

Documentation IRaMuTeQ

0.6 alpha 3

version 0.1

Auteurs :
Lucie Loubère
Pierre Ratinaud

Avertissements

Iramuteq est en cours de développement. Regardez les informations disponibles sur la page <http://www.iramuteq.org/news> pour connaître la fiabilité des différentes analyses.

19/02/14

Table des matières

1Présentation d'iramuteq.....	4
1.1Présentation générale.....	4
1.2Principe de fonctionnement du logiciel.....	4
2Analyses de textes.....	4
2.1Un peu de vocabulaire.....	4
2.2Format des données en entrée.....	4
2.2.1Généralités.....	4
2.2.2Les règles de formatages.....	4
2.3Ouverture et indexation d'un corpus texte.....	6
2.3.1Indexation.....	6
2.3.1.1Onglet « Général ».....	7
2.3.1.2 Onglet Nettoyage.....	8
2.4Traitements commun aux analyses (option de lemmatisation).....	9
2.4.1Lemmatisation.....	9
2.4.2Clés d'analyse.....	10
2.5Les différentes analyses textuelles.....	11
2.5.1Statistiques textuelles.....	11
2.5.1.1Paramétrage de l'analyse statistiques textuelles.....	11
2.5.1.2Résultats de l'analyse statistiques textuelles.....	12
2.5.1.2.1 Onglet Global.....	12
2.5.1.2.2 Onglet formes actives.....	12
2.5.1.2.3 Onglet Formes supplémentaires.....	13
2.5.1.2.4 Onglet Total.....	13
2.5.1.2.5 Onglet Hapax.....	14
2.5.1.2.6 Options Supplémentaires de l'analyse statistiques textuelles.....	14
2.5.1.2.6.1 Sur chaque forme.....	14
2.5.1.2.6.2 Sur l'analyse.....	14
2.5.1.3 Fichiers en sortie.....	15
2.5.2Spécificités et AFC.....	16
2.5.2.1 Paramétrage de l'analyse spécificités et AFC.....	16
2.5.2.2 Résultats Spécificités et AFC.....	17
2.5.2.2.1 Analyses complémentaires des Spécificités et AFC.....	18
2.5.2.3Fichiers en sortie.....	19
2.5.3Classification Méthode Reinert.....	20
2.5.3.1 Paramétrages de l'analyse Classification Méthode Reinert.....	20
2.5.3.2 Résultats de la Classification Méthode Reinert.....	21
2.5.3.2.1 Onglet CHD.....	22
2.5.3.2.2 Onglet Profils.....	23
2.5.3.2.3 Onglet AFC de la méthode Reinert.....	24
2.5.3.2.4 Options supplémentaires de la Classification Méthode Reinert.....	25
2.5.3.2.4.1 Les options disponibles pour chaque classes.....	25
2.5.3.2.4.2 Les options disponibles pour la classification.....	27
2.5.3.3Fichiers en sortie pour la classification.....	29
2.5.4Analyses de similitude.....	30
2.5.4.1 Paramétrages de l'Analyse de Similitude sur textes.....	31
2.5.4.2 Résultats de l'analyse de similitude.....	34
2.5.4.3 Fichiers en sortie.....	34
2.5.5Nuage de mots.....	35
2.5.5.1Paramétrage de l'analyse Nuage de mots.....	35
2.5.5.2 Résultats de l'analyse Nuage de mots.....	35
2.5.5.3 Fichiers en sortie pour les nuages de mots.....	35

2.6 Réédition des graphiques dans les analyses textuelles.....	36
2.6.1 Menu de réédition des dendrogrammes.....	36
2.6.2 Menu de réédition des AFC.....	36

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.

2 ANALYSES 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

**** *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
```



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

-*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

**** *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

-*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
```



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

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 :

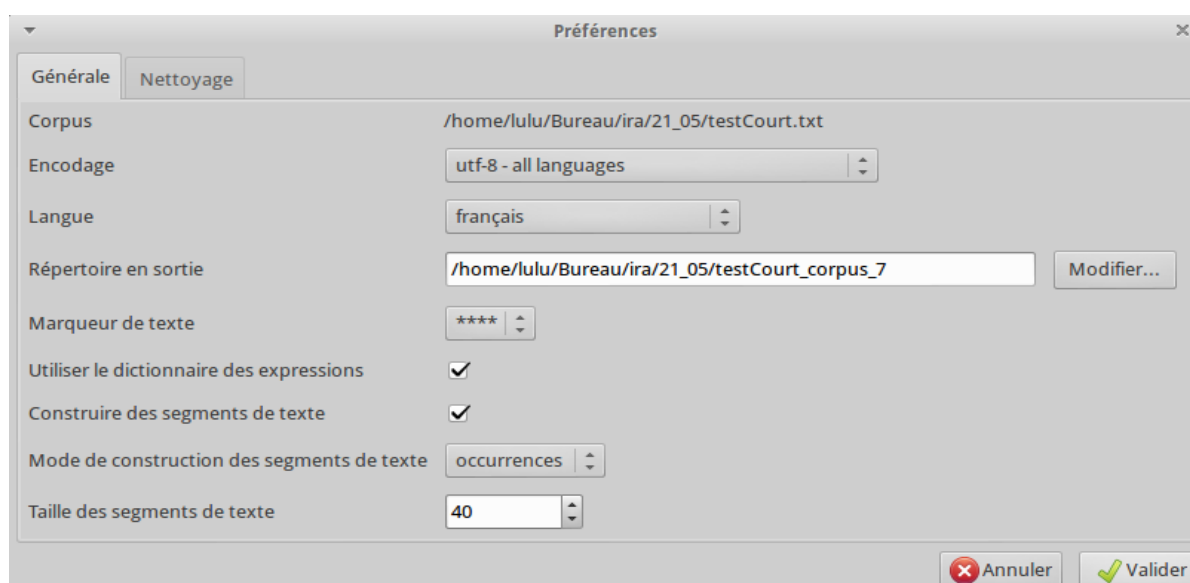


Illustration 1: Dialogue d'indexation des corpus

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)
- **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.

- **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 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. Iramuteg 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.

- **Taille des segments de textes** : définit le nombre d'unités (voir plus haut) contenus 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 segment 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 :

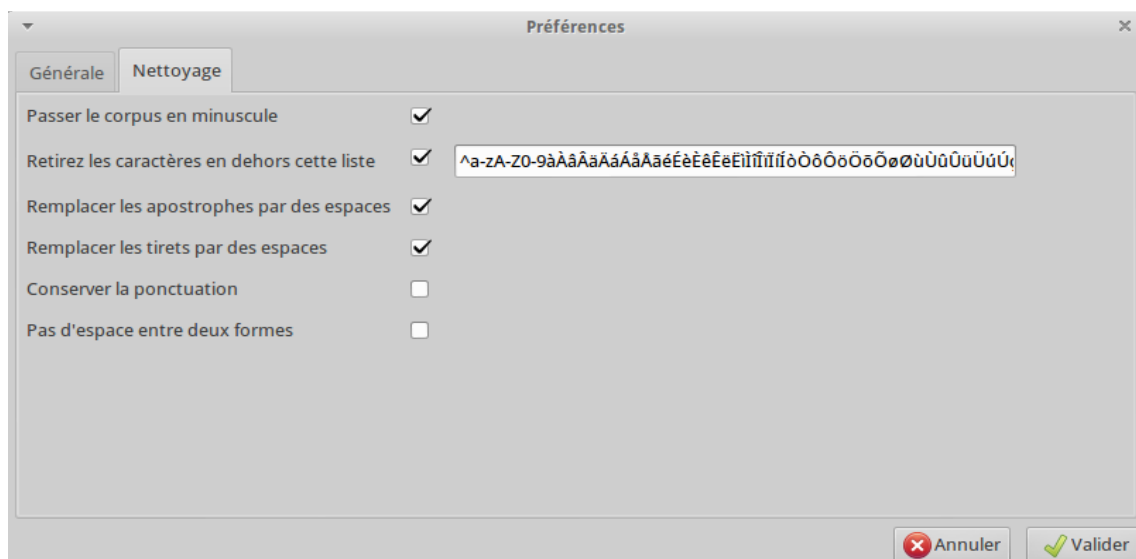


Illustration 2: Dialogue de nettoyage du corpus

- **Passer le corpus en minuscule** : Convertit toutes les majuscules en minuscules, si désactivé, Iramuteg considérera « Chat » et « chat » comme deux formes distinctes.

- **Retirer 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 Traitements commun aux analyses (option de lemmatisation)

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

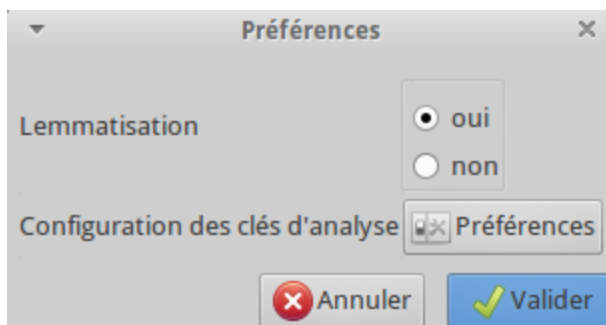


Illustration 3: Dialogue de lemmatisation

- **Lemmatisation** : Cette option choisit si le corpus doit être lemmatisé ou pas.
- **Clés d'analyse** : Permet de choisir les clés d'analyse.

2.4.1 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.



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

2.4.2 Clés d'analyse

La plupart des analyses sur corpus vont différencier 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 à chaque type grammatical.

Clés d'analyse					
Choix des clés d'analyse					
0=éliminé ; 1=active ; 2=supplémentaire					
Adjectif	1	voir liste	Conjonction	2	voir liste
Adjectif démonstratif	2	voir liste	Formes non reconnues	1	voir liste
Adjectif indéfini	2	voir liste	Nom commun	1	voir liste
Adjectif interrogatif	2	voir liste	Nom supplémentaire	2	voir liste
Adjectif numérique	2	voir liste	Onomatopée	2	voir liste
Adjectif possessif	2	voir liste	Pronom démonstratif	2	voir liste
Adjectif supplémentaire	2	voir liste	Pronom indéfini	2	voir liste
Adverbe	1	voir liste	Pronom personnel	2	voir liste
Adverbe supplémentaire	2	voir liste	Pronom possessif	2	voir liste
Article défini	2	voir liste	Pronom relatif	2	voir liste
Article indéfini	2	voir liste	Préposition	2	voir liste
Auxiliaire	2	voir liste	Verbe	1	voir liste
Chiffre	2	voir liste	Verbe supplémentaire	2	voir liste

✓ Valider

Illustration 4: Dialogue choix des clés d'analyse

Ces données sont stockées dans le fichier key.cfg accessible dans le répertoire .iramuteq. Dans ce fichier la valeur 1 signifie que le type grammatical est actif, la valeur 2 signifie que le type est supplémentaire et la valeur 0 signifie que le type n'est pas pris en compte.

Voici la liste des types grammaticaux différenciés :

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 Les différentes analyses textuelles

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 aux analyses disponibles. Elles sont aussi accessibles dans la barre de tâche sous l'option « analyse de corpus ».

2.5.1 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.

2.5.1.1 Paramétrage de l'analyse statistiques textuelles

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

2.5.1.2 Résultats de l'analyse statistiques textuelles

2.5.1.2.1 Onglet Global

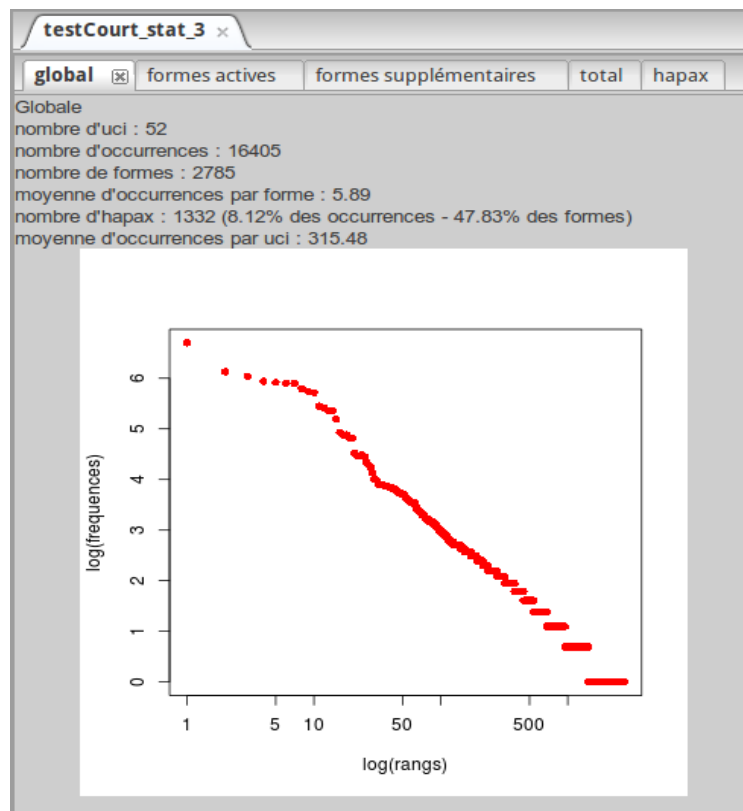


Illustration 5: Présentation des résultats de l'analyse statistique textuelle

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 : $(\text{nombre d'occurrence}) / (\text{Nombre de formes})$.
- Nombre d'Hapax : c'est le nombre de mots n'apparaissant qu'une seule fois dans tout le corpus.
- Moyenne d'occurrences par Textes: $(\text{nombre d'occurrences}) / (\text{nombre de Textes})$
- Le graphique : présente en abscisse les logarithmes des rangs et en ordonnées les logarithmes des fréquences des formes.

2.5.1.2.2 Onglet formes actives

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

testCourt_stat_3 x			
global	formes actives	formes supplémentaires	total
global	formes actives	total	hapax
forme	nb	type	
élève	125	nom	
collège	123	nom	
numérique	62	adj	
lycée	55	nom	
établissement	54	nom	
conseil	49	nom	
général	49	nom	

Illustration 6: Présentation des résultats de l'analyse textuelle, onglet formes actives

2.5.1.2.3 Onglet Formes supplémentaires

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

global	formes actives	formes supplémentaires	total	hapax
global	formes actives	formes supplémentaires	total	hapax
forme	nb	type		
de	2142	pre		
l	1084	art_def		
la	997	art_def		
le	977	art_def		
et	966	con		
les	947	art_def		
d	833	pre		

Illustration 7: Présentation des résultats de l'analyse textuelle, onglet formes supplémentaires

2.5.1.2.4 Onglet Total

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

global	formes actives	formes supplémentaires	total	hapax
forme	nb	type		
de	2142	pre		
l	1084	art_def		
la	997	art_def		
le	977	art_def		
et	966	con		
les	947	art_def		
d	833	pre		
être	820	aux		
à	800	pre		

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

2.5.1.2.5 Onglet Hapax

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

global	formes actives	formes supplémentaires	total	hapax
forme	nb	type		
évolutif	1	adj		
éveline	1	nr		
éveiller	1	ver		
éveil	1	nom		
étranger	1	nom		
étoile	1	nom		
étoffer	1	ver		
étienne	1	nr		
étendue	1	nom		

Illustration 9: Présentation des résultats de l'analyse textuelle, onglet Hapax

2.5.1.2.6 Options Supplémentaires de l'analyse statistiques textuelles

2.5.1.2.6.1 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.

2.5.1.2.6.2 Sur l'analyse

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

Informations
Exporter le dictionnaire
Exporter le dictionnaire des lemmes
Supprimer de l'historique

Illustration 10: Dialogue d'option de l'analyse statistique textuelle

Informations : Permet de récupérer les informations propres à l'analyse.

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	Forme1	Effectif forme 1	Forme 2	Effectif forme 2
Primaire	Primaire	6	Primaires	9

2.5.1.3 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 :	
total.csv	Toute les formes et leur effectif
formes_supplémentaires.csv	Les formes supplémentaires et leur effectif
formes_actives.csv	Les formes actives et leur effectif
glob.txt	Résultats de l'analyse
hapax.csv	Les hapax
Zipf.png	L'image du graphique présent dans « Global »
Analyse.ira	Fichier permettant à Iramuteq d'ouvrir l'analyse.

2.5.2 Spécificités et AFC

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

2.5.2.1 Paramétrage de l'analyse spécificités et AFC

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

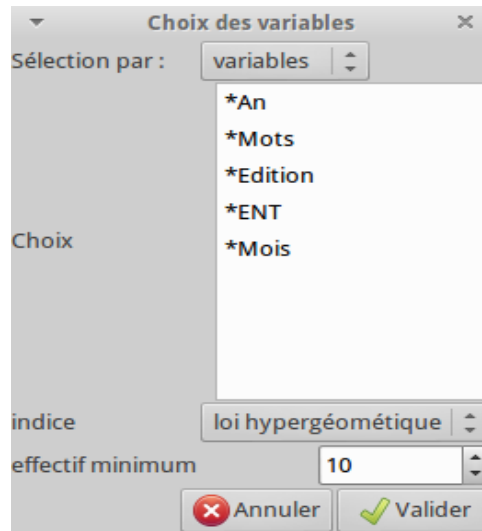


Illustration 11: Dialogue de paramétrage de l'analyse spécificité et AFC

Voici les différentes options proposées :

- 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 : Deux possibilités de calcul sont possibles, le Chi2 ou la loi hypergéométrique.
- Effectif minimum : 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.

2.5.2.2 Résultats Spécificités et AFC

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

- Les onglets **Formes et Types** : ils présentent la liste des formes et des catégories grammaticales, et leurs scores par modalité.

formes	Types	Effectifs formes	Effectifs Type	Effectifs relatifs formes	Effectifs relatifs Type	AFC
		X.An_2008	X.An_2009	X.An_2011	X.An_2012	
aide		1.5685	-1.9515	1.3851	-0.6615	
petit		1.5685	-0.4762	-0.4301	-0.3683	
difficulté		1.5625	-0.6887	-0.4393	-0.3846	
l		1.4475	-2.0832	0.4072	0.9053	
orientation		1.3216	-0.9903	1.0674	-1.1383	

Illustration 12: Présentation des résultats de l'analyse spécificités et AFC onglet Formes

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

formes	Types	Effectifs formes	Effectifs Type	Effectifs relatifs formes	Effectifs relatifs Type	AFC
		*An_2008	*An_2009	*An_2011	*An_2012	
de		326	1005	436	375	
la		184	447	199	167	
l		151	474	231	228	
et		146	399	197	224	
des		144	346	203	101	

Illustration 13: Présentation des résultats de l'analyse spécificités et AFC onglet Effectifs

- Les onglets **Effectifs relatifs** donnent les effectifs relatifs en pour mille (‰)

formes	Types	Effectifs formes	Effectifs Type	Effectifs relatifs formes	Effectifs relatifs Type	AFC
		X.An_2008	X.An_2009	X.An_2011	X.An_2012	
de		86.68	68.33	67.03	61.53	
la		48.92	30.39	30.59	27.4	
l		40.15	32.23	35.51	37.41	
et		38.82	27.13	30.28	36.75	
des		38.29	23.52	31.21	16.57	

Illustration 14: Présentation des résultats de l'analyse spécificités et AFC onglet Effectifs relatifs

- L'onglet **AFC** : Présente le graphique de l'AFC (facteurs 1/2)

formes	Types	Effectifs formes	Effectifs Type	Effectifs relatifs formes	Effectifs relatifs Type	AFC
		X.An_2008	X.An_2009	X.An_2011	X.An_2012	
aide		1.5685	-1.9515	1.3851	-0.6615	
petit		1.5685	-0.4762	-0.4301	-0.3683	
difficulté		1.5625	-0.6887	-0.4393	-0.3846	
l		1.4475	-2.0832	0.4072	0.9053	
orientation		1.3216	-0.9903	1.0674	-1.1383	

Illustration 15: Présentation des résultats de l'analyse spécificités et AFC onglet Effectifs relatifs



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

2.5.2.2.1 Analyses complémentaires des Spécificités et AFC

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, le concordancier des segments de textes où figure la forme, ainsi qu'un graphique reprenant les scores spécifiques de la forme par modalité de variable.

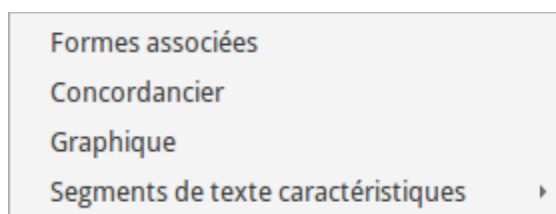


Illustration 16: Dialogue d'option de l'analyse spécificités et AFC



2.5.2.3 Fichiers en sortie

ertoire de sortie	NomDuCorpus_Spec_x
Fichiers en sortie	
Tabletypem.csv	Effectif des types grammaticaux par modalité
Tablespect.csv	Score de spécificité des Types grammaticaux par modalité
Tablespectf.csv	Score de spécificité des formes par modalité
Tableafcm.csv	Effectif des formes par modalité
liste_graph_afct.txt	Fichier Iramuteq
liste_graph_afcf.txt	Fichier Iramuteq
eff_relatif_type.csv	Effectif relatif des types grammaticaux par modalité
eff_relatif_forme.csv	Effectif relatif des formes par modalité
Analyse.ira	Permet d'ouvrir l'analyse
afct_row.png	Plan factoriel des types grammaticaux
afct_row.csv	Détail des valeurs des lignes de l'AFC sur les types grammaticaux
afct_facteur.csv	Valeurs Propres, Pourcentages, Pourcentages cumulés de chaque facteur (pour les types grammaticaux)
afct_col.png	Plan factoriel des colonnes de l'AFC sur les types grammaticaux
afct_col.csv	Détail des valeurs des colonnes de l'AFC sur les types grammaticaux
afcf_row.png	Plan factoriel des formes
afcf_row.csv	Détail des valeurs des lignes de l'AFC sur les formes
afcf_facteur.csv	Valeurs Propres, Pourcentages, Pourcentages cumulés de chaque facteur (pour les formes)
afcf_col.png	Plan factoriel des colonnes de l'AFC sur les formes
afcf_col.csv	Détail des valeurs des colonnes de l'AFC sur les formes

2.5.3 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.

Voir le détail de la classification selon la méthode Reinert en annexe.

2.5.3.1 Paramétrages de l'analyse Classification Méthode Reinert

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

Illustration 17: 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 (voir annexe). Elle permet une analyse un peu plus rapide mais un peu moins précise.

2.5.3.2 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).

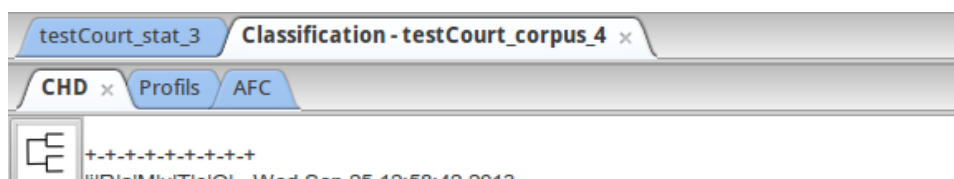
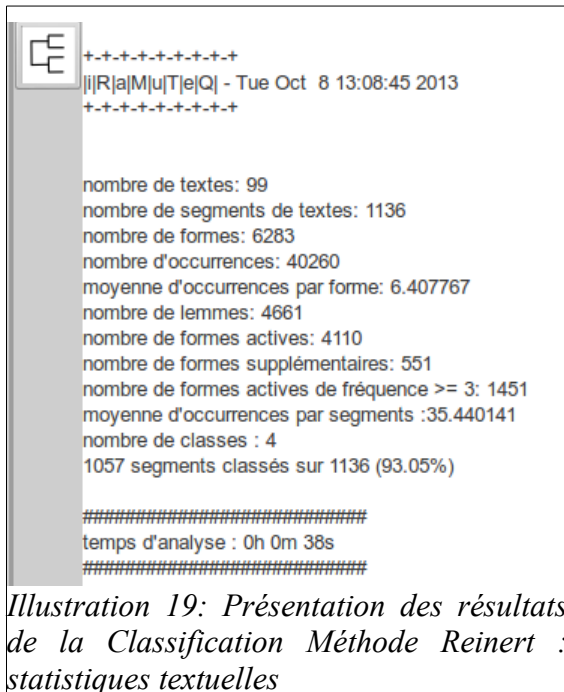


Illustration 18: Présentation des résultats de la Classification Méthode Reinert

2.5.3.2.1 Onglet CHD

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

le premier, constitué de texte, reprend les statistiques textuelles du corpus. En plus des items déjà vus au chapitre 2.5.1, vous pourrez y trouver :



- moyenne d'occurrences par segment
- le nombre de segments classés sur la totalité du corpus.

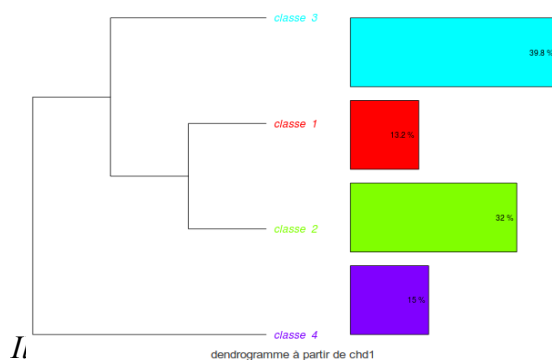


Illustration 20: Présentation des Résultats de la Classification Méthode Reinert : Dendrogramme

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, apparaît le bouton vous permettant de refaire le dendrogramme. (voir chapitre 1.3.1)

2.5.3.2.2 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							
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 22: 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 segment 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. A 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 23: 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 la touche CTRL+F
- Pour rechercher une forme dans tous les profils, vous pouvez appuyer sur la touche CTRL+Shift+F

2.5.3.2.3 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.
- 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.

2.5.3.2.4 Options supplémentaires de la Classification Méthode Reinert

2.5.3.2.4.1 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 :

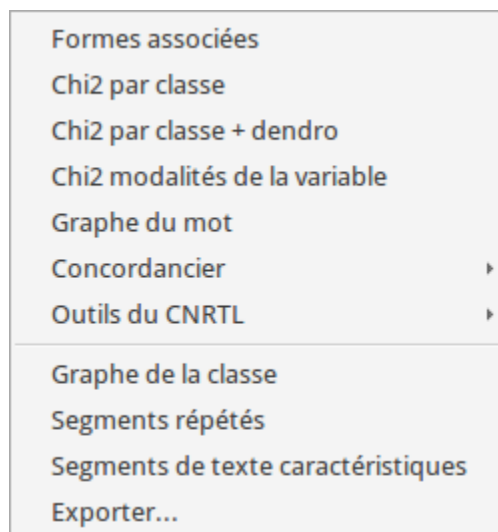


Illustration 24: Dialogue d'options disponibles pour chaque classe de la classification Méthode Reinert

- **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)

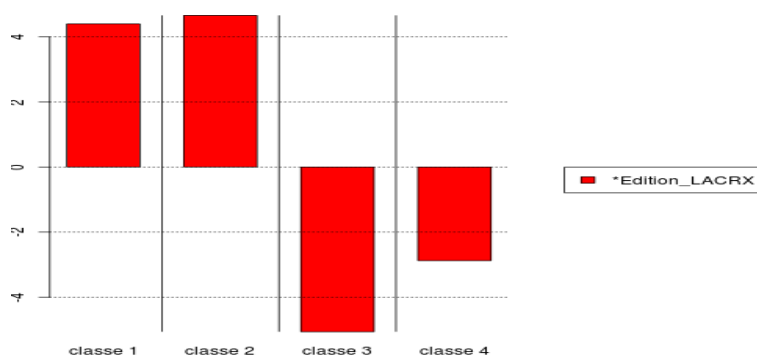


Illustration 25: Chi2 de la forme par classe

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

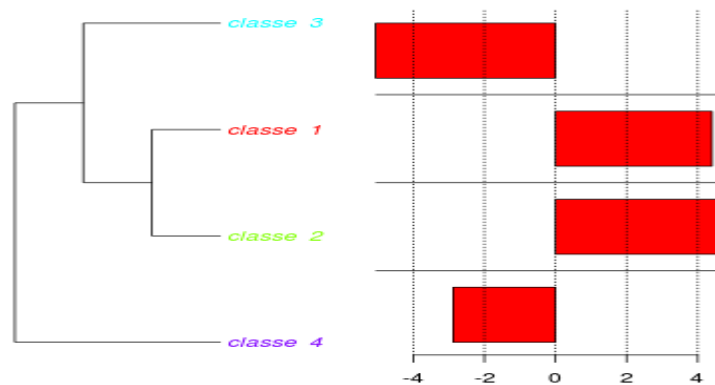


Illustration 26: Chi2 de la forme + dendro

- **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.

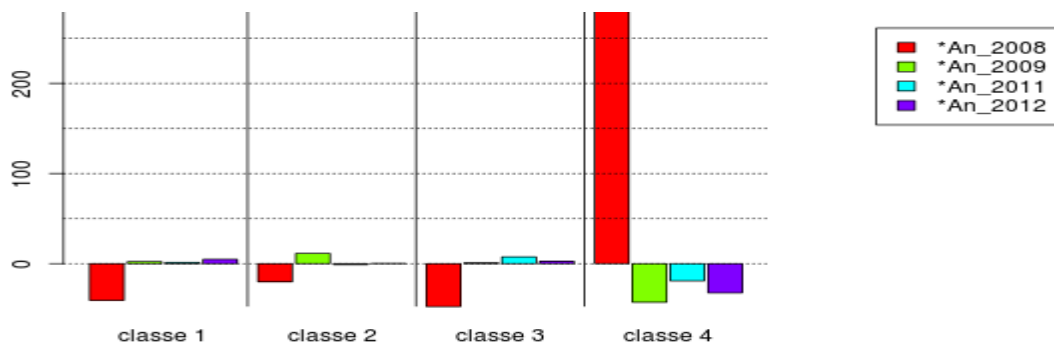
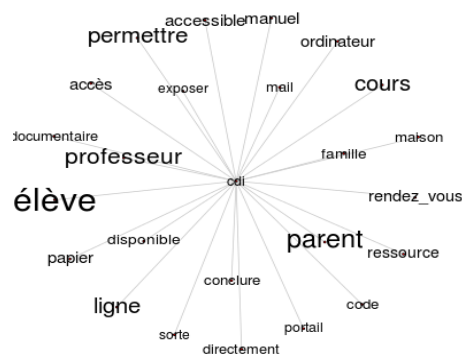


Illustration 27: Chi2 des modalités par classe

- **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.

- **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.
- **Exporter** : exporte tous les ST présents dans la classe, ils peuvent ainsi constituer un nouveau corpus.

2.5.3.2.4.2 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.

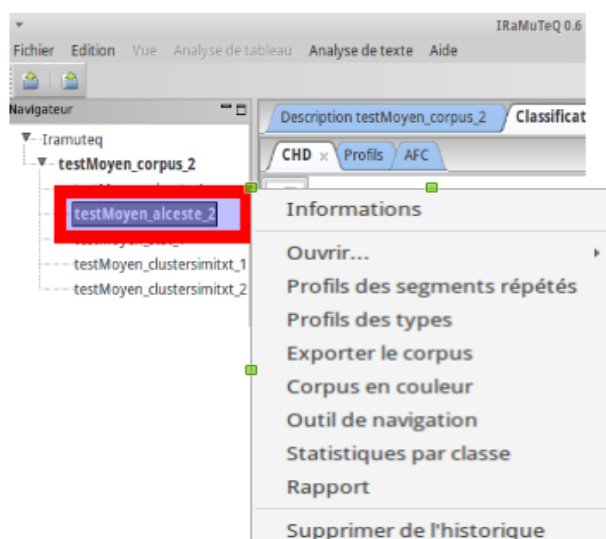


Illustration 28: Dialogue d'options disponibles pour l'analyse classification méthode Reinert

- **Informations** : reprend les caractéristiques choisies pour la classification (voir chapitre 2.5.3.1)
- **Ouvrir** -> Anti-profils, sur la même présentation que les profils seront affichés les anti-profils, ce sont les formes significativement absentes de la classe.
- **Profils des segments répétés** : Calcule les profils des segments répétés.

CHD Profils AFC Antiprofils Profils des segments répétés x Profils des types							
classe 1 classe 2 classe 3 classe 4							
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	88.00		cahier de texte	< 0,0001

Illustration 29: Option profils de segments répétés

- **Profil des types** : Calcule les profils des types grammaticaux.
- **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.

Outlook - lu_jou@hotmail.fr

E.N.T. - Accueil Etudiant

file:///home/lu...us_couleur.html

+

file:///home/lu/Bureau/ira/21_05/testMoyen_corpus_2/testMoyen_alceste_2/corpus_couleur.html

mgen

**** *ENT_Environnement *Edition_LAMONT *Mots_201a500 *Mois_Mars *An_2012

l idée de monter un club journal a germé dans la tête de johan et ismaël deux élèves de 6e au collège de marie curie de désertines à la rentrée dernière entre copains décidé de proposer ce club au principal

nous lui avons rédigé une lettre quelques jours plus tard il a dit oui depuis quatorze élèves se retrouvent chaque semaine sous la direction de christelle pelletier ense français afin de concocter ce trimestriel

il s agira du premier journal de l histoire du collège indique l enseignante nous découvrons un nouvel univers les apprentis journalistes ont été plongés dans le grand directement édito interviews reportages divers mise en page

grâce à l environnement numérique de travail nous sommes en constante relation explique la professeur nous nous adressons des mails afin de faire le point sur les avancent ou d échanger des idées

limités par la technique et par le coût du journal les collégiens vont proposer un premier numéro de huit pages nous apprenons mais c est vrai que nous avons déjà e beaucoup plus pour le second numéro assure christelle pelletier

mais qu est ce qui a donc poussé ismaël ryan dorian thomas marine noémie johan et les autres à se lancer dans cette aventure qui rogne sur leur temps libre du mid on peut exposer nos idées à tout le monde indique johan cela permet d apporter des réponses aux gens et puis c est un plaisir cela détend des arguments qui ont sédu club a même dû refuser des candidats

Illustration 30: Export du corpus en couleur

- **Outil de navigation** : 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 31: Export de l'outil navigation

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

2.5.3.3 Fichiers 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
AFC2DSL.png	graphique AFC : variables supplémentaires - coordonnées - facteurs 1 / 2
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.
antiprof_segments.csv *	Antiprofils des segments répétés
antiprof_type.csv *	Antiprofils des types grammaticaux

Antiprofile.csv	Antiprofils des formes
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
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
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 »
Tablet.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
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).

2.5.4 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.

2.5.4.1 Paramétrages de l'Analyse de Similitude sur textes

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

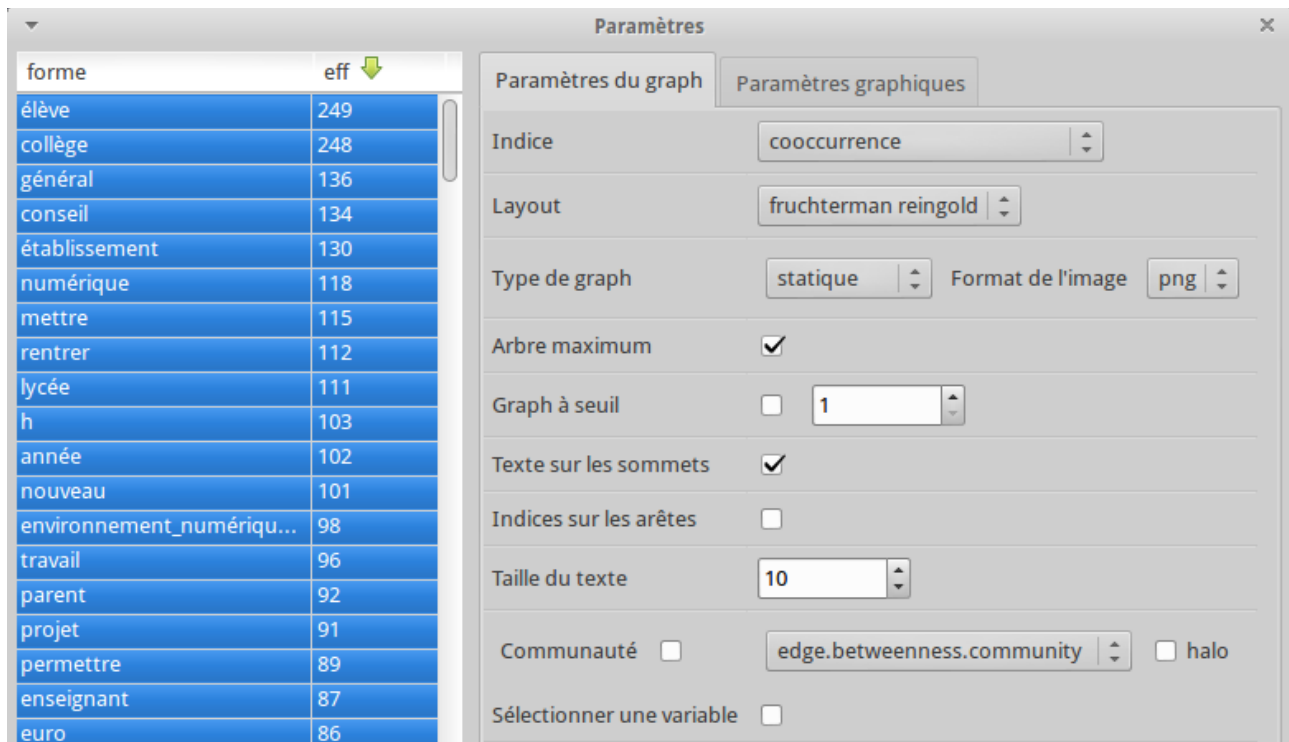


Illustration 32: Dialogue de paramétrage de l'analyse de similitude

Ce menu se compose de trois parties :

- La **liste 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-Kawaiï, graphopt. Voir la documentation de la librairie igraph pour plus de détails.

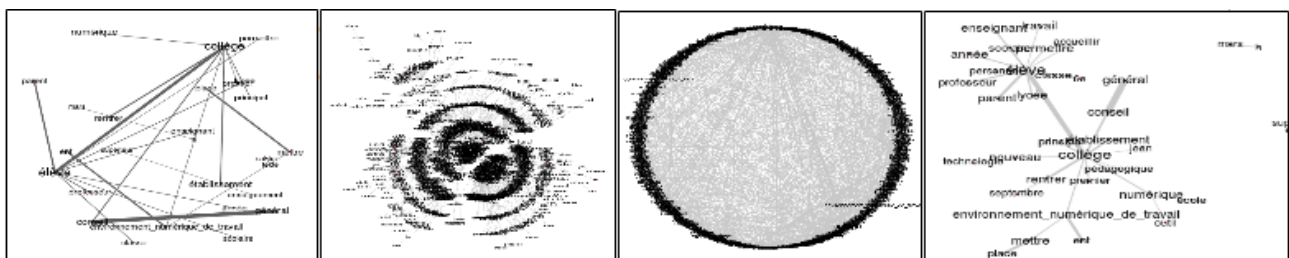


Illustration 33: Exemples des différents modes de représentations des données (Layout)

- **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
- **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.

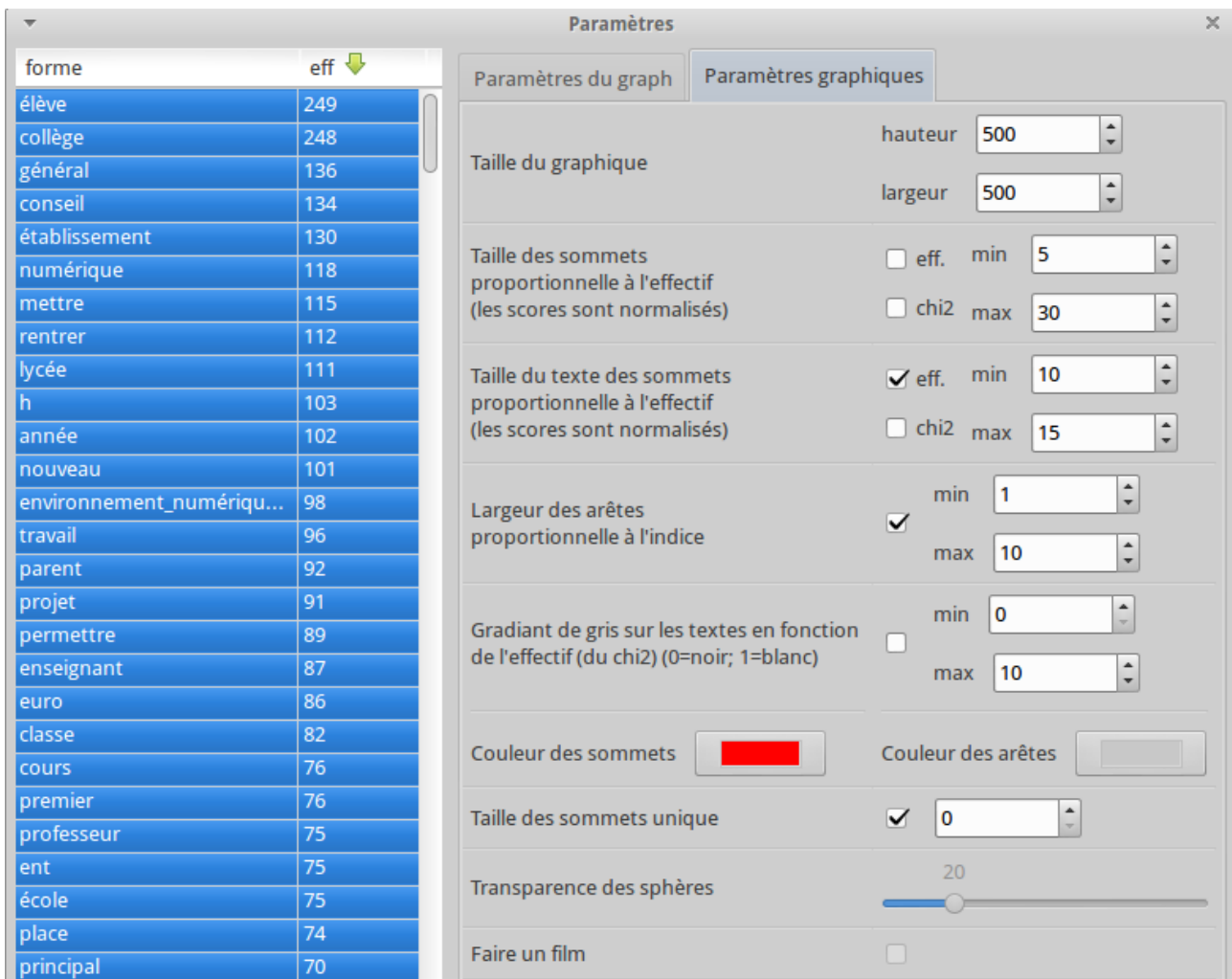


Illustration 34: Dialogue de paramétrage de graphiques de l'analyse de similitude

- Taille du graphique : taille de l'image produite (en pixels)
- Taille des sommets proportionnelle à l'effectif : 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 l'effectif : 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.
- Gradient de gris sur les textes en fonction de l'effectif (du chi2) 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

2.5.4.2 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.



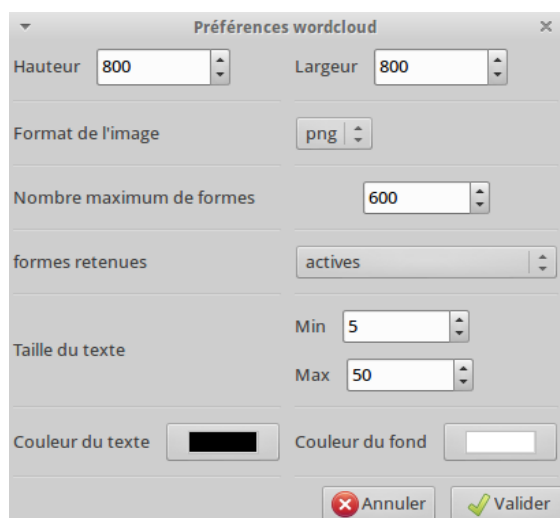
2.5.4.3 Fichiers en sortie

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_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)

2.5.5 Nuage de mots

2.5.5.1 Paramétrage 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** : définit 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)

2.5.5.2 Résultats de l'analyse Nuage de mots

Les résultats s'ouvrent dans un nouvel onglet.

2.5.5.3 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

2.6 Réédition des graphiques dans les analyses textuelles

2.6.1 Menu de réédition des dendrogrammes



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

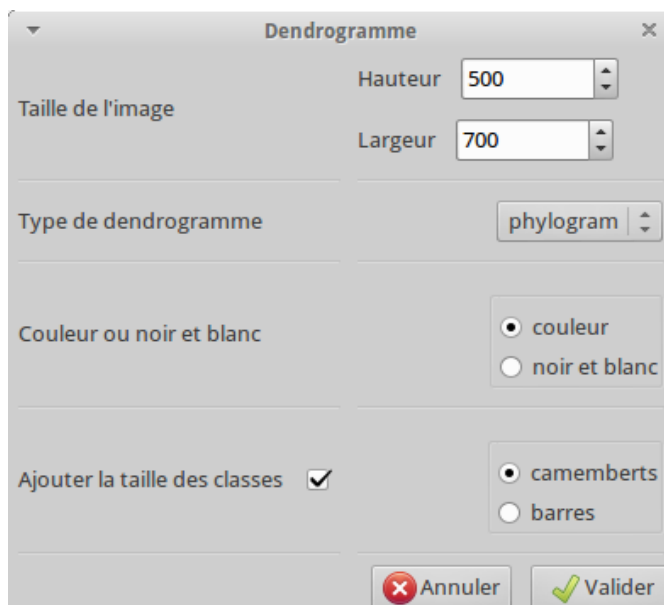


Illustration 35: Dialogue de réédition des dendrogrammes

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

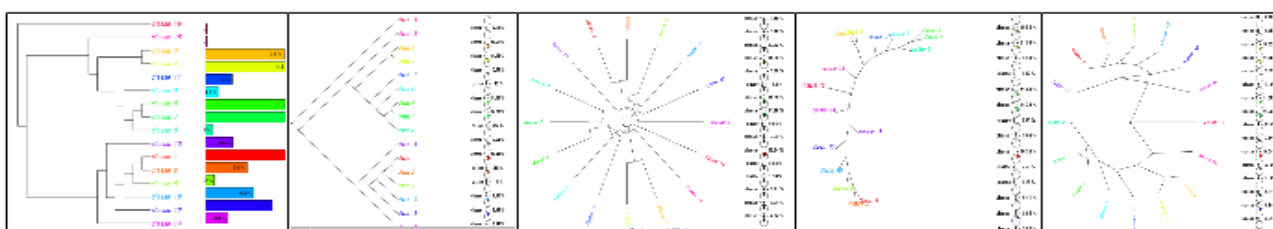


Illustration 36: Exemple des différents types de dendrogrammes

- **Ajouter la taille des classes** : permet de faire apparaître sur le dendrogramme la taille des classes (en barres ou camemberts)

2.6.2 Menu de réédition des AFC



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

Type de graph: 2D

Format de l'image: png

Représentation: coordonnées

Variables: actives

Largeur: 800 Hauteur: 800

Taille du texte: 9

Prendre les x premiers points: ☐ 50

Prendre les x premiers points par classe: ☐ 30

Limiter le nombre de points par le chi2 de liaison aux classes: ☐ 4

Éliminer les recouvrements: ☐

Taille du texte proportionnel à l'effectif de la forme: ☐ min 5 max 40

Taille du texte proportionnel au chi2 d'association de la forme: ☐ min 5 max 40

Facteur x: 1 Facteur y: 2 Facteur z: 3

Transparence des sphères: 10

Faire un film: ☐

? Annuler Valider

Illustration 37: Dialogue 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 quelles formes analyser, Actives, supplémentaires, classes (ici le titres des 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 point 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.