
Documentation utilisateur EVALOMATIC

Release 0.1

LNE

Dec 13, 2019

CONTENTS

1	Introduction	3
1.1	Objectif du document	3
1.2	Objectifs et fonctionnalités principales de EVALOMATIC	3
1.3	Organisation du document	3
2	Installer EVALOMATIC	5
2.1	Installation des packages	5
3	Lancer EVALOMATIC	7
4	Concepts	9
4.1	Vocabulaire	9
4.2	Dataframe	9
4.3	Vue Dataframe	12
4.4	Conclusion	20
5	Contenu des Dataframes par tâches	21
5.1	Colonnes globales	21
5.2	Transcription de la parole	21
5.3	Annotation des entités nommées	23
5.4	Annotation des lemmes et tokenisation	24
5.5	Vérification du locuteur	25
5.6	Traduction automatique	26
6	Interface principale	29
6.1	Tableau de bord	29
6.2	Barre des tâches	30
7	Créer une évaluation	35
7.1	Procédure de création d'une évaluation	35
7.2	Champs généraux	35
7.3	Champs spécifiques à la tâche	37
7.4	Lancement de l'évaluation	42
8	Chargement de fichiers annotations ou Dataframes	43
8.1	Formulaire de sélection des formats des fichiers	43
8.2	Formulaire de paramétrage des fichiers	43
9	Explorer une Vue d'une Dataframe	47
9.1	Opérations de la Vue	47
9.2	Autres fonctionnalités	64

9.3	Conclusions	66
10	Visualisations supplémentaires	67
10.1	Interface de visualisation des annotations	67
10.2	Interface de visualisation des traductions	68
10.3	Interface de visualisation audio	70
10.4	Nouvelle interface de visualisation audio	70
11	Créer des graphiques en utilisant EVALOMATIC	75
11.1	Les histogrammes classiques	76
11.2	Les histogrammes de distribution	76
11.3	Les histogrammes catégoriels	76
11.4	Les boîtes à moustaches	79
11.5	Les nuages de points	79
11.6	Les courbes DET	80
12	Effectuer des statistiques à l'aide de EVALOMATIC	83
12.1	Statistiques descriptives	84
12.2	Différences appariées	85
12.3	Tests de corrélation entre colonnes	87
13	Quitter EVALOMATIC	89
	Bibliography	91

Table des matières:

INTRODUCTION

1.1 Objectif du document

Ce document est la documentation utilisateur du logiciel EVALOMATIC développé par le LNE. L'objectif est de présenter les fonctionnalités d'EVALOMATIC, ainsi que les différents concepts associés au logiciel. EVALOMATIC existe par ailleurs en version ligne de commande, et une interface python permet de lancer des évaluations en s'appuyant sur des concepts d'EVALOMATIC. Dans le présent document, nous nous attachons avant tout à décrire le logiciel EVALOMATIC.

1.2 Objectifs et fonctionnalités principales de EVALOMATIC

EVALOMATIC est un logiciel permettant l'évaluation de systèmes de traitement de l'information à partir de références et d'hypothèses. Il est possible dans la version actuelle d'évaluer les tâches suivantes :

- Transcription de la parole
- Annotation des entités nommées
- Annotation des lemmes
- Tokenisation
- Vérification du locuteur
- Traduction automatique

EVALOMATIC propose les fonctionnalités suivantes :

- Lancement d'évaluation à partir de corpus exportés de Datomatic ou de fichiers annotés présent localement
- Visualisation des annotations dans un format proche d'une table SQL, avec les manipulations attendues (tri, filtrage, visibilité des colonnes)
- Ajout de colonnes calculées à partir des données déjà présentes dans l'évaluation
- Création de graphes (histogrammes, nuages de points ...) pour visualiser les caractéristiques des données
- Application d'opération statistiques, comme le test de corrélation

1.3 Organisation du document

Nous expliquons d'abord comment installer EVALOMATIC dans le chapitre *Installer EVALOMATIC*, puis comment lancer le logiciel dans *Lancer EVALOMATIC*.

Dans un premier temps, nous expliquons dans le chapitre *Concepts* les concepts intrasèques à EVALOMATIC : nous présentons d'abord le vocabulaire spécifique, puis la structure de données principale du logiciel, et sa représentation graphique. Nous illustrons par des exemples théoriques le fonctionnement des différentes fonctionnalités et mécanismes. Ensuite, nous décrivons le fonctionnement de chaque élément de l'interface graphique d'EVALOMATIC, en nous appuyant sur les concepts présentés précédemment.

Nous procédons ensuite à la description du contenu des Dataframes pour chaque tâche d'évaluation dans le chapitre *Contenu des Dataframes par tâches*. La Dataframe est la structure de données principale d'EVALOMATIC, contenant les informations extraites des fichiers sources, hypothèses et références, ainsi que les résultats des évaluations. Le contenu de ces Dataframes est donc différent pour chaque tâche.

Nous illustrons ensuite le fonctionnement des différentes parties d'EVALOMATIC par le biais d'exemples :

- L'interface principale dans le chapitre *chap-interface*
- La création d'évaluation dans le chapitre *Créer une évaluation*
- Le chargement de fichiers annotations et de Dataframes dans le chapitre *Chargement de fichiers annotations ou Dataframes*
- La manipulation de l'interface graphique représentant les Dataframes dans le chapitre *Explorer une Vue d'une Dataframe*
- Les visualisations supplémentaires associées à certaines tâches dans le chapitre *Visualisations supplémentaires*
- Les graphiques dans le chapitre *Créer des graphiques en utilisant EVALOMATIC*
- Les tests statistiques dans le chapitre *Effectuer des statistiques à l'aide de EVALOMATIC*

Nous concluons ce document en expliquant comment fermer EVALOMATIC dans *Quitter EVALOMATIC*.

INSTALLER EVALOMATIC

2.1 Installation des packages

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

- `xkb-data_2.12-1_all.deb` (standard debian 8)
- `optgcc_6.3.0-1_amd64.deb` (gcc 6.3.0 dans `/opt/gcc`)
- `optffmpeg_3.2.2-1_amd64.deb` (ffmpeg 3.2.2 dans `/opt/ffmpeg`)
- `optqt_5.8.0-1_amd64.deb` (qt 5.8.0 dans `/opt/qt`)
- `optqtav_1.11.0-1_amd64.deb` (qtav 1.11.0 dans `/opt/qt`)
- `matix_0.1-1_amd64.deb` (matix dans `/opt/matix`)

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`)

Pour plus d'informations sur l'installation d'EVALOMATIC, se référer à la documentation d'installation.

LANCER EVALOMATIC

Pour lancer EVALOMATIC, saisissez la commande suivante dans un terminal :

```
./Evalomatic
```


CONCEPTS

Le présent chapitre a pour objectif de présenter les concepts utilisés au sein d'EVALOMATIC. Nous commençons par décrire le vocabulaire principal associé à Evalomatic, puis nous développons les différents concepts inhérents à Evalomatic.

4.1 Vocabulaire

Nous listons ci-dessous le vocabulaire associé à EVALOMATIC. Certains termes, par exemple “Dataframe”, sont ensuite développées plus bas.

- Source : Fichier audio, vidéo, texte dont l’information est à extraire.
- Référence : Fichier contenant la vérité terrain ou réponse attendue suite à l’analyse d’une source.
- Hypothèse : Fichier contenant la réponse du système à évaluer.
- Evaluation : Résultat d’une comparaison entre Référence et Hypothèse, les corpus étant construit à priori.
- Dataframe : Structure de données propre à EVALOMATIC, inspirée des tables des bases de données relationnelles, contenant l’ensemble des informations relatives aux sources, références, hypothèses et évaluations.
- Vue Dataframe : Représentation graphique de la Dataframe sur laquelle des opérations peuvent être effectuées (par exemple le filtrage de certaines lignes).
- Expression : Une opération numérique ou textuelle calculée à partir de plusieurs colonnes de la Dataframe ; les résultats de l’opération permettent de remplir les cellules de la colonne à laquelle l’Expression est associée.
- Variable dépendante : Réponse mesurée du système aux différentes conditions expérimentales. [ElMethni_2017]
- Variable indépendante : Variable représentant les causes postulées des variations de la variable dépendante; le terme de facteur de variation ou de facteur expérimental peut également être employé. [ElMethni_2017]
- Granularité : Facteur de groupement pour les mesures de la Variable dépendante

4.2 Dataframe

Une Dataframe est la structure de données sous-jacente de EVALOMATIC. Elle se présente sous la forme d’une table composée de colonnes et de lignes. Elle est utilisée principalement pour stocker le contenu des sources, hypothèses et références ainsi que les résultats d’évaluation. Le contenu d’une Dataframe n’est pour le moment jamais modifiée, à l’exception de la création d’une nouvelle colonne. Une réflexion est actuellement en cours pour mettre en place des possibilités d’édition du contenu de la Dataframe.

Les Dataframes sont créées à la suite du chargement de fichiers d'annotations ou du lancement d'une Evaluation. Par exemple, une évaluation en transcription de la parole créera trois Dataframes : une contenant les informations des fichiers de référence, une autre les informations des fichiers hypothèses, et enfin une dernière contenant les résultats de l'Evaluation. Nous donnons ci-dessous une représentation du contenu de la Dataframe, pour en illustrer ensuite le fonctionnement :

Table 1: Contenu d'une Dataframe

global.fichier	global.langue	mot.id	mot.text	mot.durée
F1	eng	1	A	157
F1	eng	2	nice	213
F1	eng	3	Dataframe	120
F2	fre	4	Un	356
F2	fre	5	bien	254
F2	fre	6	beau	489
F2	fre	7	tableur	125

Cette Dataframe est composée de 5 colonnes et de 7 lignes. Nous présentons dans les parties suivantes plus en détails la représentation de l'information de la Dataframe.

4.2.1 Colonnes de la Dataframe

Chaque colonne est typée selon plusieurs paramètres. Par ailleurs, chaque colonne peut avoir un sous type optionnel. Enfin, les colonnes sont identifiées selon une convention de nommage particulière. Nous développons ces concepts dans les parties suivantes.

Typage principal des colonnes

Le typage des colonnes de la Dataframe est représentée par le biais de trois critères :

- est ce que les données sont des labels ou des valeurs, c'est à dire Label ou Value ?
- est ce que les données sont représentées sous forme de chaînes de caractères ou de nombres, c'est à dire Num ou Str ?
- est ce que les données sont le résultat d'une expression (Exp) ?

Ces trois critères nous permettent de représenter les différents cas d'utilisation de la Dataframe dans EVALOMATIC. Nous donnons une description des différents critères ci-dessous.

- **Label** qui représente des données de type catégoriels ou des identifiants. Il s'agit de données qui seront par la suite considérées comme les variables indépendantes dans nos analyses.
- **Value** qui représente des valeurs, par exemple le temps de début d'un segment audio. Il s'agit de données qui seront par la suite considérées comme des variables dépendantes dans nos analyses.
- **Num** correspond aux données de type flottant.
- **Str** correspond aux données de type chaînes de caractères.
- **Exp** correspond aux colonnes dont le contenu n'est pas directement stocké dans la Dataframe, mais dépend d'une expression calculée par rapport au contenu d'autre colonne, par exemple une durée à partir des temps de début et de fin.

Si nous appliquons ces conventions de typage à la Dataframe présentée en *Contenu d'une Dataframe*, le typage serait le suivant :

- **global.fichier** : Label+Str

- **global.langue** : Label+Str
- **mot.id** : Label+Num
- **mot.text** : Value+Str
- **mot.durée** : Value+Num

Pour illustrer le cas d’une expression, nous allons ajouter une colonne dépendant d’une expression calculée. Les expressions sont détaillées plus loin dans ce présent document, mais dans cet exemple, nous choisissons d’ajouter une colonne contenant le nombre de mots dans chaque phrase, que nous appelons “mot.nbchars”. La Dataframe résultant de cette opération est la suivante :

Table 2: Contenu d’une Dataframe avec une expression

global.fichier	global.langue	mot.id	mot.text	mot.durée	mot.nbchars
F1	eng	1	A	157	
F1	eng	2	nice	213	
F1	eng	3	Dataframe	120	
F2	fre	4	Un	356	
F2	fre	5	bien	254	
F2	fre	6	beau	489	
F2	fre	7	tableur	125	

La nouvelle colonne ajoutée reste vide : le calcul est effectué lorsque l’on affiche la Dataframe dans une Vue. Le type de la colonne est le suivant : “Value+Num+Exp”.

Typage secondaire des colonnes

En plus de leur typage principal, les colonnes de la Dataframe peuvent aussi avoir un sous type. Ce type secondaire permet de décrire plus précisément le contenu de la colonne associée. Une colonne représentant les noms de fichiers sera par exemple un label de type chaîne de caractères (Label+Str), à laquelle on a associé le sous type file (fichier). Dans la Dataframe *Contenu d’une Dataframe avec une expression*, c’est le cas de la colonne “global.fichier” qui a le sous type “file”.

Nommage des colonnes

Les colonnes de la Dataframe sont identifiées par deux paramètres : leur nom et leur groupe. La combinaison de ces deux paramètres correspond à la clé de la colonne, et cette dernière est forcément unique au sein d’une Dataframe. Dans la Dataframe *Contenu d’une Dataframe avec une expression*, si on prend la colonne “global.fichier”, le groupe est “global”, et le nom “fichier”.

Méthode de réduction des colonnes

Les colonnes peuvent aussi avoir une méthode de réduction associée. Cette méthode a un impact lors du regroupement de plusieurs lignes entre elles, ce qui peut arriver lors du changement de visibilité des colonnes. Nous expliquons plus bas le fonctionnement de la visibilité des colonnes. Les méthodes de réduction sont les suivantes : concaténation, somme, moyenne arithmétique et géométrique, minimum, maximum et médiane.

Dans la Dataframe illustré dans l’exemple *Contenu d’une Dataframe avec une expression*, les méthodes de réduction sont les suivantes :

- **mot.text** : concaténation.
- **mot.durée** : somme.

- `mot.nbchars` : somme.

4.2.2 Références

Pour chaque colonne d'une Dataframe, les informations de chaque ligne sont soit stockées directement, soit représentées sous forme de référence. Une référence représente une case de la dataframe dont le contenu est identique à une case de la même colonne de référence. C'est généralement un cas de figure que l'on voit apparaître pour des colonnes Label+Str, par exemple une colonne représentant les systèmes d'une évaluation. Attention, deux valeurs identiques d'une même colonne n'impliquent pas forcément que l'une est la référence de l'autre. Dans la Dataframe représentée dans l'exemple *Dataframe avec les références indiquées*, nous avons fait apparaître les références présentes : ainsi dans la colonne "global.fichier", les cases des lignes 2 et 3 font référence à la première ligne, d'où le "\$1".

Table 3: Dataframe avec les références indiquées

global.fichier	global.langue	mot.id	mot.text	mot.durée	mot.nbchars
F1	eng	1	A	157	
\$1	\$1	2	nice	213	
\$1	\$1	3	Dataframe	120	
F2	fre	4	Un	356	
\$4	\$4	5	bien	254	
\$4	\$4	6	beau	489	
\$4	\$4	7	tableur	125	

Cette représentation sous forme de référence est importante pour représenter précisément le contenu des Dataframes, et beaucoup de concepts d'Evalomatic s'appuient sur ces références.

4.3 Vue Dataframe

Si la Dataframe est la structure de données permettant de stocker les informations dans un format tableur, la Vue est l'interface permettant de visualiser et de manipuler les informations contenues dans la Dataframe. Les manipulations effectuées sur la vue, par exemple le filtrage de lignes, modifient la visualisation, mais pas le contenu de la Dataframe. Il peut y avoir plusieurs Vues sur une même Dataframe. La seule exception concerne la création d'une nouvelle colonne calculée, opération faite à partir d'une Vue, mais qui ajoute de nouvelles informations dans la Dataframe, et donc sur toutes les Vues liées à cette Dataframe.

Lorsqu'une Vue est créée, sa représentation est initialement proche de celle d'une Dataframe. Un exemple de Vue est donnée dans l'illustration *Vue créée initialement à partir d'une Dataframe*.

Table 4: Vue créée initialement à partir d'une Dataframe

global.fichier	global.langue	mot.id	mot.text	mot.durée	mot.nbchars
F1	eng	1	A	157	1
		2	nice	213	4
		3	Dataframe	120	9
F2	fre	4	Un	356	2
		5	bien	254	4
		6	beau	489	4
		7	tableur	125	7

En plus de l'affichage des informations contenues dans une Dataframe, la Vue permet aussi de manipuler les informations visualisées. Les opérations de manipulation possibles sont listées ci-dessous :

- Duplication de la Vue

- Visibilité des colonnes
- Tri des colonnes
- Ajout d’une colonne calculée à partir d’une Expression
- Filtrage des lignes
- Manipulation des en-têtes

Dans la suite de ce chapitre, nous présentons les différents concepts inhérents à la Vue de la Dataframe : nous expliquons d’abord les stratégies d’affichage, puis nous revenons sur les différentes manipulations et les concepts sur lesquelles ces dernières s’appuient.

4.3.1 Affichage de la Vue

La Vue de la Dataframe est présentée dans un format tableur classique. L’affichage des cases dépend de la valeur contenue à l’intérieur. Si la valeur de cette case est identique à la précédente, la case affichée est alors vide, et aucune ligne de séparation n’est visible. Dans la Vue présentée dans l’illustration *Vue créée initialement à partir d’une Dataframe*, c’est le cas dans la colonne “global.fichier”, ou les cases des lignes 2 et 3 sont vides, du fait que la valeur est identique à celle de la ligne 1. Nous nous appuyons notamment sur les références présentées dans l’illustration *Dataframe avec les références indiquées* pour déterminer ces égalités.

Par ailleurs, lorsqu’une Vue est créée, si la Dataframe correspondante contenait une ou plusieurs colonnes avec des expressions, les calculs sont effectués et affichés dans la colonne correspondante dans la Vue. C’est le cas dans notre exemple, avec la colonne “mot.nbchars”.

Par rapport à cette représentation sous la forme d’un tableur classique, un deuxième point primordial est à prendre en compte : une ligne présente dans la Vue ne correspond pas forcément à une même ligne dans la Dataframe correspondante, mais potentiellement à la fusion de plusieurs lignes de la Dataframe. Ce phénomène peut se manifester lors du changement de visibilité des colonnes, que nous présentons dans la partie suivante.

4.3.2 Gestion des colonnes visibles

Pour illustrer l’impact de la visibilité, nous nous appuyons sur une Vue différente de celle utilisée initialement, que nous présentons dans l’illustration *Vue illustrant la visibilité des colonnes*. Dans cette Vue, le temps de début et de fin d’un tour sont ajoutées.

Table 5: Vue illustrant la visibilité des colonnes

global.fichier	global.langue	mot.id	turn.text	turn.debut	turn.fin	turn.durée
F1	eng	1	A	1	16	15
		2	Dataframe	18	338	320
F2	fre	3	Un	342	364	22
		4	tableur	365	589	224

Lors du changement de visibilité des colonnes de la Vue, l’opération peut avoir un impact sur les lignes visibles de la Vue. En effet, le fait de masquer des colonnes Label, que l’on définit comme étant structurantes dans une Vue, peut entraîner des fusions de lignes. Pour déterminer quelles lignes sont fusionnées, on s’appuie sur l’ensemble des valeurs contenues dans les cases des colonnes Labels pour chaque ligne. Si plusieurs lignes ont la même combinaison de valeurs Label, ces lignes sont alors fusionnées. Pour déterminer comment doivent se calculer les fusions, on s’appuie sur les méthodes de réductions associées à chaque colonne Valeur : si une colonne Valeur n’a pas de méthode de réduction, la fusion de plusieurs lignes est considérée comme vide, excepté s’il s’agit des mêmes valeurs.

Pour illustrer ce concept, on prend la Vue illustrée dans *Vue illustrant la visibilité des colonnes*. Si on ferme la colonne “mot.id”, les lignes 1 à 2 ont la même combinaison de labels, à savoir “F1+eng”. De même, les lignes 3 à 4 ont la

même combinaison, “F2+fre”. Ces égalités nous permettent de déterminer quelles sont les fusions effectuées dans la Vue. Les méthodes de réduction sont “concaténation” pour “turn.text”, minimum pour “turn.debut”, maximum pour “turn.fin” et “somme” pour “turn.durée”. L’état de la Vue suite au changement de visibilité de la colonne “mot.id” est illustrée dans *Vue illustrant la granularité*.

Table 6: Vue illustrant la granularité

F1	eng	A Dataframe	1	338	turn.durée
					335
F2	fre	Un tableur	1324	342	589
					246

Il est à noter qu’il n’arrive normalement jamais qu’une Dataframe ait plusieurs lignes avec des combinaisons de labels identiques.

4.3.3 Fonctionnalités de la Vue

Nous listons ci-dessous les possibilités offertes par la Vue :

- Duplication de la Vue
- Visibilité des colonnes
- Tri des colonnes
- Filtrage des lignes
- Ajout d’une colonne calculée à partir d’une Expression
- Transformation de la matrice

Nous allons expliquer les différents concepts inhérents à chacune de ces fonctionnalités ci-dessous.

Duplication de la Vue

Que la Vue ait subi des modifications suite à des opérations ou non, il est possible à tout moment de dupliquer la Vue actuelle pour en créer une nouvelle. Cette dernière sera dans le même état, et les deux Vues seront indépendantes. Ainsi, une opération sur une de ces deux Vues n’aura pas d’impact sur l’autre.

Visibilité des colonnes

Chaque colonne de la Vue peut être masquée ou affichée. Pour déterminer le comportement de la Vue à la suite de cette opération, on s’appuie sur les mécanismes présentés précédemment. Ce principe s’applique aussi bien lorsque l’on cache une colonne que lorsqu’on la rend visible. On parle aussi de fold et d’unfold pour cacher ou rendre visible les colonnes.

Table 7: Vue illustrant la visibilité des colonnes

global.fichier	global.langue	mot.id	mot.text	mot.durée	mot.nbchars
F1	eng	1	A	157	1
		2	nice	213	4
		3	Dataframe	120	9
F2	fre	4	Un	356	2
		5	bien	254	4
		6	beau	489	4
		7	tableur	125	7
F3	eng	8	Hello	89	5

Dans la Vue présentée dans l’illustration *Vue illustrant la visibilité des colonnes*, la colonne “mot.text” a “concaténation” comme méthode de réduction, la colonne “mot.durée” a “somme” comme méthode, et la colonne mot.nbchars n’a pas de méthode de réduction. Si la colonne “mot.id” est masquée, le résultat est illustré dans la Vue *Vue illustrant les opérations de fold : fold de mot.id*.

Table 8: Vue illustrant les opérations de fold : fold de mot.id

global.fichier	global.langue	mot.text	mot.durée	mot.nbchars
F1	eng	A nice Dataframe	490	
F2	fre	Un bien beau tableur	1324	
F3	eng	Hello	89	5

Comme dans l’exemple *Vue illustrant la granularité*, le fold de la colonne “mot.id” conduit à une fusion de lignes. Les lignes avec les mêmes combinaisons de labels sont fusionnées. On dénombre ainsi trois combinaisons : F1+eng, F2+fre, F3+eng.

La colonne mot.nbchars n’ayant pas d’opération de réduction, le contenu des cases associées aux combinaisons F1+eng et F2+fre sont vides. Par contre, la case associée à F3+eng contient une valeur, car cette case résulte de la “fusion” d’une seule et unique ligne.

Table 9: Vue illustrant les opérations de fold : fold de global.langue

global.fichier	mot.text	mot.durée	mot.nbchars
F1	A nice Dataframe	490	
F2	Un bien beau tableur	1324	
F3	Hello	89	5

Si on décide de masquer la colonne “global.langue” par une opération de fold, il n’y aura pas de nouvelles fusions : en effet, aucune nouvelle combinaison de labels n’est créée à la suite de cette opération. Le résultat de cette opération est visible dans la Vue *Vue illustrant les opérations de fold : fold de global.langue*.

Si l’utilisateur décide de rendre la colonne “mot.id” de nouveau visible par une opération de unfold, la Vue sera cette fois modifiée et reprendra une structure équivalente à celle de l’exemple *Vue illustrant la granularité*, avec néanmoins la colonne “global.langue” toujours masquée. La Vue résultat est visible dans l’exemple *Vue illustrant les opérations de unfold : unfold de mot.id*.

Table 10: Vue illustrant les opérations de unfold : unfold de mot.id

global.fichier	mot.id	mot.text	mot.durée	mot.nbchars
F1	1	A	157	1
	2	nice	213	4
	3	Dataframe	120	9
F2	4	Un	356	2
	5	bien	254	4
	6	beau	489	4
	7	tableur	125	7
F3	8	Hello	89	5

Tri des colonnes

Par défaut, la Vue d’une Dataframe représente le contenu tel qu’il est organisé dans la Dataframe. Cet ordre a généralement un sens. Néanmoins, l’utilisateur peut vouloir réorganiser les colonnes selon ses préférences. Le tri des colonnes permet de choisir une liste de colonnes réorganisant les informations contenues par ordre croissant ou décroissant. Pour les colonnes non sélectionnées, on garde l’ordre précédent le tri. Quelque soit le tri effectué, le contenu de la Vue

n'est jamais modifié : le tri n'impacte que l'ordre des lignes, qu'il y ait eu des fusions ou non suite à un changement de visibilité de colonnes.

Table 11: Vue illustrant les opérations de tri

global.fichier	global.langue	mot.id	mot.text	mot.durée
F1	eng	1	A	157
		2	nice	213
		3	Dataframe	120
F2	fre	4	Un	356
		5	bien	254
		6	beau	489
		7	tableur	125
F3	eng	8	Hello	89

Pour illustrer les concepts derrière le tri, nous allons nous appuyer sur la Vue illustrée dans l'exemple *Vue illustrant les opérations de tri*. Le premier tri est effectué sur la colonne "global.langue", dans l'ordre croissant. Le résultat est visible dans l'exemple *Vue illustrant les opérations de tri : tri sur la colonne gloabl.langue*.

Table 12: Vue illustrant les opérations de tri : tri sur la colonne gloabl.langue

global.fichier	global.langue	mot.id	mot.text	mot.durée
F1	eng	1	A	157
		2	nice	213
		3	Dataframe	120
F3	eng	8	Hello	89
F2	fre	4	Un	356
		5	bien	254
		6	beau	489
		7	tableur	125

Le tri sur la colonne "global.langue" ne modifie que très modérément la structure de la Vue : seule la ligne contenant la combinaison F3+eng+8 est déplacée. Dans la Vue illustrée dans l'exemple *Vue illustrant les opérations de tri : tri sur les colonnes "global.fichier" et "mot.durée"*, nous donnons le résultat du tri sur les colonnes "global.fichier" en ordre croissant et "mot.durée" en ordre décroissant.

Table 13: Vue illustrant les opérations de tri : tri sur les colonnes "global.fichier" et "mot.durée"

global.fichier	global.langue	mot.id	mot.text	mot.durée
F1	eng	2	nice	213
		1	A	157
		3	Dataframe	120
F2	fre	6	beau	489
		4	Un	356
		5	bien	254
		7	tableur	125
F3	eng	8	Hello	89

Avec ce tri multiple, les lignes sont d'abord triées par fichier, puis par la durée des mots.

Filtrage des lignes

Le filtrage des lignes consiste à sélectionner des lignes de la Vue selon différents critères : cette sélection entrainera un filtrage des lignes, c'est à dire qu'elles seront masquées. Ce filtrage s'appuie là encore sur la granularité de la Vue : on filtre les lignes selon l'état d'une Vue.

Les critères vont dépendre du type de la colonne sur laquelle on applique un filtre :

- Lab+Num : opérateurs numériques de comparaisons unaires ou binaires;
- Lab+Str : utilisation d'expressions régulières ou sélection manuelle des labels;
- Val+Num : identique à Lab+Num;
- Val+Str : utilisation d'expressions régulières.

Le filtrage de lignes n'entraîne jamais de nouveaux calculs : la seule conséquence est que certaines lignes visibles de la Vue disparaissent. Nous nous appuyons sur la Vue illustrée dans l'exemple *Vue illustrant le filtrage* pour donner un exemple de filtrage.

Table 14: Vue illustrant le filtrage

global.fichier	global.langue	mot.id	mot.text	mot.durée
F1	eng	1	A	157
		2	nice	213
		3	Dataframe	120
F2	fre	4	Un	356
		5	bien	254
		6	beau	489
		7	tableur	125
F3	eng	8	Hello	89

A partir de l'état initial de cette Vue, nous effectuons une opération de filtrage sur la colonne "mot.durée" : on sélectionne les lignes dont la valeur est supérieur à 200. Le résultat de l'opération est illustré dans la Vue *Vue illustrant le filtrage : filtrage des durées de mot supérieures à 200*.

Table 15: Vue illustrant le filtrage : filtrage des durées de mot supérieures à 200

global.fichier	global.langue	mot.id	mot.text	mot.durée
F1	eng	2	nice	213
F2	fre	4	Un	356
		5	bien	254
		6	beau	489

Ajout d'une colonne calculée à partir d'une Expression

En plus des colonnes déjà présentes, il est possible d'ajouter une nouvelle colonne à partir de la Vue. Au contraire des autres opérations de la Vue, l'ajout de colonne modifie aussi le contenu de la Dataframe. Ces colonnes suivent la typologie des colonnes classiques : le type est défini lors de la création de la colonne. Pour illustrer le fonctionnement des expressions nous nous appuyons sur la Vue illustrée dans l'exemple *Vue illustrant les colonnes calculées*.

Table 16: Vue illustrant les colonnes calculées

global.système	global.fichier	tour.id	tour.deb	tour.fin
S1	F1	1	15	158
		2	190	256
	F2	3	257	289
		4	315	458
	F3	5	510	545

Une expression fait généralement référence à une ou plusieurs colonnes de la Dataframe, dont le contenu des cases est utilisé pour calculer le contenu de la colonne Expression. L'utilisateur créé son expression en utilisant des opérateurs numériques et logiques, ainsi que différentes fonctions mathématiques usuelles (moyenne, sin, somme ...). Pour faire référence à une colonne, il suffit d'utiliser la syntaxe \$clef de la colonne\$. Ainsi, le calcul de la durée d'un tour de la parole pourrait être écrit de la manière suivante :

```
$tour.fin$ - $tour.deb$
```

La granularité d'une colonne expression peut être fixée explicitement. Si une colonne expression n'a pas de granularité explicite, les calculs s'appuient directement sur les lignes visibles de la Vue. Ainsi, le calcul est effectué à partir des valeurs des colonnes associées à l'expression à la suite de la réduction de la Vue. Un exemple est présenté dans la Vue `vue-expression4`.

Table 17: Vue illustrant les colonnes calculées : fold de la colonne "tour.id"

global.système	global.fichier	tour.deb	tour.fin	tour.durée
S1	F1	15	256	241
	F2	257	458	201
	F3	510	545	35

Dans cet exemple, la durée est calculée à partir des cases fusionnées des colonnes "tour.deb" et "tour.fin". Enfin, si on masque la colonne "global.fichier", le résultat est illustré dans l'exemple *Vue illustrant les colonnes calculées : fold de la colonne "global.fichier"*.

Table 18: Vue illustrant les colonnes calculées : fold de la colonne "global.fichier"

global.système	tour.deb	tour.fin	tour.durée
S1	15	545	530

Une granularité explicite est donnée à la création de l'expression sous la forme d'une colonne de référence dont la granularité sera suivie. Une et une seule valeur sera calculée par case de la colonne de référence. Par exemple un temps de parole pourra être calculé au fichier en donnant comme granularité de référence la colonne "global.fichier". Si à la suite du fold de certaines colonnes, dont la colonne de référence, la granularité de la Vue englobe plusieurs cases de la colonne de référence, des opérations de fusion seront effectuées en s'appuyant sur la méthode de réduction de la colonne expression. Nous illustrons ce fonctionnement dans la Vue illustrée dans l'exemple *Vue illustrant les colonnes calculées : création d'une colonne calculée "tour.durée" avec comme granularité la colonne "global.fichier", et "somme" comme méthode de réduction.*

Table 19: Vue illustrant les colonnes calculées : création d'une colonne calculée "tour.durée" avec comme granularité la colonne "global.fichier", et "somme" comme méthode de réduction.

global.système	global.fichier	tour.id	tour.deb	tour.fin	tour.durée
S1	F1	1	15	158	241
		2	190	256	
	F2	3	257	289	201
		4	315	458	
	F3	5	510	545	35

Dans cet exemple, une colonne "tour.durée" est créée avec comme granularité "global.fichier" et comme méthode de réduction "somme". Les méthodes de réduction "tour.deb" et "tour.fin" sont respectivement "minimum" et "maximum". La granularité de la colonne expression étant au fichier, on calcule les valeurs des colonnes "tour.deb" et "tour.fin" à cette granularité. Par exemple, dans la case associée au fichier "F1", les valeurs sont 15 et 256 : en appliquant l'expression, la valeur calculée est donc 241.

Si l'utilisateur décide de masquer les colonnes "global.fichier" et "tour.id", la granularité a un impact sur le contenu "tour.durée". Le résultat est illustré dans l'exemple *Exemple d'une vue pour illustrer les colonnes calculées : fold des colonnes "global.fichier" et "tour.id"*.

Table 20: Exemple d'une vue pour illustrer les colonnes calculées : fold des colonnes "global.fichier" et "tour.id"

global.système	tour.deb	tour.fin	tour.durée
S1	15	545	477

Dans cet exemple, du fait du changement de granularité, le contenu de la case est calculé en appliquant la méthode de réduction "somme". Du fait d'avoir créé la colonne calculée avec une granularité de référence et une méthode de réduction, le résultat obtenu dans la case est différent de celui de la Vue *Vue illustrant les colonnes calculées : fold de la colonne "global.fichier"*.

Manipulation des en-têtes

Par défaut, l'organisation d'une Vue s'appuie sur la représentation héritée d'une Dataframe, à savoir une représentation sous forme de colonnes verticales. Néanmoins, cette représentation peut avoir ces limites pour représenter et visualiser l'information contenue dans une Dataframe. De fait, les Vues permettent de manipuler les en-têtes des colonnes pour modifier la représentation. Nous présentons dans cette partie les concepts sur lesquels s'appuient cette fonctionnalité.

La représentation graphique d'une Vue s'appuie sur deux caractéristiques : l'orientation générale de la Vue, et la position des en-têtes à l'horizontal ou à la vertical. Par défaut, l'orientation de la Vue est à la verticale, et il n'y a aucun en-têtes à l'horizontal ou à la vertical. C'est le fait de déplacer une ou plusieurs colonnes Label en-tête qui va provoquer un changement de la représentation graphique de la Vue : si les colonnes sont déplacés en en-têtes dans le sens opposé à l'orientation de la Vue, alors il y aura une réorganisation automatique des colonnes Valeurs. Pour déterminer les colonnes Valeurs réorganisées, on détecte automatique quelles sont les colonnes Valeurs dépendant des colonnes Labels déplacées en en-têtes.

Nous allons illustrer ce fonctionnement en nous appuyant sur la Vue *Exemple d'une vue pour illustrer les manipulations des en-têtes*.

Table 21: Exemple d’une vue pour illustrer les manipulations des en-têtes

global.fichier	global.durée	système.id	système.wer	système.cer
F1	1253	S1	12.5	24.8
		S2	17.8	14.2
		S3	45.2	64.5
F2	847	S1	19.5	10.4
		S2	40.5	32.8
		S3	31.2	32.5
		S4	24.2	27.9

Dans cette Vue, l’orientation générale est verticale, et aucune colonne Label n’a été manipulée : il n’y a donc aucun en-tête à l’horizontal ou à la vertical. Si on décide alors de déplacer la colonne “système.id” en en-tête horizontal, il y a opposition entre l’orientation de la Vue et orientation de l’en-tête. Il va donc y avoir un impact sur la représentation des colonnes Valeurs. Dans ce cas, les colonnes Valeurs “système.wer” et “système.cer” dépendent de la colonne “système.id”, mais pas la colonne “global.durée”. La Vue résultat suite à cette manipulation est illustrée en *Manipulation des en-têtes : passage de “global.système” en en-tête horizontal*.

Table 22: Manipulation des en-têtes : passage de “global.système” en en-tête horizontal

		Système						S1	S2
		S3	S4					WER	CER
global.fichier	global.durée	WER	CER	WER	CER	WER	CER		
F1	1253	12.5	24.8	17.8	14.2	45.2	64.5		
F2	847	19.5	10.4	40.5	32.8	31.2	32.5	24.2	27.9

La colonne “système.id” est passée en en-tête horizontal, où l’on voit apparaître chacun des systèmes. Les deux colonnes valeurs impactées, “système.wer” et “système.cer” sont alors réorganisées en fonction de cette nouvelle représentation. On peut voir que deux cellules vides apparaissent du fait cette nouvelle représentation. Par contre, si on décide de manipuler la colonne “global.fichier” et de la mettre en en-tête vertical, l’orientation de la Vue étant elle aussi vertical, cette manipulation n’aura qu’un simple impact formel. Cette manipulation est illustrée dans la Vue *Manipulation des en-têtes : passage de “global.fichier” en en-tête vertical*.

Table 23: Manipulation des en-têtes : passage de “global.fichier” en en-tête vertical

		Système								S1
		S2	S3	S4					WER	CER
Fichier	F2	WER	CER	WER	CER	WER	CER	WER	CER	
	1253	12.5	24.8	17.8	14.2	45.2	64.5			
	847	19.5	10.4	40.5	32.8	31.2	32.5	24.2	27.9	

Il est possible d’avoir plusieurs en-têtes dans un sens ou dans l’autre : l’ordre déterminera alors la hiérarchie de headers.

4.4 Conclusion

Nous avons présenté les différents concepts importants inhérents à Evalomatic dans ce chapitre. Dans les chapitres suivants, nous illustrons le fonctionnement de ces différents concepts et fonctionnalités au travers d’exemples illustrés. Nous expliquerons notamment comment créer une évaluation, puis comment explorer les résultats obtenus.

CONTENU DES DATAFRAMES PAR TÂCHES

Dans le chapitre *Concepts*, nous avons présenté le concept de Dataframe. Pour rappel, la Dataframe est la structure de données principale d'EVALOMATIC. Elle permet de représenter le contenu extrait des fichiers sources, hypothèses et références ainsi que le résultat des évaluations.

Selon la tâche, le contenu de ces Dataframes sera différent en fonction des données à représenter. Nous nous attachons dans le présent chapitre à détailler le contenu des Dataframes liées à chaque tâche : pour chacune des colonnes présentes, nous décrivons dans quelles types de Dataframe (références, hypothèses, résultats) ces données peuvent se retrouver, et si elles sont obligatoirement présentes ou non. Enfin, nous indiquons aussi le type de chaque colonne.

5.1 Colonne globales

Ces colonnes ne sont pas liées à une tâche spécifique et se retrouvent généralement dans toutes les évaluations :

- **Groupe “global” et “general” :**
 - “file” : les fichiers présents - optionnel - présent dans toutes les Dataframes
 - “system” et “system version” : les systèmes et leurs versions - optionnel - présent dans l’hypothèse et l’évaluation
 - “comments” : des commentaires à ajouter aux fichiers - optionnel - présent dans toutes les Dataframes
 - “lang” : la langue évaluée - optionnel - présent dans toutes les Dataframes
 - “spk_local” et “spk_local_regexp” : détermine la démarche à avoir par rapport aux locuteurs identiques d’un tour de parole à un autre - optionnel - présent dans les Dataframes transcription de la parole seulement

5.2 Transcription de la parole

EVALOMATIC gère les types de fichiers suivant pour l’évaluation en transcription de la parole :

- TRS
- STM
- TDF
- CTM
- ASRXML

Les colonnes présentes sont les suivantes :

- **Groupe “global” :**

- “channel” : le canal d’un fichier - présent dans toutes les Dataframes
- **Groupe “section” :**
 - “id” : l’identifiant de la section - présent dans toutes les Dataframes
 - “start time” : le temps de début d’une section - présent dans toutes les Dataframes
 - “end time” : le temps de fin d’une section - présent dans toutes les Dataframes
 - “type” : le type de la section - optionnel - présent dans toutes les Dataframes des évaluations STM et TRS
- **Groupe “speaker” :**
 - “name” : le nom du locuteur - présent dans toutes les Dataframes
 - “type” : le type du speaker - optionnel - présent dans toutes les Dataframes TRS
 - “gender” : le genre (féminin/masculin/inconnu) du locuteur - présent dans toutes les Dataframes
 - “dialect” : le dialecte du locuteur - optionnel - présent dans toutes les Dataframes des évaluations TDF et TRS
 - “accent” : l’accent du locuteur - optionnel - présent dans toutes les Dataframes TRS
- **Groupe “turn” :**
 - “id” : l’identifiant du tour de parole - présent dans toutes les Dataframes
 - “start time” : le temps de début du tour de parole - présent dans toutes les Dataframes
 - “end time” : le temps de fin du tour de parole - présent dans toutes les Dataframes
 - “raw text” : le texte brut associé au tour de parole - présent dans toutes les Dataframes TRS
 - “pronounced text” : le texte associé au tour de parole - présent dans toutes les Dataframes
- **Groupe “word” :**
 - “id” : l’identifiant du mot - présent dans les Dataframes hypothèses
 - “start time” : le temps de début du mot - présent dans les Dataframes hypothèses
 - “end time” : le temps de fin du mot - présent dans les Dataframes hypothèses
 - “text” : le mot sous forme textuel - présent dans les Dataframes hypothèses
 - “confidence” : le taux de confiance associé par le système au mot - présent dans les Dataframes hypothèses
- **Groupe “cer cd” et “wer ci”**
 - “id” : l’identifiant du mot évalué par la métrique - présent dans les Dataframes évaluations
 - “ref text” : le texte référence associé au mot - présent dans les Dataframes évaluations
 - “hyp start time” et “hyp end time” : le temps de début et de fin du mot - présent dans les Dataframes évaluations pour le WER
 - “hyp text” : le texte hypothèse associé au mot - présent dans les Dataframes évaluations
 - “correct”, “subst”, “insert”, “delete”, “delete opt”, “correct cut”, “delete cut”, “ref cnt” et “hyp cnt” : les compteurs utilisés pour calculer les métriques - présent dans les Dataframes évaluations
 - “entropie”, “hyp correct”, “hyp incorrect” et “entropie max” : ensemble de colonnes utilisées pour calculer l’entropie maximum - présent dans les Dataframes évaluations pour le WER

- “WER” ou “CER” : métrique avec laquelle l’évaluation a été lancée - présent dans les Dataframes évaluations
- “NCE” : résultat de la métrique NCE, dépend de l’entropie - présent dans les Dataframes évaluations pour le WER

5.3 Annotation des entités nommées

Actuellement, EVALOMATIC gère les fichiers au format suivant :

- CoNLL 2002, 2003 et 2003L
- Quaero IMM, Quaero V1 et Quaero V2
- ACE
- GermEval
- Apachene
- Stanford NER
- MUC Fragment

Les colonnes présentes sont les suivantes :

- **Groupe “sentence” :**
 - “id” : l’identifiant de chaque phrase - présent dans les Dataframes références et hypothèses
 - “word id” : les identifiants des mots de chaque phrase - présent dans les Dataframes références et hypothèses
 - “word text” : les mots de chaque phrase - présent dans les Dataframes références et hypothèses
 - “pos” : le Part of Speech associé à chaque mot - optionnel - présent dans les Dataframes issues des évaluations Stanford, MUC Fragment et CoNLL
 - “component id” : les identifiants des groupes syntaxiques de chaque phrase - optionnel - présent dans les Dataframes issues des évaluations Quaero et CoNLL
 - “component type” : les groupes syntaxiques - optionnel - présent dans les Dataframes issues des évaluations Quaero et CoNLL
 - “named entity id” : les identifiants de chaque entité nommée contenue dans la phrase - présent dans les Dataframes références et hypothèses
 - “named entity type” : le type d’entité nommée - présent dans les Dataframes références et hypothèses
 - “named entity category” : la catégorie de l’entité nommée - optionnel - présent dans toutes les Dataframes issues des évaluations Ace et MUC Fragment
 - “named entity metonymy” : la métonymie associé à une entité nommée - optionnel - présent dans les Dataframes références et hypothèses issues de l’évaluation Quaero

Les entités nommées des références et hypothèses peuvent avoir plusieurs niveaux de profondeurs : il peut donc y avoir un “named entity id 1” et “named entity type 1”, “named entity id 2” et “named entity type 2”, etc ...

- **Groupe “segment” :**
 - “id” : l’identifiant du segment - présent dans les Dataframes évaluations
 - “text” : le texte - présent dans les Dataframes évaluations
- **Groupe “slot” :**

- “id” : l’identifiant du slot - présent dans les Dataframes évaluations
- “reference id” : l’identifiant des entités nommées de la référence - présent dans les Dataframes évaluations
- “hypothesis id” : l’identifiant des entités nommées de l’hypothèse - présent dans les Dataframes évaluations
- “insertion cost”, “deletion cost” et “substitution cost” : les coûts des différentes opérations associées au SER - présent dans les Dataframes évaluations
- “hypothesis count” et “reference count” : le nombre d’entités nommées références et d’hypothèses - présent dans les Dataframes évaluations
- “ser” : le résultat de la métrique SER - présent dans les Dataframes évaluations
- “precision” : la précision - présent dans les Dataframes évaluations
- “recall” : le rappel - présent dans les Dataframes évaluations
- “f-measure” : la f-mesure - présent dans les Dataframes évaluations

5.4 Annotation des lemmes et tokenisation

Les deux évaluations partagent un certain nombre de colonnes en commun. Les colonnes présentes sont :

- **Groupe “sentence” :**

- “id” : l’identifiant de la phrase - présent dans toutes les Dataframes
- “word id” : l’identifiant du mot associé à la phrase - présent dans toutes les Dataframes hypothèses et références
- “word text” : le mot - présent dans toutes les Dataframes hypothèses et références
- “type” : le type du mot - optionnel - présent dans toutes les Dataframes
- “word lemma” : le lemme associé au mot - optionnel - présent dans toutes les Dataframes références lemmes
- “word prefix” : le préfixe du mot - optionnel - présent dans toutes les Dataframes
- “word suffix” : le suffixe du mot - optionnel - présent dans toutes les Dataframes
- “word schema” : le schéma du mot - optionnel - présent dans toutes les Dataframes
- “word split” : le split du mot - optionnel - présent dans toutes les Dataframes

- **Groupe “reference” :**

- “word id” : l’identifiant du mot de la référence - présent dans les Dataframes évaluations
- “word text” : le mot de la référence - présent dans les Dataframes évaluations
- “word lemma” : le lemme de la référence - présent dans les Dataframes évaluations de lemmatisation
- “word id” : l’identifiant du mot de l’hypothèse - présent dans les Dataframes évaluations
- “word text” : le mot de l’hypothèse - présent dans les Dataframes évaluations
- “word lemma” : le lemme de l’hypothèse - présent dans les Dataframes évaluations de lemmatisation

Le groupe “lemma” n’est présent que dans la tâche de lemmatisation.

- **Groupe “lemma” :**

- “id” : l’identifiant du lemme - présent dans les Dataframes évaluations
- “correct” : booléen déterminant si le lemme est correct - présent dans les Dataframes évaluations
- “lemma errors” : le nombre d’erreurs d’annotation de lemmes - présent dans les Dataframes évaluations
- “token errors” : le nombre d’erreur d’annotation de tokens - présent dans les Dataframes évaluations
- “ref count” : le compteur du nombre de références - présent dans les Dataframes évaluations
- “hyp count” : le compteur du nombre d’hypothèses - présent dans les Dataframes évaluations
- “LER” : le LER calculé lors de l’évaluation - présent dans les Dataframes évaluations
- “precision” : la précision calculée lors de l’évaluation - présent dans les Dataframes évaluations
- “rappel” : le rappel calculé lors de l’évaluation - présent dans les Dataframes évaluations
- “f-mesure” : la f-mesure calculée lors de l’évaluation - présent dans les Dataframes évaluations

Le groupe “token” n’est présent que dans la tâche de lemmatisation.

- **Groupe “token” :**

- “id” : l’identifiant du token - présent dans les Dataframes évaluations
- “correct” : booléen déterminant si le token est correct - présent dans les Dataframes évaluations
- “errors” : le nombre d’erreurs d’annotation - présent dans les Dataframes évaluations
- “ref count” : le compteur du nombre de références - présent dans les Dataframes évaluations
- “hyp count” : le compteur du nombre d’hypothèses - présent dans les Dataframes évaluations
- “TER” : le TER calculé lors de l’évaluation - présent dans les Dataframes évaluations
- “precision” : la précision calculée lors de l’évaluation - présent dans les Dataframes évaluations
- “rappel” : le rappel calculé lors de l’évaluation - présent dans les Dataframes évaluations
- “f-mesure” : la f-mesure calculée lors de l’évaluation - présent dans les Dataframes évaluations

5.5 Vérification du locuteur

Pour la tâche d’évaluation de vérification du locuteur, les colonnes présentes sont les suivantes :

- **Groupe “locuteur” :**

- “reference session” : l’identifiant de la référence - présent dans les Dataframes références et évaluations
- “reference locuteur” : le locuteur associé au fichier - présent dans toutes les Dataframes
- “exp id” : l’identifiant du locuteur - présent dans les Dataframes hypothèses et évaluations
- “model” : le modèle utilisé pour reconnaître le locuteur - présent dans les Dataframes hypothèses et évaluations
- “hypothesis” : le locuteur proposé par l’hypothèse - présent dans les Dataframes hypothèses et évaluations
- “score” : le score de confiance fourni par le système - présent dans les Dataframes hypothèses et évaluations
- “type” : le type du locuteur - présent dans les Dataframes évaluations

- “ptarget” : le ptarget associé au locuteur - présent dans les Dataframes évaluations
- “Cllr” : le Cllr associé au locuteur - présent dans les Dataframes évaluations
- **Groupes “eer”, “cdet” et “libre” :**
 - “seuil” : le seuil associé à la métrique - optionnel - présent dans les Dataframes évaluations
 - “decision” : la décision associée à la métrique - optionnel - présent dans les Dataframes évaluations
 - “pcorrect” : le pcorrect associé à la métrique - optionnel - présent dans les Dataframes évaluations
 - “pfa” : le pfa associé à la métrique - optionnel - présent dans les Dataframes évaluations
 - “cdet” : le cdet associé - optionnel - présent dans les Dataframes évaluations

Les trois groupes “eer”, “cdet” et “libre” sont optionnels et dépendent de la paramétrisation de l’évaluation par l’utilisateur.

5.6 Traduction automatique

Pour la tâche d’évaluation de traduction automatique, les colonnes présentes sont les suivantes :

- **Groupe “global” :**
 - “reference name” : le nom de la référence - présent dans toutes les Dataframes
 - “source lang” : la langue du fichier source - présent dans les Dataframes évaluations
 - “target lang” : la langue cible - présent dans les Dataframes évaluations
- **Groupe “turn” :**
 - “id” : l’identifiant du tour/paragraphe - présent dans toutes les Dataframes
 - “raw text” : la phrase associé au tour - présent dans toutes les Dataframes sources, références et hypothèses
 - “source raw text” : le texte source - optionnel - présent dans toutes les Dataframes évaluations
 - “source normalized text” : le texte source normalisé - optionnel - présent dans toutes les Dataframes évaluations
 - “reference raw text” : le texte référence - présent dans toutes les Dataframes évaluations
 - “reference normalized text” : le texte référence normalisé - présent dans toutes les Dataframes évaluations
 - “hypothesis raw text” : le texte hypothèse - présent dans toutes les Dataframes évaluations
 - “hypothesis normalized text” : le texte hypothèse normalisé - présent dans toutes les Dataframes évaluations
- **Groupe “reference” :**
 - “word count” : le nombre de mots dans le texte de référence - présent dans toutes les Dataframes évaluations
- **Groupe “hypothesis” :**
 - “global id” : l’identifiant de l’hypothèse - présent dans toutes les Dataframes évaluations
 - “word count” : le nombre de mots dans l’hypothèse - présent dans toutes les Dataframes évaluations avec la métrique BLEU

- “ngram count order 1 à 4” : le nombre de ngram dans l’hypothèse - présent dans toutes les Dataframes évaluations avec la métrique BLEU

- **Groupe “single bleu” :**

- “correct order 1 à 4” : le nombre de ngram correct dans l’hypothèse - présent dans toutes les Dataframes évaluations
- “precision order 1 à 4” : la précision des ngrams - présent dans toutes les Dataframes évaluations
- “mean precision” : la précision moyenne sur l’ensemble des ngrams - présent dans toutes les Dataframes évaluations
- “best reference” : la référence sur laquelle l’hypothèse obtient le meilleur résultat - présent dans toutes les Dataframes évaluations
- “best reference correct order 1 à 4” : le nombre de ngram correct dans la meilleure référence - présent dans toutes les Dataframes évaluations
- “best reference word count” : le nombre de mots dans la meilleure référence - présent dans toutes les Dataframes évaluations
- “best reference precision order 1 à 4” : la précision des ngrams dans la meilleure référence - présent dans toutes les Dataframes évaluations
- “best reference mean precision” : la précision moyenne sur la meilleure référence - présent dans toutes les Dataframes évaluations
- “score” : le score obtenu sur la meilleure référence - présent dans toutes les Dataframes évaluations

- **Groupe “multi bleu” :**

- “correct order 1 à 4” : le nombre de ngram corrects en prenant toutes les références - présent dans toutes les Dataframes évaluations
- “precision order 1 à 4” : la précision des ngrams en prenant toutes les références - présent dans toutes les Dataframes évaluations
- “mean precision” : la précision moyenne sur l’ensemble des références - présent dans toutes les Dataframes évaluations
- “reference word count” : le nombre de mots sur l’ensemble des références - présent dans toutes les Dataframes évaluations
- “score” : le score BLEU obtenu en prenant en compte toutes les références - présent dans toutes les Dataframes évaluations

Les groupes “single bleu” et “multi bleu” ne sont présents que si la métrique BLEU a été sélectionnée lors de l’évaluation.

- **Groupe “ter” :**

- “slot id” : l’identifiant d’alignement entre la référence et l’hypothèse - présent dans toutes les Dataframes évaluations
- “hypothesis word id” : l’identifiant des mots de l’hypothèse - présent dans toutes les Dataframes évaluations
- “hypothesis word” : les mots de l’hypothèse - présent dans toutes les Dataframes évaluations
- “reference word id” : les identifiants de la référence - présent dans toutes les Dataframes évaluations
- “reference word” : les mots de la référence - présent dans toutes les Dataframes évaluations
- “correct count” : le nombre de mots correctement traduits - présent dans toutes les Dataframes évaluations

- “substitution count”, “insertion count”, “deletion count” et “displacement count” : les compteurs des différentes opérations associées à TER - présent dans toutes les Dataframes évaluations
- “error count” : le nombre d’erreurs lors de la traduction - présent dans toutes les Dataframes évaluations
- “single reference score” : le score TER obtenu par référence - présent dans toutes les Dataframes évaluations
- “best reference name” : la meilleure référence sur TER - présent dans toutes les Dataframes évaluations
- “best reference error count” : le nombre d’erreurs sur la meilleure référence - présent dans toutes les Dataframes évaluations
- “best reference word count” : le nombre de mots de la meilleure référence - présent dans toutes les Dataframes évaluations
- “best reference score” : le score TER obtenu par la meilleure référence - présent dans toutes les Dataframes évaluations

Le groupe “ter” n’est présent que si la métrique TER a été sélectionnée lors de l’évaluation.

INTERFACE PRINCIPALE

Une fois EVALOMATIC lancé, l'interface principale apparaît, comme illustré par *Tableau de bord*. Elle se compose du tableau de bord dans la zone centrale et d'un menu et de trois icônes dans la barre des tâches. Nous présentons d'abord le tableau de bord, puis les fonctionnalités présentes dans la barre des tâches.

6.1 Tableau de bord

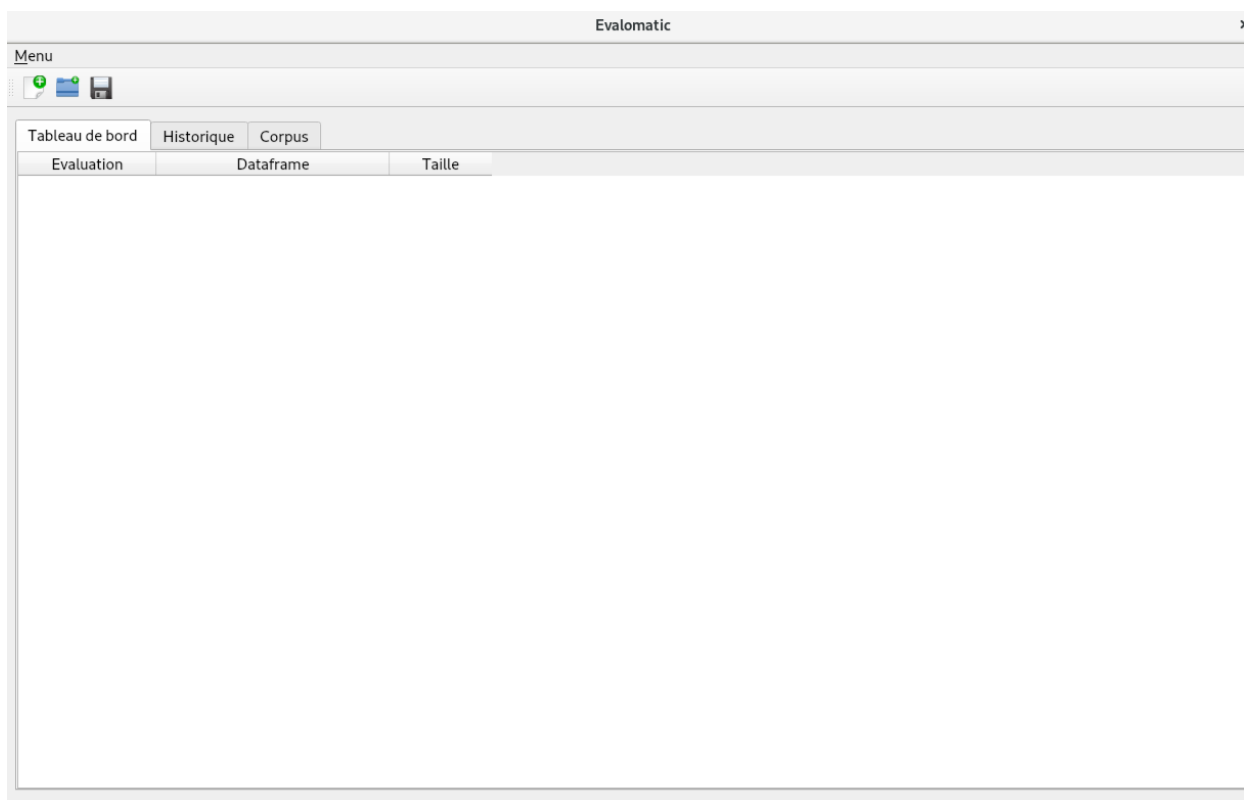


Fig. 1: Tableau de bord

Ce tableau de bord contient 3 onglets :

- L'espace de travail
- Les historiques
- Les corpus

6.1.1 L'espace de travail

L'espace de travail présente l'ensemble des Evaluations et Dataframes actuellement ouvertes par l'utilisateur comme le montre *Espace de travail : liste des dataframes actuellement ouvertes*. Il est à l'ouverture de EVALOMATIC vide et se remplit au fur et à mesure des manipulations.

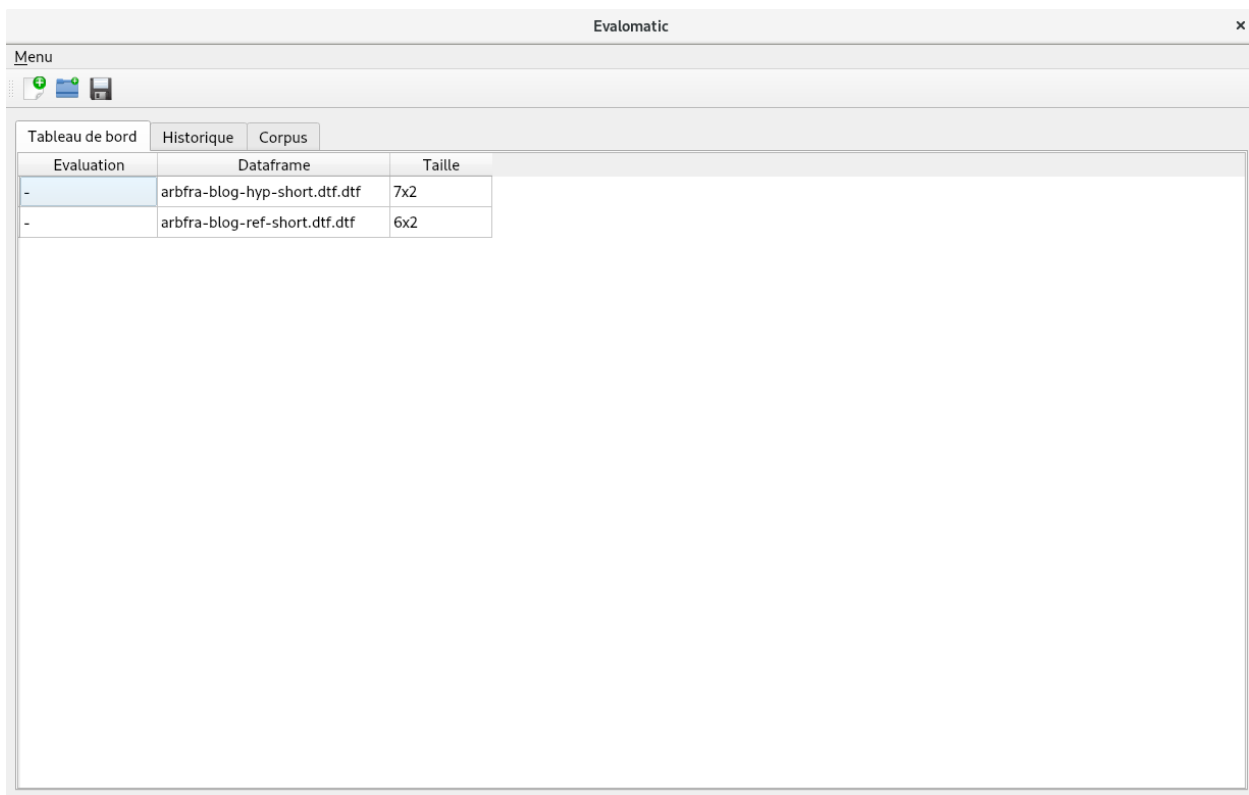


Fig. 2: Espace de travail : liste des dataframes actuellement ouvertes

6.1.2 Les historiques

L'onglet historique permet d'accéder à la liste des Evaluations et Dataframes que l'utilisateur a sauvegardées comme illustré par *Exploiter les historiques*. Comme pour l'espace de travail, l'utilisateur peut ouvrir une Dataframe en double-cliquant dessus.

6.1.3 Les corpus

L'onglet corpus permet d'accéder à la liste des corpus que l'utilisateur à récupérer depuis DATOMATIC. C'est à partir de ces corpus que l'utilisateur peut effectuer des nouvelles évaluations et explorations. Les informations suivantes sont accessibles depuis l'onglet corpus comme illustré par *Accéder à la liste des corpus récupérés*.

6.2 Barre des tâches

Le menu déroulant ainsi que les icônes de la barre des tâches correspondent aux trois mêmes fonctionnalités, le menu permettant en plus de quitter EVALOMATIC. Nous pouvons lister trois fonctionnalités :

The screenshot shows the 'Evalomatic' application window with the 'Historique' tab selected. The table below shows the data for the 'Trad Eval' evaluation.

Évaluation	Type	Références	Hypothèses	Normalisation	Date création	Date modification
Trad Eval	traduction	arbfra-blog-ref-short.dtf.dtf	arbfra-blog-hyp-short.dtf.dtf		2019-01-09	2019-01-09

Fig. 3: Exploiter les historiques

The screenshot shows the 'Evalomatic' application window with the 'Corpus' tab selected. The table below displays the list of corpora and their associated statistics.

Nom	Format	Langue	Type d'annotation	Durée du signal	Durée parole	Durée exclue	Nombre de mots
Repere source	wav	-	-	51390,4	0	0	0
Repere references	-	fre	trans	0	9761,4	40775,1	34940
Repere hypotheses SODA	-	fre	trans	0	11149	0	145335
Repere hypotheses QCOMPERE	-	fre	trans	0	36357,4	0	170699
Repere hypotheses PERCOL	-	fre	trans	0	172353	0	667620

Fig. 4: Accéder à la liste des corpus récupérés

- *Lancer une nouvelle évaluation*
- *Charger des Dataframes ou des fichiers annotations*
- *Sauvegarder la session*

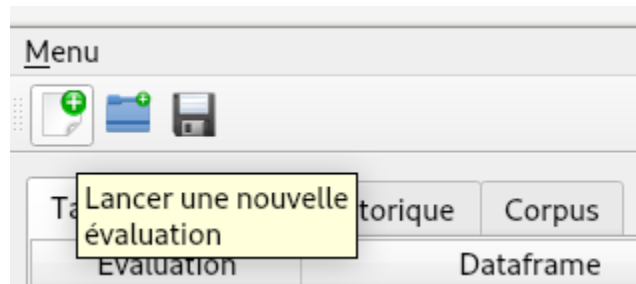


Fig. 5: Lancer une nouvelle évaluation

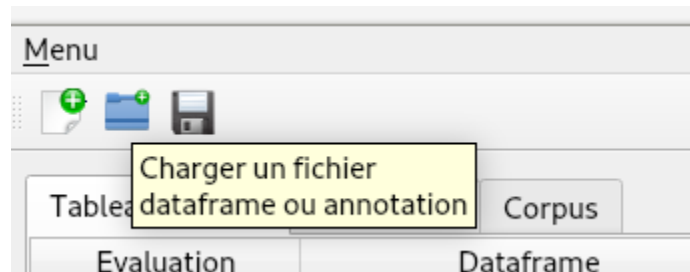


Fig. 6: Charger des Dataframes ou des fichiers annotations

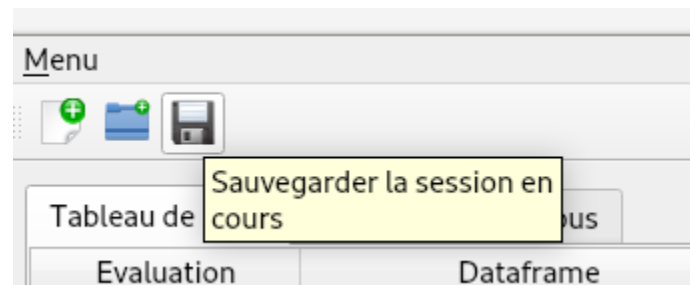


Fig. 7: Sauvegarder la session

Les deux premières fonctionnalités sont présentées en détail dans les deux parties suivantes. La sauvegarde de la session permet qu'au prochain lancement d'EVALOMATIC l'ensemble des Dataframes créées à la suite des manipulations soient présentes dans le tableau de bord pour être directement utilisables. Le menu déroulant, qui contient les mêmes fonctionnalités, est illustré dans la capture d'écran *Menu de l'interface principale*

Dans les parties suivantes, nous allons commencer par présenter le lancement d'une évaluation dans le chapitre *Créer une évaluation*. Puis nous allons décrire le fonctionnement du chargement des Dataframes et fichiers d'annotation dans le chapitre *Charger des fichiers*. Enfin, ces deux fonctionnalités conduisant à la création de nouvelles Vues de Dataframes, nous décrivons dans le chapitre *Explorer une Vue d'une Dataframe* la manipulation des Vues.

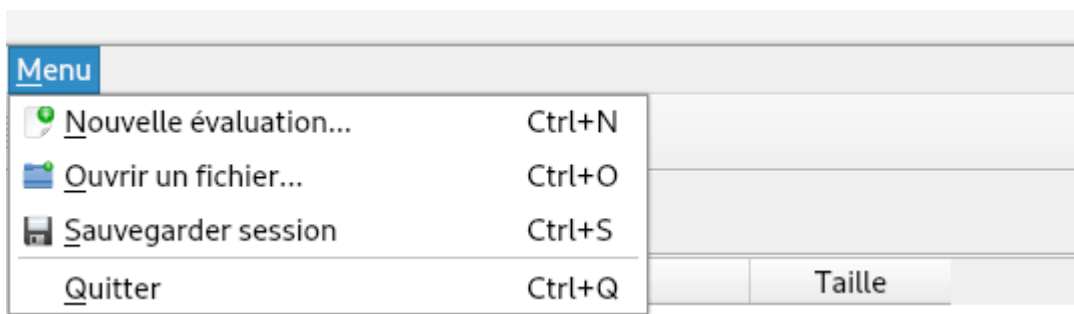


Fig. 8: Menu de l'interface principale

CRÉER UNE ÉVALUATION

7.1 Procédure de création d'une évaluation

Lors de l'activation de la fonctionnalité de création d'évaluation par le biais de l'interface principale, un menu de configuration de l'évaluation va être lancé. Ce menu permet de renseigner différentes informations, dont certaines sont obligatoires, et d'autres dépendent de la tâche sélectionnée. Nous listons ci-dessous les informations que l'on retrouve quelque soit l'évaluation.

- Nom de l'évaluation
- Tâche
- Langue
- Corpus de sources (optionnel)
- Corpus de références
- Corpus d'hypothèses

Un exemple du formulaire d'évaluation pour la tâche de lemmatisation est illustrée dans l'image eval-lemme. L'évaluation en lemmatisation ne contient que des champs généralistes.

Ces champs sont décrits ci-dessous.

7.2 Champs généraux

7.2.1 Nom de l'évaluation

Le nom de l'évaluation est un champ texte libre. L'utilisateur doit le renseigner pour lancer une évaluation. Si le nom de l'évaluation est déjà utilisé, EVALOMATIC ajoute automatiquement un nombre derrière afin de ne pas générer de doublon.

7.2.2 Tâche

EVALOMATIC est actuellement capable d'effectuer des évaluations pour les tâches de transcription de la parole, de traduction, d'entités nommées, de lemmatisation, de tokenisation et de vérification du locuteur. Comme EVALOMATIC est modulaire, il sera bientôt possible de choisir d'autres tâches. L'utilisateur doit donc sélectionner via une liste de choix, la tâche voulue pour effectuer une évaluation sur ce type.

Nouvelle évaluation

Nom de l'évaluation

Tâche

Langue

Corpus source

Corpus de référence

Corpus d'hypothèse

Autres paramètres :

Métriques :

Fig. 1: Formulaire de lancement d'une évaluation en lemmatisation

7.2.3 Langue

Le champ langue permet de sélectionner les corpus relatifs à une langue choisie. Par défaut, l'utilisateur peut choisir d'afficher les corpus pour toutes les langues. La liste de choix des langues dépend des corpus récupérés via DATOMATIC ou des fichiers annotations chargés manuellement dans EVALOMATIC. Ce choix de langue conditionne les listes de corpus sélectionnable dans les champs Corpus de référence et d'hypothèse.

7.2.4 Corpus de sources

Les corpus de sources peuvent être choisis par les corpus récupérés via DATOMATIC ou dans les fichiers chargés manuellement dans EVALOMATIC. Les corpus sources sont optionnels pour lancer une évaluation.

7.2.5 Corpus de références

Les corpus de référence peuvent être choisis parmi les corpus qui ont été récupérés via DATOMATIC ou dans les fichiers annotations chargés manuellement dans EVALOMATIC, à condition que ces corpus soient d'un type attendu par la tâche choisie. Si une langue a été préalablement sélectionnée, seuls les corpus dans cette langue apparaissent. Actuellement, on ne peut ajouter qu'un seul corpus de référence.

7.2.6 Corpus d'hypothèses

Les corpus d'hypothèse peuvent être choisis parmi les corpus qui ont été récupérés via DATOMATIC ou de fichiers annotations chargés manuellement dans EVALOMATIC. L'utilisateur peut charger autant de corpus hypothèses qu'il en a besoin. Pour ajouter un corpus, il suffit de cliquer sur l'icône +.

7.3 Champs spécifiques à la tâche

Selon la tâche, certains champs sont ajoutés.

7.3.1 Evaluation de traduction

Pour l'évaluation de traduction illustrée dans l'image eval-trad, l'utilisateur peut éditer les champs suivants :

- Ajouter un script pour normaliser les références
- Choisir un beam de référence pour la métrique TER, c'est à dire la taille de la distance d'édition
- Choisir quelles métriques doivent être calculées - TER et/ou BLEU

7.3.2 Evaluation en entités nommées

L'évaluation en entités nommées est illustrée dans l'image *Formulaire de lancement d'une évaluation en entités nommées*, l'utilisateur peut éditer les champs suivants :

- Script de mapping des entités nommées de la référence
- Script de mapping des entités nommées des hypothèses
- Configuration de la métrique SER

Pour l'évaluation des entités nommées, on utilise la métrique SER, que l'on peut configurer sur les différents coûts associés à la métrique : l'insertion, la suppression, l'erreur de type, et l'erreur de frontière.

Nouvelle évaluation

Nom de l'évaluation	<input type="text" value="Evaluation TRAD"/>
Tâche	<input type="text" value="Traduction"/>
Langue	<input type="text" value="Toutes"/>
Corpus source	<input type="text" value="Aucun"/>
Corpus de référence	<input type="text" value="arbfra-blog-hyp-short.dtf.dtf"/>
Corpus d'hypothèse	<input type="text" value="arbfra-blog-hyp-short.dtf.dtf"/> <input style="float: right;" type="button" value="+"/>

Autres paramètres :

Normalisation références (script) :

TER - Taille de beam :

Métriques : BLEU TER

Fig. 2: Formulaire de lancement d'une évaluation en traduction

Nouvelle évaluation

Nom de l'évaluation

Tâche

Langue

Corpus source

Corpus de référence

Corpus d'hypothèse

Autres paramètres :

Mapping types référence :

Mapping types hypothèse :

SER - Coût d'insertion (fausse alarme) :

SER - Coût de deletion (miss) :

SER - Coût d'erreur de type :

SER - Coût d'erreur de frontière :

Métriques : Force SER

Fig. 3: Formulaire de lancement d'une évaluation en entités nommées

7.3.3 Evaluation en transcription de la parole

Le formulaire d'évaluation en transcription de la parole est illustré dans l'image *Formulaire de lancement d'une évaluation en transcription de la parole*. L'utilisateur peut éditer les champs suivants :

- Script de normalisation de la référence
- Script nglm de normalisation des hypothèses
- Choix des métriques (WER, CER et NCE), prise en compte de la casse, et leur configuration

Nouvelle évaluation

Nom de l'évaluation : Evaluation Transcription

Tâche : Transcription

Langue : Toutes

Corpus source : Repere - Repere source

Corpus de référence : Repere - Repere references

Corpus d'hypothèse : Repere - Repere hypotheses PERCOL (+)
 Repere - Repere hypotheses SODA (-)
 Repere - Repere hypotheses QCOMPERE (-)

Autres paramètres :

Normalisation références (script) : [Parcourir...](#)

Normalisation hyp/ref (nglm) : [Parcourir...](#)

Métriques :

Sans casse Avec casse

WER Insertions

CER Omissions

NCE Substitutions

Fig. 4: Formulaire de lancement d'une évaluation en transcription de la parole

L'utilisateur peut proposer un script de normalisation des références. Par défaut, EVALOMATIC applique la règle selon laquelle un mot est une suite de caractères séparés par un blanc. Le script est appliqué par EVALOMATIC tel quel. Il est à noter que la syntaxe appliquée par EVALOMATIC est celle de la lecture des trs et des stm, ainsi, les { } sont interprétés comme présentant des alternatives dans les références et les ' permettent de protéger un caractère.

Le script de normalisation des hypothèses doit suivre les règles de format du nglm. Le nglm permet de substituer certains mots des hypothèses afin de pallier des alternatives graphiques. Il s'agit de transformer les réponses des hypothèses pour prendre des formes gérables vis-à-vis des références. Le nglm ne s'applique qu'à des mots.

Les métriques disponibles pour la transcription sont le CER, le WER, NCE et le nombre de mots ou caractères substitués, le nombre de mots ou caractères oubliés, le nombre de mots ou caractères ajoutés. Il est à noter que l'utilisateur

choisit soit d'estimer la qualité d'une transcription en CER soit en WER. En effet, l'alignement de l'hypothèse et de la référence est effectué soit au mot soit au caractère et l'utilisateur doit faire un choix.

7.3.4 Evaluation en vérification du locuteur

Le formulaire d'évaluation en transcription de la parole est illustré dans l'image *Formulaire de lancement d'une évaluation en vérification du locuteur*. L'utilisateur peut éditer les champs suivants :

- Configuration des coûts de miss, de fausse alarme, de Ptarget ainsi que le seuil libre
- Choix des métriques
- Estimation du Ptarget

Nouvelle évaluation

Nom de l'évaluation

Tâche

Langue

Corpus source

Corpus de référence

Corpus d'hypothèse

Autres paramètres :

Cdet - Coût miss :

Cdet - Coût fausse alarme :

Cdet - Ptarget :

Seuil libre :

Métriques :

Seuil EER

Seuil Cdet Estimer Ptarget

Seuil libre

Fig. 5: Formulaire de lancement d'une évaluation en vérification du locuteur

Pour la vérification du locuteur, les métriques utilisées sont les seuils EER et Cdet, et une estimation du Ptarget. Il faut pour cela configurer les coût de miss et de fausse alarme, ainsi que le seuil libre.

7.3.5 Evaluation en lemmatisation et en tokenisation

Contrairement aux autres évaluations, ces deux tâches ne demandent pas de champs supplémentaires. Elles utilisent respectivement comme métriques le LER (Lemmatisation Error Rate) et le TER (Tokenisation Error Rate).

7.4 Lancement de l'évaluation

Une fois les champs renseignés, l'utilisateur clique sur lancer pour que l'évaluation commence. Une fois l'évaluation effectuée, les résultats de l'évaluation apparaissent sous la forme d'une nouvelle Vue. Cette Vue est manipulable comme décrit en *Explorer une Vue d'une Dataframe*.

CHARGEMENT DE FICHIERS ANNOTATIONS OU DATAFRAMES

Lors de l'activation de la fonctionnalité de chargement de fichiers par l'utilisateur, un explorateur de fichiers classiques va s'ouvrir. Il est alors possible pour l'utilisateur de choisir un à plusieurs fichiers annotations ou Dataframes à ouvrir dans EVALOMATIC. Il est recommandé que ces fichiers fassent tous partis de la même tâche d'évaluation, et du même type de corpus.

8.1 Formulaire de sélection des formats des fichiers

Une fois ces fichiers sélectionnés, exceptés si les fichiers ne sont composés uniquement de fichiers au format Dataframe (généralement reconnu à leur extension .dtf), une fenêtre de sélection du format des fichiers va s'ouvrir. Cette fenêtre est illustrée dans l'image *Formulaire de sélection des formats*.

EVALOMATIC essaye de détecter automatiquement le type de chacun des fichiers choisis par l'utilisateur. Dans l'image *Formulaire de sélection des formats*, les trois fichiers ont été identifiés comme étant des fichiers trans/trs par EVALOMATIC. Néanmoins, l'utilisateur peut choisir de forcer un type en particulier sur les fichiers sélectionnés. Généralement, cela entraînera une erreur de parsing, qui sera remontée dans le formulaire. Dans l'image *Formulaire de sélection incorrect des formats*, on peut voir que le fait de forcer le troisième fichier en "trans/stm" provoque une erreur qui est affichée dans la zone de texte.

Une fois les choix de formats validés par l'utilisateur, une nouvelle fenêtre va s'ouvrir : la fenêtre de paramétrage des fichiers.

8.2 Formulaire de paramétrage des fichiers

Lorsqu'EVALOMATIC parse un fichier pour en extraire les informations, certaines informations nécessaires au fonctionnement du logiciel ne sont pas présentes dans le fichier. Ces informations dépendent du type de corpus (référence, hypothèse) ainsi que de la tâche associée au fichier. Pour pouvoir remplir ces informations, une fenêtre de paramétrage s'ouvre une fois les formats validés : l'utilisateur doit alors renseigner différents champs. L'image *Formulaire de paramétrage* illustre la fenêtre générée à la suite des fichiers ouverts dans la partie précédente.

La fenêtre étant organisée en arborescence, il est possible de propager un paramètre sur l'ensemble de fichiers à partir d'une racine. Nous donnons une liste des types de paramètres ci-dessous :

- file : permet de renommer le fichier - particulièrement important si les hypothèses n'ont pas le même nom que la référence ;
- langue : permet de sélectionner la langue du corpus de fichiers ;
- système : présent seulement dans les fichiers de corpus hypothèse - permet de donner un nom aux systèmes présents dans les fichiers ;

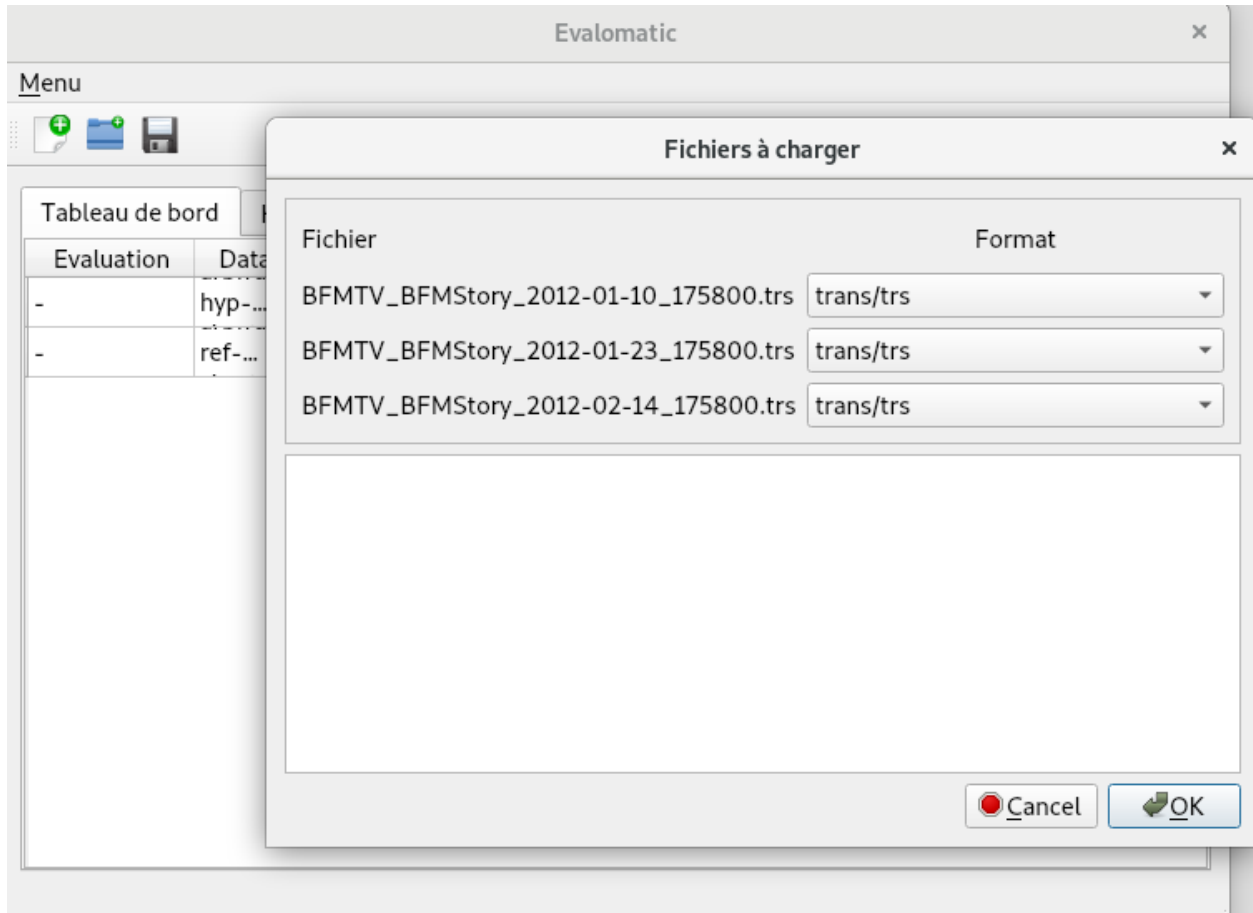


Fig. 1: Formulaire de sélection des formats

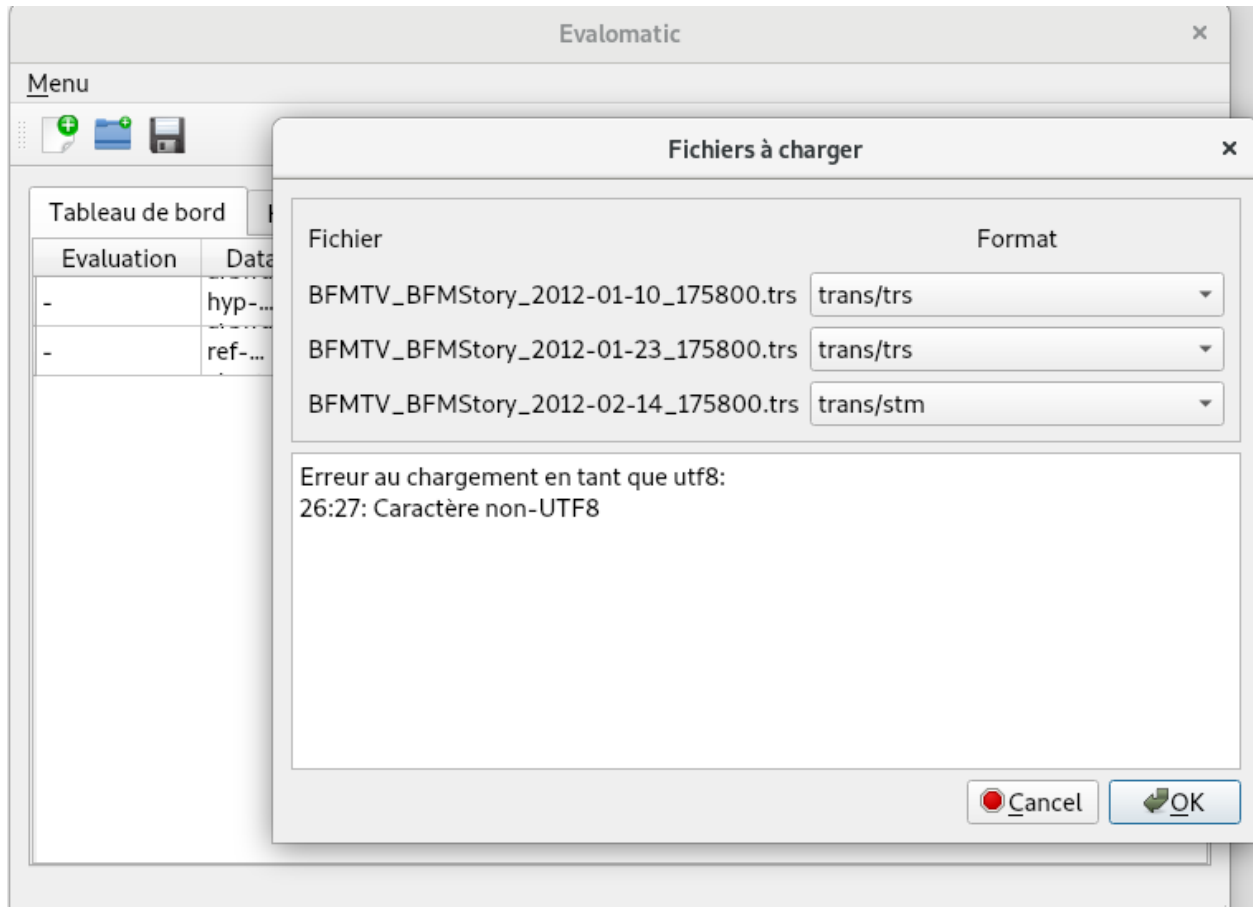


Fig. 2: Formulaire de sélection incorrect des formats

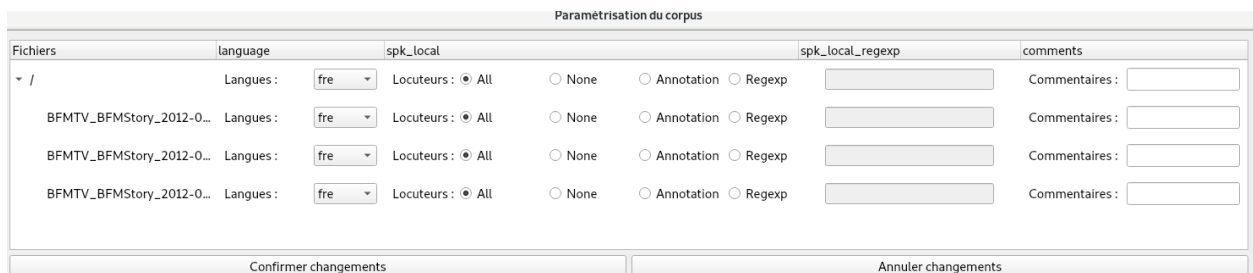


Fig. 3: Formulaire de paramétrage

- système version : est toujours présent si le paramètre “système” doit être rentré - permet de spécifier une version aux systèmes présents ;
- type : permet de définir si les fichiers sont des références ou des hypothèses ;
- comments : permet d’ajouter un commentaire à chaque fichier ;
- spk_local et spk_local_regexp : permet de définir l’approche à utiliser par aux locuteurs présent d’un tour de parole à un autre - ces paramètres ne sont présents que pour les évaluations de transcription de la parole ;

Une fois les paramètres entrés, la Dataframe correspondante est créée ainsi qu’une Vue associée, que l’utilisateur peut explorer comme décrit dans le chapitre *Explorer une Vue d’une Dataframe*.

EXPLORER UNE VUE D'UNE DATAFRAME

La Vue d'une Dataframe peut être explorée dans deux cas de figure : soit après le lancement d'une évaluation, où plusieurs Dataframes seront créées, notamment la Dataframe résultat. Soit lors de l'ouverture d'un ou plusieurs fichiers annotations ou Dataframe manuellement. Dans ces deux cas, l'utilisateur peut effectuer un certain nombre d'opérations permettant de manipuler les données affichées dans la Vue. Nous avons présenté les concepts inhérents à ces opérations dans le chapitre *Concepts*. Dans le présent chapitre nous allons illustrer toutes ces opérations. Pour rappel, les opérations actuellement possibles sont les suivantes :

- Duplication de la Vue
- Visibilité des colonnes
- Tri des colonnes
- Filtrage des lignes
- Ajout d'une colonne calculée à partir d'une Expression
- Manipulation des en-têtes

Pour l'illustration des exemples, nous nous appuyons sur la Vue illustrée en *Vue exemple*. Cette Vue est celle générée à la suite d'une évaluation en transcription de la parole sur trois consortiums ayant participé au challenge REPERE. Ainsi, la majorité des colonnes sont cachées dans l'état actuel de cette Vue.

Outre les opérations de la Vue, nous présenterons aussi deux fonctions usuelles de la Vue, la sauvegarde et l'historique.

9.1 Opérations de la Vue

9.1.1 Visibilité des colonnes

A partir d'un clic sur le bouton représenté dans l'image *Illustration du bouton de visibilité des colonnes*, l'utilisateur peut ouvrir une fenêtre lui permettant de d'afficher ou de masquer les colonnes de la Vue de son choix.

Cette fenêtre est représentée dans l'image *Exemple de la fenêtre de gestion de la visibilité des colonnes*.

En cliquant sur les checkbox, on peut choisir quelles colonnes sont visibles ou non. Les colonnes en gras correspondent aux colonnes de type Label, c'est à dire celles dont la visibilité peut avoir un impact sur l'état de la Vue. Le fait de cliquer sur les colonnes "global.file" et "speaker.name" va modifier leur visibilité, comme illustré dans l'image fenetre-visibilite2.

Comme présenté dans le chapitre *Concepts*, le fait de masquer ou de rendre visible certaines colonnes peut entraîner une modification de la Vue. Ainsi, dans l'image exemple-visibilite3, le fait de rendre visible les colonnes "global.file" et "speaker.name" annule certaines fusions présentes entre les lignes de la Vue en faisant apparaître les fichiers et les locuteurs.

global.system	global.system version	wer ci.wer ci	wer ci.nce
SODA	Kaldi	19,0431	-inf
	Sphinx	17,352	-inf
PERCOL	Chouchou	31,2414	
SODA	Cheated	15,9147	-inf
QCOMPERE	Primary	28,2534	0,0662861
SODA		16,166	-inf
PERCOL	Oreillettes	27,5699	
	Tartatatin	31,4089	
	Primary	28,1345	

Fig. 1: Vue exemple

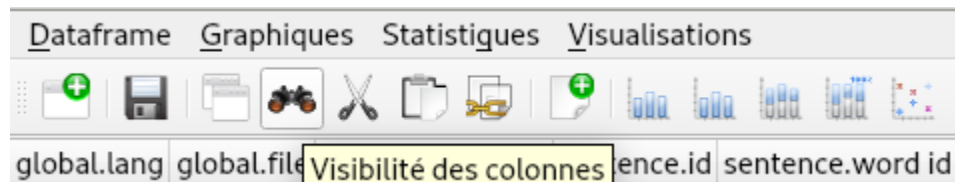


Fig. 2: Illustration du bouton de visibilité des colonnes

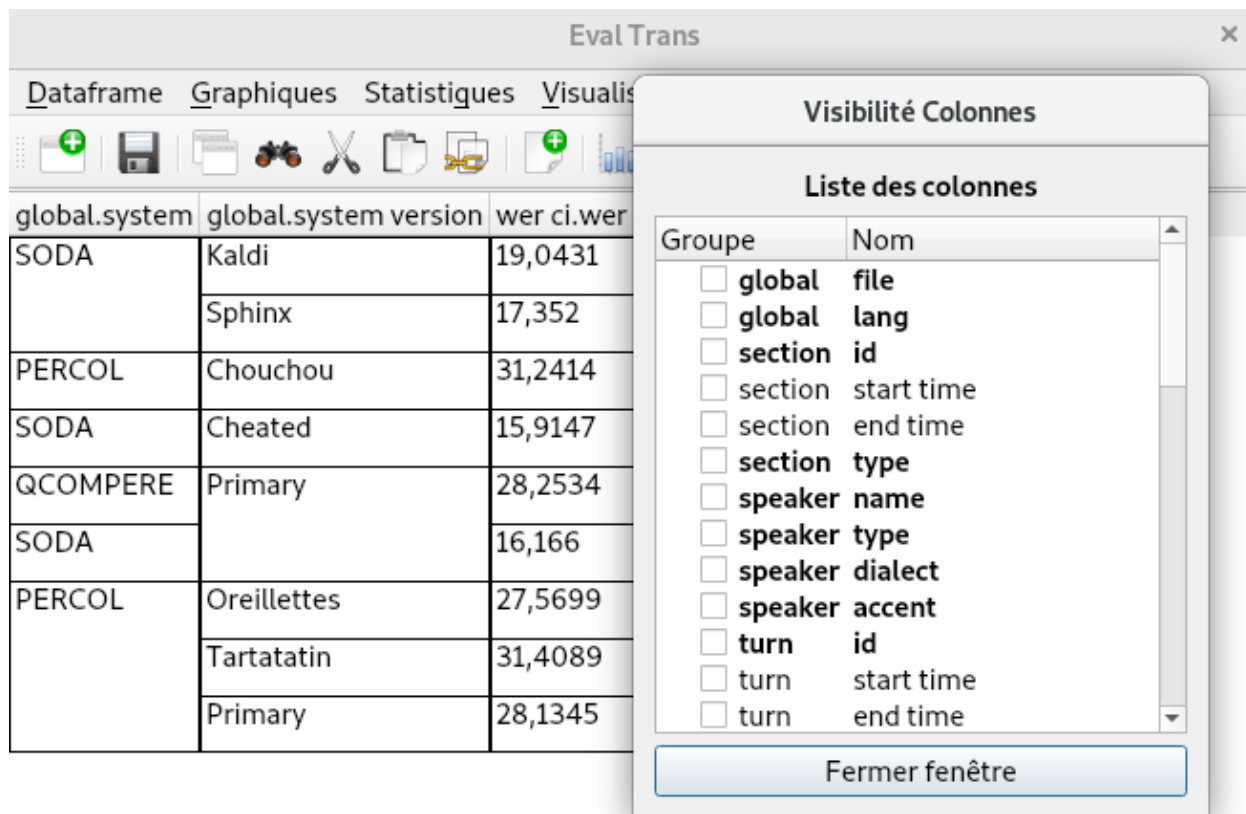


Fig. 3: Exemple de la fenêtre de gestion de la visibilité des colonnes

9.1.2 Duplication de la Vue

Il y a deux façons de dupliquer la Vue : soit à partir du tableau de bord, soit directement dans la fenêtre de la Vue elle même. Pour dupliquer à partir du tableau de bord, il suffit de double cliquer sur la Dataframe correspondante dans la liste du tableau de bord, ou alors de faire un clic droit et de sélectionner “Voir”, comme cela est présenté dans l’image *Exemple du menu de duplication de la Vue*.

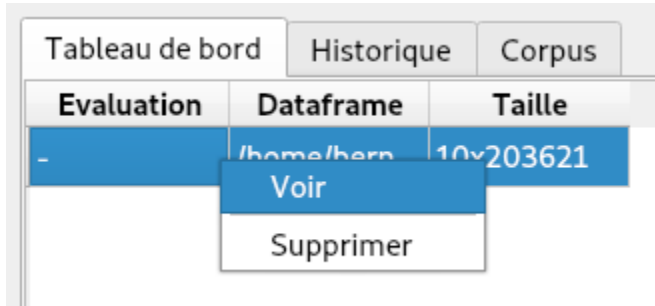


Fig. 4: Exemple du menu de duplication de la Vue

La Vue nouvellement créée représentera l’ensemble du contenu de la Dataframe, comme illustré dans l’image *Exemple du bouton de duplication de la Vue*.

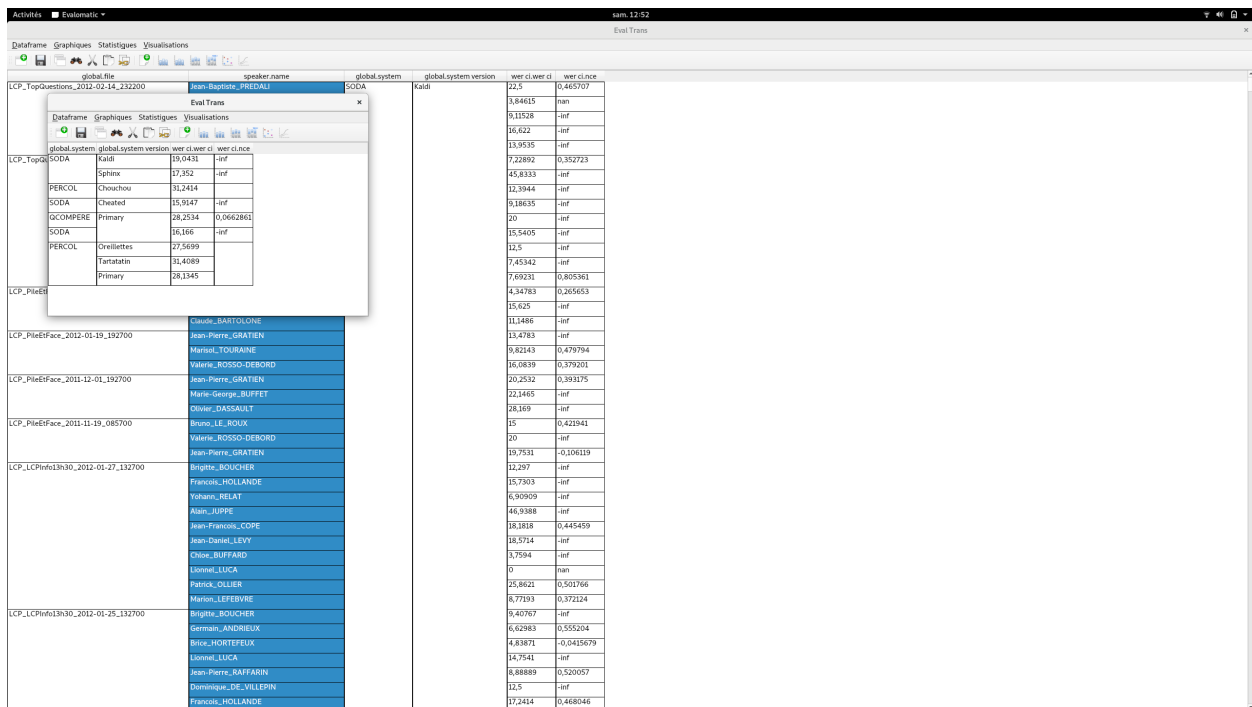


Fig. 5: Exemple du bouton de duplication de la Vue

L’autre manière de dupliquer la Vue est d’appuyer sur le bouton représenté dans l’image *Exemple du bouton de duplication de la Vue*.

La Vue créé lors de cette duplication aura la même forme que la Vue originelle. Ainsi, les deux colonnes “global.file” et “speaker.name” qui ont été rendues visibles précédemment sont visibles dans dans cette Vue, comme illustré dans l’image *Exemple du bouton de duplication de la Vue*.

9.1.3 Tri des colonnes

Pour l'exemple du tri, nous revenons à l'état initial de la Vue. Le tri des colonnes est accessible à partir du bouton représenté dans l'image *Illustration du bouton de tri*.

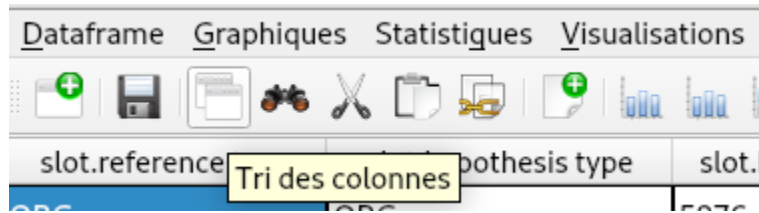


Fig. 8: Illustration du bouton de tri

Une pression de ce bouton ouvre alors la fenêtre de tri, représentée dans l'image *Illustration de la fenêtre de tri*.

Dans cet exemple, on désire trier pour chaque consortium les versions de systèmes proposés en fonction de la métrique WER. Ainsi, on effectue un tri multiple où l'on sélectionne d'abord la colonne "global.system" puis la colonne "wer ci. wer ci", comme illustré dans l'image *Illustration des colonnes de tri*. On choisit par ailleurs de prendre un ordre croissant.

Une fois le tri paramétré, la validation de la sélection entraîne la réorganisation des lignes de la Vue selon le tri sélectionné, comme illustré dans l'image *Résultat du tri*.

9.1.4 Filtrage des lignes

Pour illustrer le fonctionnement du filtrage des lignes, on ouvre une Vue dans l'état initial, on ouvre la colonne "global.file" et on ferme la colonne "wer ci.nce". L'état de la Vue est illustré dans l'image *explo-filtrage1.png*.

La fenêtre de filtrage des lignes de la Vue est accessible à partir du bouton représenté dans l'image *Illustration du bouton de filtrage*.

La fenêtre de filtrage des lignes est illustrée dans l'image *Fenêtre de tri*. Une opération de filtrage s'appuie sur un filtre, dépendant du type de la colonne sur laquelle l'utilisateur souhaite l'appliquer.

Dans cet exemple, on souhaite tout d'abord filtrer les fichiers sur lesquels les systèmes ont obtenu plus de 60 de WER. On sélectionne donc la colonne "wer ci. wer ci", on choisit l'opérateur ">=" et on entre la valeur "60", comme illustré dans l'image *Application d'un tri sur le WER*.

Le tri est effectué et l'état de la Vue modifiée, comme illustré dans l'image *Etat de la Vue suite au tri sur le WER*.

On souhaite alors enlever les lignes contenant les fichiers "LCP" et "BFM...075000". On ouvre de nouveau la fenêtre de filtrage, et on sélectionne la colonne "global.file". La colonne étant de type "Lab+Str", l'utilisateur peut choisir d'utiliser des expressions régulières, ou alors de sélectionner les labels à filtrer. Nous illustrons cette opération à partir de la deuxième possibilité, tel qu'illustré dans l'image *Application d'un tri sur les fichiers*.

Les checkbox permettent de sélectionner les labels à garder. Le résultat de ce filtrage est visible dans l'image *Etat de la Vue suite au tri sur les fichiers*.

9.1.5 Manipulation des en-têtes

Les fonctionnalités de manipulation des en-têtes sont accessibles à partir du bouton *Illustration du bouton de manipulation des en-têtes*. Pour cet exemple, on utilise la Vue obtenue à la fin des exemples illustrant le filtrage.

La fenêtre permettant de manipuler les en-têtes est illustrée dans l'image *Fenêtre de manipulation des en-têtes*. La fenêtre est composée d'une partie supérieure représentant l'ensemble des colonnes de la Vue regroupées par groupe.

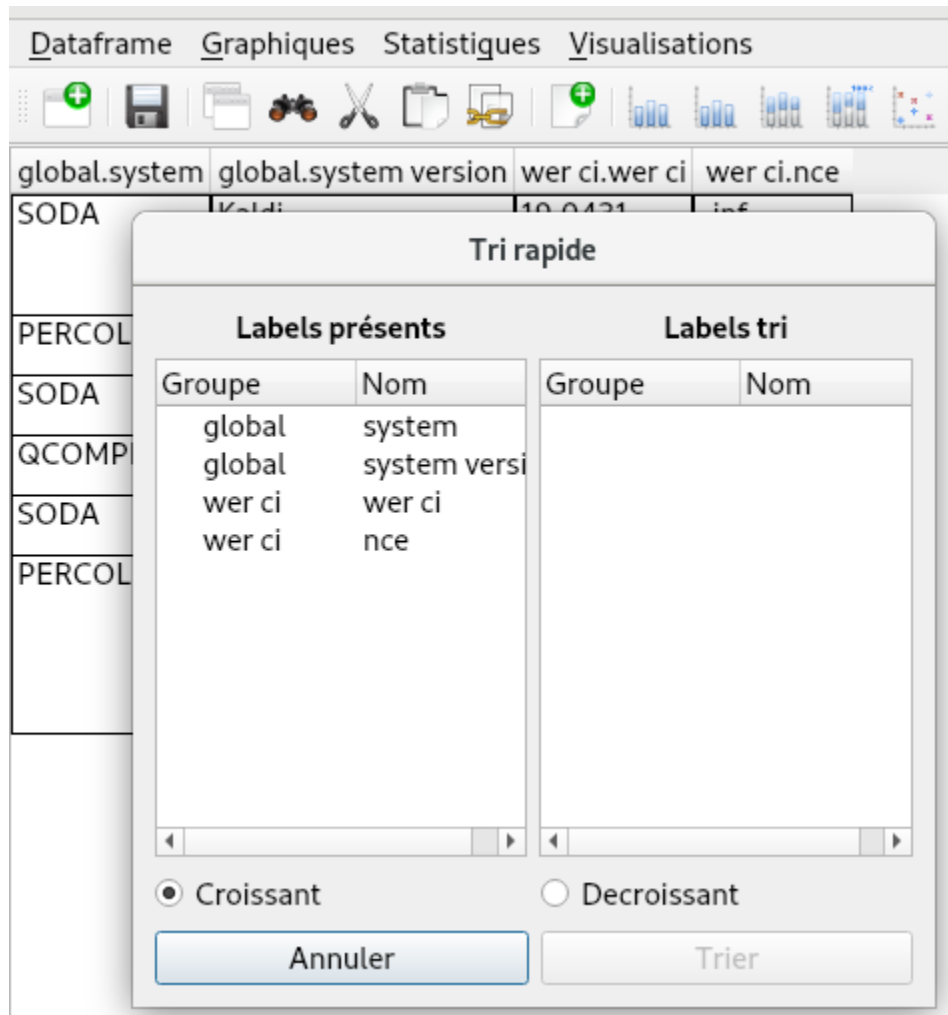


Fig. 9: Illustration de la fenêtre de tri

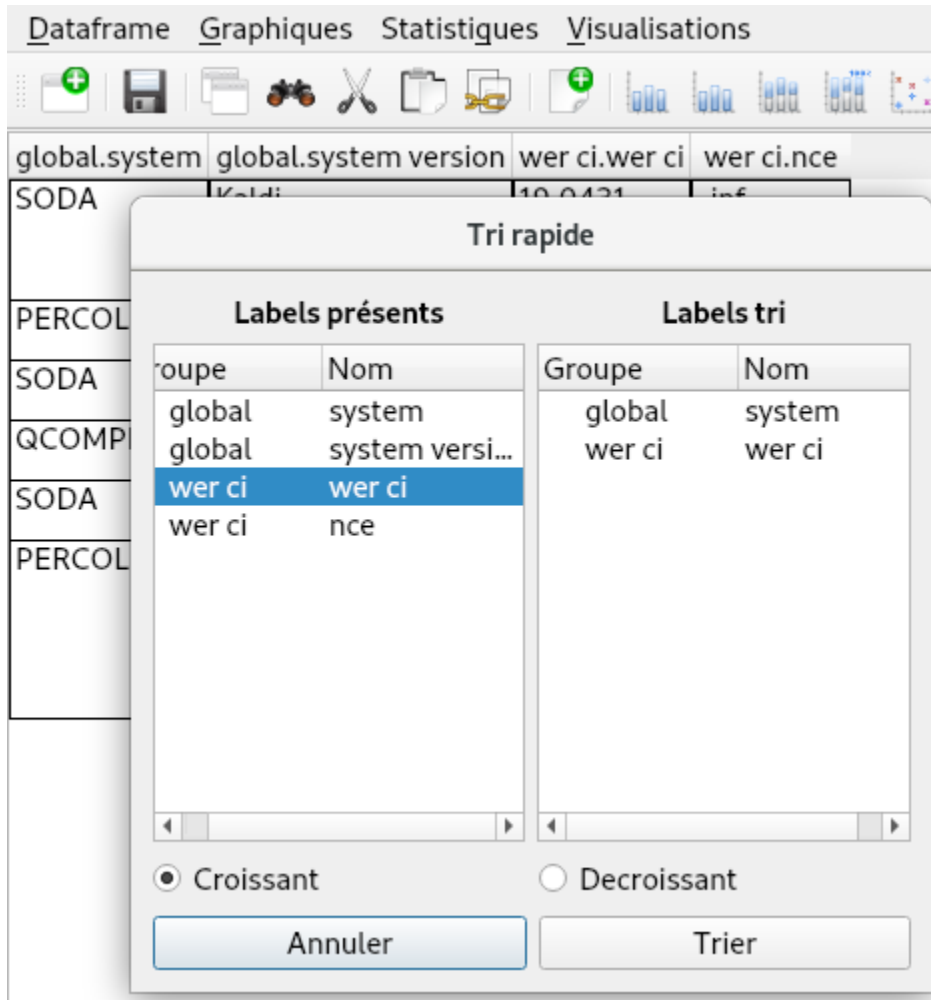


Fig. 10: Illustration des colonnes de tri

Dataframe Graphiques Statistiques Visualisations			
global.system	global.system version	wer ci.wer ci	wer ci.nce
PERCOL	Oreillettes	27,5699	
	Primary	28,1345	
	Chouchou	31,2414	
	Tartatatin	31,4089	
QCOMPERE	Primary	28,2534	0,0662861
SODA	Cheated	15,9147	-inf
	Primary	16,166	-inf
	Sphinx	17,352	-inf
	Kaldi	19,0431	-inf

Fig. 11: Résultat du tri

Dataframe Graphiques Statistiques Visualisations			
global.file	global.system	global.system version	wer ci.wer ci
LCP_TopQuestions_2012-02-14_232200	SODA	Kaldi	13,6167
LCP_TopQuestions_2012-01-25_232700			12,6851
LCP_PileEtFace_2012-01-26_192700			13,0148
LCP_PileEtFace_2012-01-19_192700			12,7303
LCP_PileEtFace_2011-12-01_192700			22,5237
LCP_PileEtFace_2011-11-19_085700			18,1953
LCP_LCPInfo13h30_2012-01-27_132700			15,5534
LCP_LCPInfo13h30_2012-01-25_132700			15,2299
LCP_LCPInfo13h30_2012-01-24_132700			11,1418
LCP_EntreLesLignes_2012-05-11_1922...			15,0558
LCP_PileEtFace_2012-01-12_192700			22,1875
BFMTV_CultureEtVous_2012-01-16_0...			46,7633
LCP_TopQuestions_2012-02-22_232700			9,70043

Fig. 12: Etat de la Vue pour le filtrage

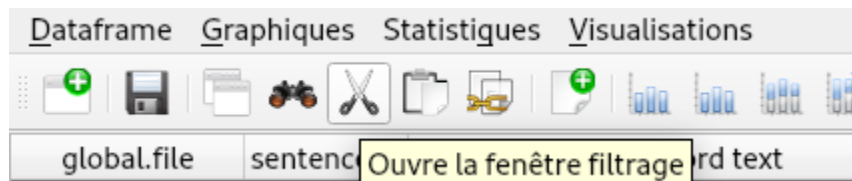


Fig. 13: Illustration du bouton de filtrage

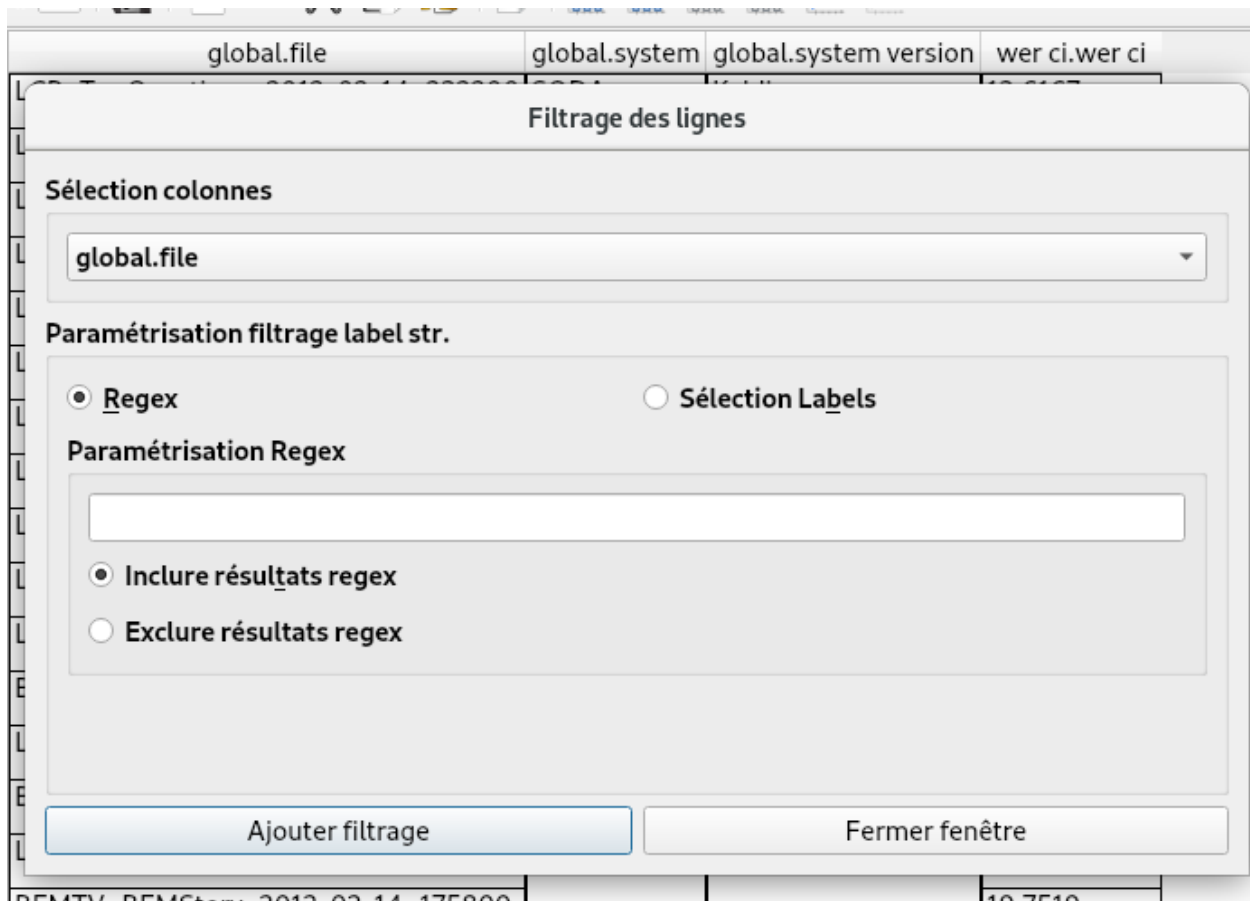


Fig. 14: Fenêtre de tri

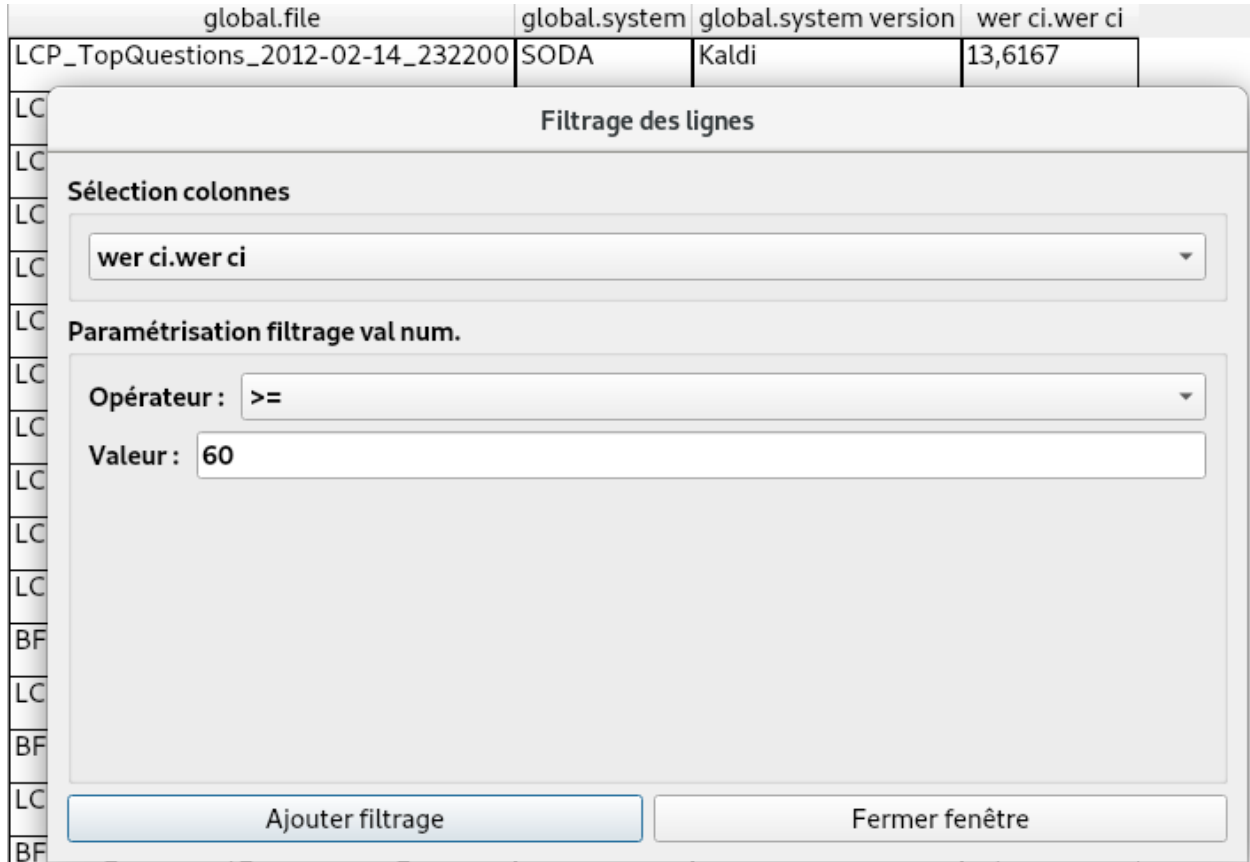


Fig. 15: Application d'un tri sur le WER

The screenshot shows the application window with a toolbar and a data table. The table displays the result of a sort operation on the 'wer ci.wer ci' column. The data is sorted in descending order of the 'wer ci.wer ci' values.

global.file	global.system	global.system version	wer ci.wer ci
BFMTV_CultureEtVous_2012-01-16_0...	PERCOL	Chouchou	76,0386
BFMTV_CultureEtVous_2012-02-15_0...			68,0912
BFMTV_CultureEtVous_2012-02-14_0...			62,6147
BFMTV_CultureEtVous_2012-02-15_0...		Oreillettes	66,0969
BFMTV_CultureEtVous_2012-02-14_0...			63,3028
LCP_PileEtFace_2012-01-12_192700		Tartatatin	65,9375
BFMTV_CultureEtVous_2012-02-15_0...			66,3818
BFMTV_CultureEtVous_2012-02-14_0...		Primary	65,1376
BFMTV_CultureEtVous_2012-02-15_0...			66,3818
BFMTV_CultureEtVous_2012-02-14_0...			65,1376

Fig. 16: Etat de la Vue suite au tri sur le WER

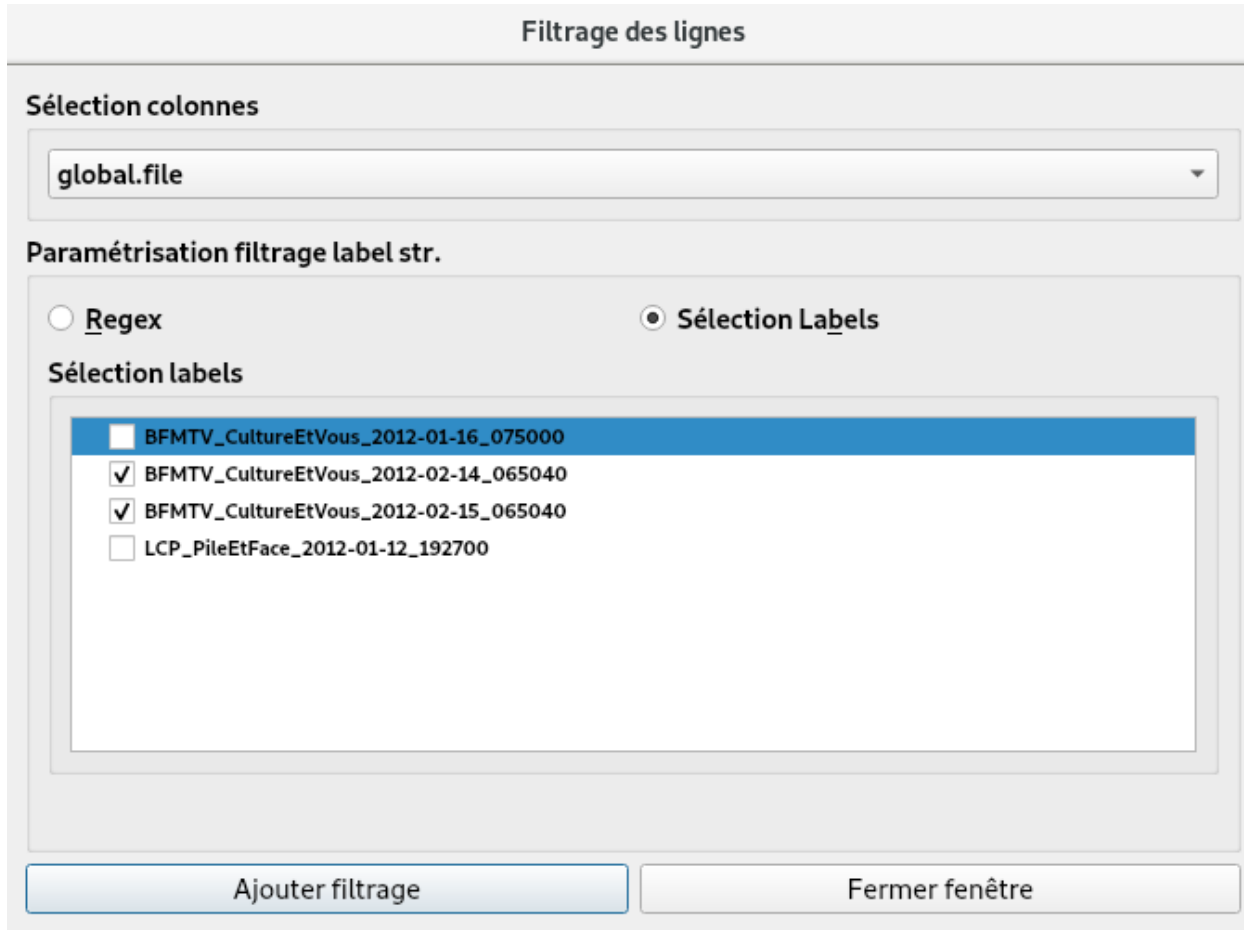


Fig. 17: Application d'un tri sur les fichiers

Datframe Graphiques Statistiques Visualisations

global.file	global.system	global.system version	wer ci.wer ci
BFMTV_CultureEtVous_2012-02-15_065040	PERCOL	Chouchou	68,0912
BFMTV_CultureEtVous_2012-02-14_065040			62,6147
BFMTV_CultureEtVous_2012-02-15_065040		Oreillettes	66,0969
BFMTV_CultureEtVous_2012-02-14_065040			63,3028
BFMTV_CultureEtVous_2012-02-15_065040		Tartatatin	66,3818
BFMTV_CultureEtVous_2012-02-14_065040			65,1376
BFMTV_CultureEtVous_2012-02-15_065040		Primary	66,3818
BFMTV_CultureEtVous_2012-02-14_065040			65,1376

Fig. 18: Etat de la Vue suite au tri sur les fichiers

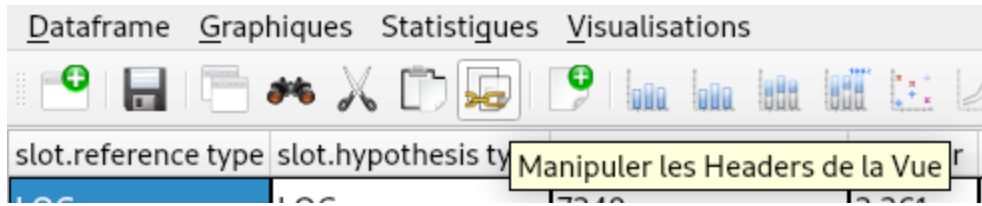


Fig. 19: Illustration du bouton de manipulation des en-têtes

Les colonnes rouges correspondent à des colonnes Valeurs et les colonnes bleues à des colonnes Labels. La partie inférieure de la fenêtre est divisée en 4 blocs. Ces quatre blocs permettent à l'utilisateur d'indiquer comment il veut modifier la représentation de la Vue :

- les colonnes Labels à placer en en-tête horizontale
- les colonnes Labels à placer en en-tête verticale
- les colonnes Valeurs à placer sous les en-têtes si c'est pertinent
- les colonnes Valeurs dépendant d'un en-tête

Les colonnes à choisir sont déplaçables par drag-and-drop.

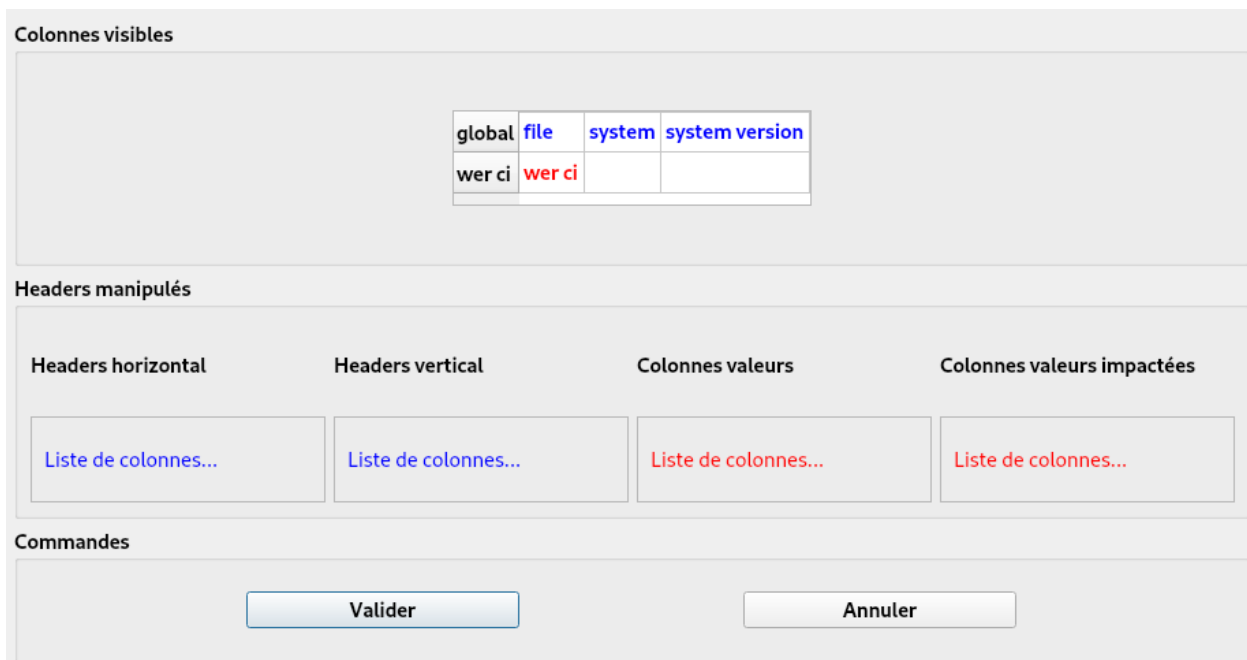


Fig. 20: Fenêtre de manipulation des en-têtes

Dans l'exemple illustré dans l'image *Drag and drop de "global.file"*, la colonne Label "global.file" est déplacée en en-tête horizontal.

L'orientation de l'en-tête étant opposée à celle de la Vue, il y aura une réorganisation potentielle de la Vue. La colonne Valeur "wer ci. wer ci" dépendant de cette colonne, elle est placée automatiquement dans le bloc "Colonnes Valeurs impactées". Cette colonne ne peut pas être enlevée du bloc tant que la colonne "global.file" est présente en en-tête horizontal. Une fois les modifications validées, la nouvelle représentation de la Vue est visible dans l'image *Vue modifiée après manipulation d'en-têtes 1*.

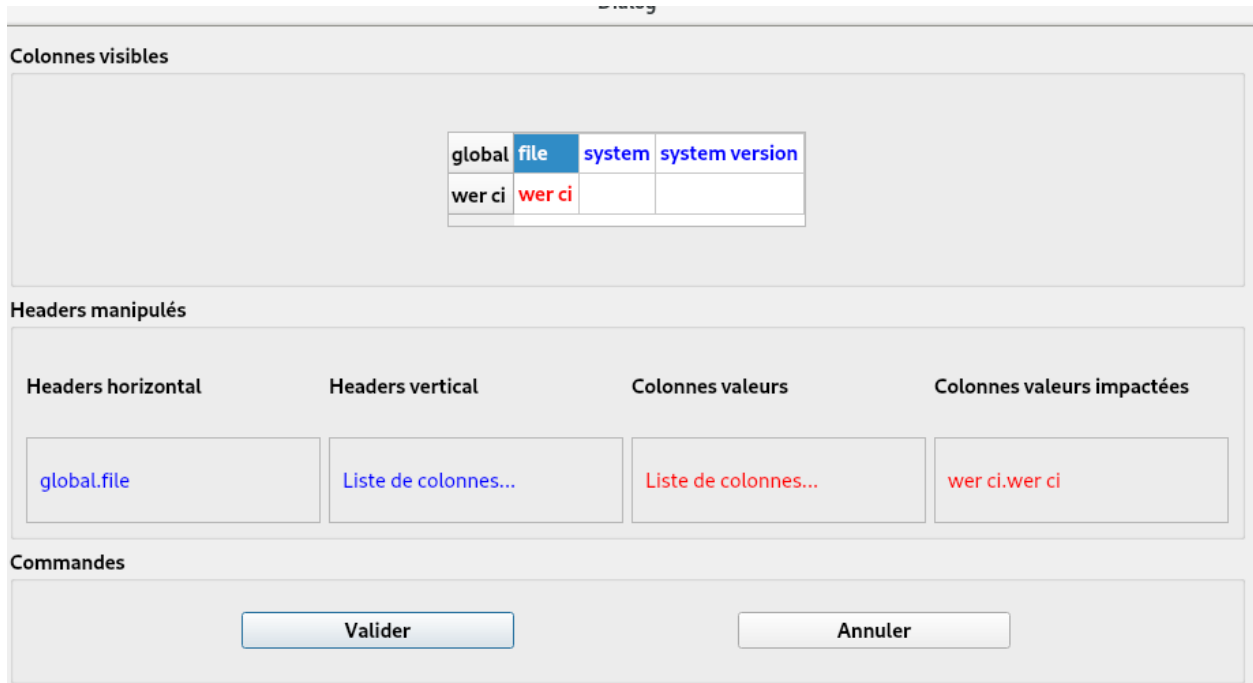


Fig. 21: Drag and drop de “global.file”

global.system	global.system version	file	
		BFMTV_CultureEtVous_2012-02-15_065040	BFMTV_CultureEtVous_2012-02-14_065040
		wer ci	wer ci
PERCOL	Chouchou	68,0912	62,6147
	Oreillettes	66,0969	63,3028
	Tartatatin	66,3818	65,1376
	Primary	66,3818	65,1376

Fig. 22: Vue modifiée après manipulation d’en-têtes 1

Si l'on ouvre de nouveau la fenêtre de manipulation, il est possible de modifier les manipulations, voir de les annuler. Un clic droit sur la colonne "global.file" permet de la supprimer. Si l'utilisateur décide de déplacer la colonne "global.system" en en-tête vertical, comme dans l'image *Drag and drop de "global.system"*, l'orientation est identique à celle de l'en-tête, il n'y aura pas de réorganisation de la Vue.

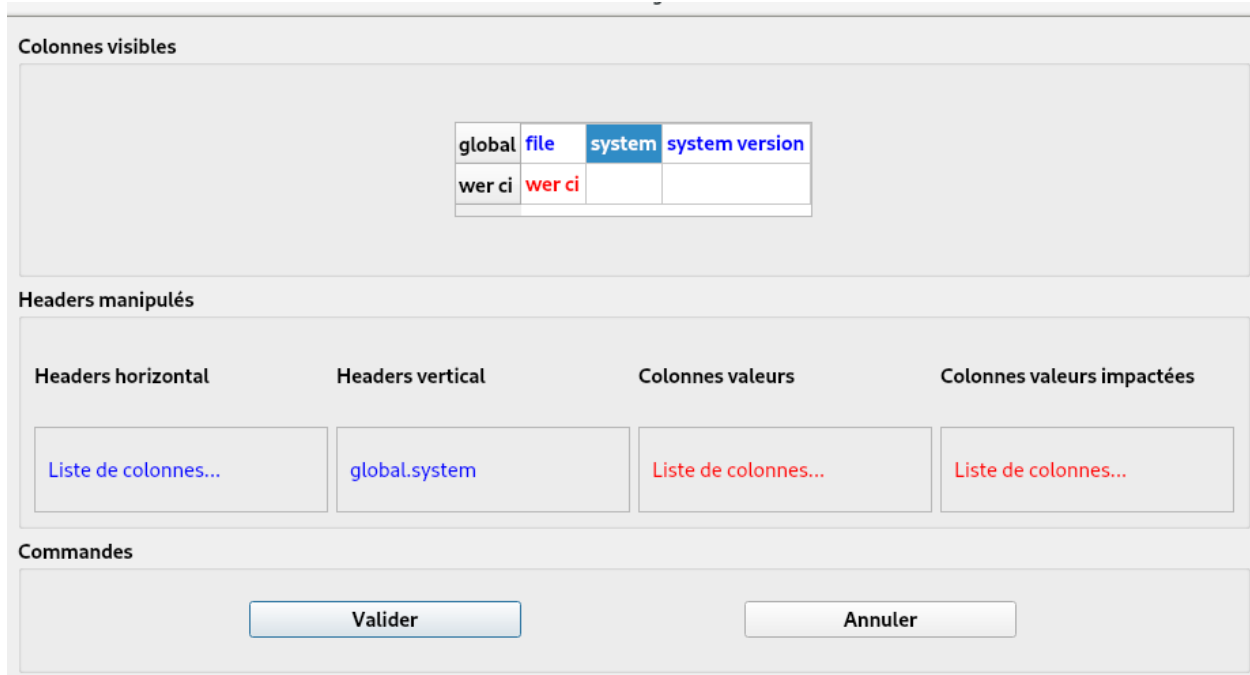


Fig. 23: Drag and drop de "global.system"

Enfin, il est tout à fait possible de faire des hiérarchies d'en-têtes par un drag-and-drop, comme présenté dans l'image *Drag and drop de plusieurs en-têtes*.

On peut observer là encore que la colonne Valeur "wer ci. wer ci" est dépendante des en-têtes manipulées. La Vue résultat est illustrée dans l'image *Vue modifiée après manipulation d'en-têtes 2*.

9.1.6 Ajouter une colonne calculée

Pour présenter le fonctionnement de l'ajout de la colonne calculée, nous allons nous appuyer sur la Vue dans son état initial, avec les colonnes "speaker.name", "turn.id", "turn.start" et "turn.end" ouvertes en plus des autres colonnes déjà visibles dans l'état initial. L'état de la Vue est illustrée dans l'image *Vue exemple ajout colonne calculée*.

Pour créer de nouvelles colonnes calculées, l'utilisateur peut appuyer sur le bouton illustré dans l'image *Illustration du bouton d'ajout d'une colonne calculée*.

La pression de ce bouton fait apparaître le formulaire de création de colonnes calculées. Ce formulaire permet d'écrire la formule correspondant à la colonne que l'utilisateur souhaite créer ainsi qu'aux différents paramètres nécessaires à la création de la colonne. La matrice située dans la partie supérieur du formulaire présente toutes les colonnes composants la Dataframe. Si l'utilisateur souhaite s'appuyer sur le contenu d'une de ces colonnes, il peut soit entrer la clef de la colonne dans la zone de texte située en dessous de la matrice, soit faire glisser le nom de la colonne de la matrice vers la zone de texte.

En plus de la formule à entrer, l'utilisateur doit renseigner quatre autres paramètres : le nom de la colonne, son groupe, son type et sa formule de réduction. Un cinquième paramètre optionnel peut être entré : il s'agit de la granularité de la colonne. Dans l'exemple présenté dans l'image *Création colonne durée sans granularité*, on crée une colonne durée

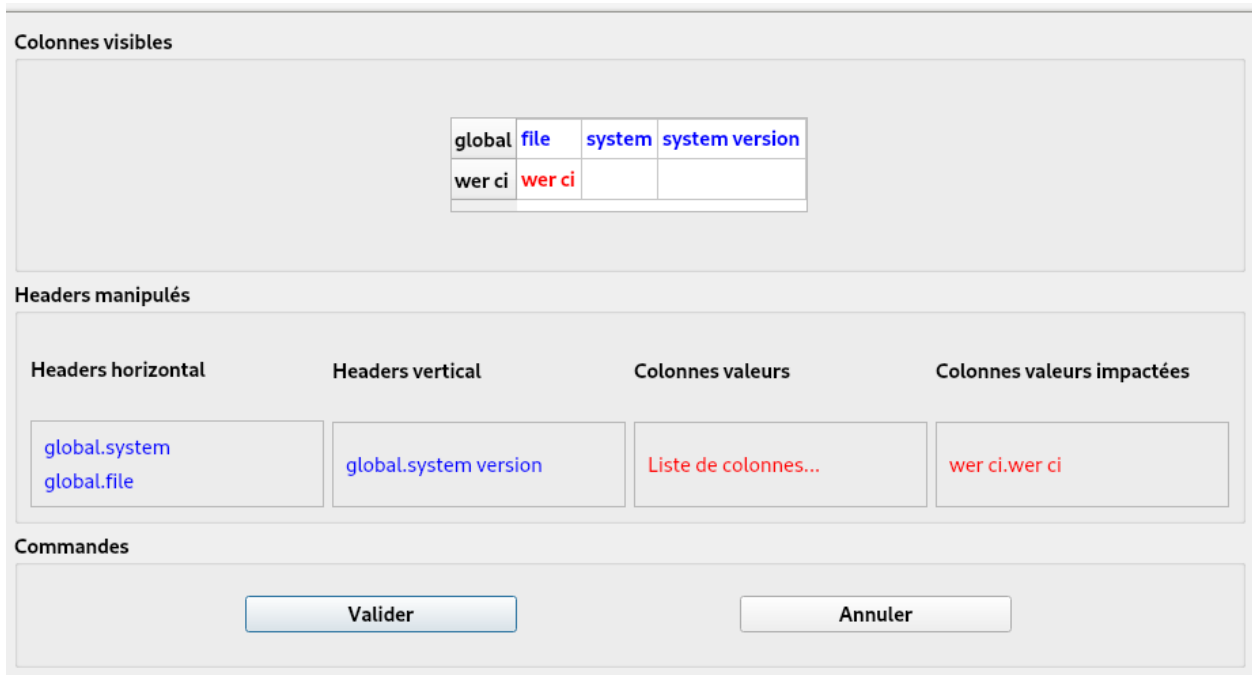


Fig. 24: Drag and drop de plusieurs en-têtes

		system	
		PERCOL	
		file	
		BFMTV_CultureEtVous_2012-02-15_065040	BFMTV_CultureEtVous_2012-02-14_065040
		wer ci	wer ci
system version	Chouchou	68,0912	62,6147
	Oreillettes	66,0969	63,3028
	Tartatatin	66,3818	65,1376
	Primary	66,3818	65,1376

Fig. 25: Vue modifiée après manipulation d'en-têtes 2

global.file	speaker.name	turn.id	urn.start time	urn.end time	global.system	global.system version	wer ci.wer ci	wer ci.nce
LCP_TopQuestions_2012-02-14_232200	Jean-Baptiste_PREDALI	1	525,962	531,769	SODA	Kaldi	14,8148	0,160528
		2	532,133	546,161			29,1667	0,508396
		3	546,491	548,595			0	nan
	Bernard_ACCOYER	4	551,168	553,359			0	nan
	Christian_PAUL	5	553,718	566,41			7,69231	0,200326
		6	566,998	579,885			7,69231	0,456802
		7	580,502	592,279			0	nan
		8	592,279	603,523			0	nan
		9	603,998	620,015			18,1818	-inf
		10	620,015	631,497			9,7561	0,227
		11	632,035	632,969			0	nan
		12	633,349	647,802			4,65116	-0,0207721
		13	647,802	662,753			14,5833	0,520776
		14	662,753	671,731			6,45161	0,0624891
		15	672,206	675,953			45,4545	0,471501
		Bernard_ACCOYER	16	677,297			680,142	0
	Valerie_PECRESSE	17	680,642	685,111			11,1111	0,813704
		18	685,554	699,776			14,2857	0,380048
		19	699,776	713,885			6,97674	0,534367
		20	713,885	728,726			29,2683	0,328524

Fig. 26: Vue exemple ajout colonne calculée



Fig. 27: Illustration du bouton d'ajout d'une colonne calculée.

en soustrayant le temps de fin d'un tour avec le temps de début. Dans cet exemple, il n'y a pas de granularité associée à cette colonne.

Colonne présentes

global	file	lang	system	system version	channel						
section	id	start time	end time	type							
speaker	name	type	dialect	accent							
turn	id	start time	end time	raw text	pronounced text	normalized text					
wer ci	id	ref text	hyp start time	hyp end time	hyp text	correct	subst	insert	delete	delete opt	correct cut

Saisie de l'opération

\$turn.end time\$-\$turn.start time\$

Paramètres

Nom : Groupe :

Type de l'opération : Label Valeur Valeur de l'opération : Num String

Réduction : Granularité :

Commandes

Fig. 28: Création colonne durée sans granularité

Par contre, dans l'image *Création colonne durée avec granularité*, la colonne est créée avec comme granularité le locuteur, "speaker.name".

Les deux colonnes résultant de ces manipulations sont visibles dans l'image *Résultats des deux colonnes expressions*. Il est à noter qu'il est possible à tout moment d'éditer une colonne calculée en effectuant un clic droit sur l'en-tête de la colonne.

On peut voir du fait de la granularité que la seconde colonne créée a un comportement différent malgré une formule identique.

9.2 Autres fonctionnalités

Deux autres fonctionnalités existent : l'historique et la sauvegarde de Dataframe.

Pour enregistrer une dataframe il suffit de cliquer sur l'icône suivante :

Un explorateur s'ouvre alors, il ne reste plus qu'à indiquer le nom et l'emplacement de la sauvegarde. Pour utiliser l'historique, il suffit de cliquer sur l'icône suivante :

Une fenêtre va alors s'ouvrir, récapitulant l'ensemble des actions effectuées par l'utilisateur. Un double clic sur une des actions permet de créer une Vue dans l'état qu'elle avait au moment de cette action. Il est à noter que cette fonctionnalité ne fonctionne pas encore très bien avec les créations de colonnes calculées : l'ajout et la suppression d'informations doivent être mieux pris en compte dans EVALOMATIC.

Colonnes présentes

global	file	lang	system	system version	channel						
section	id	start time	end time	type							
speaker	name	type	dialect	accent							
turn	id	start time	end time	raw text	pronounced text	normalized text	duree				
wer ci	id	ref text	hyp start time	hyp end time	hyp text	correct	subst	insert	delete	delete opt	co

Saisie de l'opération

\$turn.end time\$-\$turn.start time\$

Paramètres

Nom : Groupe :

Type de l'opération : Label Valeur Valeur de l'opération : Num String

Réduction : Granularité :

Commandes

Fig. 29: Création colonne durée avec granularité

Dataframe Graphiques Statistiques Visualisations

global.file	speaker.name	turn.id	urn.start time	turn.end time	wer ci.wer ci	wer ci.nce	turn.duree	turn.duree_file
LCP_TopQuestions_2012-02-14_232200	Jean-Baptiste_PREDALI	1	525,962	531,769	10,6996	0,0677419	5,807	422,902
		2	532,133	546,161	25	0,453997	14,028	
		3	546,491	548,595	4,44444	0,476674	2,104	
	Bernard_ACCOYER	4	551,168	553,359	9,25926	-1,12821	2,191	
	Christian_PAUL	5	553,718	566,41	7,97721	-inf	12,692	
		6	566,998	579,885	11,396	0,40144	12,887	
		7	580,502	592,279	2,04678	0,321417	11,777	
		8	592,279	603,523	6,17284	0,34795	11,244	
		9	603,998	620,015	14,1414	-inf	16,017	
		10	620,015	631,497	13,8211	0,362037	11,482	
		11	632,035	632,969	0	nan	0,934	
		12	633,349	647,802	6,20155	0,262745	14,453	
		13	647,802	662,753	11,1111	0,43081	14,951	
		14	662,753	671,731	15,7706	-0,0660341	8,978	
		15	672,206	675,953	38,3838	-inf	3,747	
	Bernard_ACCOYER	16	677,297	680,142	19,7531	-1,62617	2,845	
	Valérie_BREFFEE	17	680,647	695,111	20,6306	0,601050	14,460	

Fig. 30: Résultats des deux colonnes expressions



9.3 Conclusions

Nous avons présentés l'ensemble des fonctionnalités permettant d'explorer les données d'une Vue, notamment par le biais des opérations. Dans les chapitres suivantes, nous présenterons les interfaces de visualisation spécifiques aux tâches, *Visualisations supplémentaires*. Nous présenterons ensuite dans le chapitre graphiques les graphiques pouvant être générés à partir de l'état d'une Vue. Enfin, nous détaillerons les différentes opérations statistiques présentes dans le chapitre statistiques.

VISUALISATIONS SUPPLÉMENTAIRES

10.1 Interface de visualisation des annotations

Si une Vue d'une Dataframe contient des colonnes représentant des annotations textuelles, il est possible de les visualiser alors dans une sous interface spécifique. Pour ouvrir cette sous interface, il faut d'abord sélectionner la visualisation correspondante dans le menu déroulant "Visualisations".

Cette sélection permet d'accéder à un formulaire de configuration, permettant à l'utilisateur de visualiser si les colonnes nécessaires à l'affichage de cette visualisation sont présentes. Ce formulaire apparaît dans l'image *Illustration du formulaire de lancement de la fenêtre d'annotation*.

Corpus à afficher corpus

Colonne des documents Aucune sélection

Phrases : sentence.id
Mots : sentence.word text
Part of Speech : sentence.pos
Informations syntaxiques : sentence.component type
Entités nommées :
 * id 1 : sentence.named entity type 1

Valider Annuler

Fig. 1: Illustration du formulaire de lancement de la fenêtre d'annotation

L'utilisateur doit sélectionner le corpus sur lequel il veut travailler, s'il y en a plusieurs. Dans cet exemple, il n'y a qu'un seul corpus. Si la Vue contient des informations d'annotation, le nom des colonnes est affiché dans la fenêtre de dialogue inférieure. Par ailleurs, l'utilisateur peut aussi choisir une colonne de groupement des textes, par exemple le document. Dans notre exemple, aucune colonne n'est choisie : le découpage se fait alors toutes les 50 lignes, c'est à dire par rapport à "sentence.id". Nous conseillons généralement de prendre cette option : les documents peuvent être assez long, et le chargement de la visualisation très longue. Une fois la configuration du formulaire achevée, l'interface de visualisation des annotations se lance. Elle est illustrée dans l'image *Exemple de la fenêtre d'annotation*.

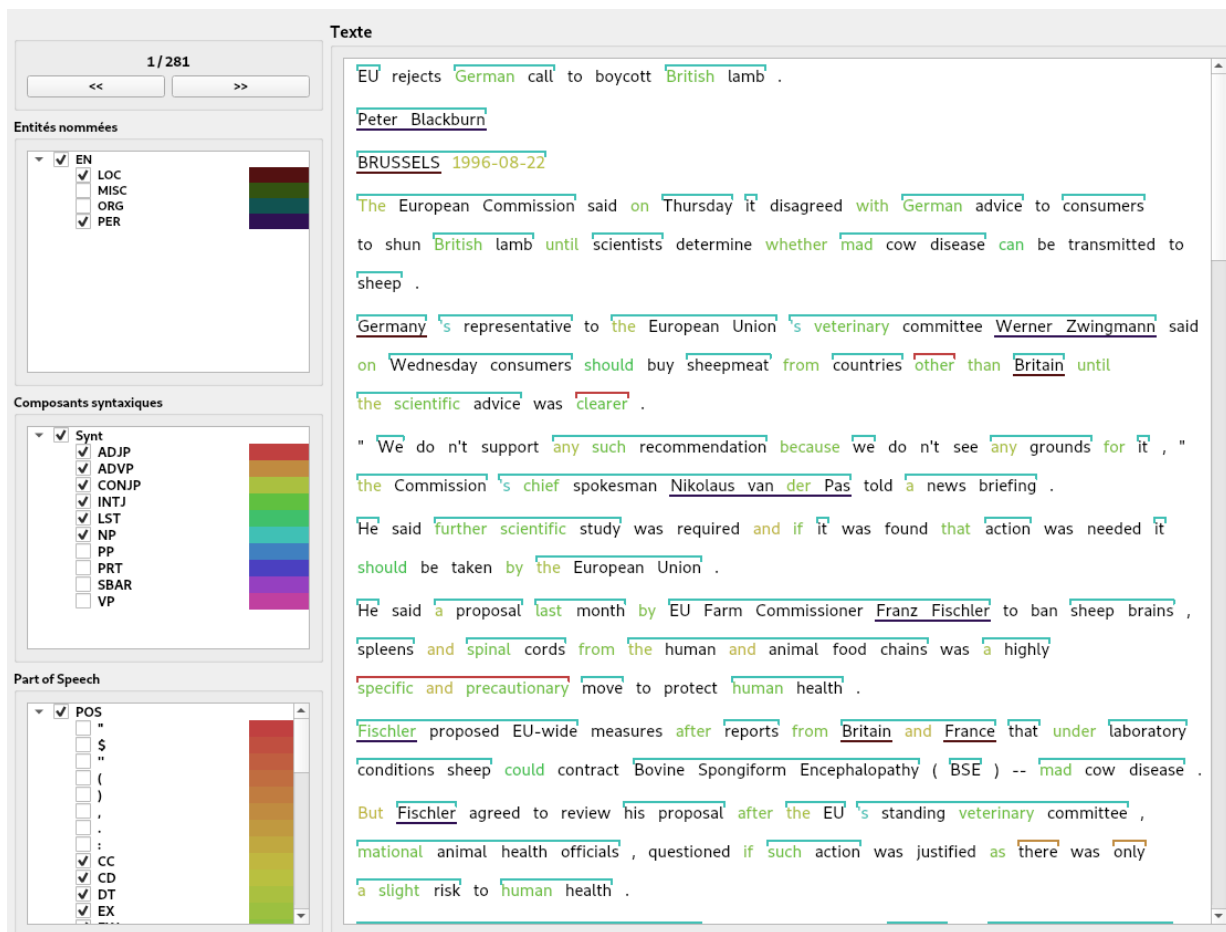


Fig. 2: Exemple de la fenêtre d'annotation

Dans cet exemple, on peut observer les différents éléments de l'interface. En haut à gauche sont situés les boutons de navigation pour explorer les ensembles de documents. Sur le côté gauche sont situés les différentes listes d'annotations, classées par catégories, dans notre cas : "Entités nommées", "Composants syntaxiques" et "Part of Speech". L'utilisateur peut choisir les annotations à afficher dans l'écran principal de l'interface. Les annotations sont représentées en fonction de leur type : les entités nommées, qui peuvent avoir plusieurs niveaux, sont soulignées; les composants syntaxiques sont encadrés par un trait supérieur; et enfin les tags Part of Speech sont surlignés.

10.2 Interface de visualisation des traductions

Si une Vue d'une Dataframe contient des colonnes représentant des informations de traduction, il est possible de visualiser les métriques TER et BLEU dans une sous interface spécifique. Pour ouvrir cette interface, il faut sélectionner la

visualisation dans le menu déroulant “Visualisations”. Cette sélection provoque l’ouverture d’une fenêtre de configuration, permettant à l’utilisateur de visualiser quelles colonnes de la Vue ont été identifiées comme utiles à l’affichage de la visualisation. Outre les colonnes correspondant aux systèmes, fichiers et tours, il est nécessaire que la Vue contienne au moins les colonnes d’une des deux métriques. Par ailleurs, l’utilisateur peut choisir sur quel hypothèse, fichier et tour l’interface doit se lancer, pour accélérer la navigation. Cette fenêtre de configuration est illustrée dans l’image *Illustration de la fenêtre de configuration de traduction*.

Fig. 3: Illustration de la fenêtre de configuration de traduction

Une fois la configuration validée par l’utilisateur, la fenêtre de visualisation se lance, à condition d’avoir les colonnes nécessaires. Si une colonne manque, une pop-up apparaît pour informer l’utilisateur. Par défaut, si les deux métriques sont présentes, la fenêtre s’ouvre sur la métrique BLEU. Le fonctionnement de la visualisation est légèrement différente en fonction de la métrique sélectionnée. Pour la métrique BLEU, l’interface affiche les n-grams entre l’hypothèse et les références. Il est possible de sélectionner grâce au menu sur la gauche de sélectionner le type de n-gram que l’on souhaite visualiser (de 1 à 4). La couleur des soulignements dans l’interface indique le score du n-gramme : la gradation va de rouge (0) à vert (1). Enfin, le score du tour en single et en multi est affiché en bas à gauche. Une illustration de l’interface pour la métrique BLEU avec des 2-grams est visible dans l’image *Visualisation traduction avec BLEU*.

A tout moment, l’utilisateur peut choisir de switcher entre deux métriques. Pour la métrique TER, l’interface représente les différentes opérations entre l’hypothèse et les références, les différents niveaux de soulignement dans l’hypothèse correspondant chacun à une référence (ex. le niveau 1 de soulignement correspond à la première référence). La colorisation permet de représenter le type d’opération associé à un mot, ou à un groupe de mots dans

Fig. 4: Visualisation traduction avec BLEU

le cas d'une opération de déplacement. Excepté dans le cas des opérations d'insertion et de suppression, les mots des références sont alignés avec ceux de l'hypothèse. A l'aide de checkbox situées sur le menu à gauche de l'interface, il est possible de cacher certains types d'opérations. Enfin, les scores TER obtenus pour chaque référence sont affichés en bas à gauche. Cette interface est illustrée dans l'image *Visualisation traduction avec TER*.

10.3 Interface de visualisation audio

Si la Vue d'une Dataframe contient des colonnes contenant des informations de transcription audio, il est possible d'ouvrir une interface de visualisation des fichiers audios. Actuellement, l'interface n'est accessible qu'en double cliquant dans la colonne des fichiers. A terme, il sera possible d'accéder à l'interface de la même manière que les autres interfaces. Une fois le double clic effectué, l'utilisateur doit patienter car il est nécessaire de récupérer depuis le serveur le fichier audio correspondant. Une fois le chargement terminé, l'interface s'ouvre et permet de visualiser le contenu du fichier audio. L'interface est illustrée dans l'image *Visualisation audio transcription de la parole*.

L'interface permet de se déplacer dans le fichier audio à partir du découpage en locuteurs ou par mots, selon le type de fichiers associés au fichier audio : référence, hypothèse, trs ... Il est aussi possible à l'utilisateur d'écouter directement le fichier audio.

10.4 Nouvelle interface de visualisation audio

La nouvelle interface de visualisation audio permet de représenter et parcourir des séries d'annotations sous la forme d'une frise temporelle. Elle permet également d'écouter l'audio associé aux éléments présentés sous forme de blocs. On y accède à partir de la fenêtre de la Dataframe et une phase de paramétrage sera nécessaire afin de permettre à l'utilisateur de définir la représentation des données la plus adaptée à ce qu'il veut observer.



Fig. 5: Visualisation traduction avec TER

Lors du premier lancement, l'interface de paramétrage sera automatiquement présentée à l'utilisateur afin de déterminer les colonnes de la Dataframe à utiliser. Par la suite, si la structure de la Dataframe est similaire, l'interface se lancera automatiquement avec la dernière configuration utilisée. On pourra à tout moment ajouter ou éditer le paramétrage des séries d'annotations à utiliser via les boutons du menu.

Un double clic sur un élément de la liste permet de l'éditer. La suppression de lignes est également possible.

La barre d'outils permet de sélectionner le fichier source observé parmi ceux définis dans la Dataframe. Si la source audio est disponible, absente ou en cours de chargement, son icône permettra de rendre compte de cet état.

La fonction de recherche (accès rapide Ctrl+F) permet de naviguer via les apparitions d'un mot dans les blocs affichés. Les blocs validant les conditions de recherche seront identifiés via un contour en pointillés.

Les panneaux latéraux permettent de visualiser le contenu de la Dataframe associé au bloc survolé à la souris. On peut en personnaliser le contenu, les dupliquer et les déplacer dans ou hors de la fenêtre principale afin d'optimiser l'espace de travail.

Concernant l'écoute, l'intervalle de lecture par défaut est la zone affichée dans la fenêtre. On peut modifier celle-ci en zoomant/dé-zoomant (Ctrl+molette), en double cliquant sur un bloc d'intérêt, ou en sélectionnant plusieurs en dessinant une zone de sélection rectangulaire autour. La touche retour arrière permet de rejouer les 5 dernières secondes avant la position du curseur de lecture.

On peut se déplacer dans la vue grâce à la barre de défilement horizontale ou en utilisant les touches gauche et droite du clavier pour du pas à pas.

Le texte des blocs étant à taille de police constante, il peut être masqué si celui-ci couvre un temps très court. On peut cependant contourner ce problème grâce à l'affichage d'une info-bulle ou la consultation du contenu via les panneaux latéraux dans les panneaux latéraux au passage de la souris, ou en zoomant à une échelle de temps plus adaptée.

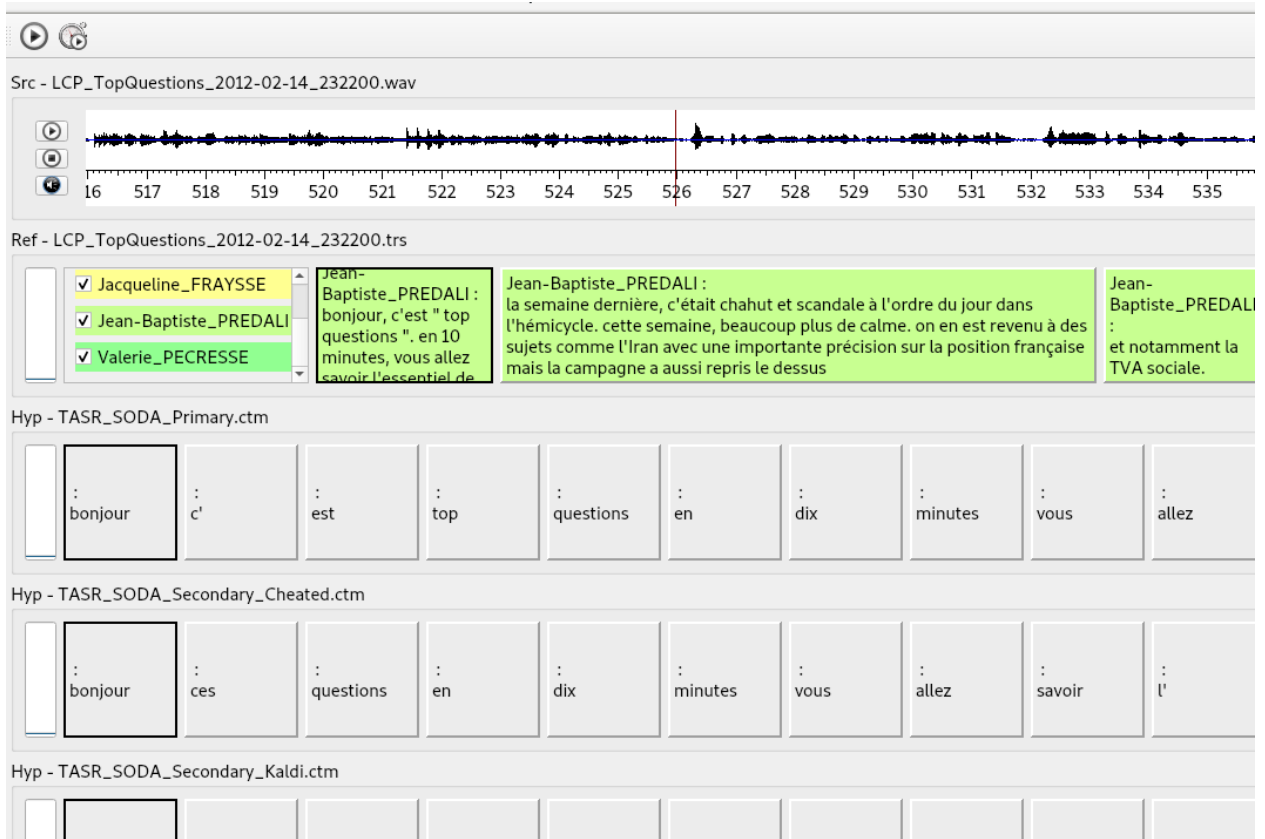


Fig. 6: Visualisation audio transcription de la parole



Formulaire paramètres annotations

Sélection type annotation

Référence Hypothèse

Système : PERCOL Version : PEROK

Sélection découpage temps

Groupe wer ci Id id Temps début hyp start time Temps fin hyp end time

Sélection sections

Groupe Label : wer ci Nom Label : ref text Groupe Couleur : wer ci Nom Couleur : correct -

Groupe Label : wer ci Nom Label : hyp text Groupe Couleur : wer ci Nom Couleur : entropie -

+ Section

Valider Annuler

Dialog

Ref, turn (id, start time, end time), x2

Hyp : SODA 1.0, wer ci (id, hyp start time, hyp end time), x2

Hyp : PERCOL PEROK, wer ci (id, hyp start time, hyp end time), x2

-

OK Close

CRÉER DES GRAPHIQUES EN UTILISANT EVALOMATIC

Il existe actuellement 6 types de graphiques accessibles dans Evalomatic :

- Les histogrammes classiques
- Les histogrammes de distribution
- Les histogrammes catégoriels
- Les boîtes à moustaches
- Les nuages de points
- Les courbes DET

Excepté les courbes DET, l'ensemble des graphiques peuvent être créés à partir de n'importe quelle Vue. Les courbes DET s'appuyant sur les valeurs de colonnes générées lors d'une évaluation en vérification du locuteur, on ne peut pas les générer dans d'autre cas. Chaque graphe doit être paramétré lors de sa création pour indiquer les données des colonnes que l'utilisateur souhaite visualiser. Dans les parties suivantes, nous décrivons les différents types de graphe et leur utilisation en détail.

Pour la plupart des exemples, nous nous appuyons sur la Vue d'une Dataframe générée à partir d'une évaluation en transcription de la parole, sur laquelle on a trié les lignes en fonction des systèmes et de leurs versions. Cette vue est illustrée en *Vue d'une Dataframe d'une évaluation en transcription de la parole* :

global.system	global.system version	wer ci.entropie	wer ci.hyp correct	wer ci.hyp incorrect	wer ci.entropie max	wer ci.wer ci	wer ci.nce
PERCOL	Chouchou					31,2414	
	Oreillettes					27,5699	
	Primary					28,1345	
	Tartatatin					31,4089	
QCOMPERE	Primary	-24765,6	28439	7485	26523,8	28,2534	0,0662861
SODA	Cheated	-inf	31342	3466	16277,8	15,9147	-inf
	Kaldi	-inf	30495	4833	20342,1	19,0431	-inf
	Primary	-inf	31238	3513	16417,9	16,166	-inf
	Sphinx	-inf	31130	4372	19112,2	17,352	-inf

Fig. 1: Vue d'une Dataframe d'une évaluation en transcription de la parole

Les nuages de points et les histogrammes distributionnels en fonction de l'état de la Vue : pour ces deux graphes, les colonnes visibles ont un impact important sur ce qui est calculé et affiché. Pour les autres graphes, les calculs et l'affichage dépendront seulement des facteurs et des variables passés en paramètre. Certains graphes permettent d'étudier plusieurs facteurs : si leur combinatoire est très élevée, le calcul peut être très long. Il convient donc de bien réfléchir au graphe que l'on veut créer.

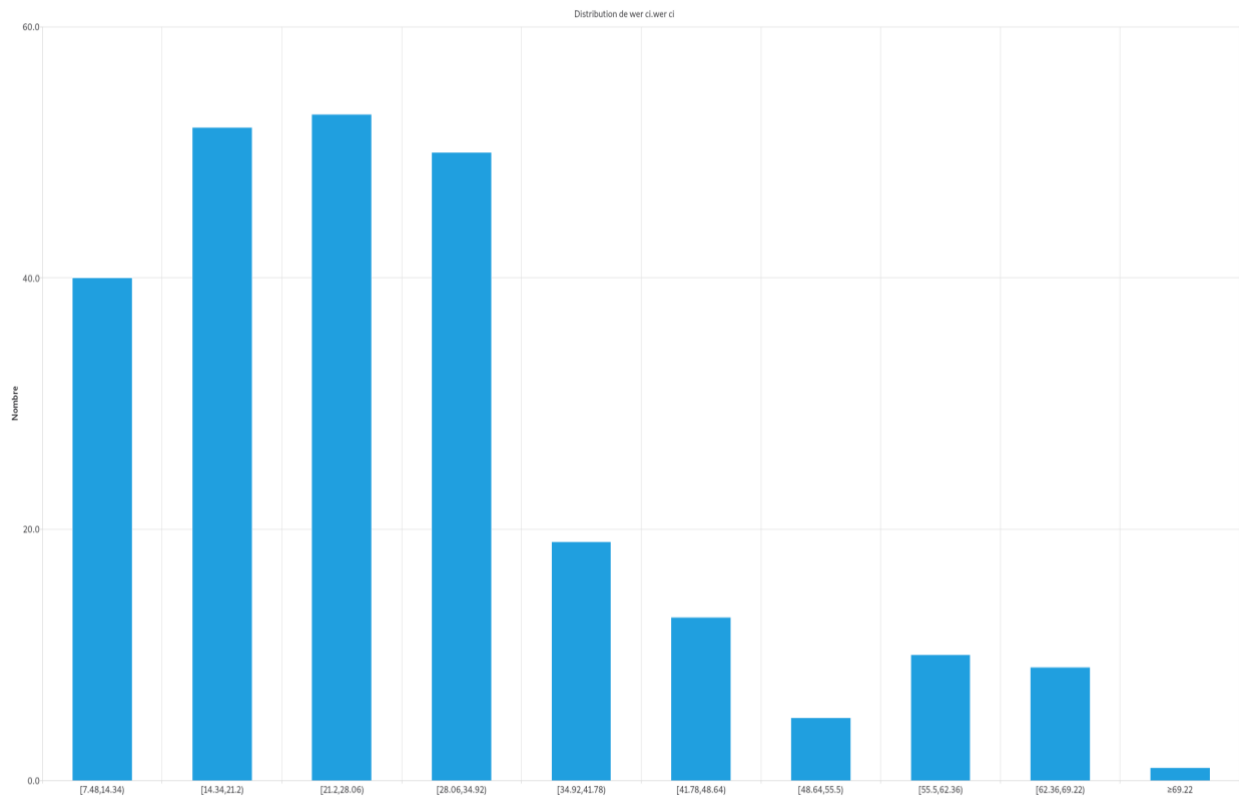


Fig. 5: Exemple d'un histogramme de distribution

global	file	lang	system	system version	channel															
section	id	start time	end time	type																
speaker	name	type	dialect	accent																
turn	id	start time	end time	raw text	pronounced text	normalized text														
wer.ci	id	ref text	hyp start time	hyp end time	hyp text	correct	subst	insert	delete	delete opt	correct cut	delete cut	ref cnt	hyp cnt	entropie	hyp correct	hyp incorrect	entropie max	wer.ci	nce

Facteurs

- global.system
- global.system version

Catégories

- speaker.type

Variables

- wer.ci.wer.ci

Fig. 6: Paramétrisation d'un histogramme de catégories

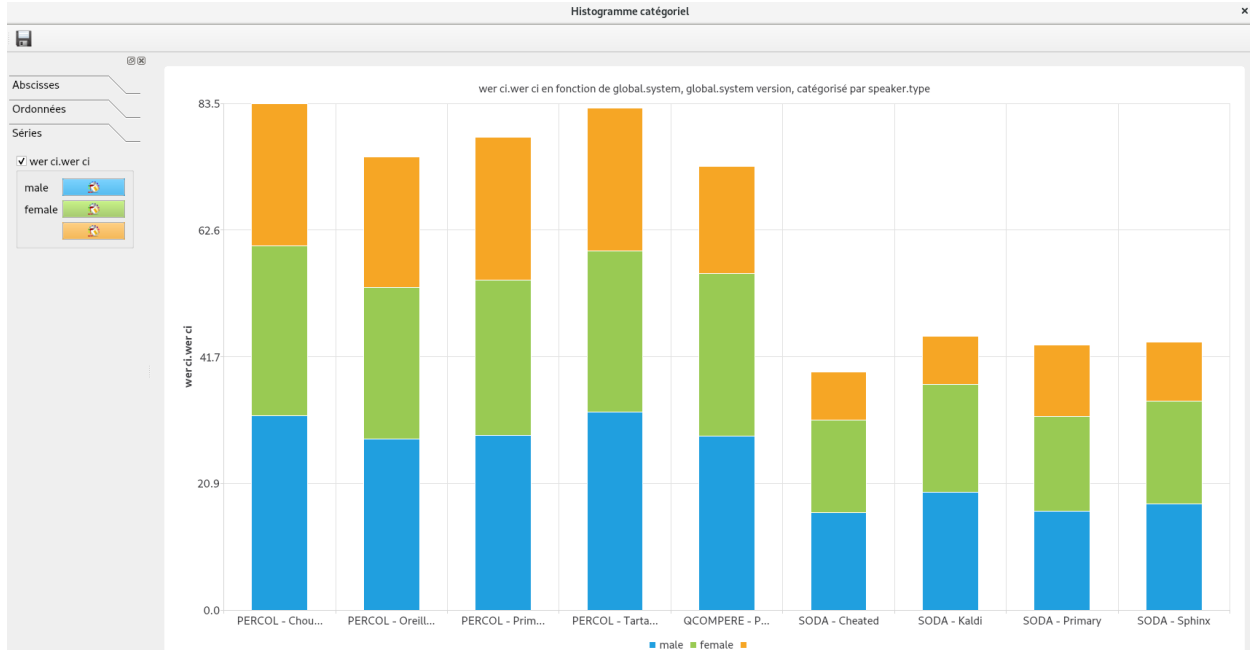


Fig. 7: Exemple d'un histogramme de catégorie

11.4 Les boîtes à moustaches

Les boîtes à moustache permettent d'illustrer différentes caractéristiques pour une valeur observée par rapport à un facteur selon une granularité donnée. Ainsi, l'utilisateur doit indiquer nécessairement la colonne contenant les valeurs à observer et la colonne label servant de référence à la granularité. La colonne label représentant le facteur est optionnelle : si elle n'est pas renseignée, le graphe ne sera composé que d'une seule boîte à moustache. Cette fenêtre est illustrée dans l'image *Paramétrisation d'un graphe de boîtes à moustaches*.

Valeur:

Facteur:

Granularité:

global	file	lang	system	system version	channel															
section	id	start time	end time	type																
speaker	name	type	dialect	accent																
turn	id	start time	end time	raw text	pronounced text	normalized text	duree													
wer.ci	id	ref text	hyp start time	hyp end time	hyp text	correct	subst	insert	delete	delete opt	correct cut	delete cut	ref cnt	hyp cnt	entropie	hyp correct	hyp incorrect	entropie max	wer.ci	nce

Fig. 8: Paramétrisation d'un graphe de boîtes à moustaches

Dans notre exemple, nous décidons d'observer les caractéristiques du WER sur les différentes versions de systèmes, avec comme granularité le fichier. Le graphe est visible dans l'image *Exemple d'un graphe de boîtes à moustache*.

11.5 Les nuages de points

Les nuages de points permettent de mettre en corrélation deux variables données tirées de colonnes valeurs. Par ailleurs, cette corrélation peut être observée en fonction d'un ou deux facteurs tirés de colonnes labels. Pour générer ce type de graphe, l'utilisateur doit donc renseigner les deux colonnes valeurs dont il souhaite observer la corrélation.

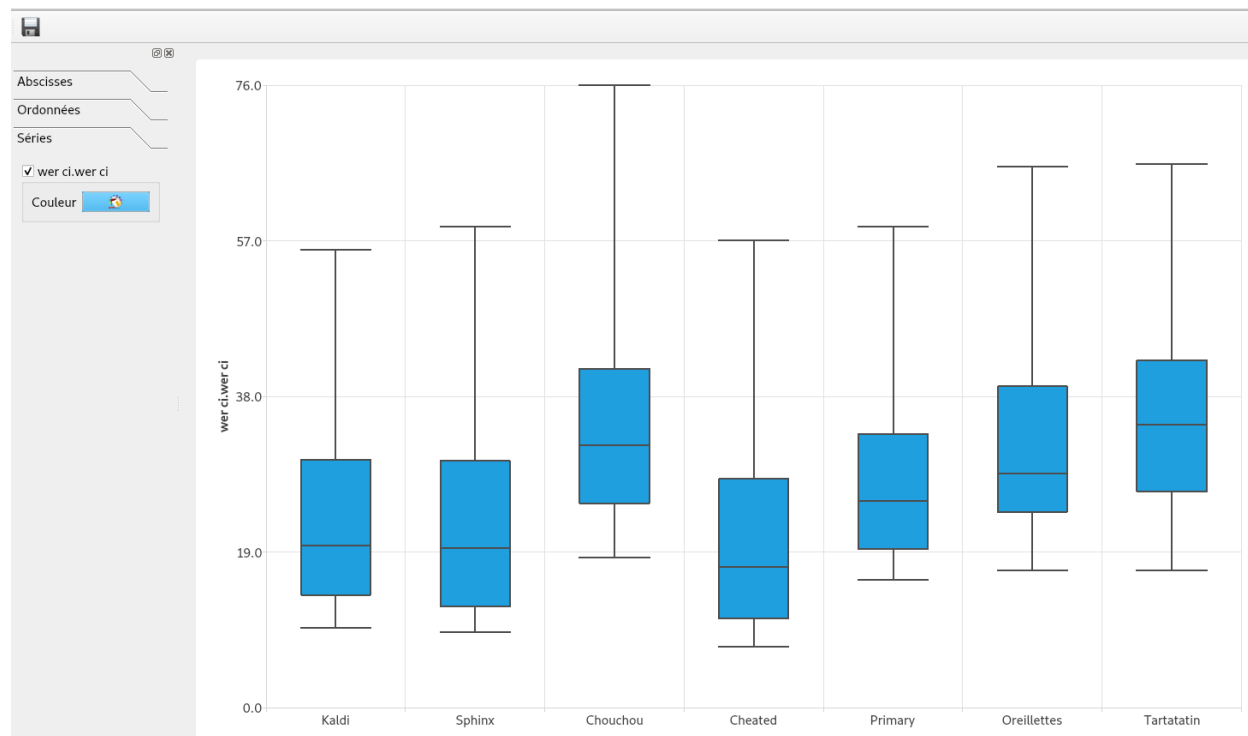


Fig. 9: Exemple d'un graphe de boîtes à moustache

Il est aussi possible à l'utilisateur d'indiquer une ou deux colonnes facteurs, qui seront représentés par des variations de couleurs et de formes. La fenêtre utilisée est représentée dans l'image *Paramétrisation d'un nuage de points*.

global	file	lang	system	system version	channel															
section	id	start time	end time	type																
speaker	name	type	dialect	accent																
turn	id	start time	end time	raw text	pronounced text	normalized text	duree													
wer.ci	id	ref text	hyp start time	hyp end time	hyp text	correct	subst	insert	delete	delete opt	correct cut	delete cut	ref cnt	hyp cnt	entropie	hyp correct	hyp incorrect	entropie max	wer.ci	nce

Abscisse: turn.duree

Ordonnée: wer.ci.wer.ci

Couleur: global.system version

Forme: Optionnel

Cancel OK

Fig. 10: Paramétrisation d'un nuage de points

Dans notre exemple, nous décidons d'observer la corrélation entre le WER et la durée en fonction des versions de systèmes, qui sont représentés par les changements de couleurs. Nous ajoutons donc préalablement une colonne calculée "turn.duree" avec comme granularité le tour de parole. Le nuage résultant de cette paramétrisation est illustré dans l'image *Paramétrisation d'un nuage de points*.

11.6 Les courbes DET

Les courbes DET ne sont accessibles qu'à partir d'une Vue résultant d'une évaluation en vérification du locuteur : il est nécessaire d'avoir un certain nombre de colonnes générées dans ces évaluations pour pouvoir créer une courbe

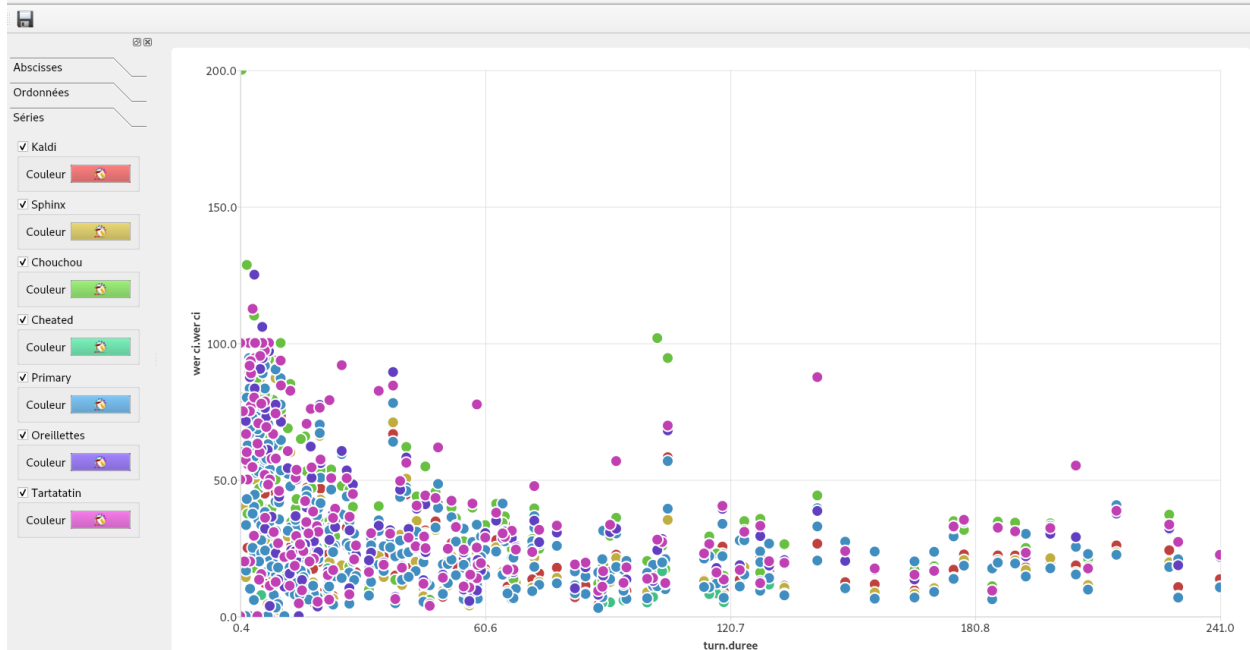


Fig. 11: Exemple d'un nuage de points

DET. Pour illustrer ces courbes, nous nous appuyons sur la Vue illustrée en *Vue d'une Dataframe d'évaluation de la vérification du locuteur*.

A partir de cette Vue, on peut alors générer la courbe DET : contrairement aux autres graphes, l'utilisateur n'a pas à indiquer les colonnes sur lesquelles travailler. La courbe résultat est présentée dans l'illustration *Exemple d'une courbe DET*.

Activités Evalomatic jeu. 23:53 full_det.dtf

Dataframe Graphiques Statistiques Visualisations

global.system	det.id	det.seuil	det.pmiss	det.pfa	det.cdet	det.cldr	seuil.EER	seuil.Cdet	seuil.Libre
h1b	1	-0,215346	1	1,81159e-05	0,020849	9,08245	1,18328	-0,215346	0,3
	2	-0,212604	1	3,62319e-05	0,0208646				
	3	-0,205709	1	5,43478e-05	0,0208802				
	4	-0,20396	1	7,24638e-05	0,0208958				
	5	-0,203541	1	9,05797e-05	0,0209115				
	6	-0,201033	1	0,000108696	0,0209271				
	7	-0,200462	1	0,000126812	0,0209427				
	8	-0,198447	1	0,000144928	0,0209583				
	9	-0,198134	1	0,000163043	0,020974				
	10	-0,196126	1	0,000181159	0,0209896				
	11	-0,19378	1	0,000199275	0,0210052				
	12	-0,187115	1	0,000217391	0,0210208				
	13	-0,186374	1	0,000235507	0,0210365				
	14	-0,184676	1	0,000253623	0,0210521				
	15	-0,184277	1	0,000271739	0,0210677				
	16	-0,183697	1	0,000289855	0,0210833				
	17	-0,181629	1	0,000307971	0,021099				
	18	-0,181406	1	0,000326087	0,0211146				
	19	-0,181057	1	0,000344203	0,0211302				
	20	-0,180968	1	0,000362319	0,0211458				
	21	-0,180946	1	0,000380435	0,0211615				

Fig. 12: Vue d'une Dataframe d'évaluation de la vérification du locuteur

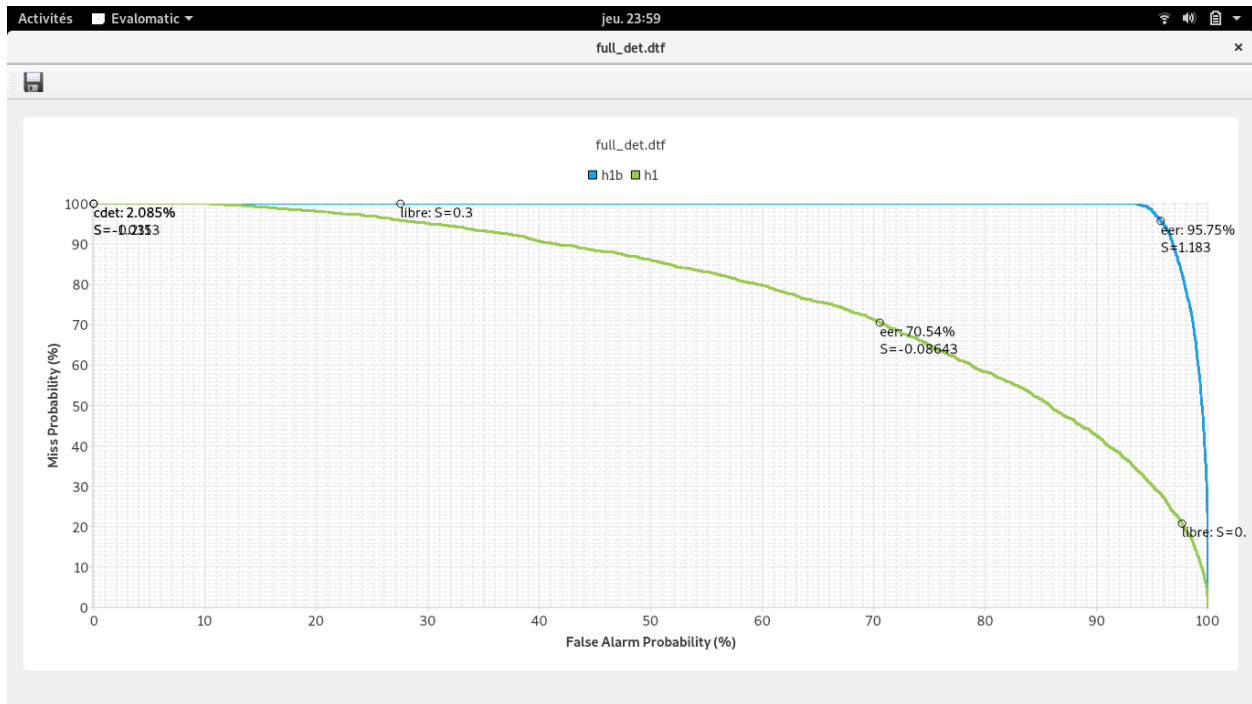
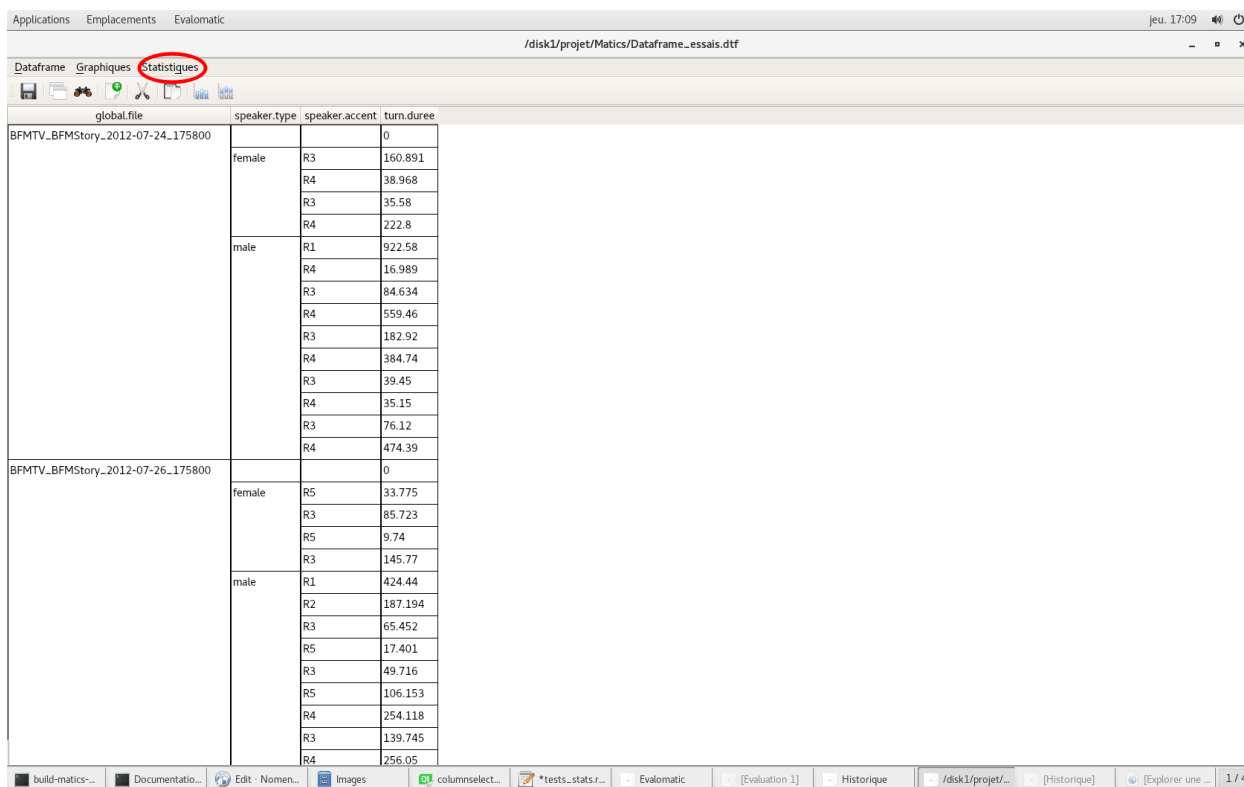


Fig. 13: Exemple d'une courbe DET

EFFECTUER DES STATISTIQUES À L'AIDE DE EVALOMATIC

Les 3 fonctionnalités statistiques suivantes sont disponibles sur EVALOMATIC dans le Menu Statistique comme l'illustre *Accéder aux fonctionnalités statistiques*:

- Statistiques descriptives
- Différences appariées
- Tests de corrélation entre colonnes



The screenshot shows the Evalomatic application window. The menu bar includes 'Applications', 'Emplacements', and 'Evalomatic'. The main window title is '/disk1/projet/Matics/Dataframe_essais.dtf'. The 'Statistiques' menu item is circled in red. Below the menu bar, there is a toolbar with icons for file operations and data analysis. The main area displays a data table with the following columns: 'global.file', 'speaker.type', 'speaker.accent', and 'turn.duree'. The table contains two main sections of data, each with 'female' and 'male' speaker types and various accents (R3, R4, R5).

global.file	speaker.type	speaker.accent	turn.duree
BFMTV_BFMStory...2012-07-24...175800	female	R3	160.891
		R4	38.968
		R3	35.58
		R4	222.8
	male	R1	922.58
		R4	16.989
		R3	84.634
		R4	559.46
		R3	182.92
		R4	384.74
		R3	39.45
		R4	35.15
		R3	76.12
		R4	474.39
BFMTV_BFMStory...2012-07-26...175800	female	R5	33.775
		R3	85.723
		R5	9.74
		R3	145.77
	male	R1	424.44
		R2	187.194
		R3	65.452
		R5	17.401
		R3	49.716
		R5	106.153
		R4	254.118
		R3	139.745
		R4	256.05

Fig. 1: Accéder aux fonctionnalités statistiques

Les résultats des opérations statistiques sont visibles dans des Vues générées à la suite du calcul. Ces Vues sont manipulables de la même façon que les autres Vues, comme décrit dans *Explorer une Vue d'une Dataframe*. Pour nos exemples, nous nous appuyons sur la Vue d'une Dataframe générée à partir d'une évaluation en transcription de la parole, sur laquelle on a trié les lignes en fonction des systèmes et de leurs versions. Cette vue est illustrée en *Vue d'une Dataframe d'une évaluation en transcription de la parole* :

global.system	global.system version	wer ci.entropie	wer ci.hyp correct	wer ci.hyp incorrect	wer ci.entropie max	wer ci.wer ci	wer ci.nce
PERCOL	Chouchou					31,2414	
	Oreillettes					27,5699	
	Primary					28,1345	
	Tartatatin					31,4089	
QCOMPERE	Primary	-24765,6	28439	7485	26523,8	28,2534	0,0662861
SODA	Cheated	-inf	31342	3466	16277,8	15,9147	-inf
	Kaldi	-inf	30495	4833	20342,1	19,0431	-inf
	Primary	-inf	31238	3513	16417,9	16,166	-inf
	Sphinx	-inf	31130	4372	19112,2	17,352	-inf

Fig. 2: Vue d'une Dataframe d'une évaluation en transcription de la parole

12.1 Statistiques descriptives

Les statistiques descriptives permettent d'obtenir les informations suivantes sur une colonne de type Valeur numérique :

- Moyenne
- Ecart type
- Asymétrie de la distribution
- Coefficient d'aplatissement
- Minimum
- 1er décile
- 1er quartile
- Médiane
- 3ème quartile
- 9ème décile
- Maximum
- Mode

Après avoir sélectionner dans le Menu la fonctionnalité "Statistiques Descriptives" l'interface *Formulaire pour obtenir les statistiques descriptives d'une colonne valeur* apparaît.

Fig. 3: Formulaire pour obtenir les statistiques descriptives d'une colonne valeur

Dataframe Graphiques Statistiques Visualisations					
factor.system 1	factor.system version 1	factor.system 2	factor.system version 2	différence.p-value	
SODA	Kaldi	SODA	Kaldi	1	
	Sphinx		Sphinx	2,32353e-05	
	Kaldi		Kaldi	2,32353e-05	
PERCOL	Chouchou	PERCOL	Chouchou	2,44547e-104	
	Kaldi		Kaldi	2,44547e-104	
	Chouchou		Chouchou	2,44547e-104	
SODA	Kaldi	SODA	Cheated	5,53331e-29	
	Cheated		Kaldi	5,53331e-29	
	Kaldi		Primary	2,02972e-75	
QCOMPERE	Primary	SODA	Kaldi	2,02972e-75	
	Kaldi		Primary	9,54607e-23	
	Primary		Kaldi	9,54607e-23	
PERCOL	Oreillettes	PERCOL	Oreillettes	1,25085e-76	
	Kaldi		Kaldi	1,25085e-76	
	Tartatatin		Tartatatin	1,01926e-96	
PERCOL	Tartatatin	SODA	Kaldi	1,01926e-96	
	Kaldi		Primary	8,28225e-79	
	Primary		Kaldi	8,28225e-79	
SODA	Sphinx	SODA	Sphinx	1	
	Chouchou		PERCOL	Chouchou	2,06665e-114
	Chouchou		SODA	Sphinx	2,06665e-114
SODA	Sphinx		Cheated	1,65986e-22	

Fig. 6: Test des différences appariées entre les systèmes et leurs versions en fonction du WER

12.3 Tests de corrélation entre colonnes

Le test de corrélation permet de déterminer si deux valeurs sont corrélées entre elles. EVALOMATIC permet de calculer la corrélation selon trois tests :

- Corrélation linéaire (Pearson)
- Corrélation de rang (Kendall)
- Corrélation de rang (Spearman)

La Vue résultat calculera pour les deux valeurs renseignées les corrélations choisies par l'utilisateur. Ces corrélations sont calculées à partir d'une granularité. La fenêtre de paramétrisation est illustrée dans l'image *Formulaire du test de corrélation entre deux valeurs*.

global	file	lang	system	system version	channel															
section	id	start time	end time	type																
speaker	name	type	dialect	accent																
turn	id	start time	end time	raw text	pronounced text	normalized text	duree													
wer ci	id	ref text	hyp start time	hyp end time	hyp text	correct	subst	insert	delete	delete opt	correct cut	delete cut	ref cnt	hyp cnt	entropie	hyp correct	hyp incorrect	entropie max	wer ci	nce

Fig. 7: Formulaire du test de corrélation entre deux valeurs

Dans notre exemple, nous souhaitons observer la corrélation entre la durée du tour de parole et le WER. Nous ajoutons donc une colonne calculée durée avec comme granularité le tour de parole. Nous choisissons par ailleurs comme granularité pour le calcul de la corrélation le tour de parole. La Vue résultat est illustrée dans l'image *Test de corrélation entre le WER et la durée du tour de parole*.

série.1	série.2	corrélation.pearson r	corrélation.kendall	corrélation.spearman
wer ci.wer ci	turn.duree	-0,39268	-0,208402	-0,307669

Fig. 8: Test de corrélation entre le WER et la durée du tour de parole

QUITTER EVALOMATIC

Pour quitter EVALOMATIC, il suffit de cliquer sur la croix en haut droite du Tableau de bord ou de sélectionner “Quitter” dans le menu déroulant.

BIBLIOGRAPHY

[ElMethni_2017] El Methni 2017 “Analyse de la variance (Niveau I)”, Mathématiques Appliquées & Sciences Humaines et Sociales, Université Pierre Mendès France,