés lors des analyses ou exports.
- **Editeur de TGEN** : permet d'afficher le menu de gestion des TGEN
- **Calculer les TGEN** : permet de calculer les scores des TGEN par modalités de variables
- **Supprimer de l'historique** : supprime l'analyse du volet historique (mais conserve les dossiers créés sur le disque dur)

#### 2.6.2.3Fichiers en sortie

| Répertoire de sortie | NomDuCorpus_Spec_x |
|---|---|
| **Fichiers en sortie** | |
| afcf_col.csv | Détail des valeurs des colonnes de l'AFC sur les formes |
| afcf_col.png | Plan factoriel des colonnes de l'AFC sur les formes |
| afcf_facteur.csv | Valeurs Propres, Pourcentages, Pourcentages cumulés de chaque facteur (pour les formes) |
| afcf_row.csv | Détail des valeurs des lignes de l'AFC sur les formes |
| afcf_row.png | Plan factoriel des formes |
| afct_col.csv | Détail des valeurs des colonnes de l'AFC sur les types grammaticaux |
| afct_col.png | Plan factoriel des colonnes de l'AFC sur les types grammaticaux |
| afct_facteur.csv | Valeurs Propres, Pourcentages, Pourcentages cumulés de chaque facteur (pour les types grammaticaux) |
| afct_row.csv | Détail des valeurs des lignes de l'AFC sur les types grammaticaux |
| afct_row.png | Plan factoriel des types grammaticaux |
| analyse.ira | Permet d'ouvrir l'analyse |
| banalites.csv | Fréquences et scores des formes banales |
| eff_relatif_forme.csv | Effectif relatif des formes par modalité |
| eff_relatif_type.csv | Effectif relatif des types grammaticaux par modalité |
| liste_graph_afcf.txt | Fichier Iramuteq |
| liste_graph_afct.txt | Fichier Iramuteq |
| RData.Rdata | Fichier R |
| tableafcm.csv | Effectif des formes par modalité |
| tablespecf.csv | Score de spécificité des formes par modalité |
| tabletypem.csv | Effectif des types grammaticaux par modalité |

| | |
|---|---|
| Tgen.csv* | Formes contenues dans les Tgens. La structure est une ligne par Tgen, la première cellule est le nom du Tgen, les colonnes suivantes sont les formes contenues dans le Tgen. |
| Tgenspec.csv* | Scores spécificité des Tgens par modalité de la variable. |

### 2.6.3 TGEN

Les TGEN sont proposés dans les analyses de spécificités et AFC ainsi que dans la classification Reinert. La seule façon d'accéder à cet outil est donc de passer par une de ces analyses, qui, une fois effectuées, proposent dans le menu d'options supplémentaires (clic droit dans le volet historique) l'éditeur de TGEN.

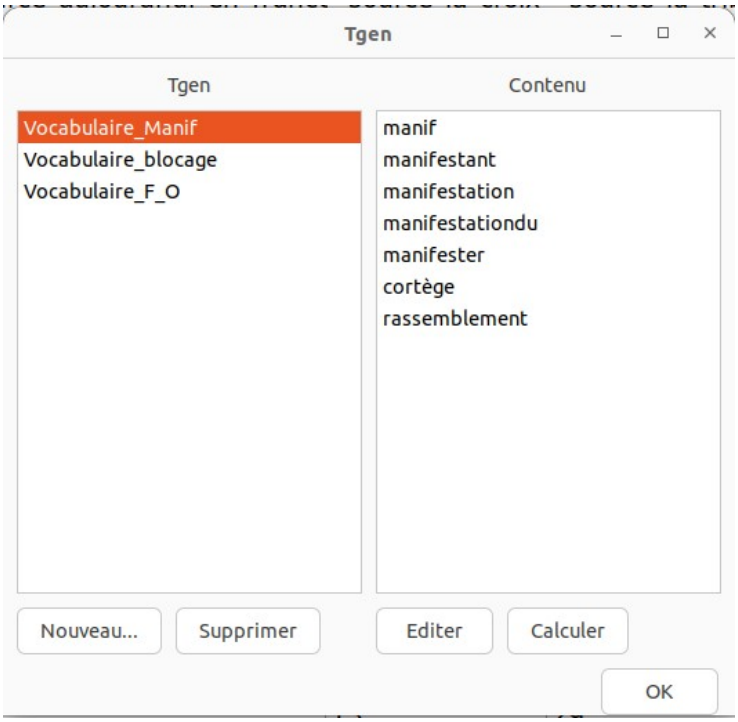



*Illustration 26: menu principal des TGEN*

Dans le menu TGEN, nous retrouvons, sur la colonne de gauche, la liste des TGEN déjà enregistrés ; sur la colonne de droite, les formes contenues dans le TGEN sélectionné, et des boutons signalant les différentes actions possibles.

**Nouveau** : permet de créer un nouveau Tgen, affiche pour cela l'éditeur de Tgen (voir 2.6.3.1)

**Supprimer** : supprime le Tgen sélectionné

**Editer** : affiche l'éditeur de Tgen prérempli avec le Tgen sélectionné

**Calculer** : calculer les scores des Tgen enregistrés

#### 2.6.3.1 Editeur de Tgen

Après avoir demandé la création d'un nouveau Tgen ou l'édition d'un Tgen existant, le menu Editeur de Tgen est ouvert.

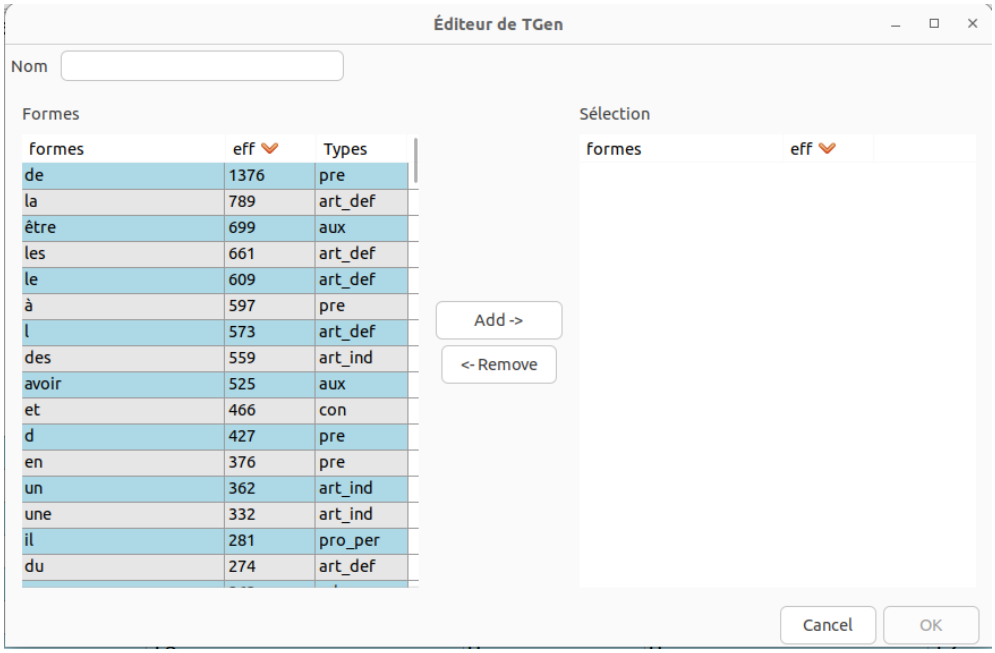


*Illustration 27: menu Editeur de Tgen*

**Nom** : c'est le nom que vous devez renseigner poru retrouver le Tgen

**Colonne formes** : nous retrouvons la liste des formes présentes dans le corpus, leur fréquence et leur catégories grammaticales

**Colonne Sélection** : c'est la liste des formes qui constituent le Tgen

**Add** : permet d'ajouter au Tgen une ou des formes sélectionnée(s) dans la colonne Formes

**Remove** : enlève une ou des formes de la liste Sélection

**Cancel** : annule le Tgen

**Ok** : valide et enregistre le Tgen

#### 2.6.3.2Résultats

Une fois le calcul des TGEN effectué, un onglet supplémentaire s'affiche dans l'analyse, reprenant les scores des TGEN dans chaque modalité pour l'analyse de spécificités et pour chaque classe dans la classification Reinert.

Types | Fréquences des formes | Fréquences des types | Fréquences relatives des formes | Fréquences relatives des types | AFC | **Spécificités des TGens**

| formes | *source_la_croix | *source_le_figaro | *source_le_monde | *source_lhumanit | *source_libération |
|---|---|---|---|---|---|
| Vocabulaire_F_O | -0.4551 | -2.808 | 1.1836 | -2.3996 | -0.7841 |
| Vocabulaire_blocage | 0.6871 | 0.3812 | 0.3794 | 0.5349 | -0.5471 |
| Vocabulaire_Manif | -0.5811 | -1.5412 | -0.6008 | 0.7391 | 0.3804 |

*Illustration 28: onglet Tgen dans l'analyse de spécificités*

CHD | Profils | AFC | **Spécificités des TGens**

| formes | cluster_001 | cluster_002 | cluster_003 |
|---|---|---|---|
| Vocabulaire_blocage | 13.29546 | -3.12746 | -5.01656 |
| Vocabulaire_Manif | 7.44904 | 46.7399 | -59.0306 |
| Vocabulaire_F_O | -1.02546 | 68.93759 | -26.29624 |

*Illustration 29: onglet Tgen dans la classification Reinert*

### 2.6.3.3Options Supplémentaires des Tgens

À partir d'un clic droit sur les résultats de Tgen, un menu s'affiche avec plusieurs options supplémentaires :

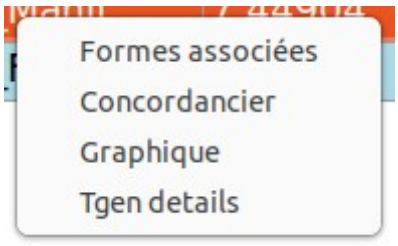



*Illustration 30: menu d'option des Tgens*

**Formes associées** : affiche les formes composant le Tgen et leur déclinaisons

**Concordancier** : affiche les segments contenant au moins une des formes constituant le Tgen

**Graphique** : affiche l'histogramme des scores du Tgen pas modalité pour les spécificités et par classe pour les CHD Reinert

**Tgen Details** : affiche un tableau avec toutes les formes constituant le Tgen et leurs scores par modalité pour les spécificités et par classe pour la CHD Reinert

| formes | cluster_001 | cluster_002 | cluster_003 |
|---|---|---|---|
| manifester | 13.53704 | -0.40137 | -9.65463 |
| manifestation | 6.56658 | 1.47562 | -11.44486 |
| cortège | 0.29479 | 7.16738 | -6.23977 |
| manifestant | 0.26113 | 42.41926 | -27.98812 |
| manifestationdu | -0.60302 | 5.23795 | -0.86564 |
| manif | -0.91688 | 25.77064 | -7.8773 |
| rassemblement | -1.63566 | 0.82254 | 0.33016 |

*Illustration 31: résutlats du détail d'un Tgen*

#### 2.6.3.4 Fichiers en sortie

La création d'un TGEN dans une analyse (CHD Reinert ou spécificités) implique la création d'un fichier tgen.csv dans le dossier de l'analyse. Il contient une ligne par TGEN enregistré, avec en première colonne le nom du TGEN et dans les colonnes suivantes les formes composant le TGEN.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Vocabulaire_Manif | manif | manifestant | manifestation | manifestationdu | manifester | cortège | rassemblement |
| Vocabulaire_blocage | bloc | blocage | bloquer | filtre | | | |
| Vocabulaire_F_O | police | policier | policieres | gendarme | crs | | |

*Illustration 32: contenu du fichier tgen.csv*

- Le fichier Tgen.csv peut être copié et collé dans une autre analyse, il suffira d'ouvrir l'éditeur de Tgen pour retrouver ceux déjà enregistrés.

### 2.6.4 Distances de Labbé

Cette analyse propose un calcul des distances lexicales entre les textes de chaque modalité d'une variable comme décrit dans Labbé et Labbé (2003) accessible ici

#### 2.6.4.1Paramètres

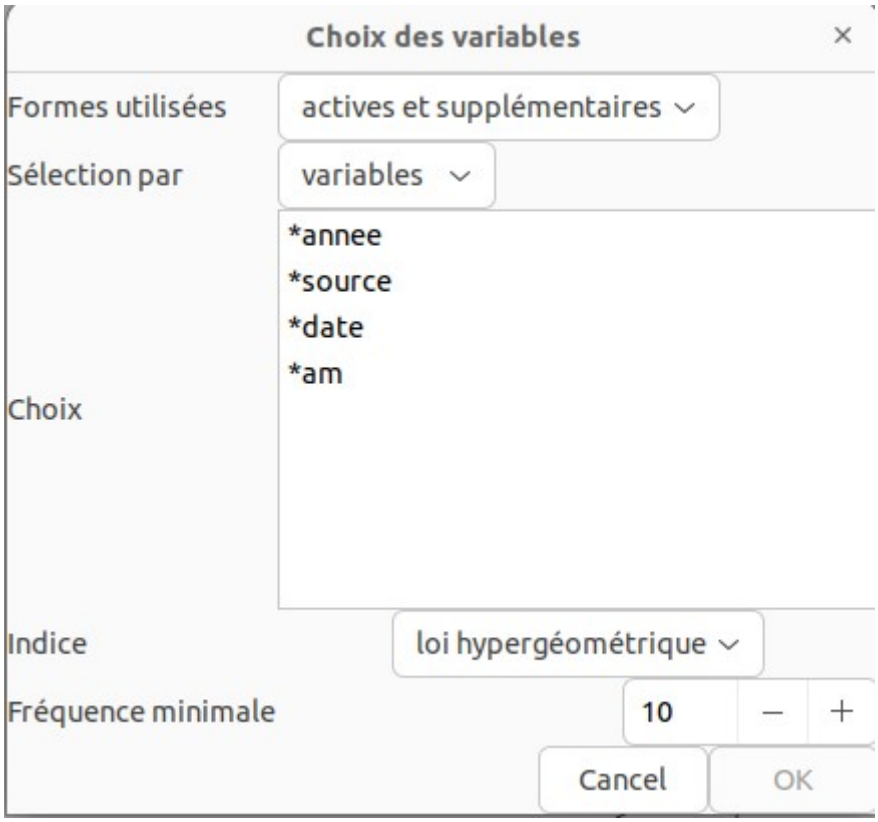



*Illustration 33: paramètres de l'analyse de distances de Labbé*

Voici les différentes options proposées :

- **Formes utilisées** : vous permet d'utiliser les formes actives et/ou supplémentaires pour l'analyse
- **Sélection par** : Vous permet de sélectionner toutes les modalités d'une variable (option « variable »), soit uniquement certaines modalités (option « modalité »).
- **Choix** : vous devez sélectionner une variable
- **Indice** : cette fonction n'a pas d'incidence dans le calcul des distances
- **Fréquence minimale** : seules les formes dont le nombre d'occurrences dans le corpus complet est supérieur à ce seuil seront affichées.



La sélection d'une variable est OBLIGATOIRE, si vous validez sans l'avoir sélectionnée, un message de Bug s'affichera. Cette variable doit avoir au minimum 2 modalités.



Si vous sélectionnez plusieurs variables, une seule sera prise en compte.



Ne jamais sélectionner 2 modalités de variables pouvant être présentes sur la même ligne (deux modalités n'appartenant pas à la même variable), l'analyse n'aurait plus aucun sens.

### 2.6.4.2Résultats de l'analyse des distances de Labbé

Les résultats directement accessibles pour cette analyse sont composés d'une classification de ward (onglet ward clustering) d'une matrice de chaleur (onglet heatmap) et de la matrice des distances.

#### 2.6.4.2.1Représentations de ward des distances de Labbé

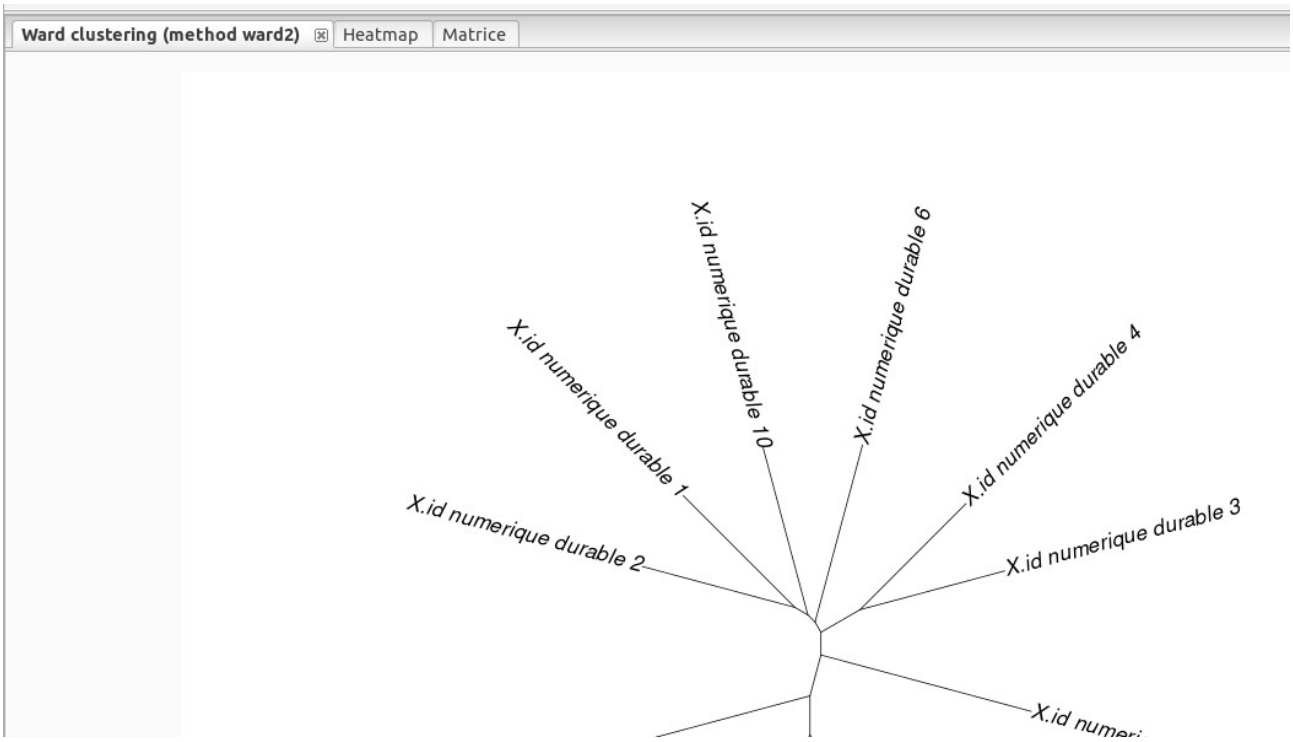



*Illustration 34: onglet Word clustering (method ward2) des distances de Labbé*

Ce graphique est une représentation arborescente d'une classification ascendante (méthode de Ward) basée sur une matrice de distances. Les modalités des variables sont regroupées en fonction de leur proximité lexicale. Plus deux modalités sont situées sur des branches proches, plus leur lexique est similaire. À l'inverse, plus les branches sont éloignées, plus les lexiques diffèrent.

#### 2.6.4.2.2 Onglet heatmap des distances de Labbé

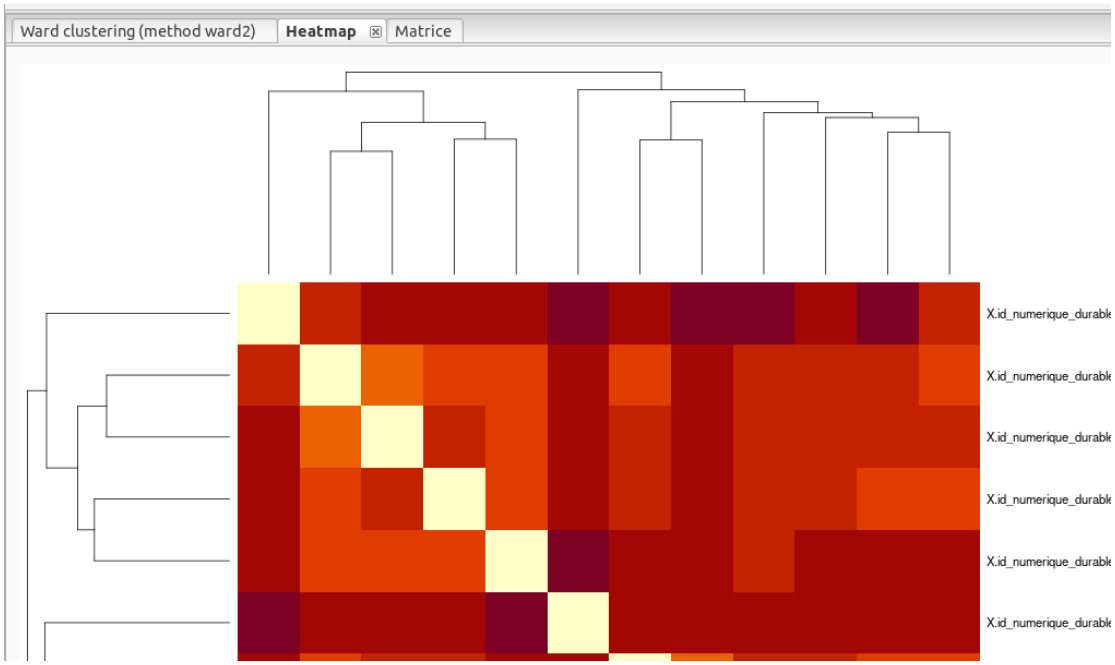



*Illustration 35: onglet heatmap de l'analyse de distance de Labbé*

Cet onglet présente une carte thermique issue de la matrice de distances. En abscisse et en ordonnée, nous retrouvons un dendrogramme résultant d'une classification des distances. Plus une case est claire, plus la distance lexicale entre les textes des deux modalités de la variable est faible (lexique similaire). À l'inverse, plus une case est foncée, plus la distance lexicale entre les textes des deux modalités est élevée (lexique différent).

#### 2.6.4.2.3 Onglet matrice des distances de Labbé




*Illustration 36: onglet matrice des distances de Labbé*

Ce dernier onglet présente la matrice de distances (indice de Labbé). Les scores sont compris entre 0 et 1. Plus les textes des deux modalités d'une variable utilisent un même lexique, plus l'indice est faible. À l'inverse, plus le lexique des textes des deux modalités diffère, plus le score est élevé.

### 2.6.4.3 Option supplémentaire des distance de Labbé

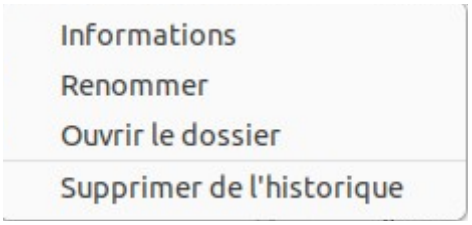



*Illustration 37: menu d'option de l'analyse des distances de Labbé*

Sur cette analyse, seul un menu d'option dans le volet historique est disponible. Nous y retrouvons les fonctions suivantes :

- **Informations** : permet de récupérer les informations propres à l'analyse.
- **Renommer** : permet de renommer le titre de l'analyse affichée dans le volet historique d'Iramuteq (ne change pas les noms des fichiers attachés).
- **Ouvrir le dossier** : permet d'ouvrir directement le dossier créé par Iramuteq. C'est dans ce dernier que nous pourrons retrouver les fichiers exportés lors des analyses ou exports.
- **Supprimer de l'historique** : supprime l'analyse du volet historique (mais

conserve les dossiers créés sur le disque dur)

#### 2.6.4.4Fichier en sortie

| Répertoire de sortie | NomDuCorpus_labbe_x |
|---|---|
| **Fichiers en sortie** | |
| Analyse.ira | Fichier analyse Iramuteq |
| Distmat.csv | Matrice des distances textuelles |
| graph_min.graphml | Fichier graphml de l'analyse de similitudes des distances (graphe minimum) |
| graph_min.png | Représentation par analyse de similitudes des distances (graphe minimum) |
| graph_tot.graphml | Fichier graphml de l'analyse de similitudes des distances (graphe complet) |
| graph_tot.png | Représentation par analyse de similitudes des distances (graphe complet) |
| Labbe-heatmap.png | Carte thermique des distances |
| Labbe-matric.png | Image de la matrice des distances |
| Labbe-tree.png | Représentation arborée de la classification de la matrice de distances |
| liste_graph_afct.txt | Fichier analyse Iramuteq |
| Tableafcm.csv | Fréquence des formes par modalités de variable |

### 2.6.5 Classification Méthode Reinert

Cette analyse propose une classification hiérarchique descendante selon la méthode décrite par Reinert (1983, 1986, 1991). Cette classification est proposée selon trois modalités :




- Classification simple sur texte : Ici, les **Textes** resteront dans leur intégralité, la classification permettra ainsi de regrouper les **Textes**

les plus proches.

- Classification simple sur segments de texte : La classification portera sur les segments de textes (ST).
- Classification double sur RST : La classification est menée sur deux tableaux dans lesquels les lignes ne sont plus des segments de texte mais des regroupements de segments de texte (RST). Le même traitement est ainsi fait deux fois, mais en changeant le nombre de formes actives par RST.

#### 2.6.5.1 Paramétrages de l'analyse Classification Méthode Reinert

Après le menu des préférences (voir chapitre 2.5), un menu « Options » s'affiche.

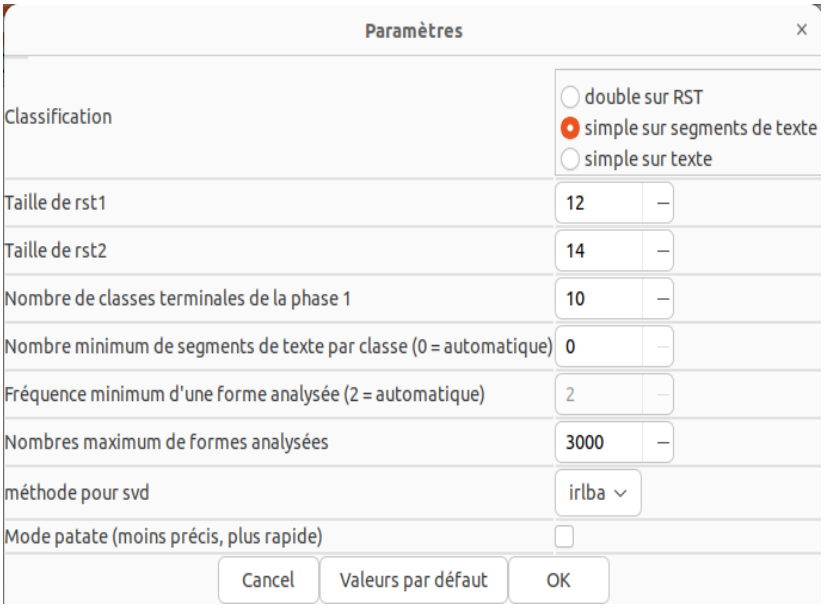



*Illustration 38: Dialogue de paramétrages de l'analyse classification méthode Reinert*

- **Classification** : permet de choisir le mode de classification (voir chapitre 2.5.3)
- **Taille rst1 et rst2** : pour la classification double sur RST permet de faire varier la taille des regroupements de segments de texte sur les deux passages, ces unités représentent le nombre de formes actives regroupées par ligne.
- **Nombre de classes terminales de la phase 1** : détermine le nombre de classes de la première partie de la classification, par défaut elle est réglée sur 10, mais doit être ajustée à la taille et à l'hétérogénéité du corpus.
- **Nombre minimum de segments de texte par classe** : détermine un seuil minimal de ST en dessous duquel les classes ne seront pas sélectionnées. Par défaut ce nombre est égal au nombre de segments de texte divisé par le nombre de classes terminales pour la classification simple, et au nombre de segments de texte divisés par 2 fois le nombre de classes terminales pour la

classification double RST.

- **Fréquence minimum d'une forme analysée** : cette option n'est pas disponible.
- **Nombre maximum de forme analysées** : Par défaut, les 3000 formes actives les plus fréquentes et les 3000 formes supplémentaires les plus fréquentes seront retenues. Une forme doit avoir au minimum une fréquence de 3 pour être retenue. Si le corpus a moins de 3000 formes, toutes les formes avec une fréquence supérieure ou égale à 3 seront retenues.




Ce paramètre a une forte incidence sur la taille des tableaux analysés et donc sur la quantité de mémoire de l'ordinateur mobilisée. Si votre ordinateur n'a pas assez de mémoire pour analyser un corpus, essayez de baisser ce paramètre. Si votre ordinateur possède « suffisamment » de mémoire pour le corpus et que le corpus possède plus de 3000 formes de fréquence ≥ 3, n'hésitez pas à l'augmenter.

- **Méthode pour SVD** : cette option permet de choisir l'algorithme utilisé dans les décompositions en valeur singulières.
- **Mode Patate** : cette option supprime la deuxième phase de chaque partition. Elle permet une analyse un peu plus rapide mais un peu moins précise.

### 2.6.5.2 Résultats de la Classification Méthode Reinert

Les résultats directement disponibles présentent un résumé de la classification (onglet CHD), les profils des classes (onglet Profils), et une analyse factorielle des correspondances menée sur le tableau de contingence croisant formes/lemmes et classes (onglet AFC).




*Illustration 39: Présentation des résultats de la Classification Méthode Reinert*

#### 2.6.5.2.1 Onglet CHD

Sur le premier onglet, vous pouvez distinguer 3 blocs distincts :

le premier, constitué de texte, reprend les statistiques textuelles du corpus. Vous pourrez y trouver :

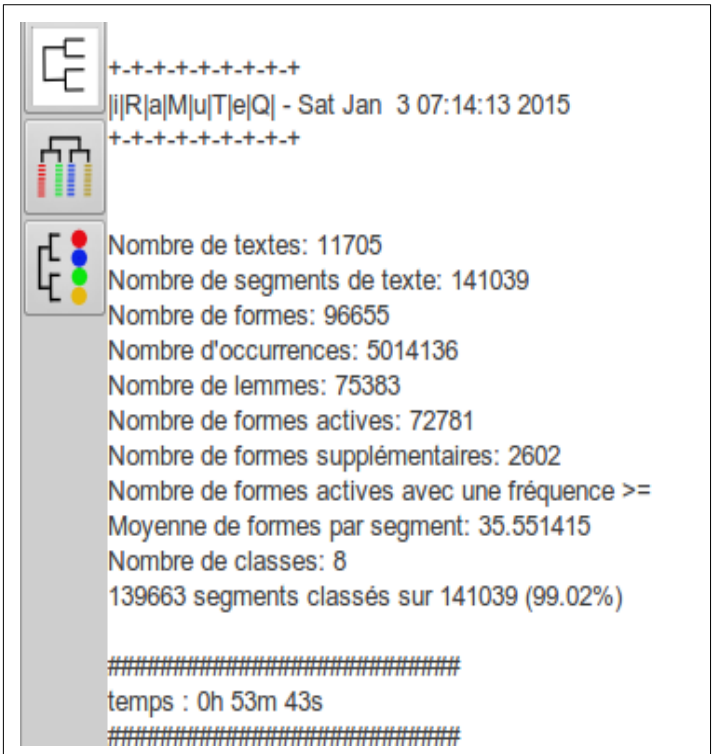



*Illustration 40: Présentation des résultats de la Classification Méthode Reinert : statistiques textuelles*

**Nombre de textes** : le nombre de textes présents dans le corpus

**Nombre de segments de textes** : le nombre de segments de textes présents dans l'ensemble du corpus

**Nombre de formes** : nombre de formes (actives et illustratives)

**Nombre d'occurrences** : nombre de mots dans le corpus

**Nombre de lemmes** : nombre de lemmes dans le corpus

**Nombre de formes actives** : nombre de formes actives (nom, verbes d'action, adjectifs et adverbes)

**Nombre de formes supplémentaires** : nombre d'articles, verbes auxiliaires, verbes supplémentaires, nombres...

**moyenne de formes par segment**

**le nombre de segments classés sur la totalité du corpus**. : affiche le nombre de segments classés et sa proportion par rapport à l'ensemble des segments contenus dans le corpus. LA différence provient des classes trop petites pour apparaître (voir paramètres sur le nombre minimum de segments par classe).

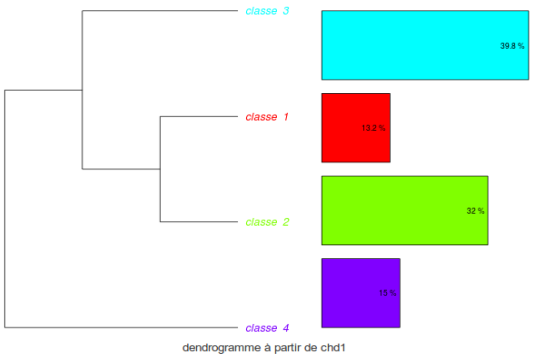



*Illustration 41: Présentation des résultats de la Classification Méthode Reinert : Dendrogramme*

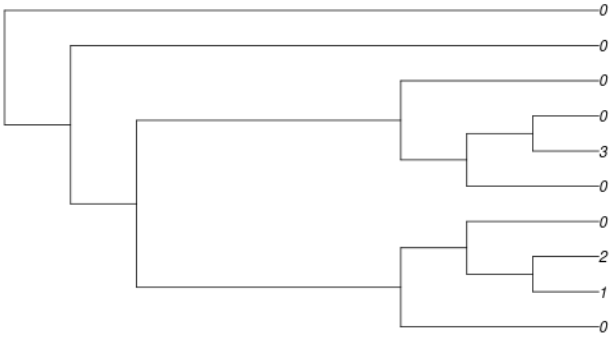



*Illustration 42: présentation des résultats de la Classification méthode Reinert : dendrogramme phase 1*

La seconde partie est constituée du dendrogramme, il représente la partition, et une indication de la taille des classes (exprimée en pourcentage du corpus classé).

Pour finir le troisième élément est le dendrogramme de la phase 1, les classes notées 0 représentent les classes non retenues

À gauche de ces éléments, apparaissent plusieurs boutons permettant d'éditer le dendrogramme sous différentes formes :

 permet de refaire le dendrogramme par défaut (voir illustration 23)

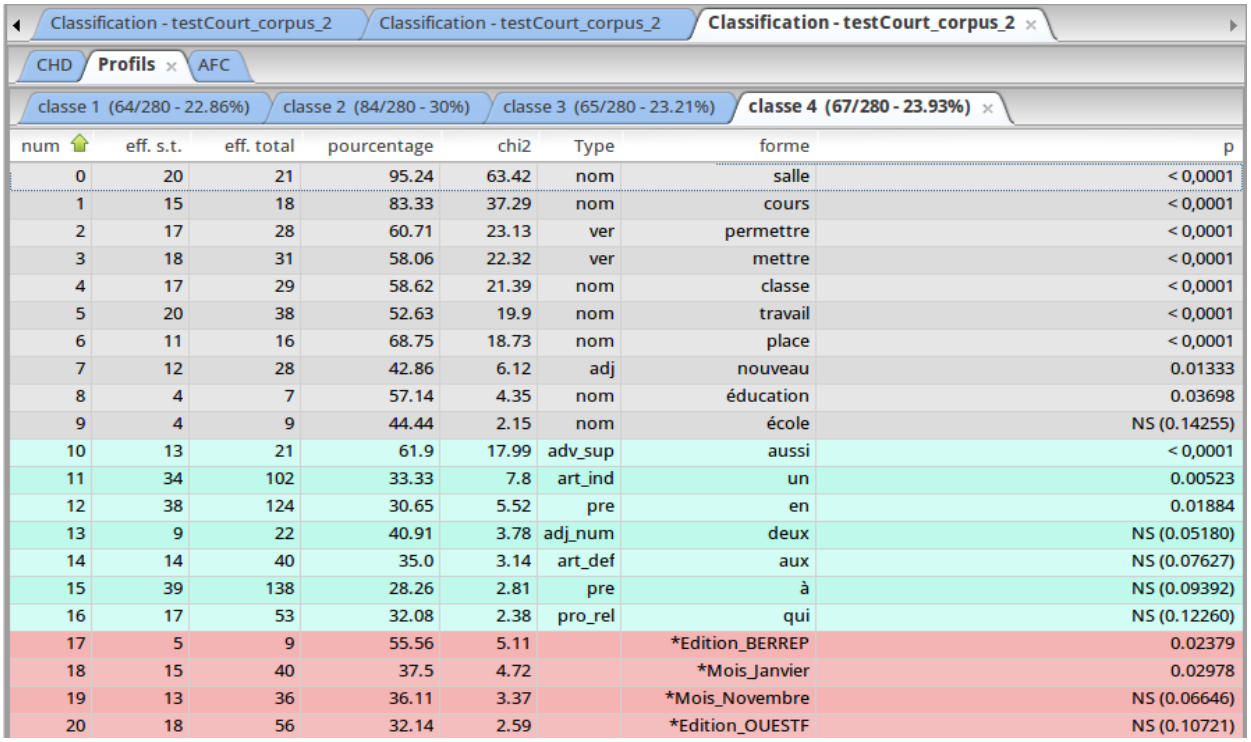
 permet d'éditer un dendrogramme avec les profils de chaque classe

 permet d'éditer un dendrogramme avec les nuages de mots de chaque classe.

### 2.6.5.2.2 Onglet Profils

Pour chaque classe sont édités les profils, ces tables présentent les formes/lemmes

caractéristiques des classes. Dans ces onglets, vous trouverez :



Classification - testCourt_corpus_2 | Classification - testCourt_corpus_2 | Classification - testCourt_corpus_2

CHD | Profils | AFC

classe 1 (64/280 - 22.86%) | classe 2 (84/280 - 30%) | classe 3 (65/280 - 23.21%) | classe 4 (67/280 - 23.93%)

| num | eff. s.t. | eff. total | pourcentage | chi2 | Type | forme | p |
|---|---|---|---|---|---|---|---|
| 0 | 20 | 21 | 95.24 | 63.42 | nom | salle | < 0,0001 |
| 1 | 15 | 18 | 83.33 | 37.29 | nom | cours | < 0,0001 |
| 2 | 17 | 28 | 60.71 | 23.13 | ver | permettre | < 0,0001 |
| 3 | 18 | 31 | 58.06 | 22.32 | ver | mettre | < 0,0001 |
| 4 | 17 | 29 | 58.62 | 21.39 | nom | classe | < 0,0001 |
| 5 | 20 | 38 | 52.63 | 19.9 | nom | travail | < 0,0001 |
| 6 | 11 | 16 | 68.75 | 18.73 | nom | place | < 0,0001 |
| 7 | 12 | 28 | 42.86 | 6.12 | adj | nouveau | 0.01333 |
| 8 | 4 | 7 | 57.14 | 4.35 | nom | éducation | 0.03698 |
| 9 | 4 | 9 | 44.44 | 2.15 | nom | école | NS (0.14255) |
| 10 | 13 | 21 | 61.9 | 17.99 | adv_sup | aussi | < 0,0001 |
| 11 | 34 | 102 | 33.33 | 7.8 | art_ind | un | 0.00523 |
| 12 | 38 | 124 | 30.65 | 5.52 | pre | en | 0.01884 |
| 13 | 9 | 22 | 40.91 | 3.78 | adj_num | deux | NS (0.05180) |
| 14 | 14 | 40 | 35.0 | 3.14 | art_def | aux | NS (0.07627) |
| 15 | 39 | 138 | 28.26 | 2.81 | pre | à | NS (0.09392) |
| 16 | 17 | 53 | 32.08 | 2.38 | pro_rel | qui | NS (0.12260) |
| 17 | 5 | 9 | 55.56 | 5.11 | | *Edition_BERREP | 0.02379 |
| 18 | 15 | 40 | 37.5 | 4.72 | | *Mois_Janvier | 0.02978 |
| 19 | 13 | 36 | 36.11 | 3.37 | | *Mois_Novembre | NS (0.06646) |
| 20 | 18 | 56 | 32.14 | 2.59 | | *Edition_OUESTF | NS (0.10721) |

*Illustration 43: Présentation des profils de la classification méthode Reinert*

- **Le titre** : il est composé du nom de la classe puis du nombre de ST de la classe, puis du nombre ST participant à la classification finale, puis du pourcentage des ST de la classe par rapport au nombre de segments de texte participant à la classification finale.




```
Sur l'onglet activé, on lit : Classe 4 (67/280 – 23,93%)  =>  sur les
280 segments de textes classés, 67 appartiennent à cette classe, ce
qui représente  23,93 % des segments de texte classés.
```

- **Le contenu** : il est présenté sous forme de tableaux, utilisant un code de 3 couleurs les formes actives apparaissent en gris, les formes supplémentaires en bleu et les variables illustratives en rose. À l'intérieur de chacune de ces catégories, les formes sont présentées dans l'ordre décroissant des Chi2 de liaison à la classe.

| num | eff. s.t. | eff. total | pourcentage | chi2 | Type | forme | p |
|---|---|---|---|---|---|---|---|
| 0 | 73 | 104 | 70.19 | 77.44 | adj | numérique | < 0,0001 |

*Illustration 44: Présentation des résultats de la Classification Méthode Reinert : Ligne d'un profil*

- **num** : le numéro de la ligne (permet à l'utilisateur de retrouver le classement initial).
- **Eff. s.t** : Effectif des segments de texte, c'est le nombre de ST de cette classe contenant au moins une fois la forme
- **eff. Total** : Effectif total des segments de texte, c'est le nombre de ST dans tout le corpus classé contenant au moins une fois la forme
- **pourcentage** : c'est pourcentage de eff. ST / eff Total
- **Chi2** : chi2 exprimant la force du lien entre la forme et la classe (voir annexe)
- **Type** : catégorie grammaticale de la forme
- **Forme** : forme ou lemme concerné
- **p** : Seuil de significativité associé au Chi2




- Pour rechercher une forme dans le profil ouvert, vous pouvez appuyer sur les touches CTRL+F
- Pour rechercher une forme dans tous les profils, vous pouvez appuyer sur les touches CTRL+Shift+F

#### 2.6.5.2.3 Onglet AFC de la méthode Reinert

Dans cet onglet vous trouverez l'AFC déduite de la classification,

- L'onglet **AFC** : présente le graphique de l'AFC, l'icône  permet d'afficher les options de graphique, et de le réafficher (possible pour les formes active ou illustratives, pour les variables étoilées et les classes).
- L'onglet **Facteurs** : présente les valeurs propres, pourcentages et pourcentages cumulés de chaque facteur.
- L'onglet **graph** 3D : produit un graphe des distances minimum sur les formes en 3 dimensions.

#### 2.6.5.2.4 Options supplémentaires de la Classification Méthode Reinert

##### 2.6.5.2.4.1 Les options disponibles pour chaque classe

Pour aider votre étude des profils, Iramuteq vous propose diverses options, vous y avez accès dans chaque classe en effectuant un clic droit sur la ligne vous intéressant. Le menu suivant apparaît :

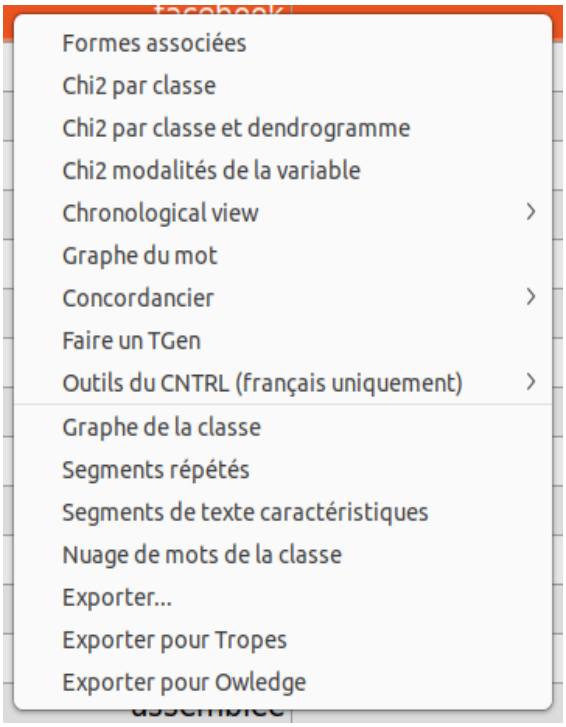



*Illustration 45: Dialogue d'options disponibles pour chaque classe de la classification Méthode Reinert*

- **Les formes associées** : donne les effectifs dans la classe des formes regroupées dans un lemme
- **Chi2 par classe** : crée un graphique qui présente le Chi2 d'association de la forme à chacune des classes. Plusieurs formes peuvent être sélectionnées en même temps (maintenir la touche Ctrl enfoncée)

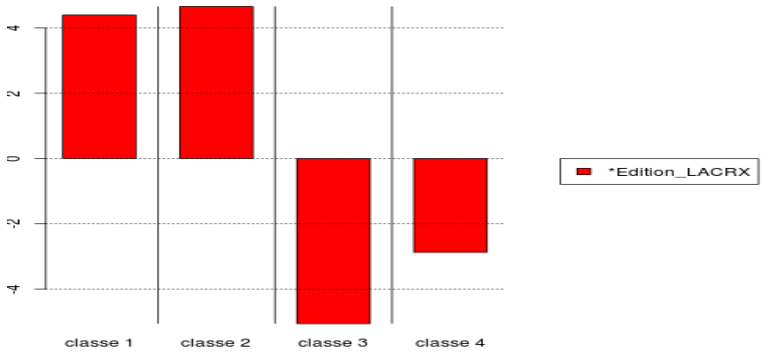



*Illustration 46: Chi2 de la forme par classe*

- **Chi2 par classe + dendro** : reprend le même graphique en barres, en le projetant sur le dendrogramme.

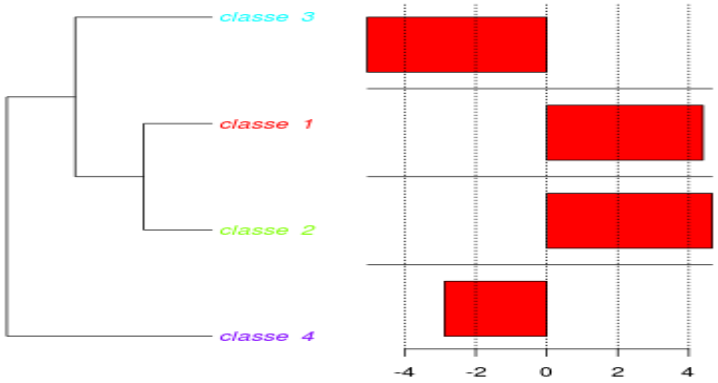



*Illustration 47: Chi2 de la forme + dendro*

- **Chi2 Modalités de la variable** : crée un graphique qui représente le chi2 d'association des modalités de la variable sélectionnée à chacune des classes. Nécessite un formatage du type variable_modalité.

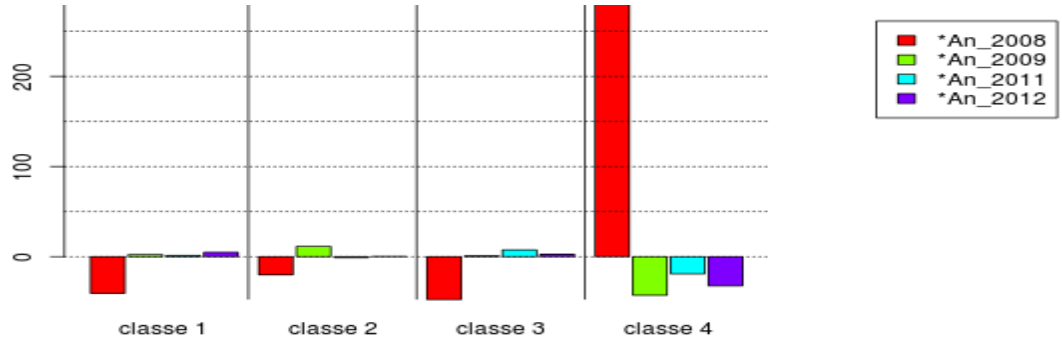



*Illustration 48: Chi2 des modalités par classe*

- **Chronogical view :** *nécessite un formatage du type variable_modalité. Cette option* permet de générer des graphiques reprenant l'évolution des contenus des modalités d'une variable selon les classes. Ces représentations sont proposées en deux versions :
  - Chi2 : Cette illustration présente un tableau croisant les classes en lignes et les modalités des variables en colonnes. Lorsqu'une modalité est sur-représentée dans une classe, la cellule correspondante du tableau est colorée. Plus la couleur est foncée, plus la corrélation est forte. La largeur des lignes représente la taille de la classe, et la largeur des colonnes représente le nombre de segments contenant la modalité.

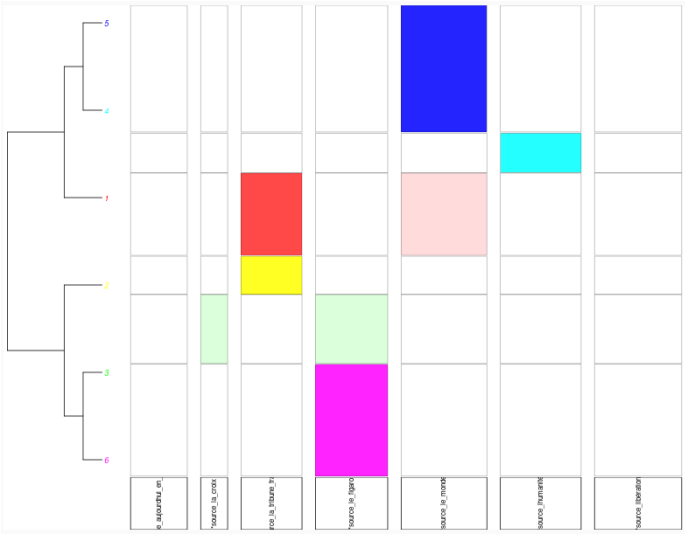


*Illustration 49: graphique chronogical view par chi2 pour la CHD Reinert*

  - Proportion : cette visualisation propose pour chaque modalité d'une variable un histogramme empilé à 100 % du nombre de segments de textes (ST) présents dans chaque classe. La proportion de ST d'une classe est représentée par la couleur et la position de la classe dans le dendrogramme. La largeur des colonnes est proportionnelle au nombre de segments de textes présents dans chaque modalité.

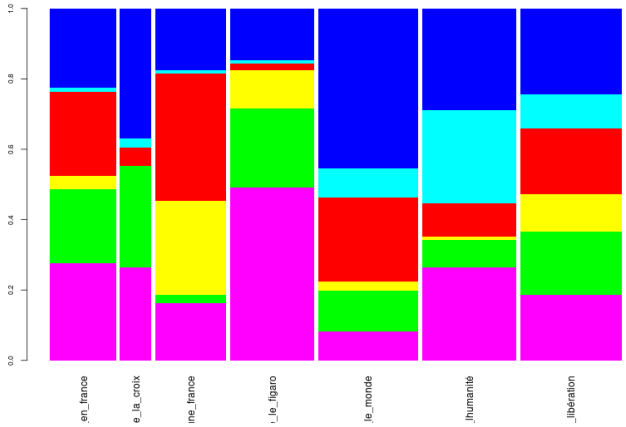
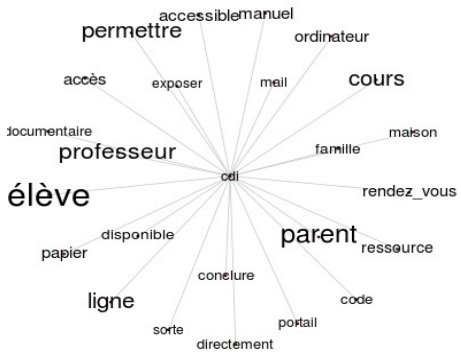



*Illustration 50: graphique chronogical view par proportion pour la CHD Reinert*

- **Graphe du mot** : crée un graphe de similitude représentant les liens de la forme sélectionnée avec les autres formes de la classe (voir #Ref  pour plus de détails).




- **Concordancier** : propose le concordancier de la (ou des) forme(s) / lemmes sélectionnée(s). Ce concordancier est disponible pour les segments de texte  de la classe, ceux classés ou tous ceux du corpus. Dans ce concordancier, sur chaque ligne étoilée, est propsé un hyperlien reprenant le numéro du texte et le numéro du ST. Il permet d’afficher le texte complet avec le ST en surbrillance.

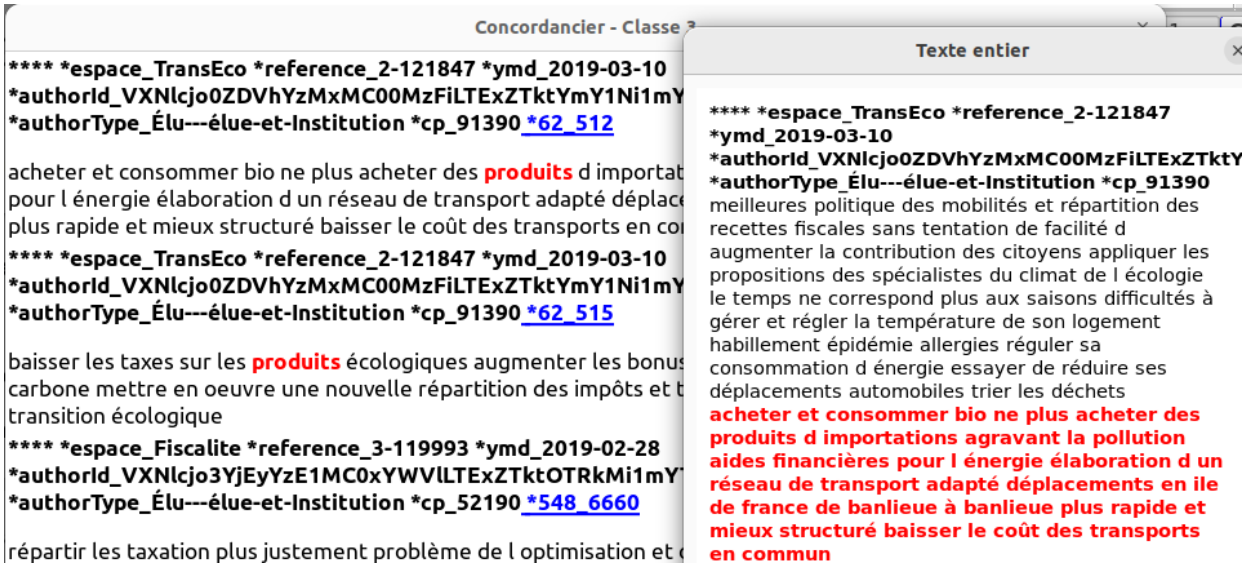


*Illustration 51: concordancier et texte complet*

- **Faire un TGEN** : permet d'afficher le menu de création de TGEN (voir chapitre 2.6.3).
- **Outils du CNRTL** : interroge la base de données du Centre National de Ressources Textuelles et Lexicales (http://www.cnrtl.fr/) à partir de la forme / du lemme sélectionné(e) (nécessite d'être connecté à Internet, ne fonctionne que pour le français). Permet d'obtenir une définition (Lexicographie), les synonymes (Synonymie), les Antonymes (Antonymie), l'étymologie (Etymologie), la morphologie (Morphologie) ou un graphe de proxémie de la forme. Les résultats s'affichent dans le navigateur internet par défaut du système. Cette option n'a de sens que sur un corpus en français.
- **Graphe de classe** : indépendant de la ligne sélectionnée. Il s'agit d'une analyse de similitude menée sur un tableau absence/présence (0/1) qui croise les unités choisies en ligne (TEXTE ou ST) et les formes actives de la classe en colonne. La matrice de similitude est construite sur les colonnes (les formes actives de la classe). Voir chapitre 2.5.4 pour plus de détails.
- **Segments répétés** : indépendant de la ligne sélectionnée. Effectifs et tailles des segments répétés de la classe. Préférez les profils des segments répétés.
- **Segments de texte caractéristiques** : indépendant de la ligne sélectionnée. Liste les ST caractéristiques de la classe. Deux scores sont proposés :
  - absolu : les ST sont classés en fonction de la somme des chi2 de liaison à la classe des formes actives qu'ils contiennent.
  - Relatif : les ST sont classés en fonction de la moyenne des chi2 de liaison à la classe des formes actives qu'ils contiennent.
- **Nuage de mots de la classe** : permet d'effectuer un nuage de mot des formes contenues dans le profil de la classe. (voir chapitre 2.5.5)
- **Exporter** : exporte tous les ST présents dans la classe, ils peuvent ainsi

constituer un nouveau corpus.

- **Exporter pour Tropes®** : exporte les ST présents dans la classe en respectant le format d'import de Tropes®
- **Exporter pour Owledge®** : exporte les ST présents dans la classe en respectant le format d'import d'owledge®

#### 2.6.5.2.4.2 Les options disponibles pour la classification

En plus de ces outils, vous trouverez d'autres options en effectuant un clic droit sur l'analyse, dans le menu « navigateur » à gauche d'Iramuteq.

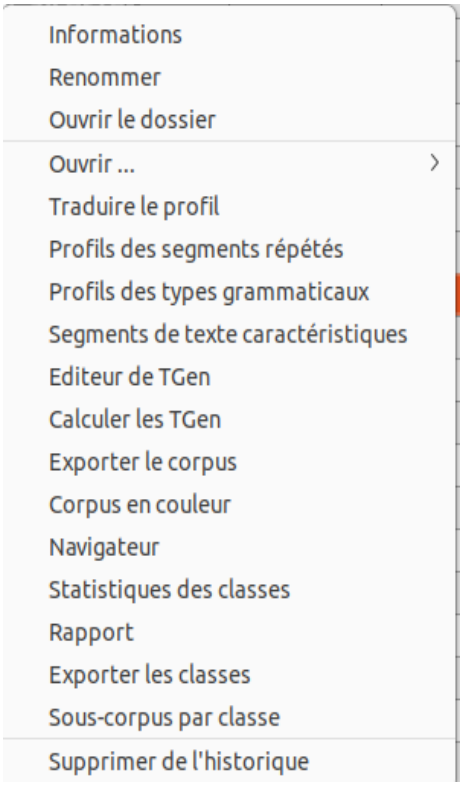
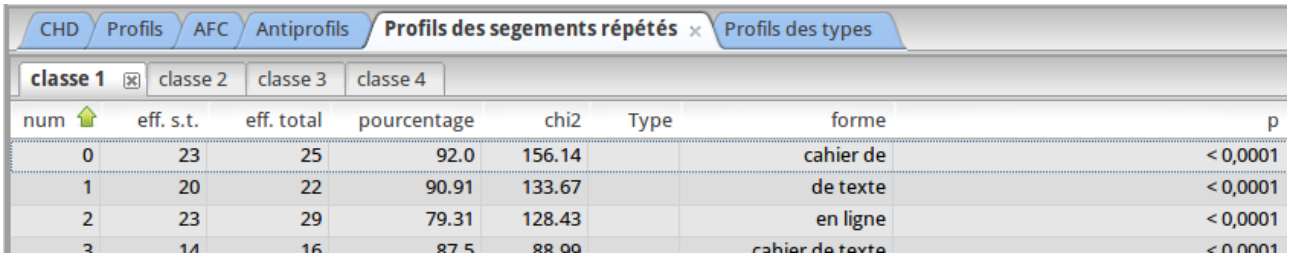



*Illustration 52: Dialogue d'options disponibles pour l'analyse classification méthode Reinert*

- **Informations** : reprend les caractéristiques choisies pour la classification (voir chapitre 2.5.3.1)
- **Renommer** : permet de renommer l'analyse dans le volet historique d'Iramuteq (ne change pas le nom des fichiers attachés).
- **Ouvrir le dossier** : permet d'ouvrir le dossier contenant tous les fichiers attachés à l'analyse
- **Ouvrir** -> Anti-profils, sur la même présentation que les profils seront affichés les anti-profils, ce sont les formes significativement absentes de la classe.
- **Translate profil** : permet de générer un nouvel onglet contenant les profils de

chaque classe avec une traduction des formes.

- **Profils des segments répétés** : Calcule les profils des segments répétés.




| num | eff. s.t. | eff. total | pourcentage | chi2 | Type | forme | p |
|---|---|---|---|---|---|---|---|
| 0 | 23 | 25 | 92.0 | 156.14 | | cahier de | < 0,0001 |
| 1 | 20 | 22 | 90.91 | 133.67 | | de texte | < 0,0001 |
| 2 | 23 | 29 | 79.31 | 128.43 | | en ligne | < 0,0001 |
| 3 | 14 | 16 | 87.5 | 88.99 | | cahier de texte | < 0,0001 |

*Illustration 53: Option profils de segments répétés*

- **Profil des types grammaticaux** : calcule les profils des types grammaticaux.
- **Segments de texte caractéristiques** : génère des fichiers HTML contenant les segments caractéristiques de chaque classe.
  - absolu : les ST sont classés en fonction de la somme des chi2 de liaison à la classe des formes actives qu'ils contiennent.
  - Relatif : les ST sont classés en fonction de la moyenne des chi2 de liaison à la classe des formes actives qu'ils contiennent.
- **Éditeur de TGEN** : permet d'afficher le menu de création de TGEN (voir chapitre 2.6.3).
- **Calculer les TGEN** : permet de calculer les TGEN.
- **Exporter le corpus** : cette fonction permet d'exporter le corpus découpé en segments de texte en ajoutant en variable étoilée la classe du ST. Les segments de texte écartés de la classification apparaîtront avec la variable *classe_0.
- **Corpus en couleur** : Cette fonction crée un fichier .html retranscrivant le corpus entier, où chaque segment de texte apparaît de la couleur de sa classe. Les ST non classés sont écrits en noir. Le fichier créé est accessible dans le répertoire de l'analyse.

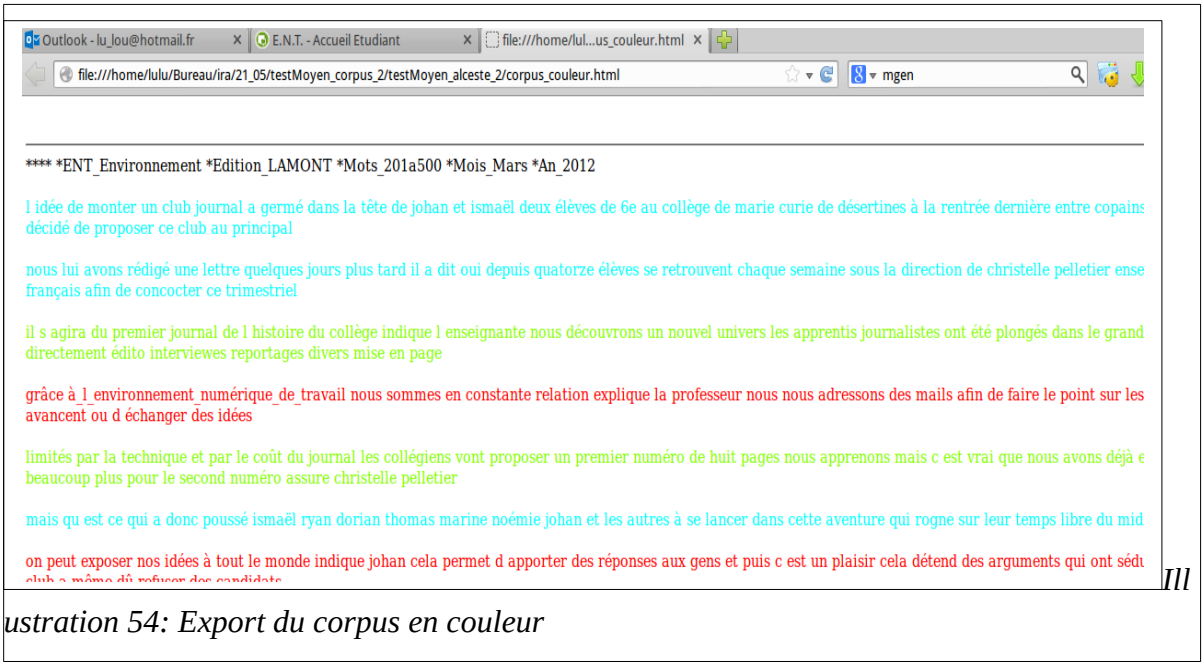



*Illustration 54: Export du corpus en couleur*

- **Navigateur** : Matrice reprenant toutes les formes avec leur Chi2 de liaison à chaque classe.

Navigation

| id | formes | classe 1 | classe 2 | classe 3 | classe 4 |
|---|---|---|---|---|---|
| 0 | élève | 11.603 | -10.347 | 30.86 | -43.97 |
| 1 | collège | -2.566 | -0.181 | 29.794 | -29.212 |
| 2 | général | -16.56 | -0.368 | -2.296 | 45.073 |
| 3 | conseil | -9.73 | -0.311 | -3.292 | 37.958 |

*Illustration 55: Export de l'outil navigation*

- **Statistiques des classes** : cette fonction crée un fichier .csv dans le répertoire de l'analyse (stat_par_classe.csv) qui contient pour chacune des classes : le nombre total d'occurrences, le nombre de formes différentes, le nombre d'hapax, le nombre de segments de texte, et le ratio nbre d'hapax/ nbre de formes.
- **Rapport** : cette fonction crée un fichier RAPPORT.TXT, qui contient les statistiques générales de la classification et les profils.
- **Exporter les classes** : permet de créer un fichier .txt par classe, comprenant tous les ST présents dans chacune d'entre elles. Les fichiers seront nommés classe_x_export.txt et seront disponibles dans le dossier de l'analyse.
- **Sous corpus par classe** : permet d'ouvrir dans Iramuteq une partition du corpus correspondant aux segments contenus dans les classes sélectionnées. Cette partition sera ouverte automatiquement et sera nommée dans le volet historique subcorpus_NomCorpus_X. Attention, cette fonction ne crée pas de

fichier correspondant à la partition.

- **Supprimer de l'historique** : efface cette analyse de l'historique d'Iramuteq.

#### 2.6.5.3Fichiers en sortie pour la classification

| Répertoire de sortie | NomDuCorpus_Alceste_x |
|---|---|
| **Fichiers en sortie :** | |
| AFC2DCL.png | graphique AFC : Classes - Coordonnées - facteur 1 / 2 |
| AFC2DEL.png | graphique AFC : Variables illustratives - Coordonnées - facteur 1 / 2 |
| AFC2DL.png | graphique AFC : Variables actives - coordonnées - facteurs 1 / 2 |
| AFC2DL.png_notplotted.csv | Matrice des colonnes non représentées dans le graphique de l'AFC. Les colonnes sont identifiant, forme, coordonnées axe 1, coordonnées axe 2, taille |
| AFC2DSL.png | graphique AFC : variables supplémentaires - coordonnées - facteurs 1 / 2 |
| AFC2DSL.png_notplotted.csv | Matrice des formes non repésentées dans le graphique AFC2DSL.png |
| afc_col.csv | Résultats de l'AFC ; Coordonnées, corrélation, MASS, contribution des classes : voir le manuel de la librairie ca de R pour plus de détails. |
| afc_facteur.csv | Résultats de l'AFC ; Valeurs propres, Pourcentage d'inertie extraite et Pourcentage cumulé des facteurs. |
| afc_row.csv | Résultats de l'AFC ; Coordonnées, corrélation, MASS, contribution des formes : voir le manuel de la librairie ca de R pour plus de détails. |
| Analyse.ira | Fichier Analyse : permet d'ouvrir une analyse. |
| Antiprofiles.csv | Antiprofils des formes pleines |
| antiprof_segments.csv * | Antiprofils des segments répétés |
| antiprof_type.csv * | Antiprofils des types grammaticaux |
| Arbre_1.png | Dendrogramme de la première CHD |
| Chistable.csv | Chi2 d'association de chaque forme aux classes |

| | |
|---|---|
| classe_mod.csv | Tableau de contingence formes actives/classes |
| corpus_couleur.html * | Corpus respectant les couleurs des classes d'appartenance. |
| Dendro1.png | Dendrogramme de la classification |
| Dendrogramme.RData | Fichier analyse |
| dendrogramme_x.png* | Réédition du dendrogramme avec une nouvelle mise en page |
| graph_afc_x.png* | Réédition du graphe d'AFC |
| Export_corpus.txt * | Fichier créé par l'option « exporter le corpus » |
| Info.txt | Résumé de la classification |
| liste_graph_afc.txt | Fichier Iramuteq |
| liste_graph_chd.txt | Fichier Iramuteq |
| ListeUCE1.csv | Fichier Iramuteq |
| n1.csv | Fichier Iramuteq |
| n1-1.csv | Fichier Iramuteq |
| prof_segments.csv * | Profils des segments répétés |
| profil_type.csv * | Profils des types grammaticaux |
| Profiles.csv | Profils des classes |
| Ptable.csv | Seuil de significativité des chi2 d'association de chaque forme aux classes. |
| RAPPORT.txt * | Fichier créé par l'option « Rapport » |
| SbyClasseOut.csv | Fichier Iramuteq |
| segment_classe.csv * | Tableau de contingence des segments répétés/classes |
| stat_par_classe.csv * | Fichier créé par l'option « Stat par classe » |
| Tableet.csv | Tableau de contingence des variables illustratives/classes |
| Tablesup.csv | Tableau de contingence des formes supplémentaires/classes |
| TableUC1.csv | Matrice utilisée pour l'analyse au format MatrixMarket |
| Tgen.csv | Formes contenues dans les Tgens. La structure est une ligne par Tgen, la première cellule est le nom du Tgen, les |

| | |
|---|---|
| | colonnes suivantes sont les formes contenus dans le tgen. |
| Tgenchi2.csv | Chi2 de liaison entre les Tgens et les classes. |
| Tgeneff.csv | Effectifs de chaque tgen dans chaque classe. |
| Tgenlemchi2.csv | Chi2 de chaque lemme contenu dans les tgens |
| Tgenlemeff.Csv | Effectifs dans chaque classe pour chaque forme contenue dans les tgens. |
| Tgenlempchi2.csv | Seuil de significativité des chi2 d'association de chaque forme contenue dans les tgens aux classes |
| Tgenpchi2.csv | Seuil de significativité des chi2 d'association de chaque Tgen aux classes |
| translation_profile_co.csv | Profils des classes avec les formes traduites |
| Translation.txt | Liste des fichiers générés pour les traductions. |
| translation_word_co.csv | Formes des profils traduites. En première colonne la version traduite, en seconde colonne la version originale. |
| type_cl.csv * | Tableau de contingence des types grammaticaux/classes |
| Uce.csv | Positionnement des segments de textes par classes |

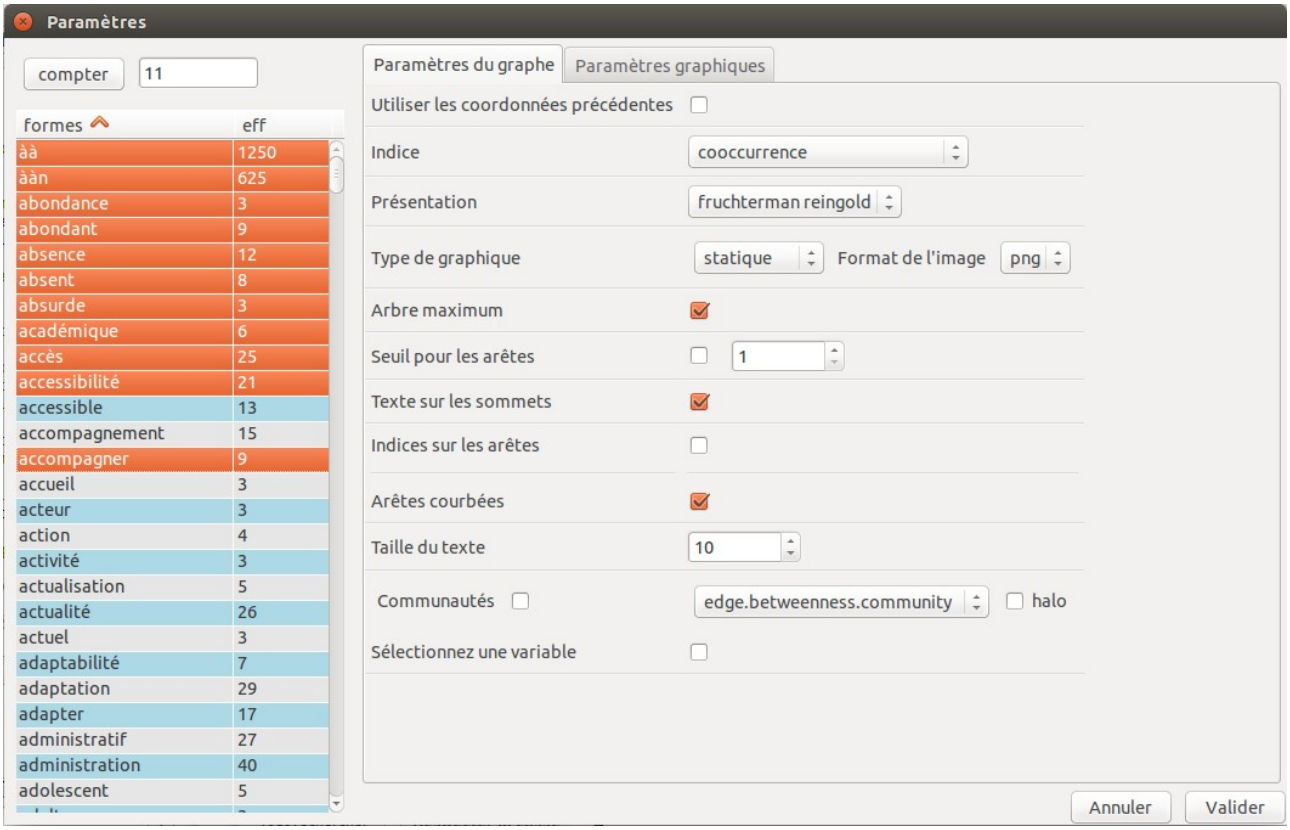
* ces fichiers ne sont accessibles qu'après avoir commandé ces options dans le menu « option d'analyse » (clic droit sur l’analyse dans le menu à gauche).

### 2.6.6 Analyses de similitude

Cette analyse produit des graphes à partir de la librairie « igraph » de R. Le tableau en entrée est un tableau de présence / absence. La matrice de similitude est calculée à partir de l'un des indices proposés. La plupart des indices proposés sont issus de la librairie « proxy » de R. Voir le détail des indices de similitude en annexe.

#### 2.6.6.1 Paramétrages de l'Analyse de Similitude sur textes

Après le menu des préférences (voir chapitre 2,4), un menu de paramétrage s'affiche.




Ce menu se compose de trois parties :

- compter : affiche le nombre de formes sélectionnées
- La l**iste des formes** : dans le cadre de gauche, sont présentées les formes et leurs fréquences, par défaut elles sont toutes sélectionnées.
- Les **paramètres du graphe** : c'est l'onglet actif lors de l'ouverture de ce menu. Il permet de choisir les caractéristiques du graphe. Dans l'ordre, nous trouvons :
  - Indice : permet de choisir le type d'indice de similitude utilisé (cooccurrence par défaut)
  - Layout : mode de présentation des données. Les options sont aléatoires (random), cercle, Fruchterman Reingold, Kamada-Kawaï, graphopt. Voir la documentation de la librairie igraph pour plus de détails.

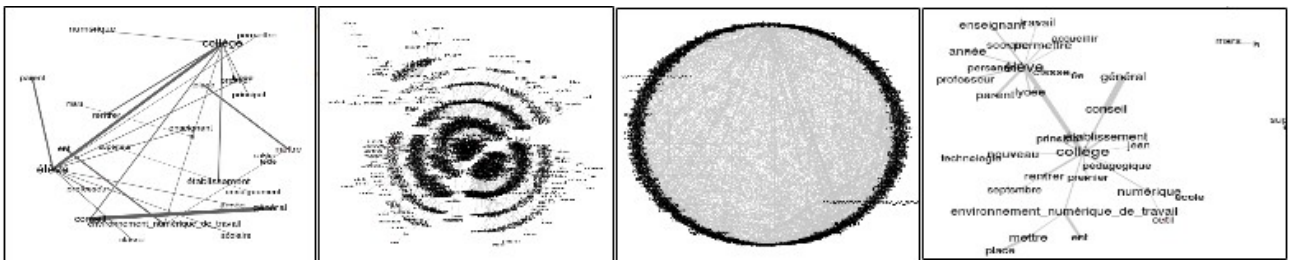



*Illustration 57: Exemples des différents modes de représentations des données (Layout)*

- **Type de graphique** : choix de la méthode de présentation du graphe.
  - Statique : produit une image PNG ou SVG
  - Dynamique : utilise l'interface tk d'igraph. Cette interface permet d'organiser les graphes et de conserver les coordonnées pour réutilisation dans un graphe statique.
  - 3D : produit un graphe en trois dimensions qui s'ouvre dans une fenêtre rgl.

- **Format de l'image** : permet d'enregistrer l'image en .PNG ou .SVG
- **Arbre maximum** : calcul de l'arbre maximum
- **Graphe à seuil** : permet d'éliminer les arêtes pour lesquelles la valeur de l'indice est inférieure ou égale au seuil choisi.
- **Texte sur les sommets** : affiche du texte sur les sommets des graphes
- **Indice sur les arêtes** : affiche le texte sur les arêtes du graphe
- Arêtes courbées : affiche les arêtes de façons courbées
- **Taille du texte** : Indique la taille du texte sur le graphe
- **communauté** : Permet un calcul de communauté (voir la documentation de la librairie igraph pour plus de détails).
- **halo** : représente les communautés par un halo de couleur.
- **sélectionner une variable** : ouvre après validation un nouveau menu permettant de choisir une variable (ou plusieurs modalités). Les formes caractéristiques des modalités seront mises en couleur. **Option sélection de variable** : pour les détails de cette fenêtre

- **Les paramètres graphiques** : c'est le second onglet, il permet de modifier les caractéristiques visuelles du graphe.

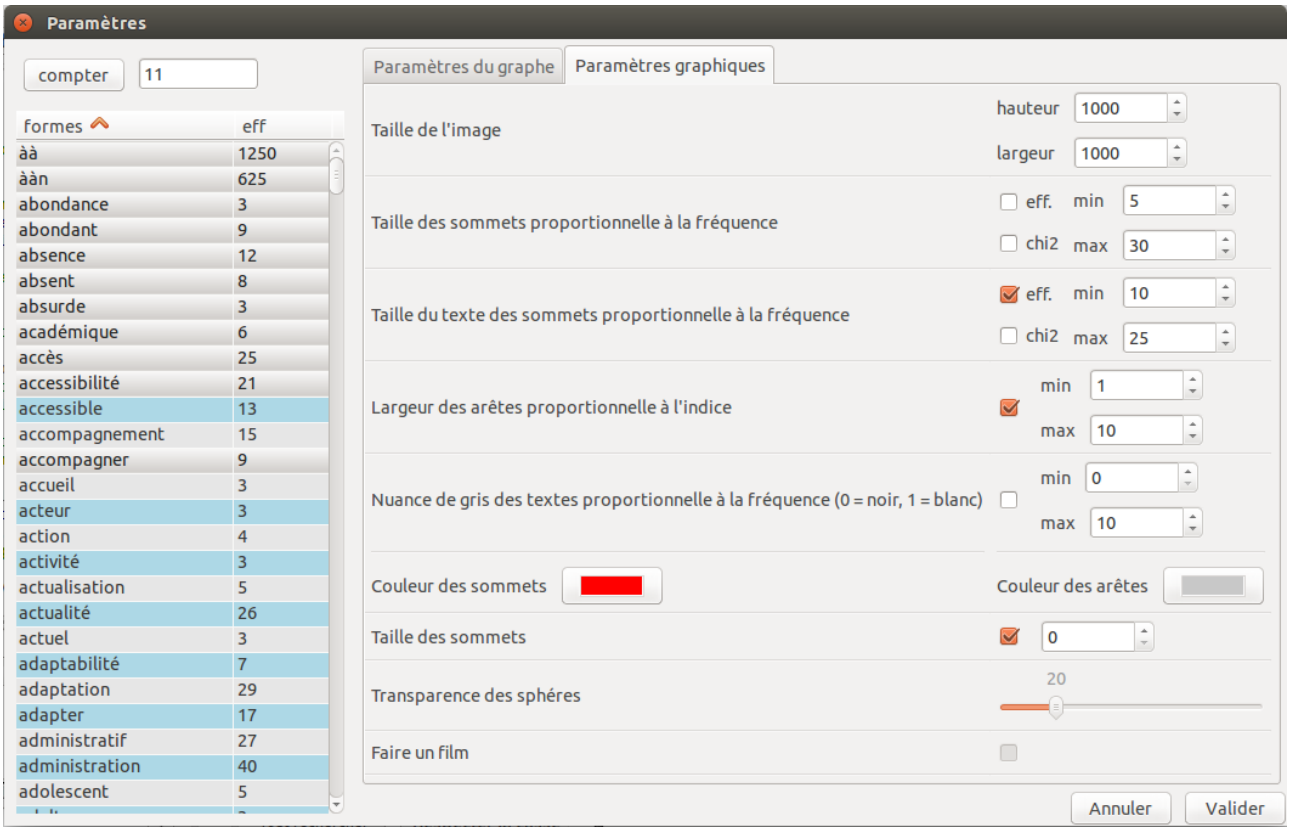



*Illustration 58: Dialogue de paramétrage de graphiques de l'analyse de similitude*

- Taille de l'image : taille de l'image produite (en pixels)
- Taille des sommets proportionnelle à la fréquence : permet d'assigner une taille des sommets proportionnelle à l'effectif de la forme (eff) ou à son chi2 (chi2).
- Taille du texte des sommets proportionnelle à la fréquence : permet d'assigner une taille du texte des sommets proportionnelle à l'effectif (eff) de la forme ou à son chi2.
- Largeur des arêtes proportionnelle à l'indice : permet d'assigner une largeur des arêtes proportionnelle à la valeur de l'indice.
- Nuance de gris des textes proportionnelle à la fréquence (0=noir, 1=blanc) : permet d'ajuster la couleur du texte des sommets en fonction de l'effectif de la forme.
- Couleur des sommets / Couleurs des arêtes : affecte les couleurs aux sommets et/ou arêtes
- Taille des sommets : assigne une taille unique aux sommets du graphe
- Transparence des sphères : règle la transparence des sphères pour les graphes en 3D

- Faire un film : permet de faire un film représentant un tour complet d'un graphe en 3D

### 2.6.6.2 Résultats de l'analyse de similitude

Les résultats s'ouvrent dans un nouvel onglet, sur la gauche sont présentes deux icônes :

 Cette icône permet d'ouvrir à nouveau le menu paramètre, en conservant les derniers choix.



Cette icône exporte le dernier graphe au format graphml. Ce graphique peut alors être ouvert dans des logiciels comme Gephi ou visone.



Cette icône exporte la matrice de similitude.

### 2.6.6.3 Fichiers en sortie

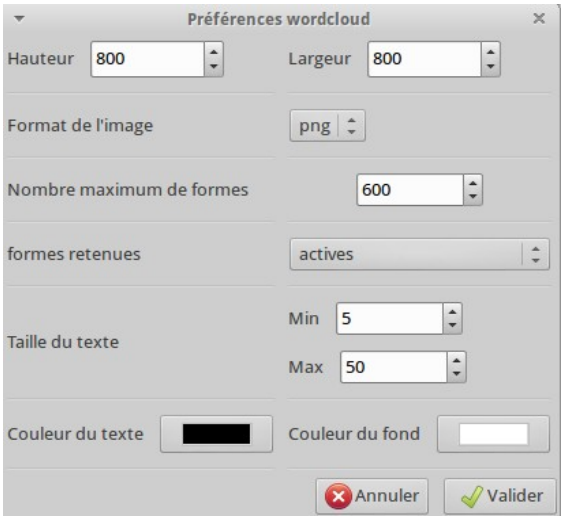



| Répertoire de sortie | NomDuCorpus_simitxt_x |
|---|---|
| **Fichiers en sortie :** | |
| Actives.csv | Liste des formes actives |
| Analyse.ira | Fichier Analyse : permet d'ouvrir une analyse. |
| graph_X.graphml | Graphique au format .graphml exploitable dans gephi |
| graph_simi_X.png | Graphique au format .PNG (à chaque réédition du graphe, un nouveau fichier se crée) |
| liste_graph.txt | Fichier Iramuteq |
| listeuce1.csv | Fichier Iramuteq |
| mat01.csv | Matrice utilisée pour l'analyse au format MarketMatrix |
| RData.RData | Fichier R |
| Selected.csv | Liste des identifiants des formes sélectionnées (se recharge à chaque réédition) |

## 2.6.7 Nuage de mots

### 2.6.7.1Paramétrage de l'analyse Nuage de mots

Après le menu des préférences (voir chapitre 2.4), un menu de paramétrage s'affiche.




- **Hauteur/Largeur** : fixe les dimensions du nuage
- **Format de l'image** : permet d'enregistrer l'image en .PNG ou .svg
- **Nombre maximum de formes** : Nombre maximum de formes présentes dans le nuage
- **Formes retenues** : permet de choisir entre Formes actives et/ou supplémentaires
- **Taille du texte** : fixe les limites supérieures et inférieures de la taille du texte
- **Couleur du texte/fond** : défini les couleurs

Après validation, la liste des formes et leur fréquence sont affichées, par défaut elles sont toutes sélectionnées, l'utilisateur peut en exclure, ou choisir sa liste (pour sélectionner plusieurs formes, maintenir la touche « Ctrl » lors de la sélection)

### 2.6.7.2 Résultats de l'analyse Nuage de mots

Les résultats s'ouvrent dans un nouvel onglet.

### 2.6.7.3 Fichiers en sortie pour les nuages de mots

| Répertoire de sortie | NomDuCorpus_worldcloud_x |
|---|---|
| **Fichiers en sortie :** | |

| | |
|---|---|
| Actives_eff.csv | Liste des formes actives avec leur fréquence |
| Analyse.ira | Fichier Analyse : permet d'ouvrir une analyse. |
| Nuage_1.png | Image de sortie du nuage |
| Selected.csv | Liste des identifiants des formes sélectionnées |

## 2.7 Réédition des graphiques dans les analyses textuelles

### 2.7.1 Menu de réédition des dendrogrammes

Cette boîte de dialogue est accessible par l'icône  à gauche des dendrogrammes

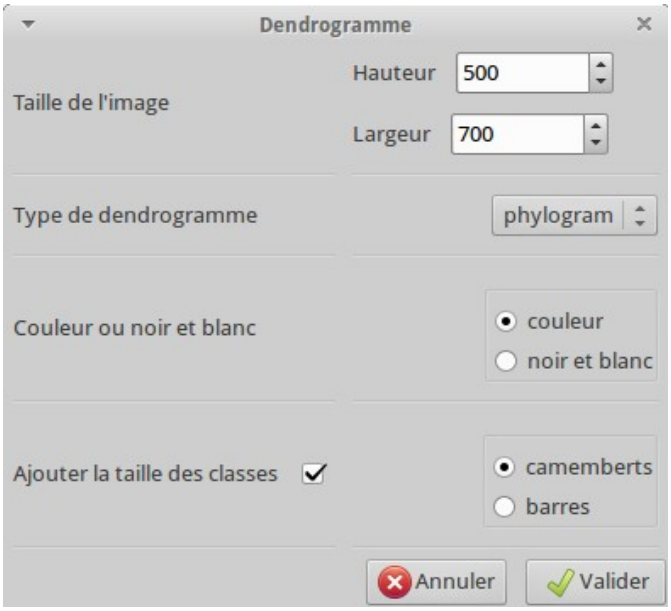



*Illustration 59: Dialogue de réédition des dendrogrammes*

- **Hauteur / largeur :** Taille de l'image en pixels
- **Type de dendrogramme :** permet de choisir le type de dendrogramme (phylogram, cladogram, fan, unrooted, radial)

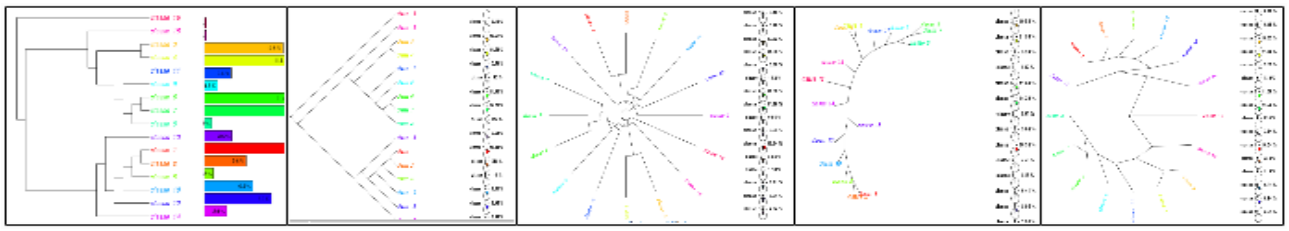


*Illustration 60: Exemple des différents types de dendrogrammes*

- **Ajouter la taille des classes :** permet de faire apparaître sur le

dendrogramme la taille des classes (en barres ou camemberts)

### 2.7.2 Menu de réédition des  AFC

Cette boîte de dialogue est accessible par l'icône  à gauche des AFC

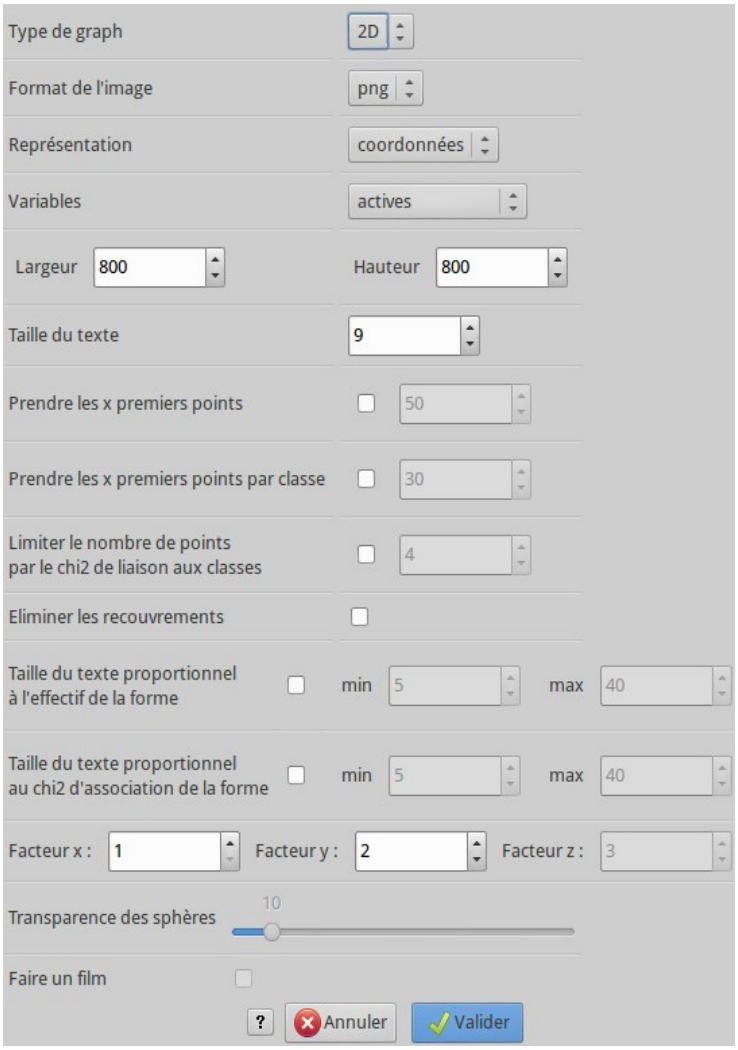



*Illustration 61: Dialoque de réédition des AFC*

- **Type de graphique** : permet de choisir les graphiques en 2D ou 3D
- **Format de l'image** : permet de choisir le format d'enregistrement de l'image
- **Représentation** : choisi le mode de représentation par coordonnées , ou corrélations
- **Variable** : choisi quelles formes analyser, Actives, supplémentaires, classes (ici le titres des colonnes)
- **Largeur/hauteur** : défini les dimensions du graphique (en px)
- **Taille du texte** : défini la taille du texte
- **Prendre les x premiers points** : Prends les x premiers points qui ont les chi2 les plus forts dans l'ensemble de la classification.
- **Prendre les x premiers points par classe :** Prend les x premiers points qui ont les chi2 les plus forts par classe
- **Limiter le nombre de points par le chi2 de liaison par classe** : seuil minimal de liaison à la classe pour apparaître
- **Éliminer les recouvrements** : cette option évite le recouvrement du label des points,  elle produit un biais, mais permet une meilleure lisibilité.
- **Taille du texte proportionnel à l'effectif de la forme** : règle la taille du

texte en fonction de l'effectif de la forme dans le corpus.

- **Taille du texte proportionnel au chi2 d'association de la forme** : règle la taille du texte en fonction du  chi2 de liaison à la classe
- **facteurX /facteurY/facteurZ** : respectivement facteur en abscisse, facteur en ordonnée, facteur  en cote
- **Transparence des sphères** : permet d'ajuster la transparence des sphères dans les graphiques en 3D
- **Faire un film** : disponible uniquement pour les graphiques en 3D, produit un .gif animé qui représente une rotation de 360° du graphique.