

---

# Documentation utilisateur DATOMATIC

*Release 2.0*

**LNE**

**Dec 13, 2019**



# CONTENTS

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Objectif du document . . . . .	3
1.2	Objectifs et fonctionnalités principales de DATOMATIC . . . . .	3
<b>2</b>	<b>Installer DATOMATIC</b>	<b>5</b>
2.1	Installation des packages . . . . .	5
2.2	Configuration du serveur . . . . .	5
2.3	Configuration du client . . . . .	6
<b>3</b>	<b>Lancer DATOMATIC</b>	<b>7</b>
<b>4</b>	<b>Tableau de bord</b>	<b>9</b>
4.1	L'espace de travail . . . . .	10
4.2	La recherche . . . . .	11
4.3	Les favoris . . . . .	11
4.4	Les historiques . . . . .	11
<b>5</b>	<b>Création de corpus</b>	<b>13</b>
5.1	Importer de nouveaux fichiers . . . . .	13
5.2	Fichiers média . . . . .	18
5.3	Changer le format d'annotation . . . . .	18
5.4	Changer les paramètres . . . . .	21
5.5	Nommer un corpus . . . . .	21
5.6	Changer les liens entre fichiers . . . . .	21
5.7	Enregistrer un corpus . . . . .	22
<b>6</b>	<b>Rechercher dans l'espace de stockage</b>	<b>25</b>
<b>7</b>	<b>Ajouter un corpus aux favoris</b>	<b>31</b>
7.1	Depuis le tableau de bord . . . . .	31
7.2	Depuis la fenêtre d'exploration de corpus . . . . .	31
<b>8</b>	<b>Visualiser les fichiers audio</b>	<b>35</b>
<b>9</b>	<b>Download de corpus</b>	<b>37</b>
9.1	Récupération de corpus . . . . .	37
9.2	Exportation de corpus . . . . .	37
<b>10</b>	<b>Supprimer un corpus</b>	<b>39</b>
10.1	L'utilisateur masque un corpus . . . . .	39
10.2	L'administrateur supprime le corpus de la base de données . . . . .	39

10.3	L'administrateur supprime le corpus dans l'espace de stockage . . . . .	39
<b>11</b>	<b>Administration de DATOMATIC</b>	<b>45</b>
11.1	Mettre à jour les champs des formulaires . . . . .	45
11.2	Supprimer des Corpus et des fichiers de l'espace de stockage . . . . .	45
11.3	Gérer les utilisateurs de DATOMATIC . . . . .	45
<b>12</b>	<b>Création de corpus en ligne de commande</b>	<b>53</b>
12.1	Commande scan . . . . .	53
12.2	Commande import . . . . .	53

Table des matières:



## INTRODUCTION

### 1.1 Objectif du document

Ce document est la documentation utilisateur du logiciel DATOMATIC développé par le LNE.

### 1.2 Objectifs et fonctionnalités principales de DATOMATIC

DATOMATIC est un système de gestion de données multimédia. Il couvre les fonctionnalités suivantes :

- Importer et indexer des corpus, des fichiers
- Rechercher des fichiers et des corpus
- Visualiser des listes de corpus, fichiers sources, versions, annotations de référence et annotations hypothèses
- Visualiser des fichiers sources de type audio accompagnés de leur annotations
- Obtenir des statistiques sur les corpus, les fichiers
- Exporter des données
- Administrer l'utilisation de DATOMATIC



## INSTALLER DATOMATIC

### 2.1 Installation des packages

Il faut, sur toutes les machines, installer les packages suivants :

- `xkb-data_2.12-1_all.deb` (sur debian 7 uniquement)
- `optgcc_8.1.0-1_amd64.deb` (gcc 8.1.0 dans `/opt/gcc`)
- `optffmpeg_4.0.2-1_amd64.deb` (ffmpeg 4.0.2 dans `/opt/ffmpeg`)
- `optqt_5.11.1-1_amd64.deb` (qt 5.11.1 dans `/opt/qt`)
- `optqtav_1.12.0-1_amd64.deb` (qtav 1.12.0 dans `/opt/qt`)
- `matix_2.0-1_amd64.deb` (matix dans `/opt/matix`)

Sur la machine serveur, il faut en plus :

- `optmongo_3.4.3-1_amd64.deb` (mongodb dans `/opt/mongo`)
- `optmongoc_1.13.0-1_amd64.deb` (interfaçage C pour mongodb dans `/opt/mongo`)
- `optmongocxx_3.3.0-1_amd64.deb` (interfaçage C++ pour mongodb dans `/opt/mongo`)

Optionnellement, pour recompiler certains packages, il est possible d'ajouter :

- `optcmake_3.8.0-1_amd64.deb` (cmake dans `/opt/cmake`)
- `optscons_2.5.1-1_amd64.deb` (scons dans `/opt/scons`)

### 2.2 Configuration du serveur

La machine serveur doit avoir dans le home directory de l'utilisateur serveur un fichier de configuration `$HOME/.config/lne.fr/Datomic-serveur.conf`. Il contient:

```
[General]
data-path=/data/datomic
cache-path=/data/datomic-cache
```

Où `/data/datomic` est le répertoire où le serveur doit stocker les corpus, et `/data/datomic-cache` est un répertoire sur le même filesystem où `datomatic` stockera les fichiers temporaires (en cours de transfert avant création d'un corpus). Il est aussi possible de rajouter une ligne `"port=nnn"` pour choisir le port de connexion des clients.

Cette machine devra être configurée pour lancer `/opt/mongo/bin/mongod` et `/opt/matix/bin/Datomic-serveur` au démarrage.

## 2.3 Configuration du client

Au premier démarrage le client ne pourra probablement pas se connecter au serveur (la machine visée est localhost par défaut). Les préférences apparaîtront automatiquement pour indiquer le nom/IP et le port de la machine contenant le serveur.

## LANCER DATOMATIC

Pour lancer DATOMATIC, saisissez `/opt/matics/bin/Datomic` dans un terminal.



## TABLEAU DE BORD

Une fois DATOMATIC lancé, le tableau de bord apparaît comme illustré par *Tableau de bord*.

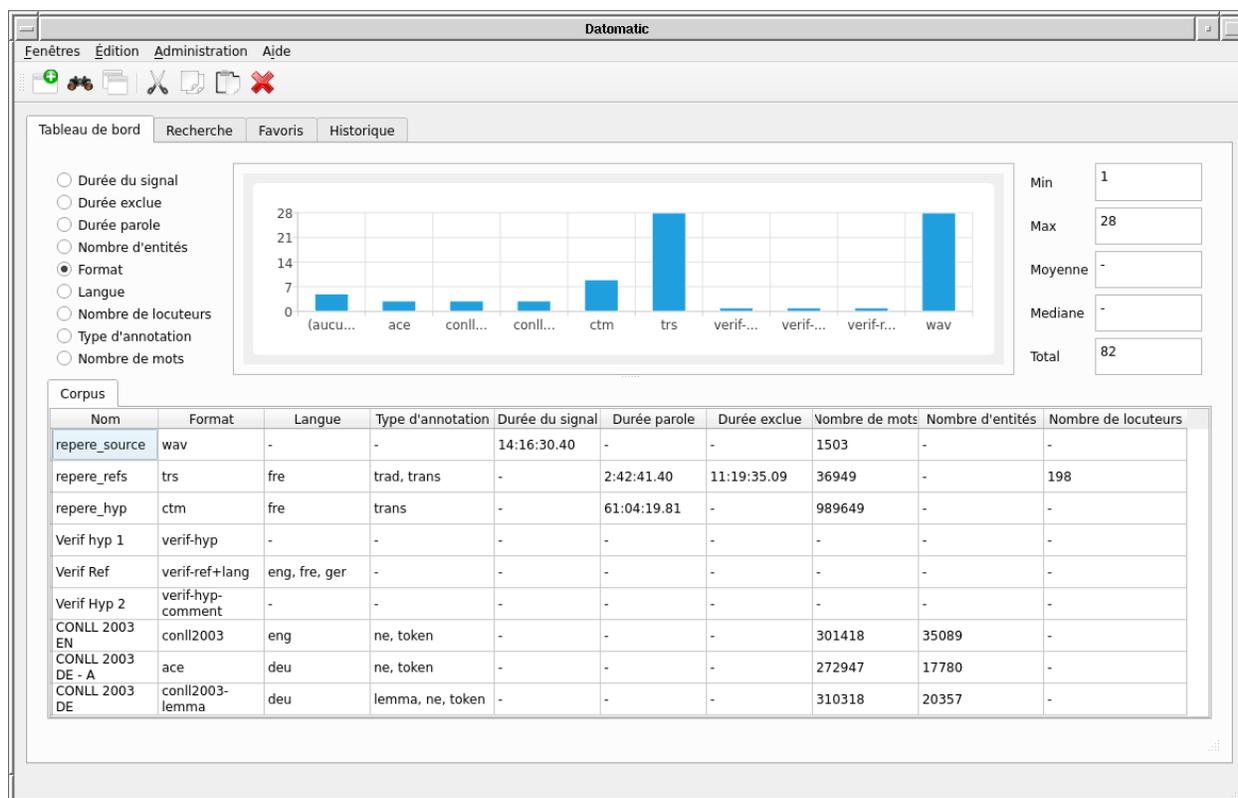


Fig. 1: Tableau de bord

Ce tableau de bord contient 4 onglets :

- L'espace de travail
- La recherche
- Les favoris
- Les historiques

## 4.1 L'espace de travail

L'espace de travail permet d'accéder à l'ensemble des corpus indexés comme le montre *Tableau de bord*

Cet espace est divisé en deux parties.

La première partie présente, sous forme d'histogrammes et de tableau des informations synthétiques, ce qui est contenu dans l'espace de stockage. Ainsi lorsque l'utilisateur choisit la visualisation des modalités, il obtient la répartition des fichiers selon leur modalité (fichiers hypothèses, fichiers références, fichiers audio, fichiers vidéo, fichiers texte...). En changeant le bouton radio, l'utilisateur peut obtenir les répartitions des fichiers selon les informations suivantes, dans la mesure où elles sont pertinentes pour les corpus considérés :

- Nombre de mots
- Nombre d'entités
- Nombre de locuteurs
- Durée de signal
- Durée de parole
- Durée exclue
- Format de fichier
- Langue
- Type d'annotation

Les tableaux pour les informations catégorielles (Format, langue, type d'annotation) contiennent les informations suivantes :

- Le nombre de fichiers pour la catégorie qui contient le moins de fichiers
- Le nombre de fichiers pour la catégorie qui contient le plus de fichiers
- Le nombre total de fichiers

Les tableaux pour les informations numériques (Nombres et durées) contiennent les informations suivantes :

- La plus petite valeur rencontrée pour un fichier
- La plus grande valeur rencontrée pour un fichier
- La valeur médiane de la distribution
- La valeur moyenne de la distribution
- Le total des valeurs.

La seconde partie liste l'ensemble des corpus présents dans la plateforme. Pour chaque corpus les informations des histogrammes sont disponibles.

Un séparateur invisible placé au bas de l'histogramme peut être déplacé avec la souris pour changer la répartition verticale de la fenêtre, avec la possibilité de supprimer un des deux panneaux, cf [:ref:tableau\\_de\\_bord\\_f:](#) et [:ref:tableau\\_de\\_bord\\_h:](#).

Cliquer sur un corpus l'ouvre en mode exploration, où sont détaillés les fichiers qui le composent et les informations associées comme illustré par *Ouvrir un corpus en mode exploration fichiers*. Cette vue contient aussi les histogrammes réduits au corpus considéré.

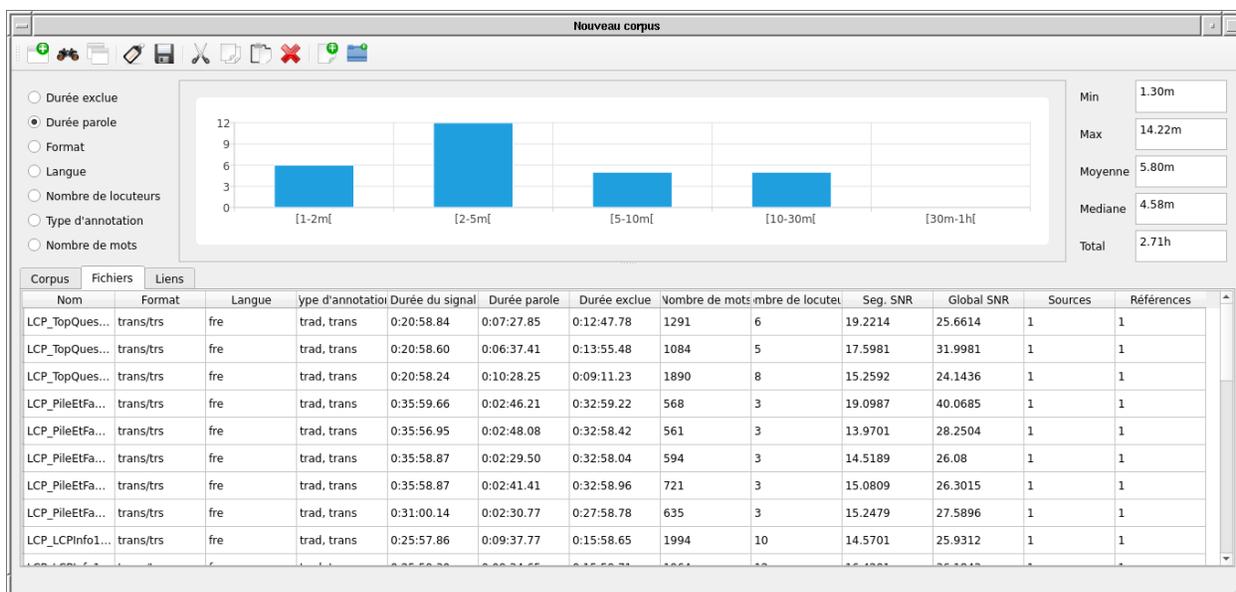


Fig. 2: Ouvrir un corpus en mode exploration fichiers

## 4.2 La recherche

La recherche permet de rechercher des fichiers et des corpus qui sont déjà indexés dans l'espace de stockage. Le détail de l'utilisation de cette fonctionnalité est accessible dans le *Rechercher dans l'espace de stockage*.

## 4.3 Les favoris

L'onglet favoris permet d'accéder aux favoris de l'utilisateur comme le montre la Figure *Exploiter les favoris*.

Elle permet donc d'accéder aux corpus que l'utilisateur a tagué comme étant ses favoris. Lorsque l'utilisateur clique sur un corpus favori, une fenêtre d'exploration de corpus s'ouvre et l'utilisateur peut explorer le corpus.

L'ajout aux favoris s'effectue selon la procédure suivante *Ajouter un corpus aux favoris*.

## 4.4 Les historiques

L'onglet historique permet d'accéder aux derniers corpus et dernières requêtes effectués par l'utilisateur. Cette page est divisée en deux parties comme l'illustre *Exploiter les historiques*.

La première partie qui se situe à gauche de l'écran permet d'accéder aux dernières recherches effectuées par l'utilisateur. Quand l'utilisateur clique sur une requête, le tableau de recherche apparaît configuré selon cette requête.

La seconde partie qui se situe à droite de l'écran permet d'accéder aux derniers corpus créés par l'utilisateur. Lorsque l'utilisateur clique sur un corpus, une fenêtre d'exploration de corpus s'ouvre et l'utilisateur peut explorer le corpus.

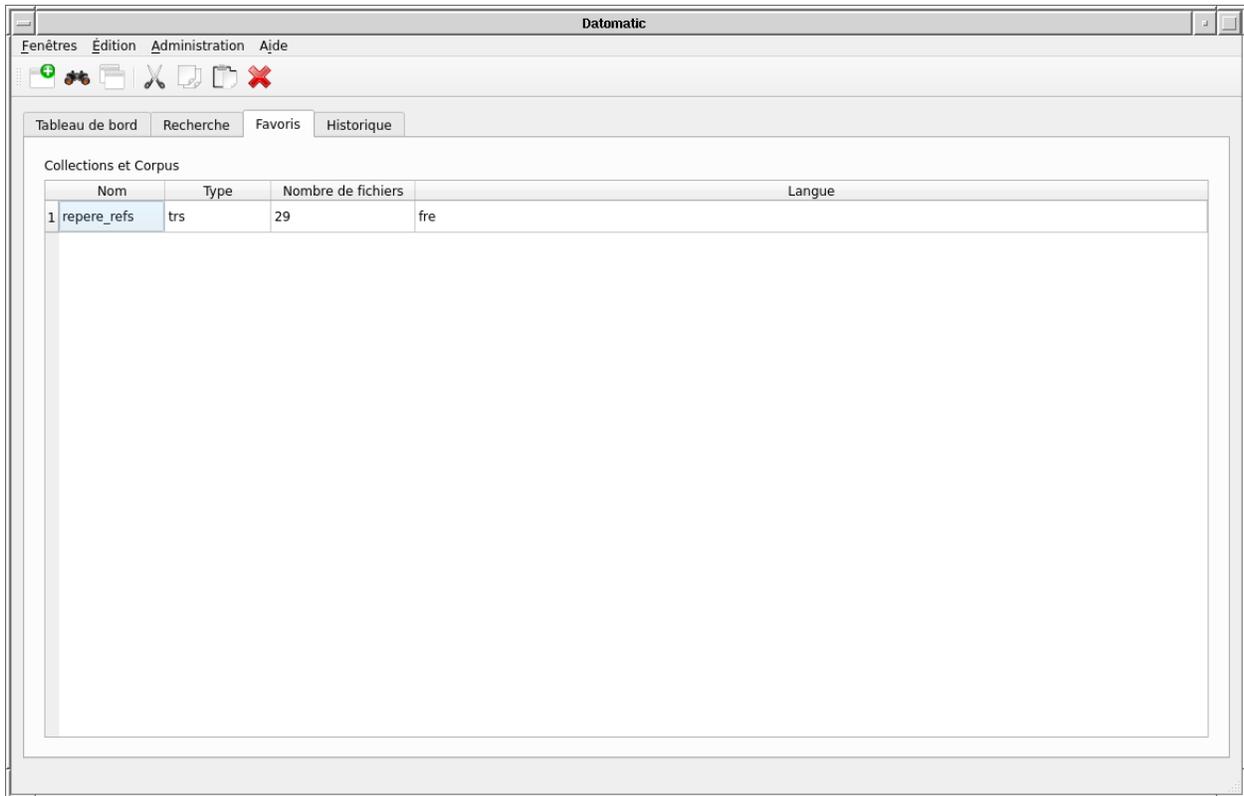


Fig. 3: Exploiter les favoris

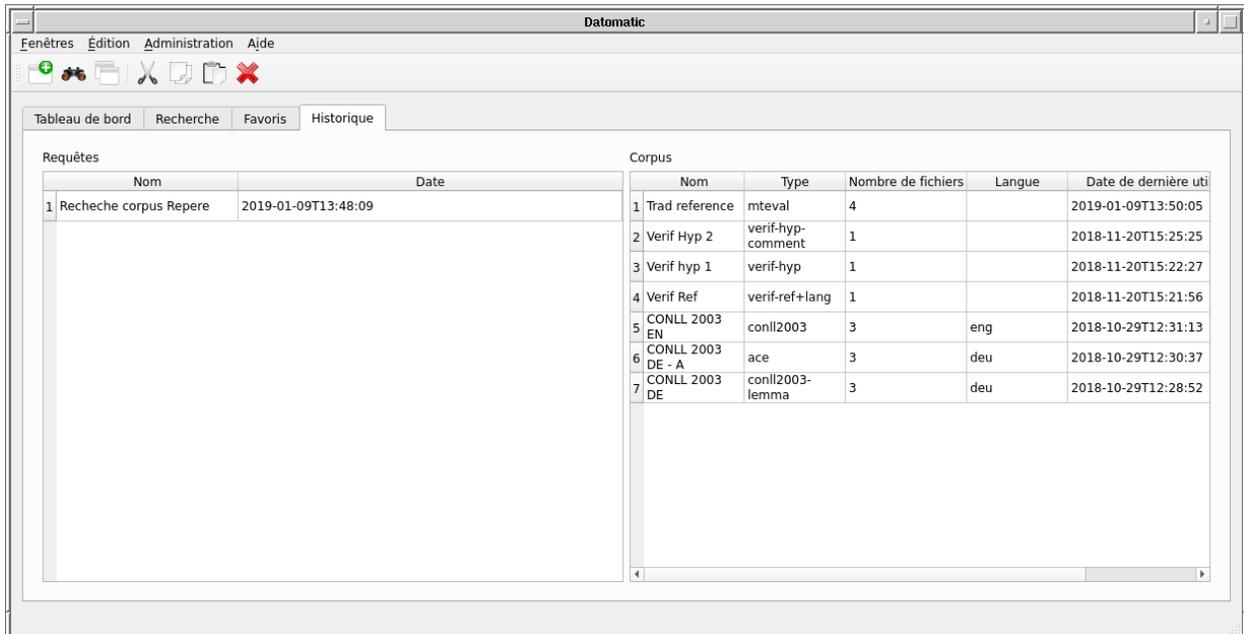


Fig. 4: Exploiter les historiques

## CRÉATION DE CORPUS

### 5.1 Importer de nouveaux fichiers

DATOMATIC permet d'importer des données de type textuel, audio, image et vidéo. Le détail des formats traités est indiqué ci-après.

Pour importer des données, il faut d'abord ouvrir une fenêtre de création de nouveau corpus en cliquant sur l'icône suivante :



La fenêtre *Interface de création d'un nouveau corpus* apparaît alors.

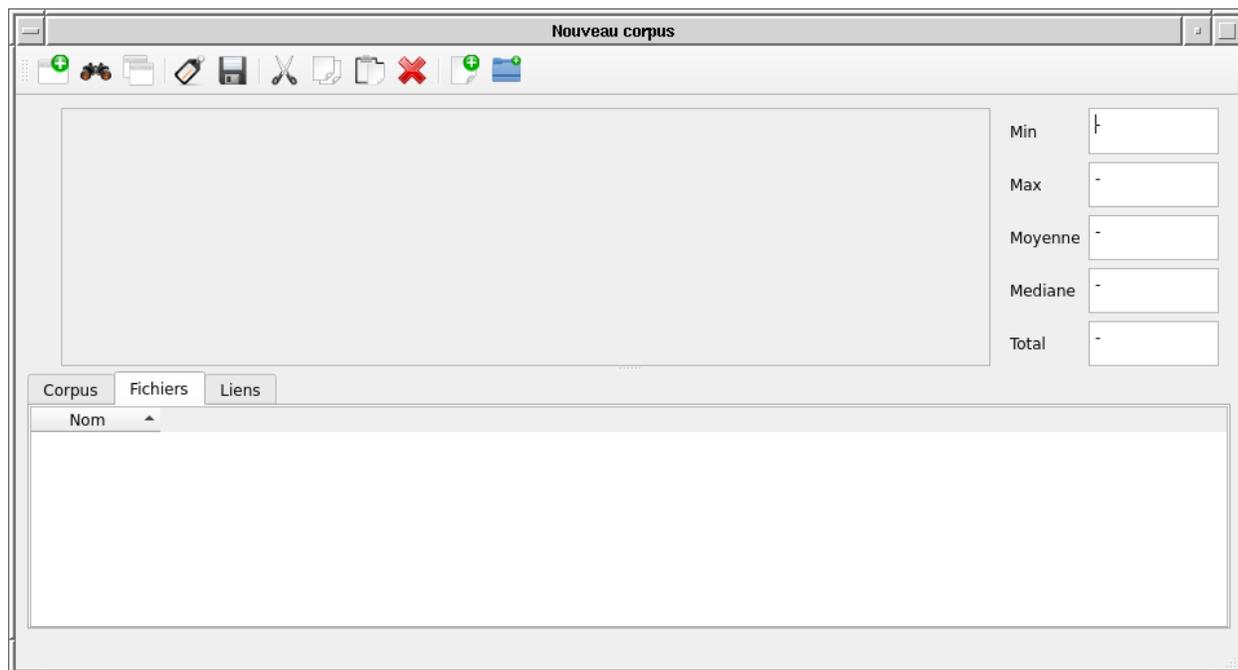


Fig. 1: Interface de création d'un nouveau corpus

Cette interface se divise en 2 parties.

La partie haute de l'interface permet de visualiser, une fois des données importées, des distributions de données au niveau du fichier selon les critères suivants :

- Les types de fichiers
- La langue
- Le nombre de mots
- La durée de signal de parole
- La durée de signal exclue
- La durée de signal

L'information est sélectionnée à l'aide d'un bouton radio sur la gauche de l'interface.

La partie basse permet de visualiser le détail des données. Elle se compose de 3 onglets

- L'onglet corpus permet de visualiser les corpus présents dans les données explorées
- L'onglet fichier permet de visualiser l'ensemble de fichiers importés
- L'onglet lien permet de visualiser les liens entre fichiers (la manipulation de cet onglet est décrit en *Nommer un corpus*)

Il est recommandé de se positionner dans l'onglet fichiers lorsque l'on souhaite importer des nouvelles données.

L'importation des données s'effectue soit au niveau répertoire en cliquant sur l'icône suivante dans la barre de menus



soit au niveau fichier en cliquant sur l'icône suivante dans la barre de menu



L'interface d'exploration de fichier apparaît alors. Si l'utilisateur est en importation fichier, il doit sélectionner les fichiers correspondants, si il est en importation répertoire, il doit sélectionner le répertoire correspondant.

L'importation commence. La plateforme calcule alors selon les types de fichiers les caractéristiques décrites ci-après. Une fois l'ensemble des caractéristiques calculées, l'utilisateur doit saisir certaines informations via l'interface de

paramétrage illustrée par *Paramétrage de l'importation*. Ces informations dépendent de la nature des fichiers importés. Le détail est donné ci-après selon les types de fichiers. La structure arborescente est utilisée pour permettre de propager une valeur à tout un ensemble de fichiers. Par exemple écrire “misc” comme nom de système sur la ligne “xml” et presser enter permet de propager cette valeur à tous les fichiers de ce sous-répertoire.

The screenshot shows the 'Paramétrisation du corpus' dialog box. It features a tree view on the left with folders like 'ctm' and 'xml'. The main area is a table with columns: 'Fichiers', 'systeme', 'version', 'langage', and 'commentaires'. Each row represents a file or folder and contains input fields for 'Nom système', 'Version système', 'Langues' (a dropdown menu), and 'Commentaires'. The 'xml' folder is expanded, showing several XML files with 'misc' entered in the 'Nom système' field. At the bottom, there are buttons for 'Confirmer changements' and 'Annuler changements'.

Fig. 2: Paramétrage de l'importation

Pour certains types de fichiers, en particulier les fichiers texte mais aussi certains formats entités nommées ou OCR, il n'est pas possible de déterminer de par son contenu quel est le rôle d'un annotation entre source, référence ou hypothèse. Il faut donc le choisir comme sur le screenshot *Paramétrage de l'importation, choix de rôle* et les champs nom de système, version et nom de référence s'activent ou se désactivent en fonction.

This screenshot shows the 'Paramétrisation du corpus' dialog box with a different configuration. The 'Fichiers' column now lists individual files like 'article4217.html.txt'. The 'type\_all' column has a dropdown menu with options 'hypothese', 'source', and 'reference'. The 'systeme' column has a dropdown menu with 'misc' selected. The 'version' column has an input field. The 'reference' column has a 'Nom Référence' input field. The 'langage' column has a dropdown menu with 'fre' selected. The 'commentaires' column has an input field. At the bottom, there are buttons for 'Confirmer changements' and 'Annuler changements'.

Fig. 3: Paramétrage de l'importation, choix de rôle

Le champ “Nom Fichier” sert à faire le lien entre référence et hypothèse. Le champ “Nom référence” sert à reconnaître les différentes références sur un même fichier dans le cadre de la traduction.

Annuler les changements annuler l'import du ou des fichiers.

Une fois les données importées, les données sont accessibles comme le montre la Figure *Importation des données*. Les histogrammes de la partie haute de l'interface se mettent automatiquement à jour.

Il est alors possible de trier les données importées à l'aide des entêtes des caractéristiques des fichiers. Il est également possible de filtrer des colonnes à l'aide du menu contextuel accessible par un clic-droit en sélectionnant la fonction Filtrer (cf. Interface *Filtrer les colonnes : Etape 1* et *Filtrer les colonnes : Etape 2*). Il est possible pour chaque colonne de sélectionner si elle doit apparaître toujours (coché), jamais (décoché), ou, le choix par défaut, si au moins un fichier a une information pour la colonne (grisé).

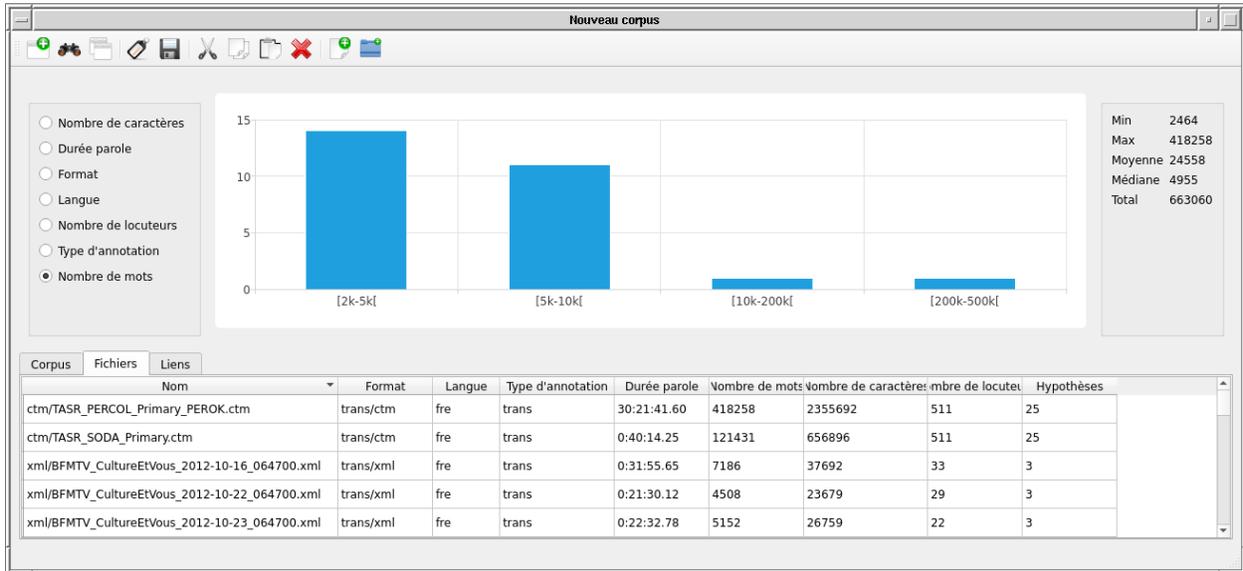


Fig. 4: Importation des données

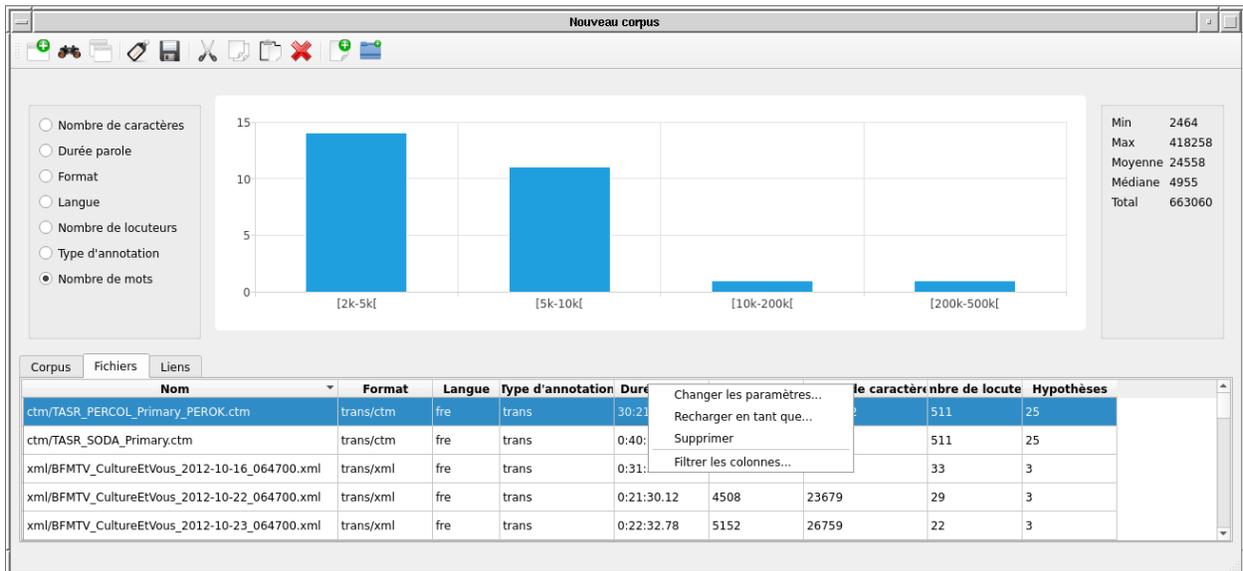


Fig. 5: Filtrer les colonnes : Etape 1

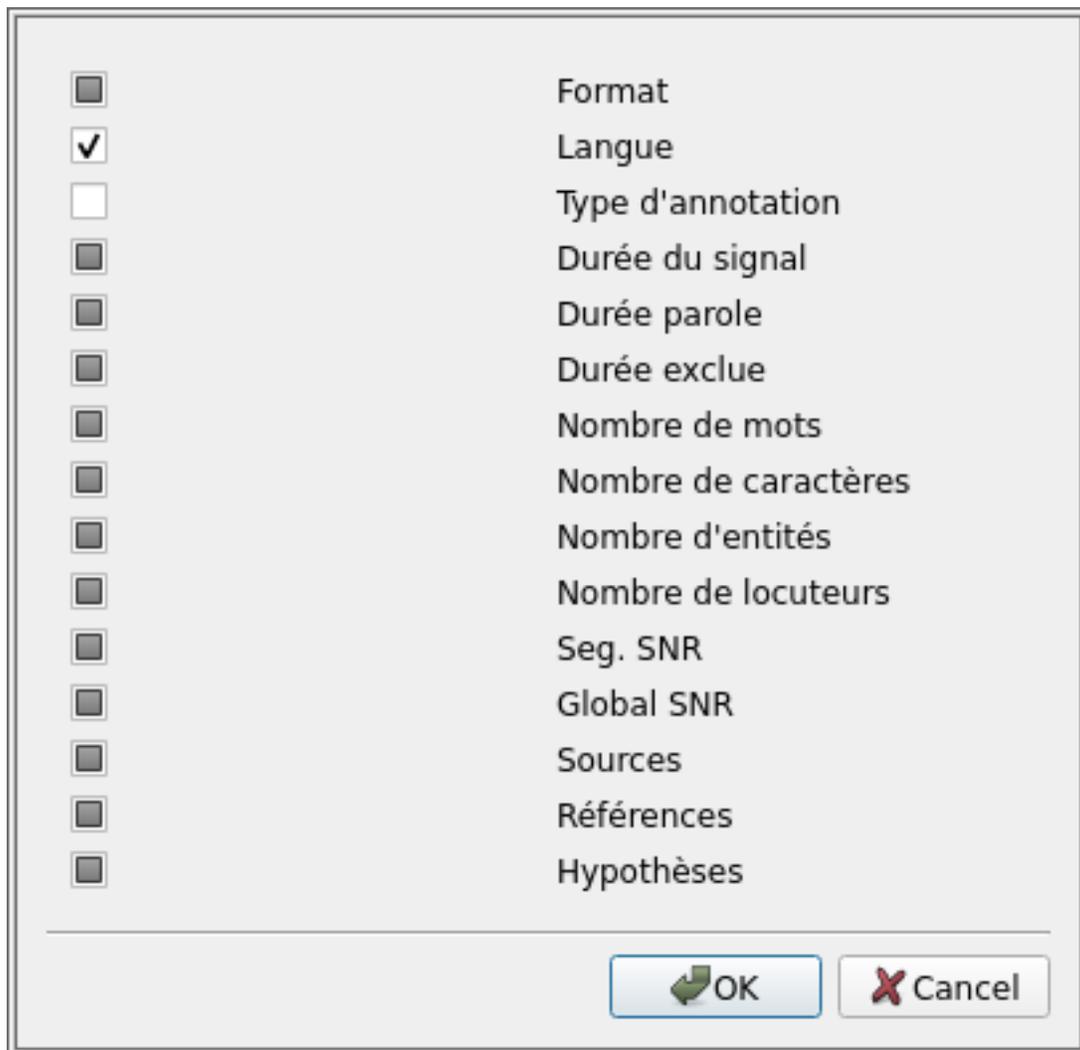


Fig. 6: Filtrer les colonnes : Etape 2

## 5.2 Fichiers média

### 5.2.1 Formats sources traités

Les formats acceptés sont tous ceux supportés par ffmpeg, qui sont juste trop nombreux pour être listés.

### 5.2.2 Caractéristiques calculées sur les fichiers sources

DATOMATIC transmet les informations fournies par ffmpeg et, de plus, calcule le rapport signal/bruit segmental et global pour les fichiers contenant de l'audio.

### 5.2.3 Formats annotations de références traités

DATOMATIC gère tous les formats d'annotation de la suite matics. À l'heure actuelle cela inclus: \* Transcription: trs, stm, tdf, ctm, xml pour asr \* Entités nommées, lemmes et tokénisation: CoNLL, Quaero, ACE, GermEval, ApacheNE, StanfordNER, MUC, format custom \* Vérification du locuteur: formats référence et hypothèse custom \* Traduction: MTEval

Du plus le format texte est en grande partie considéré comme un format de traduction.

### 5.2.4 Caractéristiques calculées sur les annotations

DATOMATIC détermine les informations suivantes à partir des fichiers annotation ou des paramètres quand elles sont pertinentes :

- types de tâche possibles
- langue
- durée de signal de parole et exclue
- nombre de locuteurs
- nombres de mots (tokens séparés par des espaces) et de caractères (codepoint unicode)
- nombre d'entités nommées

## 5.3 Changer le format d'annotation

Pour changer le format d'annotation, il faut effectuer un clic droit sur le fichier à modifier et dans le menu contextuel sélectionner la ligne "recharger en tant que" comme illustré par *Changer le format d'un fichier-étape 1*.

La fenêtre *Changer le format d'un fichier-étape 2* apparaît alors et il est possible de modifier le format de fichier.

Si le format proposé n'est pas compatible selon DATOMATIC, un message apparaît pour expliquer pourquoi (Figure *Changer le format d'un fichier-étape 3*).

Il est à noter que le fichier est rechargé du disque pour chaque essai, permettant de le corriger en live si besoin est. Si le nouveau format est compatible la fenêtre de paramétrisation apparaîtra.

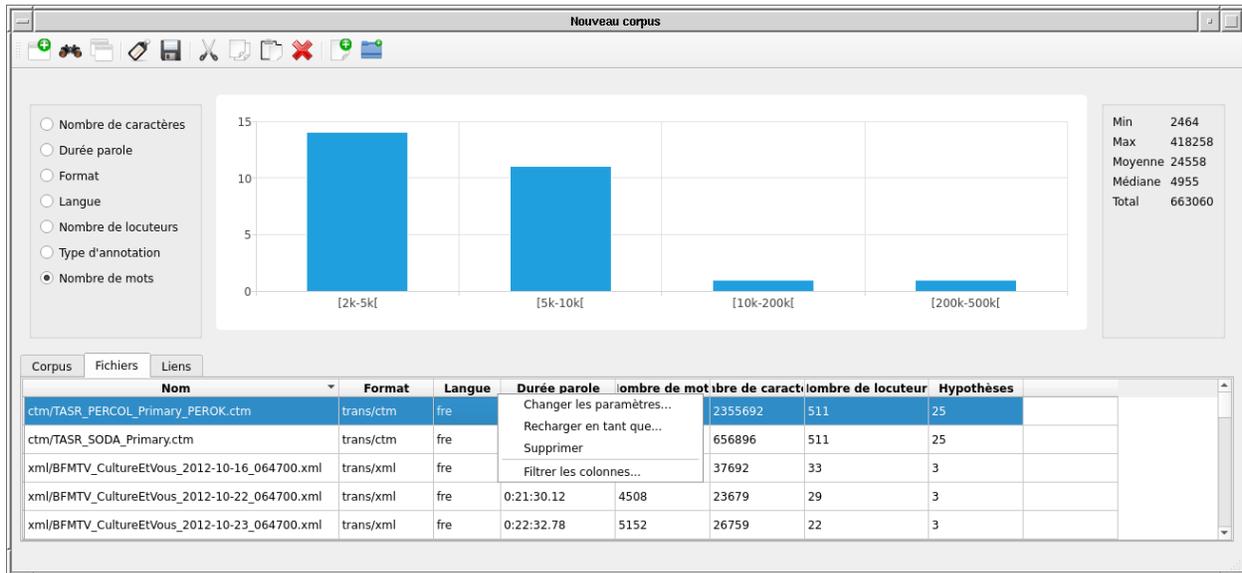


Fig. 7: Changer le format d'un fichier-étape 1

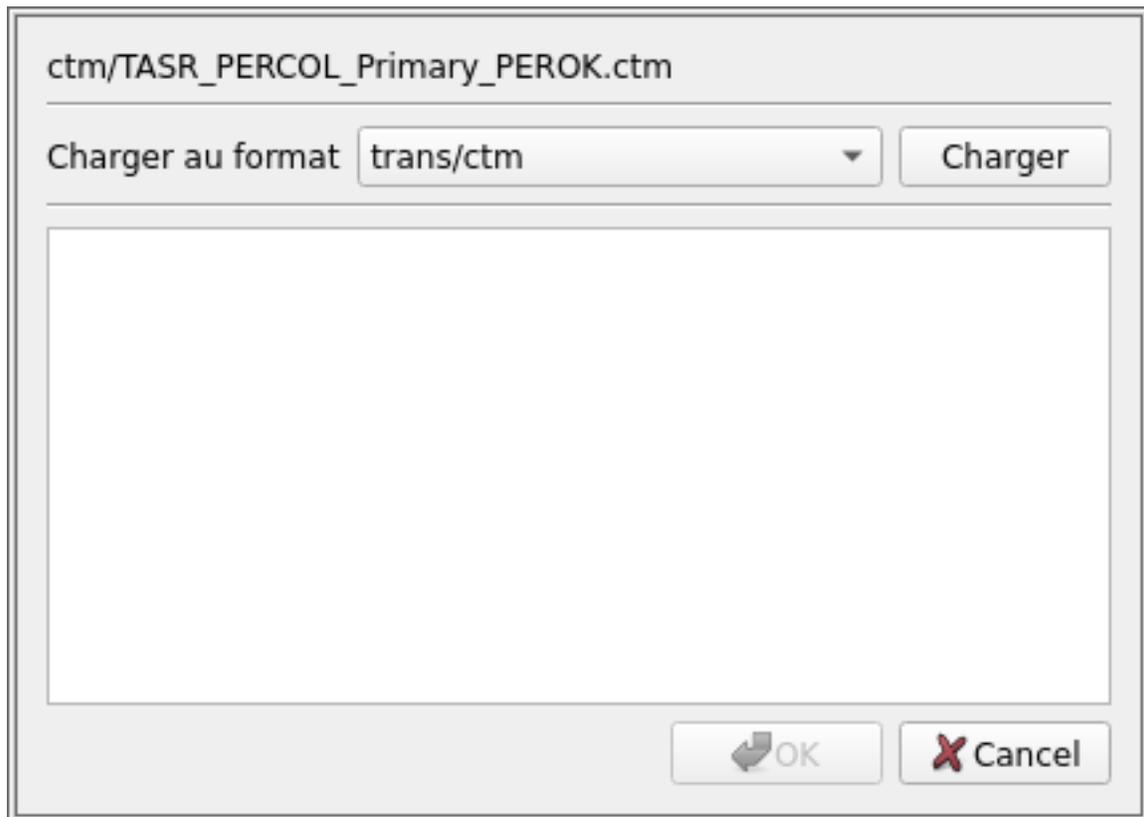


Fig. 8: Changer le format d'un fichier-étape 2

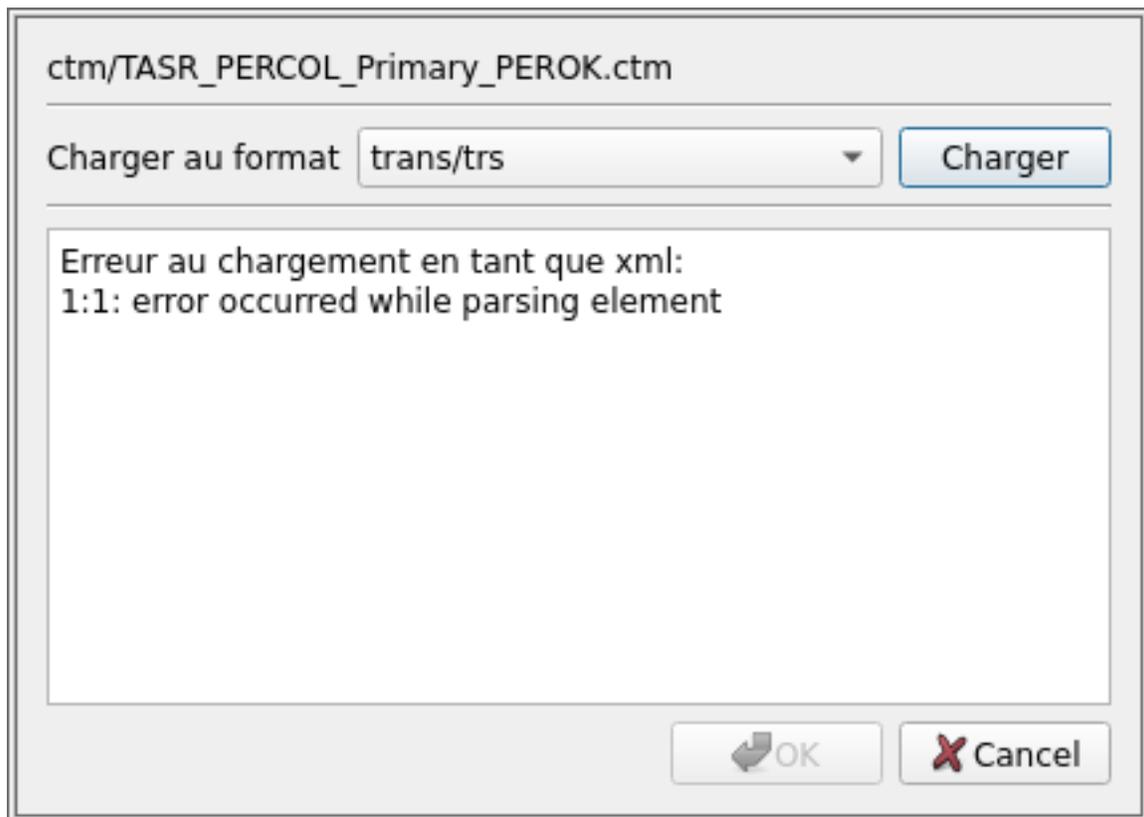


Fig. 9: Changer le format d'un fichier-étape 3

## 5.4 Changer les paramètres

Il est possible de changer les paramètres dans le même menu contextuel. Le fichier n'est pas rechargé du disque en ce cas.

## 5.5 Nommer un corpus

Le nommage d'un corpus se fait par l'icône suivante:



La fenêtre *Fenêtre de choix de nom du corpus* permet alors de le nommer et de choisir sa collection.

A screenshot of a dialog box titled 'Fenêtre de choix de nom du corpus'. It contains two input fields: 'Nom de corpus' with an empty text box, and 'Collection' with a dropdown menu showing 'Nouveau nom...'. At the bottom right, there are two buttons: 'OK' with a green checkmark icon and 'Cancel' with a red X icon.

Fig. 10: Fenêtre de choix de nom du corpus

Les collections existantes sont disponibles dans le drop-down ou en complétion avec tab. Le nom de corpus doit être unique dans la base.

## 5.6 Changer les liens entre fichiers

Les liens entre fichiers s'établissent après nommage du corpus et en particulier du choix de la collection, et peuvent être modifiés à l'aide de l'interface *Modifier les liens grâce au prendre-déplacer* accessible via l'onglet lien.

La modification de lien consiste à l'aide d'un prendre-déplacer d'associer au bon groupe, les données sources, hypothèses et/ou références chargées.

En pratique les liens ne sont pour l'instant pas pris en compte par Evalomatic, faisant perdre l'essentiel de l'intérêt de cette interface. Une réflexion est en cours sur l'évolution du concept.

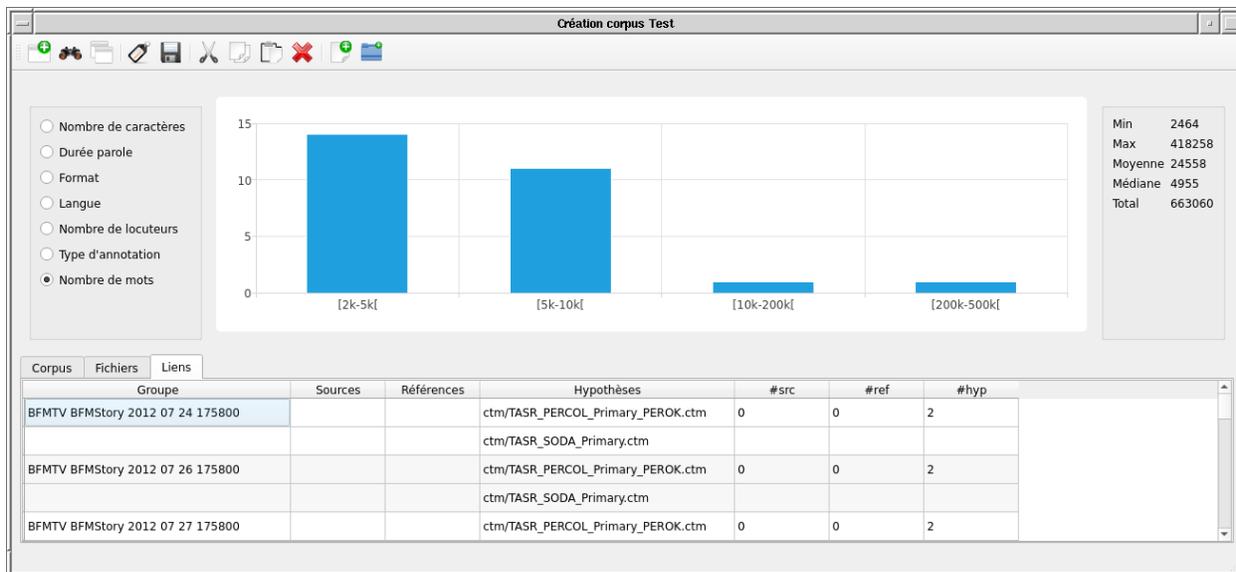


Fig. 11: Modifier les liens grâce au prendre-déplacer

## 5.7 Enregistrer un corpus

Une fois le corpus constitué, l'utilisateur l'enregistre dans DATOMATIC via l'icône suivante :



L'interface *Sauvegarde de corpus* permet alors de choisir le type de corpus et de remplir les champs associés tels que choisis par l'administrateur. Les différents types de corpus possibles sont:

- Corpus source, ne contenant que des données source
- Corpus hypothèse, ne contenant que des hypothèses de système
- Corpus référence, ne contenant que des références
- Corpus transformé, contenant des données source changées d'une manière ou d'une autre
- Corpus updaté, contenant des références changées d'une manière ou d'une autre
- Corpus mixte, pouvant contenir n'importe quoi

Les données sont alors stockées dans la base de données et dans l'espace de stockage. Le corpus créé apparaît dans le tableau de bord et peut être accessible par une recherche.

The screenshot shows a dialog box titled "Sauvegarde d'un corpus". It contains the following fields and values:

Nom du corpus :	Rerepe sources
Collection	Repere
Type de corpus	Corpus source
Date de création Thu Feb 14 2019	
Auteur	galibert
Source	Évaluation repere
Classification	Bleu

At the bottom right, there are two buttons: "Enregistrer Corpus" and "Annuler".

Fig. 12: Sauvegarde de corpus



## RECHERCHER DANS L'ESPACE DE STOCKAGE

La recherche s'effectue en sélectionnant l'onglet rechercher dans l'espace de travail ou bien en sélectionnant l'icône suivante :



La fenêtre illustrée par *Interface de recherche initiale* apparaît alors.

Une recherche peut être effectuée sur l'ensemble des critères saisis ou calculés à l'aide de DATOMATIC. L'utilisateur doit d'abord choisir si il veut rechercher des corpus ou des fichiers, puis peut accéder aux critères de recherche est accessible en cliquant sur l'icône suivante à côté des termes *ajouter un critère*:



L'utilisateur saisit alors son critère de recherche et le remplit via l'interface *Sélection du critère de recherche*.

Il existe 5 grandes familles de critères de recherche qui peuvent se résumer de la manière suivante :

- critère fondé sur une variable numérique
- critère fondé sur une date
- critère fondé sur une variable de type texte
- critère fondé sur une liste de choix
- critère fondé sur une valeur booléenne

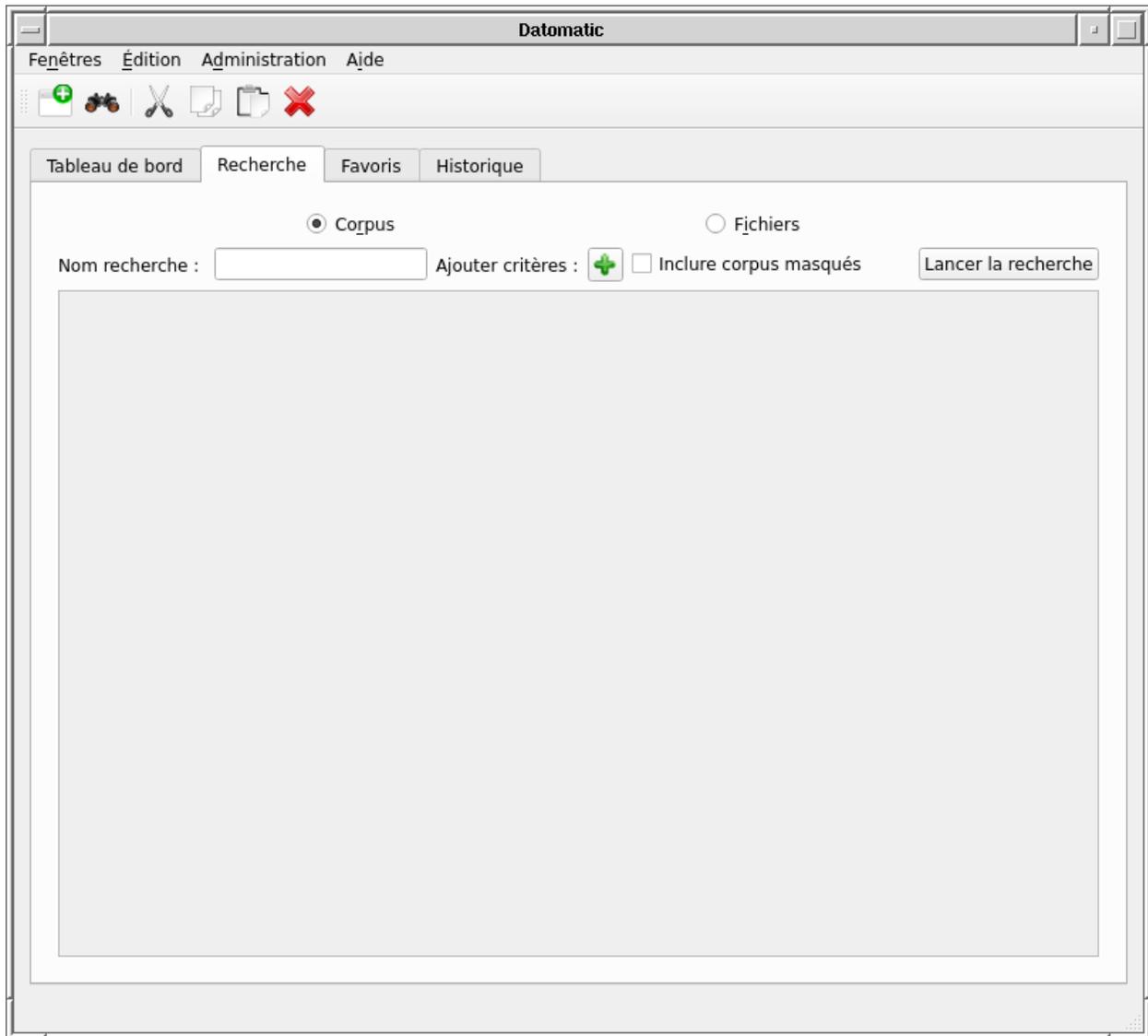


Fig. 1: Interface de recherche initiale

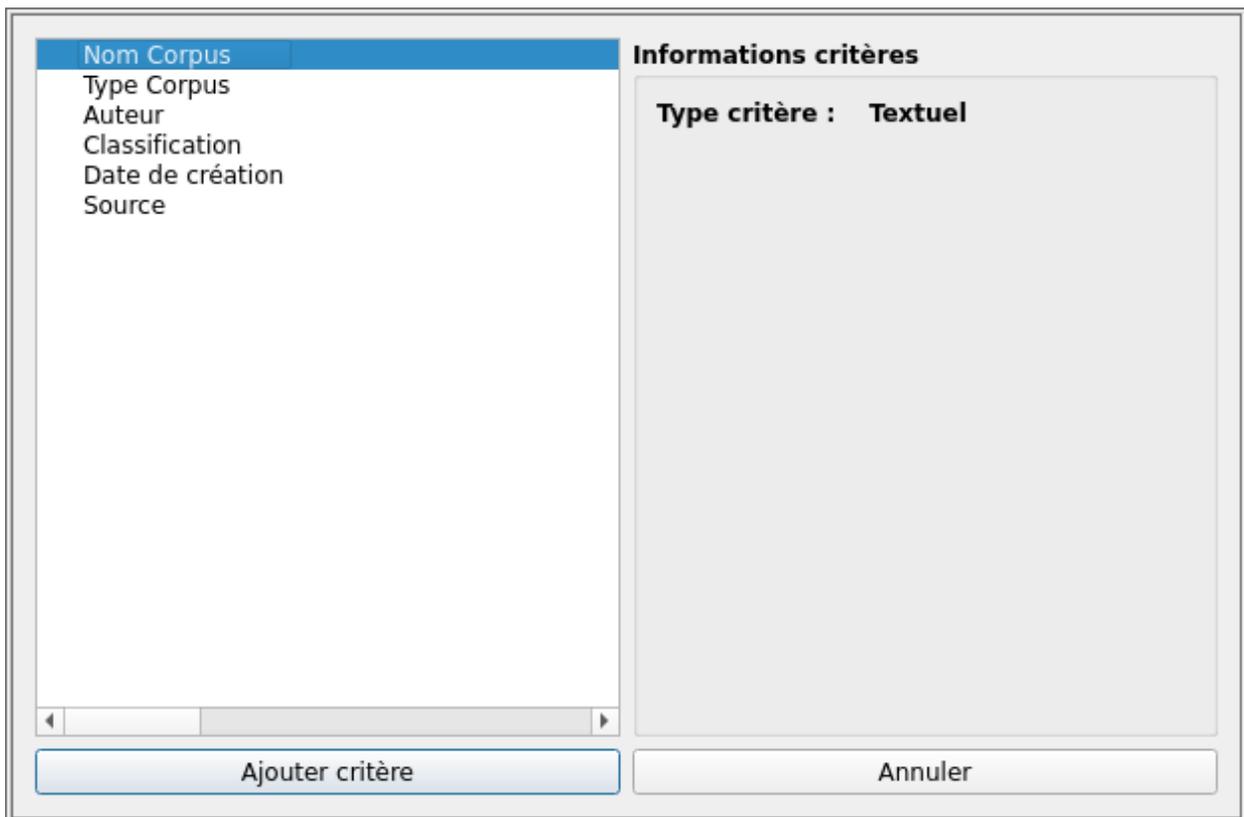


Fig. 2: Sélection du critère de recherche

L'utilisateur peut formuler sa recherche avec autant de critères qu'il le souhaite. Les critères sont liés par un opérateur logique ET. Les champs entrés à la sauvegarde des corpus sont disponibles dans les critères de recherche.

Pour chaque famille de critères, une interface spécifique est proposée.

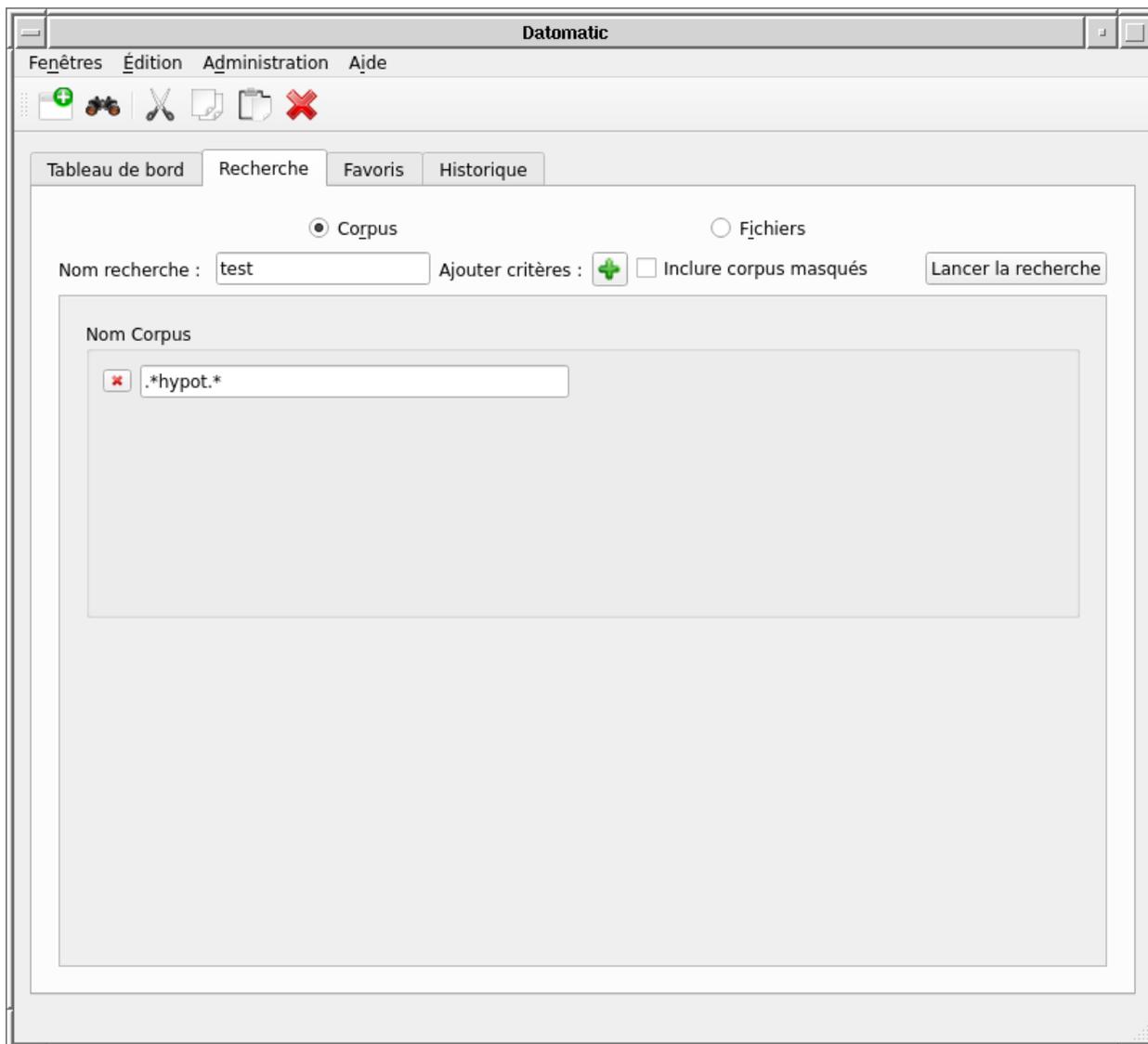


Fig. 3: Critère sur une variable textuelle

En cochant la case *Inclure les corpus masqués* il est possible d'effectuer des recherches sur les corpus qui sont en cours de suppression en plus des corpus normaux.

Pour lancer la recherche, il suffit la nommer puis de cliquer sur *lancer la recherche*. La fenêtre obtenue *Fenêtre de résultat de recherche* permet immédiatement de créer un nouveau corpus à partir du résultat.

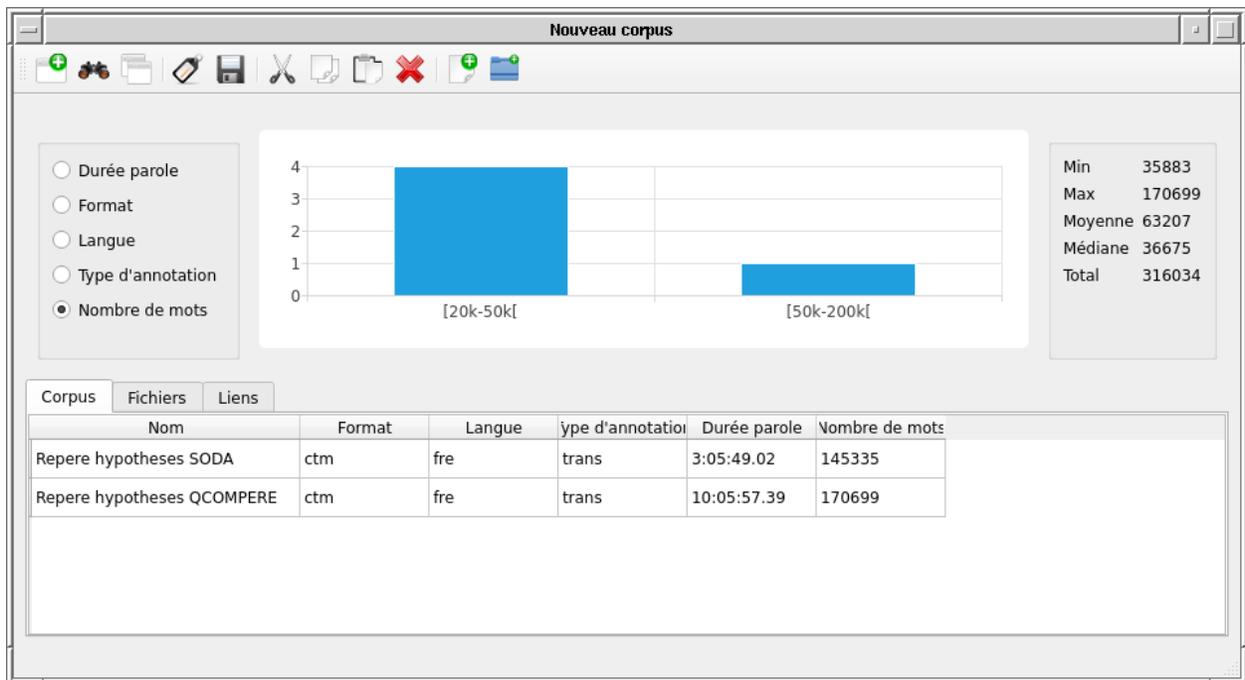


Fig. 4: Fenêtre de résultat de recherche



## AJOUTER UN CORPUS AUX FAVORIS

Ajouter un corpus aux favoris peut se faire soit depuis le tableau de bord soit depuis une fenêtre d'exploration de corpus.

### 7.1 Depuis le tableau de bord

Depuis le tableau de bord, il suffit à l'aide d'un clic droit de sélectionner **ajouter aux favoris** comme l'illustre l'interface *Créer un favori à partir du tableau de bord*

### 7.2 Depuis la fenêtre d'exploration de corpus

Depuis la fenêtre d'exploration de corpus, dans l'onglet corpus, il suffit à l'aide d'un clic droit de sélectionner **ajouter aux favoris** comme l'illustre l'interface *Créer un favori à partir de la fenêtre d'exploration de corpus*

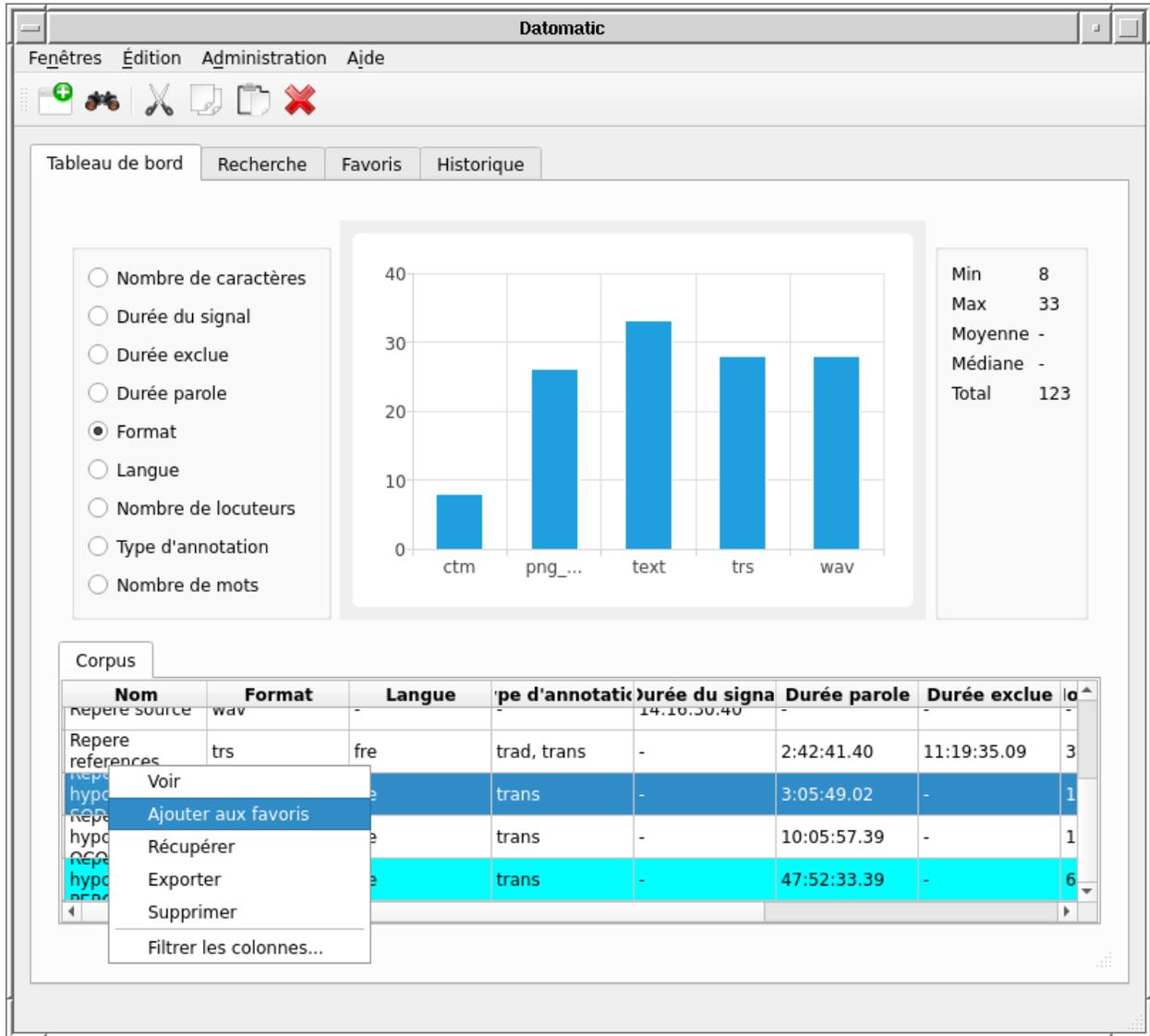


Fig. 1: Créer un favori à partir du tableau de bord

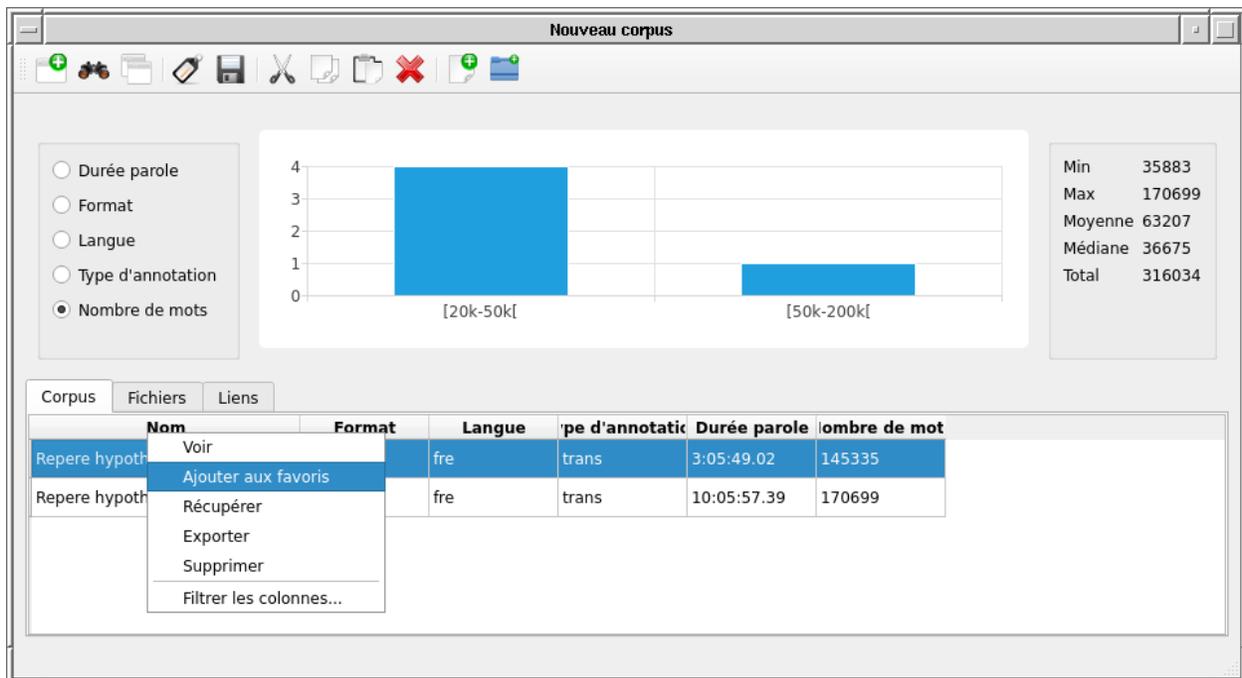


Fig. 2: Créer un favori à partir de la fenêtre d'exploration de corpus



## VISUALISER LES FICHIERS AUDIO

Pour visualiser un fichier audio, faire un clic droit sur le fichier et dans le menu contextuel choisir visualiser (cf. *Visualiser un fichier audio*).

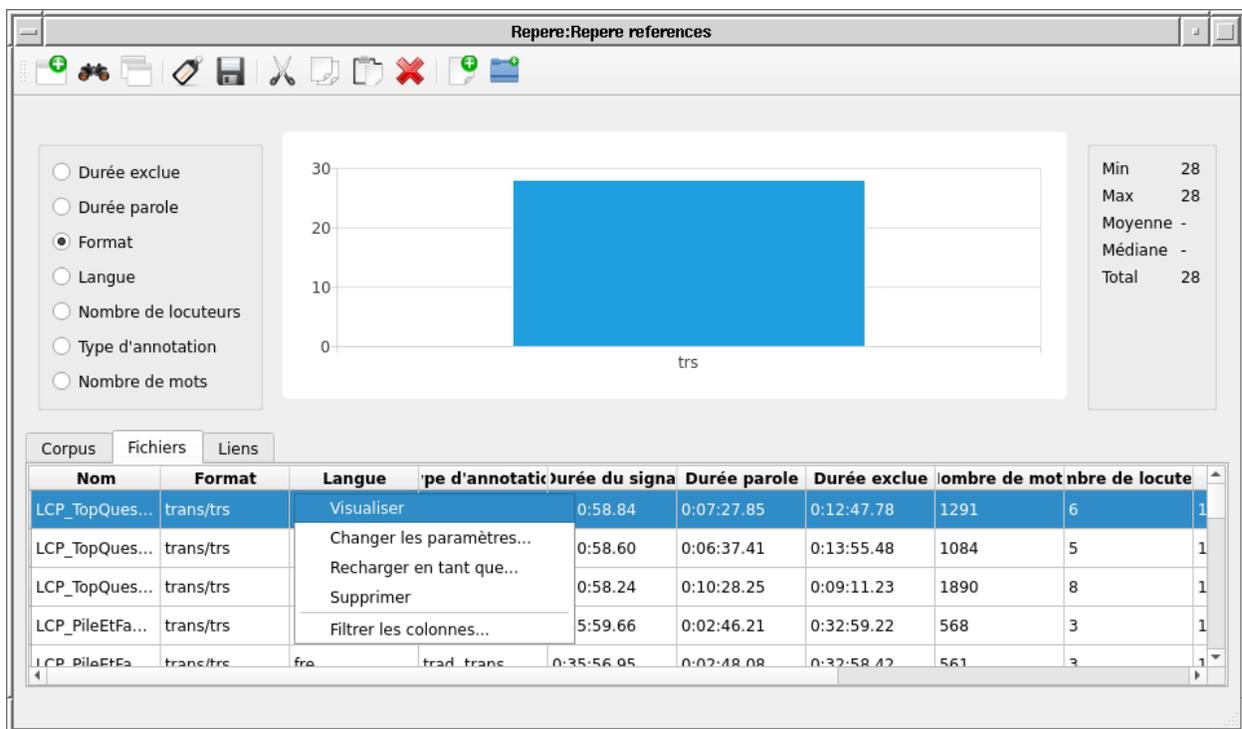


Fig. 1: Visualiser un fichier audio

L'interface illustrée par la figure *Interface de visualisation d'un fichier audio* apparaît alors.

Cette interface de visualisation propose les 2 modes suivants :

- *Mode tour de parole* qui permet d'écouter et de visualiser les tours de parole un à un.
- *Mode temps* qui permet d'écouter et de visualiser un échantillon temporel du signal.

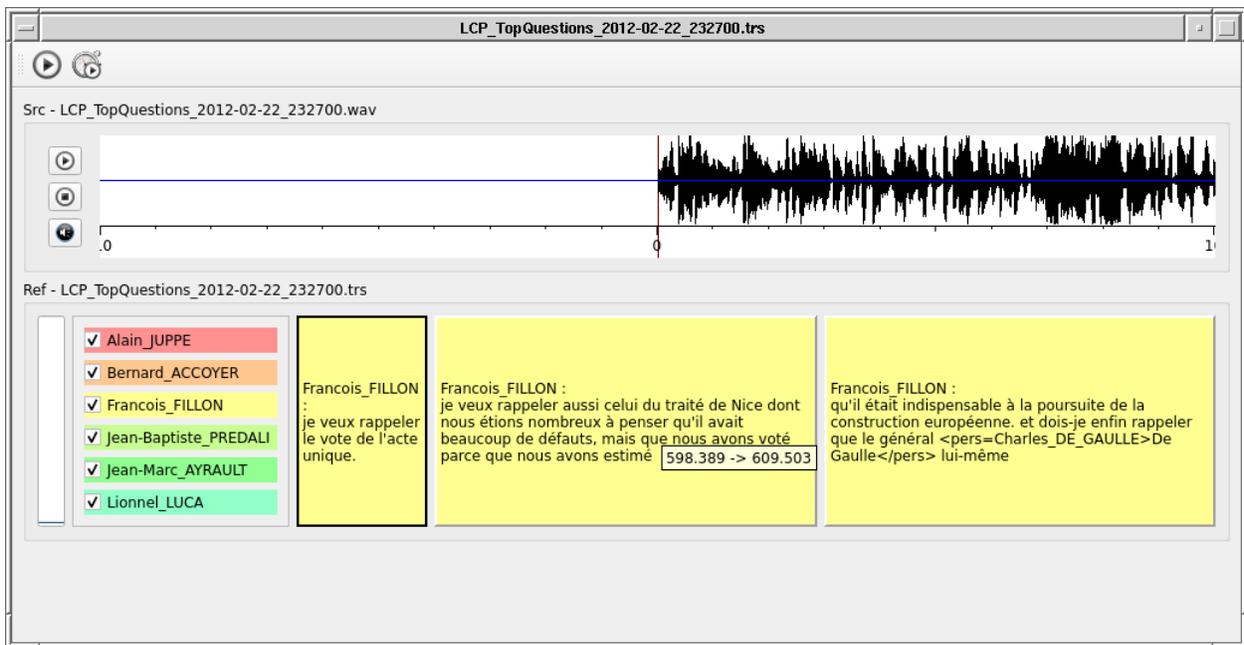


Fig. 2: Interface de visualisation d'un fichier audio

## **DOWNLOAD DE CORPUS**

Les deux modes de download de corpus s'effectue en faisant un clic-droit sur ce corpus en mode exploration corpus.

### **9.1 Récupération de corpus**

La récupération consiste à reconstituer localement l'arborescence du corpus pour l'utiliser ensuite. En particulier Evalomatic voit les corpus qui ont été exportés. L'utilisateur doit choisir le répertoire dans lequel le corpus sera écrit et remplir les champs choisis par l'administrateur.

### **9.2 Exportation de corpus**

L'exportation de corpus crée une archive zip avec le contenu du corpus dans un objectif de diffusion. L'utilisateur remplit les champs et choisit le nom du fichier archive. Exporter un corpus interdit sa suppression.



## SUPPRIMER UN CORPUS

La suppression d'un corpus se fait en 3 étapes.

1. l'utilisateur masque un corpus pour en demander la suppression.
2. L'administrateur autorise la suppression de la base de données.
3. L'administrateur supprime les données dans l'espace de stockage.

### 10.1 L'utilisateur masque un corpus

Pour masquer un corpus, l'utilisateur se met en mode corpus dans l'interface de visualisation d'un corpus puis il effectue un clic droit sur le corpus considéré et sélectionne *supprimer* comme le montre l'interface *Masquer un corpus par menu contextuel*

Le corpus est à présent masqué (couleur bleue pour l'utilisateur) comme le montre l'interface *Un corpus masqué apparaît en bleu pour l'utilisateur*. Il n'est plus visible des autres utilisateurs de DATOMATIC. Il n'est toutefois pas supprimé de la base de données tant qu'un administrateur n'a pas supprimé définitivement le corpus.

### 10.2 L'administrateur supprime le corpus de la base de données

Pour supprimer le corpus de la base de données, l'administrateur se rend dans le menu **Administration/supprimer corpus** comme le montre l'interface *La suppression des corpus se fait en mode administrateur*.

L'interface *Interface administrateur de suppression de corpus* apparaît alors.

Les corpus à supprimer peuvent alors être sélectionnés.

Une fois supprimé le corpus n'est alors plus visible pour l'utilisateur comme l'illustre l'interface *Le corpus n'est plus visible dans la base de données*.

### 10.3 L'administrateur supprime le corpus dans l'espace de stockage

La suppression de données dans l'espace de stockage n'est pas gérée en tant que tel par DATOMATIC. L'administrateur doit supprimer le répertoire contenant le corpus selon sa responsabilité.

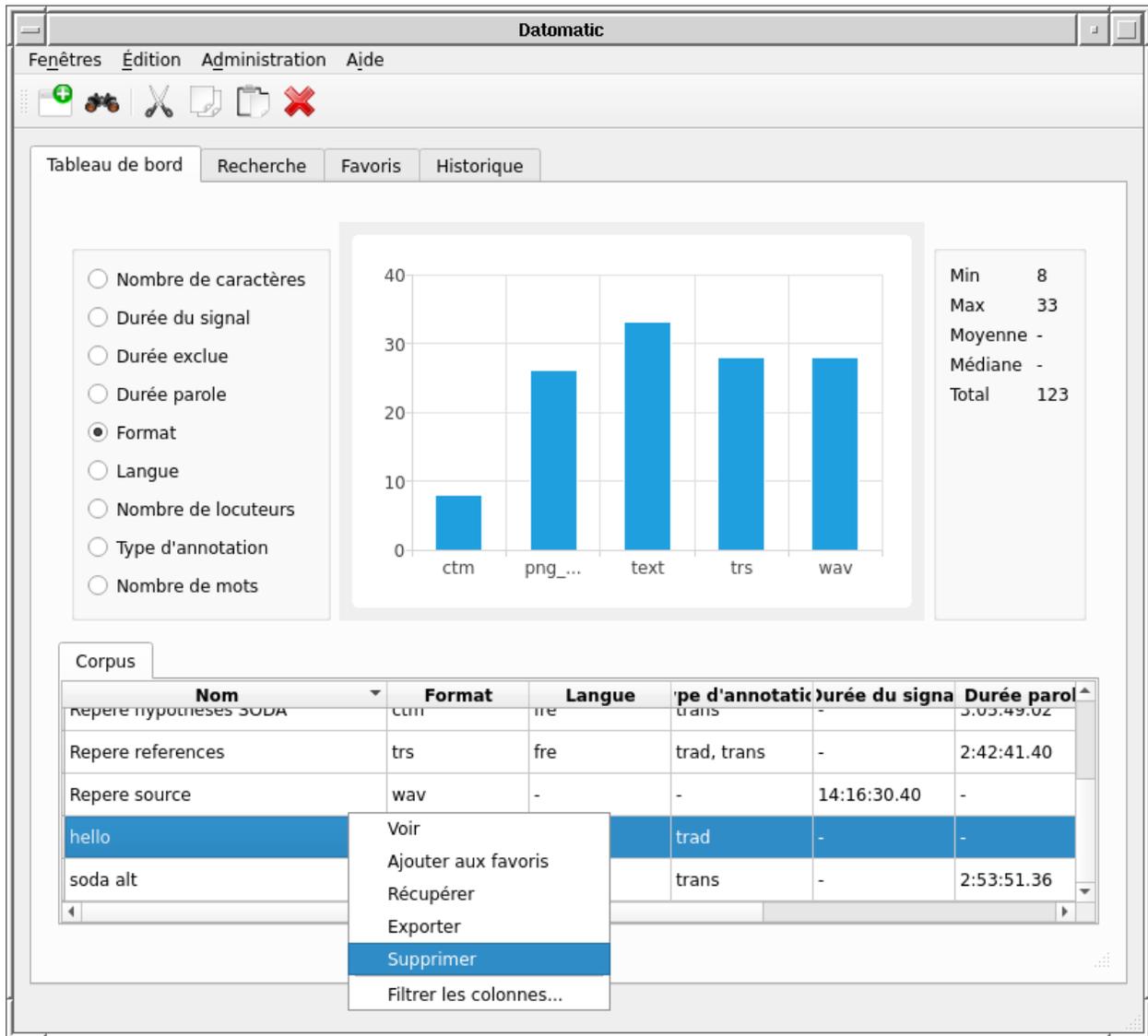


Fig. 1: Masquer un corpus par menu contextuel

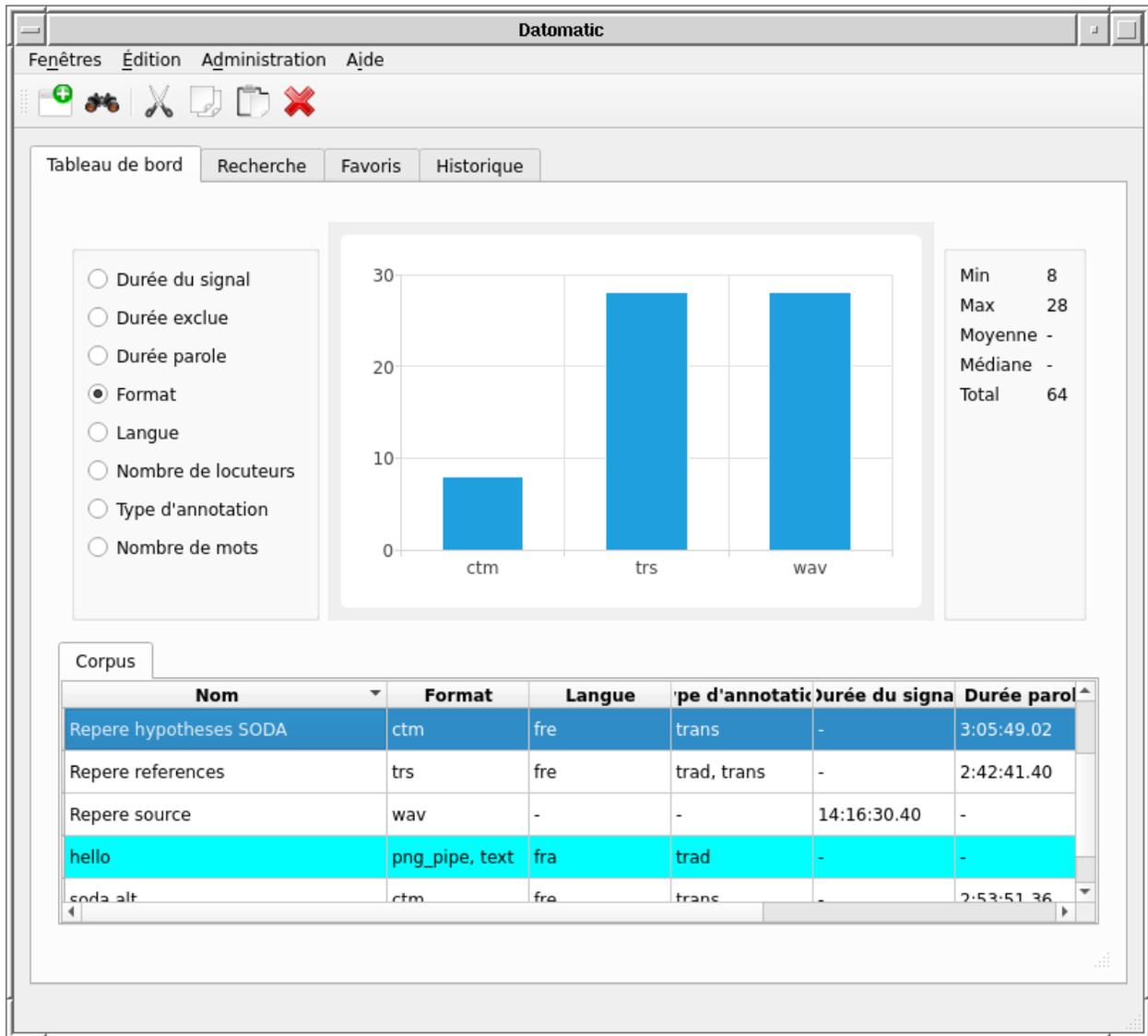


Fig. 2: Un corpus masqué apparaît en bleu pour l'utilisateur

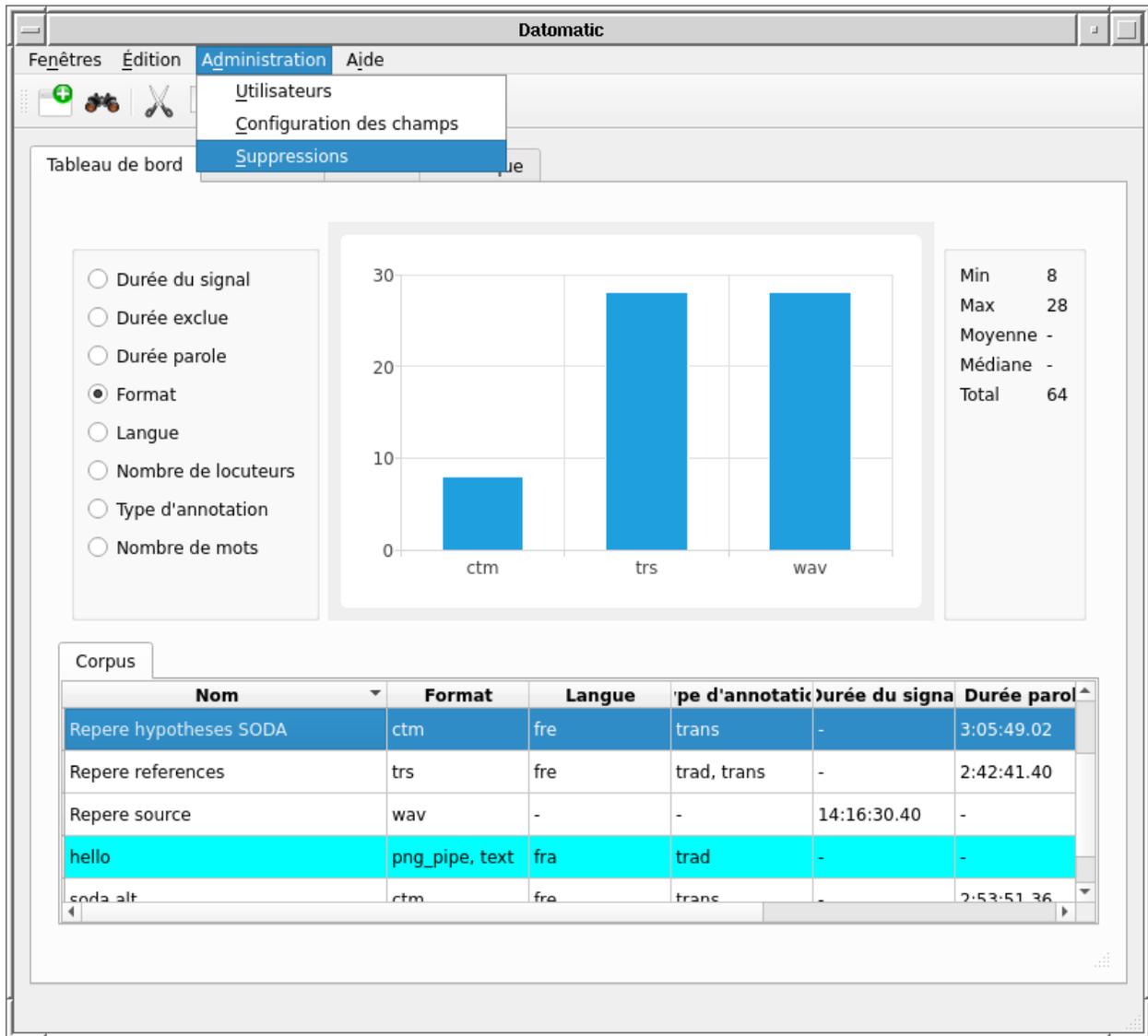


Fig. 3: La suppression des corpus se fait en mode administrateur

	Nom	Type	Langue	Annotation	Utilisateur	Créateur	Exporté	
1	hypotheses	ctm	fre	trans	galibert		OUI	<input type="checkbox"/>
2	hello	Multimodal	fra	trad	galibert		NON	<input checked="" type="checkbox"/>

Valider suppressions      Annuler suppressions

Fig. 4: Interface administrateur de suppression de corpus

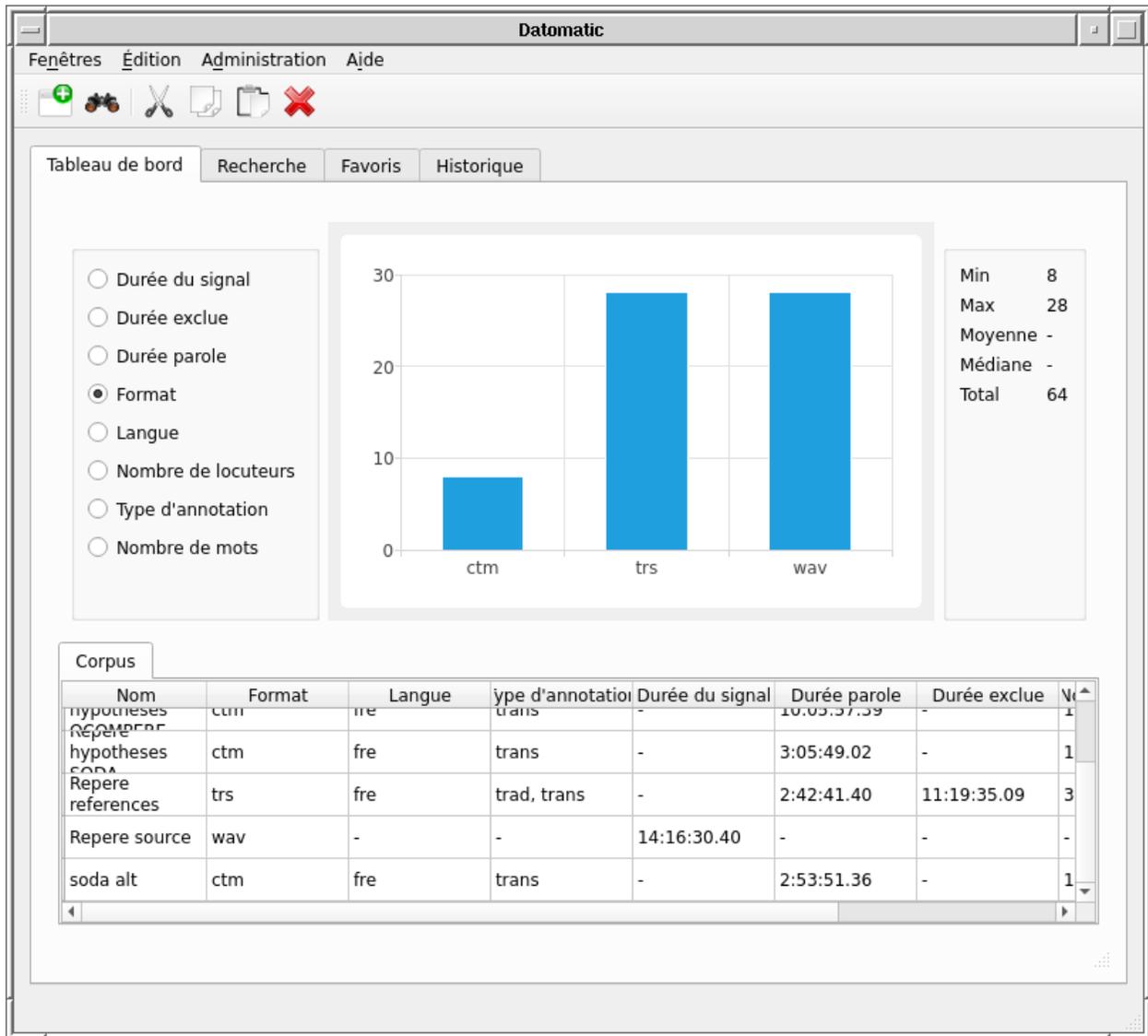


Fig. 5: Le corpus n'est plus visible dans la base de données

## ADMINISTRATION DE DATOMATIC

### 11.1 Mettre à jour les champs des formulaires

Les formulaires de saisie des corpus peuvent être mis à jour à l'aide d'une interface dédiée. Cette interface est accessible en cliquant sur Administration/gérer les formulaires comme illustré par la figure *Accéder à la définition des formulaires*.

Le formulaire illustré par l'Interface *Formulaire de définition des champs* apparaît alors.

Il existe 4 types de champs de formulaire qui peuvent être inclus :

- Un champ **texte libre** qui permet principalement d'indiquer un nom, un commentaire, un identifiant. Ce champ texte est libre pour l'utilisateur. La colonne Valideur/Options permet de mettre une expression régulière de validation de l'entrée.
- Un champ de **liste de choix** permet de choisir une option parmi une liste fermée. La colonne Valideur/Options contient l'ensemble des options possibles séparées par des points-virgules.
- Un champ **date** permet d'ajouter automatiquement la date du jour.
- Un champ **utilisateur** permet d'ajouter automatiquement le nom de l'utilisateur courant.

Ajout et suppression de champ se fait par le menu contextuel sur le bouton de droite, cf. *Formulaire de définition des champs : ajout d'un champ ou suppression d'un existant*.

L'administrateur peut ainsi choisir les champs à remplir à la sauvegarde de chaque type de corpus, à l'exportation, à la récupération et quand on crée un utilisateur.

L'import/export permet de sauvegarder la définition des champs dans un fichier json ré-importable ultérieurement.

### 11.2 Supprimer des Corpus et des fichiers de l'espace de stockage

La suppression des corpus est décrite dans la partie *Supprimer un corpus*

### 11.3 Gérer les utilisateurs de DATOMATIC

Il est possible d'ajouter et de supprimer des utilisateurs via l'interface administrateur de DATOMATIC. Pour cela, saisir dans le menu Administration/utilisateurs comme illustré par la Figure *Accès au formulaire de gestion des utilisateurs*

L'interface illustrée par la Figure *Ajouter et supprimer des utilisateurs* apparaît alors.

L'administrateur peut ajouter un utilisateur en cliquant sur ajouter. Il saisit alors les informations à renseigner et clique sur valider.

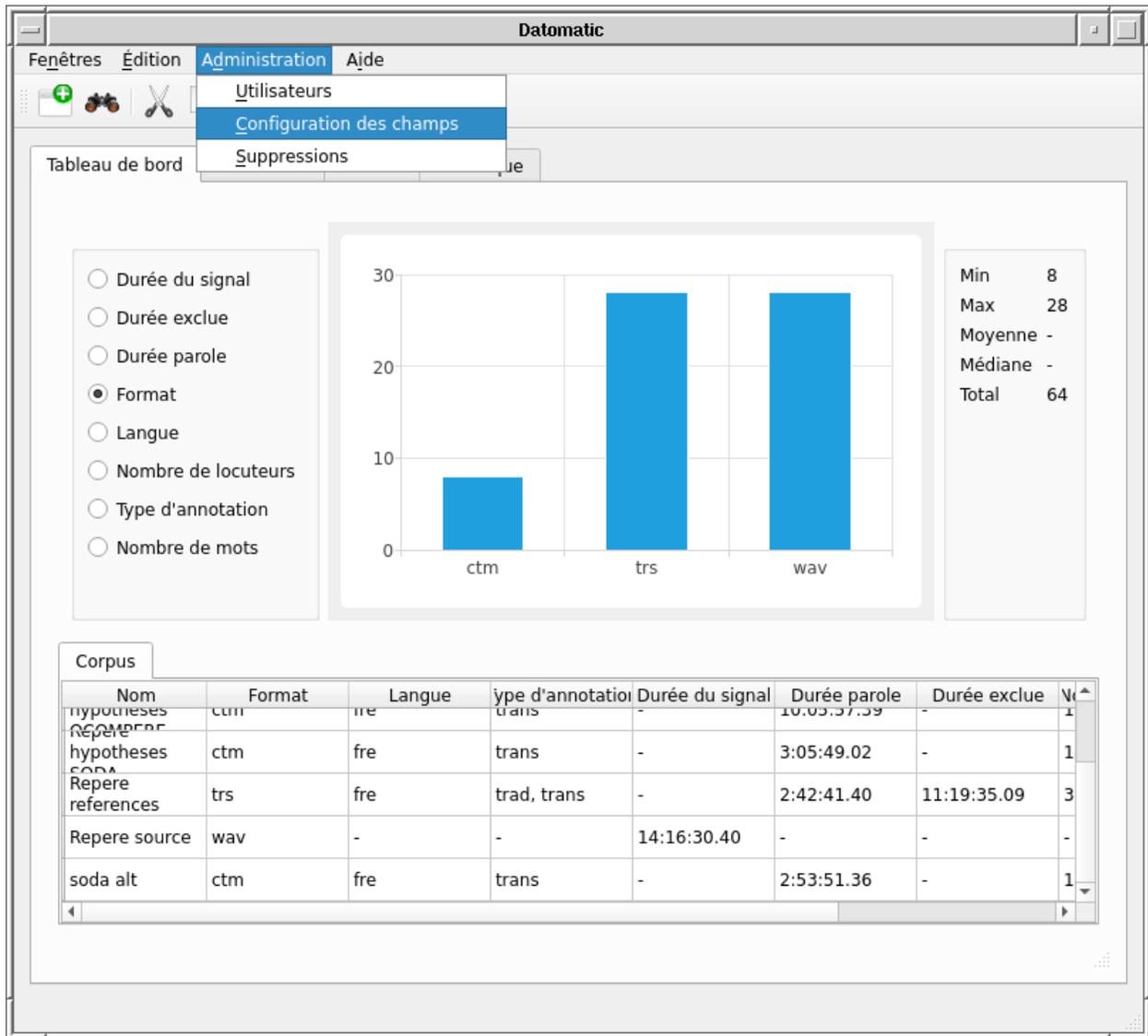


Fig. 1: Accéder à la définition des formulaires

	Nom	Type	Valideur/Options
1	Date de création	Date du jour	
2	Auteur	Auteur	
3	Source	Texte	
4	Classification	Liste de choix	Bleu;Vert;Rouge

Fig. 2: Formulaire de définition des champs

	Nom	Type	Valideur/Options
1	Date de création	Date du jour	
2	Auteur	Auteur	
3	Source	Texte	
4	Classi		Bleu;Vert;Rouge

Ajouter une ligne  
Supprimer

OK Cancel

Fig. 3: Formulaire de définition des champs : ajout d'un champ ou suppression d'un existant

The screenshot shows the Datomatic software interface. The menu bar includes 'Fenêtres', 'Édition', 'Administration', and 'Aide'. The 'Administration' menu is open, showing options: 'Utilisateurs', 'Configuration des champs', and 'Suppressions'. The 'Utilisateurs' option is highlighted.

The main window displays a dashboard with the following elements:

- Tableau de bord:** A bar chart showing the number of items for three formats: ctm, trs, and wav. The y-axis ranges from 0 to 30. The values are approximately 8 for ctm, 28 for trs, and 28 for wav.
- Statistiques:** A summary box on the right showing:
  - Min: 8
  - Max: 28
  - Moyenne: -
  - Médiane: -
  - Total: 64
- Filtres:** A list of filters on the left with radio buttons:
  - Durée du signal
  - Durée exclue
  - Durée parole
  - Format
  - Langue
  - Nombre de locuteurs
  - Type d'annotation
  - Nombre de mots
- Corpus:** A table listing corpus items with columns: Nom, Format, Langue, type d'annotation, Durée du signal, Durée parole, Durée exclue, and Nb.

Nom	Format	Langue	type d'annotation	Durée du signal	Durée parole	Durée exclue	Nb
hypothèses Repere	ctm	fre	trans	-	10:05:37.39	-	1
hypothèses SODA	ctm	fre	trans	-	3:05:49.02	-	1
Repere references	trs	fre	trad, trans	-	2:42:41.40	11:19:35.09	3
Repere source	wav	-	-	14:16:30.40	-	-	-
soda alt	ctm	fre	trans	-	2:53:51.36	-	1

Fig. 4: Accès au formulaire de gestion des utilisateurs

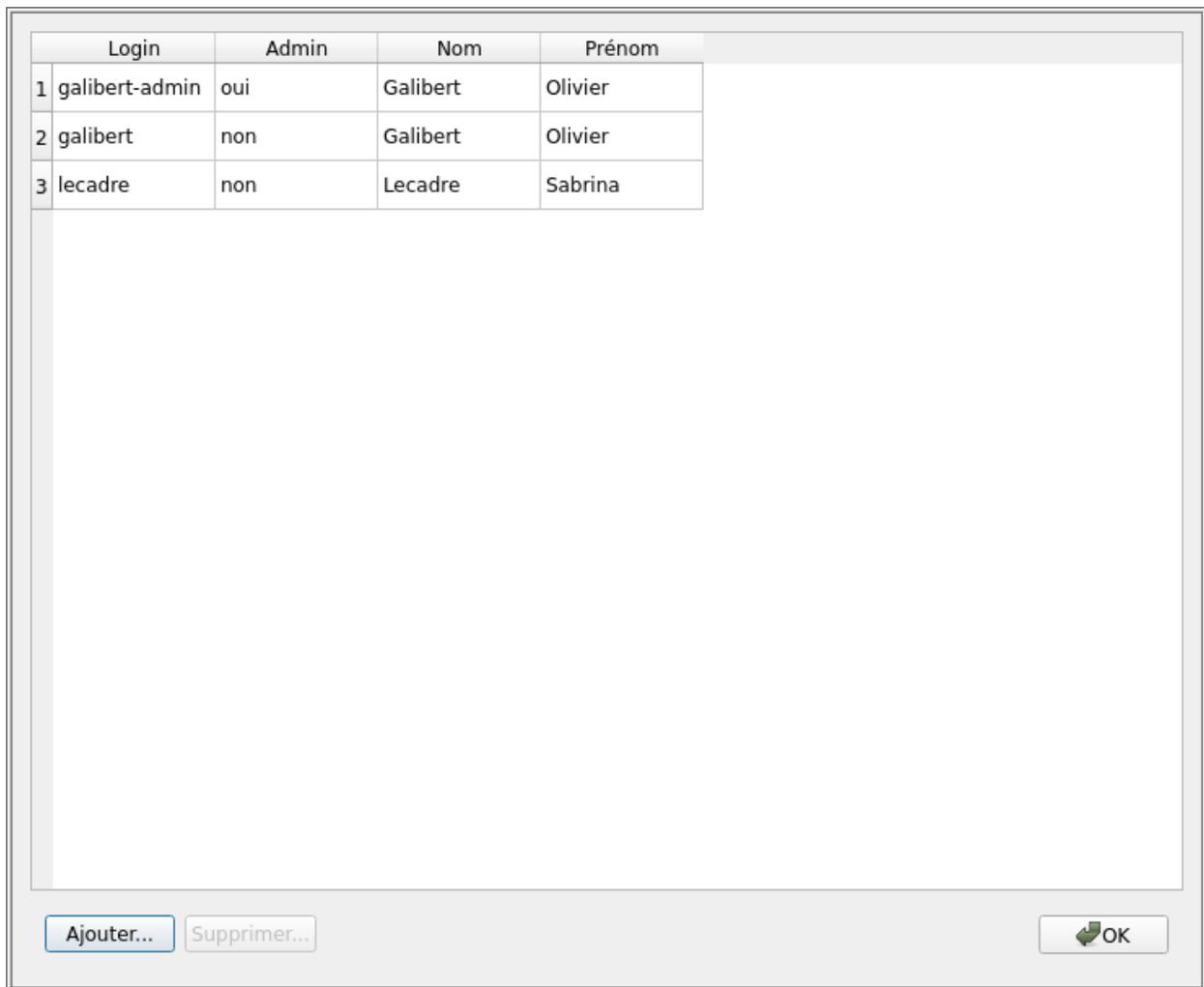


Fig. 5: Ajouter et supprimer des utilisateurs

Il peut supprimer un utilisateur.

Il ne s'agit pas d'une authentification réelle, le nom d'utilisateur pour la session est choisi dans la fenêtre de préférences. Cela permet cependant de ne pas être en mode administrateur sans le vouloir.



## CRÉATION DE CORPUS EN LIGNE DE COMMANDE

Datomic permet la création de corpus en ligne de commande sans interface graphique active. Pour cela deux commandes sont utilisées, *scan* et *import*.

### 12.1 Commande scan

```
/opt/matics/bin/Datomic scan corpus.json [fichierrépertoire] ...
```

La commande scan permet de construire un fichier json à partir d'un ou plusieurs répertoires et fichiers qui contient toutes les informations utiles pour la construction du corpus. La commande va inventorier les fichiers et va ensuite demander le nom et la collection du futur corpus avant de générer le fichier json.

L'utilisateur doit ensuite éditer le fichier json avec un éditeur de texte pour y ajouter le paramétrage des fichiers annotation. Il suffit de remplir les champs correspondants.

Pour les fichiers au type ambigu, il faut aussi choisir entre "hypothese", "reference" ou "source" sur chaque fichier.

### 12.2 Commande import

```
/opt/matics/bin/Datomic import corpus.json
```

La commande import permet de faire l'importation finale du corpus dans la base de données. Elle vérifie les paramètres mis dans le fichier json, demande à l'utilisateur de choisir le type de corpus et de remplir les champs associés, puis fait l'upload. Le serveur détecte si il a un accès direct aux fichiers et si oui évite le transfert.