

Préparation d'un corpus de focus group pour une utilisation dans Iramuteq (méthode decrowded)

Ce mode opératoire permet de passer d'un corpus de retranscription de focus group obtenu avec Sonal à un corpus au format txt utilisable dans Iramuteq (logiciel libre développé par P. Ratinaud, Lerass¹). Les logiciels utilisés sont libres : Sonal (retranscription), Libre office calc (écriture des variables), Notepad++ (modifications des intitulés) et Iramuteq (lexicométrie).

La méthode peut cependant être transférée à d'autres logiciels. Elle correspond à une approche de partition des focus groups dite individualisée ou « decrowded » (Peyrat-Guillard et al. 2014).

Les étapes sont les suivantes :

0. Export des textes des focus groups au format txt
1. Suppression des questions (expressions régulières dans Notepad++) et des lignes vides
2. Création d'un tableau de correspondance des variables (Libre Office Calc)
3. Modification des entêtes des lignes (remplacer dans Notepad++)
4. Finalisation de la préparation du corpus : vérification, points de synchronisation

Ces opérations sont répétées pour chaque focus group. Etant donné les possibilités d'erreurs et de problèmes techniques, **il est fortement conseillé de travailler groupe par groupe** plutôt que de modifier tous les groupes ensemble.

Corpus de départ au format SONAL :

```
1  [>Question?]: je vous rappelle le principe, ça s'appelle un f
2  [>Question?]: première question, sur les 5 sites que vous av
3  [>Yanis]: il y a des choses marquantes, il y a un site où or
4  [>Maurice]: Je suis d'accord avec toi. Le site de la banque I
5  [>Question?]: ça oriente votre perception.
6  [>Yanis]: la recherche d'information. Je vais plutôt cherche
```

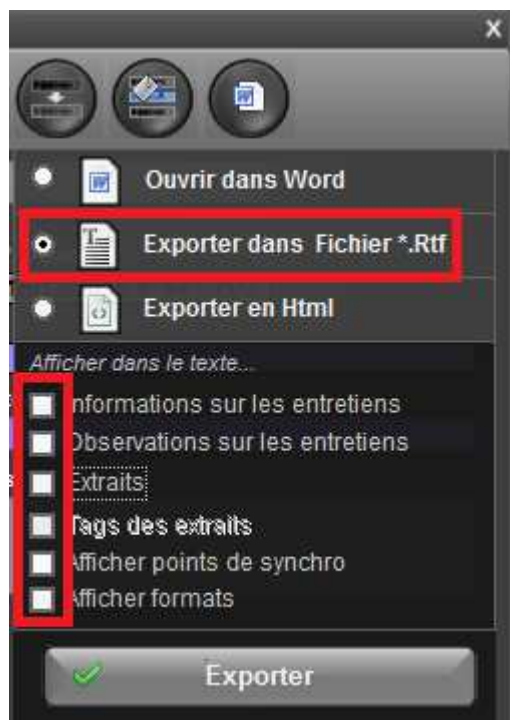
Corpus final utilisable dans IRAMUTEQ :

```
1
2  **** *num_1 *formation_UT1 *groupe_1 *sexe_M *participation_TA *initial_STS
3  il y a des choses marquantes, il y a un site où on n'a pas beaucoup d'informati
4
5  **** *num_6 *formation_UT1 *groupe_1 *sexe_M *participation_TA *initial_IUT
6  Je suis d'accord avec toi. Le site de la banque Dupuy est vraiment très pauvre.
7
8  **** *num_1 *formation_UT1 *groupe_1 *sexe_M *participation_TA *initial_STS
9  la recherche d'information. Je vais plutôt chercher là où je vais trouver pour
```

¹ Téléchargement et documentation : <http://www.iramuteq.org/>

0. Export des textes de Sonal en txt :

Sélectionner le groupe dans votre liste et exporter dans Fichier *.Rtf en décochant toutes les données :



Après avoir supprimé l'entête généré automatiquement par SONAL :

```
-----  
Entretien exporté depuis Sonal (v.2.0) le 08/07/2016 à 15:17:15  
-----  
G13IL.avi  
56 extrait(s)  
-----
```

Enregistrez au format txt, UTF8. Cet enregistrement peut-être fait avec Libre Office par exemple.

Utiliser Sonal ou un autre logiciel de retranscription permet de standardiser les noms des participants ce qui facilite beaucoup ensuite la mise en forme du corpus.

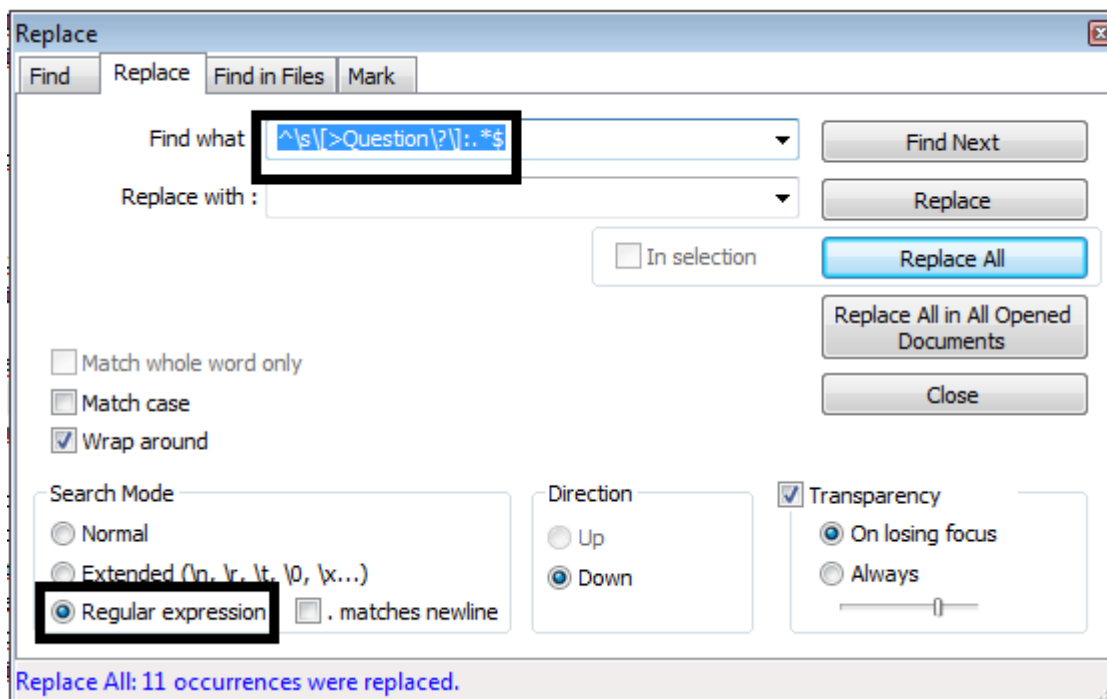
1. Supprimer les questions (expressions régulières dans Notepad++) et les lignes vides

Ouvrir le fichier obtenu dans Notepad++ puis sélectionnez 'Rechercher' puis 'Remplacer' (Search Replace).

Pour supprimer toutes les questions, nous allons utiliser une expression régulière :

```
.*\[>Question\?\]:.*\r?\n
```

à placer de cette façon dans la fenêtre :



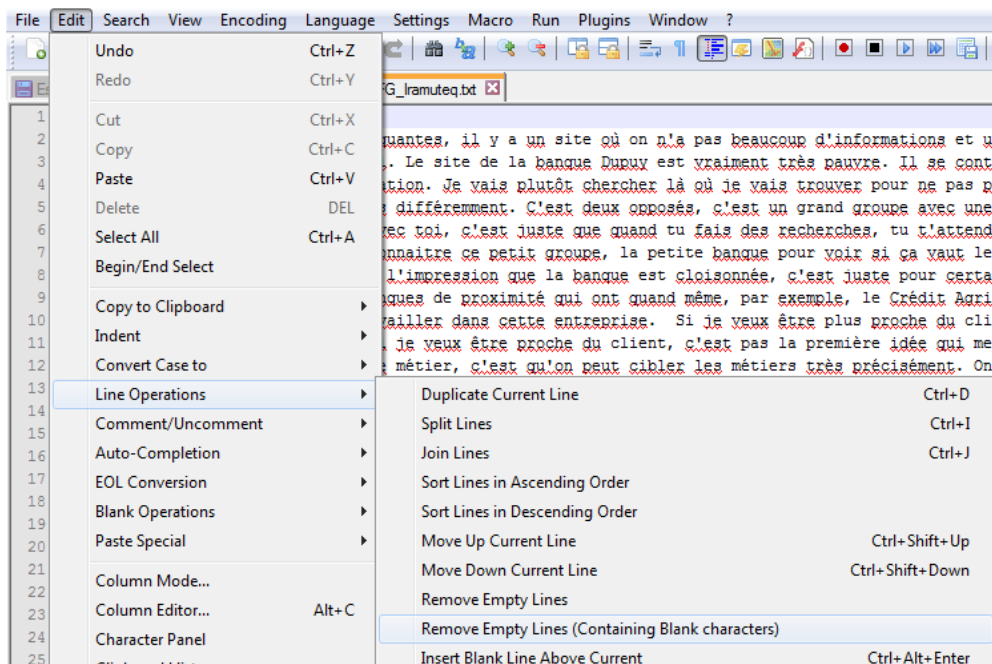
Après avoir cliqué sur « Replace All », Notepad++ vous indique le nombre de remplacements.

Normalement, les lignes contenant [`>Question ?`] ont été supprimées.

Il est recommandé de supprimer les lignes vides. Soit par une nouvelle expression régulière :

```
^\s*\t]*$\r?\n
```

Ou utiliser une fonction dans Notepad++ :



2. Créer un tableau de correspondance des variables (Libre Office Calc)

Le principe central de la préparation du corpus est de modifier les noms des participants pour les remplacer par une ligne de déclaration de texte avec variables d'Iramuteq².

Par exemple, nous allons réaliser les modifications suivantes :

Au départ DANS SONAL :

[>Valentin]:

Au final DANS IRAMUTEQ :

```
**** *num_2 *formation_UT2 *groupe_1 *sexe_M *participation_PA *initial_IUT
```

Pour générer automatiquement ces lignes, nous allons utiliser un tableur.

Le 'tableau de correspondance' peut se présenter de la façon suivante :

	A	B	C	D	E	F	G	H	I
1	SONAL							Fiche	Iramuteq **** *num_ *formation_ *groupe_ *sexe_ *participation_ *initial_
2	Num	Prénom	Code Sonal	Formation	Groupe	Sexe	Participation	Initial	Codage
3		Dorian	[>Dorian]:	3IL		M	TA	STS	An**** *num 1 *formation 3IL *groupe 1 *sexe M *participation TA *initial STS\n

En I3, une formule permet de concaténer les informations pour obtenir la ligne de déclaration dans Iramuteq :

```
=CONCATENER (" \n****"; " "; "*num_"; A3; " "; "*formation_"; D3; " "; "*groupe_"; E3; " "; "*sexe_"; F3; " "; "*participation_"; G3; " "; "*initial_"; H3; "\n")
```

² Une autre possibilité est d'ajouter des thématiques, voir le forum d'Iramuteq : <https://sourceforge.net/p/iramuteq/discussion/1068066/thread/2511ebfc/>

Tableau de correspondance, exemple :

SONAL							Fiche	Iramuteq **** *num_ *formation_ *groupe_ *sexe_ *participation_ *initial_
Num	Prénom	Code Sonal	Formation	Groupe	Sexe	Participation	Initial	Codage
1	A	[>A]:	ut1	1	M	TA	STS	\n**** *num_1 *formation_ut1 *groupe_1 *sexe_M *participation_TA *initial_STS\n
2	B	[>B]:	ut1	1	M	PA	IUT	\n**** *num_2 *formation_ut1 *groupe_1 *sexe_M *participation_PA *initial_IUT\n
3	C	[>C]:	ut1	1	M	PA	STS	\n**** *num_3 *formation_ut1 *groupe_1 *sexe_M *participation_PA *initial_STS\n
4	D	[>D]:	ut1	1	M	TA	IUT	\n**** *num_4 *formation_ut1 *groupe_1 *sexe_M *participation_TA *initial_IUT\n
5	E	[>E]:	ut1	1	M	PA	STS	\n**** *num_5 *formation_ut1 *groupe_1 *sexe_M *participation_PA *initial_STS\n
6	F	[>F]:	ut1	1	M	TA	IUT	\n**** *num_6 *formation_ut1 *groupe_1 *sexe_M *participation_TA *initial_IUT\n
7	G	[>G]:	ut1	2	M	TA	IUT	\n**** *num_7 *formation_ut1 *groupe_2 *sexe_M *participation_TA *initial_IUT\n
8	H	[>H]:	ut1	2	M	PA	IUT	\n**** *num_8 *formation_ut1 *groupe_2 *sexe_M *participation_PA *initial_IUT\n
9	I	[>I]:	ut1	2	M	A	STS	\n**** *num_9 *formation_ut1 *groupe_2 *sexe_M *participation_A *initial_STS\n
10	J	[>J]:	ut1	2	M	TA	STS	\n**** *num_10 *formation_ut1 *groupe_2 *sexe_M *participation_TA *initial_STS\n

Les prénoms sont dans cet exemple anonymés.

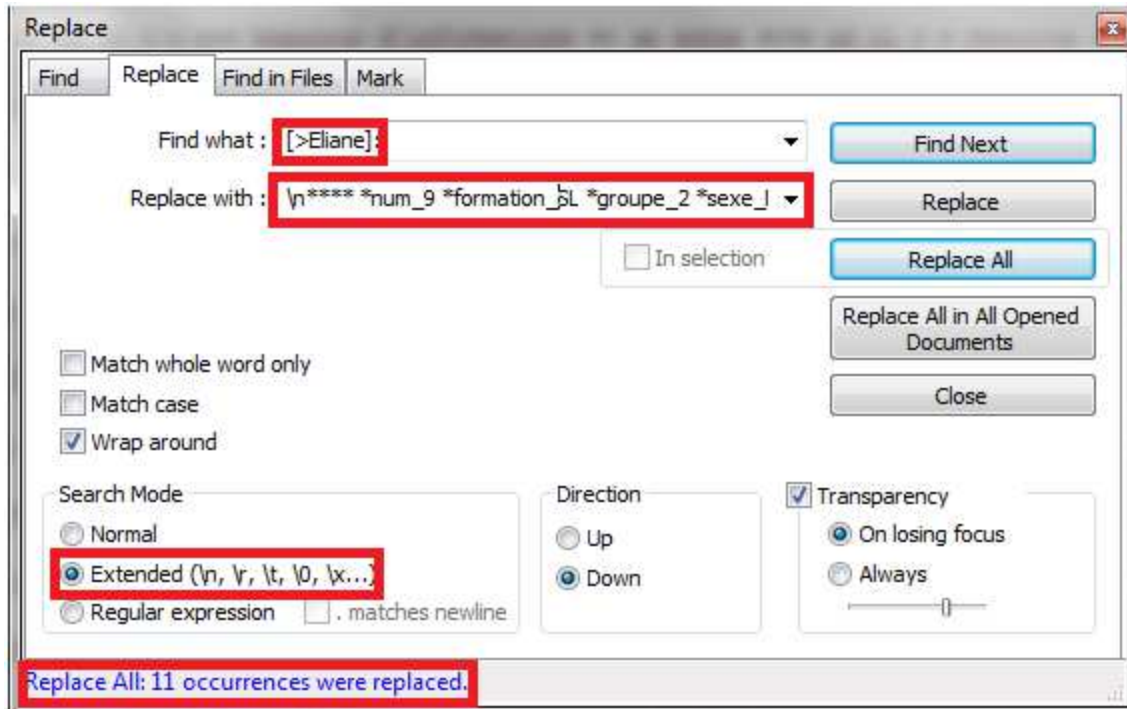
Cette procédure d'automatisation avec un tableur évite des erreurs de codage et accélère la préparation du corpus. Les colonnes SONAL sont des informations trouvées dans SONAL qui peuvent être complétées par d'autres données ('Fiche' dans cet exemple).

L'expression \n permet de sauter automatiquement une ligne avant et après la ligne de déclaration pour éviter des problèmes de repérage dans Iramuteq. Les variables sont à adapter selon le contexte de la recherche. Dans cet exemple, le nom du participant est anonymé et remplacé par un numéro.

Ce tableau permet aussi de faire correspondre ce numéro avec le prénom d'origine.

3. Modifier les entêtes des lignes (remplacer dans Notepad++)

Une fois ce tableau obtenu, il est possible de remplacer automatiquement les entêtes des lignes dans Notepad++ :



Les informations dans les zones « Find what » et « Replace with » sont dans le fichier du tableur fait à l'étape précédente (copier/coller du tableur vers Notepad++). Ne pas oublier de sélectionner « Extended » pour que le saut de ligne fonctionne.

4. Finaliser la préparation du corpus : vérification, points de synchronisation

Il est préférable de vérifier qu'il ne reste pas d'entêtes au format Sonal : faire une recherche sur les caractères [> par exemple. Puis vérifier qu'il ne reste pas de points de synchronisation générés par Sonal comme [2121, 6]. Faire une recherche sur le caractère] par exemple.

Une fois votre corpus relu, vous pouvez recommencer la même procédure avec un autre focus group. Ensuite, il vous faudra fusionner les différents fichiers³ pour obtenir le corpus analysable dans Iramuteq.

D. Pélissier

Idetcom, Lerass, université Toulouse, IUT de Rodez

³ Soit par copier/coller ; avec des commandes DOS : type *.txt >> fichier.res puis renommer le fichier.